Fundamentals of ocean climate models

Stephen M. Griffies

Foreword by Trevor J. McDougall

FUNDAMENTALS OF OCEAN CLIMATE MODELS

Stephen M. Griffies
NOAA/Geophysical Fluid Dynamics Laboratory

Version prepared on February 11, 2003
Available online at www.gfdl.noaa.gov

This book was prepared using LaTeX as described by Lamport (1994) and Goosens et al. (1994).

FOREWORD

Ocean models embody a wide range of oceanic knowledge. There are issues concerned with the numerical representation of processes and of conservation equations on different finite spatial grids. There is the need to respect the smallness of diapycnal mixing and particularly to not allow horizontal diffusion to cause false diapycnal density fluxes. Do the model variables represent the Eulerian-mean quantities or are they the result of some other type of averaging process? Is the Boussinesq approximation really made in a model or is this simply a matter of interpretation? Should the diffusion tensor be symmetric, and if not, what is the physical justification for the skew component of diffusion? Where this book really shines is that rather than simply presenting workable recipes to each of these issues (and many more), the underlying physics is expertly described, usually first in differential equations and then in the discrete representation on the numerical grid. Because of this clarity and attention to detail, the book will appeal to not only ocean modellers but to a much broader spectrum of oceanographers.

As a branch of fluid dynamics, distinguishing features of the ocean are its rather strong stratification and the smallness of its diapycnal mixing processes. For climate purposes the seemingly small diapycnal mixing processes are very important yet their faithful representation in models has been a formidable obstacle. While we have known the importance of diapycnal processes in controlling the thermohaline ocean circulation for some decades, it is only in the past decade that we have learnt how to control the amount of diapycnal mixing in z-coordinate ocean models, and then only at coarse resolution. This goal has been achieved by clear thinking about the conservation of properties in ocean models, particularly the way in which diffusion is imposed. With this book we now have the relevant averaged model equations derived carefully from first principles and the subtleties associated with interpreting averaged quantities should no longer need to be glossed over.

The present book can be described as providing comprehensive treatment of the following subjects (i) deriving the oceans conservation equations from first principles, (ii) carefully considering the issues that arise when these equations are averaged, (iii) describing the many numerical procedures that are used to integrate the averaged equations forward in time, and (iv) providing the tensorial underpinning so that the equations can be transformed consistently onto different grids on the spherical earth. This book clearly fills a void in the oceanographic literature. The usual development skips the first two topics and treats the averaged equations as given (which then begs the question of what the model variables might represent), while the subject of tensor analysis is familiar to too few oceanographers especially given the range of grids that are now in common use in ocean models.

Two key advances in ocean modelling over the past decade are the rotation of the (symmetric) diffusion tensor to be aligned with respect to the local neutrally buoyant directions, and the discovery and implementation of the so-called Gent-McWilliams mixing scheme for mesoscale eddies. These two topics are expertly described in this book and together they occupy one third of its pages. This is entirely appropriate since both these advances have been crucial for controlling the false diapycnal mixing that otherwise occurs across sloping density surfaces in low-

resolution z-coordinate models. Both advances have been hard-won by oceanographers. The solution of the first involved the subtle thermobaric nature of the equation of state of seawater, while the second relies on the action of mesoscale eddies to release potential energy, with the mixing scheme being best viewed as either an extra quasi-Stokes advection or alternatively as an additional amount of diffusion which happens to be skew diffusion. These issues are subtle and are not covered in existing oceanographic texts, and yet they have been at the root of the biggest improvements in ocean climate modelling in more than twenty years.

This book truly lives up to its title. By delving into the physical basis for the choices made in present ocean models, the book does indeed establish the fundamental basis of ocean models. Where the issues are not yet agreed by the oceanographic community, the book takes on the character of a review of these research issues. The book will prove invaluable to ocean modellers and to those concerned with how fundamental ocean physics is represented in ocean models.

Trevor J. McDougall
CSIRO Marine Research, Hobart, Australia
Dec 18, 2002

PREFACE

The purpose of this book is to contribute to a rationalization of the physical, mathematical, and numerical foundations of computer models used to understand and predict the global ocean circulation. For this purpose, the presentation is geared towards those aiming to understand the physical meaning of the equations of ocean models and the methods used to solve these equations. Much of the formulation is general, and so applicable to any primitive equation ocean model. However, for the purpose of presenting concrete examples, attention is focused on z-coordinate models, which represent the most common model class used to simulate the global ocean climate system. Beyond a solid understanding of classical mechanics and vector calculus, requirements of previous knowledge are minimized by approaching the formulations from basic principles. Nonetheless, experience with ocean modeling and the ideas of geophysical fluid dynamics (GFD) will aid the reader.

Various topics in ocean model development are complex, technical, and a bit esoteric. This book aims to reduce the mystery by keeping presentations reasonably self-contained, thorough, and starting from a fundamental perspective. Additionally, some level of mathematical sophistication, fully explained in the text, has proven useful to cleanly present physical and mathematical results in their necessary generality. It is hoped that this book provides the reader with a digestable presentation of certain key aspects of ocean climate models. In turn, the reader is encouraged to suggest improvements of any sort, or to provide corrections to the mistakes likely contained here.

Limitations allow for only a partial fulfillment of this book's goals. For example, notable topics are missing or only partially considered, and discussion is slanted towards z-coordinate models. Additionally, this book is not a text on oceanography, geophysical fluid dynamics, or numerical methods. Hence, there are no solutions described, either analytical or numerical, little mention is made of ocean observations, and some knowledge of basic numerical methods is required. For more complete treatments of these subjects, the reader is referred to Pickard and Emery (1990) and Tomczak and Godfrey (1994) for observational descriptions of the ocean, Gill (1982), Apel (1987), Pedlosky (1987), Cushman-Roisin (1994), Pedlosky (1996), and Salmon (1998) for foundations in geophysical fluid dynamics and theoretical physical oceanography, Trenberth (1992), Peixoto and Oort (1992), Chassignet and Verron (1998), Haidvogel and Beckmann (1999), Kantha and Clayson (2000a), and Kantha and Clayson (2000b) for discussions of ocean climate physics and modeling, and Haltiner and Williams (1980), O'Brien (1986), Bryan (1989), and Durran (1999) for numerical methods of use for geophysical fluid dynamics. Additionally, the recent book edited by Siedler et al. (2001) is noteworthy for its many lucid contributions affording the reader a glimpse at the state-of-the-art in global ocean observations, theory, and modeling.

The present book aims to fill a niche in the literature whereby a (mostly) first principles presentation of ocean climate models is given, with discussions extending from fundamental ocean fluid mechanics to detailed parameterization and discretization issues. That is, this book presents an exposition of certain features of ocean models, and we distinguished such from a book on ocean *modeling*, where focus is given towards simulations and their relation to the real ocean. Restricted

as this scope may appear, there are numerous topics missing that reflect on the author's limitations rather than an absence of their importance. It is hoped that this book nonetheless serves as a useful complement to other presentations, and that it provides a stepping stone for future expositions addressing shortcomings of the present document.

ORGANIZATION

This book is organized into parts, each of which have multiple chapters. It is assumed that the book will *not* be read cover-to-cover. Consequently, the contents of one part are written so that they can largely be read independently of other parts. Chapters likewise aim for such independence, though less so. Although useful to the reader aiming to pick up the book at any point in its development, this *modular* approach has not been taken to its extreme, nor has it been used to justify poor cross-referencing. Instead, this organization comes at the price of needing to summarize key results from allied parts/chapters, with cross-referencing maintained. It is hoped that this approach enhances the book's readability, accessibility, and utility as both a text and a reference.

Chapter 1 starts the book with an introduction to ocean climate models. It is here that we provide some understanding for why ocean models are of primary importance for climate science. This chapter also helps to motivate the more fundamental, and by necessity more mathematical, development presented in subsequent chapters.

Part I begins the main part of the text by focusing on the equations describing stratified fluid dynamics on a rotating sphere. Chapter 2 introduces some physical ideas underlying a continuum description of the ocean fluid. It also highlights common approximations made in ocean climate models. Chapter 3 derives the hydrodynamical equations of the ocean fluid. Chapter 4 presents aspects of ocean energetics and thermo-hydrodynamics. Chapter 5 presents some mathematical results applicable to a generalized vertical coordinate description of the ocean.

The equations described in Part I allow for a mathematical description of a particular realization of the ocean fluid given essentially an infinite level of knowledge about its state. In practice, we are always faced with less than infinite information in both space and time. Hence, it is necessary to derive alternative equations via a coarse-graining procedure. That is, we need to determine approximate, averaged, or mean-field equations for the coarsened ocean fluid.

Part II presents two chapters illustrating two averaged descriptions. Chapter 6 accepts the fact that the precise state of the turbulent ocean is not physically measurable with finite instruments, thus necessitating a statistical approach. In particular, this chapter provides a conceptual and mathematical interpretation of the fields discretized by a z-coordinate model. It is argued that these equations represent an ensemble mean of individual realizations of the ocean fluid, where the sub-grid-scales are associated with the small-scale dianeutral mixing arising from three-dimensional turbulence. In Chapter 7, we focus on the transport of tracers by mesoscale eddies. This transport is the dominant means whereby tracers are stirred within the ocean interior, with tracer gradients ultimately dissipated by small scale mixing processes. The discussion in this chapter motivates a particular interpretation of model variables for those cases where the simulations do not explicitly

resolve mesoscale eddies. Furthermore, in this context we see how the choice of vertical coordinate strongly affects the simplicity of the mathematical description of adiabatic transport. This part of the book finishes with Chapter 8, which presents a general discussion of subgrid-scale processes, with a focus on those contributing to dianeutral transport. Such transport is spread non-uniformly throughout the World Ocean, with most occurring in critical regions mostly near ocean boundaries. Dianeutral transport plays a critical role in affecting the ocean's stratification and vertical distribution of tracers, and this chapter provides an introduction to some of the issues.

Part III presents two chapters on the semi-discrete equations of a z-coordinate ocean model. Semi-discrete refers to equations that have some parts discretized, some parts continuous. The aim here is to describe in general terms the equations of a discrete z-coordinate ocean model, and to outline methods used to time step these equations. In particular, Chapter 9 discusses the semi-discrete version of the mass and tracer budgets, and highlights the importance of maintaining compatibility between these budgets. Chapter 10 provides a semi-discrete version of the momentum equations and methods used to march these equations forward in time.

Chapter 7 introduces a framework for thinking about transport in the ocean interior. The parameterization of this transport in ocean models constitutes the *neutral physics* part of the models. Notably, the integrity of z-model simulations has greatly improved since modelers started employing such schemes. Part IV of this book aims to rationalize the neutral physics schemes commonly used in z-models. We discuss some physical and mathematical aspects of neutral physics in Chapters 11 and 12. Chapter 13 focuses on issues that must be addressed to implement neutral physics operators in climate models, where in particular what is done next to boundaries proves crucial for both numerical stability and physical integrity. Chapter 14 provides a rational, with details, for discretizing neutral physics in z-models.

Horizontal friction plays an important role in determining the integrity of ocean model simulations. Additionally, the state-of-the-art in ocean climate model horizontal friction remains at an engineering stage, with modelers encouraged to use horizontal friction largely for purposes of maintaining numerical stability. The more scientifically satisfying situation is to consider horizontal friction as a parameterization of unresolved physical processes, such as the case for the kinetic theory of gases. However, there is no complete theory for how to close the momentum equation at scales relevant to ocean climate studies. Hence, horizontal friction used in ocean climate models is *ad hoc*. This situation would be of little concern were it not for the importance of friction for the simulations, such as in boundary and/or equatorial regions (i.e., regions where geostrophy breaks down). Hence, it is crucial that modelers understand the rationale underlying some of the commonly used schemes, if only to motivate approaches that lead to improvements.

Given the importance to simulations of horizontal friction, Part V is offered as a tutorial for these issues. Chapter 15 focuses on the friction force in the continuum, with some of the mathematics developed in Part VI of some utility. Chapter 16 then presents some considerations for how to choose a viscosity to set the friction force's magnitude. The issues discussed in these chapters are generic, whereas Chapter 17 finishes this part of the book with a discretization of the horizontal friction operator appropriate for B-grid models.

Tensor analysis is not commonly taught in oceanographic or atmospheric sci-

ence curricula. However, it is standard in physics and engineering studies. An understanding of general tensor analysis is necessary when formulating the equations of rotating fluid dynamics on a sphere. Hence, although of a relatively specialized nature, tensor analysis is needed by those actually building ocean climate models. Its logical place is at the start of this book. However, it is placed in Part VI to allow the less specialized reader ample opportunity to skip this material.

Two chapters are included that aim to develop tensor analysis from first principles, with examples drawn from ocean fluid dynamics. Chapter 18 introduces the basic notions, whereas Chapter 19 expands on these notions while deriving some key results from calculus on curved manifolds. The presentation is aimed at the reader who has a solid foundation in undergraduate vector calculus. No exposure to Cartesian or general tensors is assumed. The pace may be a bit slow for those who have some exposure to tensor analysis, yet it is hoped that both experienced and inexperienced readers will find something of use here. For those not having experience with these matters, some attempt has been made to allow the reader to only briefly read or to skip this material. We do so by summarizing salient points as needed in the appropriate places throughout the main text, with Section 19.12 offered as a summary of certain salient points.

Concerning ocean climate model development

Ocean climate models are not conceived one year, to be then publicly released and supported the next. Instead, they take years, indeed decades, of creative passion and obsession from many scientists and engineers. It is only via patience and persistence that an ocean model is successfully taken from its vision phase, to its prototype phase, onto its public release phase, and then to its revision phase where one begins to fully appreciate the wisdom, or lack, of decisions made years earlier. In short, the construction of an ocean climate model requires a marriage of research with development, with each phase requiring an unpredictable amount of time to debate and explore various research avenues. Allowing adequate time requires dedication and support from funding agencies and managers. Absent such, ocean climate model development is handicapped, the integrity of the simulations compromised, and the depth of the understanding shallow.

In general, it is critical to maintain a research and development environment that fosters a healthy balance between addressing the exigencies of the moment with visions extending out decades. Only by doing so will climate models and climate modeling reach a new level of respect as hard-sciences. Short of this goal, the truly massive problems of climate and climate change will remain out of the reach of hard-science. It is the profound responsibility of leaders in the international climate science field to foster this atmosphere, especially now that the questions of climate change are at the forefront of society's concern.

Stephen Griffies
Princeton, New Jersey, USA
February 11, 2003

Acknowledgments

During my ten years in climate research, the Geophysical Fluid Dynamics Laboratory (GFDL) has provided a friendly and exciting environment for studying the earth's climate. GFDL has nurtured honest and rich scientific relationships through the active exchange of knowledge, experience, and ideas, and it encourages scientists to attack critical research problems that may take decades to resolve. These are perhaps the greatest legacies of GFDL's founder Joseph Smagorinsky, whose uncompromising vision for climate science from the 1950's and 1960's is still unfolding in the 21st century.

This book draws its inspiration from the many creative researchers within the international oceanography and climate science communities. At an increasing rate, these communities share a sense that scientifically based models of sound mathematical and numerical integrity are effective tools to further understanding of the ocean and the earth's climate. I also owe a debt to the mathematical physicists encountered during my graduate education. Their lessons regarding how clear and general mathematical statements lead to deep physical understanding are hopefully reflected in these pages. Finally, this book could not have been written without the love, patience, and support of my wife, Adi. It is to her that I dedicate this work. Muchas gracías por tu amor, cariño. Sat Nam.

Stephen Griffies
Princeton, New Jersey, USA
February 11, 2003

CONTENTS

1 OCEAN CLIMATE MODELS **17**
1.1 Ocean models as tools for ocean science . . . 17
1.2 Ocean climate models . . . 18
1.3 Climate change . . . 19

I FUNDAMENTAL OCEAN EQUATIONS **21**

2 BASICS OF OCEAN HYDRODYNAMICS **25**
2.1 Some fundamental ocean processes . . . 25
2.2 The continuum hypothesis . . . 27
2.3 Kinematics of fluid motion . . . 28
2.4 Kinematical and dynamical approximations . . . 33
2.5 Averaging over scales and realizations . . . 38
2.6 Numerical discretization . . . 39

3 OCEAN HYDRODYNAMICS **41**
3.1 Orthogonal coordinates . . . 42
3.2 Geometry of the Traditional Approximation . . . 44
3.3 Volume and mass conserving kinematics . . . 45
3.4 Principles of continuum dynamics . . . 53
3.5 Dynamics of fluid parcels . . . 56
3.6 Hydrostatic pressure . . . 64
3.7 Vorticity and potential vorticity . . . 66
3.8 Dynamics of fluid columns . . . 71
3.9 Particle dynamics . . . 73
3.10 Symmetry and conservation laws . . . 79

4 OCEAN THERMO-HYDRODYNAMICS **83**
4.1 Introduction . . . 84
4.2 General types of ocean tracers . . . 84
4.3 Basic equilibrium thermodynamics . . . 87
4.4 Energy of a fluid parcel . . . 92
4.5 Global energy balances . . . 98
4.6 Basic non-equilibrium thermodynamics . . . 103
4.7 Thermodynamical tracers . . . 105

4.8 Ocean density 108

5 GENERALIZED VERTICAL COORDINATES 113
5.1 Introduction 114
5.2 Concerning the choice of vertical coordinate 114
5.3 Generalized surfaces 119
5.4 Local orthonormal coordinates 120
5.5 Generalized vertical coordinates 122
5.6 Metrics with generalized vertical coordinates 127
5.7 The dia-surface velocity component 128
5.8 Conservation of mass and volume 131
5.9 Primitive equations 134

II AVERAGED OCEAN EQUATIONS 139

6 ENSEMBLE-MEAN OCEAN EQUATIONS 143
6.1 Introduction 144
6.2 Unaveraged primitive equations 146
6.3 Averaged kinematics of a fluid parcel 148
6.4 Averaged kinematics over finite domains 149
6.5 Averaged tracer in a parcel 153
6.6 Averaged momentum budget 156
6.7 Summary of the averaged equations 158
6.8 Mapping to ocean model variables 159
6.9 Discretization basics 162

7 KINEMATICS OF AN ISENTROPIC ENSEMBLE 163
7.1 The mesoscale eddy parameterization problem 164
7.2 Advection and skewsion 166
7.3 Volume conservation in an isentropic ensemble 168
7.4 Ensemble mean tracer equation 177
7.5 Quasi-Stokes transport in z-models 180

8 CONCERNING UNRESOLVED PHYSICS 187
8.1 Represented dynamics and parameterized physics 188
8.2 Lateral (neutral) and vertical processes 190
8.3 Basic mechanisms for dianeutral transport 191
8.4 Dianeutral transport in models 193
8.5 Numerically induced spurious dianeutral transport 198

III SEMI-DISCRETE Z-EQUATIONS 199

9 SEMI-DISCRETE MASS AND TRACER BUDGETS 203
9.1 Summary of the continuous model equations 203
9.2 Tracer and mass/volume compatibility 204
9.3 Mass budget for a grid cell 205
9.4 Mass budget for a discrete fluid column 207

9.5 Tracer budget for a grid cell . . . 209
9.6 Fluxes for turbulence mixed layer schemes . . . 213
9.7 Flux plus restore boundary conditions . . . 214

10 MOMENTUM EQUATION SOLUTION METHODS 217
10.1 Summary of the continuous model equations . . . 218
10.2 Momentum budget for a grid cell . . . 219
10.3 Strategy for time stepping . . . 225
10.4 Basics of the time stepping algorithm . . . 228
10.5 Time stepping the barotropic system . . . 230
10.6 Updating the baroclinic velocity . . . 232
10.7 Summarizing a model time step . . . 233
10.8 Rigid lid streamfunction method . . . 235

IV Z-MODEL NEUTRAL PHYSICS 237

11 NEUTRAL TRANSPORT: PART I 241
11.1 The general utility of neutral physics operators . . . 241
11.2 Summary of the mass and tracer budgets . . . 243
11.3 Compatibility in the mean field budgets . . . 244
11.4 The SGS transport tensor . . . 244
11.5 Advection and skewsion . . . 246
11.6 Neutral tracer fluxes . . . 247

12 NEUTRAL TRANSPORT: PART II 251
12.1 Neutral diffusion . . . 252
12.2 Gent-McWilliams stirring . . . 261
12.3 A summary of the neutral physics fluxes . . . 264
12.4 Flow dependent diffusivities . . . 265
12.5 Biharmonic operators . . . 272

13 NEUTRAL PHYSICS NEAR BOUNDARIES 283
13.1 Linear numerical stability for neutral diffusion . . . 284
13.2 Linear stability for GM stirring . . . 287
13.3 Neutral physics near boundaries . . . 288

14 NEUTRAL PHYSICS DISCRETIZATION 299
14.1 Foundations for discrete neutral physics . . . 300
14.2 Introduction to the discretization . . . 305
14.3 A one-dimensional warm-up . . . 307
14.4 Elements of the discrete functional . . . 308
14.5 Triads and some notation . . . 314
14.6 The discrete diffusion operator . . . 316
14.7 Diffusive flux components . . . 320
14.8 Some further issues of numerical implementation . . . 323

V MODEL HORIZONTAL FRICTION 327

15 HORIZONTAL FRICTION 331
15.1 Boussinesq and non-Boussinesq friction . . . 332
15.2 Introduction and general framework . . . 332
15.3 Properties of the stress tensor . . . 333
15.4 Properties of the viscosity tensor . . . 339
15.5 Transverse isotropy . . . 341
15.6 Transverse anisotropy . . . 344
15.7 Friction in generalized orthogonal coordinates . . . 347
15.8 Dissipation functional . . . 350
15.9 Biharmonic friction . . . 354
15.10 Some mathematical details . . . 356

16 HORIZONTAL VISCOSITY 359
16.1 Stability and resolution considerations . . . 359
16.2 Smagorinsky viscosity . . . 367
16.3 Background viscosity . . . 371
16.4 Viscosities for anisotropic friction . . . 372

17 DISCRETE B-GRID HORIZONTAL FRICTION 375
17.1 Comments on the notation . . . 376
17.2 Summary of the various formulations . . . 376
17.3 Horizontal friction discretization . . . 377
17.4 Laplacian plus metric form of isotropic friction . . . 387

VI TENSOR ANALYSIS 391

18 TENSOR ANALYSIS BASICS 395
18.1 Introduction . . . 396
18.2 Some practical motivation . . . 397
18.3 Coordinates and vectors . . . 398
18.4 The metric and coordinate transformations . . . 401
18.5 Transformations of a vector . . . 403
18.6 One-forms . . . 404
18.7 Metric as the map between vectors and one-forms . . . 406
18.8 Transformation of a one-form . . . 406
18.9 Arbitrary tensors and their transformations . . . 407
18.10 Tensorial properties of the gradient operator . . . 408
18.11 The invariant volume element . . . 409
18.12 Determinants and the Levi-Civita symbol . . . 411
18.13 Surfaces embedded in Euclidean space . . . 413

19 CALCULUS ON CURVED MANIFOLDS 417
19.1 Fundamental character of tensor equations . . . 418
19.2 Covariant differentiation . . . 419
19.3 Covariant derivative of a second order tensor . . . 422
19.4 Christoffel symbols in terms of the metric . . . 423

19.5 Covariant divergence of a vector 423
19.6 Covariant divergence of a second order tensor 424
19.7 Covariant Laplacian of a scalar 425
19.8 Covariant curl of a vector 425
19.9 Covariant Laplacian of a vector 425
19.10 Integral theorems 426
19.11 Orthogonal curvilinear coordinates 426
19.12 Summary of some salient results 433

CHAPTER
ONE

Ocean climate models

Contents

1.1 Ocean models as tools for ocean science 17
1.2 Ocean climate models 18
1.3 Climate change 19

The purpose of this chapter is to introduce ocean climate models and their use in climate science.

1.1 Ocean models as tools for ocean science

The ocean covers roughly 70% of the earth's surface, and a column of ocean water only 3m thick contains as much heat capacity as the full atmospheric column above (Gill (1982)). Hence, the oceans provide a large reservoir for heat and other constituents of the earth's climate system, such as the increasing amounts of anthropogenic carbon dioxide. Through its buffering abilities and relatively slow time scales, the ocean represents the flywheel of the earth's climate system.

A scientific understanding of the ocean's mean state as well as its variability and stability represents a key goal of physical oceanography and climate science. Due to our inability to perform controlled experiments on large-scale systems studied in the geosciences, such as the earth's climate and its component subsystems, computer models represent a critical tool for rationalizing climate phenomena. For example, computer models are, at an increasing rate, becoming the primary tools used to study and predict physical, chemical, and biological characteristics of the ocean fluid. Such represents a statement about the growing power of computers, improved knowledge and observations of the ocean, and enhancements in the realism of ocean model simulations.

That ocean models are increasingly being used by all sorts of climate scientists, including those not having direct experience developing models, is a sign that the models have enhanced their physical integrity over the past decades to a level rendering a general respect within the broader climate science community. Correspondingly, as model usage climbs, model developers have a growing responsi-

bility to ensure that their codes are physically based, numerically sound, and well documented. Given this mandate, one aim of this book is to establish a level of ocean model documentation that goes beyond the usual technical discussion that assumes the model user is familiar with the fundamentals and understands the physical meanings of the mathematical symbols. Instead, we develop the equations from a (mostly) first principles perspective and take some care to nurture a physical understanding of the mathematics.

1.2 Ocean climate models

Models of the ocean range in complexity from the simplest analytic theories that can be written on the back of an envelope, to the most sophisticated and realistic global ocean circulation models requiring many thousands of lines of computer code. The main focus of this text concerns the realistic models and their formulation.

We use the term *ocean climate model* as a means to distinguish models that simulate the World Ocean over climatologically relevant time scales, from those simulating, say, regional or basin scale dynamics. The distinctions between these ocean modeling subfields is decreasing, largely due to the steady growth in computer power that allows modelers to dispense with some of the simplifications required only a few years ago. Nonetheless, there remain differences, with choices made by one subfield unacceptable to those in others. These differences will likely remain for some years, with associated questions and debates regarding the validity of the choices.

Within ocean climate modeling, there is one model class that has dominanted the field since its beginnings in the 1960's. These are the z-coordinate ocean models, or simply z-models. This is the class of model that forms the focus of the latter parts of this book. The main characteristics of this model class, as well as the other two classes in use for large-scale ocean modeling, are described in Section 5.2.

Simulations of the World Ocean over time scales appropriate for climate (e.g., decades to millenia) involve extremely rich and complex arrays of flow regimes and interactions between components of the climate system. For example, high latitude oceanography involves strong interactions between the ocean with sea ice and rivers, and intense air-sea buoyancy forces induce deep convection and the associated formation of deep water masses. Tropical oceanography involves intense equatorial current systems with rapid adjustments to wind forcing associated with equatorial Kelvin and Rossby waves, and a powerful interannual mode of air-sea variability known as El Niño in the Pacific. Oceanography in the subtropical and subpolar latitudes is dominated by large-scale gyres with meandering and eddying boundary currents forming their western margins. Indeed, amongst other differences, the oceans are distinct from the atmosphere by having leading order influences from the solid earth, where meridional boundaries block otherwise zonal flow except within the Southern Ocean, relatively steep bottom topography variations cause flows to feel the bottom throughout many crucial parts of the World Ocean, and straits and sills funnel water from marginal seas, such as the Mediterranean and Greenland, into the larger ocean basins.

As discussed in Section 5.2, each of the three dominant ocean climate model classes has advantages and disadvantages when simulating various flow regimes

encountered in ocean climate modeling. Indeed, only two of the three classes have routinely been used for global studies (see Griffies et al. (2000a) for discussion). The main reason that z-models presently dominate the field of ocean climate modeling is that their relative simplicity allowed for the early use of these models for interesting climate studies for many decades now, going back to the work of Bryan (1969), Bryan and Lewis (1979), and Cox (1984). Additionally, the z-coordinate is appropriate for the upper ocean mixed layer, which is important for coupled air-sea modeling. In contrast, other classes, such as the isopycnal models, require more sophisticated schemes whose development did not occur until the 1980's.

Ocean climate models continue to evolve. For example, many egregious problems identified with early representatives of the different model classes are now remedied by new numerical treatments. Nonetheless, as argued in Section 5.2, each class has limitations warranting some focus on developing models with generalized vertical coordinates. The hope is that these models, using a suitably defined hybrid vertical coordinate, will reduce many of the egregious problems to mere trivialities.

1.3 Climate change

A central focus of climate science is anthropogenic climate change. Each day, it seems, there is new scientific evidence that industrial society represents a nontrivial geophysical force. Common questions that the person on the street asks are: What should we expect? How much of the present climate variability is due to humans? Providing sound scientifically based answers to these, and other, questions is nontrivial. Indeed, unequivocal answers are not forthcoming. Instead, as with weather prediction, probabilistic statements are the best that the science can provide.

We are at a stage in climate science where the wide variety of climate models yield a general consensus regarding the large-scale effects of increased greenhouse gases. Quite simply, the planet will warm, with higher latitudes feeling the effects moreso than lower latitudes. However, when some more detailed questions are posed, models provide varying projections. Part of the spread is related to the chaotic nature of the climate system. Part is due to large uncertainties in future greenhouse gas emission scenarios. Yet a great deal is due to differing details of the model formulations and their parameterizations. It is to this latter issue that climate scientists can make further progress.

Given the critical importance of models for developing and understanding climate, it is incumbant on model developers to impose the highest standards on model integrity. In particular, ocean climate models should incorporate realistic parameterizations and sound numerical formulations. Yet they must do so at a level of computational expense that does not overly handicap the abilities of the climate modeler to fully investigate various scenarios. Within the ocean science community, this mandate to improve the models entrains hundreds of researchers such as process oriented physicists, chemists, and biologists, seagoing oceanographers, numerical algorithm developers, software engineers, ocean climate modelers, and others. It is anticipated that the questions of climate change will continue to strongly influence and motivate all areas of climate science for many years.

Part I

FUNDAMENTAL OCEAN EQUATIONS

Fundamental Ocean Equations

This part of the book presents the fundamental equations describing the physics of a continuous ocean fluid. These equations are based on classical continuum physics using both Newtonian mechanics and quasi-equilibrium thermodynamics. The material is for the most part independent of the niceties of ocean model numerics. Instead, we focus on the basics. Hence, the following chapters will be of interest to those aiming to understand the physical content of the mathematical equations underlying ocean climate models.

The tensor analysis described in Part VI is necessary to arrive at certain results presented in this part of the book. However, requirements of the reader to penetrate Part VI are actually quite minimal. Additionally, little previous knowledge of continuum mechanics is assumed. Instead, the presentation is aimed at the reader who has a solid foundation in undergraduate physics and vector calculus, yet without a formal course on continuum mechanics.

CHAPTER

TWO

Basics of ocean hydrodynamics

Contents

2.1 Some fundamental ocean processes **25**
2.2 The continuum hypothesis **27**
2.3 Kinematics of fluid motion **28**
2.3.1 Material and fixed-space coordinates 29
2.3.2 Lagrangian and Eulerian descriptions 30
2.3.3 ζ-space and $\mathbf{x}$-space transformations 31
2.3.4 Mass and volume conservation 32
2.4 Kinematical and dynamical approximations **33**
2.4.1 Hydrostatic approximation 34
2.4.2 Traditional Approximation 35
2.4.3 Boussinesq approximation 35
2.4.4 Rigid lid approximation 36
2.4.5 Trends toward reduced approximations 37
2.5 Averaging over scales and realizations **38**
2.6 Numerical discretization **39**

This chapter introduces some of the basic principles and assumptions forming the foundations for a mathematical description of ocean hydrodynamics. It is from this foundation that numerical models are then formulated.

2.1 Some fundamental ocean processes

This book is not a text on oceanography or geophysical fluid dynamics. Nonetheless, we find it useful to mention here some physical processes that are important for the ocean, and thus for representing or parameterizing in realistic ocean circulation models of use for climate science. The list is illustrative, not comprehensive, and is given as motivation for the interested reader to pursue an understanding by referring to the appropriate text or the research literature.

Physical processes in the ocean span space-time scales from millimeters and seconds to global and millenial. These processes can be classified into large-scale mean currents with time dependent fluctuations manifesting as waves, turbulence, large-scale fluctuations of the mean circulation, and fundamental changes in the overall ocean circulation (e.g., transitions of the overturning circulation). Important examples in the wave-like regime include gravity waves set up when the atmospheric winds perturb the ocean surface. Indeed, these surface gravity waves are crucial for transmitting atmospheric momentum into the ocean's upper planetary boundary layer since they provide a frictional element to the overwise smooth ocean surface. In the ocean interior, undulations of the stratified density surfaces are generated through a variety of forcing mechanisms, such as tidal motion over rough topography. The breaking of these waves is thought to contribute to internal mixing of density classes, and so helps determine stratification of the ocean interior. At the ocean's boundaries, waves are set up which help to transmit information around ocean basins, with similar waves set up at the equator where a vanishing Coriolis force acts as an effective boundary allowing for the existence of equatorial Kelvin waves. Finally, at planetary scales, Rossby waves owe their existence to the differential Coriolis force (the β-effect) experienced by large-scale motion on the rotating spherical earth. These waves are important for setting up the ocean's general circulation.

As with any hydrodynamical system, some fluctuations can feed off dynamical instabilities of the mean flow and grow over time. For example, a heavy fluid parcel overlying a lighter parcel rapidly moves vertically to its level of neutral buoyancy as it releases its potential energy. This effect leads to convective overturning, which is an important process for the formation of the oceans' water masses. An additional process involved with deep water formation is the cascade of heavy water down topographic slopes in the certain parts of the World Ocean. These unstable flows are quite turbulent and so entrain much of the ambient fluid. Another unstable process, which is ubiquitous in the ocean, arises when gravitationally stable density fronts, common in a rotating stratified fluid, are perturbed so that the sloped front breaks down into unstable undulations that transfer potential energy into kinetic energy. This process, known as baroclinic instability, forms one of the most important instability mechanisms in geophysical fluids.

In general, after an initial rapid growth phase, unstable fluctuations saturate via nonlinear wave breaking and tend to evolve into a highly chaotic or turbulent state. Such turbulent states occur on all scales, from the micros-scale to planetary scale, and they efficiently stir and mix ocean properties across these scales. At the largest scales, the turbulence associated with baroclinic instability, known as geostrophic turbulence, represents the large-scale storms of the ocean, analogous to the synoptic storms comprising the atmosphere's weather. Due to differences in vertical stratification, the scale of oceanic geostrophic eddies is roughly ten times smaller than the atmosphere. Hence, simulating these eddies explicitly is far more difficult than in the atmosphere due to the associated increased computational requirements.

The oceanic processes mentioned above, and numerous others, are important for understanding, modeling, and predicting the ocean circulation. Such issues are of fundamental interest scientifically and have driven the fields of physical oceanography and ocean modeling for much of their history. They have furthermore become increasingly important to society within the past one or two decades

due to the nontrivial human effects on the earth's climate system. Numerical models are thus being relied upon to help answer questions of profound importance to life on this planet. Physically sound and rational numerical models are critical as we seek an understanding of the climate's response to the CO_2 experiment now being conducted.

2.2 The continuum hypothesis

There are roughly 18 grams of pure water per mole, where a mole has roughly 6×10^{23} molecules (Avogodro's number). As noted in Section 1.2 of Batchelor (1967), a tiny parcel of pure water with dimensions on the order of 10^{-3}cm has a volume of 10^{-9}cm^3. With a density of 1g/cm^3, this parcel has roughly 3×10^{19} molecules. For macroscopic purposes, 10^{-9}cm^3 is extremely tiny in so far as macroscopic properties such as density, tracer concentration, and temperature are effectively uniform over this volume. Yet the 3×10^{19} molecules in this volume means that it is gigantic from the perspective of molecular dynamics. The same sized parcel of air, with approximately 29 grams per mole, at the earth's surface and $0°C$ has a density of $1.3 \times 10^{-3}\text{g/cm}^3$. Therefore, a 10^{-9}cm^3 parcel of air at the earth's surface has roughly 3×10^{10} molecules, which is again a huge number.

These numbers suggest that for the purposes of geophysical fluid dynamics, and for many other areas of fluid mechanics, it is extremely accurate to characterize a fluid parcel with volume on the order 10^{-9}cm^3 as *macroscopically small yet miscropically large*. The macroscopically small characterization is warranted since such parcels of fluid have a near uniform set of thermodynamic properties (temperature, salinity, tracer concentration, etc.). Hence, the first and second laws of thermodynamics can be applied to such parcels using a local thermodynamic equilibrium assumption, which constitutes the basis for classical irreversible or non-equilibrium thermodynamics (e.g., DeGroot and Mazur (1984)). That is, each fluid parcel, though generally far from equilibrium with other parcels, itself constitutes an open thermodynamic system in near-equilibrium.

The microscopically large characterization warrants our treating the fluid as a *mathematical continuum*. That is, we need not worry about the discrete nature of molecules when concerned with ocean dynamics. As such, we will be taking limits as volumes go to zero with impunity in formulating the differential laws of fluid motion. Figure 2.1 illustrates this proposition, which formally is known as the *continuum hypothesis*. The physical processes mentioned in Section 2.1, and all those of conceivable importance to ocean dynamics, can be described using the continuum hypothesis.

As the above discussion suggests, the continuum fluid equations are in principle related to more fundamental equations of motion, describing the multitude of molecular degrees of freedom, through a series of averaging operations. Hence, the laws of continuum fluid dynamics represent a *mean field theory*. A rigorous proof of this statement, however, is available only for the simplest of gaseous systems, and can be found in statistical mechanics books such as Huang (1987). Also, the reader interested in these issues may find the discussion in Chapter 1 of Salmon (1998) enlightening. For the ocean, or for other liquid media, the continuum hypothesis has not been rigorously proven. Nonetheless, its validity is evidenced by its high

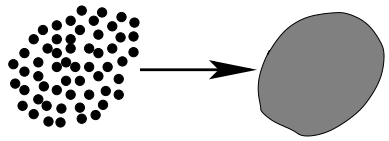

Figure 2.1: Fundamentally, a small parcel of water, say something on the order 10^{-9}cm^3, is comprised of a huge number of interacting molecules (roughly 3×10^{19}), schematically drawn in the left figure. For purposes of geophysical fluid dynamics, we can safely ignore details of the individual molecules and approximate the collection of molecules as a continuum, as drawn in the right figure. Such constitutes the continuum hypothesis.

degree of success in describing the bulk of liquid and gas motions, including those relevant for geophysical fluid dynamics.

After making the continuum hypothesis, the differential laws of thermodynamics and hydrodynamics are derived by applying Newtonian dynamics and thermodynamics to a collection of local thermodynamically equilibrated fluid parcels. The resulting field equations possess information regarding a broad spectrum of fluid motion, with spatial scales reaching from the parcel level to those relevant for global ocean dynamics, and temporal scales on the order of milliseconds to millennia. The continuum equations, often called the Navier-Stokes equations, have been shown through numerous experiments over the past 100-200 years to be very successful in describing fluid flow. That these equations are relevant over such a large space-time range represents a spectacular success in providing a rational mathematical description of natural phenomena.

2.3 Kinematics of fluid motion

Kinematics is the branch of mechanics concerned with the intrinsic aspects of motion, the geometry of space-time where motion occurs, and the conceptual and mathematical tools used to describe motion. It is not concerned with the forces or causes of motion, each of which are within the purview of *dynamics*. Kinematical results are therefore very generic and of great utility for a variety of dynamical situations.

The purpose of this section is to introduce some of the basic kinematics relevant to the motion of fluid parcels. In particular, it is here that we establish the relationship between *Eulerian* and *Lagrangian* descriptions of that motion. Z-coordinate ocean models employ an Eulerian framework, in which all spatial coordinates are fixed on a rotating sphere. In contrast, isopycnal models employ a semi-Lagrangian framework, in which the vertical coordinate is a function of space-time (hence Lagrangian) yet the horizontal coordinates are fixed (hence Eulerian). Chapter 5 discusses some mathematical issues of models using such *generalized vertical coordinates*.

2.3.1 Material and fixed-space coordinates

In describing the dynamics of fluid parcels, we may take the perspective that the parcels form a continuum of deformable "particles." In this way, many of the ideas from point particle mechanics can be transferred over to fluid parcel mechanics. To proceed in this manner, it is necessary to distinguish between the infinite number of fluid parcels. Recall that in classical particle mechanics, particles are distinguished by an integer label. For a continuum of fluid parcels, the integer label becomes continuous. That is, the fluid parcels are distinguished by ascribing a value for a continuous label, ζ, to each parcel filling up the fluid continuum. Associated with this *material spatial coordinate* is the co-moving or *material time coordinate* τ measured by an observer riding along with the fluid parcels. The coordinates (ζ, τ) are thought of as coordinates in a *material* space-time. For ideal parcel dynamics, the only dynamics in material space-time is the trivial dynamics $\partial_\tau \zeta = 0$, since the parcels maintain their integrity for all time (see Figure 2.2). For real fluids, a parcel's identity is lost over time due to the mixing of properties with other parcels. In general, material coordinates are often called *Lagrangian* or *convected* coordinates in the literature.

Measurements of fluids discussed here are taken by observers living on a generally curved non-Euclidean surface or volume embedded in a background Euclidean space-time*. In particular, for geophysical fluid dynamics, observers are on the surface of a rotating sphere (a non-Euclidean space), approximating the rotating earth, where the sphere is embedded in a Euclidean space-time background (the space-time of Newtonian mechanics).

A physical description of the fluid, motivated by the particle mechanics analog, includes the trajectory of fluid parcels as they move in Euclidean space-time. A set of coordinates can be used to specify these trajectories (e.g., cartesian, spherical, cylindrical, etc.), and these are sometimes called *physical* coordinates. This name is not preferred here since there is nothing "unphysical" about the material coordinates. Rather, the name *fixed* or *position* coordinates is prefered. In turn, it should be noted that the "fixed" coordinates for ocean modeling are actually rotating coordinates, and so are strictly not "fixed" in the sense of inertial coordinates. Rather, they are fixed relative to a terrestrial observer's rotating frame of reference.

The continuum of fluid parcel trajectories is represented by the vector field

$$\mathbf{x} = \mathbf{x}(\zeta, \tau). \tag{2.1}$$

The vector field $\mathbf{x}(\zeta, \tau)$ measures the Cartesian position of a parcel with material space-time coordinates (ζ, τ). This position is taken relative to some fixed origin in Euclidean space. Figure 3.1 illustrates the situation for a fluid parcel moving on a spherical planet. For geophysical fluids, it is sufficient to assume that fluid parcels move at speeds well below the speed of light. Therefore, the co-moving time, τ, and the time t measured by observers at rest with respect to the origin $\mathbf{x} = 0$ can be taken to be the same

$$t = \tau. \tag{2.2}$$

*For geophysical fluid dynamics, there is no reference to relativistic fluids, which would involve Minkowski space-time, the space-time of special relativity. Nor is there reference to cosmological systems, for which the curved space-times of general relativity are relevant.

The relations (2.1) and (2.2) specify a time dependent coordinate transformation, or mapping, between the material space and the fixed-space. Since the time parameters are the same, it is sufficient to distinguish these two space-times simply by calling the material space-time as $\boldsymbol{\zeta}$-space, and the fixed space-time as $\mathbf{x}$-space. Note that the number of geometric dimensions needed to describe the material space is the same as the number of dimensions in the fixed-space. For example, we may choose to define the material label as the initial position of the field of fluid parcels, $\boldsymbol{\zeta} = \mathbf{x}(\boldsymbol{\zeta}, \tau = 0)$.

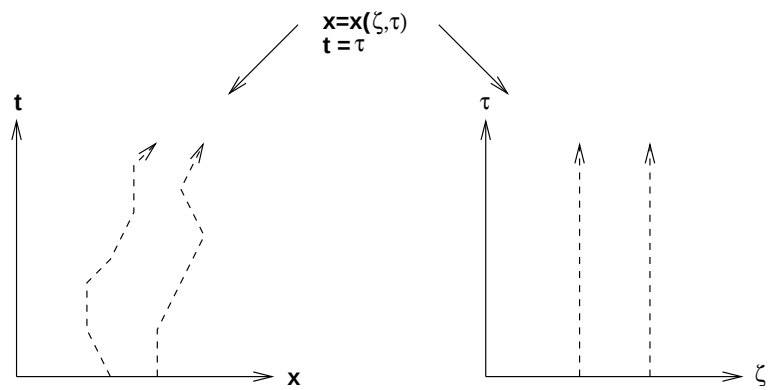

Figure 2.2: Illustration of the space-time paths or trajectories exhibited by two ideal fluid parcels. The left panel shows the trajectories as seen in a fixed-space or Eulerian perspective. The paths are typically curved, which represents motion due to waves, advection, etc. The right panel shows the same trajectories in the material space-time. These trajectories are trivial for ideal parcels since they maintain the same value for their spatial material coordinate $\boldsymbol{\zeta}$, hence exhibiting a linear path in material space-time. The two pictures are related through the time dependent coordinate transformation $\mathbf{x} = \mathbf{x}(\boldsymbol{\zeta}, \tau)$ and $t = \tau$.

2.3.2 Lagrangian and Eulerian descriptions

The *Lagrangian description* of a fluid is afforded by knowledge of the continuum of parcel trajectories $\mathbf{x}(\boldsymbol{\zeta}, \tau)$. The equations of motion are written much as the equations for a point particle, yet with forcing terms written generally as partial derivatives with respect to the material coordinate $\boldsymbol{\zeta}$ and functional derivatives with respect to the parcel trajectory $\mathbf{x}(\boldsymbol{\zeta}, \tau)$.

In contrast, an *Eulerian description* is afforded by knowledge of the velocity field

$$\mathbf{v}(\mathbf{x}(\boldsymbol{\zeta}, \tau), t) = \partial_\tau \, \mathbf{x}(\boldsymbol{\zeta}, \tau). \tag{2.3}$$

This relation defines the mapping between the Eulerian and Lagrangian descriptions of the fluid: knowledge of one description implies knowledge of the other as specified through this differential equation.

2.3.3 ζ-space and x-space transformations

As mentioned in Section 2.3.1, the relation between the space point $\mathbf{x}$ and the parcel trajectory $\mathbf{x}(\zeta, \tau)$ defines a time dependent mapping between the fixed space-time with the material space-time:

$$x^m = x^m(\zeta^a, \tau) \tag{2.4}$$

$$t = \tau, \tag{2.5}$$

where $m = 1, 2, 3$ labels coordinates in $\mathbf{x}$-space, and $a = 1, 2, 3$ labels coordinates in ζ-space. Although identification of the time parameters $t = \tau$ has been made, it is important to use different symbols. The reason is that the partial derivatives with respect to τ are taken with the material space coordinates ζ constant, whereas partial derivatives with respect to t are taken with the fixed-space coordinates $\mathbf{x}$ held fixed. As such, the material time derivative is related to the fixed-space time derivative through the expression

$$\partial_\tau = \partial_t + \dot{x}^m \partial_m, \tag{2.6}$$

which follows from the usual rules of coordinate transformations. A more familiar notation is to introduce the *total* or *material time derivative*

$$\frac{\mathrm{d}}{\mathrm{d}t} = \partial_t + \mathbf{v} \cdot \nabla, \tag{2.7}$$

where $\mathbf{v}$ is the Eulerian velocity field related to the time tendency of the parcel's trajectory

$$\dot{x}^m(\zeta, \tau) = v^m(\mathbf{x}(\zeta, \tau), t). \tag{2.8}$$

The nonlinear term $\mathbf{v} \cdot \nabla = \dot{x}^m \partial_m$ is known as the *transport*, or *advection*, term, with advection the usual terminology in geophysical fluid dynamics. Advection arises since an observer at a fixed position $\mathbf{x}$, associated with the Eulerian description, is not a co-moving or material observer. The Lagrangian observer is co-moving with the fluid parcel, and is therefore at a fixed material space coordinate ζ and measures time with τ.

The mass of a fluid parcel is given by

$$\mathrm{d}M = \varpi(\zeta, \tau)\, \mathrm{d}\zeta. \tag{2.9}$$

In this expression, $\varpi(\zeta, \tau)$ is the fluid mass density in material space, and $\mathrm{d}\zeta = \mathrm{d}\zeta^1\, \mathrm{d}\zeta^2\, \mathrm{d}\zeta^3$ is a useful shorthand for the volume element in material space. Due to the conservation of mass for each parcel

$$\partial_\tau(\mathrm{d}M) = 0, \tag{2.10}$$

which means that the mass density in material space-time must be independent of material time:

$$\varpi(\zeta, \tau) = \varpi(\zeta). \tag{2.11}$$

Therefore, $\varpi(\zeta)$ can be chosen to be the material space mass density at any convenient time.

The familiar rules of calculus* yield an expression for the mass element written using the **x**-space coordinates

$$\mathrm{d}M = \varpi\,\mathrm{d}\boldsymbol{\zeta} = \varpi\,J^{-1}\,\mathrm{d}\mathbf{x}, \tag{2.12}$$

where

$$\begin{aligned} J &= \det(x_a^m) \\ &= \frac{\partial \mathbf{x}}{\partial \boldsymbol{\zeta}} \\ &= \frac{1}{3!}\epsilon_{mnp}\epsilon^{abc}\left(\frac{\partial x^m}{\partial \zeta^a}\frac{\partial x^n}{\partial \zeta^b}\frac{\partial x^p}{\partial \zeta^c}\right) \end{aligned} \tag{2.13}$$

is the determinant of the transformation matrix between the position coordinates and material coordinates; i.e., it is the Jacobian, and ϵ is the totally anti-symmetric Levi-Civita symbol defined in Section 18.12. Therefore, the mass density of the fluid as expressed in the position space, ρ, is related to the mass density of the fluid as expressed in material space, ϖ, by the expression

$$\rho = \varpi\,J^{-1}. \tag{2.14}$$

Note that uniform mass density in material space (ϖ = constant) does not imply uniform mass density in position space, and vice-versa.

2.3.4 Mass and volume conservation

In classical fluid mechanics, mass is conserved for fluid parcels regardless the dynamical interactions. Hence, mass conservation naturally is considered as part of fluid kinematics. We present here an elementary derivation of mass conservation using an Eulerian description for a fluid devoid of mass sources and sinks. The derivation proceeds in a manner useful for establishing other fluid kinematical results.

In $\boldsymbol{\zeta}$-space, conservation of parcel mass implies that $\partial_\tau \varpi = 0$, as mentioned previously. For **x**-space, we derive the consequences of mass conservation by setting $\partial_\tau(\rho\, d\mathbf{x}) = 0$. Note that since mass is a scalar under the Galilean transformations relevant for geophysical fluid dynamics, the differential mass conservation law takes the same form in any frame of reference.

With the mass of a parcel written

$$\mathrm{d}M(\boldsymbol{\zeta}) = \rho(\mathbf{x}(\boldsymbol{\zeta},\tau),t)\,\mathrm{d}\mathbf{x}(\boldsymbol{\zeta},\tau), \tag{2.15}$$

conservation of mass for each parcel implies

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t}\ln(\mathrm{d}\,M) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\ln(\rho\,\mathrm{d}V) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\ln(\rho\,\mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3) \\ &= \frac{\mathrm{d}\ln\rho}{\mathrm{d}t} + \frac{\mathrm{d}}{\mathrm{d}x^m}\dot{x}^m, \end{aligned} \tag{2.16}$$

*See Schutz (1980), or most books on differential forms, for a more elegant proof in terms of the volume 3-form.

where

$$\mathrm{d}V = \mathrm{d}\mathbf{x} \tag{2.17}$$

is a common notation for the infinitesimal volume of a fluid parcel written in Cartesian coordinates. The parcel mass conservation equation (2.16) is satisfied by each of the parcels throughout the fluid, which implies that it holds at each point of the fluid at all times. It is therefore convenient to introduce the Eulerian velocity field $\mathbf{v} = \dot{\mathbf{x}}$, which leads to the statement of parcel mass conservation

$$\frac{\mathrm{d}\ln\rho}{\mathrm{d}t} = -\nabla\cdot\mathbf{v}. \tag{2.18}$$

Expanding the material time derivative allows this equation to be written in the form of an Eulerian conservation law

$$\rho_{,t} + \nabla\cdot(\mathbf{v}\,\rho) = 0, \tag{2.19}$$

with $\rho_{,t} = \partial_t\rho$ a useful shorthand.

For many purposes, a seawater parcel can be approximated as incompressible in the sense that changes in the volume of a parcel are small

$$\partial_\tau\left(\mathrm{d}V\right) \approx 0. \tag{2.20}$$

This is the case, in particular, when the time changes of a parcel's density are very small relative to the density

$$\frac{\mathrm{d}\ln\rho}{\mathrm{d}t} \approx 0. \tag{2.21}$$

Letting the approximation become an equality leads to a fluid whose parcels materially conserve their volume

$$\partial_\tau\left(\mathrm{d}V\right) = 0 \Rightarrow \nabla\cdot\mathbf{v} = 0. \tag{2.22}$$

Materially conserved volume is part of the *Boussinesq* approximation, and this approximation has found much use for ocean modeling. Notably, volume conservation *does not* mean the fluid maintains a materially constant density. Indeed, material changes in density affect changes in hydrostatic pressure, which in turn drives much of the ocean's current systems, especially those at depth.

2.4 Kinematical and dynamical approximations

It is by their success that the Navier-Stokes equations in turn are often cumbersome to apply in practice. That is, since they encompass such a huge spectrum of dynamical motions, the theoretician, modeler, and experimentalist typically find it difficult to focus on certain elements of these motions while using the full set of equations. Reducing the spectral range over which the equations are valid has motivated a multitude of methods to approximate the equations of fluid motion.

The purpose of this section is to summarize some common approximations made in ocean climate modeling. We have more precise descriptions of these approximations in later chapters. For now, we are content to give the general ideas.

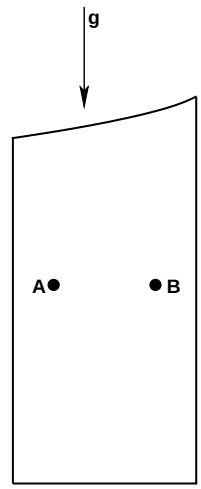

Figure 2.3: Shown here is a schematic of a column of ocean fluid within a gravitational field. The two points indicated within the fluid are at the same distance from the bottom. However, because of the slightly higher height of fluid above point B, the hydrostatic pressure at B is greater than the hydrostatic pressure at A.

2.4.1 Hydrostatic approximation

Pressure at a point within a fluid at rest in a gravitational field arises just from the weight of fluid above the point (see Figure 2.3). That is, vertical pressure gradients are given by the buoyancy at the point. This situation constitutes the *hydrostatic balance*. When the fluid is in motion, vertical pressure gradients are also affected by vertical accelerations and friction. However, for many geophysically relevant fluid motions, the dominant balance in the vertical momentum equation remains the hydrostatic balance. The level to which the hydrostatic balance remains dominant is directly proportional to the ratio of vertical H to horizontal L scales of motion. For many motions of direct interest in ocean climate modeling, the vertical scale is on the order of a few meters to a few kilometres, whereas the horizontal scale is on the order of tens to thousands of kilometers. Thus, H/L is typically far less than $1/10$.

In addition to simplifying the computation of pressure, the hydrostatic approximation filters out sound waves, which are three-dimensional pressure fluctuations (e.g., Gill (1982) and Apel (1987)). There remains, however, a strictly horizontal acoustic mode known as the Lamb wave that is not filtered by the hydrostatic approximation. The Lamb wave propagates at the speed of sound, yet has small energy and may be dissipated by numerical effects. We have more to say about acoustic modes when discussing the Boussinesq approximation in Section 2.4.3, where it is noted that the Boussinesq approximation filters all acoustic modes from both hydrostatic and non-hydrostatic models.

For studies aiming to simulate motions with large vertical accelerations and relatively small horizontal scales, such as those occuring in convective regions, the hydrostatic approximation can be an unacceptable limitation. Refer to Marshall et al. (1997) for a lucid discussion of non-hydrostatic Boussinesq ocean modeling.

2.4.2 Traditional Approximation

The ocean is a shallow layer of fluid moving on an approximately spherical earth. When measuring the distance between two points within the ocean, one must use a metric tensor (Chapter 18). The metric tensor components are functions of the latitudinal, longitudinal, and radial position within the ocean.

Recognizing the huge scale separation between the depth of the ocean fluid and the radius of the earth, the Traditional Approximation drops the radial dependence of the metric tensor components (Section 3.1). Instead, it sets the radius to a constant as given by the radius of a sphere R with the same volume as the earth. The dynamical equations for a fluid parcel moving on a sphere with this approximate metric constitute the *primitive equations* (Chapter 3). As discussed by Marshall et al. (1997), one may wish to avoid the Traditional Approximation for those situations where the hydrostatic approximation is not suitable, since here vertical and horizontal excursions of fluid parcels are comparable. Hydrostatic ocean climate models typically retain the geometric basis of the Traditional Approximation.

2.4.3 Boussinesq approximation

One common approximation employed in ocean climate modeling is the Boussinesq approximation. From a kinematic perspective, a Boussinesq fluid parcel maintains the same volume during its transport through the fluid, regardless the dynamical interactions. Without volume sources, a Boussinesq ocean maintains a constant total volume. This behaviour is to be contrasted to the more fundamental mass conservation property of non-Boussinesq fluid parcels, whereby an ocean without mass sources maintains a fixed total mass. Figure 2.4 provides a schematic of the differences.

The Boussinesq approximation has been commonly used in ocean climate models due to the near incompressibility of ocean fluid parcels, and thus the near conservation of volume maintained by these parcels. Additionally, a Boussinesq fluid does not support acoustic fluctuations (e.g., Gill (1982)), thus filtering out the climatologically unimportant acoustic waves. By filtering out such fast and unimportant waves, an ocean model based on the Boussinesq approximation can support larger time steps without violating the Courant-Friedricks-Levy (CFL) condition (e.g., Haltiner and Williams (1980) and Durran (1999). Larger time steps makes the model more efficient and thus enhances its usability for climate studies.

Nonetheless, as with any approximate description, a Boussinesq fluid has its limitations, and these limitations have been increasingly scrutinized of late. Most notably for climate purposes, a Boussinesq fluid does not render an accurate computation of the sea level height. The reason is that it does not incorporate fluctuations in the depth averaged density field (Section 3.3.3). Such *steric effects* are not incorporated into volume conserving kinematics. The example of uniformly heating the ocean illustrates this point. As with the example given in Figure 2.4, uniformly heating a non-Boussinesq ocean causes its density to decrease and its volume to increase in order to maintain a constant mass. Hence, the non-Boussinesq ocean's sea level rises under uniform heating. In contrast, uniformly heating a Boussinesq ocean causes its density to decrease and its mass to decrease in order to maintain a constant volume. Hence, the Boussinesq ocean's sea level remains constant under

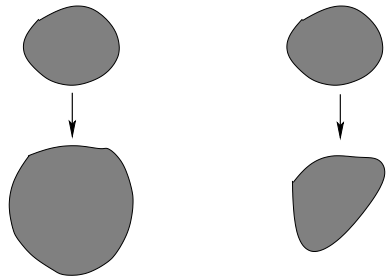

Figure 2.4: The left schematic illustrates a mass conserving fluid parcel, as in a non-Boussinesq ocean. As it is deformed by currents and affected by changes in temperature, salinity, and pressure, the parcel's volume and density generally change, yet its mass $dM = \rho\, dV$ remains constant. In contrast, a Boussinesq fluid parcel evolves without changing its volume $dV = dM/\rho$, which allows for changes in its density and mass. The right schematic is meant to illustrate such constant-volume deformations. The particular example of heating is revealing. Heating a mass conserving parcel reduces its density and so increases its volume, whereas heating a volume conserving parcel reduces its density and so reduces its mass via *virtual mass fluxes* removing mass from the parcel.

uniform heating. The sea level height prognostically computed from Boussinesq ocean models is therefore insufficient for purposes of predicting sea level changes associated with global warming.

The pressure within a Boussinesq model is compromised due to the inaccurate computation of the ocean surface height. Additionally, as volume is conserved instead of mass, Boussinesq fluid parcels generally change their mass via unphysical *virtual mass fluxes*, thus influencing the weight of a fluid column, again affecting the pressure field. We focus here on the effects on pressure since satellite measurements will soon be able to determine the absolute pressure at the ocean bottom to an extremely high level of accuracy, sufficient to distinguish a Boussinesq fluid from a non-Boussinesq fluid (e.g., Wunsch and Zlotnicki (1999) and Hughes et al. (2000)).

Although diagnostic adjustments can be made to account for inaccuracies in the Boussinesq fluid's surface height (e.g., Greatbatch (1994) and Mellor and Ezer (1995)), these adjustments have their limitations, and they do not overcome problems associated with bottom pressure errors (Greatbatch et al. (2001)). Hence, for the purpose of checking the accuracy of ocean models as compared to data, we can fully exploit the satellite bottom pressure measurements only by using a non-Boussinesq ocean model. In this book, we employ the methods of Greatbatch et al. (2001) to readily allow for the implementation of either a Boussinesq or non-Boussinesq z-coordinate ocean model. The degree to which inaccuracies in the Boussinesq approximation are an issue can therefore be directly tested in such models.

2.4.4 Rigid lid approximation

The rigid lid approximation assumes that the surface elevation of the ocean fluid is static. Doing so filters out fluctuations of the ocean associated with motions of the

ocean surface. Most notably, *external mode* gravity wave fluctuations cause undulations in the sea level height, and these waves can move at speeds some 100 times faster than the next fastest *internal* mode fluctuations. For many climate modeling purposes, the external gravity waves are only of indirect interest, and so their absence from the model is of little physical consequence yet of great computational advantage due to the reduced CFL constraints. It is for these reasons that the methods of Bryan (1969) exploited the rigid lid approximation.

The most physically unsatisfying consequence of assuming a fixed surface height is that, as commonly implemented, a rigid lid ocean does not allow the passage of fresh water across the ocean surface. That is, the typical form of a rigid lid ocean maintains a constant volume. This is a severe restriction on the utility of the ocean model, since freshwater fluxes are important in their affect on the ocean's buoyancy field. There are well known indirect methods to include hydrological forcing on the rigid lid model, yet they are *ad hoc* and physically unsatisfying (e.g., Huang (1993), and Griffies et al. (2001)). Ocean climate modelers have therefore been moving away from the rigid lid assumption, and moving towards less restrictive free surface methods. In addition, free surface methods allow for the direct simulation of tides in ocean climate models, whereas rigid lid models do not.

The most computationally unsatisfying consequence of the rigid lid approximation is that it introduces an elliptic problem with attendent global boundary conditions. Solving elliptic problems where the geometry, topography, and forcing fields are complex, as in the World Ocean, is a computationally difficult problem. Additionally, the implementation of parallel algorithms is hindered by the need to evaluate the global boundary conditions. Hence, although possessing some advanatages from a theoretical perspective (e.g., by filtering out certain modes that theoretically may be undesired), ocean climate modelers have generally been moving away from algorithms involving elliptic problems. The unapproximated hyperbolic system is proving to be both physically more accurate and computationally more tractable on parallel machines.

2.4.5 Trends toward reduced approximations

As explained in the previous discussion, approximations may be useful for many purposes, yet they often come at the cost of reducing the physical integrity and/or accuracy of the solution. Indeed, as with the rigid lid approximation, they often have poor computational efficiency. Hence, there is a general trend in ocean climate modeling towards reducing many of the common approximations (see Griffies et al. (2000a) for a review of developments in ocean climate modeling).

There are three general reasons to reduce the number of approximations. First, measurement techniques are getting quite accurate and will improve in the future. For example, as mentioned in Section 2.4.3, it is anticipated that remote methods will soon be able to measure ocean bottom pressures accurately enough to distinguish between a Boussinesq and non-Boussinesq fluid. Second, approximations such as the rigid lid approximation place restrictions on the utility of the model for realistic ocean climate simulations incorporating sea ice, rivers, hydrological cycles, and tides. Third, the motivation for making approximations in the early days of ocean modeling were largely based on insufficient computer power. Three to four decades of ocean modeling have taught us there will "never" be enough

computer power to do all the experiments desired. Nonetheless, present-day computers do allow us to run global models with increasingly realistic scenarios and improved skill. Hence, in our aim to more accurately simulate the global ocean over its broad range of temporal and spatial scales, modelers are uninterested in handicapping their abilities by making restrictive assumptions on the basic model equations. Such motivation has led to the gradual elimination of the previously ubiquitous Boussinesq and rigid lid approximations in ocean climate models. In some ten years time, global modelers will likely also dispense with the hydrostatic approximation.

2.5 Averaging over scales and realizations

Oceanic flow is generally nonlinear, chaotic, and turbulent. Correspondingly, the ocean's spatial and temporal scales are extremely broad, with fluctuations reaching from the parcel level and seconds, to the global scale and millennia. Coupling occurs between the scales due to nonlinear cascades of flow properties, such as energy and variance. Furthermore, we cannot measure every spatial point within a fluid, nor can we maintain a measurement for all time. Thus, we are limited in our ability to garner information about the turbulent fluid's spatio-temporal properties. Sensitive dependence on initial conditions in this turbulent flow then severely limits our ability to predict the precise evolution of oceanic flow. This lack of information and limited predictability motivates our formulating an averaged or mean-field set of continuum fluid equations.

To derive the mean field ocean equations, we envision a multitude of ocean flows, each differing by their initial and/or boundary conditions. Averaging over these realizations leads to the *ensemble averaged* equations of motion. Formally, there are an infinite number of realizations in an ensemble. Alternatively, averages over an imagined infinite sized ensemble are replaced by integrations over a probability distribution function setting the likelihood of a particular phase space configuration. We prefer ensemble averages to space-time averages since ensemble averages allow for an unambiguous interchange between the averaging operator and integral/differential operators appearing in the unaveraged equations of motion. However, we are admittedly non-rigorous in our usage of ensemble averages and have nothing to say about whether the ocean fluid satisfies the ergodic hypothesis.

Conceptually, we assume that the distance in phase space between members of an ensemble is directly proportional to the spatio-temporal scales that remain unresolved by the resolution of the measurements and/or discrete numerical model. For example, ensemble members may differ by the phase of a wave-like or turbulent eddying fluctuation occuring on a spatial scale of some tens of kilometres. For measurements and/or models that resolve only scales greater than some hundreds of kilometres, these *mesoscale* features remain unresolved and so constitute part of the sub-grid scale (SGS) regime (see Chapter 7 for more discussion). Other features smaller than mesoscale fluctuations also constitute dynamics within the unresolved spectrum, and so likewise must be averaged over when formulating the equations describing the large-scale features (see Chapter 6).

As discussed in Chapter 6, a very practical and useful goal for the averaging methods is to produce averaged kinematical and dynamical equations maintain-

ing the same mathematical form as the unaveraged equations. Doing so allows for mathematical properties of the unaveraged equations to be shared by the averaged equations. Although mathematically similar, the terms appearing in the two sets of equations differ in their interpretation. In particular, the averaged equations contain terms unknown to the "averaged observer." These terms take a form dependent on the dynamics occupying the SGS.

Specifying the SGS terms is a complicated problem in *turbulence closure theory*. Notably, most closures contain *ad hoc* and formally unjustified steps, thus prompting a multitude of different closures. Unfortunately, ocean model simulations have proven to be quite sensitive to the general forms of the closures.

Although no complete theory exists to tell us the full details of closures, there is some very compelling evidence, based on ocean measurements and model simulations, to guide their general form. For example, in large-scale ocean modeling, lateral SGS tracer mixing is predominantly directed parallel to neutral directions (e.g., Solomon (1971), Redi (1982), Olbers et al. (1985), McDougall and Church (1986), McDougall (1987a), Gent and McWilliams (1990), Gent et al. (1995), and Griffies et al. (1998)). In particular, ocean climate models using horizontally directed tracer mixing, even those models of very high resolution capable of resolving meso-scale eddies (Roberts and Marshall (1998)), can be of less physical integrity than those with neutrally directed mixing. Detailed aspects of neutral physics are addressed in Chapters 7 and 11–14.

2.6 Numerical discretization

A mathematical description of fluid motion starts from the continuum approximation, whereby molecular dynamical scales are averaged out to reveal a mean field set of equations. Then, depending on ones interest in dynamical regimes, selected approximations are introduced to reduce the huge scales of motion allowed by the full continuum set of equations. Next, due to incomplete information and sensitive dependence on initial conditions, averaging methods are required to form a set of fluid equations which can be integrated on a computer. Setting forms for the SGS terms renders the averaged equations closed.

After these fundamental mathematical and physical steps are finalized, the ocean modeler must formulate the averaged equations of motion in a manner accessible to finite computational capabilities. That is, the approximated, averaged, closed, continuum partial differential equations must be cast into a discrete form so that they are accessible to the numerical/algebraic methods interpretable by modern computers. In turn, solution algorithms must be developed in order to evolve numerical solutions forward in time.

There are various methods for casting the continuum equations into discrete form. One method integrates the continuum equations over finite grid cells, defining cell averaged variables, thus resulting in a set of discrete *finite volume* equations of motion (Section 6.9). Most ocean climate models follow this approach to some degree. However, there are notable exceptions. In particular, hydrostatic pressure at a model grid point is often interpreted as a point-wise value, rather than the averaged pressure over the depth of the grid cell (see Griffies et al. (2003) for discussion).

Numerical discretization breaks the continuum symmetries present in the partial differential equations of fluid mechanics. Hence, solutions realized on the computer may not always correspond to those allowed by the continuum equations. Eliminating such spurious numerical solutions is a constant concern of the numerical modeler. In general, significant care is required so that the numerical fluid is a close analog to the continuum fluid.

Assuming the continuum equations are successfully cast into a numerical algorithm, a numerically realized fluid provides the modeler with a powerful tool to probe the many dynamical regimes present in the equations of motion. That is, numerical models are powerful tools to perform investigative experiments and to extrapolate observations into the future via prediction. For example, the modeler is able to test ideas by altering sub-grid scale parameters, changing boundary forcing or domain geometry, refining the grid mesh sizes to allow the flow to become more nonlinear, simplifying the dynamical degrees of freedom via approximations, identifying dominant balances by measuring terms within the model equations, and projecting scenarios for future ocean climate change due to alterations in atmospheric radiative forcing. Such direct experimental methods are unavailable for studies of large spatio-temporal scales, due to our inability to perform reproducible and controllable experiments with Nature's climate system. Furthermore, the suite of solutions available numerically is far larger than those available analytically. Hence, when combined with analytical methods of highly idealized configurations, laboratory experiments of small-scale phenomena, and *in situ*, remote, and paleo ocean observations, numerical models round out the suite of tools for ocean climate science. Indeed, numerical models have become the repository for the observations and theories of ocean climate.

CHAPTER
THREE

Ocean hydrodynamics

Contents

3.1 Orthogonal coordinates **42**
3.1.1 Spherical and Cartesian coordinates 42
3.1.2 Horizontal orthogonal coordinates 43
3.2 Geometry of the Traditional Approximation **44**
3.3 Volume and mass conserving kinematics **45**
3.3.1 Mass and volume conservation for parcels 45
3.3.2 Volume conservation for finite domains 46
3.3.3 Mass conservation for finite domains 49
3.4 Principles of continuum dynamics **53**
3.4.1 Position, velocity, and acceleration 53
3.4.2 External/body and internal/contact forces 54
3.4.3 Local equilibrium of contact stresses 54
3.4.4 The stress tensor 55
3.4.5 Equations of motion 56
3.5 Dynamics of fluid parcels **56**
3.5.1 Time tendency of the linear momentum 56
3.5.2 Motion on a rotating sphere 58
3.5.2.1 Earth's angular velocity 58
3.5.2.2 Inertial and non-inertial frames 59
3.5.2.3 Effective gravitational force 59
3.5.2.4 Gravity for a global ocean model 61
3.5.2.5 Coriolis acceleration 61
3.5.3 Internal forces 62
3.5.4 Evolution of linear momentum density 62
3.5.5 Vector invariant velocity equation 63
3.6 Hydrostatic pressure **64**
3.6.1 Hydrostatic balance 64
3.6.2 Bottom pressure 65
3.7 Vorticity and potential vorticity **66**

3.7.1 Vorticity evolution 66
3.7.2 Potential vorticity evolution 67
3.7.3 Hydrostatic-Boussinesq potential vorticity 68
3.7.4 Planetary geostrophic potential vorticity 70
3.8 Dynamics of fluid columns 71
3.8.1 An approximate hydrostatic pressure 71
3.8.2 Vertically integrated momentum 72
3.9 Particle dynamics 73
3.9.1 A fluid parcel and a point particle 74
3.9.2 Lagrangian function 74
3.9.3 Hamiltonian dynamics 75
3.9.4 Conservation laws 77
3.10 Symmetry and conservation laws 79
3.10.1 Motion on an infinite plane 79
3.10.2 Angular momentum about the north pole 80
3.10.3 Advective and frictional metric terms 81

Hydrodynamical equations describing the ocean circulation are based on the use of Newton's Laws for a stratified fluid continuum moving on a rotating planet. Two common approximations are made to the hydrodynamical equations for use in large-scale ocean climate: (1) the fluid maintains hydrostatic balance, (2) the fluid comprises a shallow layer on a rotating sphere. With these assumptions, the fundamental equations are termed the *primitive equations*, where *primitive* refers to the computation of velocity time tendencies, rather than time tendencies for derived fields such as vorticity.

3.1 Orthogonal coordinates

The ocean primitive equations are based on taking the *Traditional Approximation* to the Navier-Stokes equations. This approximation represents a statement about the spherical geometry on which fluid parcels move. We summarize here some mathematical results from Part VI of use in the following.

3.1.1 Spherical and Cartesian coordinates

The World Ocean forms a thin layer of fluid moving on the rotating earth. For purposes of ocean climate modeling, it is sufficient to approximate the earth as a sphere (see Veronis (1973) for a discussion). Hence, a key goal of this chapter is to formulate the dynamical equations describing fluid motion on a rotating sphere. Notably, the derivations must carefully account for details of the sphere's non-Euclidean geometry (see Chapters 18 and 19 for more on differential geometry). Such constitutes a fundamental distinction from non-geophysical formulations of hydrodynamics (e.g., Landau and Lifshitz (1987)), where it may be more appropriate to work in three-dimensional Euclidean space. A rigorous and lucid treatment of fluid motion on arbitrary smooth surfaces is given by Aris (1962).

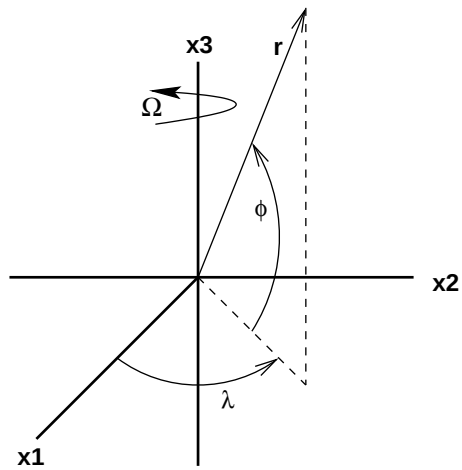

Figure 3.1: A schematic of the coordinates used for describing fluid dynamics on a rotating sphere, where the rotation axis is aligned through the north pole of the sphere. The coordinate $0 \leq \lambda \leq 2\pi$ is the longitude, with positive values measured eastward from Greenwich, England. The coordinate ϕ is the latitude, with values $\phi = 0$ at the equator and $\phi = \pi/2(-\pi/2)$ at the north (south) poles. The radial distance r is measured here with respect to the center of the sphere. The explicit coordinate transformations are $x^1 = r \cos\phi \cos\lambda$, $x^2 = r \cos\phi \sin\lambda$, and $x^3 = r \sin\phi$. Note that for many idealized geophysical fluid studies, Cartesian coordinates refer to those defined locally to a tangent plane at some point on the surface of the rotating sphere. Such β-plane or f-plane coordinates (e.g., Gill (1982); Pedlosky (1987)) are distinct from the Cartesian coordinates defined here.

Figure 3.1 provides a schematic illustrating the relation between spherical and Cartesian coordinates of use for describing fluid dynamics on a rotating sphere. The axis of rotation passes from the southern pole to the northern pole of the sphere. Both sets of coordinates are fixed on the rotating sphere (non-inertial coordinates). In the next subsection, we note that most modern global ocean climate models generalize the angular coordinates (λ, ϕ) to allow arbitrary locally orthogonal coordinates to specify angular positions on the sphere.

3.1.2 Horizontal orthogonal coordinates

Throughout this book, we employ generalized horizontal coordinates, where *horizontal* means coordinates within a locally defined tangent plane on the surface of a spherical earth (see Figure 3.1). As discussed in Chapter 19, the use of orthogonal curvilinear coordinates (ξ^1, ξ^2) allows for the squared infinitesimal distance between two points in the ocean to be written as

$$(\mathrm{d}s)^2 = (h_1 \, \mathrm{d}\xi^1)^2 + (h_2 \, \mathrm{d}\xi^2)^2 + (\mathrm{d}z)^2, \tag{3.1}$$

where the metric, or stretching functions, h_1 and h_2 are non-negative. In terms of the dimensionful *physical* horizontal distances

$$\mathrm{d}x = h_1 \, \mathrm{d}\xi^1 \tag{3.2}$$
$$\mathrm{d}y = h_2 \, \mathrm{d}\xi^2, \tag{3.3}$$

the line element takes the compact form

$$(\mathrm{d}s)^2 = (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2, \tag{3.4}$$

the volume of an infinitesimal parcel is given by

$$\begin{aligned} \mathrm{d}V &= (h_1 \, \mathrm{d}\xi^1)\,(h_2 \, \mathrm{d}\xi^2)\,\mathrm{d}z \\ &= \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z, \end{aligned} \tag{3.5}$$

and the physical components of the horizontal partial derivatives are

$$\partial_x = h_1^{-1} \, \partial_1 \tag{3.6}$$
$$\partial_y = h_2^{-1} \, \partial_2. \tag{3.7}$$

Figure 3.2 illustrates these formulae. Notably, although the introduction of physical displacements brings the metric tensor into a form analogous to that for three dimensional Euclidean space, the non-Euclidean nature of the sphere manifests by the non-vanishing commutator

$$\begin{aligned} \left[\partial_x, \partial_y\right] &= \partial_x \, \partial_y - \partial_y \, \partial_x \\ &= (\partial_x \, \ln \, \mathrm{d}y) \, \partial_y - (\partial_y \, \ln \, \mathrm{d}x) \, \partial_x, \end{aligned} \tag{3.8}$$

which vanishes only when the horizontal geometry is flat instead of curved. In particular, the use of spherical coordinates leads to

$$\left[\partial_x, \partial_y\right] = \left(\frac{\tan \phi}{R}\right) \partial_x. \tag{3.9}$$

3.2 Geometry of the Traditional Approximation

The Traditional Approximation assumes that the metric functions h_1 and h_2 are dependent *only* on the horizontal coordinates (ξ^1, ξ^2). Radial dependence of the metric functions is reduced to a constant radial factor R, where $R = 6.371 \times 10^6$ m is the radius of a sphere with the same volume as the earth. Therefore, distances, used to compute partial and covariant derivatives and areas and volumes, are determined by a metric tensor whose components are functions only of the horizontal position on the sphere.

This approximation is accurate for many purposes because the ocean forms a thin layer of fluid moving on the outer part of the earth. For example, vertical motions of fluid parcels in stratified regions of the World Ocean occur over distances of only a few centimeters to meters per day, which is far smaller than the typical

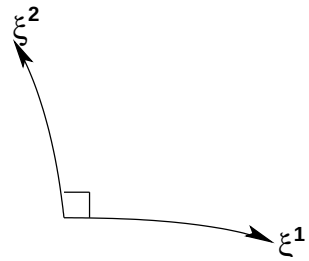

Figure 3.2: Illustrating the generalized horizontal orthogonal coordinates of use for global ocean models. The coordinate lines intersect at right angles, but generally do not follow lines parallel to constant longitude or latitude on the sphere. An infinitesimal horizontal region has area given by $\mathrm{d}A = (h_1\,\mathrm{d}\xi^1)\,(h_2\,\mathrm{d}\xi^2) = \mathrm{d}x\,\mathrm{d}y$. For spherical coordinates $(\xi^1, \xi^2) = (\lambda, \phi)$, the infinitesimal horizontal distances are $\mathrm{d}x = (r\,\cos\phi)\,d\lambda$ and $\mathrm{d}y = r\,d\phi$.

horizontal speeds that reach up to a meter per second. For larger vertical deviations, such as those occuring in convective regions, the Traditional Approximation may be unwarranted (see Marshall et al. (1997) for a discussion).

In the remainder of this chapter, we derive the differential equations of motion for fluid parcels moving in the space defined by the Traditional Approximation metric tensor. Notably, the resulting energy and angular momentum conservation laws follow as a consequence of casting the dynamical laws within this geometry. Marshall et al. (1997) summarize controversies related to the precise form of energy and angular momentum conservation laws for a fluid based on the Traditional Approximation. Maintaining a focus on the geometric essence of the Traditional Approximation clarifies these controversies. We provide added comment on this point in Section 3.9.

3.3 Volume and mass conserving kinematics

As discussed in Section 2.4.3, Boussinesq fluid parcels conserve their volume, whereas non-Boussinesq fluid parcels conserve their mass. In this section, we highlight the kinematic issues related to volume and mass conservation. In turn, we derive the kinematic boundary conditions applicable for Boussinesq and non-Boussinesq fluids.

3.3.1 Mass and volume conservation for parcels

To get started, let us summarize the discussion given in Section 2.3.4 where we derived constraints that the velocity field must satisfy for mass and volume conserving fluid parcels. With the mass of a parcel of seawater* given by $\mathrm{d}M = \rho\,\mathrm{d}V$, mass conservation for the parcel leads to

$$\rho_{,t} + \nabla \cdot (\mathbf{u}\,\rho) + (w\,\rho)_{,z} = 0, \tag{3.10}$$

*The mass of a parcel of seawater includes the mass of freshwater plus the mass of dissolved salts.

whereas volume conservation leads to

$$\nabla \cdot \mathbf{u} + w_{,z} = 0. \tag{3.11}$$

Both of these equations are referred to as *continuity equations*, with the relevant one implied by knowing whether the fluid is mass or volume conserving. We employ these results in the following discussion of mass and volume conservation applied over finite regions of fluid.

Boussinesq fluid parcels have mass given by $\mathrm{d}M = \rho_o \, \mathrm{d}V$ for purposes of developing the momentum and tracer budgets, except when the mass is multiplied by gravity, at which point they have the non-Boussinesq $\mathrm{d}M = \rho \, \mathrm{d}V$. This formulation is somewhat *ad hoc*, although it has proven to be quite useful. The *ad hoc* nature is clearly revealed when considering the gravitational potential energy budget in Chapter 4. In this case, volume conserving Boussinesq parcel's can have their potential energy budgets affected by changes in the parcel's mass: $D(\rho \, \mathrm{d}V)/Dt = \mathrm{d}V \, (D\rho/Dt) \neq 0$. These changes in mass are termed *virtual mass fluxes* due to their unphysical nature.

3.3.2 Volume conservation for finite domains

The purpose of this section is to apply the constraint of volume conservation over a finite control volume, where the surface of the volume represents the free ocean surface. The result will be expressions for the time tendency of the free ocean surface as well as kinematic boundary conditions satisfied at the ocean surface and bottom.

For this purpose, consider the control volume of fluid shown in Figure 3.3. The position $z = \eta$ is the time dependent vertical deviation from a resting ocean state $z = 0$, and $z = -H$ is the stationary position of the impenetrable bottom topography. The volume of fluid within the control volume is given by

$$\begin{aligned} V &= \int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z \\ &= \int D \, \mathrm{d}A, \end{aligned} \tag{3.12}$$

where

$$D = H + \eta \tag{3.13}$$

is the vertical thickness of a fluid column as a function of horizontal position and time, and $\mathrm{d}A = \mathrm{d}x \, \mathrm{d}y$ is the horizontal cross-sectional area element. Assuming the control volume is filled with volume conserving fluid parcels, conservation of the domain volume implies that the volume's time tendency

$$V_{,t} = \int \mathrm{d}A \, D_{,t} \tag{3.14}$$

equals the sum of all volume fluxes passing across the domain boundaries

$$V_{,t} = \int \mathrm{d}A \left(q_w + \hat{\mathbf{N}} \cdot \mathbf{v}(-H) - \nabla \cdot \int_{-H}^{\eta} \mathrm{d}z \, \mathbf{u} \right). \tag{3.15}$$

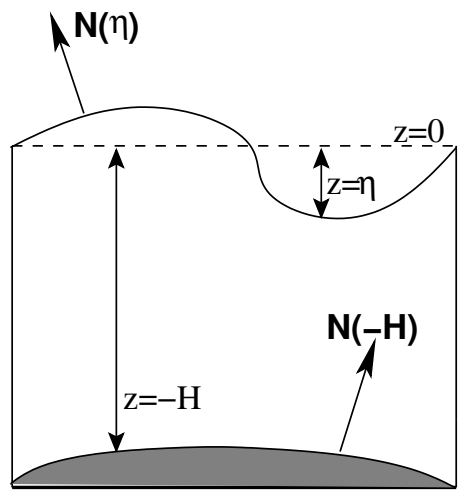

Figure 3.3: A control volume of fluid for use in discussing volume and mass conservation and kinematic boundary conditions. The vertical side-boundaries are assumed stationary and penetrable, the bottom topography at $z = -H$ is generally non-flat, impenetrable, and stationary, and the free surface at $z = \eta$ is generally non-flat, penetrable, and time dependent. The surface orientation vector is given by $\hat{\mathbf{N}} = (-\nabla\eta, 1)$. To a high degree of accuracy for climate modeling purposes, $\hat{\mathbf{N}} \approx \hat{\mathbf{z}}$, although we do not rely on this approximation when formulating the boundary conditions. The bottom orientation vector is given by $\hat{\mathbf{N}} = (\nabla H, 1)$. Undulations of the surface height are exagerated here for illustrative purposes. In the real ocean, topographic variations can reach from the ocean bottom to top (thousands of meters), whereas the surface height variations are on the order of a few centimeters to meters. Furthermore, for mathematical simplicity, the bottom and surface boundaries are assumed to never "overturn" or break. Hence, the orientation vector always has a positive vertical component. In particular, this restriction precludes our formulation from providing an explicit framework to study breaking surface waves.

Convergence of the horizontal flux captures the contribution from volume entering horizontally through the sides of the control volume. The term $\int \mathrm{d}A\, \hat{\mathbf{N}} \cdot \mathbf{v}(-H)$ is the tendency of volume entering through the bottom topography, with

$$\hat{\mathbf{N}} = (\nabla H, 1) \tag{3.16}$$

an orientation vector pointing from the bottom topography into the ocean domain (see Figure 3.3). However, $\hat{\mathbf{N}} \cdot \mathbf{v}(-H)$ vanishes since the bottom is assumed to be impenetrable, and so the velocity satisfies a no-normal flow or bottom kinematic boundary condition

$$\hat{\mathbf{N}} \cdot \mathbf{V} = \mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H. \tag{3.17}$$

In contrast, $q_w\, \mathrm{d}A$ is the generally nonzero tendency of volume crossing the penetrable free surface, with $q_w > 0$ signaling volume entering the ocean domain. In a volume conserving model, $q_w\, \mathrm{d}A$ is assumed equivalent to the volume of fresh water per unit time vertically crossing the free surface

$$\begin{aligned} q_w &= \hat{\mathbf{z}} \cdot \hat{\mathbf{N}} \left(\frac{\text{volume/time}}{\mathrm{d}A} \right) \\ &= \hat{\mathbf{z}} \cdot \hat{\mathbf{N}}\, (P - E + R). \end{aligned} \tag{3.18}$$

In this equation,

$$\hat{\mathbf{N}} = (-\nabla \eta, 1) \tag{3.19}$$

is an orientation vector at the ocean surface (see Figure 3.3). $P - E + R$ represents the volume input from precipitation, evaporation, and/or river runoff.

Equating the volume tendencies given by equations (3.14) and (3.15) leads to the balance of thickness within each vertical fluid column in the control volume

$$\eta_{,t} + \nabla \cdot \left(\int_{-H}^{\eta} \mathrm{d}z\, \mathbf{u} \right) = q_w. \tag{3.20}$$

Defining

$$\mathbf{U} = \int_{-H}^{\eta} \mathrm{d}z\, \mathbf{u} \tag{3.21}$$

as the vertically integrated horizontal velocity leads to the surface height equation for a volume conserving fluid

$$\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w. \tag{3.22}$$

The convergence term affects surface height values via changes in the dynamical fields within the ocean model, whereas the fresh water term affects surface height via changes in boundary forcing. Figure 3.4 illustrates this balance.

Integrating the free surface height equation (3.22) horizontally over the full ocean domain leads to the volume balance

$$\partial_t \left(\int \mathrm{d}A\, D \right) = \int \mathrm{d}A\, q_w. \tag{3.23}$$

Therefore, the ocean's volume changes only due to the passage of fresh water across the free surface. Changes in the interior dynamics cancel out when integrated over the full domain. This is an important relation to maintain in the numerical Boussinesq ocean model to ensure that total volume is indeed conserved.

Applying the horizontal divergence operator onto the vertical integral in equation (3.20) leads to

$$(\partial_t + \mathbf{u}(\eta) \cdot \nabla)\, \eta + \mathbf{u}(-H) \cdot \nabla H + \int_{-H}^{\eta} \mathrm{d}z\, \nabla \cdot \mathbf{u} = q_w. \tag{3.24}$$

As discussed in Section 3.3.1, volume conserving fluid parcels have a velocity field satisfying $\nabla \cdot \mathbf{u} + w_{,z} = 0$ (equation (3.11)). Use of this divergence-free condition, as well as the bottom kinematic boundary condition (3.17), yields the surface kinematic boundary condition for a volume conserving fluid

$$(\partial_t + \mathbf{u} \cdot \nabla)\, \eta = q_w + w \quad \text{at } z = \eta. \tag{3.25}$$

This boundary condition can be written in the equivalent form

$$\eta_{,t} = \hat{\mathbf{N}} \cdot \mathbf{v} + q_w \quad \text{at } z = \eta, \tag{3.26}$$

where $\hat{\mathbf{N}} = (-\nabla\eta, 1)$ is the surface orientation vector. We see that local time tendencies of the surface height are driven by fresh water fluxes at the surface as well as the projection of the surface velocity onto the surface orientation vector. Comparing the boundary condition (3.26) with the column integrated budget (3.22) leads to the balance

$$\hat{\mathbf{N}} \cdot \mathbf{v} = -\nabla \cdot \mathbf{U} \quad \text{at } z = \eta. \tag{3.27}$$

This expression says that convergence of the vertically integrated transport in a Boussinesq fluid is balanced by the projection of the surface velocity onto the surface orientation vector. For a *rigid lid* Boussinesq ocean (see Section 2.4.4), $\nabla \cdot \mathbf{U} = 0$ which then leads to $\hat{\mathbf{N}} \cdot \mathbf{v} = 0$. We have more to say about the rigid lid ocean in Section 10.8.

3.3.3 Mass conservation for finite domains

We now proceed in a manner parallel to the previous section, yet here consider a control volume full of mass conserving fluid parcels. The total mass of fluid inside the control volume is given by

$$M = \int \mathrm{d}A \int_{-H}^{\eta} \rho\, \mathrm{d}z, \tag{3.28}$$

where ρ is the seawater's mass density. Mass conservation implies that the time tendency of mass

$$M_{,t} = \int \mathrm{d}A\, \partial_t \left(\int_{-H}^{\eta} \mathrm{d}z\, \rho \right) \tag{3.29}$$

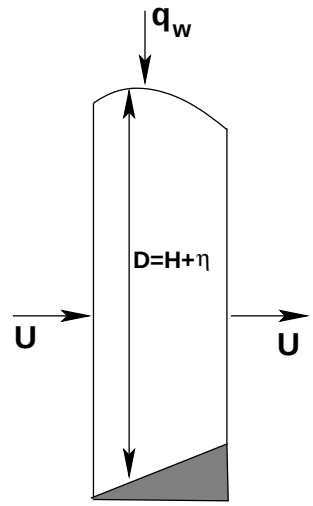

Figure 3.4: The balance of volume within a column of Boussinesq fluid extending from the free surface to the ocean bottom is determined by three processes as given by equation (3.23): $\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w$. These effects are (1) the input of volume across the ocean surface from fresh water, (2) the convergence of volume associated with the vertically integrated horizontal currents, (3) and the time dependent fluctuations of the ocean surface height.

changes due to imbalances in the mass flux passing across the boundaries

$$M_{,t} = \int \mathrm{d}A \left(q_w + \hat{\mathbf{N}} \cdot \mathbf{v}(-H)\, \rho(-H) - \nabla \cdot \int_{-H}^{\eta} \mathrm{d}z\, \rho\, \mathbf{u} \right). \tag{3.30}$$

In this equation, $\rho(-H)$ is the density of water entering the domain through the bottom topography at $z = -H$. Yet, as in the volume conserving case, the bottom is assumed impenetrable, thus leading to the no-normal flow condition $\hat{\mathbf{N}} \cdot \mathbf{v}(-H) = 0$. The term

$$\begin{aligned} q_w &= \hat{\mathbf{z}} \cdot \hat{\mathbf{N}} \left(\frac{\text{mass/time}}{dA} \right) \\ &= (\hat{\mathbf{z}} \cdot \hat{\mathbf{N}})\, (P - E + R)\, \rho_w \\ &= q_w\, \rho_w \end{aligned} \tag{3.31}$$

represents the mass flux of fresh water (mass per unit time per unit area) crossing the free surface. Equivalently, it is the vertical momentum density of the fresh water. Equating the time tendencies given by equations (3.29) and (3.30) leads to a mass balance within each vertical column of fluid in the control volume

$$\partial_t \left(\int_{-H}^{\eta} \mathrm{d}z\, \rho \right) + \nabla \cdot \left(\int_{-H}^{\eta} \mathrm{d}z\, \rho \mathbf{u} \right) = q_w\, \rho_w. \tag{3.32}$$

Likewise, integrating horizontally over the full ocean domain and using no-flux or periodic side boundary conditions leads to the mass balance

$$\partial_t \left(\int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z \, \rho \right) = \int_{z=\eta} \mathrm{d}A \, q_w \, \rho_w. \tag{3.33}$$

Hence, the ocean's mass changes only due to the passage of fresh water across the free surface. This is an important relation to check in the numerical non-Boussinesq model to ensure that total mass is indeed conserved. It is the non-Boussinesq analog to the volume conservation statement given by equation (3.23).

In deriving the free surface equation for a mass conserving fluid, it is useful to introduce the density weighted velocity field

$$\rho_o \, \tilde{\mathbf{v}} = \rho \, \mathbf{v} \tag{3.34}$$

and the vertical integral of its horizontal components

$$\tilde{\mathbf{U}} = \int_{-H}^{\eta} \mathrm{d}z \, \tilde{\mathbf{u}}. \tag{3.35}$$

Notably, $\rho \, \mathbf{v} = \rho_o \, \tilde{\mathbf{v}}$ represents a mass flux (mass per unit time per unit area), and so $\tilde{\mathbf{v}}$ is the mass flux normalized by the constant Boussinesq reference density

$$\rho_o = 1035 \, \mathrm{kg/m^3} \tag{3.36}$$

(see page 47 of Gill (1982)). Equivalently, $\rho_o \, \tilde{\mathbf{v}}$ is the linear momentum density (momentum per unit volume) carried by a fluid parcel. Likewise, $\rho_o \, \tilde{\mathbf{U}}$ is the vertically integrated horizontal mass flux, or vertically integrated horizontal momentum density. With these definitions, the mass balance (3.32) takes the form

$$\partial_t \left(\int_{-H}^{\eta} \mathrm{d}z \, \rho \right) = -\rho_o \, \nabla \cdot \tilde{\mathbf{U}} + q_w \, \rho_w. \tag{3.37}$$

This equation should be compared to the analogous result (3.22) valid for Boussinesq fluids.

To isolate the surface height tendency, and thus to highlight the main differences from equation (3.22), two approaches are apparent. One approach applies the time derivative in equation (3.37) to render

$$\rho(\eta) \, \eta_{,t} = -\rho_o \, \nabla \cdot \tilde{\mathbf{U}} + q_w \, \rho_w - \int_{-H}^{\eta} \mathrm{d}z \, \rho_{,t}. \tag{3.38}$$

An alternative arises from introducing the vertically averaged density

$$\overline{\rho}^z = \frac{1}{H + \eta} \int_{-H}^{\eta} \mathrm{d}z \, \rho, \tag{3.39}$$

thus leading to the surface height equation for a mass conserving fluid

$$\overline{\rho}^z \, \eta_{,t} = -\rho_o \, \nabla \cdot \tilde{\mathbf{U}} + \rho_w \, q_w - D \, \partial_t \, \overline{\rho}^z, \tag{3.40}$$

where $D = H + \eta$ is the total thickness of the fluid column. Figure 3.5 illustrates the terms in this equation.

Comparison of equation (3.40) with the analogous surface height equation (3.22) for a volume conserving fluid reveals the following main differences. First, it is the convergence $-(\rho_o/\overline{\rho}^z) \, \nabla \cdot \tilde{\mathbf{U}}$ rather than $-\nabla \cdot \mathbf{U}$ that drives the surface height in mass conserving fluids, and the density ratio $(\rho_w/\overline{\rho}^z)$ weights the fresh water contribution. These two processes have direct analogs with the Boussinesq fluid, where the convergence term arises from changes in the vertically integrated dynamics, and the fresh water term arises from boundary forcing. Additionally, $-D \, \partial_t \ln \overline{\rho}^z$ represents a fundamentally new dynamical process that expresses a change in free surface height upon changing the vertically averaged density. If the vertically averaged density decreases, the surface height increases, and vice versa. This *steric effect* is absent from the Boussinesq fluid's surface height equation, thus necessitating a diagnostic correction to compute the absolute surface height for sea level studies with Boussinesq models (Greatbatch (1994); Mellor and Ezer (1995)). However, as noted by Greatbatch et al. (2001), the diagnostic correction misses a potentially significant level of spatial variability in the surface height that is captured by the non-Boussinesq surface height. Such differences may be important for sea level studies, especially those associated with climate change.

To derive the surface kinematic boundary condition, we apply the time and space derivatives throughout equation (3.32) to yield

$$\rho(\eta) \, (\partial_t + \mathbf{u}(\eta) \cdot \nabla) \, \eta + \mathbf{u}(-H) \cdot \nabla H + \int_{-H}^{\eta} \mathrm{d}z \, (\rho_{,t} + \nabla \cdot (\mathbf{u} \, \rho)) = q_w \, \rho_w. \tag{3.41}$$

Recall from Section 3.3.1 that a mass conserving parcel of fluid satisfies $\rho_{,t} + \nabla \cdot (\mathbf{u} \, \rho) + \partial_z \, (w \, \rho) = 0$ (equation (3.10)). Using this result with the bottom kinematic boundary condition (3.17) leads to the surface kinematic boundary condition applicable for a mass conserving fluid

$$(\partial_t + \mathbf{u} \cdot \nabla) \, \eta = (\rho_w/\rho) \, q_w + w \quad \text{at } z = \eta. \tag{3.42}$$

Comparison with the analogous result (3.25) applicable to a volume conserving fluid reveals that the only difference is the density ratio (ρ_w/ρ) weighting the fresh water flux. Introducing the linear momentum density $\rho_o \, \tilde{\mathbf{v}} = \rho \, \mathbf{v}$ leads to the equivalent form

$$(\rho \, \partial_t + \rho_o \, \tilde{\mathbf{u}} \cdot \nabla) \, \eta = \rho_w \, q_w + \rho_o \, \tilde{w} \quad \text{at } z = \eta. \tag{3.43}$$

Finally, introducing the surface orientation vector $\hat{\mathbf{N}} = (-\nabla \eta, 1)$ leads to

$$\rho \, \eta_{,t} = \rho_o \, \hat{\mathbf{N}} \cdot \mathbf{v} + \rho_w \, q_w \quad \text{at } z = \eta. \tag{3.44}$$

Comparing the column integrated mass budget (3.38) with the surface kinematic boundary condition (3.44) leads to the balance

$$\rho_o \, \hat{\mathbf{N}} \cdot \mathbf{v} = -\rho_o \, \nabla \cdot \tilde{\mathbf{U}} - \int_{-H}^{\eta} \mathrm{d}z \, \rho_{,t} \quad \text{at } z = \eta. \tag{3.45}$$

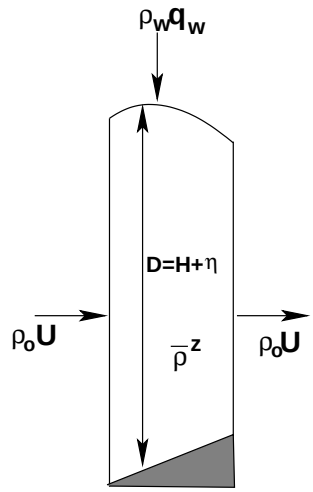

Figure 3.5: The balance of mass within a column of non-Boussinesq fluid extending from the free surface to the ocean bottom is determined by four processes as given by equation (3.40): $\overline{\rho}^z \, \eta_{,t} = -\rho_o \, \nabla \cdot \tilde{\mathbf{U}} + \rho_w \, q_w - h \, \partial_t \, \overline{\rho}^z$. These effects are (1) the input of mass across the ocean surface from fresh water, (2) the convergence of mass associated with the vertically integrated horizontal linear momentum, (3) time tendencies in the vertically averaged density, (4) and the time dependent fluctuations of the ocean surface height.

This expression says that projection of the surface velocity onto the surface orientation vector is balanced by convergence of vertically integrated momentum per volume and the depth integrated time tendencies in the seawater density. The second term, as noted already, accounts for the steric effect of density changes within a fluid column, and this term is absent in the Boussinesq fluid.

3.4 Principles of continuum dynamics

The purpose of this section is to summarize some general dynamical principles forming the basis of classical continuum mechanics. The book by Aris (1962) is recommended for a more thorough discussion. By establishing these ideas, we help to support our application of Newton's Laws of motion to infinitesimal fluid parcels in Section 3.5.

3.4.1 Position, velocity, and acceleration

The position of a parcel of a continuous media is described by a general coordinate $\vec{\xi} = (\xi^1, \xi^2, \xi^3)$, with $\xi^3 = z$ the choice for z-coordinate ocean models. As discussed in Chapter 18, coordinates carry contravariant labels, hence the upstairs placement of the indices on ξ^a. In other words, the position of a parcel is given by the *vector* position $\vec{\xi} = \xi^a \, \vec{e}_a$ with respect to an appropriate set of basis vectors $\vec{e}_a$. Note that the arrow over an object, thus symbolizing a vector, is typically replaced in the following by the boldface for brevity in notation.

The time rate of change of the position defines the parcel's velocity. As time does not carry tensorial properties in non-relativistic dynamics, velocity $\mathbf{v} = \mathrm{d}\boldsymbol{\xi}/\mathrm{d}t = \mathrm{d}(\xi^a \vec{e}_a)/\mathrm{d}t$ is also a vector. Likewise, the parcel's acceleration $\mathbf{a} = \mathrm{d}\mathbf{v}/\mathrm{d}t$ is a vector. In order for Newton's law of motion to be tensorially consistent, it follows that the forces per unit mass acting on a parcel, which drive the acceleration, are vector quantities. The tensorial properties of the equations for fluid flow are based on these statements about position, velocity, and acceleration.

3.4.2 External/body and internal/contact forces

The force vectors acting on an arbitrary volume of a continuous media are of two kinds. The first are *external or body* forces, such as gravitation, Coriolis, and electromagnetic forces. These forces act throughout the extent of the media. The total body force acting on the volume is given by the integral of the external force per unit mass $\mathbf{f}$ multiplied by the mass of the media

$$\mathbf{F}_{ext} = \int \mathbf{f}\,\rho\,\mathrm{d}V. \tag{3.46}$$

For example, the gravitational force acting on a volume of fluid is given by

$$\mathbf{F}_{gravity} = \int \mathbf{g}\,\rho\,\mathrm{d}V, \tag{3.47}$$

where $\mathbf{g}$ is the acceleration of gravity, or equivalently the gravitational force per unit mass.

The second kind of forces are *internal or contact* forces, such as pressure forces and frictional forces. These forces act on a volume of the continuous media through affecting boundaries of the volume. Note that point particles, having no finite extent, do not experience internal or contact forces. Cauchy's Stress Principle (e.g., Chapter 5 of Aris (1962)) states that the force per unit area exterted by the material outside the volume is a function of the position within the volume, the time, and the orientation of the boundaries of the surface. We do not justify this principle since its consequences are familiar and intuitive. For internal boundaries, the total contact force exerted on the volume V through its boundaries ∂V is given by

$$\mathbf{F}_{int} = \int \mathbf{t}_{(\hat{\mathbf{n}})}\,\mathrm{d}A \tag{3.48}$$

where $\hat{\mathbf{n}}$ is the outward normal* to the domain boundary, $\mathbf{t}_{(\hat{\mathbf{n}})} = \mathbf{t}_{(\hat{\mathbf{n}})}(\boldsymbol{\xi}, t, \hat{\mathbf{n}})$ is the force per unit area, or *contact stress*, exterted on the boundary ∂V by the surrounding material, and $\mathrm{d}A$ is the area element of the boundary. The functional dependence of the contact stress on the normal direction can be solved through invoking Newton's Second Law, as now described.

3.4.3 Local equilibrium of contact stresses

The conservation of linear momentum (i.e., Newton's Second Law) states that the sum of the external and internal forces acting on a fluid volume are equal to the acceleration

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \mathbf{v}\,\rho\,\mathrm{d}V = \int \mathbf{f}\,\rho\,\mathrm{d}V + \int \mathbf{t}_{(\hat{\mathbf{n}})}\,\mathrm{d}A. \tag{3.49}$$

*A normal to a surface is a one-form, as discussed in Chapter 18

There is an important consequence of this relation that follows from allowing the volume to smoothly go to zero without changing the shape. Assuming the integrands are well behaved, the volume integrals vanish as L^3, where L is a characteristic dimension of the volume. The surface integral, however, vanishes only as L^2. Consequently, dividing equation (3.49) by L^2 yields the relation

$$\lim_{L\to 0} \frac{1}{L^2} \int \mathbf{t}_{(\hat{\mathbf{n}})} \, \mathrm{d}A = 0. \tag{3.50}$$

This result implies that the *stresses acting on the volume are locally in equilibrium*. In other words, the sum of the internal stresses acting on all sides of an infinitesimal volume vanish. In particular, $\mathbf{t}_{(\hat{\mathbf{n}})} = -\mathbf{t}_{-(\hat{\mathbf{n}})}$, which is a statement of Newton's Third Law.

3.4.4 The stress tensor

We now apply the local equilibrium property to an infinitesimal volume with four flat sides, three of which are aligned along the negative coordinate axes directions, and the fourth directed in an arbitrary slanted direction (see Figure 5.1 of Aris (1962)) defined by the normal $\hat{\mathbf{n}} = n_a \, \tilde{e}^a$, with $\tilde{e}^a$ for $a = 1, 2, 3$ representing a basis of one-forms (see Section 18.6 for discussion of one-forms). Local equilibrium implies

$$\mathbf{t}_{(\hat{\mathbf{n}})} \, \mathrm{d}A - \mathbf{t}_{(1)} \, \mathrm{d}A_1 - \mathbf{t}_{(2)} \, \mathrm{d}A_2 - \mathbf{t}_{(3)} \, \mathrm{d}A_3 = 0, \tag{3.51}$$

where $\mathrm{d}A$ is the area of the slanted face, $\mathrm{d}A_a$ are the areas of the other three faces, and each face has a corresponding force per unit area directed normal to the respective face. The negative signs arise from the projection of the areas onto the negative coordinate directions. Now the projection of the areas $\mathrm{d}A_a$ onto the direction defined by $\hat{\mathbf{n}}$ is given by $\mathrm{d}A_a = n_a \, \mathrm{d}A$. Hence, equation (3.51) takes the form

$$\mathbf{t}_{(\hat{\mathbf{n}})} = \mathbf{t}_{(1)} \, n_1 + \mathbf{t}_{(2)} \, n_2 + \mathbf{t}_{(3)} \, n_3. \tag{3.52}$$

As such, the stress in an arbitrary direction defined by $\hat{\mathbf{n}}$ can be written as the linear sum of the three stresses projected along the coordinate axes, each weighted by the projection of the normal along those axes. We can summarize this equation by introducing T^{ba} as the $a'th$ component of the stress $\mathbf{t}_{(b)}$, and $t^b_{(\hat{\mathbf{n}})}$ as the b'th component of the stress $\mathbf{t}_{(\hat{\mathbf{n}})}$. With this notation, local equilibrium implies the stress along a direction defined by $\hat{\mathbf{n}}$ is linearly related to $\hat{\mathbf{n}}$ through

$$t^b_{(\hat{\mathbf{n}})} = T^{ba} \, n_a. \tag{3.53}$$

Since $t^b_{(\hat{\mathbf{n}})}$ are components to a vector, and n_a are components to a one-form, the linear proportionality coefficients T^{ba} form the components to a second order *stress tensor*. We are mostly concerned with two types of stress: diagonal stresses associated with pressure, and stresses associated with friction. We discuss pressure more in this chapter, and Chapter 15 details the symmetric stresses associated with horizontal deformations of fluid parcels.

The reasoning leading to the stress tensor implies that the internal stresses $\mathbf{t}_{(\hat{\mathbf{n}})}$, which are generally functions $\mathbf{t}_{(\hat{\mathbf{n}})}(\boldsymbol{\xi}, t, \hat{\mathbf{n}})$ of space, time, and the normals, are actually quite simple in that they can be summarized by nine quantities $T^{ab}(\boldsymbol{\xi}, t)$ that

provide the linear proportionality between stresses and the normals. Additionally, since the result of this reasoning led to a proper tensorial or covariant expression (3.53), the result remains form invariant under an arbitrary change in coordinates.

3.4.5 Equations of motion

The relation (3.53) between the contact stresses and the stress tensor brings the equations of motion (3.49) to the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \rho\, v^a\, \mathrm{d}V = \int \rho\, f^a\, \mathrm{d}V + \int T^{ab}\, n_b\, \mathrm{d}A. \tag{3.54}$$

As described in Section 19.10, the covariant form of Gauss's Law, or the Divergence Theorem, can be applied to the area integral to yield

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \rho\, v^a\, \mathrm{d}V = \int \rho\, f^a\, \mathrm{d}V + \int T^{ab}_{;b}\, \mathrm{d}V, \tag{3.55}$$

where the semi-colon denotes a covariant derivative (Section 19.2.2). Since the volume under consideration is arbitrary, this integral relation is satisfied only if it is satisfied as a differential balance for infinitesimal fluid parcels. Such provides a theoretical basis for the derivations provided in Section 3.5.

3.5 Dynamics of fluid parcels

Given the theory developed in Section 3.4, we apply Newton's Laws with impunity to infinitesimal fluid parcels. That is, we derive the dynamical equations for a fluid parcel traveling on a general trajectory $\mathbf{x}(\zeta,\tau)$ with velocity $\mathbf{v} = \partial_\tau \mathbf{x}(\zeta,\tau)$, where (ζ,τ) is the material space-time coordinates of the parcel (see discussion in Section 2.3). We use Newton's Second Law by equating the time tendency of the fluid parcel's linear momentum $(\rho\,\mathrm{d}V)\,\mathbf{v}$ to forces acting on the parcel. Although this approach is Lagrangian in character, we derive Eulerian equations which consider time tendencies for the linear momentum density $\rho\,\mathbf{v} = \rho(\mathbf{x},t)\,\mathbf{v}(\mathbf{x},t)$ instead of the parcel's trajectory $\mathbf{x}(\zeta,\tau)$.

Many details in this section rely on the tensor analysis presented in Chapters 18 and 19. Nonetheless, these steps in the derivation can be easily skipped without sacrificing continuity in the overall presentation.

3.5.1 Time tendency of the linear momentum

The fluid parcel's linear momentum has a time tendency given by

$$\begin{aligned}\partial_\tau(\mathbf{v}\,\rho\,\mathrm{d}V) &= (\rho\,\mathrm{d}V)\,\partial_\tau\mathbf{v}\\ &= (\rho\,\mathrm{d}V)\,\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t},\end{aligned} \tag{3.56}$$

where $\partial_\tau = \mathrm{d}/\mathrm{d}t$ is the material time derivative following the fluid parcel, and $\partial_\tau(\rho\,\mathrm{d}V) = 0$ since the parcel conserves its mass. For parcel's moving on the sphere, it is necessary to use the tensor analysis tools presented in Chapters 18 and 19 to express the parcel's acceleration $\mathrm{d}\mathbf{v}/\mathrm{d}t$ in general coordinates fixed on the sphere.

Let us start by noting that a tensorial component of the parcel's acceleration is given by

$$\begin{aligned}\frac{\mathrm{d}v^a}{\mathrm{d}t} &= v^a_{,t} + v^b\, v^a_{;b} \\ &= v^a_{,t} + (v^b\, v^a)_{;b} - v^a\, v^b_{;b}\end{aligned} \tag{3.57}$$

where $a = 1, 2, 3$ is a label for the three spatial coordinates, and the semi-colon notation signifies the covariant derivative introduced in Section 19.2.2. The second step introduced the covariant divergence of the velocity $v^b_{;b}$ and this can be replaced by the following through use of mass conservation (3.10)

$$\rho\, v^b_{;b} = -(\partial_t + v^a\, \partial_a)\rho \tag{3.58}$$

to yield

$$\rho\, \frac{\mathrm{d}v^a}{\mathrm{d}t} = (\rho\, v^a)_{,t} + (\rho v^a\, v^b)_{;b}. \tag{3.59}$$

The product $\rho v^a\, v^b$ comprises the elements to a second order symmetric tensor. As shown in Section 19.6, the covariant divergence of such a tensor is given by

$$(\rho v^a\, v^b)_{;b} = \frac{1}{\sqrt{\mathcal{G}}}\, \left(\sqrt{\mathcal{G}}\, v^a\, v^b\right)_{,b} + \Gamma^a_{bc}\, v^b\, v^c. \tag{3.60}$$

In this expression, $\sqrt{\mathcal{G}} = h_1\, h_2\, h_3$ is the square-root of the metric tensor's determinant (see Section 3.1). Multiplication by h_a, without an implied sum over a, leads to the physical components of the density weighted acceleration

$$\begin{aligned}\rho\, \frac{\mathrm{d}v^{(a)}}{\mathrm{d}t} &= (\rho\, v^{(a)})_{,t} + \nabla\cdot(\rho\, v^{(a)}\, \mathbf{v}) \\ &\qquad + \rho\, v^b\, h_a\, (v^c\, \Gamma^a_{bc} - v^a \partial_b\, \ln h_a),\end{aligned} \tag{3.61}$$

where

$$v^{(a)} = (\mathbf{u}, w) = (u, v, w) \tag{3.62}$$

are the physical velocity components, each with dimensions length per time. Use of the Christoffel symbols from Section 19.2.2 leads to

$$\rho\, \frac{\mathrm{d}u}{\mathrm{d}t} = (\rho\, u)_{,t} + \nabla\cdot(\rho\, u\, \mathbf{v}) - \rho\, v\, (v\, \partial_x \ln h_2 - u\, \partial_y \ln h_1) \tag{3.63}$$

$$\rho\, \frac{\mathrm{d}v}{\mathrm{d}t} = (\rho\, v)_{,t} + \nabla\cdot(\rho\, v\, \mathbf{v}) + \rho\, u\, (v\, \partial_x \ln h_2 - u\, \partial_y \ln h_1) \tag{3.64}$$

$$\rho\, \frac{\mathrm{d}w}{\mathrm{d}t} = (\rho\, w)_{,t} + \nabla\cdot(\rho\, w\, \mathbf{v}). \tag{3.65}$$

As in Section 3.1, introduce the physical displacements $\mathrm{d}x = h_1\, \mathrm{d}\xi^1$, $\mathrm{d}y = h_2\, \mathrm{d}\xi^2$ and use $\partial_1(\mathrm{d}\xi^2) = \partial_2(\mathrm{d}\xi^1) = 0$. Doing so eliminates the stretching functions in favor of the physical displacements

$$\rho\, \frac{\mathrm{d}u}{\mathrm{d}t} = (\rho\, u)_{,t} + \nabla\cdot(\rho\, u\, \mathbf{v}) - \rho\, v\, \left(v\, \partial_x \ln \mathrm{d}y - u\, \partial_y \ln \mathrm{d}x\right) \tag{3.66}$$

$$\rho\, \frac{\mathrm{d}v}{\mathrm{d}t} = (\rho\, v)_{,t} + \nabla\cdot(\rho\, v\, \mathbf{v}) + \rho\, u\, \left(v\, \partial_x \ln \mathrm{d}y - u\, \partial_y \ln \mathrm{d}x\right) \tag{3.67}$$

$$\rho\, \frac{\mathrm{d}w}{\mathrm{d}t} = (\rho\, w)_{,t} + \nabla\cdot(\rho\, w\, \mathbf{v}), \tag{3.68}$$

which can be written in the short-hand form

$$\rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = (\rho\,\mathbf{v})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathbf{v}) + \mathcal{M}\,(\hat{\mathbf{z}}\wedge\rho\,\mathbf{v}), \tag{3.69}$$

where

$$\mathcal{M} = v\,\partial_x \ln \mathrm{d}y - u\,\partial_y \ln \mathrm{d}x \tag{3.70}$$

defines an *advective metric frequency*. In spherical coordinates where $\mathrm{d}x = (R\,\cos\phi)\,d\lambda$, $\mathrm{d}y = R\,d\phi$, $\partial_x = (R\,\cos\phi)^{-1}\partial_\lambda$, and $\partial_y = R^{-1}\partial_\phi$,

$$\mathcal{M} = (u/R)\,\tan\phi \tag{3.71}$$

and

$$\rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = (\rho\,\mathbf{v})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathbf{v}) + (u/R)\,\tan\phi\,(\hat{\mathbf{z}}\wedge\rho\,\mathbf{v}). \tag{3.72}$$

The advective metric frequency is typically quite small due to the size of the earth's radius

$$R = 6.371\times 10^6\,\mathrm{m}. \tag{3.73}$$

For example, a zonal current of $0.1\mathrm{m\,s^{-1}}$ at 45°N latitude has an advective metric frequency $\mathcal{M} \approx 1.6\times 10^{-8}\mathrm{s}^{-1}$, or roughly 10^{-4} times smaller than the Coriolis frequency $f = 1.03\times 10^{-4}\mathrm{s}^{-1}$ (Section 3.5.2).

The advective metric force density $\mathcal{M}\,(\hat{\mathbf{z}}\wedge\rho\,\mathbf{v})$ arises from the nontrivial metric functions h_1 and h_2, and it appears as an added rotational force analogous to the Coriolis force (Section 3.5.2). It arises from our choice of coordinate frame of reference, whereas the Coriolis force appears since we choose to work in a non-inertial rotating reference frame. As with the Coriolis force, the advective metric force does no work on a fluid parcel since it is perpendicular to the parcel's velocity: $\mathbf{v}\cdot(\hat{\mathbf{z}}\wedge\mathbf{v}) = 0$. Hence, it does not affect the parcel's kinetic energy. This result provides a good check on the manipulations, since work and kinetic energy are scalars, and so cannot be influenced by coordinate dependent metric components. Additionally, the frequency $\mathcal{M}$ vanishes for some coordinate choices and remains nonzero for others. Hence, $\mathcal{M}$ is not a tensor (see Section 19.1).

3.5.2 Motion on a rotating sphere

As discussed in Section 3.1, we approximate the earth as a sphere with angular velocity $\mathbf{\Omega}$ about an axis through the north pole (Figure 3.1). We introduce here the issues relevant for dynamics in such a rotating reference frame.

3.5.2.1 Earth's angular velocity

The earth's angular velocity is comprised of two main contributions: the spin of the earth about its axis, and the orbit of the earth about the sun (see Figure 3.6). Other astronomical motions can be neglected. Therefore, in the course of a single period of $24\times 3600 = 86400s$, the earth experiences an angular rotation of $(2\pi + 2\pi/365.24)$ radians. As such, the angular velocity of the earth is given by

$$\begin{aligned}\Omega &= \left(\frac{2\pi + 2\pi/365.24}{86400\mathrm{s}}\right)\\ &= \left(\frac{\pi}{43082}\right)\mathrm{s}^{-1}\\ &= 7.292\times 10^{-5}\mathrm{s}^{-1}.\end{aligned} \tag{3.74}$$

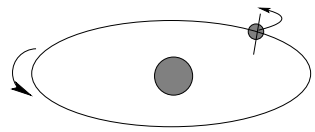

Figure 3.6: Illustrating the earth-sun system, whereby angular velocity of the earth is associated both with rotation about the earth's axis, and orbits about the sun. This angular velocity determines the strength of the Coriolis force.

To an extremely high degree of accuracy, this velocity is assumed constant in time for purposes of ocean climate modeling.

3.5.2.2 Inertial and non-inertial frames

The preferred frame of reference for describing geophysical fluids is that for terrestrial observers fixed on the rotating spherical planet. A rotating frame is not inertial, and so there are non-inertial terms appearing in the equations of motion that must be taken into account. To derive the relations between inertial and non-inertial observers, we employ the general operator relation (e.g., Marion and Thornton (1988) and Gill (1982))

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{inertial} = \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{rotating} + \mathbf{\Omega} \wedge . \tag{3.75}$$

Applying this relation to the position vector $\mathbf{r}$ of a fluid parcel leads to

$$\mathbf{v}_{inertial} = \mathbf{v} + \mathbf{\Omega} \wedge \mathbf{r}, \tag{3.76}$$

where $\mathbf{v} = \mathbf{v}_{rotating}$ is the parcel's velocity in the rotating frame. Note that because $\mathbf{r}$ is the position vector for a fluid parcel, the time derivative $\mathrm{d}/\mathrm{d}t$ is the material time derivative. Applying the relation twice leads to

$$\begin{aligned}\left(\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}t^2}\right)_{inertial} &= \left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{rotating} + \mathbf{\Omega}\wedge\right)\left(\left(\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right)_{rotating} + \mathbf{\Omega} \wedge \mathbf{r}\right) \\ &= \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} + 2\,\mathbf{\Omega} \wedge \mathbf{v} + \mathbf{\Omega} \wedge (\mathbf{\Omega} \wedge \mathbf{r}) \\ &= \dot{\mathbf{v}} + 2\,\mathbf{\Omega} \wedge \mathbf{v} - \nabla\,(\Omega\, r\, \cos\phi)^2/2.\end{aligned} \tag{3.77}$$

where $\dot{\mathbf{v}}$ is the parcel's acceleration in the rotating frame, and $\mathbf{\Omega}$ is assumed to be constant in time.

3.5.2.3 Effective gravitational force

Excursions by a fluid parcel in the radial direction away from the center of the sphere introduce changes to the particle's gravitational and centrifugal potential energies. It is useful to combine these two accelerations into an *effective gravitational force*.

The gravitational potential energy of a point mass m at a radius $r = z + R$ from the earth's center is given by

$$\begin{aligned} \Phi &= -\frac{G\,M_e\,m}{z+R} \\ &\approx -\frac{G\,M_e\,m}{R} + \left(\frac{G\,M_e}{R^2}\right)\,m\,z \\ &= \text{constant} + g_e\,m\,z, \end{aligned} \tag{3.78}$$

where

$$g_e = G\,M_e/R^2. \tag{3.79}$$

is the gravitational acceleration. Agreeing to measure the zero of gravitational potential energy at the earth's surface ($z = 0$) allows us to ignore the dynamically irrelevant constant term in Φ. Note that $G = 6.67 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}}$ is Newton's gravitational constant, $M_e = 5.98 \times 10^{24}\,\mathrm{kg}$ is the mass of the earth, and $R = 6.371 \times 10^6\,\mathrm{m}$ is the radius of a sphere with the same volume as the earth. The restriction of $z << R$ is motivated by our interest in fluid parcel motion in the ocean, where this inequality is maintained to a high degree of accuracy for all motions. For this reason, we can safely drop terms going as $(z/R)^2$ or higher order in the gravitational potential energy.

Combining the gravitational potential for a point particle at rest with that from the centrifugal force leads to the effective gravitational force acting on the particle (see e.g., Fetter and Walecka (1980) pages 40 and 48)

$$\begin{aligned} m\,\mathbf{g} &= -m\,\nabla(r\,g_e - (\Omega\,r\,\cos\phi)^2/2) \\ &= -m(g_e - r\,\Omega^2\cos^2\phi)\,\hat{\mathbf{r}} - (m\,r\,\Omega^2\,\cos\phi\,\sin\phi)\,\hat{\boldsymbol{\phi}}\,. \end{aligned} \tag{3.80}$$

The effective gravitational force is non-central due to the effects of rotation. Hence, if the earth were an ideal fluid, matter would flow from the poles towards the equator, thus ensuring that the earth's surface would everywhere be perpendicular to the effective gravitational acceleration $\mathbf{g}$. Indeed, the earth does exhibit a slight equatorial bulge, yet inhomogeneities in the earth's composition and surface loading by continents and glaciers make its shape differ from the ideal case.

For purposes of ocean climate modeling, and consistent with the Traditional Approximation, we make the following assumptions regarding the effective gravitational force. Firstly, we assume that the local vertical direction $\hat{\mathbf{z}}$ is defined by the direction of the effective gravitational acceleration

$$\begin{aligned} \mathbf{g} &= -\nabla\Phi \\ &= -g\,\hat{\mathbf{z}}, \end{aligned} \tag{3.81}$$

where

$$\Phi = g\,z \tag{3.82}$$

is the gravitational potential energy per mass. Secondly, we assume that the effective gravitational acceleration $\mathbf{g}$ has a magnitude given by

$$g = 9.7963\,\mathrm{m\,s^{-2}}, \tag{3.83}$$

which is the areal average of g_e taken over the surface of the ocean (see Section 3.5.2.4). Notably, we ignore changes in the effective gravitational acceleration associated with centrifugal accelerations or changes in the distance from the earth's center. Thirdly, we assume that the angular directions $(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\phi}})$ remain locally orthogonal to the local vertical direction $\hat{\mathbf{z}}$, thus neglecting non-sphericity of the earth.

3.5.2.4 Gravity for a global ocean model

The real earth is not perfectly spherical, yet we assume such for purposes of ocean climate modeling. Thus, consistent with the geometrical assumptions of the primitive equations (Section 3.1), it is appropriate to use a gravitational acceleration as determined by a mean over the ocean surface. As reported on page 132 of Moritz (2000), the gravitational acceleration at the earth's surface as a function of latitude is given by

$$g_e = 9.780327\,[1 + 0.0053024\,\sin^2\phi - 0.0000058\,(\sin 2\phi)^2]. \tag{3.84}$$

At a latitude of 45°, $g_e = 9.806\,\mathrm{m\,s^{-2}}$, which is the value used by ocean models based on Cox (1984). The areal average over the earth's surface is $9.7976\,\mathrm{m\,s^{-2}}$ as reported on page 597 of Gill (1982). Based on a suggestion by Trevor McDougall and David Jackett (2002, personal communication), it may be more appropriate to take an areal average over the ocean surface, which yields $g = 9.7963\,\mathrm{m\,s^{-2}}$ as given by equation (3.83). It is likely that the sensitivity of large-scale ocean circulation to which of these three values is used will be negligible.

3.5.2.5 Coriolis acceleration

Besides the centrifugal acceleration, motion in a rotating frame introduces the Coriolis acceleration

$$\mathbf{a}_c = -2\,\boldsymbol{\Omega} \wedge \mathbf{v}. \tag{3.85}$$

The Coriolis acceleration plays a crucial role in the fluid motion on the earth. Notably, it is a function of the velocity, and is directed perpendicular to the velocity direction and so does not perform work on the fluid parcel. In this way, the Coriolis force

$$\mathbf{F}_c = -2\,(\rho\,\mathrm{d}V)\,\boldsymbol{\Omega} \wedge \mathbf{v} \tag{3.86}$$

acting on a fluid parcel is akin to the Lorentz force acting on a charged particle in an electromagnetic field (e.g., Jackson (1975) as well as Section 3.9).

As discussed in Section 3.1, the Traditional Approximation sets the metric tensor components equal to their values at the earth's surface $r = R$. This assumption has an influence on the corresponding angular momentum that is associated with the motion in this geometry. Namely, the angular momentum about the earth's center is computed with a moment-arm that has a fixed radius $r = R$. Hence, motion in the radial/vertical direction does not alter angular momentum within the Traditional Approximation. Correspondingly, for self-consistency, we must drop the non-radial component of the earth's angular rotation vector when computing the Coriolis force, since this component alters angular momentum when moving

in the vertical direction. Thus, within the Traditional Approximation, the Coriolis force is given by

$$\mathbf{F}_c = -f\,(\rho\,\mathrm{d}V)\,\hat{\mathbf{z}} \wedge \mathbf{v} \tag{3.87}$$

In generalized horizontal coordinates, the Coriolis parameter

$$f = 2\,\Omega\,\sin\phi \tag{3.88}$$

is generally a function of the horizontal coordinates ξ^1, ξ^2 since

$$\phi = \phi(\xi^1, \xi^2). \tag{3.89}$$

3.5.3 Internal forces

As discussed in Section 3.4.2, there are two internal forces of interest when formulating the equations for an ocean model: pressure and friction.

The force due to pressure acting on the boundaries of a fluid parcel is given by the volume of the parcel times the gradient of pressure

$$\mathbf{F}_p = -\mathrm{d}V\,\nabla p. \tag{3.90}$$

When parcels exchange momentum with other parcels and/or boundaries, we can represent this exchance via the components to a stress tensor, whose covariant divergence leads to a force. The forces arise from molecular viscosity, and can generally be expressed as the divergence of a symmetric stress tensor. However, when averaging over turbulent realizations of the ocean fluid, sub-grid-scale transfers of momentum can be far larger than those associated with molecular viscosity. More discussion of this point is postponed until Chapter 6 when presenting the ensemble averaged equations. Further details of friction used in ocean models are given in Chapters 15–17. For now, the physical friction components are written

$$(\rho\,\mathrm{d}V)\,\mathbf{F}^{(\mathbf{v})} = (\rho\,\mathrm{d}V)\,(F^{(u)}, F^{(v)}, F^{(w)}). \tag{3.91}$$

3.5.4 Evolution of linear momentum density

Setting the time tendency of a fluid parcel's linear momentum equal to the forces acting on the parcel yields the equations of motion

$$\begin{aligned}\mathrm{d}V\,[(\rho\,\mathbf{v})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathbf{v}) + \mathcal{M}\,(\hat{\mathbf{z}} \wedge \rho\,\mathbf{v})] \\ = \mathrm{d}V\,[-\rho\,g\,\hat{\mathbf{z}} - f\,\rho\,\hat{\mathbf{z}} \wedge \mathbf{v} - \nabla p + \rho\,\mathbf{F}^{(\mathbf{v})}].\end{aligned} \tag{3.92}$$

Dividing by the parcel's volume leads to the time tendency for the linear momentum density

$$\begin{aligned}(\rho\,\mathbf{v})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathbf{v}) + \mathcal{M}\,(\hat{\mathbf{z}} \wedge \rho\,\mathbf{v}) \\ = -\rho\,g\,\hat{\mathbf{z}} - f\,\hat{\mathbf{z}} \wedge \rho\,\mathbf{v} - \nabla p + \rho\,\mathbf{F}^{(\mathbf{v})}.\end{aligned} \tag{3.93}$$

This equation forms the basis for averaging the equations of motion over an ensemble of realizations (Chapter 6) as well as for discretizing the dynamics (Chapter 10).

3.5.5 Vector invariant velocity equation

Another form of the equation describing fluid parcel dynamics appears in many contexts, such as when deriving the vorticity equation discussed in Section 3.7. It also serves as an alternative starting point for numerical discretization. This equation goes by the name *vector invariant velocity equation* since it is written in a coordinate invariant form. Notably, the advection frequency $\mathcal{M}$ in equation (3.93) vanishes for some coordinate choices, thus indicating the non-tensorial character of this object (see Section 3.5.1). In contrast, all terms in the vector invariant velocity equation remain in tact regardless the coordinates.

To derive the vector invariant velocity equation, return to the derivation in Section 3.5.1. Instead of focusing in the linear momentum density, consider a parcel's velocity, whose time tendency is given by

$$\frac{\mathrm{d}v^a}{\mathrm{d}t} = v^a_{\ ,t} + v^b\, v^a_{\ ;b} \tag{3.94}$$

where again v^a is a tensorial component of the velocity field and all tensor labels run over the spatial range $1, 2, 3$. The nonlinear advection term can be written

$$\begin{aligned} v^b\, v^a_{\ ;b} &= v^b\, g^{ac}\, v_{c;b} \\ &= v^b\, g^{ac}\, (v_{c;b} - v_{b;c}) + v^b\, g^{ac}\, v_{b;c} \end{aligned} \tag{3.95}$$

where we used the metricity condition $g^{ac}_{\ ;b} = 0$ (Section 19.2.4). The second term in equation (3.95) can be written as the covariant derivative of the kinetic energy per mass

$$\begin{aligned} g^{ac}\, v^b\, v_{b\,;c} &= (g^{ac}/2)\, (v^b\, v_b)_{;c} \\ &= (g^{ac}/2)\, (v^b\, v_b)_{,c} \end{aligned} \tag{3.96}$$

where the last step noted that the covariant derivative of a scalar equals to its partial derivative (equation (19.11)). Use of expression (19.14) for the covariant derivative of a one-form component leads to the first term taking the form

$$\begin{aligned} v^b\, g^{ac}\, (v_{c;b} - v_{b;c}) &= v^b\, g^{ac}\, (v_{c,b} - \Gamma^p_{cb}\, v_p - v_{b,c} + \Gamma^p_{bc}\, v_p) \\ &= g^{ac}\, v^b\, (v_{c,b} - v_{b,c}) \end{aligned} \tag{3.97}$$

where the last step used the symmetry property $\Gamma^p_{bc} = \Gamma^p_{cb}$ of the Christoffel symbols (Section 19.4). The anti-symmetric term $v_{c,b} - v_{b,c}$ is reminiscient of vorticity. Indeed, introducing the covariant curl given in Section 19.8 leads to

$$\begin{aligned} (\vec{\omega} \wedge \vec{v})_a &= (\mathrm{curl}\, \vec{v} \wedge \vec{v})_a \\ &= \varepsilon_{abc}\, (\mathrm{curl}\, \vec{v})^b\, v^c \\ &= \varepsilon_{abc}\, \varepsilon^{bpq}\, v_{q,p}\, v^c \\ &= (\delta^p_c\, \delta^q_a - \delta^p_a\, \delta^q_c)\, v_{q,p}\, v^c \\ &= v^c\, (v_{a,c} - v_{c,a}) \end{aligned} \tag{3.98}$$

where we used properties of the covariant Levi-Civita symbol ε_{abc} defined in Section 18.12, and $\vec{v} = (v^1, v^2, v^3)$ form the three tensor components of the velocity. This result then renders

$$\frac{\mathrm{d}v^a}{\mathrm{d}t} = v^a_{\ ,t} + (g^{ac}/2)\, (v^b\, v_b)_{,c} + g^{ac}\, (\vec{\omega} \wedge \vec{v})_c. \tag{3.99}$$

As in Section 3.5.1, we specialize the result (3.99) to the locally orthogonal coordinates commonly used in ocean models, in which case the metric tensor components can be written $g_{ab} = h_a\, h_b\, \delta_{ab}$ with no implied summation. Assuming a time independent metric, it is straightforward to derive an expression for the time tendency of a physical velocity component for a parcel

$$\frac{\mathrm{d}v^{(a)}}{\mathrm{d}t} = v^{(a)}_{,t} + (1/2)\,(\mathbf{v}\cdot\mathbf{v})_{,(a)} + (\boldsymbol{\omega}\wedge\mathbf{v})_{(a)} \tag{3.100}$$

where again the physical components of the velocity are given by $v^{(a)} = h_a\, v^a$ with no implied summation, a boldface symbol characterizes the physical components to a vector such as the velocity $\mathbf{v} = (v^{(1)}, v^{(2)}, v^{(3)})$, and the covariant Levi-Civita symbol ε_{abc} is given by $\varepsilon_{abc} = h_1\, h_2\, h_3\, \epsilon_{abc}$, with ϵ_{abc} the flat-space Levi-Civita symbol (Section 18.12).

The expression (3.100) is more compactly written as

$$\rho\,\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \rho\,(\partial_t + \boldsymbol{\omega}\wedge)\,\mathbf{v} + (\rho/2)\,\nabla(\mathbf{v}\cdot\mathbf{v}). \tag{3.101}$$

Hence, bringing together all the forces acting on a fluid parcel leads to the vector-invariant form of the velocity equation

$$[\partial_t + (\boldsymbol{\omega} + \mathbf{f})\wedge]\,\mathbf{v} = \nabla[\Phi - \mathbf{v}\cdot\mathbf{v}/2] - \rho^{-1}\,\nabla p + \mathbf{F}^{(\mathbf{v})}, \tag{3.102}$$

where $\mathbf{g} = -\nabla\,\Phi = g\,\hat{\mathbf{z}}$ is the gravitational acceleration (Section 3.5.2.3). Comparison of this equation with that for the linear momentum density (3.93) highlights the different treatment of nonlinear transport. The vector invariant velocity equation (3.102) exposes vorticity and kinetic energy per mass, whereas the linear momentum equation (3.93) focuses on nonlinear self-advection along with the coordinate dependent advection metric.

As stated earlier, the vector invariant velocity equation is useful for certain theoretical developments, such as deriving the vorticity equation. Additionally, some modelers use it as their starting point for numerical discretization, with those building Boussinesq C-grid models generally prefering this approach (see Griffies et al. (2000a) for a summary). However, in Chapter 6 we prefer the linear momentum equation (3.93) as the starting point for obtaining the ensemble averaged dynamical equations, as such renders a straightforward discretization of a non-Boussinesq fluid in z-models following the methods of McDougall et al. (2002a) and Greatbatch and McDougall (2002).

3.6 Hydrostatic pressure

We discuss here the hydrostatic balance, which is a balance maintained for the fluid discretized by most global ocean models.

3.6.1 Hydrostatic balance

A fluid in static equilibrium maintains a balance between the vertical pressure gradient and buoyancy force

$$p_{,z} = -\rho\, g. \tag{3.103}$$

This equation is known as the *hydrostatic balance*. Scaling analysis (e.g., Gill (1982)) shows that the hydrostatic balance remains the dominant balance within the vertical momentum equation so long as the vertical scales of motion are much smaller than the horizontal. As such scales are relevant for large-scale ocean climate modeling, global ocean models typically assume a hydrostatic balance. Integrating the hydrostatic balance vertically from the ocean surface determines the pressure at a point in the ocean

$$p(z) = p_a + g \int_z^\eta \mathrm{d}z' \, \rho, \tag{3.104}$$

where p_a is the atmospheric pressure at the sea surface.

3.6.2 Bottom pressure

As mentioned in Section 2.4.3, satellite missions may soon be able to provide very precise and accurate measurements of the ocean bottom pressure. Assuming hydrostatic balance in a constant gravitational field, bottom pressure is a function of the density within the overlying vertical fluid column, the thickness of the column, and the atmospheric pressure. Hence, there are two reasons that the bottom pressure differs for a Boussinesq and non-Boussinesq fluid: (1) differences in sea level, (2) differences in density within the column.

To further understand these differences, let us start with two fluid columns of equal depth and equal density, thus producing equal bottom pressures

$$p_{bot} = g \int_{-H}^\eta \mathrm{d}z \, \rho. \tag{3.105}$$

Let one column be filled with a volume conserving fluid, and the other with a mass conserving fluid. We consider the different evolutions of bottom pressure for these two columns in the case when atmospheric effects are ignored.

The time tendency for bottom pressure in the mass conserving fluid is given by

$$\begin{aligned} \partial_t \, p_{bot} &= g \, \partial_t \left(\int_{-H}^\eta \mathrm{d}z \, \rho \right) \\ &= g \, (-\rho_o \, \nabla \cdot \tilde{\mathbf{U}} + \rho_w \, q_w), \end{aligned} \tag{3.106}$$

where the second equality follows from the mass balance (3.37). This equation says that bottom pressure changes according to changes in the mass per unit area of fluid within the vertical column. In particular, for a mass conserving fluid at rest with zero currents, uniform heating of the fluid, which maintains the fluid at rest, does not change the bottom pressure since the mass of fluid in the column remains the same, only the thickness of the column changes. Integration of the bottom pressure tendency over the full ocean bottom area leaves only a contribution from surface mass fluxes, regardless the buoyancy fluxes within the ocean interior. This result makes sense because bottom pressure integrated over the full area of ocean bottom represents the total gravitational force applied to the ocean bottom from the mass

of fluid. With a constant gravitational acceleration, this force changes in time only when the ocean fluid mass changes.

Now consider the same issues for a volume conserving fluid, in which case the time tendency of bottom pressure is

$$\begin{aligned} \partial_t \, p_{bot} &= g \left(\rho(\eta) \, \eta_{,t} + \int_{-H}^{\eta} \mathrm{d}z \, \rho_{,t} \right) \\ &= g \left(\rho(\eta) \, (-\nabla \cdot \mathbf{U} + q_w) + \int_{-H}^{\eta} \mathrm{d}z \, \rho_{,t} \right) \end{aligned} \tag{3.107}$$

where the volume budget (3.22) was used. Comparison to equation (3.107) for the mass conserving fluid reveals some differences in details of the first and second terms, plus an extra term $\int_{-H}^{\eta} \mathrm{d}z \, \rho_{,t}$ associated with depth integrated density time tendency. When uniformly heating a volume conserving fluid at rest, the extra density tendency term leads to a reduction in bottom pressure. We can understand this result in the following way. Recall from Section 2.4.3 that uniformly heating a column of volume conserving fluid reduces its density while at the same time keeping its surface height fixed in order to maintain the same volume. Therefore, the mass within the column is reduced. Reducing the mass within the column, without changing its thickness, leads to a reduction in bottom pressure, as revealed by equation (3.107).

These examples illustrate some of the unphysical effects associated with the Boussinesq fluid. Again, if we are able to precisely and accurately measure the ocean's bottom pressure using satellites, then comparison of this data with a numerical simulation is handicapped if the model discretizes a Boussinesq fluid. In particular, assimilation of bottom pressure data into the Boussinesq model is awkward since it has a flawed bottom pressure budget.

3.7 Vorticity and potential vorticity

Vorticity is one of the most important dynamical variables in fluid mechanics, with the associated potential vorticity scalar key to understanding and predicting aspects of geophysical fluid flows. We aim in this section to introduce these vorticities by deriving their evolution equations. More complete discussions are available in many places, with Gill (1982), Pedlosky (1987), Müller (1995), and Salmon (1998) recommended for a geophysical fluid focus. To simplify the mathematics, we assume Cartesian tensors. The resulting equations are easily translated into general coordinates by writing them in a covariant manner as detailed in Chapter 19.

3.7.1 Vorticity evolution

We derive the vorticity equation by starting from the vector invariant form of the velocity equation derived in Section 3.5.5, rewritten here for completeness

$$\partial_t \mathbf{v} + \underline{\boldsymbol{\omega}} \wedge \mathbf{v} = \nabla \left(\Phi - \mathbf{v} \cdot \mathbf{v}/2 \right) - (1/\rho) \, \nabla \, p + \mathbf{F}^{(\mathbf{v})} \tag{3.108}$$

where

$$\underline{\boldsymbol{\omega}} = \boldsymbol{\omega} + \mathbf{f} \tag{3.109}$$

is the absolute vorticity. Taking the curl eliminates the conservative forcing from gravity and the kinetic energy density. To tidy-up the resulting equation, we use the identity

$$\begin{aligned}
[\nabla \wedge (\underline{\boldsymbol{\omega}} \wedge \mathbf{v})]_p &= \epsilon_{pqr}\, \epsilon_{rst}\, (\underline{\omega}_s\, v_t)_{,q} \\
&= (\delta_{ps}\, \delta_{qt} - \delta_{pt}\, \delta_{qs})\, (\underline{\omega}_s\, v_t)_{,q} \\
&= (\underline{\omega}_p\, v_q)_{,q} - (\underline{\omega}_q\, v_p)_{,q} \\
&= (\mathbf{v} \cdot \nabla)\underline{\omega}_p + \underline{\omega}_p\, \nabla \cdot \mathbf{v} - (\underline{\boldsymbol{\omega}} \cdot \nabla) v_p
\end{aligned} \tag{3.110}$$

where, as in Section 3.5.5, we used properties of the Levi-Civita symbol ϵ_{pqr} defined in Section 18.12, and we noted that the absolute vorticity has zero divergence $\nabla \cdot \underline{\boldsymbol{\omega}} = 0$. Bringing these results together yields the material evolution of absolute vorticity

$$\begin{aligned}
\frac{\mathrm{d}\, \underline{\boldsymbol{\omega}}}{\mathrm{d}t} = -\underline{\boldsymbol{\omega}}\, (\nabla \cdot \mathbf{v}) + (\underline{\boldsymbol{\omega}} \cdot \nabla)\, \mathbf{v} \\
+ \rho^{-2}\, (\nabla \rho \wedge \nabla p) + \nabla \wedge \mathbf{F}^{(v)}.
\end{aligned} \tag{3.111}$$

The discussion in Chapter 2 of Pedlosky (1987) is recommended for interpreting the forcing terms on the right hand side.

3.7.2 Potential vorticity evolution

Now introduce a scalar function that is materially conserved $\mathrm{d}\psi/\mathrm{d}t = 0$, and consider evolution of vorticity projected onto a direction normal to surfaces of constant ψ. For this purpose, we note that

$$\frac{\mathrm{d}\, \psi_{,a}}{\mathrm{d}t} = -\mathbf{v}_{,a} \cdot \nabla \psi \tag{3.112}$$

which leads to

$$\nabla \psi \cdot \frac{\mathrm{d}\, \underline{\boldsymbol{\omega}}}{\mathrm{d}t} = \frac{\mathrm{d}\, (\nabla \psi \cdot \underline{\boldsymbol{\omega}})}{\mathrm{d}t} + \nabla \psi \cdot (\underline{\boldsymbol{\omega}} \cdot \nabla) \mathbf{v}. \tag{3.113}$$

More than an arbitrary conserved scalar, we are interested in those satisfying $\nabla \psi \cdot (\nabla \rho \wedge \nabla p) = 0$. That is, where the *baroclinicity vector* $\nabla \rho \wedge \nabla p$ is parallel to surfaces of constant ψ. This is the case for functions $\psi = \psi(\rho, p)$ where $\nabla \psi = \psi_{,\rho}\, \nabla \rho + \psi_{,p}\, \nabla p$. For such scalars, the vorticity equation leads to

$$\frac{\mathrm{d}\, (\nabla \psi \cdot \underline{\boldsymbol{\omega}})}{\mathrm{d}t} = -\nabla \psi \cdot \underline{\boldsymbol{\omega}}\, (\nabla \cdot \mathbf{v}) + \nabla \psi \cdot \nabla \wedge \mathbf{F}^{(v)}. \tag{3.114}$$

Mass conservation $(\partial_t + \mathbf{v} \cdot \nabla)\, \rho = -\rho\, \nabla \cdot \mathbf{v}$ allows us to eliminate the divergence $\nabla \cdot \mathbf{v}$ to render

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}t} = \rho^{-1}\, \nabla \psi \cdot \nabla \wedge \mathbf{F}^{(v)} \tag{3.115}$$

where

$$\mathcal{P} = \rho^{-1}\, \underline{\boldsymbol{\omega}} \cdot \nabla \psi \tag{3.116}$$

is the *potential vorticity*. As defined, $\rho\,\mathcal{P}$ is the projection of the absolute vorticity onto the direction normal to surfaces of constant ψ. Equation (3.116) thus says that inviscid fluid motion materially conserves the potential vorticity $\mathcal{P}$. This is an especially important and useful conservation property of ideal oceanic fluid dynamics.

In closing this subsection, we note that the density weighted potential vorticity $\rho\,\mathcal{P}$ can be written as a total divergence

$$\rho\,\mathcal{P} = \nabla \cdot (\underline{\boldsymbol{\omega}}\,\psi) \tag{3.117}$$

since $\nabla \cdot \underline{\boldsymbol{\omega}} = 0$. Haynes and McIntyre (1987, 1990) show that this form of $\rho\,\mathcal{P}$ implies that for general fluid flow, even nonideal and diabatic flow, potential vorticity sources are present *only* at domain boundaries. This result distinguishes potential vorticity from other tracers whose sources can generally be anywhere in the fluid.

3.7.3 Hydrostatic-Boussinesq potential vorticity

Potential vorticity conservation is maintained for many approximate fluid flows with a corresponding potential vorticity variable. To illustrate such, let us derive the potential vorticity equation appropriate for an adiabatic, inviscid, Boussinesq, hydrostatic fluid. Many of the following steps have their analog in the more general discussion above, yet they are presented here for pedagogical purposes.

The dynamical and thermodynamical equations satisfied by the fluid are

$$\rho_o\,(\mathrm{d}/\mathrm{d}t + f\,\hat{\mathbf{z}}\wedge)\,\mathbf{u} = -\nabla_z p \tag{3.118}$$

$$p_{,z} = -g\,\rho \tag{3.119}$$

$$\nabla \cdot \mathbf{v} = 0 \tag{3.120}$$

$$\frac{\mathrm{d}\,\rho}{\mathrm{d}t} = 0, \tag{3.121}$$

where $\mathbf{u} = (u, v, 0)$ is the horizontal velocity, $\nabla_z = (\partial_x, \partial_y, 0)$ is the horizontal gradient, equation (3.119) is the hydrostatic balance, ρ_o is the constant Boussinesq reference density, and equation (3.121) expresses the conservation of potential density for the adiabatic fluid. Note that we equate potential density with *in situ* density, thus allowing the same symbol to be used in the hydrostatic equation as the potential density equation.

The vector invariant form of the velocity equation is derived by noting that the advection of horizontal momentum can be written

$$(\mathbf{v}\cdot\nabla)\,\mathbf{u} = (1/2)\,\nabla_z\,(\mathbf{u}\cdot\mathbf{u}) + \zeta\,\hat{\mathbf{z}}\wedge\mathbf{u} + w\,\partial_z\,\mathbf{u} \tag{3.122}$$

which leads to

$$(\partial_t + w\,\partial_z + \underline{\zeta}\,\hat{\mathbf{z}}\wedge)\,\mathbf{u} = -\nabla_z\,(\mathbf{u}\cdot\mathbf{u}/2 + p/\rho_o), \tag{3.123}$$

where

$$\begin{aligned} \underline{\zeta} &= \zeta + f \\ &= \hat{\mathbf{z}}\cdot\nabla\wedge\mathbf{v} + f \end{aligned} \tag{3.124}$$

is the vertical component of the absolute vorticity. Operating on equation (3.123) with $\hat{\mathbf{z}} \cdot \nabla \wedge$ leads to the material evolution

$$\frac{\mathrm{d}\underline{\zeta}}{\mathrm{d}t} = w_{,z}\, \underline{\zeta} + w_{,y}\, u_{,z} - w_{,x}\, v_{,z}. \tag{3.125}$$

This equation can be put in a more concise form by noting that the non-hydrostatic relative vorticity

$$\boldsymbol{\omega} = \hat{\mathbf{x}}\,(w_{,y} - v_{,z}) + \hat{\mathbf{y}}\,(u_{,z} - w_{,x}) + \hat{\mathbf{z}}\,(v_{,x} - u_{,y}) \tag{3.126}$$

becomes

$$\boldsymbol{\omega} = -\hat{\mathbf{x}}\, v_{,z} + \hat{\mathbf{y}}\, u_{,z} + \hat{\mathbf{z}}\,(v_{,x} - u_{,y}) \tag{3.127}$$

upon moving to the hydrostatic primitive equations as dictated by the Traditional Approximation (Section 3.2). In this case, we have

$$\frac{\mathrm{d}\underline{\zeta}}{\mathrm{d}t} = (\boldsymbol{\omega} \cdot \nabla)\, w, \tag{3.128}$$

which can be deduced from the general vorticity equation (3.111) by making the Boussinesq and Traditional approximations.

For an adiabatic Boussinesq fluid, potential density is a conserved scalar whose gradient we project onto the absolute vorticity to define potential vorticity. Hence, we are led to consider

$$\rho_o\, \mathcal{P} = \underline{\boldsymbol{\omega}} \cdot \nabla \rho \tag{3.129}$$

with the relative vorticity given by the hydrostatic form (3.127) and the ρ^{-1} factor present in the potential vorticity (3.116) reduced to ρ_o^{-1} for the Boussinesq fluid. As defined, $\mathcal{P}$ is proportional to the projection of the absolute vorticity onto the three-dimensional potential density stratification, and its units are those of vorticity.

To verify that $\mathcal{P}$ given by equation (3.129) is materially conserved by the ideal flow, we proceed in a direct manner. That is, we write

$$\rho_o\, \mathcal{P} = f\, \rho_{,z} + (v_{,x}\, \rho_{,z} - v_{,z}\, \rho_{,x}) + (u_{,z}\, \rho_{,y} - u_{,y}\, \rho_{,z}) \tag{3.130}$$

and compute the material evolution of each term on the right hand side. We note here some of the intermediate steps necessary for this calculation

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\,(f\, \rho_{,z}) &= v\, \beta\, \rho_{,z} - \mathbf{v}_{,z} \cdot \nabla \rho \\
\frac{\mathrm{d}}{\mathrm{d}t}\,(v_{,x}\, \rho_{,z}) &= \rho_{,z}\,(-f\, u_{,x} - p_{,xy}/\rho_o - \mathbf{v}_{,x} \cdot \nabla v) - v_{,x}\, \mathbf{v}_{,z} \cdot \nabla \rho \\
\frac{\mathrm{d}}{\mathrm{d}t}\,(v_{,z}\, \rho_{,x}) &= \rho_{,x}\,(-f\, u_{,z} + g\, \rho_{,y}/\rho_o - \mathbf{v}_{,z} \cdot \nabla v) - v_{,z}\, \mathbf{v}_{,x} \cdot \nabla \rho \\
\frac{\mathrm{d}}{\mathrm{d}t}\,(u_{,z}\, \rho_{,y}) &= \rho_{,y}\,(f\, v_{,z} + g\, \rho_{,x}/\rho_o - \mathbf{v}_{,z} \cdot \nabla u) - u_{,z}\, \mathbf{v}_{,y} \cdot \nabla \rho \\
\frac{\mathrm{d}}{\mathrm{d}t}\,(u_{,y}\, \rho_{,z}) &= \rho_{,z}\,(\beta\, v + f\, v_{,y} - p_{,xy}/\rho_o - \mathbf{v}_{,y} \cdot \nabla u) - u_{,y}\, \mathbf{v}_{,z} \cdot \nabla \rho
\end{aligned}$$

where $\beta = f_{,y}$ is the planetary vorticity gradient. Expansion of these terms reveals that indeed $\mathrm{d}\mathcal{P}/\mathrm{d}t = 0$.

3.7.4 Planetary geostrophic potential vorticity

To finish this brief excursion into potential vorticity, we consider potential vorticity for the inviscid planetary geostrophic equations

$$f\,\hat{\mathbf{z}} \wedge \mathbf{u} = -\nabla_z p/\rho_o \tag{3.131}$$

$$p_{,z} = -g\,\rho \tag{3.132}$$

$$\nabla \cdot \mathbf{v} = 0 \tag{3.133}$$

$$\frac{\mathrm{d}\,\rho}{\mathrm{d}t} = 0. \tag{3.134}$$

The momentum equation is a statement of *geostrophic balance* between the Coriolis force and pressure force. It neglects relative vorticity and the material evolution of velocity. Geostrophy is the dominant balance for steady, large-scale, mid-latitude, open ocean circulations. Notably, it represents a diagnostic relation, and so cannot be used to predict future evolution of the velocity field. The remaining equations are the same as considered in the ideal hydrostatic Boussinesq system in Section 3.7.3. The equations of planetary geostrophy have been studied extensively in oceanography, especially when considering thermocline theory (e.g., Pedlosky (1987)).

Potential vorticity for the planetary geostrophic system is given by the planetary vorticity times the vertical stratification

$$\begin{aligned} \rho_o \mathcal{P} &= f\,\rho_{,z} \\ &= -(f\,\rho_o/g)\,N^2 \end{aligned} \tag{3.135}$$

with $N^2 = -(g/\rho_o)\,\rho_{,z}$ the buoyancy frequency. That is, we neglect all contributions to the hydrostatic Boussinesq potential vorticity $\rho_o\,\mathcal{P} = \underline{\boldsymbol{\omega}} \cdot \nabla\rho$ that arise from relative vorticity, and so we are left with just the contribution from the Coriolis term $\hat{\mathbf{z}}\,f \cdot \nabla\rho = f\,\rho_{,z}$.

Proof that $\mathcal{P}$ is materially conserved is straightforward, and we expose here the main steps. First, note that

$$\rho_o \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}t} = \beta\,v\,\rho_{,z} - f\,\mathbf{u}_{,z} \cdot \nabla\rho - f\,w_{,z}\,\rho_{,z} \tag{3.136}$$

where we used $\mathrm{d}\rho/\mathrm{d}t = 0$ and $f_{,y} = \beta$ defines the meridional gradient of the planetary vorticity. Next take the vertical derivative of the geostrophic balance to derive the *thermal wind* equations

$$f\,\mathbf{u}_{,z} = (g/\rho_o)\,\hat{\mathbf{z}} \wedge \nabla\rho. \tag{3.137}$$

Consequently, the vertical shear of the geostrophic velocity $\mathbf{u}_{,z}$ is orthogonal to the horizontal density gradient

$$\begin{aligned} \mathbf{u}_{,z} \cdot \nabla\rho &= (g/f\,\rho_o)\,(\hat{\mathbf{z}} \wedge \nabla\rho) \cdot \nabla\rho \\ &= 0. \end{aligned} \tag{3.138}$$

Now use the continuity equation to replace $w_{,z}$ with the horizontal convergence $w_{,z} = -\nabla \cdot \mathbf{u}$, and use geostrophy to relate this convergence to the meridional

advection of planetary vorticity

$$\begin{aligned} \nabla \cdot (f\,\mathbf{u}) &= \beta\, v + f\, \nabla \cdot \mathbf{u} \\ &= 0. \end{aligned} \tag{3.139}$$

This result substituted into equation (3.136) leads to the desired conservation law $\mathrm{d}\mathcal{P}/\mathrm{d}t = 0$.

3.8 Dynamics of fluid columns

In Section 3.3, we considered the kinematics of infinitesimal fluid parcels as well as finite domains of fluid, including columns of ocean fluid. We now consider the dynamics of a column of hydrostatically balanced ocean fluid extending from the bottom to the free surface. To derive the dynamical equations for a fluid column, we vertically integrate the equations for a fluid parcel over the full ocean depth.

The dynamics of fluid columns is represented by equations in two-spatial dimensions, instead of the three dimensions available for parcels. The two dimensional vertically integrated dynamics, when combined with two dimensional fluid column kinematics, approximates the dynamics of the *external* or *barotropic* mode of the linear primitive equations. The external mode, rigorously defined only for a flat bottom ocean, represents the zeroth eigenmode of an infinite number of vertical modes possessed by the linearized primitive equations (e.g., Gill (1982) and Killworth et al. (1991)). The external mode is nearly two-dimensional, yet for a free surface ocean it has a small vertical dependence, thus breaking the exact analogy with the vertically integrated dynamics.

For the World Ocean, the speed of external mode disturbances is roughly 100 times faster than the speed of the next fastest disturbances. These *external mode gravity waves* propogate such physical effects as tidal fluctuations and tsunami's. To explicitly resolve such waves in the full three-dimensional numerical model would place a severe restriction on the efficiency of the model, since the model's time step would need to be small enough to resolve the fast waves. Hence, there are various methods to extract or split the fast external mode dynamics from the slower, *baroclinic* or *internal* mode dynamics. Over the years, ocean climate modelers have implemented a handfull of methods. The *split-explicit* free surface method is presently quite popular, and details of a particular implementation are discussed in Chapter 10.

3.8.1 An approximate hydrostatic pressure

Recall that the hydrostatic pressure is given by equation (3.104): $p(z) = p_a + g \int_z^\eta \mathrm{d}z'\, \rho$. Fluctuations in this pressure are associated with fluctuations in the atmospheric pressure, the ocean surface height, and the ocean density. Fluctuations in the surface height generally occur on the faster external mode time scale, whereas fluctuations in density generally occur on the slower internal mode time scale. Because we aim to incorporate the fast dynamics into the vertically integrated mode, it is useful to separate these two contributions to the pressure field prior to vertically integrating the equations of motion.

For this purpose, we write

$$\int_z^\eta \rho \, \mathrm{d}z' = \int_z^0 \rho \, \mathrm{d}z' + \int_0^\eta \rho \, \mathrm{d}z'. \tag{3.140}$$

For ocean climate modeling, to within a very good approximation, the fluid between $z = 0$ and $z = \eta$ is well mixed, which is expected due to air-sea interactions. It is therefore appropriate to set

$$\int_0^\eta \rho \, \mathrm{d}z \approx \eta \, \rho(z = 0). \tag{3.141}$$

Doing so then brings the hydrostatic pressure to the form

$$p = p_a + p_s + p_{(b)}, \tag{3.142}$$

where

$$p_{(b)} = g \int_z^0 \rho \, \mathrm{d}z \tag{3.143}$$

is known as the *baroclinic* or *internal* pressure field, and

$$p_s = g \, \eta \, \rho(z = 0) \tag{3.144}$$

is known as the *surface pressure* field. Figure 3.7 illustrates this partitioning of the pressure field.

The baroclinic pressure is the hydrostatic pressure associated with density in the vertical column between some depth z and a resting ocean surface $z = 0$. In turn, the horizontal gradient of this field, $\nabla p_{(b)} = g \int_z^0 \nabla \rho \, \mathrm{d}z'$, arises from baroclinic effects (i.e., horizontal density gradients) in that part of the ocean between the resting ocean surface and the depth z.

Care should be taken to distinguish the surface pressure p_s, which is the pressure due to ocean fluid between $z = 0$ and $z = \eta$, with the atmospheric pressure p_a, which is the pressure felt at the ocean surface $z = \eta$. Both the atmospheric and surface pressures are independent of depth, whereas the baroclinic pressure is depth dependent. Notably, when $\eta < 0$, the surface pressure is negative. Many Boussinesq models approximate the surface pressure with $p_s \approx g \, \eta \, \rho_o$, where ρ_o is the constant Boussinesq density. This approximation only saves a small amount of computer time, and does not simplify the equations in a fundamental manner. Hence, even when making the Boussinesq approximation, modelers may choose to maintain the more accurate hydrostatic result $p_s = g \, \eta \, \rho(z = 0)$ (e.g., Griffies et al. (2001)).

3.8.2 Vertically integrated momentum

Vertically integrating the momentum equation given in Section 3.8.1, and using the surface and bottom kinematic boundary conditions (Section 3.3) leads to

$$(\partial_t + f \, \hat{\mathbf{z}} \wedge) \, \tilde{\mathbf{U}} = -\nabla \, (p_a + p_s)/\rho_o + \mathbf{G}, \tag{3.145}$$

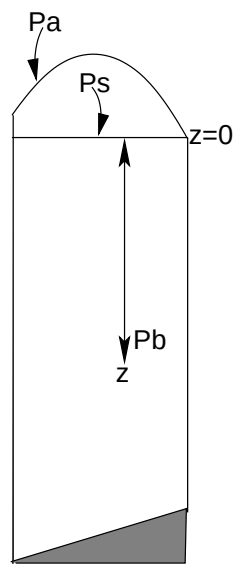

Figure 3.7: Illustrating the partitioning of the hydrostatic pressure given by equation (3.142). The pressure at a depth z is given by the sum of (1) the baroclinic pressure $p_{(b)}$, which is the pressure associated with the density between the depth z and the resting ocean surface $z = 0$, (2) the surface pressure p_s, which is the pressure associated with the vertically uniform ocean density between the resting ocean surface $z = 0$ and the free surface $z = \eta$, (3) the atmospheric pressure p_a.

where

$$\mathbf{G} = \mathbf{u}(\eta)\,\rho_w\,q_w + \int_{-H}^{\eta} \mathrm{d}z\,[-\nabla\,(p_{(b)}/\rho_o) \\ - (\nabla\cdot(\rho\,\mathbf{v}\,\mathbf{u}) + \mathcal{M}\,\hat{\mathbf{z}}\,\wedge\,\rho\,\mathbf{u}) + \rho\,\mathbf{F}^{\mathbf{v}}]. \quad (3.146)$$

With the nonlinear term $\mathbf{G} = 0$, the equations describe the dynamics of linear shallow water inertia-gravity waves (Gill (1982)). The nonlinear forcing imparted by $\mathbf{G} \neq 0$ couples the vertically integrated dynamics to the depth dependent dynamics.

We have much more to say about the vertically integrated dynamics when discussing solution methods in Chapter 10.

3.9 Particle dynamics

The purpose of this section is to help highlight some dynamical issues related the Traditional Approximation (Section 3.1). Namely, what is the form for the energy and angular momentum of a fluid parcel moving in the geometry defined by the Traditional Approximation? Both of these questions can be addressed with manipulations of the primitive equations derived in the previous sections. However, we find it interesting to provide a complementary discussion that is not commonly presented in the geophysical fluids literature. Note that the content in this section lies

somewhat outside the main focus of this chapter, and so it may easily be skipped without loss of continuity.

In this section, we deviate from our focus on continuum fluid parcel mechanics, and instead consider dynamics of a point particle freely moving around a rotating massive sphere. As mentioned in Section 3.5, a Lagrangian perspective on fluid parcel mechanics shares much in common with point particle mechanics. Hence, many of the results for the point particle apply to the fluid parcel. However, the fundamental distinction, as pointed out in Section 3.4.2, is that the boundary of a continuous fluid parcel feels the effects of other parcels through internal or contact forces such as pressure and friction. Correspondingly, a continuous fluid has internal energy, whereas a point particle does not.

To study the point particle motion, we employ Hamilton's Principle instead of Newton's Second Law. Hamilton's Principle leads to the same dynamical equations as Newton's Second Law, yet it formulates the dynamics from Hamilton's variational principle of least action. The action principle serves our purpose since it readily highlights conservation laws and their connection to geometric symmetries. These issues are well known from analytical mechanics (e.g., Fetter and Walecka (1980) and Marion and Thornton (1988)) and have their direct analog in fluid dynamics (Salmon (1988), Salmon (1998), Müller (1995)). In the following, we assume some experience with these issues, at least in so far as they arise in particle mechanics.

3.9.1 A fluid parcel and a point particle

Newton's Law was applied in Section 3.5, which led to the following equation of motion for a fluid parcel

$$(\rho\,\mathrm{d}V)\left(\frac{\mathrm{d}}{\mathrm{d}t}+f\,\hat{\mathbf{z}}\wedge\right)\mathbf{v} = -g\,(\rho\,\mathrm{d}V)\,\hat{\mathbf{z}} - \mathrm{d}V\,\nabla p + (\rho\,\mathrm{d}V)\,\mathbf{F}^{(\mathbf{v})}. \quad (3.147)$$

If we now consider a point particle of mass $\rho\,\mathrm{d}V$ moving around the rotating sphere, it obeys the same equation of motion, yet with the following differences: (1) internal forces (i.e., pressure and friction) are absent since these forces arise from the finite extent of the fluid parcel, (2) the mass times acceleration $(\rho\,\mathrm{d}V)\,\mathrm{d}\mathbf{v}/\mathrm{d}t$ becomes $m\,\dot{\mathbf{v}} = m\,\ddot{\mathbf{x}}$ for the particle, where the dot represents a time derivative with time measured within the frame of the moving particle. Hence, the point particle satisfies

$$(\rho\,\mathrm{d}V)\left(\frac{\mathrm{d}}{\mathrm{d}t}+f\,\hat{\mathbf{z}}\wedge\right)\dot{\mathbf{x}} = -g\,(\rho\,\mathrm{d}V)\,\hat{\mathbf{z}}. \quad (3.148)$$

In the following, we derive this equation from alternative methods for the purpose of clarifying how the Traditional Approximation affects the energy and angular momentum of the particle.

3.9.2 Lagrangian function

To derive the dynamical equations using Hamilton's Principle, we require expressions for the kinetic and potential energy of the particle. The gravitational potential

energy has been discussed in Section 3.5.2.3. We focus here on the kinetic energy and thus derive the particle's Lagrangian function.

If the particle is at a fixed position with respect to the rotating reference frame, it is experiencing a constant *solid body* rotation represented by the velocity

$$\mathbf{U} = \mathbf{\Omega} \wedge \mathbf{x} \tag{3.149}$$

where $\mathbf{\Omega}$ is the angular velocity for the sphere, and $\mathbf{x}$ is the Cartesian position vector of the particle with respect to the earth's center (see Figure 3.1 for a definition of the Cartesian coordinates). Allowing the particle to move with respect to the rotating sphere yields the particle's inertial velocity

$$\dot{\mathbf{x}}_{inertial} = \dot{\mathbf{x}} + \mathbf{U}, \tag{3.150}$$

where $\dot{\mathbf{x}}$ is the particle's velocity as measured in the rotating reference frame. Correspondingly, the particle's inertial kinetic energy is given by

$$K = (m/2)\,(\dot{\mathbf{x}} + \mathbf{U})^2, \tag{3.151}$$

which leads to the Lagrangian function (difference of kinetic and potential energies)

$$\begin{aligned} L &= K - \Phi \\ &= (m/2)\,(\dot{\mathbf{x}} + \mathbf{U})^2 - m\,g_e\,r. \end{aligned} \tag{3.152}$$

Use of the spherical coordinates defined in Figure 3.1 leads to the solid body velocity

$$\mathbf{U} = \Omega\,r\,\cos\phi\,(-\sin\lambda, \cos\lambda, 0), \tag{3.153}$$

the velocity measured in the rotating frame

$$\begin{aligned} \dot{\mathbf{x}} &= \frac{\mathrm{d}}{\mathrm{d}t}\,(x, y, z) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\,(r\,\cos\phi\cos\lambda, r\,\cos\phi\sin\lambda, r\,\sin\phi), \end{aligned} \tag{3.154}$$

and the kinetic energy

$$\begin{aligned} K = (m/2)\,\Big[&(\dot{r}^2 + r^2\dot{\phi}^2 \\ &+ r^2\cos^2\phi\,\dot{\lambda}^2) + (\Omega^2\,r^2\,\cos^2\phi) + (2\,\Omega\,r^2\dot{\lambda}\,\cos^2\phi)\Big]. \end{aligned} \tag{3.155}$$

3.9.3 Hamiltonian dynamics

The Hamiltonian function for the particle is generally given by

$$H = \mathbf{\Pi} \cdot \dot{\mathbf{x}} - L, \tag{3.156}$$

where

$$\mathbf{\Pi} = \frac{\partial L}{\partial \dot{\mathbf{x}}} \tag{3.157}$$

is the *canonical momentum*. For the point particle in the rotating reference frame

$$\mathbf{\Pi} = m\,(\dot{\mathbf{x}} + \mathbf{U}), \tag{3.158}$$

thus leading to the Hamiltonian

$$\begin{aligned} H &= m\,(\dot{\mathbf{x}} + \mathbf{U}) \cdot \dot{\mathbf{x}} - (m/2)\,(\dot{\mathbf{x}} + \mathbf{U})^2 + m\,g_e\,r \\ &= (m/2)\,\dot{\mathbf{x}}^2 + m\,r\,(g_e - \mathbf{U}^2/(2r)). \end{aligned} \tag{3.159}$$

Hence, the Hamiltonian is the sum of the kinetic energy of the particle as measured in the rotating reference frame, plus an effective gravitational potential energy arising from the earth's gravitational attraction and the centrifugal force

$$\Phi_{eff} = m\,r\,[g_e - \mathbf{U}^2/(2r)]. \tag{3.160}$$

The gradient of Φ_{eff} defines the local vertical direction $\hat{\mathbf{z}}$ pointing away from geopotential surfaces (Section 3.5.2.3).

Because the physical system exhibits symmetry with respect to constant translations in time $\tau \rightarrow \tau + \delta\tau$, general principles of analytic mechanics show that the Hamiltonian remains unchanged during the particle's motion. Additionally, the physical system remains invariant under rotations about the north-pole. This symmetry leads to yet another conserved quantity, the angular momentum, as shown in Section 3.9.4.

To employ the formalism of Hamiltonian dynamics for deriving the equations of motion, it is necessary to write the Hamiltonian function in terms of the canonical momentum

$$H = \frac{(\mathbf{\Pi} - m\,\mathbf{U})^2}{2\,m} + \Phi_{eff}. \tag{3.161}$$

This form for the Hamiltonian suggests an interpretation of the physical system motivated by its analog to a charged particle moving in an electromagnetic field (e.g., Section 33 of Fetter and Walecka (1980)). Both systems have the same mathematical form for the kinetic energy, where $\mathbf{U}$ is the analog of the electromagnetic vector potential and $\nabla \wedge \mathbf{U}$ is the analog of the magnetic field. Recall that this correspondence is to be expected since the Coriolis force is mathematically analogous to the Lorentz force acting on a moving charged particle (Section 3.5.2). It suggests an interpretation of the particle moving around the rotating sphere as a particle moving within the *field* associated with the solid body velocity $\mathbf{U}$ as well as the gravitational field.

Hamilton's canonical equations of motion are given by

$$\frac{\partial H}{\partial \mathbf{\Pi}} = \dot{\mathbf{x}} \tag{3.162}$$

$$\frac{\partial H}{\partial \mathbf{x}} = -\dot{\mathbf{\Pi}}. \tag{3.163}$$

For the particle, the first equation leads to an identity

$$\frac{\partial H}{\partial \mathbf{\Pi}} = (1/m)\,(\mathbf{\Pi} - m\,\mathbf{U}) = \dot{\mathbf{x}}, \tag{3.164}$$

whereas the second equation leads to

$$\begin{aligned}\frac{\partial H}{\partial x^a} &= -m\,\dot{\mathbf{x}}\cdot\frac{\partial \mathbf{U}}{\partial x^a} + \frac{\partial \Phi_{eff}}{\partial x^a} \\ &= -\dot{\Pi}_a.\end{aligned} \tag{3.165}$$

Taking the time derivative of equation (3.164) and substituting equation (3.165) leads to the second order equation of motion

$$m\,\ddot{\mathbf{x}} = m\,\dot{\mathbf{x}} \wedge (\nabla \wedge \mathbf{U}) - \nabla\Phi_{eff}. \tag{3.166}$$

To reach this equation, we noted that the time derivative of the rotating frame's velocity is required from the frame of the moving particle. Hence, it is given by the total derivative

$$\dot{\mathbf{U}} = (\partial_t + \dot{\mathbf{x}}\cdot\nabla)\,\mathbf{U}, \tag{3.167}$$

where $\partial_t \mathbf{U} = 0$ since the sphere is rotating at a constant rate. Standard vector identities lead to

$$\nabla \wedge \mathbf{U} = 2\,\mathbf{\Omega}, \tag{3.168}$$

where we assumed $\mathbf{\Omega}$ is spatially constant. Thus, the equation of motion takes the form

$$m\left(\frac{\mathrm{d}}{\mathrm{d}t} + 2\,\mathbf{\Omega}\wedge\right)\dot{\mathbf{x}} = -\nabla\Phi_{eff}, \tag{3.169}$$

as expected from our derivation of equation (3.148).

3.9.4 Conservation laws

We reflect the system's spatial symmetry by choosing spherical coordinates, in which case the Lagrangian L takes the form

$$\begin{aligned}&L + \Phi_{eff} \\ &\qquad = (m/2)\left[(\dot{r}^2 + r^2\dot{\phi}^2 + r^2\cos^2\phi\,\dot{\lambda}^2) + (2\,\Omega\,r^2\dot{\lambda}\,\cos^2\phi)\right],\end{aligned} \tag{3.170}$$

with

$$\Phi_{eff} = m\,g_e\,r - (m/2)\,(\Omega\,r\,\cos\phi)^2. \tag{3.171}$$

Closer connection to the fluid case is realized by introducing the velocity vector's physical components

$$(u, v, w) = ((r\,\cos\phi)\,\dot{\lambda}, r\,\dot{\phi}, \dot{r}), \tag{3.172}$$

which leads to

$$L = (m/2)(u^2 + v^2 + w^2 + 2\,\Omega\,u\,r\,\cos\phi) - \Phi_{eff}. \tag{3.173}$$

Spherical components to the canonical momentum

$$\begin{aligned}\mathbf{\Pi} &= (\Pi_r, \Pi_\lambda, \Pi_\phi) \\ &= \left(\frac{\delta L}{\delta\dot{r}}, \frac{\delta L}{\delta\dot{\lambda}}, \frac{\delta L}{\delta\dot{\phi}}\right)\end{aligned} \tag{3.174}$$

are given by

$$\Pi_r = m\dot{r} \tag{3.175}$$

$$\Pi_\lambda = m\,(r\,\cos\phi)^2\,(\dot{\lambda} + \Omega) \tag{3.176}$$

$$\Pi_\phi = m\,r^2\,\dot{\phi}, \tag{3.177}$$

which can be brought into a more familiar form by introducing the velocity components (u, v, w)

$$\Pi_r = m\,w \tag{3.178}$$

$$\Pi_\lambda = m\,(r\,\cos\phi)\,(u + \Omega\,r\,\cos\phi) \tag{3.179}$$

$$\Pi_\phi = m\,r\,v. \tag{3.180}$$

Note the different physical dimensions of the momentum components, reflecting the different dimensions of the spherical coordinates (r, λ, ϕ).

The zonal canonical momenta Π_λ represents the angular momentum of the particle about the rotation axis extending from the south to north poles of the sphere, with $r\,\cos\phi$ the distance to the rotation axis, and $u + \Omega\,r\,\cos\phi$ the particle's total zonal velocity. Because the Lagrangian is independent of the zonal angle λ, reflecting the system's symmetry about the north-pole, Π_λ is a constant of the motion

$$\frac{\mathrm{d}\,\Pi_\lambda}{\mathrm{d}t} = 0. \tag{3.181}$$

The corresponding Hamiltonian takes the form

$$\begin{aligned} H &= (m/2)\,(\dot{r}^2 + r^2\dot{\lambda}^2\cos^2\phi + r^2\dot{\phi}^2) + \Phi_{eff} \\ &= (m/2)\,(u^2 + v^2 + w^2) + \Phi_{eff} \\ &= \frac{\Pi_r^2}{2\,m} + \frac{[\Pi_\lambda - m\,\Omega\,(r\,\cos\phi)^2]^2}{2\,m\,(r\,\cos\phi)^2} + \frac{\Pi_\phi^2}{2\,m\,r^2} + \Phi_{eff}, \end{aligned} \tag{3.182}$$

which is also conserved because of time translation symmetry

$$\frac{\mathrm{d}\,H}{\mathrm{d}t} = 0. \tag{3.183}$$

Following the notions summarized in Salmon (1998), approximations made within the Hamiltonian enable one to directly track the approximations' effects on conservation laws. For that reason, we choose to make the Traditional Approximation for the particle's motion by dropping the radial component to the canonical momentum, which then leads to the Hamiltonian

$$H_{trad} = (m/2)\,(u^2 + v^2) + \Phi_{eff}. \tag{3.184}$$

Again, this approximation is based on noting that for large-scale geophysical fluid motion, the kinetic energy is dominated by the horizontal motions. This Hamiltonian can be used to generate the dynamical equations for the particle moving in the Traditional Approximation geometry. In particular, by dropping the radial canonical momenta component, we reduce the conserved angular momentum Π_λ to the following form

$$\Pi_\lambda^{trad} = m\,(R\,\cos\phi)\,(u + \Omega\,R\,\cos\phi). \tag{3.185}$$

Thus establishes the form of the conserved energy and angular momentum for the Traditional Approximated point particle. As noted in Section 3.10, the corresponding results for the fluid parcel are modified by noting that the conservation laws apply (1) over the extent of the finite fluid domain, (2) only in the absence of friction, (3) and angular momentum is conserved in the absence of zonal boundaries which break the zonal symmetry. It is because of zonal boundaries introduced by land/sea boundaries that angular momentum is not commonly discussed in ocean modeling. However, it does play a key role in meteorology (e.g., Holton (1992)).

3.10 Symmetry and conservation laws

When discretizing physical processes, it is desirable to formulate these processes in terms of the finite difference of a flux across the faces of a grid cell. For this purpose, the spatial derivative terms in the continuous equations should be in the form of a divergence of a flux. However, when considering motion on a sphere, the presence of the advective metric term in the budget for linear momentum (Section 3.5.1) precludes such. Additionally, as seen in Section 17.4, horizontal friction acting on linear momentum on the sphere cannot be fully written as the divergence of a flux across grid cells. The purpose of this section is to describe the physical and geometric reasons for this form of the linear momentum equation.

3.10.1 Motion on an infinite plane

To get started, consider fluid motion on an infinite flat plane. In the absence of external forces that act to make a particular horizontal direction special, the environment maintains translational symmetry in either of the horizontal directions. Hence, the total horizontal momentum in either direction is conserved. Mathematically, this result means that the linear momentum of a fluid parcel satisfies a conservation equation. That is, the forces affecting the time tendency of this momentum are represented as a total divergence. These statements take their mathematical form as the time tendency for the momentum density

$$(\rho\, u^m)_{,t} = (T^{mn} - \rho\, u^m\, u^n)_{,n} + \rho\, f^m \tag{3.186}$$

where the tensor labels extend over the horizontal directions, we use Cartesian coordinates x, y for convenience, and the conservation of mass was used in the form

$$\rho_{,t} + (\rho\, u^n)_{,n} = 0. \tag{3.187}$$

Additionally, the stress tensor has been written

$$T^{mn} = \tau^{mn} - \delta^{mn}\, p, \tag{3.188}$$

thus representing the sum of the symmetric frictional stress tensor and a diagonal pressure stress tensor. In the absence of external forces f^m, or in the case when these forces can be derived as the divergence of a scalar, the total horizontal momentum $\int \rho\, u^m\, \mathrm{d}V$ is a constant in time.

In addition to momentum in a particular direction, the discussion in Section 15.3.3 shows that so long as the stress tensor is symmetric and there is an absence of external forces, there is an angular momentum conservation law. For motion on the plane, this conservation law arises from symmetry of the unforced motion under rotations about the vertical axis. That is, angular momentum about the vertical direction is conserved in the absence of external forces or boundary effects. Mathematically, the conservation of angular momentum can be derived from the momentum equation in a similar manner to that used in Section 15.3.3. For completeness, the derivation is summarized here.

The angular momentum per unit volume has components given by

$$\rho\,\mathcal{L}_m = \rho\,\epsilon_{mnp}\,x^n\,u^p. \tag{3.189}$$

As the motion is restricted to a horizontal plane, the labels n, p are here limited to $1, 2$. Hence, $m = 3$ represents the angular momentum about the vertical axis. Using the conservation of mass, the momentum equations, and symmetry of the stress tensor, it is straightforward to determine the conservation law

$$(\rho\,\mathcal{L}_m)_{,t} + (\rho\,\mathcal{L}_m\,u^p)_{,p} = \epsilon_{mnp}\,[(x^n\,T^{pq})_{,q} + x^n\,\rho\,f^p]. \tag{3.190}$$

The first term on the right hand side takes the form of a total divergence, and the second term represents external torques. In the absence of external torques and boundary effects, $\int \rho\,\mathcal{L}_m\,\mathrm{d}V$ is a constant in time.

3.10.2 Angular momentum about the north pole

In general, for unforced ideal motion on a manifold containing a translational symmetry, momentum in the direction of this symmetry is conserved. Likewise, if the manifold contains an axis of symmetry, the angular momentum about this axis is conserved. In either case, the form of the equation describing the evolution of the density of the conserved quantity takes the form of a conservation law. To be more specific, for unforced motion on a sphere, the three components of angular momentum about three independent axes of the sphere are conserved. For unforced motion on a rotating sphere, angular momentum about the axis of rotation is conserved. We saw such in Section 3.9 for the motion of a particle around a sphere.

For either the rotating or non-rotating sphere, linear momentum in any direction along the surface of the sphere is not conserved. The reason is that the manifold is not flat, thus removing the translational symmetry required for linear momentum conservation. Consequently, the time tendency of the linear momentum for a fluid parcel on a sphere does not appear in a conservative form. This is the reason that linear momentum evolution cannot be written in a purely conservative form for fluid motion on a sphere.

Let us illustrate these points by recalling the spherical coordinate form of the budget for the zonal and meridional linear momentum density (Section 3.5)

$$(\rho\,\mathbf{u})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathbf{u}) + (\mathcal{M} + f)\,(\hat{\mathbf{z}}\wedge\rho\,\mathbf{u}) = -\nabla p + \rho\,\mathbf{F}^{\mathbf{u}} \tag{3.191}$$

with $\mathcal{M} = (u/R)\,\tan\phi$ the spherical coordinate advection metric term. Use of mass conservation in the form $\rho_{,t} + \nabla\cdot(\rho\,\mathbf{v}) = 0$ leads to the budget for zonal velocity

$$(\partial_t + \mathbf{v}\cdot\nabla)\,u - v\,(\mathcal{M} + f) = -\rho^{-1}\,p_{,x} + F^u. \tag{3.192}$$

Multiplication by the distance to the polar axis, $R\,\cos\phi$, yields

$$(\partial_t + \mathbf{v}\cdot\nabla)\,(u\,R\,\cos\phi) \\ = -(R\,\cos\phi/\rho)\,p_{,x} + v\,f\,R\,\cos\phi + R\,\cos\phi\,F^u. \quad (3.193)$$

Use of the identity

$$(\partial_t + \mathbf{v}\cdot\nabla)\,(\Omega\,R^2\cos^2\phi) = -f\,v\,R\,\cos\phi \quad (3.194)$$

renders

$$(\partial_t + \mathbf{v}\cdot\nabla)[\,R\,\cos\phi\,(u + R\,\Omega\,\cos\phi)] \\ = (R\,\cos\phi)\,(-\rho^{-1}\,p_{,x} + F^u). \quad (3.195)$$

Alternatively, use of mass conservation gives

$$(\rho\,\mathcal{L})_{,t} + \nabla\cdot(\rho\,\mathcal{L}\,\mathbf{v}) = (R\,\cos\phi)\,(-p_{,x} + \rho\,F^u) \quad (3.196)$$

where

$$\mathcal{L} = R\,\cos\phi\,(u + R\,\Omega\,\cos\phi) \quad (3.197)$$

is the angular momentum per unit mass about the north polar axis for a fluid parcel subject to the Traditional Approximation. With $(R\,\cos\phi)\,p_{,x} = p_{,\lambda}$ and $\rho\,F^u$ the zonal friction force per volume discussed in Section 17.4, we see that $\int(\rho\,\mathrm{d}V)\,\mathcal{L} = \int(\rho\,R^2\,\cos\phi\,d\lambda\,d\phi)\,\mathcal{L}$ is a constant for latitudes and depths where a full zonal circuit about the sphere is possible. Such geometry is common in the atmosphere, yet holds only for the Circumpolar Current in the ocean.

3.10.3 Advective and frictional metric terms

We here provide a mathematical statement concerning the origin of the advective and frictional metric terms. For this purpose, consider the linear momentum of a fluid parcel moving on an arbitrary manifold. To describe such motion, we employ the tensor analysis considered in Sections 3.5, 19.12, and Chapters 18 and 19 to write

$$\rho\frac{\mathrm{d}u^m}{\mathrm{d}t} = T^{mn}_{;n} + \rho\,f^m. \quad (3.198)$$

Acceleration of the parcel takes the form

$$\rho\,\frac{\mathrm{d}u^m}{\mathrm{d}t} = \rho(u^m_{,t} + u^n\,u^m_{;n}) \\ = (\rho\,u^m)_{,t} + (\rho\,u^m\,u^n)_{;n} \quad (3.199)$$

where conservation of mass on a curved manifold has been used $\rho_{,t} + (\rho\,u^n)_{;n} = 0$. Hence, the time tendency for the momentum density is given by

$$(\rho\,u^m)_{,t} = (T^{mn} - \rho\,u^m\,u^n)_{;n} + \rho\,f^m. \quad (3.200)$$

This equation is written in the same form as the Cartesian equivalent (3.186), except that now the derivatives are covariant and so contain information about the generally curved manifold. Expanding the covariant divergence using equation (15.167) yields

$$(\rho\, u^m)_{,t} = (\sqrt{\mathcal{G}})^{-1}\, [\sqrt{\mathcal{G}}\, (T^{mn} - \rho\, u^m\, u^n)]_{,n} + (T^{ab} - \rho\, u^a\, u^b)\, \Gamma^m_{ab} + \rho\, f^m. \quad (3.201)$$

The Christoffel symbol Γ^m_{np} accounts for the spatial dependence of the basis vectors. The $\Gamma^m_{ab}\, \tau^{ab}$ term is the "frictional metric term" and the $\rho\, u^a\, u^b\, \Gamma^m_{ab}$ term is the "advective metric term."

As discussed in Section 19.10, integration of a quantity over the volume of a finite grid cell in arbitrary coordinates means performing an integral of the form

$$\int \mathrm{d}V \psi = \int \sqrt{\mathcal{G}}\, \mathrm{d}\xi^1\, \mathrm{d}\xi^2\, \mathrm{d}\xi^3\, \psi, \quad (3.202)$$

where

$$\mathrm{d}V = \sqrt{\mathcal{G}}\, \mathrm{d}\xi^1\, \mathrm{d}\xi^2\, \mathrm{d}\xi^3 \quad (3.203)$$

is the invariant volume element. For example, in spherical coordinates

$$\mathrm{d}V = R^2 \cos\phi\, d\lambda\, d\phi\, \mathrm{d}z. \quad (3.204)$$

Hence, integration of the metric term over a grid cell does not lead to its being transformed into the difference of terms across grid cells faces. The remaining terms can be written such. For flat space using Cartesian coordinates, the metric terms drop out, thus recovering the results discussed in Section 3.10.1.

CHAPTER
FOUR

Ocean thermo-hydrodynamics

Contents

4.1 Introduction **84**
4.2 General types of ocean tracers **84**
4.2.1 Material tracers 84
4.2.2 Thermodynamical tracers 86
4.2.3 Dynamical tracers 87
4.3 Basic equilibrium thermodynamics **87**
4.3.1 The First Law of thermodynamics 87
4.3.2 Fundamental thermodynamical relation 88
4.3.3 Internal energy and homogeneous functions 89
4.3.4 Fundamental relation and specific quantities 90
4.3.5 Adiabatic lapse rate 90
4.4 Energy of a fluid parcel **92**
4.4.1 Kinematics and dynamics 92
4.4.2 Kinetic energy 94
4.4.3 Gravitational potential energy 95
4.4.3.1 Non-Boussinesq parcels 95
4.4.3.2 Boussinesq parcels 96
4.4.3.3 Distinctions between budgets 96
4.4.4 Total mechanical energy 97
4.4.5 Mechanical plus internal energies 97
4.5 Global energy balances **98**
4.5.1 Generic form of the global energy budget 99
4.5.2 Kinetic energy budget 100
4.5.3 Gravitational potential energy budgets 101
4.5.3.1 Non-Boussinesq fluid 102
4.5.3.2 Boussinesq fluid 102
4.6 Basic non-equilibrium thermodynamics **103**
4.7 Thermodynamical tracers **105**
4.7.1 Potential temperature 105

4.7.2 Potential enthalpy 106
4.8 Ocean density 108
4.8.1 *in situ* density 108
4.8.2 Potential density 109
4.8.3 Idealized equations of state 110
4.8.4 Quasi-non-Boussinesq approximation 110

We combine here the hydrodynamics of Chapter 3 with equilibrium and non-equilibrium thermodynamics, thus leading to a *thermo-hydrodynamical* description of the ocean fluid.

4.1 Introduction

In chapter 3 we deduced the equations of motion for a fluid parcel moving in a stratified fluid on a rotating sphere. In this chapter, we combine the mechanical equations with the notions of thermodynamics to formulate a *thermo-hydrodynamical* description of the ocean fluid. Along the way, we present energetics of the ocean fluid as well as the evolution equation for material constituents of the ocean fluid, such as salts and biological matter.

4.2 General types of ocean tracers

In our considerations of non-equilibrium thermodynamics later in this chapter, we introduce a temperature-like scalar field that measures the heat content within a parcel of seawater. This field evolves in a manner similar to other ocean *tracers*. Therefore, it is useful to introduce the various forms of ocean tracers and to derive their evolution equation.

There are three general types of ocean tracers: (1) tracers representing the concentration of material constituents in the fluid (e.g., passive tracers and salinity), (2) tracers representing the thermodynamic properties of the fluid (e.g., entropy, enthalpy, and temperature), (3) and tracers embodying dynamical properties of the fluid (e.g., potential vorticity discussed in Section 3.7).

4.2.1 Material tracers

Fluid parcels generally contain material constituents that are transported as the parcel moves through the ocean. These constituents are commonly termed *tracers* since they are useful for *tracing* pathways of fluid transport.*

Important *passive tracers* include the biological species of phytoplankton and zooplankton, which form the base of the ocean's food chain; radio-active isotopes input to the ocean due to nuclear bomb testing active in the 1950's and 1960's; and purposeful releases of tracers, such as sulpher hexaflouride (SF_6), used to quantify rates of mixing (Ledwell et al. (1993)). *Passive* here refers to the near inability of

*A similar terminology is commonly used in medicine for probing human physiology.

these tracers to alter ocean density. Hence, they have negligible effects on ocean dynamics.

Salinity s is a generic term referring to the large number of ionic constituents in seawater. The dominant ions are sodium Na^{+}, chlorine Cl^{-}, sulphate SO_4^{--}, and magnesium Mg^{++}. At the common salinity of 35 parts "salt" per thousand parts solution, Table 4.1 of Apel (1987) indicates there are roughly 19.345 grams of chlorine ion per kilogram of water, 10.752 grams/kilogram of sodium ion, 1.295 gram/kilogram of magnesium ion, and 2.701 grams/kilogram of sulfate ion. These, and the other trace ions, maintain a near constant ratio throughout the World Ocean, hence allowing for the single term "salinity" to have near global validity.

Material tracers have a total mass within a fluid parcel written

$$\mathrm{d}M_C = (\mathrm{d}M)\, C = (\rho\, \mathrm{d}V)\, C, \tag{4.1}$$

where $\mathrm{d}M = \rho\, \mathrm{d}V$ is the total mass of the parcel of seawater, and $C = \mathrm{d}M_C/\mathrm{d}M$ is the mass of tracer per mass of seawater. C is often termed the *tracer concentration* or *tracer mass fraction*. Notably, since the seawater mass and tracer mass are both scalar quantities, the tracer concentration is likewise a scalar field. Hence, tracers are often called *scalars* in the fluid dynamics literature (see Section 18.5 for a mathematical definition of scalar fields). Additionally, when the tracer concentration C is uniform throughout the ocean, the budget for tracer mass reduces to the budget for seawater mass. This observation, which follows from the definition of C, provides an important compatibility constraint between mass and tracer budgets.

Conserving tracer mass within a parcel of seawater leads to the conservation statement

$$\begin{aligned} \frac{\mathrm{d}\left[(\rho\, \mathrm{d}V)\, C\right]}{\mathrm{d}t} &= (\rho\, \mathrm{d}V)\, \frac{\mathrm{d}C}{\mathrm{d}t} \\ &= 0, \end{aligned} \tag{4.2}$$

where we used mass conservation $\mathrm{d}(\rho\, \mathrm{d}V)/\mathrm{d}t = 0$. Combining the Eulerian form of material tracer conservation $\mathrm{d}C/\mathrm{d}t = (\partial_t + \mathbf{v}\cdot\nabla)\, C = 0$ with mass conservation $\rho_{,t} + \nabla\cdot(\rho\, \mathbf{v}) = 0$ leads to the Eulerian conservation law for tracer mass per volume $\rho\, C$ within in a seawater parcel

$$(\rho\, C)_{,t} + \nabla\cdot(\rho\, C\, \mathbf{v}) = 0. \tag{4.3}$$

As discussed below, additional terms appear on the right hand side when sources or sinks of tracer are found within the fluid or at the boundaries.

In most numerical ocean models, tracer concentration is time stepped by discretizing the Eulerian equation (4.3) for the tracer mass per volume, and separately time stepping the conservation for ocean mass per volume $\rho_{,t} + \nabla\cdot(\rho\, \mathbf{v}) = 0$. Compatibility between the two discretized conservation laws is maintained by having the discrete equation for tracer mass per volume reduce to the discrete mass per volume equation upon setting tracer concentration to unity: $C = 1$. Such compatibility is very important to maintain numerically.

For purposes of establishing the momentum and tracer budgets, the mass of a Boussinesq seawater parcel is given by $\mathrm{d}M = \rho_o\, \mathrm{d}V$. Material conservation of tracer mass $(\rho_o\, \mathrm{d}V)\, C$ still leads to material conservation of tracer concentration:

$dC/dt = 0$, since now the parcel's volume is conserved. Combining material conservation of tracer concentration with parcel volume conservation, $\nabla \cdot \mathbf{v} = 0$, leads to the Eulerian conservation law

$$C_{,t} + \nabla \cdot (\mathbf{v}\, C) = 0. \tag{4.4}$$

Again, this combined equation is especially useful as a starting point numerically since it manifests the need to maintain compatibility between conservation of volume and tracer concentration.

There are two main ways that the mass of tracer within a seawater parcel can change. First, there can be tracer-dependent sources or sinks of tracer mass $\rho\,\mathcal{S}$ within the fluid domain or at the boundaries. Second, there can be irreversible molecular or turbulent mixing effects that move tracer mass between parcels. Many of these mixing effects can be mathematically represented by the convergence of a density weighted tracer flux $\rho\,\mathbf{F}$. Hence, for a non-Boussinesq fluid, the time tendency for density weighted tracer concentration is written

$$(\rho\, C)_{,t} = -\nabla \cdot (\rho\, C\, \mathbf{v} + \rho\, \mathbf{F}) + \rho\, \mathcal{S}, \tag{4.5}$$

whereas for a Boussinesq fluid it is

$$C_{,t} = -\nabla \cdot (C\, \mathbf{v} + \mathbf{F}) + \mathcal{S}. \tag{4.6}$$

4.2.2 Thermodynamical tracers

Along with salinity and pressure, the density of a seawater parcel is a function temperature. The relation

$$\rho = \rho(T, s, p) \tag{4.7}$$

is known as the equation of state (Section 4.8), with T the *in situ* temperature. Since density affects pressure via the hydrostatic balance, variations in temperature and salinity yield important *thermo-haline* forces for ocean currents. Therefore, it is crucial to provide an accurate and precise description of the temperature and salinity distributions to understand and simulate ocean currents.

Developing an understanding of heat in the ocean requires some of the tools of thermodynamics to be discussed in subsequent sections. Thermodynamics is a phenomenological discipline whose fundamentals lie in statistical mechanics (e.g., Huang (1987) or Reichl (1987)), which itself can be considered a branch of information theory (e.g., Jaynes (1957a,b)). As discussed in Section 2.2, geophysical fluid dynamical systems of interest for ocean climate modeling involve parcels of seawater in quasi-thermodynamical equilibrium. Hence, although these fluid parcels evolve in time, the laws of thermodynamics can be used to establish balances of heat, work, internal energy, and entropy within the fluid. Use of equilibrium thermodynamics for time dependent phenomenon falls under the area of quasi-equilibrium thermodynamics (or linear irreversible thermodynamics). We use the methods of irreversible thermodynamic to derive equations expressing the evolution of heat within a parcel of seawater.

The traditional manner for describing heat within the ocean is via *potential temperature*, and its evolution is typically simulated by ocean models. The reason for

preferring potential temperature is that it is nearly conserved under adiabatic and isentropic motion, whereas *in situ* temperature changes due to pressure effects even without the transfer of heat. Conservative equations are generally more amendable to straightforward physical interpretations in terms of flux divergences leading to local temporal changes. They also lend themselves to standard numerical discretizations of the transport operators. Hence, we prefer a conservative quantity when aiming to describe heat transport in the ocean, and potential temperature has served this purpose for sometime.

Even though potential temperature is standard in the oceanographic literature, work by McDougall (2002) highlights the fact that potential temperature is less conservative in the ocean than *potential enthalpy*. Hence, a more accurate description of ocean heat content, and hence the density of a fluid parcel in an ocean model, is obtained by using potential enthalpy rather than potential temperature. We discuss this point further in Section 4.7.

4.2.3 Dynamical tracers

The third class of tracers is termed the *dynamical tracers*. These tracers are built from generally nonlinear combinations of the dynamical fields. Potential vorticity, discussed in Section 3.7, is the most important dynamical tracer for geophysical fluid dynamics.

4.3 Basic equilibrium thermodynamics

When discussing the continuum hypothesis in Section 2.2, we noted that a fluid parcel represents a macroscopically small, yet microscopically large, system in quasi-thermodynamical equilibrium. Hence, there are innumerable molecular degrees of freedom that are averaged over when describing a fluid as a continuum. Internal energy embodies the energy of these internal degrees of freedom, such as arise from thermal agitation and molecular interactions. The total energy of a macroscopic system in quasi-thermodynamical equilibrium is thus the sum of the system's kinetic, potential, and internal energies.

The methods of equilibrium thermodynamics allow us to relate small changes in work and heat applied to a parcel of seawater to changes in its internal energy. For our purposes, seawater can be considered a two component fluid consisting of freshwater and salt, where salt is a generic term representing a suite of dissolved ions in seawater whose ratio is nearly constant over the ocean (Section 4.2.1). We therefore consider thermodynamical relations for a multi-component fluid and then specialize to the two-component case.

4.3.1 The First Law of thermodynamics

The First Law of thermodynamics establishes a relationship between infinitesimal changes of a system's internal energy to changes in the system's heat, work, and composition

$$\mathrm{d}\mathcal{I}^e = \delta q + \delta w + \sum_k \mu_k \, \mathrm{d}M_k. \tag{4.8}$$

In this equation, $\mathcal{I}^e$ is the system's internal energy, q is the heat applied to the system, w is the work applied to the system, μ_k is the chemical potential associated with species component k, and M_k is the mass of component k. As discussed in Section 4.3.3, the internal energy is proportional to the size of the system, with systems having more volume and mass having more energy. Quantities of this sort are termed *extensive* and are labelled with the superscript e in the following.

Changes in heat and work applied to a system, δq and δw, depend on details of the path taken to realize the changes. This path dependence motivates the use of the δ symbol to signify an *inexact* differential. Their sum is path independent, thus identifying $\mathrm{d}\mathcal{I}^e$ as an *exact* differential just as differential changes in mass of a species, $\mathrm{d}M_k$, or differential changes in the mechanical energy, $\mathrm{d}\left[M\left(\mathbf{v}^2/2+\Phi\right)\right]$. Correspondingly, internal energy is known as a thermodynamic *state function* since its value depends only on properties of the equilibrium state, not the path used to reach that state.

There are many ways that forces can do work to a parcel of seawater. However, we limit our focus to work associated with quasi-static* changes to the parcel's volume V^e

$$\delta w = -p\,\mathrm{d}V^e \tag{4.9}$$

where p is the pressure applied to the boundaries of the parcel.† The negative sign in the pressure-work relation (4.9) arises since compressing the fluid parcel into a smaller volume ($\mathrm{d}V^e < 0$) requires positive work be applied to the parcel ($\delta w > 0$). Pressure is termed an intensive variable, measuring the *intensity* of a force conjugate to the extensive variable V^e. Pressure also provides an *integrating factor* that relates the inexact differential δw to the exact differential $\mathrm{d}V^e$. Note that for a Boussinesq seawater parcel, its incompressibility means that no pressure-work can be applied to this parcel.

Just as for work, heating applied to the parcel is assumed to occur in a quasi-static manner. For such processes, we can relate heating to changes in entropy via

$$\delta q = T\,\mathrm{d}\zeta^e, \tag{4.10}$$

with T the absolute *in situ* temperature (an intensive variable as well as an integrating factor) and ζ^e the entropy of the system (an extensive variable and a state function).

4.3.2 Fundamental thermodynamical relation

With work limited to quasi-static pressure-work associated with volume changes, and heat applied quasi-statically, substitution of relations (4.9) and (4.10) into the First Law (4.8) leads to

$$\mathrm{d}\mathcal{I}^e = T\,\mathrm{d}\zeta^e - p\,\mathrm{d}V^e + \sum_k \mu_k\,\mathrm{d}M_k. \tag{4.11}$$

Quasi-static refers to an idealized situation whereby a thermodynamic system moves from one state to another via an infinite number of intermediate equilibrium states. See Callen (1985) for details.

†In this section we prefer the symbol V^e instead of $\mathrm{d}V$ for the parcel volume in order to maintain consistency with the thermodynamic literature.

This relation holds between infinitesimal changes in thermodynamical state functions. Hence, although derived for quasi-static processes from the First Law using connections to work and heat, equation (4.11) holds for arbitrary infinitesimal changes; its connection to the First Law holds only for quasi-static processes.

Equation (4.11) represents an important *fundamental thermodynamic* relation, often called the *first Gibbs relation*. It is the starting point for many manipulations in thermodynamics. In particular, it renders the following relations between intensive variables and the partial derivatives of internal energy with respect to extensive quantities

$$\left(\frac{\partial \mathcal{I}^e}{\partial \zeta^e}\right)_{V^e, M_k} = T \tag{4.12}$$

$$\left(\frac{\partial \mathcal{I}^e}{\partial V^e}\right)_{\zeta^e, M_k} = -p \tag{4.13}$$

$$\left(\frac{\partial \mathcal{I}^e}{\partial M_k}\right)_{\zeta^e, V^e} = \mu_k. \tag{4.14}$$

In these expressions, partial derivatives are taken with the noted variables held constant.

4.3.3 Internal energy and homogeneous functions

The fundamental thermodynamic relation (4.11) indicates that internal energy is naturally considered a function of entropy, volume, and mass

$$\mathcal{I}^e = \mathcal{I}^e(\zeta^e, V^e, M_k). \tag{4.15}$$

If we scale the system by an arbitrary parameter λ, the extensive variables entropy, volume, and mass also scale by this factor. Through the fundamental relation (4.11), the internal energy scales likewise, thus rendering

$$\mathcal{I}^e(\lambda\, \zeta^e, \lambda\, V^e, \lambda\, M_k) = \lambda\, \mathcal{I}^e(\zeta^e, V^e, M_k). \tag{4.16}$$

A function that scales in this way is termed a *homogenous function of degree one*. Differentiating both sides of this identity with respect to λ, setting λ to unity, and using the partial derivative identities (4.12)–(4.14), yields

$$\mathcal{I}^e = T\, \zeta^e - p\, V^e + \sum_k \mu_k\, M_k. \tag{4.17}$$

This result represents a special case of Euler's Theorem of homogeneous functions. Taking the differential of this equation, and using the fundamental relation (4.11) leads to the *Gibbs-Duhem* relation

$$\zeta^e\, \mathrm{d}T - V^e\, \mathrm{d}p + \sum_k M_k\, \mathrm{d}\mu_k = 0. \tag{4.18}$$

4.3.4 Fundamental relation and specific quantities

For many purposes in fluid dynamics, it proves convenient to consider fundamental thermodynamic relations for a system of unit mass. For this purpose, we scale away the mass of the system by setting the scale-factor $\lambda = M^{-1}$ and introduce the *specific* quantities

$$\mathcal{I}^e = M\,\mathcal{I} \tag{4.19}$$
$$\zeta^e = M\,\zeta \tag{4.20}$$
$$V^e = M\,v_s \tag{4.21}$$
$$M_k = M\,C_k \tag{4.22}$$

where

$$v_s = \rho^{-1}. \tag{4.23}$$

is the specific volume* and C_k is the mass fraction or concentration of species k in the seawater parcel. Substituting the specific quantities into the fundamental relation (4.11), and using expression (4.17) for the internal energy leads to the fundamental relation in terms of intensive state variables

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\zeta - p\,\mathrm{d}v_s + \sum_k \mu_k\,\mathrm{d}C_k. \tag{4.24}$$

As mentioned earlier, seawater can for our purposes be considered a binary system of salt and freshwater. In this case,

$$\begin{aligned} C_1 + C_2 &= C_{salt} + C_{water} \\ &= 1. \end{aligned} \tag{4.25}$$

Introducing this constraint into (4.24) leads to

$$\mathrm{d}\mathcal{I} = T\,\mathrm{d}\zeta - p\,\mathrm{d}v_s + \mu\,\mathrm{d}C, \tag{4.26}$$

where $C = C_{salt}$ is the concentration of salt, and

$$\mu = \mu_{salt} - \mu_{water} \tag{4.27}$$

is the relative chemical potential. Note that the *salinity* s is related to C_{salt} via

$$s = 1000\,C_{salt}. \tag{4.28}$$

4.3.5 Adiabatic lapse rate

Consider a finite mass of water with uniform and fixed salinity (e.g., a fresh water lake). Let this water be static and rest in a very thin horizontal layer so each water parcel has approximately the same gravitational potential energy. Consequently, all fluid parcels experience the same pressure and hence the same temperature. Now isentropically (i.e., adiabatically and reversibly) rearrange the fluid into a vertical column by introducing gravitational potential energy. Gravity imparts a

*The subscript s distinguishes the specific volume v_s from the meridional component of the velocity field v.

hydrostatic pressure (Section 3.6) that makes pressure at the bottom more than the top. We wish to know how this difference in pressure affects the temperature in the fluid. For this purpose, we introduce some standard thermodynamic manipulations valid for a single component fluid, such as fresh water (all manipulations here are for C =constant).

To start, introduce the specific heat at constant pressure

$$\begin{aligned} C_p &\equiv (\delta q/\mathrm{d}T)_p \\ &= T\,(\partial\zeta/\partial T)_p \end{aligned} \tag{4.29}$$

where the p subscript indicates that pressure is held fixed when varying temperature. Specific heat measures the change in heat associated with a change in temperature, here taken with the fluid pressure held fixed. The second form of this expression arises by assuming that changes in heat occur quasi-statically, hence allowing for use of the relation between heating and entropy $\delta q = T\,d\zeta$. With entropy a function of temperature and pressure, infinitesimal changes in entropy are given by

$$\mathrm{d}\zeta = (\partial\zeta/\partial T)_p\,\mathrm{d}T + (\partial\zeta/\partial p)_T\,\mathrm{d}p. \tag{4.30}$$

Substituting the definition of heat capacity leads to

$$T\,\mathrm{d}\zeta = C_p\,\mathrm{d}T + T\,(\partial\zeta/\partial p)_T\,\mathrm{d}p. \tag{4.31}$$

We now eliminate $(\partial\zeta/\partial p)_T$ in favor of a more easily measurable quantity. For this purpose, note that use of the fundamental thermodynamic relation (4.26) (with $\mathrm{d}C = 0$) leads to

$$T\,(\partial\zeta/\partial T)_p = (\partial\mathcal{I}/\partial T)_p + p\,(\partial v_s/\partial T)_p. \tag{4.32}$$

Likewise, equation (4.26) implies

$$T\,(\partial\zeta/\partial p)_T = (\partial\mathcal{I}/\partial p)_T + p\,(\partial v_s/\partial p)_T. \tag{4.33}$$

Applying $(\partial/\partial p)_T$ to equation (4.32) and $(\partial/\partial T)_p$ to equation (4.33), and then subtracting, leads to the identity

$$(\partial\zeta/\partial p)_T = -(\partial v_s/\partial T)_p. \tag{4.34}$$

Introducing the thermal expansion coefficient

$$\begin{aligned} \alpha_T &= -\rho^{-1}\,(\partial\rho/\partial T)_p \\ &= \rho\,(\partial v_s/\partial T)_p \end{aligned} \tag{4.35}$$

yields an expression for changes in entropy in terms of changes in temperature and pressure

$$\begin{aligned} T\,\mathrm{d}\zeta &= C_p\,\mathrm{d}T - T\,(\partial v_s/\partial T)_p\,\mathrm{d}p \\ &= C_p\,\mathrm{d}T - (T\,\alpha_T/\rho)\,\mathrm{d}p. \end{aligned} \tag{4.36}$$

C_p and α_T are readily measurable *response functions*, thus making equation (4.36) a very useful expression for infinitesimal entropy changes. Note that we add a subscript T to α in order to distinguish this expansion coefficient from that defined according to potential temperature (Section 4.8.1).

We are now ready to answer the question posed at the start of this section. Namely, equation (4.36) tells us that a change in temperature associated with isentropic motion through a pressure field is given by

$$\mathrm{d}T = (T\,\alpha_T/\rho\,C_p)\,\mathrm{d}p \tag{4.37}$$

where $d\zeta = 0$ for isentropic changes. Hence, temperature indeed changes when pressure changes, even though there has been no heat exchanged with the environment. The combination

$$\Gamma = \frac{T\,\alpha_T}{\rho\,C_p} \tag{4.38}$$

is termed the adiabatic lapse rate for temperature as a function of pressure. Alternatively, we note that a static fluid in a gravity field maintains hydrostatic balance, thus allowing the total pressure differential to be directly related to the vertical differential

$$\mathrm{d}p = -g\,\rho\,\mathrm{d}z, \tag{4.39}$$

which leads to the temperature change

$$\mathrm{d}T = -(g\,T\,\alpha_T/C_p)\,\mathrm{d}z. \tag{4.40}$$

In this case, we introduce the adiabatic lapse rate

$$\widehat{\Gamma} = \frac{g\,T\,\alpha_T}{\rho\,C_p} \tag{4.41}$$

for temperature as a function of depth. The lapse rates determine the change in temperature (the *lapse*) a fluid parcel experiences as it is isentropically moved vertically through a hydrostatically balanced pressure field. Note that McDougall and Feistel (2002) provide a thorough discussion of the lapse rate in terms of molecular dynamics.

4.4 Energy of a fluid parcel

In this section we derive energy balances of a dynamic fluid parcel maintaining hydrostatic balance, as is the case for a parcel respecting the primitive equations. Both non-Boussinesq and Boussinesq parcels are considered.

4.4.1 Kinematics and dynamics

Before getting started, it is useful to summarize the kinematical and dynamical equations discussed in Chapter 3. In particular, mass conservation for a non-Boussinesq fluid in its Eulerian form is given by (Section 3.3)

$$\rho_{,t} + \nabla\cdot(\rho\,\mathbf{v}) = 0, \tag{4.42}$$

where

$$\mathbf{v} = (\mathbf{u}, w) = (u, v, w) \tag{4.43}$$

is the three-dimensional velocity with $\mathbf{u} = (u, v)$ the horizontal components. Mass conservation in its Lagrangian form

$$\begin{aligned} \frac{1}{\mathrm{d}M}\frac{\mathrm{d}(\mathrm{d}M)}{\mathrm{d}t} &= \frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t} + \frac{1}{\mathrm{d}V}\frac{\mathrm{d}(\mathrm{d}V)}{\mathrm{d}t} \\ &= 0 \end{aligned} \tag{4.44}$$

allows us to write

$$\begin{aligned} \frac{1}{\mathrm{d}V}\frac{\mathrm{d}(\mathrm{d}V)}{\mathrm{d}t} &= \frac{1}{v_s}\frac{\mathrm{d}v_s}{\mathrm{d}t} \\ &= \nabla \cdot \mathbf{v}, \end{aligned} \tag{4.45}$$

where the second equality follows from the Eulerian form of mass conservation (4.42). Equation (4.45) says that the relative change in volume of a fluid parcel equals the divergence of the three-dimensional velocity field. This relation then leads to the volume conservation relation appropriate for incompressible Boussinesq fluids

$$\frac{1}{\mathrm{d}V}\frac{\mathrm{d}(\mathrm{d}V)}{\mathrm{d}t} = \nabla \cdot \mathbf{v} = 0. \tag{4.46}$$

Mass conservation can be used to relate the Lagrangian and Eulerian forms of time evolution for a scalar via

$$\rho\frac{\mathrm{d}\psi}{\mathrm{d}t} = (\rho\,\psi)_{,t} + \nabla \cdot (\mathbf{v}\,\rho\,\psi). \tag{4.47}$$

This is a useful relation that will be used frequently in the following. Its Boussinesq analog is given by

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \psi_{,t} + \nabla \cdot (\mathbf{v}\,\psi). \tag{4.48}$$

In addition to constraints imposed by mass and volume conservation for parcels within the ocean interior, the fluid maintains kinematic constraints at the bottom and surface taking the form

$$\mathbf{v} \cdot (\nabla H, 1) = 0 \qquad \text{at } z = -H \tag{4.49}$$

$$\mathbf{v} \cdot (-\nabla \eta, 1) = \eta_{,t} - (\rho_w/\rho)\, q_w \qquad \text{at } z = \eta, \tag{4.50}$$

which are more frequently written

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H \tag{4.51}$$

$$(\partial_t + \mathbf{u} \cdot \nabla)\,\eta = (\rho_w/\rho)\, q_w + w \qquad \text{at } z = \eta. \tag{4.52}$$

For a Boussinesq fluid, the $z = \eta$ boundary condition is recovered by setting $\rho_w = \rho$. These boundary conditions are relevant when considering energy budgets over the full ocean domain.

The dynamics of interest here result from making the hydrostatic approximation (Section 3.6), where the vertical momentum equation reduces to

$$p_{,z} = -\rho\, g, \tag{4.53}$$

and the horizontal momentum equation (section 3.5) takes the form

$$(\rho\,\mathbf{u})_{,t} + \nabla \cdot (\rho\,\mathbf{v}\,\mathbf{u}) + (f + \mathcal{M})\,\hat{\mathbf{z}} \wedge (\rho\,\mathbf{v}) = -\nabla p + \rho\,\mathbf{F}^{(\mathbf{u})}. \tag{4.54}$$

The momentum balance for a Boussinesq fluid results by setting $\rho = \rho_o$.

4.4.2 Kinetic energy

For a dynamic fluid maintaining hydrostatic balance, we are only concerned with kinetic energy in the horizontal currents

$$(\rho \, \mathrm{d}V) \, \mathcal{K} = (\rho \, \mathrm{d}V) \, \mathbf{u}^2/2, \tag{4.55}$$

where $\mathbf{u}^2 = \mathbf{u} \cdot \mathbf{u}$ and

$$\mathcal{K} = \mathbf{u}^2/2 \tag{4.56}$$

is the kinetic energy per unit mass. The kinetic energy associated with vertical motions is many orders smaller, and so is ignored for fluids respecting the primitive equations. Furthermore, as noted by Bokhove (2000), the Hamiltonian function for a hydrostatic fluid has a kinetic energy associated just with horizontal motions.

The material time tendency for the parcel's horizontal kinetic energy is

$$\begin{aligned} \frac{\mathrm{d}(\rho \, \mathrm{d}V \, \mathcal{K})}{\mathrm{d}t} &= (\rho \, \mathrm{d}V) \, \frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} \\ &= \mathrm{d}V \, [(\rho \, \mathcal{K})_{,t} + \nabla \cdot (\rho \, \mathbf{v} \, \mathcal{K})], \end{aligned} \tag{4.57}$$

where the first step used mass conservation in its Lagrangian form $\mathrm{d}(\rho \, \mathrm{d}V)/\mathrm{d}t = 0$, and the last step used the identity (4.47) resulting from mass conservation in its Eulerian form (4.42). The momentum equation (4.54) leads to

$$\begin{aligned} \rho \, \frac{\mathrm{d}\mathcal{K}}{\mathrm{d}t} &= \mathbf{u} \cdot \rho \, \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} \\ &= \mathbf{u} \cdot (-\nabla p + \rho \, \mathbf{F}^{\mathbf{u}}), \end{aligned} \tag{4.58}$$

thus rendering

$$(\rho \, \mathcal{K})_{,t} + \nabla \cdot (\rho \, \mathbf{v} \, \mathcal{K}) = \mathbf{u} \cdot (-\nabla p + \rho \, \mathbf{F}^{\mathbf{u}}). \tag{4.59}$$

The pressure term can be interpreted as work by horizontal currents against horizontal pressure gradients. For example, if currents are directed down the horizontal pressure gradient ($-\mathbf{u} \cdot \nabla p > 0$), then kinetic energy increases as the flow speed increases. The opposite occurs for flow directed up the pressure gradient. Flow directed parallel to isobars ($\mathbf{u} \cdot \nabla p = 0$), such as those in geostrophic balance, do not alter kinetic energy.

The friction term can be written in a manner exposing its dissipative aspects. For this purpose we anticipate results from Chapter 15, where we show that the friction force per volume acting on the horizontal momentum takes the general form

$$\rho \, F^m = \tau^{mn}_{;n} \tag{4.60}$$

where τ^{mn} are components to the symmetric frictional stress tensor $\boldsymbol{\tau}$, the semicolon denotes a covariant derivative (see Section 19.2), and the summation convention is employed where repeated indices are summed over their range. This form for friction leads to the inner product

$$\begin{aligned} \rho \, \mathbf{u} \cdot \mathbf{F}^{\mathbf{u}} &= \rho \, g_{mn} \, u^m \, F^n \\ &= g_{mn} \, u^m \, \tau^{mp}_{;p} \\ &= (g_{mn} \, u^m \, \tau^{mp})_{;p} - g_{mn} \, u^m_{;p} \, \tau^{mp} \\ &= (g_{mn} \, u^m \, \tau^{mp})_{;p} - e_{mp} \, \tau^{mp} \\ &\equiv \nabla \cdot (\mathbf{u} \cdot \boldsymbol{\tau}) - \rho \, \epsilon_M. \end{aligned} \tag{4.61}$$

In these equations, the labels m, n are summed over their horizontal range $m, n = 1, 2$ whereas the label p is summed over the full three-dimensional range $p = 1, 2, 3$. We introduced components to the strain tensor

$$2\, e_{mp} = u_{m\,;p} + u_{p\,;m} \tag{4.62}$$

whose properties are described in Section 15.3.1. We also noted that the covariant derivative of all metric tensor components vanish

$$g_{mn\,;p} = 0, \tag{4.63}$$

as discussed in Section 19.2.4.

The divergence $\nabla \cdot (\mathbf{u} \cdot \boldsymbol{\tau})$ vanishes against solid walls, yet locally it redistributes momentum via frictional stresses. As described in Section 15.4, the $\rho\, \epsilon_M$ term satisfies

$$e_{mp}\, \tau^{mp} \equiv \rho\, \epsilon_M \geq 0, \tag{4.64}$$

with zero resulting in the absence of strain. Hence, $-\rho\, \epsilon_M$ is a local sink of kinetic energy. Integrating $\rho \mathbf{u} \cdot \mathbf{F}^{\mathbf{u}}$ over a closed fluid system leaves just the dissipation term, since the total divergence term drops out. Therefore, as anticipated, friction dissipates the kinetic energy of a closed fluid system.

In the ocean, frictional dissipation associated with molecular viscosity is on the order

$$\epsilon_M \approx 10^{-9}\, \mathrm{W\, kg^{-1}}. \tag{4.65}$$

This dissipation causes a warming, through frictional heating, of only $10^{-3} K$ per hundred years (McDougall (2002)). This term is thus very small. For large-scale ocean modeling, kinetic energy dissipation is large and nontrivial due to the need to maintain numerical stability (see Chapters 15 and 16). Notably, it is not appropriate to transfer this kinetic energy dissipation in the models into a heating source. Doing so would create a large, and unphysical, diapycnal mass flux spuriously altering simulated water mass properties. It will be some decades before the levels of kinetic energy dissipation in global ocean models reach the small levels appropriate for the real ocean.

4.4.3 Gravitational potential energy

We consider here the gravitational potential energy of a fluid parcel, taken with respect to the reference level $z = 0$ at the surface of a resting ocean.

4.4.3.1 Non-Boussinesq parcels

The gravitational potential energy of a non-Boussinesq fluid parcel is

$$(\rho\, \mathrm{d}V)\, \mathcal{P} = (\rho\, \mathrm{d}V)\, g\, z, \tag{4.66}$$

where

$$\mathcal{P} = g\, z \tag{4.67}$$

is the potential energy per mass, and the gravitational acceleration g is assumed to be constant over the ocean fluid (Section 3.5.2.3). The material evolution of potential energy is thus given by

$$(\rho \, \mathrm{d}V) \, \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}t} = (\rho \, \mathrm{d}V) \, g \, w. \tag{4.68}$$

That is, only motion in the vertical direction, parallel to the gravitational force, alters the parcel's gravitational potential energy. Mixing, surface forcing, and SGS processes affect potential energy only so far as they affect the vertical velocity. Use of mass conservation in the form (4.42) leads to the Eulerian conservation law

$$(\rho \, \mathcal{P})_{,t} + \nabla \cdot (\rho \, \mathbf{v} \, \mathcal{P}) = w \, \rho \, g. \tag{4.69}$$

4.4.3.2 Boussinesq parcels

When developing budgets for a Boussinesq fluid parcel, the parcel is assumed to have mass $\rho_o \, \mathrm{d}V$, *except* when the parcel feels the gravitational force, where its mass is that of a non-Boussinesq parcel, $\rho \, \mathrm{d}V$. Hence, for purposes of deriving the potential energy budget, we proceed as for the non-Boussinesq case, yet constrain the parcel to maintain a constant volume instead of a constant mass

$$\begin{aligned} \frac{\mathrm{d}\,(\rho \, \mathrm{d}V \, \mathcal{P})}{\mathrm{d}t} &= g \, \mathrm{d}V \, \frac{\mathrm{d}\,(\rho \, z)}{\mathrm{d}t} \\ &= \mathrm{d}V \, \left(\rho \, g \, w + g \, z \, \frac{\mathrm{d}\rho}{\mathrm{d}t} \right). \end{aligned} \tag{4.70}$$

The $(\rho \, \mathrm{d}V) \, g \, w$ term is the same as for the non-Boussinesq parcel. Yet the $g \, z \, D\rho/\mathrm{d}t$ term is new. Again, it arises from the assumption that the Boussinesq parcel's volume remains constant, not its mass. Hence, if the parcel changes its density, as through fluxes of buoyancy, the parcel's mass must change to maintain its fixed volume. The mass changes via the presence of *virtual mass fluxes*. Since Boussinesq fluids are so common in ocean modeling, it is worth discussing these issues a bit further.

4.4.3.3 Distinctions between budgets

As just mentioned, the $\mathrm{d} \ln \rho/\mathrm{d}t$ term means that the gravitational potential energy of a Boussinesq fluid parcel is directly affected by buoyancy fluxes, including those at the sub-grid scale (SGS). A particular example is the SGS flux arising from the Gent and McWilliams (1990) closure, which provides a sink for potential energy, so long as there is baroclinicity (horizontal density gradients) in the fluid domain (see also Section 12.2.1). In contrast, strictly horizontal fluxes do not alter the gravitational potential energy, since they flux fluid along surfaces of constant geopotential.

To further discuss the different potential energy budgets, consider a parcel that is heated within a stratified ocean. Heating reduces the parcel's density, which increases its buoyancy and so causes it to develop an upward velocity until it reaches a new level of neutral buoyancy. If the parcel is non-Boussinesq, then its mass $(\rho \, \mathrm{d}V)$ remains fixed, so its gravitational potential energy increases upon moving upward: $g \, (\rho \, \mathrm{d}V) \, \mathrm{d}z/\mathrm{d}t > 0$. If the parcel is Boussinesq, its volume is fixed, so the potential energy change is given by $g \, \mathrm{d}V \, \mathrm{d}(z \, \rho)/\mathrm{d}t$. Notably, for a parcel at the

ocean surface with the rigid lid assumption, heating the Boussinesq parcel leads to a reduction in potential energy, whereas heating the non-Boussinesq parcel leads to no change. In general, it is unclear how the potential energy will change.

The distinction between the potential energy budgets for Boussinesq and non-Boussinesq fluids was emphasized by Xin Huang (personal communication, 1998), where he cautioned that because of the added $\mathrm{d}\ln\rho/\mathrm{d}t$ term, the Boussinesq budget can be quite unphysical. However, it remains to be determined how important for global ocean circulation it is to properly represent the potential energy budget.

4.4.4 Total mechanical energy

To derive a budget for the total mechanical energy (kinetic plus potential) of a seawater parcel, we start by providing further physical interpretations of the pressure term $-\mathbf{u}\cdot\nabla p$ appearing in the kinetic energy budget of Section 4.4.2. Use of the hydrostatic balance yields

$$\begin{aligned}\mathbf{u}\cdot\nabla p &= \mathbf{v}\cdot\nabla p - w\,p_{,z}\\ &= \nabla\cdot(\mathbf{v}\,p) - p\,\nabla\cdot\mathbf{v} + \rho\,g\,w,\end{aligned} \tag{4.71}$$

thus bringing the horizontal kinetic energy budget to the form

$$(\rho\,\mathcal{K})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathcal{K}) = -\nabla\cdot(\mathbf{v}\,p) + p\,\nabla\cdot\mathbf{v} - w\,\rho\,g + \rho\,\mathbf{u}\cdot\mathbf{F}^{\mathbf{u}}, \tag{4.72}$$

or

$$\rho\,\frac{\mathrm{d}(\mathcal{K}+\mathcal{P})}{\mathrm{d}t} = -\nabla\cdot(\mathbf{v}\,p) + p\,\nabla\cdot\mathbf{v} + \rho\,\mathbf{u}\cdot\mathbf{F}^{\mathbf{u}}, \tag{4.73}$$

where we brought the kinetic and gravitational energies per mass onto the left hand side. To facilitate comparison with subsequent results, we use (4.45) to relate the velocity divergence with changes in seawater specific volume, and introduce relation (4.61) for frictional dissipation to render

$$\rho\,\frac{\mathrm{d}(\mathcal{K}+\mathcal{P})}{\mathrm{d}t} = -\nabla\cdot(\mathbf{v}\,p - \mathbf{u}\cdot\boldsymbol{\tau}) - \rho\,\epsilon_M + \frac{p}{v_s}\frac{\mathrm{d}\,v_s}{\mathrm{d}t}, \tag{4.74}$$

where $v_s = \rho^{-1}$ is the specific volume. This budget indicates that mechanical energy is redistributed locally by the fluxes of pressure and friction, $\mathbf{v}\,p - \mathbf{u}\cdot\boldsymbol{\tau}$, and affected by sources due to pressure-work $p\,v_s^{-1}\,(\mathrm{d}\,v_s/\mathrm{d}t)$ and frictional dissipation $-\rho\,\epsilon_M$. Closure for the total energy budget comes about by noting that modifications to the mechanical energy by source terms are compensated by opposite signed sources for the internal energy budget, thus maintaining a conserved total mechanical energy plus internal energy. We further discuss these issues in Section 4.4.5.

4.4.5 Mechanical plus internal energies

As discussed in Secction 4.3, internal energy represents the energy of the molecular degrees of freedom that are averaged-out when formulating a continuum description of fluid motion. Consequently, the total energy per mass $\mathcal{E}$ (specific energy) of a fluid parcel is written

$$\mathcal{E} = \mathbf{v}^2/2 + \Phi + \mathcal{I} \tag{4.75}$$

with $\mathbf{v}^2/2$ the kinetic energy per mass, Φ the gravitational potential energy per mass, and $\mathcal{I}$ the internal energy per mass. Macroscopically, this relation defines the specific internal energy $\mathcal{I}$ as the total energy minus the mechanical energy. Microscopically, $\mathcal{I}$ embodies the energy of molecular thermal agitation and molecular interactions. For a hydrostatic fluid using the Traditional Approximation,

$$\mathcal{E} = \mathbf{u}^2/2 + g\,z + \mathcal{I}. \tag{4.76}$$

Energy conservation for a closed fluid system means that total energy per mass of a fluid parcel evolves according to

$$\rho\,\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t} = -\nabla\cdot(\rho\,\mathbf{J}_{\mathcal{E}}) \tag{4.77}$$

for some flux of energy $\mathbf{J}_{\mathcal{E}}$. Closure of the full fluid system means that the normal component to the energy flux vanishes at system boundaries. Based on our considerations of mechanical energy flux, we find it convenient to define the flux of total energy as

$$\rho\,\mathbf{J}_{\mathcal{E}} = \mathbf{v}\,p - \mathbf{u}\cdot\boldsymbol{\tau} + \rho\,\mathbf{J}_q. \tag{4.78}$$

In this equation, we introduced a *heat flux* $\mathbf{J}_q$ which in general is a function of temperature as well as concentration (see Fofonoff (1962), Gregg (1984), and Davis (1994a), and especially Sections 58 and 59 of Landau and Lifshitz (1987) for discussion). Subtracting equation (4.74) for mechanical energy evolution from equation (4.77) leads to the expression for internal energy evolution

$$\rho\,\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} = -\nabla\cdot(\rho\,\mathbf{J}_q) - \frac{p}{v_s}\frac{\mathrm{d}\,v_s}{\mathrm{d}t} + \rho\,\epsilon_M. \tag{4.79}$$

Internal energy of a parcel is thus affected by the convergence of heat fluxes, and sources due to pressure-work and frictional dissipation. In the absence of irreversible effects due to heat transport and momentum friction, internal energy is affected only by pressure work. Consequently, Boussinesq parcels, which are incompressible and so cannot receive pressure work, maintain a constant internal energy when undergoing reversible transport.

4.5 Global energy balances

We present here the mechanical energy balances maintained over the full ocean domain by a continuum hydrostatic fluid. The global balances for mechanical plus internal energy follow analogously. This section is somewhat outside the main line of development in this chapter. However, it is presented here since it is useful to consider analogous balances in their discrete form when constructing the numerical algorithms in an ocean model.

4.5.1 Generic form of the global energy budget

The generic form of energy within a finite region is

$$\begin{aligned}\Psi &= \int (\rho \, \mathrm{d}V) \, \psi \\ &= \int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z \, \rho \, \psi,\end{aligned} \tag{4.80}$$

where $\mathrm{d}A = \mathrm{d}x \, \mathrm{d}y$ is the horizontal area element, and ψ the energy per mass. Time tendencies of Ψ are affected by the passage of energy across the domain boundaries (e.g., wind stresses, fresh water fluxes, heat fluxes, etc.), and changes in the amount of energy within the ocean domain (e.g., source/sink terms, pressure work, dissipation, etc.).

To mathematically express these changes, we first consider the integral of mass weighted material transport of energy

$$\int (\rho \mathrm{d}V) \, \frac{\mathrm{d}\psi}{\mathrm{d}t} = \int \mathrm{d}V [(\rho \, \psi)_{,t} + \nabla \cdot (\rho \, \mathbf{v} \, \psi)]. \tag{4.81}$$

We now use Leibnitz's Rule to interchange differentiation and integration. Use of this rule is quite common when deriving budgets over finite regions, and so it is useful to become familiar with the following manipulations. The local time derivative term becomes

$$\int_{-H}^{\eta} \mathrm{d}z \, (\rho \, \psi)_{,t} = -\eta_{,t} \, (\rho \, \psi)_{z=\eta} + \partial_t \left(\int_{-H}^{\eta} \mathrm{d}z \, \rho \, \psi \right) \tag{4.82}$$

and the horizontal divergence term

$$\begin{aligned}\int_{-H}^{\eta} \mathrm{d}z \, \nabla \cdot (\rho \, \mathbf{u} \, \psi) = \nabla \cdot \left(\int_{-H}^{\eta} \mathrm{d}z \, \rho \, \mathbf{u} \, \psi \right) \\ - (\rho \, \psi \, \mathbf{u} \cdot \nabla \eta)_{z=\eta} - (\rho \, \psi \, \mathbf{u} \cdot \nabla H)_{z=-H}\end{aligned} \tag{4.83}$$

and the vertical divergence term

$$\int_{-H}^{\eta} \mathrm{d}z \, \partial_z \, (\rho \, w \, \psi) = (\rho \, \psi \, w)_{z=\eta} - (\rho \, \psi \, w)_{z=-H}. \tag{4.84}$$

Combining these results leads to

$$\begin{aligned}\int (\rho \mathrm{d}V) \, \frac{\mathrm{d}\psi}{\mathrm{d}t} = &- \int_{z=\eta} \mathrm{d}A \, [\rho \, \psi \, (\eta_{,t} - w + \mathbf{u} \cdot \nabla \eta)] \\ &- \int_{z=-H} \mathrm{d}A \, [\rho \, \psi \, (w + \mathbf{u} \cdot \nabla \eta)] \\ &+ \int \mathrm{d}A \left(\partial_t \int_{-H}^{\eta} \mathrm{d}z \, \rho \, \psi + \nabla \cdot \int_{-H}^{\eta} \mathrm{d}z \, \rho \, \mathbf{u} \, \psi \right).\end{aligned} \tag{4.85}$$

Use of the bottom and surface kinematic boundary conditions (Section 3.3) yields

$$\int (\rho \mathrm{d}V)\, \frac{\mathrm{d}\psi}{\mathrm{d}t} = -\int\limits_{z=\eta} \mathrm{d}A\, \rho_w\, q_w\, \psi + \int \mathrm{d}A \left(\partial_t \int\limits_{-H}^{\eta} \mathrm{d}z\, \rho\, \psi + \nabla \cdot \int\limits_{-H}^{\eta} \mathrm{d}z\, \rho\, \mathbf{u}\, \psi \right). \quad (4.86)$$

The horizontal boundary conditions include either no-normal flux solid walls or periodic conditions. In either case, the horizontal divergence term vanishes. Rearrangement thus leads to

$$\Psi_{,t} = \int\limits_{z=\eta} \mathrm{d}A\, \rho_w\, q_w\, \psi + \int (\rho \mathrm{d}V)\, \frac{\mathrm{d}\psi}{\mathrm{d}t}, \quad (4.87)$$

where we noted that $\mathrm{d}A = dy\, dy$ is independent of time. The left hand side represents the time tendency of domain integrated energy. In the first term on the right hand side, $\mathrm{d}A\, \rho_w\, q_w$ is the mass per time of fresh water crossing the ocean surface, and so $(\mathrm{d}A\, \rho_w\, q_w)\, \psi$ is the energy per time crossing the surface. The final term represents material transport of energy integrated over the ocean domain.

The manipulations leading to the balance (4.87) are quite general and apply to other scalar densities, such as tracer concentrations. Notably, with $\psi = 1$ the balance reduces to mass conservation discussed in Section 3.3.3. For Boussinesq fluids, $\rho_w \to \rho$ in the $z = \eta$ surface term, and for the volume integrated term, $\rho \to \rho_o$, except for the potential energy budget.

4.5.2 Kinetic energy budget

The kinetic energy of a hydrostatic fluid is given by

$$K = \int (\rho\, \mathrm{d}V)\, \mathcal{K}, \quad (4.88)$$

where $\mathcal{K} = \mathbf{u}^2/2$ is the kinetic energy per mass of the parcel arising from horizontal motion. With $\psi = \mathcal{K}$, results from Section 4.5.1 yield

$$K_{,t} = \int\limits_{z=\eta} \mathrm{d}A\, \rho_w\, q_w\, \mathcal{K} + \int (\rho\, \mathrm{d}V)\, \mathbf{u} \cdot \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}. \quad (4.89)$$

The first term represents the passage of kinetic energy through the ocean surface with fresh water fluxes. The momentum equation (4.54) leads to $\mathbf{u} \cdot \rho\, \mathrm{d}\mathbf{u}/\mathrm{d}t = \mathbf{u} \cdot (-\nabla p + \rho\, \mathbf{F}^{\mathbf{u}})$, and so

$$K_{,t} = \int\limits_{z=\eta} \mathrm{d}A\, \rho_w\, q_w\, \mathcal{K} + \int \mathrm{d}V\, \mathbf{u} \cdot (-\nabla p + \rho\, \mathbf{F}^{\mathbf{u}}). \quad (4.90)$$

The interpretation of the second term follows that given for an individual fluid parcel in Section 4.4.2. The kinetic energy budget for a Boussinesq fluid results from setting $\rho_w \to \rho$ in the surface term, and $\rho \to \rho_o$ in the volume term.

The globally integrated contribution from $\mathbf{u} \cdot \mathbf{F}^{\mathbf{u}}$ in the ocean interior is often negative semi-definite, thus representing a dissipation. We encountered such when discussing the frictional dissipation of a fluid parcel in Section 4.4.2. For the present discussion, we highlight the effects of vertical deformations, in which case the friction force typically takes the form

$$\rho \, \mathbf{F}^{\mathbf{u}}_{\text{vert}} = (\kappa \, \rho \, \mathbf{u}_{,z})_{,z} \tag{4.91}$$

where $\kappa > 0$ is a vertical viscosity. Hence,

$$\begin{aligned} \mathbf{u} \cdot \rho \, \mathbf{F}^{\mathbf{u}} &= \mathbf{u} \cdot (\kappa \, \rho \, \mathbf{u}_{,z})_{,z} + \mathbf{u} \cdot \rho \, \mathbf{F}^{\mathbf{u}}_h \\ &= (\kappa \, \rho \, \mathbf{u} \cdot \mathbf{u}_{,z})_{,z} - \kappa \, \rho \, (\mathbf{u}_{,z})^2 + \mathbf{u} \cdot \rho \, \mathbf{F}^{\mathbf{u}}_h, \end{aligned} \tag{4.92}$$

where $\rho \, \mathbf{F}^{\mathbf{u}}_h$ is the friction force associated with horizontal deformations. Vertical integration of the $(\kappa \, \rho \, \mathbf{u} \cdot \mathbf{u}_{,z})_{,z}$ term picks out the contribution from the ocean surface and bottom boundaries. These terms arise from boundary layer interactions between the ocean and the overlying atmosphere/ice systems, or the underlying solid earth. The surface stress is typically that imparted by the atmospheric winds which increases the ocean's kinetic energy. The bottom stress is often associated with currents moving over rough small scale topography as well as tidal dissipation. However, for coarse resolution ocean models containing subgrid-scale processes, one can have such interactions between topography and meso-scale eddies act to increase the kinetic energy, as described by Holloway (1992). Further details of the stresses remain part of the difficult sub-grid scale parameterization problem. For now, they remain unspecified and will be written

$$\rho \, \kappa \, \mathbf{u}_{,z} = \boldsymbol{\tau}^{wind} \qquad z = \eta \tag{4.93}$$

for the ocean surface, and

$$\rho \, \kappa \, \mathbf{u}_{,z} = \boldsymbol{\tau}^{bottom} \qquad z = -H \tag{4.94}$$

for the ocean bottom, thus bringing the kinetic energy budget to the form

$$\begin{aligned} K_{,t} = \int_{z=\eta} \mathrm{d}A \, (\mathbf{u} \cdot \boldsymbol{\tau}^{wind} + \rho_w \, q_w \, \mathcal{K}) - \int_{z=-H} \mathrm{d}A \, \mathbf{u} \cdot \boldsymbol{\tau}^{bottom} \\ + \int \mathrm{d}V \, \mathbf{u} \cdot (-\nabla p + \rho \, \mathbf{F}^{\mathbf{u}}_h - \kappa \, \rho \, (\mathbf{u}_{,z})^2). \end{aligned} \tag{4.95}$$

The $-\kappa \, \rho \, (\mathbf{u}_{,z})^2 \leq 0$ term dissipates kinetic energy via vertical deformations interior to the ocean domain, consistent with our discussions in Section 4.4.2.

4.5.3 Gravitational potential energy budgets

We consider here the potential energy budgets for non-Boussinesq and Boussinesq fluid parcels.

4.5.3.1 Non-Boussinesq fluid

A finite domain of fluid has gravitational potential energy

$$P = \int (\rho \, \mathrm{d}V) \, \mathcal{P} \tag{4.96}$$

where $\mathcal{P} = g\,z$ is the potential energy per mass of a fluid parcel. Results from Section 4.5.1, with $D\psi/\mathrm{d}t = g\,w$, lead to the global potential energy budget for a non-Boussinesq fluid

$$P_{,t} = \int_{z=\eta} \mathrm{d}A \, (g\,z\,\rho_w\,q_w) + \int (\rho \mathrm{d}V) \, g\,w. \tag{4.97}$$

Hence, the potential energy increases when the height of the ocean surface increases upon the introduction of fresh water (first term) or when the vertical motion causes mass within the ocean domain to rise (second term).

4.5.3.2 Boussinesq fluid

The total gravitational potential energy of a Boussinesq fluid has time evolution

$$P_{,t} = g \int_{z=\eta} \mathrm{d}A \, (\rho\,z\,\eta_{,t}) + g \int \mathrm{d}V \, (z\,\rho_{,t}). \tag{4.98}$$

The mass conservation identity $\rho_{,t} = -\nabla \cdot (\mathbf{v}\,\rho)$ is not generally available here for the volume conserving Boussinesq fluid. Instead, for density written as a function of potential temperature, salinity, and pressure, $\rho = \rho(\theta, s, p)$, we have

$$\rho_{,t} = -\rho\,\alpha\,\theta_{,t} + \rho\,\beta\,s_{,t} + c_s^{-2}\,p_{,t}, \tag{4.99}$$

where $\alpha = -(\partial \ln \rho/\partial\theta)_{p,s}$ is the thermal expansion coefficient, $\beta = (\partial \ln \rho/\partial s)_{p,\theta}$ is the saline contraction coefficient, and $c_s^2 = (\partial p/\partial\rho)_{s,\theta}$ is the squared speed of sound (see Section 4.8.1).

To proceed, we consider the special case with density equal to potential density, and we linearize the equation of state (see discussion in Section 4.8.3). For this case, density's time tendency can be written

$$\rho_{,t} = -\nabla \cdot (\rho\,\mathbf{v} + \mathbf{F}), \tag{4.100}$$

where $\mathbf{F}$ represents any fluxes that act to mix or stir density. With a nonlinear equation of state, the non-flux form processes of cabbeling, thermobaricity, and halobaricity (see Section 12.1.7 as well as McDougall (1987a,b) must be included on the right hand side of the density equation. We drop these, and other source/sink terms, for brevity. Without the mixing/stirring fluxes, the evolution equation (4.100) is mathematically the same as the mass conservation equation. Notably, however, for the present case, besides linearizing the equation of state, the velocity field is assumed non-divergent $\nabla \cdot \mathbf{v} = 0$.

Manipulations similar to those in Section 4.5.1 lead to

$$\int \mathrm{d}V \, z \, \nabla \cdot \mathbf{F} = \int \mathrm{d}A \, (\eta \, \hat{\mathbf{N}} \cdot \mathbf{F})_{z=\eta} + \int \mathrm{d}A \, (H \, \hat{\mathbf{N}} \cdot \mathbf{F})_{z=-H} - \int \mathrm{d}V \, F^z, \tag{4.101}$$

where

$$\hat{\mathbf{N}} = (-\nabla z, 1) \tag{4.102}$$

is an orientation vector. At the ocean surface where $z = \eta$, $\hat{\mathbf{N}}$ points away from the ocean domain. In contrast, at the solid boundaries where $z = -H$, $\hat{\mathbf{N}}$ points into the ocean domain. Assuming no density enters the ocean through the solid boundaries allows us to drop the $(H\,\hat{\mathbf{N}} \cdot \mathbf{F})_{z=-H}$ term.

Combining with the results from Section 4.5.1 leads to the potential energy budget

$$P_{,t} = g \int\limits_{z=\eta} \mathrm{d}A\, \eta\, (\rho\, q_w - \hat{\mathbf{N}} \cdot \mathbf{F}) + g \int \mathrm{d}V\, (w\, \rho + F^z). \tag{4.103}$$

This budget affords a physical interpretation largely consistent with that discussed for the Boussinesq parcel in Section 4.4.3.2. Notably, $-\hat{\mathbf{N}} \cdot \mathbf{F}$ represents surface buoyancy fluxes that either increase or decrease the Boussinesq fluid's potential energy, depending on the sign of the flux and its height relative to $z = 0$. Additionally, the volume integrated vertical flux $\int \mathrm{d}V\, F^z$ acts in a similar manner as the vertical velocity, transporting density against or with the gravitational field. In particular, Griffies (1998) showed that the vertical skew-flux arising the Gent and McWilliams (1990) closure is negative semi-definite, thus acting to reduce potential energy, so long as there remains a non-zero horizontal density gradient in the fluid (see also Section 12.2.1). Horizontal fluxes do not alter gravitational potential energy, since they flux fluid mass parallel to geopotential surfaces.

4.6 Basic non-equilibrium thermodynamics

The purpose of this section is to introduce some basic notions of non-equilibrium thermodynamics. In particular, we derive the equation for the evolution of entropy within a fluid parcel. For this purpose, we start with the fundamental thermodynamic relation (4.26) $\mathrm{d}\mathcal{I} = T\,\mathrm{d}\zeta - p\,\mathrm{d}v_s + \mu\,\mathrm{d}C$. This relation holds for a system infinitesimally close to thermodynamic equilibrium. If we assume that each fluid parcel is in local thermodynamic equilibrium, yet allow the full ocean system to be out of equilibrium, we can determine a parcel's internal energy time evolution via

$$\rho\, \frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} = \rho\, T\, \frac{\mathrm{d}\zeta}{\mathrm{d}t} - \frac{p}{v_s}\, \frac{\mathrm{d}v_s}{\mathrm{d}t} + \mu\, \rho\, \frac{\mathrm{d}C}{\mathrm{d}t}. \tag{4.104}$$

In this manner we are able to transfer the methods of equilibrium thermodynamics to the linear irreversible thermodyanmics of moving fluid parcels. Note that the term *linear* here refers to our assumption that the system is close to thermodynamic equilibrium. In this case, the dissipative thermodynamic fluxes are linear functions of the gradients of the thermodynamic state variables. Nonlinear effects are *not* absent, however, due to the nonlinear advective transport, nonlinear source terms, nonlinear equations of state, and nonlinear dependence of the transport coeffecients. DeGroot and Mazur (1984) provide a full accounting of this subject, and Gregg (1984) and Davis (1994a) apply these methods to small ocean scale mixing. Slightly different formulations can be found in Landau and Lifshitz (1987) and Batchelor (1967), and we prefer their approaches for the following.

Inserting equation (4.79) for the evolution of internal energy within equation (4.104) leads to the expression for evolution of entropy in a seawater parcel

$$T\,\rho\,\frac{\mathrm{d}\zeta}{\mathrm{d}t} = -\nabla\cdot(\rho\,\mathbf{J}_q) + \rho\,\epsilon_M - \mu\,\rho\,\frac{\mathrm{d}C}{\mathrm{d}t}. \tag{4.105}$$

Entropy of a fluid parcel evolves by three irreversible mixing processes: (1) convergence of heat fluxes, (2) mixing of momentum in a viscous fluid, thus creating frictional dissipation sources which ultimately increase a parcel's heat content, and (3) mass exchange resulting in salinity mixing. Correspondingly, a parcel generally maintains constant entropy if processes associated with its evolution are adiabatic, frictionless, and isohaline. Since the friction source is very small in the ocean, adiabatic isohaline transport is very nearly isentropic. Indeed, when ocean modelers refer to adiabatic and isohaline processes, they typically assume this to be synonymous with isentropic. It is important to note that the parcel's entropy may remain unchanged even in the event of heating and salinity diffusion. The reason is that these effects can compensate one another, thus cancelling out. This process does not break the Second Law of thermodynamics, since the parcel is an open system and so its entropy need not always increase when mixing events occur. Consequently, a frictionless, adiabatic, isohaline transport is isentropic, yet isentropic transport can involve non-adiabatic and non-isohaline effects.

To identify local entropy sources and fluxes of entropy, let us assume salinity mixing occurs via the convergence of a salinity flux*

$$\rho\,\frac{\mathrm{d}C}{\mathrm{d}t} = -\nabla\cdot(\rho\,\mathbf{J}_C) \tag{4.106}$$

thus leading to

$$\begin{aligned}\rho\,\frac{\mathrm{d}\zeta}{\mathrm{d}t} &= -\frac{1}{T}\,\nabla\cdot(\rho\,\mathbf{J}_q) + \frac{\mu}{T}\,\nabla\cdot(\rho\,\mathbf{J}_C) + (\rho/T)\,\epsilon_M \\ &= -\nabla\cdot(\rho\,\mathbf{J}_\mathbf{q}/T - \mu\,\rho\,\mathbf{J}_C/T) \\ &\quad + \rho\,\mathbf{J}_q\cdot\nabla(1/T) - \rho\,\mathbf{J}_C\cdot\nabla(\mu/T) + (\rho/T)\,\epsilon_M.\end{aligned} \tag{4.107}$$

We can thus identify the entropy flux

$$\rho\,\mathbf{J}_\zeta = \rho\,\mathbf{v}\,\zeta + \rho\,\mathbf{J}_\mathbf{q}/T - \mu\,\rho\,\mathbf{J}_C/T \tag{4.108}$$

and entropy source

$$\sigma = \rho\,\mathbf{J}_q\cdot\nabla(1/T) - \rho\,\mathbf{J}_C\cdot\nabla(\mu/T) + (\rho/T)\,\epsilon_M. \tag{4.109}$$

The entropy source vanishes when all parcels are in thermodynamic equilibrium with one another, which is the case when the temperature and chemical potential are uniform throughout the ocean, and there is an absence of strain thus eliminating frictional dissipation. Gregg (1984) details the form of entropy sources in the ocean associated with small scale mixing processes.

*Non-flux terms causing salinity mixing are not considered here.

4.7 Thermodynamical tracers

Heating and cooling as well as mass exchange predominantly occur near the ocean surface (neglecting geothermal effects). In contrast, transport in the interior is nearly adiabatic and isohaline (and so nearly isentropic; e.g., equation (4.105)). Hence, the surface ocean experiences irreversible processes that set characteristics of the water masses moving quasi-isentropically within the ocean interior. Useful labels for these water masses maintain their values when moving within the largely ideal ocean interior. Salinity is a good tracer for such purposes since it is altered predominantly by mixing between waters of varying concentrations. Such constitutes a basic property of material tracers as considered in Section 4.2.1. We discuss here desirable properties of a thermodynamic tracer that tags the heat within a water parcel and evolves analogously to material tracers.

4.7.1 Potential temperature

Vertical isentropic motion in the ocean changes a fluid parcel's hydrostatic pressure, which thus causes its *in situ* temperature to change in proportion to the adiabatic lapse rate as given by equation (4.37), $\mathrm{d}T = \Gamma \mathrm{d}p$. Consequently, *in situ* temperature is not a useful thermodynamic variable to label water parcels of common origin. Instead, it is more useful to remove the adiabatic pressure effects.

Removing adiabatic pressure effects from *in situ* temperature leads to the concept of *potential temperature*, which is the *in situ* temperature that a water parcel of fixed composition would have if it were isentropically transported from its *in situ* pressure to a reference pressure p_r, with the reference pressure typically taken at the ocean surface. Mathematically, potential temperature θ is the reference temperature obtained via an integration of $\mathrm{d}T = \Gamma \mathrm{d}p$ for an isentropic *in situ* temperature change with respect to pressure (e.g., Apel (1987))

$$T = \theta(s, T, p_r) + \int_{p^r}^{p} \Gamma(s, \theta, p')\, \mathrm{d}p' \tag{4.110}$$

with Γ the lapse rate defined in terms of changes in pressure (equation (4.38)). By definition, *in situ* temperature T equals the potential temperature θ at the reference pressure $p = p^r$. Elsewhere, they differ by an amount determined by the adiabatic lapse rate. Beneath the diabatic surface mixed layer, a vertical profile of potential temperature is far more constant than *in situ* temperature, where *in situ* temperature increases at depth according to the lapse rate.

Potential temperature of a parcel is constant when the parcel's entropy is constant, by definition. We see this fact mathematically by noting that when entropy changes at a fixed pressure $p = p_r$ so that temperature equals potential temperature, equation (4.36) leads to

$$\mathrm{d}\zeta = C_p \, \mathrm{d} \ln \theta, \tag{4.111}$$

implying $\mathrm{d}\zeta = 0$ if and only if $\mathrm{d}\theta = 0$.

4.7.2 Potential enthalpy

The above properties of potential temperature are quite convenient and so have proved useful for many oceanographic purposes. However, we have yet to ask whether it is a convenient variable to mark the heat content in a parcel of seawater. Traditionally, it is the potential temperature multiplied by the heat capacity that is used for this purpose. Bacon and Fofonoff (1996) provide a review with suggestions for this approach. In contrast, McDougall (2002) argues that potential temperature multiplied by heat capacity is less precise, by some two orders of magnitude, than an alternative thermodyanamic tracer called *potential enthalpy*.

To understand this issue from a mathematical perspective, consider the evolution equation for potential temperature

$$\rho \frac{\mathrm{d}\theta}{\mathrm{d}t} = -\nabla \cdot (\rho \mathbf{F}_\theta) + \rho \mathcal{S}, \tag{4.112}$$

where $\mathbf{F}_\theta$ is a flux due to molecular diffusion, and $\mathcal{S}$ is a source. That potential temperature evolves in this manner is ensured by its being a scalar field. Consider the mixing of two seawater parcels at the same pressure where the parcels have different potential temperature and salinity. In the absence of the source term, the equlibrated state consists of a single parcel with mass equal to the sum of the two separate masses, and a potential temperature and salinity determined by their respective mass weighted means. Does this actually happen in the real ocean? That is, can we actually ignore source terms? Fofonoff (1962) and McDougall (2002) note that it is indeed the case for salinity (and any other material tracer), yet it is not the case for potential temperature. Instead, potential temperature contains source terms that alter the mass weighted average equilibrated state. In contrast, potential enthalpy (discussed below) maintains the desired conservative behaviour when mixing at constant pressure, and nearly maintains this behaviour if mixing parcels at different pressure. Hence, ocean modelers that set the source term to zero upon mixing potential temperature are in error, and McDougall (2002) quantifies this error.

Potential enthalpy is defined analogously to potential temperature. What motivates the use of potential enthalpy is the observation that the fundamental relation between thermodynamic state variables takes a nearly conservative form when written in terms of potential enthalpy. To see this point, we return to the evolution of internal energy given by equation (4.79). Introducing the enthalpy per mass (specific enthalpy)

$$\mathcal{H} = \mathcal{I} + p\, v_s \tag{4.113}$$

leads to

$$\rho \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = -\nabla \cdot (\rho \mathbf{J}_q) + \frac{\mathrm{d}p}{\mathrm{d}t} + \rho\, \epsilon_M. \tag{4.114}$$

Dropping the negligible frictional dissipation (Section 4.4.2) leads to the approximate statement

$$\rho \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} - \frac{\mathrm{d}P}{\mathrm{d}t} \approx -\nabla \cdot (\rho \mathbf{J}_q). \tag{4.115}$$

To proceed, note that the fundamental thermodynamic relation (4.26) becomes

$$\mathrm{d}\mathcal{H} = T\, \mathrm{d}\zeta + v_s\, \mathrm{d}p + \mu\, \mathrm{d}C \tag{4.116}$$

in terms of enthalpy. Thus, enthalpy can be written as a function of entropy, pressure, and salinity

$$\mathcal{H} = \mathcal{H}(\zeta, v_s, C). \tag{4.117}$$

Transport of a seawater parcel without changing heat, salt, or momentum occurs without change in entropy (equation (4.105)), thus allowing us to write

$$\left(\frac{\partial \mathcal{H}}{\partial p}\right)_{\zeta,C} = v_s. \tag{4.118}$$

Keeping salinity and entropy fixed (or equivalently fixed salinity and potential temperature; see equation (4.111)) leads to

$$\mathcal{H}(\theta, s, p) = \mathcal{H}^o(\theta, s, p_r) + \int_{p^r}^{p} v_s(\theta, s, p')\, \mathrm{d}p' \tag{4.119}$$

with $\mathcal{H}^o(\theta, s, p_r)$ defining the potential enthalpy of a parcel with potential temperature θ and salinity s. Taking the time derivative and using the approximate relation (4.115) renders

$$\frac{\mathrm{d}\mathcal{H}^o}{\mathrm{d}t} \approx -v_s\, \nabla \cdot (\rho\, \mathbf{J}_q) + \int_{p}^{p^r} \mathrm{d}p'\, \frac{D v_s(\theta, s, p')}{\mathrm{d}t}. \tag{4.120}$$

McDougall (2002) shows that for the ocean, the integral

$$\begin{aligned} \int_{p}^{p^r} \mathrm{d}p'\, \frac{D v_s(\theta, s, p')}{\mathrm{d}t} &= \int_{p}^{p^r} \mathrm{d}p' \left(\frac{\partial v_s}{\partial \theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} + \frac{\partial v_s}{\partial s}\frac{\mathrm{d}s}{\mathrm{d}t}\right) \\ &= \frac{\mathrm{d}\theta}{\mathrm{d}t} \int_{p}^{p^r} \mathrm{d}p'\, v_s\, \alpha - \frac{\mathrm{d}s}{\mathrm{d}t} \int_{p}^{p^r} \mathrm{d}p'\, v_s\, \beta \end{aligned} \tag{4.121}$$

has a size on the order of the ocean's frictional dissipation. In these expressions, we introduced the thermal expansion coefficient $\alpha = \partial \ln v_s / \partial \theta$ and saline contraction coefficient $\beta = -\partial \ln v_s / \partial s$ (see Section 4.8). Note we removed the time derivatives of the potential temperature and salinity from the pressure integrals, since they are each independent of pressure. Given the smallness of $\int_p^{p^r} \mathrm{d}p'\, D v_s / \mathrm{d}t$, we can write the approximate potential enthalpy equation

$$\rho\, \frac{\mathrm{d}\mathcal{H}^o}{\mathrm{d}t} \approx -\nabla \cdot (\rho\, \mathbf{J}_q). \tag{4.122}$$

Hence, potential enthalpy is a state function that approximately specifies the heat in a parcel of seawater, and it evolves analogously to a material tracer such as salinity. See McDougall (2002) for a proof that $\mathcal{H}^o$ more accurately sets the heat for a parcel of seawater than $C_p\, \theta$. Given that it does, McDougall suggest that *conservative temperature*

$$\theta_{\mathcal{H}} \equiv \frac{\mathcal{H}^o(\theta, s, p_r)}{C_p^o} \tag{4.123}$$

with $p_r = 0$ is more appropriate than potential temperature as a thermodynamic tracer for use in an ocean model, and generally for measuring heat in the ocean. In this equation

$$\begin{aligned} C_p^o &= \frac{\mathcal{H}(\theta = 25^\circ C, s = 35\text{psu}, p_r = 0)}{25^\circ C} \\ &= 3989.24495292815 \,\text{Joules}\,\text{kg}^{-1}\,\text{K}^{-1} \end{aligned} \tag{4.124}$$

is a heat capacity chosen to minimize the difference between $C_p^o \theta$ and potential enthalpy $\mathcal{H}^o(\theta, s, p_r)$ when averaged over the sea surface.

In the remainder of this book, we maintain the notation θ, recognizing the fact that $\theta_{\mathcal{H}}$ may instead be used. All formulae for density and thermodynamic fluxes can be generalized, as shown by McDougall (2002). From a fundamental perspective, McDougall (2002) provides a compelling case for the use of conservative temperature. Nonetheless, it remains a research topic to determine the significance to simulated ocean circulation of errors made in numerical models using potential temperature rather than conservative temperature.

4.8 Ocean density

Density is an important variable to measure in the ocean and to compute in an ocean model. In particular, variations in the density field, via the hydrostatic balance (equation (4.53)), provide one of the most important driving forces for the large-scale circulation. In this section we summarize various forms of density as used in oceanography and ocean models.

4.8.1 *in situ* density

As mentioned earlier, density at a point in the ocean (often called the *in situ* density) is generally a function of temperature, salinity, and pressure

$$\rho = \rho(T, s, p). \tag{4.125}$$

This equation is known as the *equation of state*, and its form is determined empirically. Recall that equation (4.110) provides a unique relation between potential temperature θ and temperature, given salinity and pressure. Hence, density can just as well be expressed as a function of potential temperature, salinity, and pressure

$$\rho = \rho(\theta, s, p). \tag{4.126}$$

Likewise, McDougall (2002) indicates that density can be written as a function of conservative temperature $\theta_{\mathcal{H}}$ defined in Section 4.7.2.

Writing the equation of state in terms of θ (or $\theta_{\mathcal{H}}$) is more convenient for ocean models than in terms of *in situ* temperature, because ocean models prognostically solve for potential temperature (or conservative temperature). The most accurate equation of state for use in ocean models using potential temperature has been given by McDougall et al. (2002b), based on the work of Feistel (1993) and Feistel and Hagen (1995). Most ocean models are now switching to such accurate formulae since the earlier approximate forms, such as Bryan and Cox (1972), maintain

a relatively narrow range of salinity variations over which the equation is valid. With ocean models of refined grid resolution and realistic fresh water forcing, it is desirable to remove such limitations since model salinity can vary quite widely, especially near river mouths.

Given the functional relation $\rho = \rho(\theta, s, p)$, we can develop the material time derivative of density

$$\frac{\mathrm{d}\ln\rho}{\mathrm{d}t} = \left(\frac{\partial \ln\rho}{\partial p}\right)_{s,\theta} \frac{\mathrm{d}P}{\mathrm{d}t} + \left(\frac{\partial \ln\rho}{\partial s}\right)_{p,\theta} \frac{\mathrm{d}s}{\mathrm{d}t} + \left(\frac{\partial \ln\rho}{\partial \theta}\right)_{p,s} \frac{\mathrm{d}\theta}{\mathrm{d}t}. \tag{4.127}$$

Introducing the thermal expansion and saline contraction coefficients

$$\alpha = -\left(\frac{\partial \ln\rho}{\partial \theta}\right)_{p,s} \tag{4.128}$$

$$\beta = \left(\frac{\partial \ln\rho}{\partial s}\right)_{p,\theta} \tag{4.129}$$

and the squared speed of sound

$$c_s^2 = \left(\frac{\partial p}{\partial \rho}\right)_{s,\theta} \tag{4.130}$$

leads to

$$\begin{aligned} \frac{\mathrm{d}\ln\rho}{\mathrm{d}t} &= \frac{1}{\rho\, c_s^2}\frac{\mathrm{d}p}{\mathrm{d}t} + \beta\frac{\mathrm{d}s}{\mathrm{d}t} - \alpha\frac{\mathrm{d}\theta}{\mathrm{d}t} \\ &= -\nabla\cdot\mathbf{v} \end{aligned} \tag{4.131}$$

where the last step used mass conservation (Section 3.3.1). For fluid parcels undergoing general isentropic motion (frictionless motion at constant potential temperature and constant salinity; see equation (4.105)), pressure satisfies

$$\begin{aligned} \rho\,\mathrm{d}p/\mathrm{d}t &= (\rho\, p)_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,p) \\ &= -(\rho\, c_s)^2\,\nabla\cdot\mathbf{v}, \end{aligned} \tag{4.132}$$

whose linearized fluctuations are known as *acoustic modes*.

4.8.2 Potential density

Since isentropic motion of a fluid parcel generally occurs at constant potential temperature and constant salinity (see equation (4.105)), it is convenient to combine the evolution of these two active tracers into the evolution of a single variable. *Potential density* is one such combination. By definition, potential density ρ_{pot} is the density a fluid parcel would have if isentropically moved to a reference pressure p_r, often taken as pressure at the ocean surface

$$\rho_{pot} = \rho(\theta, s, p_r). \tag{4.133}$$

Hence, the material evolution of potential density is given by

$$\frac{\mathrm{d}\ln\rho_{pot}}{\mathrm{d}t} = \beta\,\frac{\mathrm{d}s}{\mathrm{d}t} - \alpha\,\frac{\mathrm{d}\theta}{\mathrm{d}t}, \tag{4.134}$$

where α and β are evaluated at the reference pressure p_r. We see that for general isentropic motion where potential temperature and salinity are materially conserved and friction dissipation is negligible (see equation (4.105)), potential density is likewise materially conserved. This behaviour is in contrast to *in situ* density, whose evolution is given by equation (4.131).

In many parts of the ocean, especially those close to the reference pressure, potential density is uniformly stacked in the vertical and thus forms a useful method for layering the ocean. Many physical processes related to tracer transport naturally occur within, rather than across, potential density layers. More precisely, these processes occur predominantly along *neutral directions* (e.g., McDougall (1987a); Gent and McWilliams (1990); Griffies et al. (1998)), which are directions tangent to the locally defined (local values of θ, s, p) potential density surface. We have more to say about *neutral physics* processes in Chapters 7, 11–14.

4.8.3 Idealized equations of state

For certain purposes, it is useful to approximate the equation of state used in ocean models. One common idealization is to equate the *in situ* density to the potential density

$$\rho(\theta, s, p) = \rho_{pot}(\theta, s, p_r). \tag{4.135}$$

Models using this idealized thermodynamics remove pressure effects from the *in situ* density, thus becoming incompressible or Boussinesq. The use of potential density in such models to compute hydrostatic pressure causes inaccuracies in the horizontal pressure gradient. In particular, the thermal wind relations (Section 3.7.4), valid for geostrophic flows, are compromised. These inaccuracies are typically unimportant for idealized studies, such as dynamical studies with adiabatic isopycnal layered models (e.g., Chapter 5). Yet the inaccuracies can be important for realistic simulations, thus necessitating a more accurate approach (e.g., Sun et al. (1999)).

To further idealize the Boussinesq fluid's thermodynamics, it is convenient to compute potential density as a linear function of potential temperature and salinity

$$\rho_{pot} = \rho_o\,[1 - \alpha\,(\theta - \theta_o) + \beta\,(s - s_o)], \tag{4.136}$$

where α, β, θ_o and s_o are constants. Indeed, some authors consider this expression for density to be part of the Boussinesq Approximation (e.g., Chandrasekhar (1961)), although such is distinctly not the case in the oceanographic literature. Those wishing to avoid complications of two active tracers further simplify this expression by dropping the salinity dependence.

4.8.4 Quasi-non-Boussinesq approximation

An intermediate approximation for the equation of state, often used in early ocean climate models (e.g., Bryan and Cox (1972)), replaces the dependence on *in situ*

pressure with the hydrostatic pressure felt by a fluid of density ρ_o

$$\rho(\theta, s, p) \approx \rho(\theta, s, p = g\, z\, \rho_o), \tag{4.137}$$

where $z < 0$ is the depth of the fluid parcel and ρ_o is the constant Boussinesq reference density. Notably, this expression for density removes local time dependence from the pressure field within the non-Boussinesq continuity equation. Hence, a non-Boussinesq fluid using the equation of state (4.137) is transformed into a *quasi-non-Boussinesq fluid*, where the term "quasi" signals that the fluid admits no acoustic modes. Durran (1999) refers to this approximation as *pseudo incompressibility*. Notably, this method for filtering acoustic modes does not introduce a new elliptic problem, in contrast to certain other wave filtering methods such as the rigid lid approximation of Bryan (1969).

Use of the approximate density (4.137) in the continuity equation must be accompanied by its use in the tracer and velocity budgets to maintain self-consistency. However, one recovers significant accuracy in quasi-non-Boussinesq fluids by employing the proper pressure dependence within the equation of state for purposes of computing horizontal pressure gradients. The importance of such accuracy was emphasized by Dewar et al. (1998).

Greatbatch et al. (2001) showed that traditional Boussinesq ocean models can be easily generalized to non-Boussinesq models, thus removing many of the inaccuracies of the Boussinesq approximation without introducing acoustic modes. Note that for hydrostatic fluids, the only acoustic mode admitted is the Lamb wave (Section 2.4.1). Numerical dissipation is likely to be sufficient to damp the Lamb wave, so that use of the accurate density variable throughout the model should not incur onerous time stepping constraints. We have more to say on these matters in Chapter 10 when discussing methods to numerical solve the momentum equations.

CHAPTER
FIVE

Generalized vertical coordinates

Contents

5.1 Introduction **114**
5.2 Concerning the choice of vertical coordinate **114**
5.2.1 Z-coordinate ocean models 115
5.2.2 Sigma-coordinate ocean models 116
5.2.3 Isopycnal-coordinate ocean models 118
5.2.4 Summary of the vertical coordinates 118
5.3 Generalized surfaces **119**
5.3.1 Constraining the vertical coordinate 119
5.3.2 Specific thickness 119
5.3.3 Slope of a generalized layer 120
5.4 Local orthonormal coordinates **120**
5.5 Generalized vertical coordinates **122**
5.5.1 Projected aspect of vertical coordinates 122
5.5.2 Transformation matrix 123
5.5.3 Transformation of partial derivatives 124
5.5.4 Further useful identities 125
5.5.5 Material time derivative and vertical velocity 126
5.6 Metrics with generalized vertical coordinates **127**
5.6.1 Metric on the surface 127
5.6.2 Metric for Euclidean space 128
5.7 The dia-surface velocity component **128**
5.8 Conservation of mass and volume **131**
5.8.1 Z-coordinates 132
5.8.2 Generalized vertical coordinates 132
5.8.3 Volume conservation in ρ-coordinates 133
5.8.4 Mass conservation in pressure coordinates 133
5.9 Primitive equations **134**
5.9.1 Momentum balance 134
5.9.2 Summary of the equations 135

5.9.3 Non-Boussinesq in pressure coordinates 135
5.9.4 Isentropic Boussinesq in ρ-coordinates 136
5.9.5 Neutral physics in generalized coordinates 136

There is a trend in the ocean model development towards the use of generalized vertical coordinate models. This chapter presents some salient mathematical issues.

5.1 Introduction

The choice of vertical coordinate is the most important aspect of an ocean model's design because it strongly prejudices the model's representation of various resolved and parameterized processes. Hence, there is a significant amount of fundamental physical, mathematical, and numerical research in ocean model development aimed at optimizing the choice for vertical coordinate. Currently, there is no clear best single choice of use for realistic global ocean climate simulations. Each has certain advantages and disadvantages, many of which are complementary. This result argues for the use of *hybrid* vertical coordinates, with choices based on the flow regime.

Generalized vertical coordinates as used in geophysical fluid models are often termed *non-orthogonal projected coordinates*, for reasons which will become clear in the following. As the issues of vertical coordinate are quite distinct from horizontal coordinates, simplicity of presentation warrants the use of Cartesian (x, y) for the horizontal coordinates. Transformations to generalized horizontal coordinates follow the techniques discussed in Chapters 18 and 19.

The standard references for this chapter include the pioneering paper by Starr (1945). Starr was the first to systematically present the mathematical aspects of generalized vertical coordinates. He also argued for their utility in GFD modeling. For some reason, this paper appears to be little recognized in the oceanography literature. The paper by Bleck (1978) illustrates the geometric aspects of generalized vertical coordinates and shows how to construct an energetically consistent numerical model using generalized vertical coordinates. One of the appendices in McDougall (1995) provides further words in support Bleck's geometric interpretation.

5.2 Concerning the choice of vertical coordinate

To get a sense for the issues related to vertical coordinates, consider a schematic of an ocean basin shown in Figure 5.1. This figure illustrates three fundamental regimes of ocean dynamics. First, there is the surface mixed layer, also known as the planetary boundary layer (PBL). This is a region of strong coupling to the overlying atmosphere, ice, river, etc. It is typically very well mixed in the vertical, and the physical processes active here are mostly three-dimensional boundary layer processes. A parameterization of these processes is necessary in a primitive equation ocean climate model, since explicit representation requires very high resolution (order meters to centimeters) and non-hydrostatic physics. In contrast to the PBL, the ocean interior is close to ideal (i.e., isentropic as defined in Section

4.6). Most of the processes in this region occur along surfaces of constant potential density. The ocean's bottom topography acts as a strong forcing on the overlying currents through interactions with the pressure field. Additionally, there are many regions of the ocean where turbulent bottom boundary layer (BBL) processes act to strongly mix and transport water masses.

Each of these regimes suggest a corresponding vertical coordinate. We now discuss these coordinates and their pros and cons.

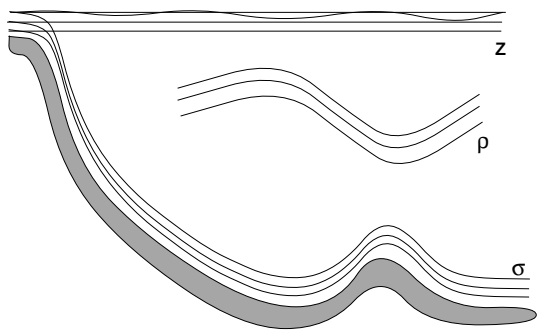

Figure 5.1: Schematic of an ocean basin, illustrating three fundamental regimes of ocean dynamics. The surface mixed layer is naturally represented using z-coordinates; the adiabatic interior is naturally represented using isopycnal coordinates; and the bottom topography is naturally represented using sigma coordinates. Outside of each coordinate's natural regime of applicability, there are many problems and complications. This figure is based on Figure 1 from Griffies et al. (2000a).

5.2.1 Z-coordinate ocean models

The simplest and oldest choice of vertical coordinate is z, which represents the vertical distance from a resting ocean surface at $z = 0$, with z positive upwards and $z = -H(x, y)$ the ocean bottom. The vertical direction $\hat{\mathbf{z}}$ is orthogonal to the two horizontal directions. Note that z-coordinate models are also often referred to as *geopotential* coordinate models, in which z is the vertical displacement with respect to a local approximation to surfaces of constant geopotential (Section 3.5.2.3). Z-models are presently the model class widely used for the study of ocean climate, hence the focus on this class in subsequent parts of this book (see Griffies et al. (2000a) for a tabulation of publicly supported ocean climate models).

$Z-$coordinate ocean models have been around for many decades, with the pioneering work of Bryan (1969) and Cox (1984) at GFDL providing the first ocean climate model. The main advantages of z-coordinate models are

- Simple (and often naive) numerical discretization methods have been used, to some success, in this framework. Other frameworks, such as isopycnal coordinate models, require more sophisticated numerics.

- The PBL is naturally parameterized using a z-coordinate.

- The equation of state for ocean water, which is highly nonlinear and very important for determining water mass properties, is cleanly represented.

- In general, thermodynamical effects are well represented in z-coordinate ocean models.

There are three main disadvantages of the z-coordinate models

- The representation of the bottom is cumbersome, as it is discretized with rectangular steps instead of piecewise linear fits to the topography. Rectangular steps impose a distinction between side and bottom. There are means to overcome aspects of this awkward representation (Adcroft et al. (1997); Pacanowski and Gnanadesikan (1998); see Figure 5.2). However, there remains the issue of parameterizing the bottom boundary layer flow, which is quite important in the ocean. Z-coordinates provide a cumbersome framework for representing such processes.

- The representation of dynamics and physics in the ideal ocean interior, away from the side and bottom boundaries, requires great care in z-models (see Part IV).

- Use of a free surface algorithm, which is desirable for physical reasons (see Chapter 10), imposes a limitation on deviations of the surface height to avoid a vanishing surface model grid cell (see Figure 9.1). As ocean climate models refine their vertical grid spacing, this limitation restricts ones ability to simulate large deviations in tidal fluctuations or large displacements by sea ice resting on the ocean surface.

5.2.2 Sigma-coordinate ocean models

The sigma coordinate is given by

$$\sigma = \frac{z - \eta}{H + \eta}, \tag{5.1}$$

where $\eta(x, y, t)$ is the displacement of the ocean surface from its resting position $z = 0$, and $z = -H(x, y)$ is the ocean bottom. Note that $\sigma = 0$ at the ocean surface and $\sigma = -1$ at the bottom. σ is monotonic, and so the relation (5.1) defines a unique mapping between depth z and σ, thus allowing for σ to be a valid vertical coordinate.

The main advantages of the σ coordinate are the following:

- It provides a smooth representation of the ocean bottom topography. This choice provides for a natural representation of bottom boundary layer physics.

- Thermodynamic effects associated with the equation of state are well represented.

There are three main problems with the sigma coordinate models:

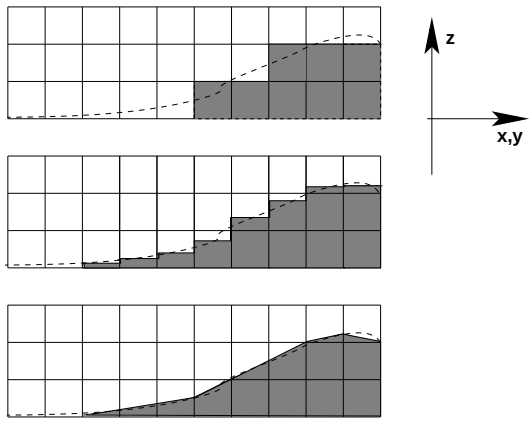

Figure 5.2: Comparing the representation of the ocean bottom in z-coordinate models. Top: The old-fashioned "full cell" approach in which vertical thicknesses of all cells are independent of latitude and longitude. Middle: The "partial cell" approach discussed by Adcroft et al. (1997) and Pacanowski and Gnanadesikan (1998). Here, the vertical thicknesses of the bottom cells can vary according to the topographic features. Bottom: The "shaved cell" approach of Adcroft et al. (1997), in which the bottom cell is a piece-wise linear fit to the topography. Both full and partial cells have discontinuous representations of the bottom, whereas the shaved cell has a continous depth, but discontinuous gradient of the depth. This figure is based on Figure 3 of Griffies et al. (2000a) (see also Figure 4 from Adcroft et al. (1997)).

- The PBL can be less well represented using σ as with the z-coordinate. The reason is that the vertical distance between grid points generally increases as one moves away from the side continental shelf regions, hence leaving the PBL with potentially small vertical resolution in the middle of an ocean basin.

- As with the z-coordinate models, the representation of interior adiabatic dynamics is unnatural in sigma coordinates.

- Sigma coordinate models have a difficult time accurately representing the horizontal pressure gradient. Because the surfaces of constant sigma are not generally flat, the horizontal pressure gradient, which is perpendicular to the local vertical direction $\hat{\mathbf{z}}$ as defined by gravity, has a projection along and across sigma surfaces. The result is a horizontal pressure gradient consisting of two terms

$$\begin{aligned} \nabla_z p &= (\nabla_\sigma - \nabla_\sigma z \, \partial_z) \, p \\ &= \nabla_\sigma p + \rho \, g \, \nabla_\sigma z. \end{aligned} \tag{5.2}$$

  To reach this result, we used the transformation (5.32) discussed in Section 5.5.3 relating horizontal gradients, and use has been made of the hydrostatic relation $p_{,z} = -\rho \, g$ to reach the second equality. In words, $\nabla_z p$ is the horizontal pressure gradient taken along surfaces of constant depth z, $\nabla_\sigma p$ is the pressure gradient along surfaces of constant σ, and $\nabla_\sigma z$ is the slope of the sigma surface relative to the constant depth surfaces. In the ocean, especially

next to the continental shelves, the slope of sigma surfaces can reach 1/100 to 1/10, which is a huge slope and so can cause the "sigma coordinate correction term" $\rho g \nabla_\sigma z$ to be on the order of $\nabla_\sigma p$. The numerical representation of the two terms is quite important. In general, errors in either one can contribute to spurious pressure gradients that drive unphysical currents.

5.2.3 Isopycnal-coordinate ocean models

Another choice for vertical coordinate is the potential density $\rho^{(p)}$, where p represents a pressure at which the potential density is referenced. In most cases, the p label is omitted for brevity, and we will do so in this book. This coordinate is a close analog to the atmosphere's entropy or potential temperature (see Section 4.8). In an idealized ocean, consisting of a simplified equation of state, the potential density defines a monotonic layering of the ocean fluid, and so it is a valid choice of vertical coordinate. In an isentropic ocean, the potential density is materially constant (see Section 4.8). Furthermore, physical processes, such as turbulent stirring and mixing from mesoscale eddies, have a strong tendency to occur along these surfaces, rather than across. Indeed, the ocean interior can, to a very good approximation, be considered an ideal fluid of stacked immiscible potential density layers. Hence, isopycnal ocean models are very useful for studying the ocean's isentropic dynamics.

With a realistic equation of state, however, there is no materially conserved density coordinate that is also monotonic with depth. For example, $p = 0$ was originally the common choice for a reference pressure used to define the potential density. Yet this pressure is inadequate for the deep ocean. Recently, $p = 2000$ decibars has been chosen for realistic isopycnal models since it leads to fewer regions with coordinate inversions. In general, the equation of state, as well as wide variations in ocean salinity such as may arise from river runoff and ice melt, add complications to isopycnal coordinate models.

Additionally, isopycnal models are an inappropriate framework for implementing parameterizations of the PBL, since the PBL is approximately vertically unstratified in density. Bulk mixed layer models are often appended to isopycnal models to represent the surface mixed layer. Additionally, some have even taken to using a hybrid model in which a pressure-coordinate PBL region is coupled to the interior isopycnal model.

5.2.4 Summary of the vertical coordinates

As mentioned at the beginning of this section, there appears to be no single vertical coordinate that is without its problems when aiming to simulate the World Ocean's circulation over time scales relevant for climate. Hence, there is presently a great deal of research focused on developing ocean models that allow for any continuously defined vertical coordinate. Given such freedom, researchers can examine the benefits of certain hybrid choices of vertical coordinates, such as those that are pressure-like near the ocean surface (effectively z-like), isopycnal in the interior, and sigma-like near the bottom.

For the remainder of this chapter, we leave behind these particular examples of vertical coordinates. Our focus will be on the mathematical elements of the ocean equations as formulated with generalized vertical coordinates.

5.3 Generalized surfaces

We consider the partitioning of the ocean into stacked layers of smooth surfaces. We define these surfaces mathematically by giving the value of a *generalized vertical coordinate*, denoted in the following by s. The value of this coordinate is generally a function of space-time, and so it maintains the functional dependence

$$s = s(x, y, z, t). \tag{5.3}$$

From Section 5.2, we have three examples of s: z-coordinates models have $s = z$, terrain following sigma-coordinate models have $s = \sigma$ defined according to equation (5.1), and isopycnal coordinate models have $s = \rho^{(p)}$, with $\rho^{(p)}$ the potential density referenced to a particular pressure. Note that s is a common notation used in the literature, and so it will be used in this chapter. Care should be taken, however, not to confuse this symbol for ocean salinity. The context should be sufficient to make the distinction.

5.3.1 Constraining the vertical coordinate

To make generalized surfaces useful for modeling GFD flows, we must constrain them to never have an undulation that is parallel to the local vertical direction. Mathematically, we insist that

$$s_{,z} \neq 0 \tag{5.4}$$

throughout the fluid. That is, the surfaces must retain a nontrivial vertical stratification. Correspondingly, $s_{,z}$ must remain single-signed and so s is monotonic with depth. We assume the generalized vertical coordinates discussed in this chapter satisfy this constraint.

The constraint (5.4) allows us to unambiguously orient the surfaces of constant generalized vertical coordinate within the ocean fluid. It is trivially satisfied for z-models. It is also satisfied for vertically stably stratified adiabatic fluids when the generalized coordinate surfaces are equated to surfaces of constant potential density. And it is satisfied for terrain following sigma surfaces, so long as we do not attempt to represent "over-hangs" or "caves" in the bottom topography.

5.3.2 Specific thickness

Mathematically, the reason we impose the above constraint on the surfaces is to ensure that the slopes of the surfaces never become infinite, in which case the transformation between depth z and generalized vertical coordinate s becomes singular. That is, the Jacobian of transformation $\partial z/\partial s = z_{,s}$ (see Section 5.5.2) becomes infinite.

The Jacobian of transformation $z_{,s}$ is often called the *specific thickness*, or *thickness* for short. We can motivate this name by the following considerations. Consider stacked set of s-layers. Invertibility of the transformation between z and s requires

$z_{,s}$ to be single-signed. When multiplied by an infinitesimal increment of the generalized vertical coordinate, $\delta s > 0$, the product

$$\delta h = z_{,s}\, \delta s \tag{5.5}$$

represents the infinitesimal *thickness* separating the layers with generalized coordinate s and $s + \delta s$. The finite difference version of the thickness is appropriate for numerical models using generalized vertical coordinates.

5.3.3 Slope of a generalized layer

In the following, we find the opportunity to characterize the generalized coordinate surfaces according to how they slope relative to the horizontal surfaces used in z-coordinate ocean models. For this purpose, we introduce here the slope vector

$$\begin{aligned} \mathbf{S} &= \nabla_s z \\ &= -z_{,s}\, \nabla_z s. \end{aligned} \tag{5.6}$$

In this equation, $\nabla_s z$ is the horizontal gradient of the depth as taken along surfaces of constant generalized vertical coordinate s. The second equality represents the horizontal gradient of the generalized surface multiplied by the Jacobian of transformation between the s and z coordinate systems. This result follows from a coordinate transformation described in Section 5.5.

The two components of the slope vector represent the projection of the slope of surfaces of constant generalized vertical coordinate s onto the two horizontal coordinate directions. The squared magnitude of this vector will be written $S^2 = \mathbf{S} \cdot \mathbf{S}$. The components to the slope vector vanish when the generalized surface coincides with the horizontal (x, y) plane. The slope is infinite where the surface is unstratified in the vertical. That is, where $s_{,z} = 0$.

5.4 Local orthonormal coordinates

Before delving into the mathematics of the generalized vertical coordinates used in GFD modeling, we find it useful to describe another set of coordinates. We term these coordinates *local orthonormal coordinates*. They are defined at each point on an arbitrary smooth surface by the locally orthonormal set of basis directions

$$\mathbf{e}_{\bar{1}} = \frac{\hat{\mathbf{y}} \wedge \nabla s}{|\hat{\mathbf{y}} \wedge \nabla s|} \tag{5.7}$$

$$\mathbf{e}_{\bar{2}} = \mathbf{e}_{\bar{3}} \wedge \mathbf{e}_{\bar{1}} \tag{5.8}$$

$$\mathbf{e}_{\bar{3}} = \frac{\nabla s}{|\nabla s|}. \tag{5.9}$$

Note that the basis directions are each well defined regardless the orientation of the surface. However, for those special cases where the local normal $\mathbf{e}_{\bar{3}}$ has a single-signed projection onto the vertical direction $\hat{\mathbf{z}}$, which is the case with generalized

vertical coordinates discussed in this chapter (see Section 5.3.1), then we can introduce the slope vector $\mathbf{S} = (S_x, S_y)$ defined in Section 5.3.3 to yield

$$\mathbf{e}_{\bar{1}} = \frac{\hat{\mathbf{x}} + \hat{\mathbf{z}}\, S_x}{\sqrt{1 + S_x^2}} \tag{5.10}$$

$$\mathbf{e}_{\bar{2}} = \frac{-\hat{\mathbf{x}}\, S_x\, S_y + \hat{\mathbf{y}}\,(1 + S_x^2) + \hat{\mathbf{z}}\, S_y}{\sqrt{(1 + S^2)(1 + S_x^2)}} \tag{5.11}$$

$$\mathbf{e}_{\bar{3}} = \frac{(-\mathbf{S}, 1)}{\sqrt{1 + S^2}}. \tag{5.12}$$

Figure 5.3 shows a two-dimensional schematic of these basis vectors. "Lateral" distances are measured in the direction of the vectors $\mathbf{e}_{\bar{1}}$ and $\mathbf{e}_{\bar{2}}$, whereas "vertical" distances are measured parallel to the normal vector $\mathbf{e}_{\bar{3}}$. When the surfaces are horizontally aligned, the generalized basis reduces to the Cartesian basis $(\mathbf{e}_{\bar{1}}, \mathbf{e}_{\bar{2}}, \mathbf{e}_{\bar{3}}) = (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$. More generally, all three basis vectors change directions according to the dynamics and geometry of the surfaces.

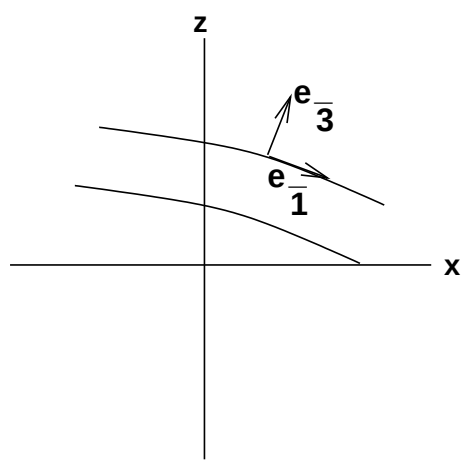

Figure 5.3: Schematic of local orthonormal coordinates. Two generalized surfaces are shown, with the unit vectors indicated that are appropriate for a point on one layer. The unit vector $\mathbf{e}_{\bar{3}}$ is normal to the surface, and $\mathbf{e}_{\bar{1}}$ lies in the plane. The third direction $\mathbf{e}_{\bar{2}}$ is perpendicular to the page. Lateral distances are measured in the plane of the surface, and vertical distances are measured normal to the surfaces.

The transformation matrix between the locally orthonormal coordinates and the fixed Cartesian basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ is generally a space-time dependent rotation, as can be deduced since both bases are orthonormal. Consequently, the transformation

$$\mathbf{e}_{\bar{a}} = \Lambda^{a}{}_{\bar{a}}\, \mathbf{e}_a \tag{5.13}$$

is comprised of directional cosines, thus rendering

$$
\begin{aligned}
(\mathbf{e}_{\overline{1}}, \mathbf{e}_{\overline{2}}, \mathbf{e}_{\overline{3}}) &= (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}) \begin{pmatrix} \hat{\mathbf{x}} \cdot \mathbf{e}_{\overline{1}} & \hat{\mathbf{x}} \cdot \mathbf{e}_{\overline{2}} & \hat{\mathbf{x}} \cdot \mathbf{e}_{\overline{3}} \\ \hat{\mathbf{y}} \cdot \mathbf{e}_{\overline{1}} & \hat{\mathbf{y}} \cdot \mathbf{e}_{\overline{2}} & \hat{\mathbf{y}} \cdot \mathbf{e}_{\overline{3}} \\ \hat{\mathbf{z}} \cdot \mathbf{e}_{\overline{1}} & \hat{\mathbf{z}} \cdot \mathbf{e}_{\overline{2}} & \hat{\mathbf{z}} \cdot \mathbf{e}_{\overline{3}} \end{pmatrix} \\
&= (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}) \begin{pmatrix} \frac{1}{\sqrt{1+S_x^2}} & \frac{-S_x\, S_y}{\sqrt{(1+S^2)(1+S_x^2)}} & \frac{-S_x}{\sqrt{1+S^2}} \\ 0 & \frac{\sqrt{1+S_x^2}}{\sqrt{1+S^2}} & \frac{-S_y}{\sqrt{1+S^2}} \\ \frac{S_x}{\sqrt{1+S_x^2}} & \frac{S_y}{\sqrt{(1+S^2)(1+S_x^2)}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} .
\end{aligned}
\tag{5.14}
$$

As a check, note that the determinant of the transformation is unity and its inverse is given by its transpose, thus making it a rotation matrix.

The locally orthonormal basis is used in formulating neutral physics operators of use in z-coordinate ocean models. We detail this formualtion in Section 12.1. However, this basis is not so useful as a means to formulate the full equations of motion of an ocean model. The problem is that each basis vector changes in time, and this proves inconvenient in practice. Additionally, the hydrostatic and geostrophic balances are very important for the large-scale ocean. For numerical accuracy, it is *crucial* to have these balances described by separate diagnostic equations, and to not have any overlap or coupling within the equations describing these balances. The locally orthonormal unit vectors do not satisfy this constraint in the general case. Hence, this choice does not represent a practical set of coordinates for GFD modeling.

5.5 Generalized vertical coordinates

When using generalized vertical coordinates in GFD modeling, the horizontal distance between two points is typically measured in the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ directions, hence crossing through the vertical planes $x = const$ and $y = const$. That is, the horizontal coordinates for a parcel are the *same* coordinates (x, y) also used for geopotential or z-coordinate ocean models. The vertical distance is measured parallel to the $x = const$ and $y = const$ planes; that is, it is measured in the $\hat{\mathbf{z}}$ direction, again just as for z-coordinate models. However, the vertical *position* is not specified by giving the value of the depth z. Instead, it is specified by giving the value for the monotonic vertical coordinate $s = s(x, y, z, t)$.

It is notable that these coordinates have found use in other areas of theoretical physics. In particular, condensed matter physicists and biophysicists studying the dynamics of fluctuating membranes use these coordinates, where the coordinates go by the name *Monge gauge*. Their mathematical aspects are lucidly described in Section 10.4 of Chaikin and Lubensky (1995).

5.5.1 Projected aspect of vertical coordinates

As it is the generalized vertical coordinate s, not the vertical distance z, that is specified, lateral property gradients are taken along surfaces of constant s instead of surfaces of constant z. It is this property that prompts the adjective "projected" when referring to these coordinates (see Bleck (1978) and McDougall (1995) for further

discussion). Mathematically, these ideas imply that the transformation between geopotential and generalized vertical coordinates takes the form

$$t = t \tag{5.15}$$
$$x = x \tag{5.16}$$
$$y = y \tag{5.17}$$
$$s = s(x, y, z, t). \tag{5.18}$$

Figure 5.4 further illustrates and describes these important, and subtle, points.

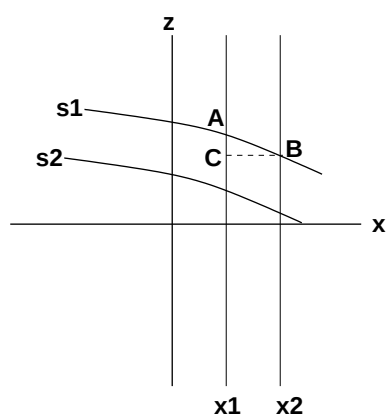

Figure 5.4: Schematic of projected coordinates as used in generalized vertical coordinate ocean models. The lateral distance between two points is measured in the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ directions, hence crossing through the vertical planes $x = const$ and $y = const$ just as for z-coordinate ocean models. For example, the lateral distance between points A and B is not measured along the constant s surface. Instead, it is the same as the horizontal distance between C and B. The vertical distance is measured parallel to the $x = const$ and $y = const$ planes; that is, it is measured in the $\hat{\mathbf{z}}$ direction *just* as for geopotential coordinates. However, the vertical *position* is specified not by giving the depth z. Rather, it is specified by giving the value of the generalized vertical coordinate $s = s(x, y, z, t)$. Lateral property gradients are thus taken along surfaces of constant s instead of surfaces of constant z. It is this property that prompts the adjective "projected" when referring to these coordinates. This figure is taken after Figure 1 of Bleck (1978).

5.5.2 Transformation matrix

To develop equations describing the ocean fluid using generalized vertical coordinates, and to help understand more of their mathematical properties, it is useful to consider properties of transformations between geopotential coordinates

$$(\xi^0, \xi^1, \xi^2, \xi^3) = (t, x, y, z) \tag{5.19}$$

and generalized vertical coordinates

$$(\xi^{\bar{0}}, \xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) = (t, x, y, s). \tag{5.20}$$

The abstract notation helps to organize the transformation properties according to the tensor calculus discussed in Chapters 18 and 19.

Again, it is only the manner in which we measure the vertical position that differs between these two sets of coordinates. That is, the coordinate transformation is just between the vertical coordinates. The horizontal coordinates and time are measured the same in both coordinate systems. Hence, the vertical coordinate in one coordinate system is generally dependent on all the coordinates in the other system

$$z = z(x, y, s, t) \tag{5.21}$$

$$s = s(x, y, z, t). \tag{5.22}$$

This coordinate dependence leads to the transformation matrix between the "unbarred" geopotential coordinates and the "barred" generalized vertical coordinates

$$\Lambda^a{}_{\bar{a}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ z_{,t} & z_{,x} & z_{,y} & z_{,s} \end{pmatrix}. \tag{5.23}$$

The determinant of this transformation matrix, also knowns as the Jacobian, is given by

$$\begin{aligned} \det(\Lambda^a{}_{\bar{a}}) &= z_{,s} \\ &= (s_{,z})^{-1}. \end{aligned} \tag{5.24}$$

We presented a geometric interpretation of the Jacobian in Section 5.3.2, where we motivated calling it the specific thickness. The inverse transformation is given by

$$\Lambda^{\bar{a}}{}_a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ s_{,t} & s_{,x} & s_{,y} & s_{,z} \end{pmatrix}. \tag{5.25}$$

5.5.3 Transformation of partial derivatives

Now that we have the transformation matrix, we can use the tensor rules established in Chapters 18 and 19 to see how various tensor objects transform. First, we consider how partial derivatives transform. As discussed in Section 18.10, the transformation of the partial derivative operator takes the form

$$\partial_{\bar{a}} = \Lambda^a{}_{\bar{a}}\, \partial_a. \tag{5.26}$$

In matrix-vector notation, with the transformation matrix (5.23), this transformation can be written

$$(\partial_{\bar{0}}, \partial_{\bar{1}}, \partial_{\bar{2}}, \partial_{\bar{3}}) = (\partial_0, \partial_1, \partial_2, \partial_3) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ z_{,t} & z_{,x} & z_{,y} & z_{,s} \end{pmatrix}. \tag{5.27}$$

Note that when translating a tensor equation to a matrix-vector equation, one should be careful not to interchange the matrix with its transpose. Multiplication yields

$$\left(\frac{\partial}{\partial t}\right)_s = \left(\frac{\partial}{\partial t}\right)_z + z_{,t}\,\partial_z \tag{5.28}$$

$$\left(\frac{\partial}{\partial x}\right)_s = \left(\frac{\partial}{\partial x}\right)_z + z_{,x}\,\partial_z \tag{5.29}$$

$$\left(\frac{\partial}{\partial y}\right)_s = \left(\frac{\partial}{\partial y}\right)_z + z_{,y}\,\partial_z \tag{5.30}$$

$$\partial_s = z_{,s}\,\partial_z. \tag{5.31}$$

In words, terms on the left hand side are partial derivative operators taken in the generalized vertical coordinate frame, and those on the right hand side are in the geopotential frame. The spatial portion of these transformations can be written in the more tidy form

$$\nabla_s + \hat{\mathbf{z}}\,\partial_s = (\nabla_z + \mathbf{S}\,\partial_z) + \hat{\mathbf{z}}\,z_{,s}\,\partial_z. \tag{5.32}$$

In this relation, ∇_z is the horizontal gradient taken on surfaces of constant depth z, ∇_s is the horizontal gradient taken along surfaces of constant generalized vertical coordinate s, and $\mathbf{S} = \nabla_s z$ is the slope vector introduced in Section 5.3.3. Recall that the horizontal portion of this relation was used in Section 5.2.2 where we discussed the pressure gradient calculation used in sigma-coordinate models.

5.5.4 Further useful identities

Further relations of use in subsequent manipulations are the triple products

$$\left(\frac{\partial s}{\partial z}\right)_t \left(\frac{\partial z}{\partial t}\right)_s \left(\frac{\partial t}{\partial s}\right)_z = -1 \tag{5.33}$$

$$\left(\frac{\partial s}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_s \left(\frac{\partial x}{\partial s}\right)_z = -1 \tag{5.34}$$

$$\left(\frac{\partial s}{\partial z}\right)_y \left(\frac{\partial z}{\partial y}\right)_s \left(\frac{\partial y}{\partial s}\right)_z = -1. \tag{5.35}$$

These relations are familiar from manipulations in thermodynamics such as those used in deriving the Maxwell relations (e.g., Callen (1985)). Using the more compact notation with the implicit understanding of what frame the derivatives are acting in, the triple products imply

$$z_{,t} = -s_{,t}\,z_{,s} \tag{5.36}$$

$$z_{,x} = -s_{,x}\,z_{,s} \tag{5.37}$$

$$z_{,y} = -s_{,y}\,z_{,s} \tag{5.38}$$

where use was made of the identities

$$s_{,t} = (t_{,s})^{-1} \tag{5.39}$$

$$s_{,x} = (x_{,s})^{-1} \tag{5.40}$$

$$s_{,y} = (y_{,s})^{-1} \tag{5.41}$$

$$s_{,z} = (z_{,s})^{-1}. \tag{5.42}$$

5.5.5 Material time derivative and vertical velocity

When operating on a scalar field, the material time derivative $\mathrm{d}/\mathrm{d}t$ is itself a scalar under the usual non-relativistic transformations relevant for GFD. Thus it takes the same form regardless the coordinate system

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} &= \partial_0 + v^a \partial_a \\ &= \partial_{\bar{0}} + v^{\bar{a}} \partial_{\bar{a}}.\end{aligned} \tag{5.43}$$

Here, the velocity components v^a are those for the geopotential coordinate frame

$$(v^1, v^2, v^3) = (\dot{x}, \dot{y}, \dot{z}) = (u, v, w) \tag{5.44}$$

with $\dot{\psi} = \mathrm{d}\psi/\mathrm{d}t$ a shorthand for the material time derivative. The velocity components $v^{\bar{a}}$ are those for the generalized vertical frame

$$(v^{\bar{1}}, v^{\bar{2}}, v^{\bar{3}}) = (\dot{x}, \dot{y}, \dot{s}) = (u, v, \dot{s}). \tag{5.45}$$

In a non-tensorial notation, the material time derivative is written

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} &= \left(\frac{\partial}{\partial t}\right)_z + \mathbf{u} \cdot \nabla_z + w \left(\frac{\partial}{\partial z}\right) \\ &= \left(\frac{\partial}{\partial t}\right)_s + \mathbf{u} \cdot \nabla_s + \dot{s} \left(\frac{\partial}{\partial s}\right).\end{aligned} \tag{5.46}$$

As discussed in Section 5.5.1, both geopotential and generalized vertical models measure the same horizontal distances and the same time. Therefore, the horizontal velocity components are the same. This point is worth emphasizing, as in Bleck (1978). Indeed, often in the literature one can read phrases such as "the horizontal velocity along the isopycnal surface." Phrases such as this lend one to picture a velocity field that is oriented parallel to the isopycnal surface. However, as seen in Section 5.7, this phrase is sensible only when the flow is static and adiabatic.

Although the horizontal velocity is the same in both coordinate systems, the vertical velocity is distinct. In z-coordinates, the vertical velocity component is

$$\begin{aligned}w &= \frac{\mathrm{d}z}{\mathrm{d}t} \\ &= \left[\left(\frac{\partial}{\partial t}\right)_s + u\left(\frac{\partial}{\partial x}\right)_s + v\left(\frac{\partial}{\partial y}\right)_s + \dot{s}\left(\frac{\partial}{\partial s}\right)\right] z \\ &= z_{,t} + \mathbf{u} \cdot \mathbf{S} + \dot{s}\, z_{,s}\end{aligned} \tag{5.47}$$

where $\mathbf{S} = \nabla_s z = -z_{,s} \nabla_z s$ is the slope vector introduced in Section 5.3.3. In the generalized vertical coordinate system, the vertical velocity component is

$$\begin{aligned}\dot{s} &= \frac{\mathrm{d}s}{\mathrm{d}t} \\ &= \left[\left(\frac{\partial}{\partial t}\right)_z + u\left(\frac{\partial}{\partial x}\right)_z + v\left(\frac{\partial}{\partial y}\right)_z + w\left(\frac{\partial}{\partial z}\right)\right] s \\ &= s_{,t} + \mathbf{u} \cdot \nabla_z s + w\, s_{,z}.\end{aligned} \tag{5.48}$$

Use of the triple products (5.33)–(5.35) allows us to directly relate the vertical velocity components

$$s_{,z}\,\dot{z} = (\mathrm{d}/\mathrm{d}t - \mathbf{u}\cdot\nabla_z)\,s - s_{,t} \tag{5.49}$$

$$z_{,s}\dot{s} = (\mathrm{d}/\mathrm{d}t - \mathbf{u}\cdot\nabla_s)\,z - z_{,t} \tag{5.50}$$

where again $\dot{z} = w$.

5.6 Metrics with generalized vertical coordinates

We consider in this section two metric tensors relevant when working with generalized coordinate surfaces.

5.6.1 Metric on the surface

A particular generalized surface can be defined by the algebraic equation

$$s(x, y, z, t) - s_{const} = 0, \tag{5.51}$$

where s_{const} is a constant. If we are performing measurements on this surface, it is relevant to ask what is the metric tensor to be used to measure distances.

To answer this question, note that on the surface,

$$\mathrm{d}s = s_{,a}\,\mathrm{d}x^a = 0 \tag{5.52}$$

since $s = s_{const}$. As a result, the vertical coordinate differential on the surface is related to the horizontal differentials through

$$\mathrm{d}z = \mathbf{S}\cdot\mathrm{d}\mathbf{x}, \tag{5.53}$$

where $\mathbf{S}$ is the slope vector with magnitude S, as defined in Section 5.3.3.

To measure the distance between two points on the surface, it is sufficient to specify their horizontal positions since we know they live somewhere on the two-dimensional surface. For GFD purposes we are concerned with generalized coordinate surfaces embedded in three-dimensional Euclidean space. Hence, we can write the squared distance between two infinitesimally close points in the form

$$\mathrm{d}\mathbf{x}\cdot\mathrm{d}\mathbf{x} = (1 + S_x^2)(\mathrm{d}x)^2 + (1 + S_y^2)(\mathrm{d}y)^2 + 2S_x\,S_y\,\mathrm{d}x\,\mathrm{d}y, \tag{5.54}$$

where we used equation (5.53). We thus identify the metric tensor for the two-dimensional surface

$$g_{\bar{a}\bar{b}} = \begin{pmatrix} 1 + S_x^2 & S_x\,S_y \\ S_x\,S_y & 1 + S_y^2 \end{pmatrix}. \tag{5.55}$$

Likewise, the components to the inverse metric are given by

$$g^{\bar{a}\bar{b}} = (1 + S^2)^{-1/2} \begin{pmatrix} 1 + S_y^2 & -S_x\,S_y \\ -S_x\,S_y & 1 + S_x^2 \end{pmatrix}. \tag{5.56}$$

Using a result from the tensor analysis discussion in Section 18.13.2, we note that the area of an infinitesimal region on the surface of constant generalized vertical coordinate is given by

$$\begin{aligned} \mathrm{d}A_{(s)} &= [\det(g_{\bar{a}\bar{b}})]^{1/2}\,\mathrm{d}x\,\mathrm{d}y \\ &= (1+S^2)^{1/2}\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{5.57}$$

where $\det(g_{\bar{a}\bar{b}}) = 1 + S^2$ is the determinant of the metric. This result is used in Section 5.7 when discussing the dia-surface velocity component.

In many geophysical applications, the slope vector has components smaller than 10^{-2}. In this case, the generally non-diagonal metric tensor is well approximated by the unit (or Kronecker) tensor δ_{ab}, which is the metric for a flat manifold (Section 18.4).

5.6.2 Metric for Euclidean space

Now consider the metric for three dimensional Euclidean space, in which the vertical position of a point is measured by the generalized vertical coordinate s. A straightforward way to compute this metric is to transform the Cartesian metric δ_{ab} to generalized coordinates using

$$g_{\bar{a}\,\bar{b}} = \delta_{ab}\,\Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}} \tag{5.58}$$

with just the spatial part of the transformation matrix equation (5.23). The result is the metric tensor

$$g_{\bar{a}\,\bar{b}} = \begin{pmatrix} 1+S_x^2 & S_x\,S_y & S_x\,z_{,s} \\ S_x\,S_y & 1+S_y^2 & S_y\,z_{,s} \\ S_x\,z_{,s} & S_y\,z_{,s} & z_{,s}^2 \end{pmatrix}. \tag{5.59}$$

The two-dimensional horizontal subspace of this metric tensor agrees with that given by equation (5.55), which was derived in a more direct manner from ideas of surface differential geometry. The metric (5.59) is that which would be used by someone measuring distances in $\mathcal{R}^3$ with (x, y, s) as coordinates. As with the two-dimensional metric (5.55), the off-diagonal terms in (5.59) indicate the non-orthogonality of the (x, y, s) coordinates.

5.7 The dia-surface velocity component

We are often concerned in geophysical fluid dynamics with the amount of fluid passing through a surface. We define the flux of fluid parcels crossing a surface as the *dia-surface* flux, with units volume per time per area (i.e., units of velocity). In determining this flux, it is important to note that surfaces generally move in time through the fluid and have arbitrarily changing shapes. The purpose of this section is to describe the kinematics appropriate for determining the dia-surface flux for a slightly restricted class of surfaces, those being defined by constant values of the generalized vertical coordinate s. Some of the material in this section parallels Section 4 of Haynes and McIntyre (1990). See also Section 3.2 of Pedlosky (1996) for a complementary discussion.

At an arbitrary point on the surface of constant generalized vertical coordinate, the flux of fluid parcels in the direction normal to the surface is computed as

$$\text{seawater flux in direction } \hat{\mathbf{n}} = \mathbf{v} \cdot \hat{\mathbf{n}} \tag{5.60}$$

with*

$$\hat{\mathbf{n}} = \nabla s \, |\nabla s|^{-1} \tag{5.61}$$

the surface unit normal. Introducing the material time derivative $\dot{s} = s_{,t} + \mathbf{v} \cdot \nabla s$ leads to the equivalent expression

$$\mathbf{v} \cdot \hat{\mathbf{n}} = (\dot{s} - s_{,t}) |\nabla s|^{-1}. \tag{5.62}$$

Since the surface is moving, the net flux of seawater passing through the surface is obtained by subtracting the Eulerian velocity of the surface $\mathbf{v}^{(ref)}$ in the $\hat{\mathbf{n}}$ direction from the Eulerian velocity component $\mathbf{v} \cdot \hat{\mathbf{n}}$ of the fluid parcels

$$\text{net flux of seawater through surface} = \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(ref)}). \tag{5.63}$$

The velocity $\mathbf{v}^{(ref)}$ is the Eulerian velocity of a reference point fixed on the surface, and it is written

$$\mathbf{v}^{(ref)} = \mathbf{u}^{(ref)} + w^{(ref)} \, \hat{\mathbf{z}}. \tag{5.64}$$

Since the reference point remains on the same $s = const$ surface, $\dot{s} = 0$ for the reference point. Consequently, the vertical velocity component $w^{(ref)}$ is given by

$$w^{(ref)} = -z_{,s} \, (s_{,t} + \mathbf{u}^{(ref)} \cdot \nabla_z s), \tag{5.65}$$

where equation (5.48) was used with $\dot{s} = 0$. Hence, we can write

$$\begin{aligned} \hat{\mathbf{n}} \cdot \mathbf{v}^{(ref)} &= \hat{\mathbf{n}} \cdot \mathbf{u}^{(ref)} + \hat{\mathbf{n}} \cdot \hat{\mathbf{z}} \, w^{(ref)} \\ &= |\nabla s|^{-1} (\nabla_z s \cdot \mathbf{u}^{(ref)} - s_{,t} - \nabla_z s \cdot \mathbf{u}^{(ref)}) \\ &= -s_{,t} \, |\nabla s|^{-1} \end{aligned} \tag{5.66}$$

which then leads to the net flux of seawater crossing the surface

$$\begin{aligned} \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(ref)}) &= (\mathbf{v} \cdot \nabla s + s_{,t}) \, |\nabla s|^{-1} \\ &= \dot{s} \, |\nabla s|^{-1} \\ &= z_{,s} \, \dot{s} \, (1 + S^2)^{-1/2} \end{aligned} \tag{5.67}$$

where we used $z_{,s} \, |\nabla s| = \sqrt{1 + S^2}$ in the last step.

The area normalizing the volume flux in the above discussion is the area $\mathrm{d}A_{(s)}$ of an infinitesimal region living on the surface of constant generalized vertical coordinate, and this area is given by equation (5.57). Hence, the volume per time of fluid passing through the generalized surface is given by

$$\begin{aligned} \text{vol/time fluid through surface} &= \hat{\mathbf{n}} \cdot (\mathbf{v} - \mathbf{v}^{(ref)}) \, \mathrm{d}A_{(s)} \\ &= z_{,s} \, \dot{s} \, \mathrm{d}x \, \mathrm{d}y \\ &\equiv w^{(s)} \, \mathrm{d}x \, \mathrm{d}y. \end{aligned} \tag{5.68}$$

*Note that in equation (5.12), we used the symbol $\mathbf{e}_{\bar{3}}$ instead of $\hat{\mathbf{n}}$. For the present discussion, we prefer the more conventional notation $\hat{\mathbf{n}}$.

As defined,

$$w^{(s)} = z_{,s}\,\dot{s} \tag{5.69}$$

measures the volume of fluid passing through the surface, per unit area $\mathrm{d}x\,\mathrm{d}y$ of the horizontal projection of the surface, per unit time. Since $w^{(s)}$ is not a vector, referring to it as the dia-surface *velocity* is a misnomer, although such is often the case in the literature. Instead, we prefer the name dia-surface velocity *component*.

The expression (5.69) for $w^{(s)}$ allows us to write the material time derivative (5.46) using the generalized vertical coordinate s as

$$\frac{\mathrm{d}}{\mathrm{d}t} = \left(\frac{\partial}{\partial t}\right)_s + \mathbf{u}\cdot\nabla_s + w^{(s)}\left(\frac{\partial}{\partial z}\right) \tag{5.70}$$

where we used $\partial_s = z_{,s}\,\partial_z$. This form for the material time derivative motivates some to consider $w^{(s)}$ as a vertical velocity component that measures the rate at which fluid parcels penetrate the surface of constant generalized coordinate (see Appendix A to McDougall (1995)). One should be mindful, however, to distinguish $w^{(s)}$ from the generally different vertical velocity component $w = \mathrm{d}z/\mathrm{d}t$ used in z-models.

To make contact with the literature, it is useful to note the many equivalent forms of $w^{(s)}$

$$\begin{aligned} w^{(s)} &= z_{,s}\,\dot{s} \\ &= z_{,s}\left(s_{,t} + \mathbf{v}\cdot\nabla s\right) \\ &= -z_{,t} + (\mathbf{v}\cdot\hat{\mathbf{n}})\,z_{,s}\,|\nabla s| \\ &= -z_{,t} + (\mathbf{v}\cdot\hat{\mathbf{n}})\sqrt{1+S^2} \\ &= -z_{,t} + \mathbf{v}\cdot(-\mathbf{S},1) \\ &= w - (\partial_t + \mathbf{u}\cdot\nabla_s)\,z \end{aligned} \tag{5.71}$$

where the identity (5.36) was used for the third equality, and $\mathbf{S} = \nabla_s z$ is the slope from Section 5.3.3. The last two forms expose some special cases rendering useful experience with certain forms for $w^{(s)}$. For example, with horizontal surfaces, $\mathbf{S} = 0$ and so

$$w^{(s)} = -z_{,t} + w \qquad \text{when } \mathbf{S} = 0. \tag{5.72}$$

For the case when the horizontal surface is static ($z_{,t} = 0$), as in geopotential coordinate models, fluid parcels penetrate the surface at a rate set by the vertical velocity component w. For the case when horizontal surfaces change their depth at a rate $z_{,t}$ equal to that of the fluid parcels w, the surfaces retreat from fluid parcels at a rate ensuring that no parcels penetrate the surfaces. Now consider a static non-horizontal surface of constant generalized vertical coordinate, in which case

$$w^{(s)} = w - \mathbf{u}\cdot\mathbf{S} \qquad \text{when } z_{,t} = 0. \tag{5.73}$$

Hence, even if fluid parcels are moving upwards ($w > 0$), the dia-surface velocity component can be negative so long as the horizontal velocity projected onto the slope dominates ($\mathbf{u}\cdot\mathbf{S} < 0$ with $|\mathbf{u}\cdot\mathbf{S}| > |w|$). Indeed, if the slope becomes very steep, thus reducing the generalized surface's projection onto the horizontal plane, and the horizontal velocity remains nontrivial, then $w^{(s)}$ can become quite large.

When the generalized surface is a surface of constant potential density (i.e., an isopycnal), then $w^{(\rho)}$ is termed the *diapycnal* velocity component. For the more general case with neutral surfaces, $w^{(\rho)}$ is the *dianeutral* velocity component. In these cases,

$$w^{(\rho)} = z_{,\rho}\, \dot{\rho}. \tag{5.74}$$

Hence, for flow materially conserving potential temperature and salinity so that potential density is materially conserved, $\dot{\rho} = \mathrm{d}\rho/\mathrm{d}t = 0$, the diapycnal velocity component vanishes

$$w^{(\rho)} = 0 \qquad \text{when } \mathrm{d}\rho/\mathrm{d}t = 0. \tag{5.75}$$

That is, no fluid parcels penetrate isopycnal surfaces for the case of ideal fluid flow. Additionally, the material time derivative simplifies for ideal flow using isopycnal coordinates, since its spatial component reduces from three to two components

$$\frac{\mathrm{d}}{\mathrm{d}t} = \left(\frac{\partial}{\partial t}\right)_{\rho} + \mathbf{u}\cdot\nabla_{\rho} \qquad \text{when } \mathrm{d}\rho/\mathrm{d}t = 0. \tag{5.76}$$

These are very important results that motivate the use of an isopycnal coordinate description of the ocean interior, where flow is nearly ideal (Bleck (1998) highlights this point in his review of isopycnal modeling). We see more advantages of isopycnal coordinates when discussing the kinematics of mesoscale eddy motion in Chapter 7.

In closing this discussion, we note that only when the potential density surfaces are static does ideal flow lead to a three-dimensional velocity oriented parallel to potential density surfaces

$$\mathbf{v}\cdot\nabla\rho = 0 \qquad \text{when } \mathrm{d}\rho/\mathrm{d}t = 0 \text{ and } \rho_{,t} = 0. \tag{5.77}$$

This flow is often considered in idealized studies.

5.8 Conservation of mass and volume

We have established a suite of mathematical results for the generalized vertical coordinate system. Now, it is time to consider how the description of the ocean fluid is formulated using these coordinates. To start, we consider expressions for the conservation of mass and volume.

Recall from our discussions in Chapter 3 that the mass of an infinitesimal parcel of fluid is written

$$\mathrm{d}M = \rho\,\mathrm{d}V, \tag{5.78}$$

where $\mathrm{d}V$ is the infinitesimal volume of the parcel and ρ is the *in situ* density. Following the discussion in Section 2.3.4, we note that mass conservation for each parcel implies

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\ln(\mathrm{d}M) &= \frac{\mathrm{d}}{\mathrm{d}t}\ln(\rho\,\mathrm{d}V) \\ &= 0. \end{aligned} \tag{5.79}$$

The Eulerian form of mass conservation, known as the *continuity equation*, depends on the coordinates used to describe the parcel's volume.

5.8.1 Z-coordinates

Using the geopotential z-coordinate system (x, y, z), the volume of a fluid element is $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$, thus leading to

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t}\ln(\rho\,\mathrm{d}V) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\ln(\rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z) \\ &= \frac{\mathrm{d}\ln\rho}{\mathrm{d}t} + \frac{\mathrm{d}}{\mathrm{d}x^a}\dot{x}^a. \end{aligned} \tag{5.80}$$

This expression for parcel mass conservation is satisfied by each of the parcels throughout the fluid, which implies that it holds at each point of the fluid at all times. It is therefore convenient to introduce the Eulerian velocity field $\mathbf{v} = \dot{x}$ leading to the statement of parcel mass conservation

$$\frac{\mathrm{d}\ln\rho}{\mathrm{d}t} = -\nabla\cdot\mathbf{v}. \tag{5.81}$$

Expanding the material time derivative allows this equation to be written as an Eulerian conservation law

$$\rho_{,t} + \nabla\cdot(\mathbf{v}\,\rho) = 0. \tag{5.82}$$

Assuming the parcel's volume is conserved instead of its mass leads to the non-divergence condition on the velocity used in the Boussinesq z-coordinate ocean model

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{d}V) = 0 \Rightarrow \nabla\cdot\mathbf{v} = 0. \tag{5.83}$$

5.8.2 Generalized vertical coordinates

With a generalized vertical coordinate s and Cartesian horizontal coordinates, conservation of mass states that

$$\begin{aligned} \rho\,\mathrm{d}V &= \rho\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ &= \rho\,\mathrm{d}x\,\mathrm{d}y\,z_{,s}\,\mathrm{d}s \end{aligned} \tag{5.84}$$

is materially constant. Therefore,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(\ln\rho\,\mathrm{d}V) &= \frac{\mathrm{d}(z_{,s})/\mathrm{d}t}{z_{,s}} + \frac{\dot{\rho}}{\rho} + \nabla_s\cdot\mathbf{u} + \partial_s\,\dot{s} \\ &= 0, \end{aligned} \tag{5.85}$$

where the material time derivative (Section 5.5.5) is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} = \partial_t + \mathbf{u}\cdot\nabla_s + \dot{s}\,\partial_s \tag{5.86}$$

and horizontal derivatives are taken with constant s. Rearrangment leads to the mass continuity equation in generalized vertical coordinates

$$(\rho\,z_{,s})_{,t} + \nabla_s\cdot(\rho\,z_{,s}\mathbf{u}) + \partial_s(\rho\,z_{,s}\,\dot{s}) = 0. \tag{5.87}$$

5.8.3 Volume conservation in ρ-coordinates

Volume conservation states that $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \mathrm{d}x\,\mathrm{d}y\,z_{,s}\,ds$ is a material constant, which leads to

$$(z_{,s})_{,t} + \nabla_s \cdot (z_{,s}\,\mathbf{u}) + \partial_s(z_{,s}\,\dot{s}) = 0. \tag{5.88}$$

This equation is often called the thickness equation, since it provides a prognostic equation for the layer specific thickness $z_{,s}$ (Section 5.3.2). In the special case where $s = \rho^{(p)}$ is the potential density, volume conservation takes the form

$$(z_{,\rho})_{,t} + \nabla_\rho \cdot (z_{,\rho}\,\mathbf{u}) + \partial_\rho(z_{,\rho}\,\dot{\rho}) = 0, \tag{5.89}$$

where we dropped the p label from potential density for brevity. For isentropic Boussinesq flow, $\dot{\rho} = 0$, thus bringing the thickness equation to the form commonly used in papers discussing closure theories for mesoscale eddies (e.g., Gent and McWilliams (1990))

$$(z_{,\rho})_{,t} + \nabla_\rho \cdot (z_{,\rho}\,\mathbf{u}) = 0. \tag{5.90}$$

Hence, in isopycnal coordinates, the thickness and potential density equations are explicitly coupled. For isentropic flow, this coupling results in a simplification of the thickness equation since its transport operator reduces from three-dimensional down to two-dimensional.

5.8.4 Mass conservation in pressure coordinates

In a hydrostatic fluid, pressure always increases with depth, and so pressure is a valid choice for vertical coordinate. In this case, it is appropriate to write the specific thickness weighted density in terms of pressure

$$\begin{aligned} \rho\, z_{,s} &= \rho\, z_{,p}\, p_{,s} \\ &= -g^{-1}\, p_{,s} \end{aligned} \tag{5.91}$$

where the hydrostatic balance

$$p_{,z} = -\rho\, g \tag{5.92}$$

was used for the second equality. The hydrostatic approximation therefore singles out pressure as being special, as can be seen by the absence of a time tendency term for mass conservation written in pressure coordinates

$$\nabla_p \cdot \mathbf{u} + \partial_p \dot{p} = 0. \tag{5.93}$$

Marshall et al. (1997) and deSzoeke and Samelson (2002) exploit this simple relation by mapping a Boussinesq z-coordinate ocean model to a non-Boussinesq pressure coordinate model. Atmospheric modelers have used pressure, or functions of pressure, for many years to simulate the atmosphere (e.g., Haltiner and Williams (1980)). The main reason is that the atmosphere is somewhat more compressible than the ocean, thus making the Boussinesq approximation less relevant.

The simplicity of the continuity equation (5.93) prompts our making a few words aiming to lend some intuition. For this purpose, recall that the hydrostatic pressure at a point is a function of the mass per area of fluid above this point. If we follow

a fluid parcel that maintains the hydrostatic balance, then, by definition, this parcel is bounded above by the same mass per area of fluid. If we then follow two parcels within the same vertical column of fluid, each maintaining their respective hydrostatic pressures, then the mass of fluid between the parcels remains constant, again by definition. Hence, convergence of mass into this column leads to a spreading of the pressure surfaces. This is the physical content of the continuity equation (5.93). It is directly analogous to the physical content of the z-coordinate continuity equation for a volume conserving fluid, where "mass" in the preceeding discussion is replaced by "volume."

5.9 Primitive equations

The purpose of this section is to derive the primitive equations in generalized vertical coordinates.

5.9.1 Momentum balance

Recall that generalized vertical coordinates are non-orthogonal, as exemplified by the non-diagonal metric tensor in Section 5.6. However, the "projected" nature of the coordinates (see Figure 5.4) allows us to formulate the momentum equations as a simple transformation from the geopotential form to the generalized vertical coordinate form. There are no non-trivial Christoffel symbols appearing in these equations, contrary to one's suspicion when noting the non-orthogonality of the coordinate system. Bleck (1978) emphasizes this point since there appeared to be some confusion in the literature prior to his paper, even though Starr (1945) wrote down the proper equations more than thirty years earlier.

To proceed, recall that in Section 3.5 we derived the equation describing the dynamics of a fluid parcel using geopotential coordinates. Assuming hydrostatic balance, this equation is given by

$$(\rho\,\mathbf{u})_{,t} + \nabla\cdot(\rho\,\mathbf{v}\,\mathbf{u}) = -f\,\hat{\mathbf{z}}\wedge\rho\,\mathbf{u} - \nabla_z p, \tag{5.94}$$

where we dropped friction and sphericity for simplicity. Use of mass conservation in the form (5.82) leads to

$$\rho\,\mathrm{d}\mathbf{u}/\mathrm{d}t = -f\,\hat{\mathbf{z}}\wedge\rho\,\mathbf{u} - \nabla_z p. \tag{5.95}$$

Use of the material time derivative in terms of the generalized vertical coordinate (Section 5.5.5) as well as mass conservation in the form (5.87) renders

$$(\rho\,z_{,s}\,\mathbf{u})_{,t} + \nabla_s\cdot(\rho\,z_{,s}\,\mathbf{u}\,\mathbf{u}) + (\rho\,z_{,s}\,\dot{s}\,\mathbf{u})_{,s} = -f\,\hat{\mathbf{z}}\wedge\rho\,z_{,s}\,\mathbf{u} - z_{,s}\,\nabla_z p. \tag{5.96}$$

From Section 5.5.3, we have for the horizontal pressure gradient

$$\begin{aligned}\nabla_z p &= \nabla_s p - \nabla_s z\,p_{,z}\\ &= \nabla_s p + \rho\,g\,\nabla_s z\end{aligned} \tag{5.97}$$

where the hydrostatic approximation was used for the second equality. Again, using the hydrostatic balance, we saw in Section 5.8.4 that $g\,\rho\,z_{,s} = -p_{,s}$, thus leading to the pressure gradient expression

$$z_{,s}\,\nabla_z p = z_{,s}\,\nabla_s p - p_{,s}\,\nabla_s z \tag{5.98}$$

and the momentum equation

$$(\rho\, z_{,s}\, \mathbf{u})_{,t} + \nabla_s \cdot (\rho\, z_{,s}\, \mathbf{u}\, \mathbf{u}) + (\rho\, z_{,s}\, \dot{s}\, \mathbf{u})_{,s} = -f\, \hat{\mathbf{z}} \wedge \rho\, z_{,s}\, \mathbf{u} - (z_{,s}\, \nabla_s p - p_{,s}\, \nabla_s z). \quad (5.99)$$

The term $p_{,s}\, \nabla_s z$ was seen in Section 5.2.2 when discussing the pressure gradient calculation in sigma-coordinate models.

5.9.2 Summary of the equations

It is a simple matter to derive the tracer balance in generalized vertical coordinates. Doing so yields the full set of primitive equations

$$(\rho\, z_{,s}\, \mathbf{u})_{,t} + \nabla_s \cdot (\rho\, z_{,s}\, \mathbf{u}\, \mathbf{u}) + (\rho\, z_{,s}\, \dot{s}\, \mathbf{u})_{,s} = -f\, \hat{\mathbf{z}} \wedge \rho\, z_{,s}\, \mathbf{u} - z_{,s}\, \nabla_z p \quad (5.100)$$

$$(\rho\, z_{,s}\, C)_{,t} + \nabla_s \cdot (\rho\, z_{,s} \mathbf{u}\, C) + \partial_s(\rho\, z_{,s}\, \dot{s}\, C) = 0 \quad (5.101)$$

$$(\rho\, z_{,s})_{,t} + \nabla_s \cdot (\rho\, z_{,s} \mathbf{u}) + \partial_s(\rho\, z_{,s}\, \dot{s}) = 0 \quad (5.102)$$

$$p_{,s} = -z_{,s}\, \rho\, g, \quad (5.103)$$

where we continue to ignore sub-grid-scale processes and sphericity for brevity, and where the horizontal pressure gradient is written

$$\nabla_z p = \nabla_s p + \rho\, g\, \nabla_s z. \quad (5.104)$$

When $s = z$, these equations reduce to the non-Boussinesq z-coordinate form considered in Chapters 3, 4, and 6. In the general case, we have an extra factor of the specific thickness $z_{,s}$ as well as the cumbersome two terms needed to compute the horizontal pressure gradient.

5.9.3 Non-Boussinesq in pressure coordinates

For the special case of pressure coordinates with $s = p$, we have

$$z_{,p}\, \nabla_z p = -\nabla_p z \quad (5.105)$$

and

$$\rho\, z_{,p} = -g^{-1}. \quad (5.106)$$

That is, the specific thickness weighted pressure gradient force is given by minus the slope of the isobaric surfaces. Hence, the non-Boussinesq hydrostatic primitive equations in pressure coordinates are given by the simple form

$$\mathbf{u}_{,t} + \nabla_p \cdot (\mathbf{u}\, \mathbf{u}) + (\dot{p}\, \mathbf{u})_{,p} = -f\, \hat{\mathbf{z}} \wedge \mathbf{u} - g\, \nabla_p z \quad (5.107)$$

$$C_{,t} + \nabla_p \cdot (\mathbf{u}\, C) + \partial_p(\dot{p}\, C) = 0 \quad (5.108)$$

$$\nabla_p \cdot \mathbf{u} + \partial_p \dot{p} = 0. \quad (5.109)$$

5.9.4 Isentropic Boussinesq in ρ-coordinates

In the case where we are interested in Boussinesq isentropic flows with a linear equation of state, we argued in Section 5.2.3 that isopycnal models are quite useful. In this case, we equate potential density to *in situ* density, thus leading to the horizontal pressure gradient

$$\begin{aligned} \nabla_z p &= \nabla_\rho p + \rho\, g\, \nabla_\rho z \\ &= \nabla_\rho\, (p + \rho\, g\, z). \end{aligned} \tag{5.110}$$

Defining the Montgomery potential

$$M = p + \rho\, g\, z \tag{5.111}$$

leads to the Boussinesq isentropic primitive equations written in isopycnal coordinates

$$(z_{,\rho}\, \mathbf{u})_{,t} + \nabla_\rho \cdot (z_{,\rho}\, \mathbf{u}\, \mathbf{u}) + (z_{,\rho}\, \dot{\rho})_{,\rho} = -f\, \hat{\mathbf{z}} \wedge z_{,\rho}\, \mathbf{u} - \nabla_\rho M \tag{5.112}$$

$$(z_{,\rho}\, C)_{,t} + \nabla_\rho \cdot (z_{,\rho}\, \mathbf{u}\, C) + (z_{,\rho}\, \dot{\rho}\, C)_{,\rho} = 0 \tag{5.113}$$

$$(z_{,\rho})_{,t} + \nabla_\rho \cdot (z_{,\rho}\, \mathbf{u}) + (z_{,\rho}\, \dot{\rho}\,)_{,\rho} = 0 \tag{5.114}$$

$$M_{,\rho} = \rho\, g. \tag{5.115}$$

Note the presence of only a single term for the horizontal pressure gradient force. Hence, in contrast to sigma coordinate models, for ideal flow where the equation of state is linear, isopycnal models do not suffer from difficulties numerically representing the horizontal pressure force.

5.9.5 Neutral physics in generalized coordinates

Fundamental to the primitive equation ocean models is the parameterization of processes that are unresolved. Part IV of this book focuses on the issues of neutral physics as implemented and parameterized within a z-model. Key to this implementation is the slope of a neutral direction. For a neutral diffusion scheme, where the diffusion tensor is oriented relative to the neutral direction, the slope of the neutral direction relative to the surface of constant generalized vertical coordinate is required. For schemes requiring a measure of baroclinicity, such as the Gent and McWilliams (1990) scheme, the slope slope of the neutral direction relative to the geopotential surface is required, as this slope reflects on baroclinicity in the fluid and its available potential energy.

As discussed in Section 5.3.3, and further detailed in Section 11.6, the neutral slope relative to a geopotential is given by

$$\mathbf{S}_{(\rho/z)} = -z_{,\rho}\, \nabla_z \rho \tag{5.116}$$

with ρ the locally referenced potential density. For the present purposes, it is useful to use the special (ρ/z) subscript notation to highlight that the neutral slope is relative to a geopotential. In generalized vertical coordinates, the horizontal gradient ∇_z is computed using the transformation (5.32) so that the neutral slope becomes

$$\begin{aligned} \mathbf{S}_{(\rho/z)} &= -z_{,\rho}\, (\nabla_s - \mathbf{S}_{(s/z)} \partial_z)\, \rho \\ &= -z_{,s} \left(\frac{\nabla_s \rho}{\rho_{,s}} \right) + \mathbf{S}_{(s/z)} \\ &= \mathbf{S}_{(\rho/s)} + \mathbf{S}_{(s/z)}, \end{aligned} \tag{5.117}$$

where $\mathbf{S}_{(s/z)} = \nabla_s z$ is the slope of the generalized vertical coordinate surface relative to the geopotential, and $\mathbf{S}_{(\rho/s)}$ is the slope of the neutral direction relative to the generalized vertical coordinate surface. In words, this equation reflects the simple statement that the slope of the neutral direction relative to the geopotential equals to the slope of the neutral direction relative to the generalized vertical coordinate surface, plus the slope of the generalized vertical coordinate surface relative to the geopotential. In isopycnal models, the slope $\mathbf{S}_{(\rho/s)}$ is very small for the most part, thus prompting isopycnal modelers to dispense with much of the numerical machinery described in Part IV needed to implement neutral physical processes in z-models. For σ-models, $\mathbf{S}_{(s/z)}$ can be nontrivial in much of the model domain affected by topography, whereas in pressure coordinate models $\mathbf{S}_{(s/z)}$ is typically less than 10^{-4}.

(ILLUSTRATION NEEDED).

Part II

AVERAGED OCEAN EQUATIONS

Averaged Ocean Equations

The ocean is a turbulent fluid with dynamical scales over a huge range in space and time, and we are privy to very limited information about the fluid. Consequently, we must base a practical description, such as that arising from an ocean model, on a statistical-dynamical approach. This part of the book is far from a complete treatment of this problem, which arguably is the most intractable problem in classical physics, that being the problem of fluid turbulence. Instead, the aim here is to illustrate some of the basics of how to formulate a description of ensemble-mean fluid motions.

We consider here two different ensembles. The first ensemble focuses on the small scales of motion currently unapproachable by even the finest of global ocean model resolutions. Our viewpoint is fully Eulerian, in that we position ourselves at a fixed point in space-time and average over the ensemble. The result of averaging over small-scale (order meters) overturns and irreversible mixing processes is a smooth averaged density profile.

The second ensemble focuses on the reversible stirring of fluid density in the nearly ideal ocean interior. The dynamics of this stirring are best viewed from the perspective of an isopycnal reference frame, where the isopycnals are well defined stably stratified layers resulting from an average over the small scales considered in the first ensemble. For the isopycnal reference frame, the horizontal position and time are the same as the Eulerian frame, yet the vertical position is Lagrangian as it moves up and down based on the undulations of a particular potential density surface. Mapping the isopycnal perspective back to the Eulerian perspective, a mapping needed when we interpret coarse z-model simulations, requires some tools analogous to those used in the generalized lagrangian mean formalism.

Corresponding to each of the two ensembles is an interpretation for the fields carried by the ocean model, with the interpretation depending on the physical regimes explicitly represented. Our dual ensemble approach is largely motivated by the work of deSzoeke and Bennett (1993).

Determining a transparent and tidy set of averaged equations requires care in nomenclature, frames of reference, and interpretations. Some of the steps are subjective and based on convenience and desire. Indeed, how to average over an ensemble of fluid states is largely an art. One can certainly perform a traditional Reynolds decomposition, but will the resulting correlation terms all be worth parameterizing? As seen in our discussions, mindfulness in choosing coordinates and weights can reduce the number of terms unavailable to the coarsened state, thus easing the burden on what we need to parameterize. These kinematic issues become critical when attacking a statistical-dynamical problem for rotating-stratified flows. That is, care in setting up the problem goes a long way towards exposing the key physical results and rendering tidy mean-field equations.

Sub-grid scale (SGS) transport operators arise from averaging the equations of motion. These operators must be parameterized prior to integrating the mean-field equations. Parameterizing SGS processes remains one of the most complex, rich, and critical areas of theoretical oceanography. There will always be regimes that an ocean model does not resolve, and how one parameterizes these regimes is a key element determining the physical integrity of simulations. Some of the issues involved with SGS parameterizations are introduced in Chapter 8. We also ex-

pose some issues important for parameterizing SGS dianeutral processes, which are processes that transfer properties across the smoothed neutral directions. We have more to say on the parameterization of certain lateral SGS processes in Parts IV and V of this book.

CHAPTER
SIX

Ensemble-mean ocean equations

Contents

6.1 Introduction **144**
6.1.1 Ensemble averages 144
6.1.2 Interpreting ocean model equations 144
6.1.3 Mean kinematics independent of dynamics 145
6.2 Unaveraged primitive equations **146**
6.3 Averaged kinematics of a fluid parcel **148**
6.4 Averaged kinematics over finite domains **149**
6.4.1 Averaged seawater mass and tracer mass 149
6.4.2 Modified mean surface height 150
6.4.3 Kinematic boundary condition 151
6.4.4 Mass within a column 152
6.5 Averaged tracer in a parcel **153**
6.5.1 Conventional approach 153
6.5.2 Density weighted approach 154
6.5.3 Sub-grid scale tracer fluxes 155
6.5.4 Boussinesq limit in the averaged fluid 155
6.5.5 A not-so-useful alternative 156
6.6 Averaged momentum budget **156**
6.6.1 Time tendency for $\langle \mathbf{v} \rangle$ 156
6.6.2 Time tendency for $\langle \tilde{\mathbf{v}} \rangle$ 157
6.7 Summary of the averaged equations **158**
6.8 Mapping to ocean model variables **159**
6.9 Discretization basics **162**

The purpose of this chapter is to ensemble average the ocean primitive equations, where the average is taken at a fixed point in space-time. We also motivate a particular interpretation of the variables used in numerical ocean models as a discrete representation of the ensemble averaged equations.

6.1 Introduction

The ocean is fundamentally turbulent. Given sensitive dependence on initial conditions, and the limitations of ocean measurements, it is not possible to obtain complete knowledge of the ocean state. Therefore, when formulating the ocean's governing equations, it is necessary to recognize our limited access to information.

6.1.1 Ensemble averages

A common way to formulate a dynamical theory recognizing an incomplete level of information is to consider ensemble averages. As in statistical mechanics, ensemble averages in oceanography are obtained by formally considering an infinite number of ocean states, each of which is described by the kinematic and dynamic balances of Chapter 3 and the thermodynamic and tracer balances of Chapter 4. Our interpretation of ensemble averages follows that given in Section 2.5, where the distance in phase space between members of an ensemble is directly related to the space-time scales that are not resolved by the "averaged observer."

The ensemble averages considered here are taken at a fixed point in space-time. These averages commute with space-time derivatives and integrals. As noted by deSzoeke and Bennett (1993), such averages are appropriate when coarsening over micro-structure processes (order meters or less) that induce small-scale turbulent mixing between density classes. We have more to say about these physical processes in Section 8.3. A particular result of averaging over the micro-scale is a smoothed mean density field suitable for use as a vertical coordinate in ocean climate models (see Chapters 5 and 7 for more discussion and use of isopycnal coordinates).

Even though we specify averaging at a particular space-time point, there remain ambiguities in details. Different methods reveal different aspects of the averaged or mean dynamics, and deviations from the mean. *Density weighted averaging* plays a central role in our preferred method, with density here referring to *in situ* density. Such averaging has recently become more commonplace in the ocean modeling literature (e.g., Smith (1999), McDougall et al. (2002a), and Lu (2001). However, it has been used in various other oceanographic studies such as Osborn and Cox (1972) and McDougall and Garrett (1992), as well as other areas of the compressible fluids literature (e.g., Favre (1965) as noted in the footnote on pages 21-22 of Hinze (1975)). It appears that Favre (1965) is the first to use density weighted averages, and so they are sometimes referred to as "Favre averages."

6.1.2 Interpreting ocean model equations

Besides their importance from a fundamental perspective, ensemble averaged ocean equations form the basis for the equations discretized in an ocean model. That is, a consistent interpretation of the ocean model equations is afforded by considering them to be discretizations of ensemble averaged continuum equations, where the average is taken over space-time scales smaller than the space-time scale of the model grid.

In this chapter, we focus on averages taken at a fixed point in space-time. This average is appropriate when averaging over small scales of motion where the dy-

namics is mostly isotropic in all three dimensions. At scales approaching the mesoscale, processes tend to occur along neutral directions, in which case the isopycnal averaging techniques discussed in Chapter 7 are far more suitable. This dual approach to averaging follows the notions described by deSzoeke and Bennett (1993). For still larger scales, one typically uses multiple realizations from numerical ocean models to explicitly form the average, such as when considering the averaged response of the large-scale ocean circulation to a suite of atmospheric states (e.g., Griffies and Bryan (1997)). Providing a theory for the resulting averaged ocean climate, its variability, and its predictability constitutes a theory for the ocean climate itself (e.g., Peixoto and Oort (1992)).

6.1.3 Mean kinematics independent of dynamics

There are two forms of fluid parcel kinematics considered in this book: parcels conserving their volume and parcels conserving their mass. As described for the unaveraged ocean fluid in Chapter 3, kinematic relations provide constraints maintained regardless the dynamics. Therefore, we believe it to be key to the integrity and usability of the equations describing the ensemble averaged ocean, and the ocean model, that kinematics of the averaged fluid remain independent of dynamical assumptions. In particular, we do not wish to require specification of unknown closure terms, whose form depends on dynamical details, in order to determine kinematic relations satisfied by the averaged ocean or the corresponding ocean model. Care in formulating and interpreting the averaged equations is required to satisfy this principle.

A key motivation for using density weighted averaging is that it assists in our desire to keep the averaged parcel kinematics independent of dynamical closure assumptions. Providing such a simple mapping between unaveraged and averaged kinematics generally does not require much thought when averaging the Boussinesq equations in z-coordinates, since volume conservation $\nabla \cdot \mathbf{v} = 0$ is a linear constraint. Yet for non-Boussinesq equations, mass conservation $\rho_{,t} + \nabla \cdot (\mathbf{v}\,\rho) = 0$ is a nonlinear constraint, thus requiring extra consideration. Note that the converse is true when working in pressure coordinates, since mass conservation is a linear constraint for a non-Boussinesq hydrostatic fluid using pressure coordinates (see Section 5.8.4). This point illustrates the coordinate dependent aspects of averaging, whereby certain coordinates more convenient than others when averaging over certain scales. Arguably, pressure coordinates are ideal for non-Boussinesq fluids given the simplicity of the resulting non-Boussinesq primitive equations. Nonetheless, we focus here on z-coordinates based on their dominance in present-day ocean climate modeling.

Besides the mathematical utility of the density weighed approach for the non-Boussinesq system, McDougall et al. (2002a) argued for maintaining this average even when considering the averaged Boussinesq equations. Their reasoning is based on noting that the resulting Boussinesq system is far more accurate than the mean-field equations resulting from non-density weighted averages. In this context, accuracy is based on comparing with the small levels of diapycnal mixing in the ocean interior.

6.2 Unaveraged primitive equations

Before considering the averaged equations, it is useful to summarize the unaveraged equations obtained in Chapters 3 and 4. Start by recalling the non-Boussinesq equations written in z-coordinates. Mass conservation or continuity reflect the kinematics of the fluid parcels

$$\rho_{,t} + \nabla \cdot (\rho \, \mathbf{v}) = 0, \tag{6.1}$$

and tracer conservation reflects the conservation of material and/or thermodynamic tracers

$$(\rho \, C)_{,t} + \nabla \cdot (\rho \, C \, \mathbf{v}) = -\nabla \cdot (\rho \, \mathbf{F}) + \rho \, \mathcal{S}. \tag{6.2}$$

Finally, Newton's Second Law applied to a fluid parcel leads to the linear momentum balance

$$(\rho \, \mathbf{v})_{,t} + \nabla \cdot (\rho \, \mathbf{v} \, \mathbf{v}) + \mathcal{M} \, (\hat{\mathbf{z}} \wedge \rho \, \mathbf{v}) = -\rho \, g \, \hat{\mathbf{z}} - f \, \hat{\mathbf{z}} \wedge \rho \, \mathbf{v} - \nabla p + \rho \, \mathbf{F}^{(\mathbf{v})}. \tag{6.3}$$

We interpret the tracer flux $\mathbf{F}$ as that arising from sub-grid scale (SGS) molecular processes, such as molecular diffusion. Likewise, the friction vector $\mathbf{F}^{(\mathbf{v})}$ is associated with momentum transport due to molecular viscosity. Averaging introduces far more significant SGS processes associated with turbulent mixing and stirring. These processes comprise the bulk of the SGS tracer and momentum fluxes appearing in the averaged balances.

As noted in Section 5.9.2, the primitive equations written in generalized vertical coordinates have ρ mapped to $z_{,s} \, \rho$, with $z_{,s}$ the specific thickness. The pressure gradient force is also generalized to two terms. Although there are many important details that depend on the choice of vertical coordinates, much of the following focus on z-coordinates can be straightforwardly generalized to arbitary vertical coordinates, as illustrated by Greatbatch and McDougall (2002). Our focus on z-coordinates is largely based on simplicity in presentation.

The linear momentum density (momentum per volume)

$$\rho \, \mathbf{v} = \rho_o \, \tilde{\mathbf{v}} \tag{6.4}$$

plays a fundamental role in the dynamical balances. It proves to be especially useful when formulating the averaged equations. Additionally, we will find it useful to introduce an analogous form for the subgrid-scale molecular fluxes of tracer and momentum

$$\rho \, \mathbf{F} = \rho_o \, \tilde{\mathbf{F}} \tag{6.5}$$

$$\rho \, \mathbf{F}^{(\mathbf{v})} = \rho_o \, \tilde{\mathbf{F}}^{(\mathbf{v})}. \tag{6.6}$$

Introducing these fields leads to the mass and tracer budgets

$$\rho_{,t} + \rho_o \, \nabla \cdot \tilde{\mathbf{v}} = 0 \tag{6.7}$$

$$(\rho \, C)_{,t} + \rho_o \, \nabla \cdot (C \, \tilde{\mathbf{v}}) = -\rho_o \, \nabla \cdot \tilde{\mathbf{F}} + \rho \, \mathcal{S}, \tag{6.8}$$

as well as the linear momentum budget

$$\tilde{\mathbf{v}}_{,t} + \nabla \cdot [(\rho_o/\rho)\, \tilde{\mathbf{v}}\, \tilde{\mathbf{v}}] + (\rho_o/\rho)\, \tilde{\mathcal{M}}\, \hat{\mathbf{z}} \wedge \tilde{\mathbf{v}} = -(\rho/\rho_o)\, g\, \hat{\mathbf{z}} - f\, \hat{\mathbf{z}} \wedge \tilde{\mathbf{v}} - \nabla(p/\rho_o) + \tilde{\mathbf{F}}^{(\mathbf{v})} \quad (6.9)$$

where we wrote the advection metric frequency as

$$\begin{aligned} \mathcal{M} &= v\, \partial_x \ln \mathrm{d}y - u\, \partial_y \ln \mathrm{d}x \\ &= (\rho_o/\rho)\, (\tilde{v}\, \partial_x \ln \mathrm{d}y - \tilde{u}\, \partial_y \ln \mathrm{d}x) \\ &= (\rho_o/\rho)\, \tilde{\mathcal{M}}. \end{aligned} \quad (6.10)$$

Integrating the continuity equation over the full ocean column of thickness $D = H + \eta$, and assuming mass sources/sinks only at the ocean surface, leads to the balance of mass per unit area within an ocean column

$$(\overline{\rho}^z\, D)_{,t} = -\rho_o\, \nabla \cdot \tilde{\mathbf{U}} + \rho_w\, q_w \quad (6.11)$$

as well as the surface and bottom kinematic boundary conditions

$$\rho\, (\partial_t + \mathbf{u} \cdot \nabla)\, \eta = \rho_w\, q_w + \rho\, w \qquad \text{at } z = \eta \quad (6.12)$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H. \quad (6.13)$$

In these expressions, $\rho_o\, \tilde{\mathbf{U}} = \int_{-H}^{\eta} \mathrm{d}z\, \rho\, \mathbf{u}$ is the depth integrated horizontal momentum density and $D\, \overline{\rho}^z = \int_{-H}^{\eta} \mathrm{d}z\, \rho$ is the depth integrated density. The mass per unit time per unit horizontal area of fresh water crossing the ocean surface is given by $\rho_w\, q_w$, which is also a linear momentum per volume. As for the linear momentum per volume of ocean fluid $\rho\, \mathbf{v} = \rho_o\, \tilde{\mathbf{v}}$, we find it useful to introduce the analogous quantity for the surface fresh water flux

$$\rho_w\, q_w \equiv \rho_o\, \tilde{q}_w, \quad (6.14)$$

which leads then to

$$(\overline{\rho}^z\, D)_{,t} = -\rho_o\, \nabla \cdot \tilde{\mathbf{U}} + \rho_o\, \tilde{q}_w \quad (6.15)$$

$$\rho\, (\partial_t + \mathbf{u} \cdot \nabla)\, \eta = \rho_o\, \tilde{q}_w + \rho\, w. \qquad \text{at } z = \eta \quad (6.16)$$

The Boussinesq equations are recovered by setting $\rho \to \rho_o$ wherever it appears, except when multiplying gravity. The resulting continuity equation becomes a constraint on the three-dimensional velocity field

$$\nabla \cdot \mathbf{v} = 0, \quad (6.17)$$

the conservative form of the tracer equation takes the form

$$C_{,t} + \nabla \cdot (C\, \mathbf{v}) = -\nabla \cdot \mathbf{F} + \mathcal{S}, \quad (6.18)$$

and linear momentum balance becomes

$$\mathbf{v}_{,t} + \nabla \cdot (\mathbf{v}\, \mathbf{v}) + \mathcal{M}\, \hat{\mathbf{z}} \wedge \mathbf{v} = -(\rho/\rho_o)\, g\, \hat{\mathbf{z}} - f\, \hat{\mathbf{z}} \wedge \mathbf{v} - \nabla(p/\rho_o) + \mathbf{F}^{(\mathbf{v})}. \quad (6.19)$$

Note that $\nabla \cdot \mathbf{v} = 0$ means the Boussinesq fluid parcels conserve their volume instead of their mass. Integrating the continuity equation over a full ocean column, and assuming volume sources/sinks only at the ocean surface, leads to the balance of volume per unit area within an ocean column

$$\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w \tag{6.20}$$

as well as the surface and bottom kinematic boundary conditions

$$(\partial_t + \mathbf{u} \cdot \nabla)\, \eta = q_w + w \qquad \text{at } z = \eta \tag{6.21}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H, \tag{6.22}$$

where $\mathbf{U} = \int_{-H}^{\eta} \mathrm{d}z\, \mathbf{u}$ is the depth integrated horizontal velocity field, and q_w is the volume per unit time per unit horizontal area of fresh water crossing the ocean surface.

An important point to note is that the linear momentum density of a Boussinesq fluid is $\rho_o\, \mathbf{v}$, instead of the non-Boussinesq $\rho\, \mathbf{v}$. Hence, there is no difference between the velocity fields $\tilde{\mathbf{v}}$ and $\mathbf{v}$. The unaveraged Boussinesq equations then take the same form whether written in terms of $\tilde{\mathbf{v}}$ and $\mathbf{v}$. Averaging, however, breaks the symmetry between the averaged value of $\tilde{\mathbf{v}}$ and the averaged $\mathbf{v}$, as discussed next.

6.3 Averaged kinematics of a fluid parcel

The averaged continuity equation using velocity (linear momentum per mass) as the fundamental transport field is given by

$$\rho_{,t} + \nabla \cdot (\rho\, \mathbf{v}) = 0 \rightarrow \langle \rho \rangle_{,t} + \nabla \cdot (\langle \mathbf{v} \rangle \langle \rho \rangle + \langle \mathbf{v}' \rho' \rangle) = 0, \tag{6.23}$$

where the angled brackets signify averages and primed variables are deviations from the average. The nonlinear correlation term in the averaged mass balance (6.23) is extremely cumbersome since, to study the averaged kinematics, it is necessary to specify the unknown correlation term $\nabla \cdot \langle \mathbf{v}' \rho' \rangle$. The form of this term is dependent on the space-time scales averaged over, and the associated SGS dynamics. In contrast, the averaged momentum per unit volume $\rho_o\, \tilde{\mathbf{v}}$ absorbs the correlation terms

$$\begin{aligned} \rho_o \langle \tilde{\mathbf{v}} \rangle &= \langle \rho\, \mathbf{v} \rangle \\ &= \langle \rho \rangle \langle \mathbf{v} \rangle + \langle \rho'\, \mathbf{v}' \rangle, \end{aligned} \tag{6.24}$$

thus leading to the more tidy averaged continuty equation

$$\rho_{,t} + \rho_o\, \nabla \cdot \tilde{\mathbf{v}} = 0 \rightarrow \langle \rho \rangle_{,t} + \rho_o\, \nabla \cdot \langle \tilde{\mathbf{v}} \rangle = 0. \tag{6.25}$$

Notably, we have not determined the value of the correlation term $\langle \rho'\, \mathbf{v}' \rangle$. Rather, we simply formed a mean-field mass budget that does not need its value by focusing on the averaged linear momentum per unit volume instead of the averaged linear momentum per unit mass. There is nothing fundamental that warrants the use of one or the other. Instead, we are guided by convenience in the resulting equations.

These observations reflect the subjective nature of what turbulent fluctuations are needed for closing the mean-field equations. For purposes of working with the averaged continuity equation in z-coordinates, the single linear term $\rho_o \langle \tilde{\mathbf{v}} \rangle$ is more convenient than the nonlinear alternative $\langle \rho \rangle \langle \mathbf{v} \rangle + \langle \rho' \, \mathbf{v}' \rangle$. This result promotes density weighted averaging as a key aspect of averaging the remaining equations.

6.4 Averaged kinematics over finite domains

Recall from Section 3.3.3 that the kinematic boundary condition for the ocean surface was derived by considering the time tendency of mass within a finite ocean domain. We procede here in a similar manner for the averaged non-Boussinesq fluid.

6.4.1 Averaged seawater mass and tracer mass

The average mass within the full ocean domain is a functional of the ocean geometry, density and surface height. Assuming the geometry to be fixed over the average leads to the functional relation

$$\begin{aligned} \langle M(\rho, \eta) \rangle &= \left\langle \int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z \, \rho \right\rangle \\ &= \int \mathrm{d}A \left\langle \int_{-H}^{\eta} \mathrm{d}z \, \rho \right\rangle , \end{aligned} \tag{6.26}$$

where the second step follows since the horizontal boundaries of the ocean are fixed over the ensemble. More generally, the total tracer mass within the ocean is given by

$$\begin{aligned} M_C &= \int (\rho \, \mathrm{d}V) \, C \\ &= \int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z \, \rho \, C, \end{aligned} \tag{6.27}$$

and its average is

$$\begin{aligned} \langle M_C(\rho, \eta, C) \rangle &= \left\langle \int \mathrm{d}A \int_{-H}^{\eta} \mathrm{d}z \, \rho \, C \right\rangle \\ &= \int \mathrm{d}A \left\langle \int_{-H}^{\eta} \mathrm{d}z \, \rho \, C \right\rangle . \end{aligned} \tag{6.28}$$

When $C = 1$, the ocean tracer mass reduces to the total mass of water in the ocean.

In general, the averaged mass $\langle M \rangle$ of ocean fluid *does not* equal to the mass of a fluid with average density $\langle \rho \rangle$ and averaged surface height $\langle \eta \rangle$. That is,

$$\langle M(\rho, \eta) \rangle \neq M(\langle \rho \rangle, \langle \eta \rangle). \tag{6.29}$$

Likewise, and more generally,

$$\langle M_C(\rho, \eta, C)\rangle \neq M_C(\langle\rho\rangle, \langle\eta\rangle, \langle C\rangle). \tag{6.30}$$

Instead, it is necessary to introduce the averaged surface height $\langle\eta\rangle = \eta - \eta'$, the averaged density $\langle\rho\rangle = \rho - \rho'$, and averaged tracer $\langle C\rangle = C - C'$. For the ocean mass, these terms lead to

$$\begin{aligned}\left\langle \int_{-H}^{\eta} \mathrm{d}z\, \rho \right\rangle &= \left\langle \int_{-H}^{\langle\eta\rangle+\eta'} \mathrm{d}z\, (\langle\rho\rangle + \rho') \right\rangle \\ &= \int_{-H}^{\langle\eta\rangle} \mathrm{d}z\, \langle\rho\rangle + \left\langle \int_{\langle\eta\rangle}^{\eta'} \mathrm{d}z\, \rho' \right\rangle,\end{aligned} \tag{6.31}$$

where $\langle\rho'\rangle = 0$ and $\langle\eta'\rangle = 0$ were used. The second term can be expanded in a Taylor series around $\eta' = 0$ to yield

$$\left\langle \int_{\langle\eta\rangle}^{\eta'} \mathrm{d}z\, \rho' \right\rangle \approx \langle\eta'\,\rho'\rangle + (1/2)\,\langle(\eta')^2\,\partial_z\,\rho'\rangle \tag{6.32}$$

where ρ' on the right hand side is evaluated at $z = 0$, and higher order correlations were dropped. Therefore, in order to formulate expressions for the averaged mass, it is necessary to introduce a closure assumption for the infinite number of correlation terms. Similar assumptions are needed for the averaged tracer mass.

6.4.2 Modified mean surface height

Although the above approach is systematic and rigorous, it is unsatisfying since it produces an averaged kinematics differing fundamentally from the unaveraged kinematics, and the averaged kinematics depends on dynamical details determining the correlations. This result is symptomatic of our working with the mean surface height field $\langle\eta\rangle$ and mean tracer $\langle C\rangle$, both of which are inconvenient.

A more satisfying approach starts by introducing the *modified mean* surface height η^*, defined according to

$$\begin{aligned}\langle M_C\rangle &= \int \mathrm{d}A \int_{-H}^{\eta^*} \mathrm{d}z\, \langle\rho\, C\rangle \\ &= \int \mathrm{d}A \int_{-H}^{\eta^*} \mathrm{d}z\, \langle\rho\rangle\, \langle C\rangle^{\rho},\end{aligned} \tag{6.33}$$

where $\langle\rho\rangle\langle C\rangle^{\rho} = \langle\rho\, C\rangle$ introduced the density weighted ensemble averaged tracer concentration $\langle C\rangle^{\rho}$. This average proves to be of further use when considering the averaged tracer within a fluid parcel in Section 6.5. In short, the introduction of η^* is defined so that

$$\langle M_C(\rho, \eta, C)\rangle \equiv M_C(\langle\rho\rangle, \eta^*, \langle C\rangle^{\rho}). \tag{6.34}$$

In words, the average tracer mass in the ocean (left hand side) equals to the mass of tracer in an ocean with average density $\langle \rho \rangle$, density weighted tracer $\langle C \rangle^{\rho}$, and surface height η^* (right hand side). Setting $\langle C \rangle^{\rho}$ to unity reduces the average tracer mass to average ocean fluid mass, as expected from compatibility between tracer and fluid masses. It is notable that the modified surface height is analogous to the modified density field introduced in Section 7.3.4, although from different motivation.

It is possible, through pursuit of the Taylor series methods discussed previously, to express η^* in terms of $\langle \eta \rangle$, $\langle \rho \rangle$, $\langle C \rangle$, and correlations between η', C', and ρ'. Analogous derivations were provided by McIntosh and McDougall (1996). However, there is no need to pursue such algebra since η^* is sufficient to formulate the averaged kinematics of finite fluid domains. It therefore serves as our mean-field surface height.

6.4.3 Kinematic boundary condition

Continuing to use the modified surface height, we are led to develop the time tendency for the averaged mass

$$\langle M \rangle_{,t} = \int \mathrm{d}A \left(\langle \rho(\eta^*) \rangle \, \eta^*_{,t} + \int_{-H}^{\eta^*} \mathrm{d}z \, \langle \rho \rangle_{,t} \right) . \tag{6.35}$$

Use of the averaged continuity equation (6.25) yields

$$\begin{aligned} \langle M \rangle_{,t} = & \int_{z=\eta^*} \mathrm{d}A \left(\langle \rho \rangle \, \eta^*_{,t} + \rho_o \, \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* - \rho_o \, \langle \tilde{w} \rangle \right) \\ & + \int_{z=-H} \mathrm{d}A \, \rho_o \, \left(\langle \tilde{w} \rangle + \langle \tilde{\mathbf{u}} \rangle \cdot \nabla H \right) - \int \mathrm{d}A \, \nabla \cdot \int_{-H}^{\eta^*} \mathrm{d}z \, \rho_o \, \langle \tilde{\mathbf{u}} \rangle . \end{aligned} \tag{6.36}$$

Averaging the bottom kinematic boundary condition (3.17) leads to

$$\langle \tilde{w} \rangle + \langle \tilde{\mathbf{u}} \rangle \cdot \nabla H = 0 . \tag{6.37}$$

Furthermore, either no normal flow or periodic side boundary conditions allows us to drop the convergence term, thus leading to the balance

$$\langle M \rangle_{,t} = \int_{z=\eta^*} \mathrm{d}A \left(\langle \rho \rangle \, \eta^*_{,t} + \rho_o \, \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* - \rho_o \, \langle \tilde{w} \rangle \right) . \tag{6.38}$$

To proceed, recall that the ocean mass is altered only through passage of fresh water through the ocean surface. Averaging the budget for mass in the full ocean, equation (3.33), leads to

$$\begin{aligned} \langle M \rangle_{,t} &= \left\langle \int_{z=\eta} \mathrm{d}A \, \rho_w \, q_w \right\rangle \\ &= \rho_o \left\langle \int_{z=\eta} \mathrm{d}A \, \tilde{q}_w \right\rangle \end{aligned} \tag{6.39}$$

where

$$\begin{aligned} \rho_w \, q_w &= \rho_o \, \tilde{q}_w \\ &= (\hat{\mathbf{z}} \cdot \hat{\mathbf{N}}) \, (P - E + R) \, \rho_w , \end{aligned} \tag{6.40}$$

with $\hat{\mathbf{N}} = (-\nabla \eta, 1)$ the orientation vector at the ocean surface. Defining a modified mean surface fresh water flux according to

$$\left\langle \int_{z=\eta} \mathrm{d}A \, \tilde{q}_w \right\rangle \equiv \int_{z=\eta^*} \mathrm{d}A \, \tilde{q}_w^* , \tag{6.41}$$

leads to

$$\int_{z=\eta^*} \mathrm{d}A \left(\langle \rho \rangle \, \eta_{,t}^* + \rho_o \, \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* - \rho_o \, \langle \tilde{w} \rangle \right) = \rho_o \int_{z=\eta^*} \mathrm{d}A \, \tilde{q}_w^* . \tag{6.42}$$

As the horizontal area is arbitrary, we are led to the surface kinematic boundary condition

$$\langle \rho \rangle \, \eta_{,t}^* + \rho_o \, \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* = \rho_o \, \langle \tilde{w} \rangle + \rho_o \, \tilde{q}_w^* \quad \text{at } z = \eta^* . \tag{6.43}$$

In the Boussinesq limit this boundary condition becomes

$$\eta_{,t}^* + \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* = \langle \tilde{w} \rangle + \tilde{q}_w^* \quad \text{at } z = \eta^* . \tag{6.44}$$

Note the absence of turbulence correlation terms in both surface boundary conditions. Therefore, both are directly analogous to the unaveraged equations given in Section 6.2, which was our motivation for introducing the modified mean fields η^* and $\tilde{q}_w^*$.

6.4.4 Mass within a column

We now consider the budget for the mass per unit area of fluid contained within a single ocean column. For this case, maintain the convergence term in equation (6.36) and use the surface kinematic boundary condition (6.43) to find

$$\partial_t \left(\int_{-H}^{\eta^*} \mathrm{d}z \, \langle \rho \rangle \right) + \rho_o \, \nabla \cdot \left(\int_{-H}^{\eta^*} \mathrm{d}z \, \langle \tilde{\mathbf{u}} \rangle \right) = \rho_o \, \tilde{q}_w^* . \tag{6.45}$$

which is directly analogous to the unaveraged budget (3.32). Introducing the vertically integrated average density

$$D^* \, \overline{\langle \rho \rangle}^z = \int_{-H}^{\eta^*} \mathrm{d}z \, \langle \rho \rangle , \tag{6.46}$$

with

$$D^* = H + \eta^* \tag{6.47}$$

the total thickness of the modified mean fluid column, and the averaged horizontal momentum density of a column

$$\rho_o \, \langle \tilde{\mathbf{U}} \rangle = \rho_o \int_{-H}^{\eta^*} \mathrm{d}z \, \langle \tilde{\mathbf{u}} \rangle , \tag{6.48}$$

leads to the mass balance for a column of fluid

$$\left(D^* \, \overline{\langle \rho \rangle}^z\right)_{,t} = -\rho_o \, \nabla \cdot \langle \tilde{\mathbf{U}} \rangle + \rho_o \, \tilde{q}^*_w. \tag{6.49}$$

The Boussinesq limit recovers

$$\eta^*_{,t} = -\nabla \cdot \langle \tilde{\mathbf{U}} \rangle + \tilde{q}^*_w. \tag{6.50}$$

Both balances are directly analogous to the unaveraged balances given in Section 6.2.

6.5 Averaged tracer in a parcel

Our goal here is to derive a mean-field tracer equation which is compatible with the preferred mean-field continuity equation derived in Section 6.3.

6.5.1 Conventional approach

Let us first consider a conventional Reynolds decomposition of the tracer equation for non-Boussinesq fluids. The product of density and tracer becomes

$$\langle \rho \, C \rangle = \langle \rho \rangle \, \langle C \rangle + \langle \rho' \, C' \rangle, \tag{6.51}$$

and the triple product becomes

$$\begin{aligned} \langle \rho \, \mathbf{v} \, C \rangle = \langle \rho \rangle \, (\langle \mathbf{v} \rangle \langle C \rangle + \langle \mathbf{v}' \, C' \rangle) \\ + \langle \mathbf{v} \rangle \langle \rho' \, C' \rangle + \langle C \rangle \langle \rho' \, \mathbf{v}' \rangle + \langle \rho' \, \mathbf{v}' \, C' \rangle, \end{aligned} \tag{6.52}$$

thus leading to the mean tracer equation

$$\begin{aligned} (\langle \rho \rangle \, \langle C \rangle + \langle \rho' \, C' \rangle)_{,t} = -\nabla \cdot [\langle \rho \rangle \, (\langle \mathbf{v} \rangle \langle C \rangle + \langle \mathbf{v}' \, C' \rangle) \\ + \langle \mathbf{v} \rangle \langle \rho' \, C' \rangle + \langle C \rangle \langle \rho' \, \mathbf{v}' \rangle + \langle \rho' \, \mathbf{v}' \, C' \rangle] \end{aligned} \tag{6.53}$$

where the source and molecular diffusion terms were temporarily dropped for brevity. Note the presence of five turbulence correlation terms, whose forms are generally unknown. Additionally, by setting $\langle C \rangle = 1$ and $C' = 1$ we recover the mean continuity equation

$$\langle \rho \rangle_{,t} = -\nabla \cdot (\langle \rho \rangle \, \langle \mathbf{v} \rangle + \langle \mathbf{v}' \, \rho' \rangle). \tag{6.54}$$

Recall that in Section 6.3, we argued that this form of the averaged continuity equation is not convenient since it leads to a mean kinematics involving unknown turbulence correlation terms. Therefore, we conclude that a conventional Reynolds decomposition of the tracer equation is unsatisfying since it leads to multiple turbulence terms and an averaged tracer equation that is not consistent with the preferred form of the averaged continuity equation.

6.5.2 Density weighted approach

To motivate an alternative approach, let us start by recalling that in Section 4.2.1, material tracer concentration C is interpreted as the mass of tracer substance within a parcel of seawater per mass of seawater

$$\begin{aligned} C &= \frac{dM_C}{dM} \\ &= \frac{dM_C}{\rho \, \mathrm{d}V}. \end{aligned} \tag{6.55}$$

As defined, C is a scalar since the tracer mass and seawater parcel mass are both scalars. If instead of following the parcel, we take an Eulerian perspective and consider the ensemble averaged tracer mass within a *fixed* volume, such as a discrete model grid cell, then we are led to

$$\begin{aligned} \langle dM_C \rangle &= \langle \rho \, C \rangle \, \mathrm{d}V \\ &= \langle \rho \rangle \langle C \rangle^{\rho} \, \mathrm{d}V, \end{aligned} \tag{6.56}$$

where

$$\langle \rho \rangle \, \langle C \rangle^{\rho} \equiv \langle \rho \, C \rangle \tag{6.57}$$

introduced the *density weighted* ensemble average tracer concentration $\langle C \rangle^{\rho}$. We already saw in Section 6.4.2 how this density weighted tracer field appeared naturally when considering the averaged tracer mass in the full ocean domain. We use this average here, along with the density weighted velocity field

$$\begin{aligned} \langle \rho \rangle \, \langle \mathbf{v} \rangle^{\rho} &\equiv \langle \rho \, \mathbf{v} \rangle \\ &= \rho_o \langle \tilde{\mathbf{v}} \rangle \end{aligned} \tag{6.58}$$

and seek a mean-field tracer equation in terms of these ensemble means.

To start, we note that deviations from a density weighted average, defined according to

$$C = \langle C \rangle^{\rho} + C'_{\rho} \tag{6.59}$$

$$\mathbf{v} = \langle \mathbf{v} \rangle^{\rho} + \mathbf{v}'_{\rho}, \tag{6.60}$$

satisfy

$$\langle \rho \, C'_{\rho} \rangle = 0 \tag{6.61}$$

$$\langle \rho \, \mathbf{v}'_{\rho} \rangle = 0. \tag{6.62}$$

Hence, the triple product term in the tracer equation takes the form

$$\begin{aligned} \rho \, C \, \mathbf{v} &= \rho \left(\langle C \rangle^{\rho} + C'_{\rho} \right) \left(\langle \mathbf{v} \rangle^{\rho} + \mathbf{v}'_{\rho} \right) \\ &= \rho \, \langle C \rangle^{\rho} \, \langle \mathbf{v} \rangle^{\rho} + \rho \, \langle C \rangle^{\rho} \, \mathbf{v}'_{\rho} + \rho \, C'_{\rho} \, \langle \mathbf{v} \rangle^{\rho} + \rho \, C'_{\rho} \, \mathbf{v}'_{\rho}, \end{aligned} \tag{6.63}$$

which has an average given by

$$\begin{aligned} \langle \rho \, C \, \mathbf{v} \rangle &= \langle C \rangle^{\rho} \, \langle \rho \rangle \, \langle \mathbf{v} \rangle^{\rho} + \langle \rho \, C'_{\rho} \, \mathbf{v}'_{\rho} \rangle \\ &= \rho_o \, \langle C \rangle^{\rho} \, \langle \tilde{\mathbf{v}} \rangle + \langle \rho \, C'_{\rho} \, \mathbf{v}'_{\rho} \rangle \end{aligned} \tag{6.64}$$

where the identities (6.61) and (6.62) were used. Notably, the five turbulence correlation terms from the conventional approach (equation (6.52)) have been reduced to a single unknown correlation term in equation (6.68) via use of density weighted averages. This is a tremendous simplification of the closure problem.

6.5.3 Sub-grid scale tracer fluxes

In addition to the molecular diffusive flux of tracer, the ensemble mean of

$$\rho \, \mathbf{F}_{sgs} = \rho \, C'_\rho \, \mathbf{v}'_\rho \tag{6.65}$$

provides a sub-grid scale (SGS) contribution to the evolution of ensemble mean tracer. This term arises from the density weighted SGS fluctuations of tracer concentration and velocity. For ocean climate modeling, molecular diffusion is far smaller than the SGS contribution from $C'_\rho \, \mathbf{v}'_\rho$. Hence, molecular diffusion contributions will be formally incorporated into $\rho \, \mathbf{F}_{sgs}$ in the following development.

In a manner similar to our introduction of the density weighted velocity field $\tilde{\mathbf{v}}$ via the definition $\rho \, \mathbf{v} = \rho_o \, \tilde{\mathbf{v}}$, we now introduce the SGS tracer flux $\tilde{\mathbf{F}}_{sgs}$ via

$$\rho \, \mathbf{F}_{sgs} \equiv \rho_o \, \tilde{\mathbf{F}}_{sgs}. \tag{6.66}$$

With this definition, the ensemble mean tracer transport term takes the form

$$\langle \rho \, C \, \mathbf{v} \rangle = \rho_o \, (\langle C \rangle^\rho \, \langle \tilde{\mathbf{v}} \rangle + \langle \tilde{\mathbf{F}}_{sgs} \rangle), \tag{6.67}$$

thus yielding the time tendency for the ensemble averaged tracer concentration in a non-Boussinesq fluid

$$\partial_t \, (\langle \rho \rangle \langle C \rangle^\rho) + \rho_o \, \nabla \cdot (\langle \tilde{\mathbf{v}} \rangle \langle C \rangle^\rho) = -\rho_o \, \nabla \cdot \langle \tilde{\mathbf{F}}_{sgs} \rangle + \langle \rho \rangle \langle \mathcal{S} \rangle^\rho. \tag{6.68}$$

The Boussinesq results are recovered by setting $\langle \rho \rangle \to \rho_o$

$$\partial_t \, \langle C \rangle^\rho + \nabla \cdot (\langle \tilde{\mathbf{v}} \rangle \langle C \rangle^\rho) = -\nabla \cdot \langle \tilde{\mathbf{F}}_{sgs} \rangle + \langle \mathcal{S} \rangle^\rho. \tag{6.69}$$

Note that we chose to keep the density weighted form of the averaged Boussinesq tracer equation (6.69). According to the analysis of McDougall et al. (2002a), doing so allows for a self-consistent set of equations compatible with the small levels of diapycnal mixing in the ocean interior. Also note that setting the averaged tracer concentration to unity, $\langle C \rangle^\rho = 1$, assuming SGS fluxes vanish and dropping source terms, reduces the tracer equation for both the non-Boussinesq and Boussinesq systems to their respective averaged continuity equations. Such *compatibility* between tracer and mass/volume conservation is important to maintain in the numerical model. If the sources are nontrivial when tracer is unity; i.e., if the sources introduce mass or volume, then compatibility requires sources to be part of the continuity equations as well.

6.5.4 Boussinesq limit in the averaged fluid

We highlight here an important formal point. As we have seen, the transition from averaged non-Boussinesq to averaged Boussinesq fluids arises by first averaging the non-Boussinesq equations and *then* setting $\langle \rho \rangle \to \rho_o$, except when multiplying gravity. Correspondingly, averaged Boussinesq fluids satisfy $\langle \tilde{\mathbf{v}} \rangle = \langle \mathbf{v} \rangle^\rho$, whereas $\langle \tilde{\mathbf{v}} \rangle \neq \langle \mathbf{v} \rangle$. Hence, this approach breaks the symmetry between $\tilde{\mathbf{v}}$ and $\mathbf{v}$ present in the unaveraged Boussinesq equations.

6.5.5 A not-so-useful alternative

The density multiplier in the time tendency term for the non-Boussinesq tracer equation (6.68) may be eliminated by introducing the new tracer variable

$$\rho_o \, \tilde{C} = \rho \, C. \tag{6.70}$$

This variable is directly analogous to the velocity variable $\rho \, \mathbf{v} = \rho_o \, \tilde{\mathbf{v}}$. The ensemble mean of the new tracer variable is given by

$$\rho_o \, \langle \tilde{C} \rangle = \langle \rho \rangle \langle C \rangle^{\rho}, \tag{6.71}$$

thus leading to the time tendency and advective portions of the tracer equation taking the form

$$\partial_t \left(\langle \rho \rangle \, \langle C \rangle^{\rho} \right) + \rho_o \, \nabla \cdot \left(\langle \tilde{\mathbf{v}} \rangle \, \langle C \rangle^{\rho} \right) = \langle \tilde{C} \rangle_{,t} + \nabla \cdot \left(\langle \mathbf{v} \rangle^{\rho} \langle \tilde{C} \rangle \right). \tag{6.72}$$

This form is mathematically similar to the Boussinesq tracer equation. However, this form does not manifest compatibility between tracer and mass conservation. Recall from the end of Section 6.5.2 that compatibility is manifest when setting $\langle C \rangle^{\rho} = 1$ in the tracer equation reduces the tracer budget of a parcel to its mass budget. If we instead naively set $\langle \tilde{C} \rangle = 1$ throughout the fluid, then equation (6.72) renders $\nabla \cdot \langle \mathbf{v} \rangle^{\rho} = 0$. However, $\nabla \cdot \langle \mathbf{v} \rangle^{\rho} = 0$ is not an expression of mass conservation. Rather, it is a solution generally realized only for the trivial case of uniform $\langle C \rangle^{\rho}$ and $\langle \rho \rangle$, as can be seen by the definition (6.71). Hence, for purposes of maintaining a manifest compatibility between tracer and mass conservation, it is more convenient to use $\langle C \rangle^{\rho}$ instead of $\langle \tilde{C} \rangle$ as the fundamental tracer variable.

6.6 Averaged momentum budget

When deriving an averaged momentum budget, the question arises whether we should work with the averaged form of the budget (6.3), written in terms of the linear momentum per mass $\mathbf{v}$, or the budget (6.9) written in terms of the linear momentum per volume $\rho_o \, \tilde{\mathbf{v}}$. As we have seen thus far, the linear momentum per volume $\rho_o \, \tilde{\mathbf{v}}$ is a very convenient transport variable for the averaged conservation equations of mass and tracer. Hence, we aim to derive an averaged momentum budget where time tendencies of $\langle \tilde{\mathbf{v}} \rangle$ can be determined. Since this is not the traditional method, we start by pointing out the form of the averaged momentum budget where the time tendency of $\langle \mathbf{v} \rangle$ is determined.

6.6.1 Time tendency for $\langle \mathbf{v} \rangle$

To isolate the time tendency of velocity, use mass conservation and divide by density in equation (6.3) to find

$$(\partial_t + \mathbf{v} \cdot \nabla + \mathcal{M} \, \hat{\mathbf{z}} \wedge) \mathbf{v} = -g \, \hat{\mathbf{z}} - f \, \hat{\mathbf{z}} \wedge \mathbf{v} - \nabla p / \rho + \mathbf{F}^{(\mathbf{v})}. \tag{6.73}$$

Taking an ensemble average leads to

$$(\partial_t + \langle \mathbf{v} \rangle \cdot \nabla + \langle \mathcal{M} \rangle \, \hat{\mathbf{z}} \wedge) \langle \mathbf{v} \rangle = -f \, \hat{\mathbf{z}} \wedge \langle \mathbf{v} \rangle - g \, \hat{\mathbf{z}} - \langle \nabla p / \rho \rangle + \langle \mathbf{F}^{(\mathbf{v})} \rangle. \tag{6.74}$$

To reach this expression, we assumed that nonlinear correlation terms arising from the advection terms are absorbed into the averaged friction vector

$$\langle \mathbf{F}^{(\mathbf{v})} \rangle = -\langle (\mathbf{v}' \cdot \nabla)\, \mathbf{v}' \rangle - \hat{\mathbf{z}} \wedge \langle \mathcal{M}'\, \mathbf{v}' \rangle, \tag{6.75}$$

where the molecular term is dropped since it is relatively small for averages over scales appropriate for ocean climate models. The pressure term can be written in terms of an average and correlations

$$\langle \nabla p/\rho \rangle \approx \frac{\nabla \langle p \rangle}{\langle \rho \rangle} + \frac{\langle \rho' \, \nabla p' \rangle}{\langle \rho \rangle^2} + \frac{\langle \rho' \, \rho' \rangle \, \nabla \langle p \rangle}{\langle \rho \rangle^3}, \tag{6.76}$$

where higher order correlations were dropped. The pressure-density turbulence correlation terms cannot be absorbed by the SGS friction vector $\langle \mathbf{F}^{(\mathbf{v})} \rangle$, since $\langle \mathbf{F}^{(\mathbf{v})} \rangle$ is meant to represent SGS transport terms, not pressure-density fluctuations. Hence, the pressure-density terms must be parameterized separately, where the parameterization depends on the dynamics at the SGS.

6.6.2 Time tendency for $\langle \tilde{\mathbf{v}} \rangle$

Now consider an average of the momentum equation (6.9) written in terms of the momentum density $\rho_o \, \tilde{\mathbf{v}}$ and using density weighted averages. As with the previous manipulations, averaging produces nonlinear correlations from the averaged nonlinear advection

$$\begin{aligned} \langle \rho \, \mathbf{v} \, \mathbf{v} \rangle &= \langle \rho \rangle \langle \mathbf{v} \rangle^{\rho} \langle \mathbf{v} \rangle^{\rho} + \langle \rho \, \mathbf{v}'_{\rho} \, \mathbf{v}'_{\rho} \rangle \\ &= \rho_o \, \langle \tilde{\mathbf{v}} \rangle \langle \mathbf{v} \rangle^{\rho} + \langle \rho \, \mathbf{v}'_{\rho} \, \mathbf{v}'_{\rho} \rangle \end{aligned} \tag{6.77}$$

and advection metric

$$\hat{\mathbf{z}} \wedge \langle \mathcal{M} \, \rho \, \mathbf{v} \rangle = \rho_o \, \hat{\mathbf{z}} \wedge \langle \tilde{\mathcal{M}} \rangle \langle \mathbf{v} \rangle^{\rho} + \hat{\mathbf{z}} \wedge \langle \rho \, \mathcal{M}'_{\rho} \, \mathbf{v}'_{\rho} \rangle, \tag{6.78}$$

where again $\rho_o \, \langle \tilde{\mathbf{v}} \rangle = \langle \rho \rangle \langle \mathbf{v} \rangle^{\rho}$.

To massage the SGS momentum terms, we proceed in a manner similar to that done in Section 6.5.3 for the SGS tracer fluxes. Thus, we introduce the SGS momentum flux

$$\begin{aligned} \rho \, \mathbf{F}^{(\mathbf{v})}_{sgs} &= \nabla \cdot (\rho \, \mathbf{v}'_{\rho} \, \mathbf{v}'_{\rho}) + \hat{\mathbf{z}} \wedge \rho \, \mathcal{M}'_{\rho} \, \mathbf{v}'_{\rho} \\ &\equiv \rho_o \, \tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs} \end{aligned} \tag{6.79}$$

and assume the much smaller molecular diffusion contribution to be absorbed by this flux. This form of the SGS momentum flux leads to the averaged non-Boussinesq momentum balance

$$\begin{aligned} &\langle \tilde{\mathbf{v}} \rangle_{,t} + \nabla \cdot (\langle \mathbf{v} \rangle^{\rho} \, \langle \tilde{\mathbf{v}} \rangle) + \langle \tilde{\mathcal{M}} \rangle \, \hat{\mathbf{z}} \wedge \langle \mathbf{v} \rangle^{\rho} \\ &\qquad = -(\langle \rho \rangle / \rho_o) \, g \, \hat{\mathbf{z}} - f \, \hat{\mathbf{z}} \wedge \langle \tilde{\mathbf{v}} \rangle - \nabla (\langle p \rangle / \rho_o) + \langle \tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs} \rangle. \end{aligned} \tag{6.80}$$

Notably, there is no need to parameterize pressure-density correlations appearing in the pressure gradient force $\langle \nabla p/\rho \rangle$, as needed with the equation for $\langle \mathbf{v} \rangle_{,t}$. We

consider this result yet another reason for using $\langle \tilde{\mathbf{v}} \rangle$ as our fundamental transport field.

The momentum equation for a Boussinesq fluid arises by setting $\langle \rho \rangle \to \rho_o$, except when multiplying gravity

$$\langle \tilde{\mathbf{v}} \rangle_{,t} + \nabla \cdot (\langle \tilde{\mathbf{v}} \rangle \langle \tilde{\mathbf{v}} \rangle) + \langle \tilde{\mathcal{M}} \rangle \, \hat{\mathbf{z}} \wedge \langle \tilde{\mathbf{v}} \rangle$$
$$= -(\langle \rho \rangle / \rho_o) \, g \, \hat{\mathbf{z}} - f \, \hat{\mathbf{z}} \wedge \langle \tilde{\mathbf{v}} \rangle - \nabla (\langle p \rangle / \rho_o) + \langle \tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs} \rangle. \quad (6.81)$$

The pressure gradient force is identical in form to that appearing in the non-Boussinesq equation for $\langle \tilde{\mathbf{v}} \rangle_{,t}$. Consistent with the averaged tracer equation, we interpret the averaged velocity field used in the Boussinesq momentum equation as $\langle \tilde{\mathbf{v}} \rangle$, which is distinct from $\langle \mathbf{v} \rangle$.

6.7 Summary of the averaged equations

We now summarize the averaged equations. First, mass conservation for a fluid parcel and for a vertical column of fluid is given by

$$\langle \rho \rangle_{,t} + \rho_o \, \nabla \cdot \langle \tilde{\mathbf{v}} \rangle = 0 \quad (6.82)$$
$$\left(D^* \, \overline{\langle \rho \rangle}^z \right)_{,t} = -\rho_o \, \nabla \cdot \langle \tilde{\mathbf{U}} \rangle + \rho_o \, \tilde{q}^*_w. \quad (6.83)$$

For the Boussinesq fluid, volume conservation for the parcel and column are given by

$$\nabla \cdot \langle \tilde{\mathbf{v}} \rangle = 0 \quad (6.84)$$
$$\eta^*_{,t} = -\nabla \cdot \langle \tilde{\mathbf{U}} \rangle + \tilde{q}^*_w. \quad (6.85)$$

The surface and bottom kinematic boundary conditions for the non-Boussinesq fluid are

$$\langle \rho \rangle \, \eta^*_{,t} + \rho_o \, \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* = \rho_o \, \langle \tilde{w} \rangle + \rho_o \, \tilde{q}^*_w \qquad \text{at } z = \eta^* \quad (6.86)$$
$$\langle \tilde{\mathbf{u}} \rangle \cdot \nabla H + \langle \tilde{w} \rangle = 0 \qquad \text{at } z = -H \quad (6.87)$$

and for the Boussinesq fluid we have

$$\eta^*_{,t} + \langle \tilde{\mathbf{u}} \rangle \cdot \nabla \eta^* = \langle \tilde{w} \rangle + \tilde{q}^*_w \qquad \text{at } z = \eta^* \quad (6.88)$$
$$\langle \tilde{\mathbf{u}} \rangle \cdot \nabla H + \langle \tilde{w} \rangle = 0 \qquad \text{at } z = -H. \quad (6.89)$$

The tracer budget for the non-Boussinesq fluid is

$$(\langle \rho \rangle \, \langle C \rangle^\rho)_{,t} + \rho_o \, \nabla \cdot (\langle \tilde{\mathbf{v}} \rangle \, \langle C \rangle^\rho) = -\rho_o \, \nabla \cdot \langle \tilde{\mathbf{F}}_{sgs} \rangle + \langle \rho \rangle \langle \mathcal{S} \rangle^\rho \quad (6.90)$$

and for the non-Boussinesq fluid

$$\partial_t \, \langle C \rangle^\rho + \nabla \cdot (\langle \tilde{\mathbf{v}} \rangle \langle C \rangle^\rho) = -\nabla \cdot \langle \tilde{\mathbf{F}}_{sgs} \rangle + \langle \mathcal{S} \rangle^\rho. \quad (6.91)$$

Finally, the non-Boussinesq momentum budget is

$$\langle \tilde{\mathbf{v}} \rangle_{,t} + \nabla \cdot (\langle \mathbf{v} \rangle^\rho \, \langle \tilde{\mathbf{v}} \rangle) + \langle \tilde{\mathcal{M}} \rangle \, \hat{\mathbf{z}} \wedge \langle \mathbf{v} \rangle^\rho$$
$$= -(\langle \rho \rangle / \rho_o) \, g \, \hat{\mathbf{z}} - f \, \hat{\mathbf{z}} \wedge \langle \tilde{\mathbf{v}} \rangle - \nabla (\langle p \rangle / \rho_o) + \langle \tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs} \rangle \quad (6.92)$$

whereas the Boussinesq budget is

$$\langle\tilde{\mathbf{v}}\rangle_{,t} + \nabla\cdot(\langle\tilde{\mathbf{v}}\rangle\langle\tilde{\mathbf{v}}\rangle) + \langle\tilde{\mathcal{M}}\rangle\,\hat{\mathbf{z}}\wedge\langle\tilde{\mathbf{v}}\rangle = \\ -(\langle\rho\rangle/\rho_o)\,g\,\hat{\mathbf{z}} - f\,\hat{\mathbf{z}}\wedge\langle\tilde{\mathbf{v}}\rangle - \nabla(\langle p\rangle/\rho_o) + \langle\tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs}\rangle. \quad (6.93)$$

Recall that $\langle\tilde{\mathbf{v}}\rangle = \langle\mathbf{v}\rangle^{\rho}$ for a Boussinesq fluid, whereas $\rho_o\,\langle\tilde{\mathbf{v}}\rangle = \langle\rho\rangle\,\langle\mathbf{v}\rangle^{\rho}$ for the non-Boussinesq case. In the tracer equation, the SGS flux term is given by

$$\begin{aligned}\rho\,\mathbf{F}_{sgs} &= \rho\,C'_{\rho}\,\mathbf{v}'_{\rho}\\ &\equiv \rho_o\,\tilde{\mathbf{F}}_{sgs}\end{aligned} \quad (6.94)$$

which dominates the molecular diffusion flux. The friction vector likewise incorporates SGS turbulence terms

$$\begin{aligned}\rho\,\mathbf{F}^{(\mathbf{v})}_{sgs} &= \nabla\cdot(\rho\,\mathbf{v}'_{\rho}\,\mathbf{v}'_{\rho}) + \hat{\mathbf{z}}\wedge\rho\,\mathcal{M}'_{\rho}\,\mathbf{v}'_{\rho}\\ &\equiv \rho_o\,\tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs}.\end{aligned} \quad (6.95)$$

The averaged equations have the same mathematical form as the unaveraged equations given in Section 6.2. Precisely, the mapping between unaveraged and averaged fields is given by

$$\rho \to \langle\rho\rangle \quad (6.96)$$
$$p \to \langle p\rangle \quad (6.97)$$
$$\mathbf{v} \to \langle\mathbf{v}\rangle^{\rho} \quad (6.98)$$
$$\tilde{\mathbf{v}} \to \langle\tilde{\mathbf{v}}\rangle \quad (6.99)$$
$$C \to \langle C\rangle^{\rho} \quad (6.100)$$
$$\tilde{\mathbf{F}} \to \langle\tilde{\mathbf{F}}_{sgs}\rangle \quad (6.101)$$
$$\mathcal{S} \to \langle\mathcal{S}\rangle^{\rho} \quad (6.102)$$
$$\tilde{\mathbf{F}}^{(\mathbf{v})} \to \langle\tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs}\rangle \quad (6.103)$$
$$\eta \to \eta^{*} \quad (6.104)$$
$$\tilde{q}_w \to \tilde{q}^{*}_w. \quad (6.105)$$

This mapping is very useful for purposes of extending properties of the unaveraged system to the averaged system, such as the energetic balances discussed in Chapter 4.

6.8 Mapping to ocean model variables

Having established a set of self-consistent averaged equations, we are ready to write down the equations to be discretized in the numerical model. As discussed in Section 6.1.2, discretization is applied to the averaged continuous equations, which are written down in this section in a bit more tidy manner than in the previous section.

First, the density variable to be discretized by the ocean model is the Eulerian mean density

$$\langle\rho\rangle \to \rho_{\text{model}}. \quad (6.106)$$

Again, this is the *in situ* density used for the mass continuity equation. It is generally distinct from potential density described in Section 4.8. At this point the "model" suffix refers to the continuous ocean model, since no discretization has yet occured. Through the hydrostatic approximation, $\langle \rho \rangle \to \rho_{\text{model}}$ leads to

$$\langle p \rangle \to p_{\text{model}}. \tag{6.107}$$

As argued in Section 6.3, maintaining a tidy form of the averaged continuity equation motivates our discretizing $\langle \tilde{\mathbf{v}} \rangle$, instead of the conventional $\langle \mathbf{v} \rangle$. The distinction is nontrivial for both non-Boussinesq and Boussinesq ocean models. Hence, we make the correspondence

$$\langle \tilde{\mathbf{v}} \rangle \to \mathbf{v}_{\text{model}}. \tag{6.108}$$

Analogously, we choose to use $\langle C \rangle^{\rho}$ in our model, instead of the conventional $\langle C \rangle$

$$\langle C \rangle^{\rho} \to C_{\text{model}}. \tag{6.109}$$

Again, the distinction between $\langle C \rangle^{\rho}$ and $\langle C \rangle$ is nontrivial for both non-Boussinesq and Boussinesq ocean models. Finally, the surface height in the ocean model corresponds to the modified mean surface height η^* described in Section 6.4.2

$$\eta^* \to \eta_{\text{model}}, \tag{6.110}$$

as will the modified mean surface fresh water flux

$$\tilde{q}_w^* \to (q_w)_{\text{model}}. \tag{6.111}$$

These mappings lead to the following equations to be discretized in the non-Boussinesq ocean model (dropping the "model" subscript for brevity)

$$\rho_{,t} + \rho_o \, \nabla \cdot \mathbf{v} = 0 \tag{6.112}$$

$$(D \, \overline{\rho}^z)_{,t} = -\rho_o \, \nabla \cdot \mathbf{U} + \rho_o \, q_w \tag{6.113}$$

$$\begin{aligned} \mathbf{v}_{,t} + \nabla \cdot [(\rho_o/\rho) \, \mathbf{v} \, \mathbf{v}] + (\rho_o/\rho) \, \mathcal{M} \, \hat{\mathbf{z}} \wedge \mathbf{v} = \\ - (\rho/\rho_o) \, g \, \hat{\mathbf{z}} - f \, \hat{\mathbf{z}} \wedge \mathbf{v} - \nabla (p/\rho_o) + \mathbf{F}^{(\mathbf{v})} \end{aligned} \tag{6.114}$$

$$(\rho \, C)_{,t} + \rho_o \, \nabla \cdot (\mathbf{v} \, C) = -\rho_o \, \nabla \cdot \mathbf{F} + \rho \, \mathcal{S}, \tag{6.115}$$

where we also made the correspondence

$$\langle \tilde{\mathbf{F}}_{sgs} \rangle \to \mathbf{F}_{\text{model}} \tag{6.116}$$

$$\langle \mathcal{S} \rangle^{\rho} \to \mathcal{S}_{\text{model}} \tag{6.117}$$

$$\langle \tilde{\mathbf{F}}_{sgs}^{(\mathbf{v})} \rangle \to \mathbf{F}_{\text{model}}^{(\mathbf{v})}. \tag{6.118}$$

Surface and bottom kinematic boundary conditions are set according to

$$\rho \, \eta_{,t} + \rho_o \, \mathbf{u} \cdot \nabla \eta = \rho_o \, w + \rho_o \, q_w \qquad \text{at } z = \eta \tag{6.119}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H \tag{6.120}$$

The mapping from unaveraged to averaged fields, and then from averaged to model fields, is summarized in Table 6.1. This table is the key result from this chapter.

Unaveraged	Averaged	Model	Model Discrete
ρ	$\langle\rho\rangle$	ρ_{model}	ρ
p	$\langle p\rangle$	p_{model}	p
$\tilde{\mathbf{v}}$	$\langle\tilde{\mathbf{v}}\rangle$	$\mathbf{v}_{\text{model}}$	$\mathbf{v}$
$\mathbf{v}$	$\langle\mathbf{v}\rangle^{\rho}$	$\rho_{\text{model}}\,\mathbf{v}^{\rho}_{\text{model}} = \rho_o\,\mathbf{v}_{\text{model}}$	$\rho\,\mathbf{v}^{\rho} = \rho_o\,\mathbf{v}$
η	η^*	η_{model}	η
$\tilde{q}_w$	$\tilde{q}^*_w$	$(q_w)_{\text{model}}$	q_w
C	$\langle C\rangle^{\rho}$	C_{model}	C
$\mathcal{S}$	$\langle\mathcal{S}\rangle^{\rho}$	$\mathcal{S}_{\text{model}}$	$\mathcal{S}$
$\tilde{\mathbf{F}}$	$\langle\tilde{\mathbf{F}}_{sgs}\rangle$	$\mathbf{F}_{\text{model}}$	$\mathbf{F}$
$\tilde{\mathbf{F}}^{(\mathbf{v})}$	$\langle\tilde{\mathbf{F}}^{(\mathbf{v})}_{sgs}\rangle$	$\mathbf{F}^{(\mathbf{v})}_{\text{model}}$	$\mathbf{F}^{(\mathbf{v})}$

Table 6.1: Correspondence between unaveraged continuous fields, ensemble averaged continuous fields, continuous model fields, and discrete model fields. This table will be referred to throughout this book.

The continuous model equations presented above are *identical* in form to the continuous unaveraged non-Boussinesq equations summarized in Section 6.2. Although in the end somewhat trivial (i.e., what a round-about way to get back to the same equations!), the intermediate steps reveal a nontrivial interpretation of the fields discretized in a z-coordinate ocean model.

The Boussinesq model equations arise by setting $\rho_{\text{model}} \to \rho_o$, except when multiplying gravity

$$\nabla\cdot\mathbf{v} = 0 \tag{6.121}$$

$$\eta_{,t} = -\nabla\cdot\mathbf{U} + q_w \tag{6.122}$$

$$\mathbf{v}_{,t} + \nabla\cdot(\mathbf{v}\,\mathbf{v}) + \mathcal{M}\,\hat{\mathbf{z}}\wedge\mathbf{v} = -(\rho/\rho_o)\,g\,\hat{\mathbf{z}} - f\,\hat{\mathbf{z}}\wedge\mathbf{v} - \nabla(p/\rho_o) + \mathbf{F}^{(\mathbf{v})} \tag{6.123}$$

$$C_{,t} + \nabla\cdot(\mathbf{v}\,C) = -\nabla\cdot\mathbf{F} + \mathcal{S} \tag{6.124}$$

with the surface and bottom kinematic boundary conditions

$$\eta_{,t} + \mathbf{u}\cdot\nabla\eta = w + q_w \qquad \text{at } z = \eta \tag{6.125}$$

$$\mathbf{u}\cdot\nabla H + w = 0 \qquad \text{at } z = -H. \tag{6.126}$$

As emphasized by McDougall et al. (2002a) and Greatbatch et al. (2001), upon making the hydrostatic approximation, these equations for the Boussinesq ocean model are identical to those integrated by traditional Boussinesq z-coordinate ocean models, with the exception of details that have been absorbed by the turbulence tracer and momentum fluxes $\mathbf{F}$ and $\mathbf{F}^{(\mathbf{v})}$. Additionally, McDougall et al. (2002a) argue that the interpretation of model fields as proposed here allows for the Boussinesq equations to be far more accurate than the alternative interpretation. Hence, for this reason, and for reasons of mathematical elegance, we prefer the interpretation summarized by Table 6.1 for the variables carried by the Boussinesq and non-Boussinesq versions of z-coordinate ocean climate models.

6.9 Discretization basics

We have spent the bulk of this chapter motivating an interpretation of the continuous equations discretely represented by a z-coordinate ocean model. It is now appropriate to consider how to discretize the continuous model equations. This book is short on pedagogical discussions of general discretization techniques, since the computational fluids literature has many examples (e.g., Haltiner and Williams (1980) and Durran (1999)). However, for completeness, we introduce here some of the issues, preserving more detail for later chapters.

The discretization of continuous equations often starts by integrating these equations over the finite space-time lattice onto which the equations are to be numerically integrated. For example, introduce a space-time averaged field via the averaging operator

$$\overline{\psi}^{\mathbf{x},t} = \frac{1}{\Delta V}\frac{1}{\Delta t}\int_{-\Delta V/2}^{\Delta V/2} \mathrm{d}\mathbf{x} \int_{-\Delta t/2}^{\Delta t/2} \mathrm{d}t\, \psi(x,y,z,t), \tag{6.127}$$

where $\Delta V = \Delta x\, \Delta y\, \Delta z$ is the volume of the lattice grid cell, and $\mathrm{d}\mathbf{x} = \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$ is the infinitesimal volume element. If the cell volume is constant in space and time, then such averaging leads to the discrete expression for a time derivative operator

$$\overline{\psi_{,t}}^{\mathbf{x},t} = \frac{\overline{\psi}^{\mathbf{x}}(t+\Delta t/2) - \overline{\psi}^{\mathbf{x}}(t-\Delta t/2)}{\Delta t}, \tag{6.128}$$

and similar expressions for spatial derivatives. Other averaging operators exist whereby a kernal is used to weight the averaged field, with the present example employing a "box-car" kernal. In either approach, the numerical model fields are interpreted as discrete representations of continuous averages of the continuous fields. This is the *finite-volume* approach. Another approach interprets the numerical model fields as point-wise discrete representations of the continuous fields, thus prompting Taylor series methods for approximating derivative operators.

Most z-coordinate models do not strictly adhere to either one interpretation. That is, there is no attempt to provide a uniform method to transfer the continuous equations to the lattice. Doing so would restrict the freedom of choosing various discretizations of operators. However, such ambiguity in transferring the ensemble averaged continuous equations to the lattice *does not* introduce ambiguity into how to interpret the fundamental continuous model equations. The interpretation proposed in this chapter, and summarized in Table 6.1, remains basic to all discretization methods used in z-coordinate ocean models.

In general, discretization breaks the symmetries maintained by the continuous partial differential equations. Consequently, the discrete equations can have solutions whose qualitative and/or quantitative behaviour is not reflective of the continuous equations. It is the task of the numerical algorithm developer to ensure that any mis-representation of the continuous system is of negligible physical consequence. Otherwise, the numerical model is of little use for physical modeling.

CHAPTER
SEVEN

Kinematics of an isentropic ensemble

Contents

7.1 The mesoscale eddy parameterization problem 164
7.1.1 Posing the parameterization problem 164
7.1.2 Averaging at fixed depth versus at fixed neutral density . 165
7.1.3 Good modeling practice 165
7.2 Advection and skewsion 166
7.2.1 The vector streamfunction 166
7.2.2 Advective and skew fluxes 166
7.2.3 Complementary aspects of the fluxes 167
7.3 Volume conservation in an isentropic ensemble 168
7.3.1 Two perspectives on volume conservation 168
7.3.2 Exact ensemble mean kinematics in isopycnal coordinates 169
7.3.3 The non-utility of an Eulerian mean description 170
7.3.4 Transformed residual mean (TRM) 171
7.3.5 Exact ensemble mean kinematics in z-coordinates 172
7.3.5.1 The total transport and the quasi-Stokes transport 172
7.3.5.2 Evolution of modified mean density 174
7.3.6 Approximate ensemble mean kinematics in z-coordinates 175
7.4 Ensemble mean tracer equation 177
7.4.1 Thickness weighted means 177
7.4.2 Mean thickness weighted tracer equation 177
7.4.3 Tracer mixing tensor 178
7.4.4 Mean tracer transport beneath a density surface 178
7.4.5 Summary of the mean field tracer equation 179
7.5 Quasi-Stokes transport in z-models 180
7.5.1 Rotational and divergent components 180
7.5.1.1 Flux divergence 181
7.5.1.2 Zonal symmetry and zonal means 181
7.5.1.3 The rotational component 182
7.5.2 GM90 skew-diffusion 183

7.5.3 Effective meridional volume transport 184
7.5.4 GM90 and diffusion of isopycnal thickness 184

This chapter presents a set of kinematic tools and applies them to an ensemble of isentropic Boussinesq fluid parcels. This ensemble is useful when formulating the mesoscale eddy problem in the ocean interior, since these eddies are well approximated by ideal fluid flow. Much of the material in this chapter requires patience and persistence from the reader. It certainly required such from the author.

For the purposes of this chapter, isentropic transport of a parcel is frictionless, adiabatic, and isohaline. Hence, we do not consider the special case of heating that compensates salt diffusion so that a fluid parcel's entropy (or potential density) remains constant.

7.1 The mesoscale eddy parameterization problem

Understanding and parameterizing mesoscale eddies in the ocean remains a topic of intense research within the physical oceanography community. The space scale characterizing the mesoscale eddies is largely determined by the first baroclinic Rossby radius (e.g., Stammer (1997a), Smith and Vallis (2002)), and the time scale is on the order of a few weeks to months. The space scale is smaller than the level of grid resolution in typical global ocean climate models, thus necessitating parameterizations aiming to capture eddy effects on the resolved scale flow. This parameterization problem is important due to the nontrivial effects that eddies have on setting tracer distributions and water mass properties, especially within the Southern Ocean. Additionally, and quite uncomfortably, ocean climate models exhibit a high degree of sensitivity to their sub-grid-scale operators. For these reasons, many research efforts have over the past two decades focused on improving the mathematical, physical, and numerical integrity of the sub-grid scale operators aiming to parameterize the eddies.

7.1.1 Posing the parameterization problem

The mesoscale eddy problem has typically been posed in the context of ideal Boussinesq fluid dynamics. We follow this route here. Notable further simplifications to the dynamical equations are rendered via the quasi-geostrophic approximation, and such has proven useful for understanding eddy mechanisms and scalings. Given our focus on ideal flow, we are not concerned with processes that alter the entropy of a fluid parcel, such as those associated with the convergence of mixing fluxes, or sources due to nonlinear effects in the equation of state for seawater density. Generalizations are straightforward, as discussed by McIntosh and McDougall (1996) and McDougall and McIntosh (2001), whereby potential density and its associated specific thickness is translated into locally referenced potential density, or neutral density, and its associated specific thickness.

The parameterization problem involves many considerations, such as numerical constraints (e.g., dissipation operators are needed for numerical stability), physical considerations (e.g., mesoscale eddies mix predominantly along neutral direc-

tions), physical postulates (e.g., eddies homogenize potential vorticity), and statistical postulates (e.g., eddies maximize entropy). Different theoretical approaches have been tried with little consensus in the community as to what is optimal both theoretically and from the perspective of "good" modeling practice, largely defined as a practice that leads to unambiguous inprovement in model realism. Griffies et al. (2000a) provides a review with numerous references.

In this chapter, our purpose is modest in the sense that we do not attempt to establish a dynamical theory leading to a parameterization scheme. Instead, we concern ourselves only with volume and tracer budgets and focus on basic kinematic issues important for interpreting simulations. These considerations provide a starting point for framing the mesoscale eddy parameterization problem, especially when posed in z-coordinate models, yet they are not sufficient. For more complete treatments, refer to some of the many references given in Griffies et al. (2000a).

7.1.2 Averaging at fixed depth versus at fixed neutral density

Chapter 6 provides a method to systematically map between ocean model variables and *in situ* density weighted ensemble averaged fields (e.g., Section 6.8). The ensemble average from that chapter takes place at a fixed point in space, and in particular at a fixed depth.

A basic property of mesoscale eddies is that they stir and mix properties predominantly parallel to neutral directions. It is important to build into the mathematical formulation of the mesoscale eddy closure problem this basic empirical fact. Consequently, when averaging over realizations of mesoscale eddies for the purpose of garnering a parameterization, we are prompted to situate ourselves at a horizontal position that maintains a constant locally referenced potential density. Doing so leads to specific thickness weighted means, as discussed later in this chapter.

We therefore find it convenient to distinguish the two forms of ensemble averaging: this chapter focuses on averaging over the mesoscale whereas Chapter 6 averages over smaller scales. An alternative is to average over the mesoscale and small scales at once, thus combining the *in situ* density weighted averaging of Chapter 6 with the specific thickness weighted means of this chapter. Greatbatch and McDougall (2002) pursue that approach, which turns out to be a straightforward exercise in the techniques introduced here and in Chapter 6.

7.1.3 Good modeling practice

A practical outcome of this chapter is a kinematic interpretation for sub-grid scale transport operators in the tracer equation integrated by models not resolving mesoscale eddies. In Chapter 11, it is argued that the neutral transport operators, motivated from eddy closure ideas, are useful even for regimes where the mesoscale eddies are explicitly resolved. The main reason is that alternative transport operators that may be appropriate at fine resolution tend to introduce unphysically large levels of spurious mixing between water masses (Roberts and Marshall (1998) and Griffies et al. (2000b)). Hence, from this perspective, the operators motivated from the mesoscale

parameterization problem represent *good modeling practice* for all resolutions. We further discuss these *neutral transport operators* in Chapters 11–13 and 14.

7.2 Advection and skewsion

There are two complementary ways of interpreting the transport operator used to reversibly stir a tracer. One can either consider the convergence of an advective flux, thus leading to *advection*, or the convergence of a skew-flux, thus leading to *skewsion*. Our focus in this chapter is on skewsion as it lends itself to a more streamlined theoretical development, and it provides a more robust numerical foundation (see Chapters 11–13 and 14 for numerical developments).

7.2.1 The vector streamfunction

Consider an arbitrary three-dimensional divergence-free velocity field

$$\nabla \cdot \mathcal{V} = 0. \tag{7.1}$$

The divergence-free condition represents a diagnostic relation, or constraint, that reduces the functional degrees of freedom for the velocity field from three to two. This constraint can be satisfied identically by introducing a vector streamfunction

$$\mathcal{V} = \nabla \wedge \boldsymbol{\Upsilon}. \tag{7.2}$$

The vector streamfunction $\boldsymbol{\Upsilon}$ is not completely specified by this relation, since the equally valid streamfunction

$$\boldsymbol{\Upsilon}' = \boldsymbol{\Upsilon} + \nabla\lambda \tag{7.3}$$

corresponds to the same velocity field $\mathcal{V}$. The arbitrary scalar function λ is known as a *gauge function*, and the freedom to modify the vector streamfunction through the addition of λ is known as *gauge freedom*. A similar symmetry is present in Maxwell's equations of electrodynamics (e.g., Jackson (1975)).

7.2.2 Advective and skew fluxes

The *advective tracer flux*

$$\mathbf{F}_{(a)} = \mathcal{V}\, C = (\nabla \wedge \boldsymbol{\Upsilon})\, C \tag{7.4}$$

can be related to a skew flux

$$\mathbf{F}_{(s)} = -\nabla C \wedge \boldsymbol{\Upsilon} = \mathbf{F}_{(a)} - \mathbf{F}_{(r)} \tag{7.5}$$

through exploitation of the identity

$$\mathcal{V}\, C = (\nabla \wedge \boldsymbol{\Upsilon})\, C = -\nabla C \wedge \boldsymbol{\Upsilon} + \nabla \wedge (C\, \boldsymbol{\Upsilon}). \tag{7.6}$$

$\mathbf{F}_{(s)}$ represents a *skew* flux, which is a flux defined such that it is directed perpendicular to the tracer gradient $\mathbf{F}_{(s)} \cdot \nabla C = 0$. Since the *rotational flux*

$$\mathbf{F}_{(r)} = \nabla \wedge (C\, \boldsymbol{\Upsilon}) \tag{7.7}$$

has a vanishing divergence, the divergence of the skew and advective fluxes is identical

$$\nabla \cdot \mathbf{F}_{(s)} = \nabla \cdot \mathbf{F}_{(a)}. \tag{7.8}$$

Hence, if these fluxes enter into the evolution of a tracer, one may choose to use the skew flux or advective flux in describing the evolution. That is,

$$\begin{aligned} C_{,t} + \nabla \cdot (\mathbf{v}\, C) &= -\nabla \cdot (\mathcal{V}\, C) \\ &= \nabla \cdot (\nabla C \wedge \boldsymbol{\Upsilon}). \end{aligned} \tag{7.9}$$

It is often convenient to introduce the anti-symmetric transport tensor

$$F^m_{(s)} = -\epsilon^{mnp}\, C_{,n}\, \Upsilon_p = -A^{mn}\, C_{,n} \tag{7.10}$$

where A^{mn} represents a reorganization of the vector streamfunction

$$A^{mn} = \epsilon^{mnp}\, \Upsilon_p = \begin{pmatrix} 0 & \Upsilon_3 & -\Upsilon_2 \\ -\Upsilon_3 & 0 & \Upsilon_1 \\ \Upsilon_2 & -\Upsilon_1 & 0 \end{pmatrix}, \tag{7.11}$$

and ϵ^{mnp} is the totally anti-symmetric Levi-Civita tensor defined in Section 18.12. Given the gauge freedom introduced in Section 7.2.1, we can always choose to specify a gauge so that

$$\Upsilon_3 \equiv 0. \tag{7.12}$$

This gauge is referred to as the *vertical* gauge and will be used in the following.

7.2.3 Complementary aspects of the fluxes

We consider here two aspects of the advective and skew tracer fluxes which turn out to be complementary to one another. First, since the advective flux is directly proportional to the velocity, this flux vanishes when the velocity vanishes. In contrast, a skew flux vanishes when the tracer gradient vanishes, as occurs for diffusive tracer fluxes.

Second, the orientation of the advective flux $\mathcal{V}\, C$ is determined by the velocity field, which is oriented according to the trajectories of fluid parcels. This orientation is the same for all tracers. The fluid parcels carry along with them a particular amount of tracer mass $(\rho\, \mathrm{d}V)\, C$, hence leading to the usual notion of a flux representing the passage of a quantity of matter across an area per unit time. Furthermore, it is the component of the velocity (and hence the advective tracer flux) aligned parallel to gradients of the tracer that creates local changes in the tracer concentration. We see this property from the advection equation

$$C_{,t} = -\mathcal{V} \cdot \nabla C. \tag{7.13}$$

Hence, an advective flux of tracer in an adiabatic flow *is not* generally oriented parallel to surfaces of constant tracer. Indeed, if it were, then the tracer field would be locally static since then $\mathcal{V} \cdot \nabla C = 0$. This statement holds for all tracers, including potential density.

As a complement to the advective flux, the skew tracer flux is directed along lines of constant tracer. That is, the skew flux is neither up-gradient nor down-gradient since

$$\nabla C \cdot \mathbf{F}_{(s)} = -\nabla C \cdot (\nabla C \wedge \boldsymbol{\Upsilon}) = 0. \tag{7.14}$$

Hence, orientation of the skew flux is directly tied to the tracer field, with each tracer yielding a generally different orientation, all parallel to the respective tracer isolines. This property does not allow for the interpretation of a skew flux in terms of the passage of a quantity of matter across an area per unit time.

7.3 Volume conservation in an isentropic ensemble

We now consider the reversible stirring of volume in an ideal stratified Boussinesq fluid. Stirring by ocean baroclinic eddies provides a good approximation to such ideal situations. Over long space and time scales, mesoscale eddy stirring can be considered chaotic, which motivates a stochastic perspective in which an ensemble of eddies is considered. The goal is to describe the ensemble mean properties of the fluid, with this section focusing on the kinematics of parcel rearrangement.

In the following, we use the overbar with a density label, $\overline{(\ )}^{\rho}$, to denote the ensemble mean over an ensemble of fluid parcels, each having the same potential density ρ, the same horizontal position (x, y), and the same time t. Since the isopycnals are undulating, each ensemble member has a vertical position that is generally distinct from the fixed depth z. When the context is clear, we often drop the (x, y, t) dependence in order to highlight the dependence on potential density and/or vertical position. Averaging over parcels maintaining the same potential density ρ is often termed *isopycnal averaging* in the literature.

7.3.1 Two perspectives on volume conservation

Consider a stably stratified ideal Boussinesq fluid. Isopycnal coordinates (x, y, ρ, t) introduced in Section 5.9.4 are ideally suited to describe the dynamics of this fluid. In these coordinates, the volume of a fluid parcel is written

$$\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \mathrm{d}x\,\mathrm{d}y\,z_{,\rho}\,\mathrm{d}\rho. \tag{7.15}$$

Since the specific thickness $z_{,\rho}$ appears quite frequently in the following, it is conventional to introduce the shorthand notation

$$h = z_{,\rho}. \tag{7.16}$$

As discussed in Section 5.3.2, specific thickness is the Jacobian of transformation between geo-potential (x, y, z, t) and isopycnal (x, y, ρ, t) coordinates. For stably stratified fluids, h is one-signed, hence making the coordinate transformation well defined. Geometrically, the product $h\,\mathrm{d}\rho$ represents the vertical distance, or *thickness*, between the two infinitesimally close density classes ρ and $\rho + \mathrm{d}\rho$. For an ideal Boussinesq fluid parcel, material conservation of volume and potential density implies conservation of the product of specific thickness and horizontal area $\mathrm{d}x\,\mathrm{d}y\,h$, which leads to the thickness equation

$$h_{,t} + \nabla_\rho \cdot (h\,\mathbf{u}) = 0, \tag{7.17}$$

with $\mathbf{u}$ the horizontal velocity field.

An Eulerian z-coordinate description of the reversible rearrangement of Boussinesq parcels is rendered via a combination of volume conservation $\nabla \cdot \mathbf{v} = 0$, and material conservation of potential density $D\rho/Dt = 0$. When written as a skewsion process, the natural gauge is the vertical gauge $\mathbf{\Psi} = \mathbf{U} \wedge \hat{\mathbf{z}}$ introduced in Section 7.2.2 since this gauge only requires the same horizontal velocity field $\mathbf{u}$ used with the isopycnal coordinate description. This gauge has an associated skew flux of potential density $\mathbf{F} = -\nabla\rho \wedge \mathbf{\Psi}$, which leads to the evolution

$$\rho_{,t} = \nabla \cdot (\nabla\rho \wedge \mathbf{\Psi}), \tag{7.18}$$

where all derivatives are here taken with fixed Eulerian coordinates, and the divergence operator is three-dimensional.

As an aside, we note that Kushner and Held (1999) considered the case with potential vorticity replacing density and meridional replacing vertical stratification. As pointed out by Held and Schneider (1999), such a description might also be relevant for a surface layer of density where vertical density stratification can be weak yet meridional stratification nonzero. In this case, volume conservation using isopycnal coordinates is described by the alternate form of the thickness equation

$$\partial_t y_{,\rho} + \partial_x \left(y_{,\rho}\, u \right) + \partial_z \left(y_{,\rho}\, w \right) = 0, \tag{7.19}$$

where the partial derivatives are still taken along surfaces of constant density. Because the thickness in this case is only affected by the zonal and vertical velocities, the natural gauge for a skewsion description is a *meridional* gauge.

7.3.2 Exact ensemble mean kinematics in isopycnal coordinates

Consider an ensemble of ideal Boussinesq fluid parcels with the same infinitesimal volume $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,h\,\mathrm{d}\rho$, and same potential density ρ. Lacking any other marker, such as a tracer concentration, the ensemble members are distinguished from one another by values of their horizontal area $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$, and their specific thickness h; that is, their geometric attributes. The ensemble members are assumed to be stirred by different stochastic realizations of the fluid flow. Since each flow realization alters the geometric properties of the parcels, a mean-field description focuses on the mean of these geometric properties.

In isopycnal coordinates (x, y, ρ, t), the thickness equation is satisfied by each ensemble member

$$h_{,t} + \nabla_\rho \cdot (h\,\mathbf{u}) = 0. \tag{7.20}$$

The ensemble mean over parcels with potential density ρ satisfies

$$\partial_t\, \overline{h}^\rho + \nabla_\rho \cdot \left(\overline{h}^\rho\, \overline{\mathbf{u}}^\rho + \overline{h'\,\mathbf{u}'}^\rho \right) = 0, \tag{7.21}$$

where primed variables represent deviations from the mean. The mean specific thickness $\overline{h}^\rho$ of parcels with potential density ρ therefore satisfies the conservation equation

$$\partial_t\, \overline{h}^\rho + \nabla_\rho \cdot (\overline{h}^\rho\, \widehat{\mathbf{u}}) = 0, \tag{7.22}$$

where

$$\begin{aligned}\widehat{\mathbf{u}} &= \frac{\overline{h\,\mathbf{u}}^{\rho}}{\overline{h}^{\rho}} \\ &= \overline{\mathbf{u}}^{\rho} + \frac{\overline{h'\,\mathbf{u}'}^{\rho}}{\overline{h}^{\rho}} \\ &= \overline{\mathbf{u}}^{\rho} + \mathbf{u}^{b}\end{aligned} \tag{7.23}$$

is an effective horizontal velocity. The first equality defines a *thickness weighted* averaging operator, and the last equality defines the horizontal *bolus velocity* $\mathbf{u}^b$ as introduced by Rhines (1982). Since each ensemble member is taken from a stably stratified fluid, the mean specific thickness $\overline{h}^{\rho}$ is single signed and non-vanishing. The bolus velocity

$$\mathbf{u}^b(\rho) = \widehat{\mathbf{u}}(\rho) - \overline{\mathbf{u}}^{\rho}(\rho) \tag{7.24}$$

arises from the along-isopycnal correlations between thickness and horizontal velocity.

Quite conveniently, the mean conservation equation (7.22) takes the *same* form as the conservation equation (7.20) satisfied by each ensemble member. The key difference, of course, is that the ensemble mean thickness $\overline{h}^{\rho}$ is stirred by the effective horizontal velocity $\widehat{\mathbf{u}}(\rho)$, whereas the thickness of each ensemble member is stirred by a randomly different realization of the horizontal velocity $\mathbf{u}$. As emphasized by Kushner and Held (1999), the simplicity of the resulting mean-field description is afforded by use of the Lagrangian vertical coordinate ρ.

7.3.3 The non-utility of an Eulerian mean description

Although ρ-coordinates provide a natural description of the isentropic rearrangement of volume, it is useful to understand how to transform to z-coordinates for purposes such as interpreting results from z-coordinate ocean models.

Let us first illustrate why it is more useful to employ fixed potential density means (which are Lagrangian in the vertical) instead of fixed depth means (which are Eulerian). For this purpose, it is sufficient to note that when sitting at a fixed Eulerian grid point, one measures a mean skew flux $\overline{\mathbf{F}}^{z}$ that has a nonzero component across the mean potential density surface $\overline{\rho}^{z}$. Here, $\overline{(\ )}^{z}$ denotes an Eulerian average obtained by an observer at a fixed point (x, y, z, t). Precisely, although

$$\mathbf{F}(z) \cdot \nabla\rho(z) = 0 \tag{7.25}$$

for each realization of the isentropic flow, the Eulerian average satisfies

$$\overline{\mathbf{F}}^{z} \cdot \nabla\overline{\rho}^{z} = -\overline{\mathbf{F}' \cdot \nabla\rho'}^{z} \neq 0. \tag{7.26}$$

Herein lies the cumbersome nature of the Eulerian perspective: the behaviour of the Eulerian mean skew flux differs fundamentally from that of each flow realization. This behaviour of the ensemble mean is in contrast to what we just saw from the isopycnal description, where the ensemble mean thickness equation has the same mathematical form as the thickness equation satisfied by each ensemble member. The problem, of course, is that an average taken at a fixed depth samples members

from different ensembles defined by different potential densities. That is, the fixed depth average cannot be an isentropic average, and so it is of little use when aiming to describe a mean over an isentropic ensemble.

7.3.4 Transformed residual mean (TRM)

To find a more convenient z-coordinate formulation, we follow approaches used by deSzoeke and Bennett (1993), McIntosh and McDougall (1996), Kushner and Held (1999), and McDougall and McIntosh (2001). Here, we introduce the displacement vector in the vertical

$$\boldsymbol{\xi} = \xi\, \hat{\mathbf{z}}. \tag{7.27}$$

This displacement vector is appropriate for the case where we position ourselves at a point (x, y, t) in a stably stratified fluid and focus on the vertical position of a parcel with a fixed potential density ρ. That is, we let $\xi(x, y, \rho, t)$ be the deviation of a potential density surface ρ from its mean vertical position

$$z(\rho) = \overline{z}^{\rho}(\rho) + \xi(\rho). \tag{7.28}$$

The ensemble mean of this deviation vanishes

$$\overline{\xi}^{\rho} = 0 \tag{7.29}$$

by definition.

We now define an isopycnally averaged version of a field Φ by fixing ourselves at (x, y, ρ, t) and computing the ensemble mean. In symbols, we measure the average $\overline{\Phi(x, y, \overline{z}^{\rho} + \xi, t)}^{\rho}$. The question arises how does $\overline{\Phi}^{\rho}$ correspond to fields measured at the mean depth $\overline{z}^{\rho}$ of the isopycnal? This question leads to our defining a *modified mean* field satisfying*

$$\widetilde{\Phi}(x, y, \overline{z}^{\rho}, t) = \overline{\Phi(x, y, \overline{z}^{\rho} + \xi, t)}^{\rho}. \tag{7.30}$$

That is, the modified mean field† $\widetilde{\Phi}$, when measured at the mean depth $\overline{z}^{\rho}$, is equal to the isopycnally averaged field $\overline{\Phi}^{\rho}$. A particularly important modified mean field is the potential density

$$\widetilde{\rho}(x, y, \overline{z}^{\rho}, t) = \overline{\rho(x, y, \overline{z}^{\rho} + \xi, t)}^{\rho}. \tag{7.31}$$

Since the average is defined over constant ρ, we can remove the average operation to render the equality

$$\widetilde{\rho}(x, y, \overline{z}^{\rho}, t) = \rho(x, y, \overline{z}^{\rho} + \xi, t). \tag{7.32}$$

In words, this equality says that the modified mean density $\widetilde{\rho}$, when measured at the mean depth $\overline{z}^{\rho}$, is equal to the density ρ of the ensemble members. Generalizations of the modified mean potential density to a nonlinear equation of state

*A similar equality arises when considering Generalized Lagrangian Mean fields (Andrews and McIntyre (1978)).

†Do not confuse the tilde symbol used here for modified mean field with that used in Chapter 18 for one-forms. The meaning is completely distinct.

lead to a modified mean neutral density (see McDougall and McIntosh (2001) for discussion).

The relation $\widetilde{\rho}(\overline{z}^{\rho}) = \rho(\overline{z}^{\rho} + \xi)$ implies that $\widetilde{\rho}$ is the functional inverse of the mean depth $\overline{z}^{\rho}$ of the ρ potential density surface (deSzoeke and Bennett (1993)). From equation (7.30), it follows that

$$\widetilde{\Phi}(x, y, \overline{z}^{\rho}, t) = \overline{\Phi(x, t, \overline{z}^{\rho} + \xi, t)}^{\rho} = \overline{\Phi}^{\rho}(x, y, \widetilde{\rho}, t). \tag{7.33}$$

This is an important relation between an ensemble mean field defined at the modified mean potential density surface (right hand side), and the modified mean field defined at the mean depth of the isopycnal (left hand side). In an analogous fashion, we *define* a relation between a thickness weighted field

$$\widehat{\Phi} = \frac{\overline{h\,\Phi}^{\rho}}{\overline{h}^{\rho}}, \tag{7.34}$$

defined as a function of density ρ, and a *transformed residual mean* (TRM) field $\overline{\Phi}^{\#}$ defined as a function of the mean depth (McDougall and McIntosh (2001))

$$\widehat{\Phi}(x, y, \widetilde{\rho}, t) = \overline{\Phi}^{\#}(x, y, \overline{z}^{\rho}, t). \tag{7.35}$$

Without thickness weighting, such as when h =constant for all ensemble members, the TRM field $\overline{\Phi}^{\#}(\overline{z}^{\rho})$ reduces to the modified mean field $\widetilde{\Phi}(\overline{z}^{\rho})$. Generally, these fields are distinct.

7.3.5 Exact ensemble mean kinematics in z-coordinates

We now apply the previous formalism to garner an exact z-coordinate description of the isentropic ensemble. For this purpose, we interpret a vertical position z as the ensemble mean vertical position $\overline{z}^{\rho}$ of a potential density surface. Mean fields defined at the fixed vertical position correspond to either modified mean fields when not thickness weighted (equation (7.33)), or TRM fields when thickness weighted (equation 7.35)). This interpretation of the vertical position, and fields defined at a vertical position, will be transferred to the discrete Eulerian lattice of a coarse resolution z-coordinate ocean model (see Section 7.5). Such an interpretation of coarse resolution z-models was originally proposed by McDougall and McIntosh (2001).

7.3.5.1 The total transport and the quasi-Stokes transport

As described using isopycnal coordinates, the mean specific thickness $\overline{h}^{\rho}$ of the ρ surface evolves via equation (7.22) through the stirring effects of the effective horizontal velocity $\widehat{\mathbf{u}}$. This velocity is thickness weighted, as defined by equation (7.23). According to our definition of a TRM field given by equation (7.35), we have an associated horizontal TRM velocity

$$\widehat{\mathbf{u}}(x, y, \widetilde{\rho}, t) = \overline{\mathbf{u}}^{\#}(x, y, \overline{z}^{\rho}, t). \tag{7.36}$$

So far, we have needed only the horizontal velocity components, either from the thickness weighted or TRM velocity. When working in z-coordinate models, it is useful to see how these velocities relate to a three-dimensional velocity.

As we are concerned with Boussinesq flow, a three-dimensional velocity used to transport properties is preferably non-divergent. Given the skewsion formalism of Section 7.2.1, we associate the horizontal TRM velocity components with a three-dimensional non-divergent TRM velocity field

$$\overline{\mathbf{v}}^{\#} = \nabla \wedge \overline{\mathbf{\Psi}}^{\#}, \tag{7.37}$$

where $\overline{\mathbf{\Psi}}^{\#}$ is a vector streamfunction.

The definition of $\overline{\mathbf{v}}^{\#}$ is shown to be reasonable *a posteriori* based on the following interpretation of the streamfunction $\overline{\mathbf{\Psi}}^{\#}$. Choosing the vertical gauge (Section 7.2.2) allows us to write

$$\overline{\mathbf{\Psi}}^{\#}(\overline{z}^{\rho}) = \int_{-H}^{\overline{z}^{\rho}} (\overline{\mathbf{u}}^{\#}(z) \wedge \hat{\mathbf{z}})\, \mathrm{d}z \equiv \overline{\mathbf{U}}^{\#}(\overline{z}^{\rho}) \wedge \hat{\mathbf{z}}, \tag{7.38}$$

where we introduced the mean horizontal transport $\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho})$. Vertical integration is over a column of the ocean from the bottom $z = -H$ to the ensemble mean depth $z = \overline{z}^{\rho}$ of the isopycnal ρ. We now show that $\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho})$ is the ensemble mean transport of fluid beneath the potential density surface $\rho = \widetilde{\rho}$. In so doing, we familiarize ourselves with the formalism in the process of providing a kinematic interpretation for the transport and justifying our choice for $\overline{\mathbf{v}}^{\#}$.

For these purposes, start by noting that

$$\begin{aligned} \overline{\mathbf{U}}^{\#}(\overline{z}^{\rho}) &= \int_{-H}^{\overline{z}^{\rho}} \overline{\mathbf{u}}^{\#}(z)\, \mathrm{d}z \\ &= \int_{\rho(-H)}^{\widetilde{\rho}(\overline{z}^{\rho})} \widehat{\mathbf{u}}(\sigma)\, \overline{h}^{\sigma}\, \mathrm{d}\sigma. \end{aligned} \tag{7.39}$$

This relation follows from a change in variables from depth to density using $\mathrm{d}z = \overline{h}^{\sigma}\, \mathrm{d}\sigma$, and by noting the inverse relation between $\overline{z}^{\rho}$ and $\widetilde{\rho}(\overline{z}^{\rho})$. We also used equation (7.36) relating the horizontal thickness weighted velocity $\widehat{\mathbf{u}}(\sigma)$, which is a function of density, and the TRM velocity $\overline{\mathbf{u}}^{\#}(\overline{z}^{\rho})$, which is a function of depth. Next, introduce definition (7.23) for the thickness weighted velocity to yield

$$\begin{aligned} \overline{\mathbf{U}}^{\#}(\overline{z}^{\rho}) &= \int_{\rho(-H)}^{\widetilde{\rho}(\overline{z}^{\rho})} \overline{\mathbf{u}\, h}^{\sigma}\, \mathrm{d}\sigma \\ &= \int_{\rho(-H)}^{\rho(\overline{z}^{\rho}+\xi)} \overline{\mathbf{u}\, h}^{\sigma}\, \mathrm{d}\sigma, \end{aligned} \tag{7.40}$$

where the second step follows from the conservation property (7.32) satisfied by the modified mean potential density. The integrand $\overline{\mathbf{u}\, h}^{\sigma}\, \mathrm{d}\sigma$ is the ensemble mean of the horizontal transport passing within the infinitesimal density layer between σ and $\sigma + \mathrm{d}\sigma$. It is vertically integrated over each of the continuum of potential

densities $\rho \leq \sigma \leq \rho(-H)$, thus yielding the ensemble mean transport of fluid beneath the potential density surface $\rho = \widetilde{\rho}$. This is the result we aimed to show.

Transferring the expression (7.40) to depth coordinates leads to the expression

$$\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho}) = \overline{\int_{-H}^{\overline{z}^{\rho}+\xi} \mathbf{u}\,\mathrm{d}z}. \tag{7.41}$$

In words, without the average operation, this expression represents a vertical integral of the horizontal transport, where the vertical integral extends from the ocean bottom $z = -H$ to a depth $z = \overline{z}^{\rho} + \xi$ determined by one of the members of an ensemble whose members all have potential density $\rho = \widetilde{\rho}$. With the average operation, we compute the mean of this vertical integral over all the members of the ρ ensemble.

To help with our interpretation of the transport $\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho})$, we split the vertical integral into two parts

$$\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho}) = \overline{\int_{-H}^{\overline{z}^{\rho}} \mathbf{u}\,\mathrm{d}z} + \overline{\int_{\overline{z}^{\rho}}^{\overline{z}^{\rho}+\xi} \mathbf{u}\,\mathrm{d}z}. \tag{7.42}$$

The first expression represents the ensemble mean horizontal transport between the bottom $z = -H$ and the fixed depth $z = \overline{z}^{\rho}$. Hence, it can be interpreted in the usual Eulerian average manner. The second term is the ensemble mean transport between the mean depth $z = \overline{z}^{\rho}$ of the ρ isopycnal, and the depth $z = \overline{z}^{\rho} + \xi$ of the various ensemble members. As a shorthand, we write this split as

$$\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho}) = \overline{\mathbf{U}}(\overline{z}^{\rho}) + \mathbf{U}^{qs}(\overline{z}^{\rho}). \tag{7.43}$$

Following McDougall and McIntosh (2001), we call

$$\mathbf{U}^{qs}(\overline{z}^{\rho}) \equiv \overline{\int_{\overline{z}^{\rho}}^{\overline{z}^{\rho}+\xi} \mathbf{u}\,\mathrm{d}z} \tag{7.44}$$

the *quasi-Stokes transport* for the ρ isopycnal. This transport arises from eddy effects accounting for differences between the Eulerian mean transport $\overline{\mathbf{U}}(\overline{z}^{\rho})$ and the TRM transport $\overline{\mathbf{U}}^{\#}(\overline{z}^{\rho})$. Such is characteristic of other forms of Stokes transport, which are defined between Eulerian and Lagrangian means (e.g., Andrews and McIntyre (1978) and Plumb (1979)). Note that as the surface or bottom of the ocean is approached, the quasi-Stokes transport vanishes since there is no more fluid to be transported at such boundaries. McDougall and McIntosh (2001) emphasize this point (see their Section 8) as it provides an unambiguous statement regarding the boundary conditions to be used for parameterizations of the quasi-Stokes transport (see Section 7.5).

7.3.5.2 Evolution of modified mean density

Following the skewsion formulation from Section 7.2, at the mean depth $z = \overline{z}^{\rho}$ the streamfunction $\overline{\mathbf{\Psi}}^{\#}$ defines an effective skew flux of potential density $\rho = \widetilde{\rho}$ given

by

$$\overline{\mathbf{F}}^{\#} = -\nabla\widetilde{\rho} \wedge \overline{\mathbf{\Psi}}^{\#}. \tag{7.45}$$

Using $\overline{\mathbf{\Psi}}^{\#} = \overline{\mathbf{U}}^{\#} \wedge \hat{\mathbf{z}}$, it is often useful to write this expression in one of the following forms

$$\begin{aligned} \overline{\mathbf{F}}^{\#} &= -\overline{\mathbf{U}}^{\#}\,\partial_z\widetilde{\rho} + \hat{\mathbf{z}}\,\overline{\mathbf{U}}^{\#}\cdot\nabla_z\widetilde{\rho} \\ &= -(\overline{\mathbf{U}}^{\#} + \hat{\mathbf{z}}\,\mathbf{S}\cdot\overline{\mathbf{U}}^{\#})\,\partial_z\widetilde{\rho}, \end{aligned} \tag{7.46}$$

where

$$\mathbf{S} = -\left(\frac{\nabla_z\widetilde{\rho}}{\partial_z\widetilde{\rho}}\right) \tag{7.47}$$

is the slope of the modified mean density field, and $\nabla_z = (\partial_x, \partial_y, 0)$ is the horizontal gradient operator taken with constant depth $z = \overline{z}^{\rho}$. The convergence of the effective skew flux leads to a stirring of the modified mean density $\widetilde{\rho}$ at the mean depth $z = \overline{z}^{\rho}$

$$\widetilde{\rho}_{,t} = \nabla\cdot(\nabla\widetilde{\rho}\wedge\overline{\mathbf{\Psi}}^{\#}). \tag{7.48}$$

This equation represents an exact z-coordinate specification of the evolution of the modified mean density due to stirring by the mean eddies. It corresponds directly to the evolution equation (7.18) satisfied at depth z by a single member of the ensemble. We contrast this elegant correspondence between unaveraged and averaged stirring processes with the lack of correspondence seen in the Eulerian description in Section 7.3.3.

7.3.6 Approximate ensemble mean kinematics in z-coordinates

Equation (7.48) represents an exact z-coordinate description of the stirring of modified mean potential density. However, when working in a discrete z-coordinate model, all that is available is Eulerian information. The Lagrangian information used to realize this exact description must be approximated.

As mentioned at the start of Section 7.3.5, we interpret values of potential density carried by a coarsely resolved z-model, at fixed model grid depths z, as the modified mean potential density $\widetilde{\rho}(z)$. Correspondingly, we associate model depth levels with the mean depth of the potential density surface: $z = \overline{z}^{\rho}$. This interpretation is in contrast to the usual association of the coarse model's Eulerian mean potential density $\overline{\rho}^{z}$ with the model potential density. The latter interpretation is not useful for coarse resolution z-models, as indicated by the discussion surrounding equation (7.26).

The approximation problem for z-coordinate models is a problem of how to compute the quasi-Stokes transport $\mathbf{U}^{qs}(z)$, where again $z = \overline{z}^{\rho}$ is assumed throughout this section. To do so in a z-model, the TRM transport is expanded in a Taylor series about the fixed model depth z via

$$\begin{aligned} \overline{\mathbf{U}}^{\#}(z) &= \overline{\int_{-H}^{z+\xi} \mathbf{u}(s)\,\mathrm{d}s} \\ &\approx \overline{\mathbf{U}}(z) + \overline{\mathbf{u}\,\xi}^{z} + \frac{1}{2}\,\overline{\partial_z\mathbf{u}\,\xi^2}^{z}, \end{aligned} \tag{7.49}$$

where neglected terms are third order in deviation quantities. Note that the averages are taken at fixed vertical position, which accords with our taking a Taylor series about the depth $z = \overline{z}^{\rho}$.

The averages in equation (7.49) are interpreted as follows. The first term is the usual Eulerian mean horizontal transport passing beneath the depth z. The second term, $\overline{\mathbf{u}(z)\,\xi}$, which is first order in deviation ξ, is the horizontal velocity evaluated at the mean depth $z = \overline{z}^{\rho}$ multiplied by the deviation ξ of the potential density surface from its mean depth, all averaged at fixed depth. An Eulerian split of the horizontal velocity $\mathbf{u}(z)$ into its Eulerian mean $\overline{\mathbf{u}}^z$ and deviation $\mathbf{u}'(z)$ leads to the correlation

$$\overline{\mathbf{u}\,\xi}^z = \overline{\mathbf{u}'\,\xi}^z. \tag{7.50}$$

For the second order term, similar considerations lead to

$$\overline{\partial_z \mathbf{u}\,\xi^2}^z \approx \partial_z \overline{\mathbf{u}}^z\,\overline{\xi^2}^z \tag{7.51}$$

where neglected terms are third order and higher. Combining these relations leads to the second order accurate expression

$$\overline{\mathbf{U}}^{\#}(z) \approx \overline{\mathbf{U}}(z) + \overline{\mathbf{u}'\,\xi}^z + (1/2)\,\overline{\xi^2}^z\,\partial_z \overline{\mathbf{u}}^z, \tag{7.52}$$

which can also be written in the suggestive form

$$\overline{\mathbf{U}}^{\#}(z) \approx \overline{\mathbf{U}}(z) + \overline{\xi\,(1 + (1/2)\,\xi\,\partial_z)\,\mathbf{u}}^z. \tag{7.53}$$

What remains is to determine the deviation ξ of the isopycnal in terms of fields at constant depth. For this purpose, use the relation (7.32) to render

$$\begin{aligned} \widetilde{\rho}(z) &= \rho(z + \xi) \\ &= \rho(z) + \partial_z \rho(z)\,\xi + \frac{1}{2}\partial_{zz}\rho(z)\,\xi^2, \end{aligned} \tag{7.54}$$

where terms of third and higher order were neglected. Subtracting the Eulerian mean of equation (7.54) from the un-averaged equation (7.54), and noting that $\widetilde{\rho}$ is already a mean-field, leads to the second order accurate expression for the deviation

$$\xi = -\rho'(z)/\partial_z \overline{\rho}^z, \tag{7.55}$$

where

$$\rho(z) = \overline{\rho}^z + \rho'(z). \tag{7.56}$$

To within the same order, the deviation can be written

$$\xi = -\rho'(z)/\partial_z \widetilde{\rho}(z), \tag{7.57}$$

which is more useful for applications in z-models since coarse versions of such models carry $\widetilde{\rho}$ rather than $\overline{\rho}$. Substituting the deviation (7.57) into the approximate expression (7.49) for the effective transport yields an approximate expression for the quasi-Stokes transport

$$\mathbf{U}^{qs} \approx -\frac{\overline{\mathbf{u}'\,\rho'}}{\partial_z \widetilde{\rho}} + \frac{\overline{\phi}\,\partial_z \overline{\mathbf{u}}}{(\partial_z \widetilde{\rho})^2}, \tag{7.58}$$

where

$$\overline{\phi}(z) = \frac{1}{2}\,\overline{(\rho'(z))^2} \tag{7.59}$$

is half the mean density variance, and recall that all depths are set to $z = \overline{z}^{\rho}$.

Substituting the deviation (7.57) into the approximate expression (7.54) yields, to within terms of third order, the relation

$$\widetilde{\rho}(z) \approx \overline{\rho}(z) - \partial_z \left(\frac{\overline{\phi}}{\partial_z \overline{\rho}} \right). \tag{7.60}$$

As for the quasi-Stokes transport, the modified mean density and Eulerian mean density, when evaluated at the same depth, differ by terms that are second order in eddy amplitude.

7.4 Ensemble mean tracer equation

In this section, we attach a tracer quantity to the ideal Boussinesq parcel and determine a mean field description for the tracer. The effects of the eddies have both a reversible stirring component and an irreversible mixing component. Much in this section follows from Smith (1999) and McDougall and McIntosh (2001).

7.4.1 Thickness weighted means

Equation (7.23) introduced a specific thickness weighted mean operator, which will prove to be quite useful when considering the mean tracer equation. In general, for any field Φ associated with a potential density layer ρ, define the decomposition

$$\Phi(\rho) = \widehat{\Phi}(\rho) + \Phi''(\rho) = \frac{\overline{h\,\Phi}^{\rho}}{\overline{h}^{\rho}} + \Phi''. \tag{7.61}$$

The quantity $\widehat{\Phi}$ is referred to as the mean thickness weighted field. It follows by definition that

$$\overline{h\,\Phi''}^{\rho} = 0. \tag{7.62}$$

7.4.2 Mean thickness weighted tracer equation

When attaching a tracer value to fluid parcels, each ensemble member satisfies the isopycnal tracer equation

$$(\partial_t + \mathbf{u} \cdot \nabla_\rho)\, C = 0. \tag{7.63}$$

Combining the tracer and thickness equations leads to the thickness weighted tracer equation

$$(h\,C)_{,t} + \nabla_\rho \cdot (h\,\mathbf{u}\,C) = 0. \tag{7.64}$$

We pause here to comment on an interpretation of this equation in relation to the orientation of the advective flux. Using isopycnal coordinates, the adiabatic evolution of thickness weighted tracer occurs via the isopycnally oriented convergence of the thickness weighted horizontal advective flux $h\,\mathbf{u}\,C$. As this flux is two-dimensional and proportional to $\mathbf{u}$, it is horizontally oriented. Hence, it is not generally oriented parallel to isopycnal surfaces. We made this point in Section

7.2.3 when discussing the complementary aspects of advective and skew fluxes of tracer as described using z-coordinates.

Returning now to the problem of describing the ensemble mean tracer equation in isopycnal coordinates, we decompose the tracer and velocity field into their thickness weighted average and deviation to render

$$[h\,(\widehat{C} + C'')]_{,t} + \nabla \cdot [h\,(\widehat{\mathbf{u}} + \mathbf{u}'')\,(\widehat{C} + C'')] = 0. \tag{7.65}$$

Taking an ensemble average over parcels with the same potential density, and using equation (7.62), renders the mean thickness weighted tracer equation

$$(\overline{h}^{\rho}\,\widehat{C})_{,t} + \nabla_{\rho} \cdot (\overline{h}^{\rho}\,\widehat{C}\,\widehat{\mathbf{u}}) = -\nabla_{\rho} \cdot (\overline{h\,C''\,\mathbf{u}''}^{\rho}). \tag{7.66}$$

With

$$\overline{h\,C''\,\mathbf{u}''}^{\rho} = \overline{h}^{\rho}\,\widehat{C''\,\mathbf{u}''}, \tag{7.67}$$

and $\partial_t \overline{h}^{\rho} + \nabla_{\rho} \cdot (\overline{h}^{\rho}\,\widehat{\mathbf{u}}) = 0$, we have the evolution equation for the mean thickness weighted tracer concentration

$$(\partial_t + \widehat{\mathbf{u}} \cdot \nabla_{\rho})\,\widehat{C} = -\frac{1}{\overline{h}^{\rho}}\,\nabla_{\rho} \cdot (\overline{h}^{\rho}\,\widehat{C''\,\mathbf{u}''}). \tag{7.68}$$

7.4.3 Tracer mixing tensor

The correlation term on the right hand side of the mean thickness weighted tracer equation (7.68) is typically parameterized via a mixing tensor

$$\widehat{C''\,\mathbf{u}''} = -\mathbf{J} \cdot \nabla_{\rho}\widehat{C}, \tag{7.69}$$

which leads to the evolution equation

$$(\partial_t + \widehat{\mathbf{u}} \cdot \nabla_{\rho})\,\widehat{C} = \frac{1}{\overline{h}^{\rho}}\,\nabla_{\rho} \cdot (\overline{h}^{\rho}\,\mathbf{J} \cdot \nabla_{\rho}\widehat{C}). \tag{7.70}$$

The ocean modeling community has traditionally assumed that the mixing tensor $\mathbf{J}$ is a positive-definite and symmetric diffusion tensor, which is also assumed isotropic in the isopycnal plane. It corresponds to the isopycnal coordinate version of the small angle diffusion tensor of Redi (1982). However, one can generally expect $\mathbf{J}$ to have both a symmetric and an anti-symmetric component, where the anti-symmetric component stirs the tracer via skew diffusion (e.g., Plumb (1979), Plumb and Mahlman (1987), Middleton and Loder (1989)). Importantly, such stirring by sub-grid scale correlations appears *in addition* to that associated with the bolus transport arising from $\mathbf{u}^b$, where again $\mathbf{u}^b$ arises from correlations between velocity and thickness.

7.4.4 Mean tracer transport beneath a density surface

It is useful to further elucidate the relevance of mean thickness weighted fields. For this purpose, we proceed as in Section 7.3.5.1 to consider the mean horizontal tracer transport occurring beneath a particular potential density surface $\rho = \widetilde{\rho}$

$$\overline{\mathbf{C}}^{\#}(\overline{z}^{\rho}) = \overline{\int_{-H}^{\overline{z}^{\rho}+\xi} C\,\mathbf{u}\,\mathrm{d}z}. \tag{7.71}$$

Setting tracer concentration to unity recovers the expression (7.41) for the TRM transport. As in Section 7.3.5.1, where we considered the mean horizontal transport of fluid beneath $\widetilde{\rho}$, we now have

$$\begin{aligned}
\overline{\mathbf{C}}^{\#}(\overline{z}^{\rho}) &= \int_{\rho(-H)}^{\widetilde{\rho}(\overline{z}^{\rho})} \overline{C\,\mathbf{u}\,h}^{\sigma}\,\mathrm{d}\sigma \\
&= \int_{\rho(-H)}^{\widetilde{\rho}(\overline{z}^{\rho})} \overline{h}^{\sigma}\,\mathrm{d}\sigma\,(\widehat{C}\,\widehat{\mathbf{u}} + \widehat{C''\,\mathbf{u}''}) \\
&= \int_{\rho(-H)}^{\widetilde{\rho}(\overline{z}^{\rho})} \overline{h}^{\sigma}\,\mathrm{d}\sigma\,(\widehat{C}\,\widehat{\mathbf{u}} - \mathbf{J}\cdot\nabla_{\rho}\widehat{C}) \\
&= \int_{-H}^{\overline{z}^{\rho}} \mathrm{d}z\,(\widehat{C}\,\widehat{\mathbf{u}} - \mathbf{J}\cdot\nabla_{\rho}\widehat{C}).
\end{aligned} \tag{7.72}$$

Hence, the mean thickness weighted fields naturally appear when considering such physically interesting quantities as the mean horizontal transport of a tracer beneath the modified mean potential density surface.

7.4.5 Summary of the mean field tracer equation

In summary, the parameterization problem for the mean thickness weighted tracer in isopycnal coordinates reduces to a parameterization of the bolus velocity $\mathbf{u}^b$, related to the effective horizontal velocity field

$$\widehat{\mathbf{u}}(\rho) = \frac{\overline{h\,\mathbf{u}}^{\rho}}{\overline{h}^{\rho}} = \overline{\mathbf{u}}^{\rho} + \frac{\overline{h'\,\mathbf{u}'}^{\rho}}{\overline{h}^{\rho}} = \overline{\mathbf{u}}^{\rho} + \mathbf{u}^b, \tag{7.73}$$

as well as the tracer mixing tensor

$$\widehat{C''\,\mathbf{u}''} = -\mathbf{J}\cdot\nabla_{\rho}\widehat{C}, \tag{7.74}$$

which generally has symmetric (diffusive) and anti-symmetric (stirring) components.

For a mean z-coordinate description, we use equation (7.35) relating thickness weighted mean fields, defined as a function of ρ, and TRM fields, defined as a function of the mean depth of ρ, to write for the tracer field

$$\widehat{C}(x, y, \widetilde{\rho}, t) = \overline{C}^{\#}(x, y, \overline{z}^{\rho}, t). \tag{7.75}$$

It is the TRM tracer concentration $\overline{C}^{\#}$ that is discretized by coarse resolution z-coordinate ocean models. Equation (7.75) and our developed formalism leads to the exact mean-field tracer equation in z-coordinates

$$\partial_t \overline{C}^{\#} = \nabla\cdot(\nabla\overline{C}^{\#} \wedge \overline{\mathbf{\Psi}}^{\#}) + R(\overline{C}^{\#}), \tag{7.76}$$

where $R(\overline{C}^{\#})$ is the z-coordinate form of the mixing/stirring operator on the right hand side of equation (7.70). In z-coordinates, this operator is generally comprised of the Redi (1982) isopycnal diffusion tensor, which produces a diffusive flux of $\overline{C}^{\#}$ oriented along surfaces of constant modified mean potential density $\widetilde{\rho}$, and an anti-symmetric piece which has yet to be studied in the ocean modeling literature. $\overline{\mathbf{\Psi}}^{\#}$ is the TRM transport streamfunction defined by equation (7.38). We further discuss its parameterization in Section 7.5.

7.5 Quasi-Stokes transport in z-models

The goal of a parameterization is to represent eddy correlation terms by expressions dependent only on the mean fields. Two correlation terms have been identified in the previous sections: the quasi-Stokes transport $\mathbf{U}^{qs}$ and the tracer mixing/stirring operator $R(\overline{C}^{\#})$. We focus here on the quasi-Stokes transport and its parameterization in z-models. As our focus here is strictly z-coordinates, all averages refer to Eulerian averages and so the z label on the overbar is dropped for brevity.

As stated earlier, we assume that modified mean potential density $\widetilde{\rho}$ is discretized by a coarse resolution Boussinesq z-model. Hence, the skew flux from equation (7.46)

$$\overline{\mathbf{F}}^{\#} = -\overline{\mathbf{U}}^{\#}\,\partial_z\widetilde{\rho} + \hat{\mathbf{z}}\,\overline{\mathbf{U}}^{\#}\cdot\nabla_z\widetilde{\rho} \tag{7.77}$$

must be parameterized. Again, for the isentropic ensemble with kinematic rearrangement of parcels, the horizontal TRM transport is written

$$\overline{\mathbf{U}}^{\#}(z) = \overline{\mathbf{U}}(z) + \mathbf{U}^{qs}(z), \tag{7.78}$$

with $z = \overline{z}^{\rho}$ the mean depth of the potential density, $\overline{\mathbf{U}}(z)$ the Eulerian mean horizontal transport explicitly carried by the z-model that passes beneath the depth z, and

$$\mathbf{U}^{qs}(z) = \overline{\int_{\overline{z}^{\rho}}^{\overline{z}^{\rho}+\xi} \mathbf{u}\,\mathrm{d}z} \approx -\frac{\overline{\mathbf{u}'\,\rho'}}{\partial_z\widetilde{\rho}} + \frac{\overline{\phi}\,\partial_z\overline{\mathbf{u}}}{(\partial_z\widetilde{\rho})^2} \tag{7.79}$$

is the horizontal eddy-induced, or quasi-Stokes, transport which must be parameterized. The second, approximate, expression is accurate to third order in eddy amplitude. Notably, one advantage of working with the skewsion or streamfunction approach is that it is only a horizontal transport $\mathbf{U}^{qs}$ that must be parameterized, in contrast to the conventional Eulerian approach which aims to parameterize a three dimensional eddy correlation $\overline{\mathbf{v}'\,\rho'}$.

7.5.1 Rotational and divergent components

The analysis so far has found the vertical gauge (Section 7.2.2) to be a natural means to represent both the resolved and unresolved potential density flux. However, as it is the convergence of the flux that determines the evolution of the modified mean potential density field $\widetilde{\rho}$, one can add various curls, or rotational terms, to the flux to put it into a more convenient form. For example, when representing the flux due

to the resolved scale flow, it is possible to add the total rotational term $\nabla \wedge (\widetilde{\rho}\,\overline{\mathbf{U}} \wedge \hat{\mathbf{z}})$ in order to change the skew flux to an advection flux.

For the unresolved part of the flux, proportional to the quasi-Stokes transport $\mathbf{U}^{qs}$, we remain within the skewsion framework as it is more convenient. However, there remains a gauge ambiguity that allows for the addition of an arbitrary rotational term to the horizontal skew flux $-\partial_z \widetilde{\rho}\ \mathbf{U}^{qs}$. Indeed, when diagnosing eddy correlations in fine resolution models, there is generally a nontrivial rotational term (e.g., Lau and Wallace (1979), Marshall and Shutts (1981), Bryan et al. (1999)) which is associated with the advection of potential density variance. This term is not easily parameterized, nor does it require parameterization since its convergence vanishes.

7.5.1.1 Flux divergence

Various approaches present themselves for handling the unresolved rotational term. The first approach is to not concern oneself with details of the diagnosed fluxes, but instead to focus on their convergence. This approach formed part of the analysis of Bryan et al. (1999), and it provides unambiguous information about what drives the mean flow. However, in some occasions it may produce noisy results which makes it difficult to make strong conclusions. This difficulty may be dependent on details of the problem considered.

7.5.1.2 Zonal symmetry and zonal means

The second approach is to work in a zonally symmetric channel and to employ zonal averages instead of, or in addition to, time and/or ensemble averages. All averaged fields then become functions of just the meridional and vertical spatial directions, and the effective transport reduces to just the meridional component.

More precisely, consider a zonally averaged form of the thickness equation using isopycnal coordinates

$$\partial_t \overline{h}^{\rho} + \partial_y (\overline{h}^{\rho}\, \widehat{v}) = 0, \tag{7.80}$$

with

$$\widehat{v} = \overline{v}^{\rho} + \frac{\overline{v'\, h'}^{\rho}}{\overline{h}^{\rho}} \tag{7.81}$$

the effective meridional velocity field. The $\partial_x(\overline{u\, h}^{\rho})$ term vanishes due to zonal symmetry and use of zonal averages. The corresponding exact z-coordinate description thence considers the two-dimensional effective skew flux

$$\overline{\mathbf{F}}^{\#} = \overline{V}^{\#}(-\hat{\mathbf{y}}\, \partial_z \widetilde{\rho} + \hat{\mathbf{z}}\, \partial_y \widetilde{\rho}) \tag{7.82}$$

whose convergence drives the evolution of the modified mean potential density field $\widetilde{\rho}(y, z, t)$. The same approximations used to reach the approximate quasi-Stokes streamfunction (7.79) can be applied here to yield the third order accurate form for the effective transport passing beneath the averaged depth of the modified mean potential density field:

$$\int_{-H}^{\overline{z}^{\rho}} \widehat{v}\, \mathrm{d}z \approx \overline{V}(z) - \frac{\overline{v'\, \rho'}}{\partial_z \widetilde{\rho}} + \frac{\overline{\phi}\, \partial_z \overline{v}}{(\partial_z \widetilde{\rho})^2}. \tag{7.83}$$

The only ambiguity that remains in the meridional skew flux $-\overline{V}^{\#}\,\partial_z\widetilde{\rho}$ arises from the ability to add an arbitrary function of zonal position. Yet since all averaged fields are only functions of meridional and depth positions, this ambiguity is not a problem. The zonally symmetric case is therefore quite simple to consider.

7.5.1.3 The rotational component

The third approach is to actually take the effort to compute the rotational component. Doing so requires the specification of boundary conditions, and their form is not obvious, except for the case of periodicity in which both rotational and divergent components are periodic.

To expose the main issue, consider the case when $-\mathbf{U}^{qs}\,\partial_z\widetilde{\rho}$ is diagnosed. For brevity, at each depth level, write

$$\mathbf{A}(x,y,t) = -\mathbf{U}^{qs}\,\partial_z\widetilde{\rho}. \tag{7.84}$$

There is a self-consistent method for splitting $\mathbf{A}$ into its rotational and divergent components, and this method is standard in the fluid mechanics literature (e.g., Section 1.1 of Saffman (1992) and Section 17.2 of Panton (1996)). Since the vector $\mathbf{A}$ is two-dimensional, the decomposition generally takes the form

$$\mathbf{A} = \mathbf{A}_{div} + \mathbf{A}_{rot} = \nabla\Phi + \hat{\mathbf{z}}\wedge\nabla\chi, \tag{7.85}$$

where the scalar fields Φ and χ satisfy

$$\nabla^2\Phi = \nabla\cdot\mathbf{A} \tag{7.86}$$

$$\nabla^2\chi = \hat{\mathbf{z}}\cdot\nabla\wedge\mathbf{A}. \tag{7.87}$$

Because of the no-normal flow boundary condition on vertical side walls

$$\mathbf{A}\cdot\hat{\mathbf{n}} = 0, \tag{7.88}$$

where $\hat{\mathbf{n}}$ is the outward normal at the boundary. When decomposing $\mathbf{A}$ into its rotational and divergent components, the boundary condition becomes

$$\hat{\mathbf{n}}\cdot\nabla\Phi + \hat{\mathbf{n}}\cdot(\hat{\mathbf{z}}\wedge\nabla\chi) = 0. \tag{7.89}$$

Notably, each component in general does *not* separately satisfy the no-normal flow condition.

To proceed, the scalar field χ is assumed to maintain regularity at infinity, periodicity on the sphere, yet no specified boundary conditions on the side walls. Upon solving for χ via a numerical elliptic solver or a Green's function approach, its value is known everywhere. In particular, it is known on the side boundaries, which in turn allows for the scalar field Ψ to be found by solving the boundary value problem

$$\nabla^2\Psi = \nabla\cdot\mathbf{A} \tag{7.90}$$

$$\hat{\mathbf{n}}\cdot\nabla\Psi = -\hat{\mathbf{n}}\cdot(\hat{\mathbf{z}}\wedge\nabla\chi), \tag{7.91}$$

again using a numerical elliptic solver or a Green's function approach. Thus specifies the divergent and rotational components.

Although the above algorithm is self-consistent, it is not unique. Fox-Kemper et al. (2003) provides a thorough summary of this issue, and suggests some integral diagnostics which are unique and so of use for the eddy-closure problem.

7.5.2 GM90 skew-diffusion

A popular parameterization of the quasi-Stokes streamfunction is that proposed by Gent and McWilliams (1990) and Gent et al. (1995). We discuss this parameterization more thoroughly in Section 12.2.1. For now, we note that this scheme closes the divergent part of the quasi-Stokes transport by setting

$$\mathbf{U}^{qs} = -\kappa \mathbf{S}. \tag{7.92}$$

In this expression, $\mathbf{S}$ is the slope of the modified mean potential density surfaces (equation (7.47)), and $\kappa > 0$ is a diffusivity. The corresponding three-dimensional non-divergent velocity is given by

$$\mathbf{v}^* = -\partial_z \left(\kappa \, \mathbf{S}\right) + \hat{\mathbf{z}} \, \nabla_z \cdot \left(\kappa \, \mathbf{S}\right), \tag{7.93}$$

and the anti-symmetric stirring tensor (Griffies (1998)) is

$$A^{mn} = \begin{pmatrix} 0 & 0 & -\kappa S_x \\ 0 & 0 & -\kappa S_y \\ \kappa S_x & \kappa S_y & 0 \end{pmatrix}. \tag{7.94}$$

The parameterized skew flux of modified mean potential density due to the quasi-Stokes transport is given by

$$\overline{\mathbf{F}}^{qs} = -\mathbf{U}^{qs} \, \partial_z \widetilde{\rho} + \hat{\mathbf{z}} \, \mathbf{U}^{qs} \cdot \nabla_z \widetilde{\rho} = -\kappa \left(\nabla_z \widetilde{\rho} - \hat{\mathbf{z}} \, S^2 \, \partial_z \widetilde{\rho} \right). \tag{7.95}$$

This parameterization leads to horizontal downgradient diffusion of modified mean potential density, combined with a vertical upgradient diffusion. Additionally, Gent et al. (1995) prescribe a diffusivity that vanishes on all boundaries, including the ocean surface. McIntosh and McDougall (1996) and McDougall and McIntosh (2001) present more discussion of vertical boundary conditions, which are understood transparently when considering the exact form of the quasi-Stokes transport defined by equation (7.44).

To see what the Gent et al. (1995) parameterization corresponds to in terms of sub-grid scale correlations, we expand the quasi-Stokes transport to second order (equation (7.79)) and thus equate to the parameterization (7.92)

$$\left(-\overline{\mathbf{u}' \, \rho'} + \frac{\overline{\phi} \, \partial_z \overline{\mathbf{u}}}{\partial_z \widetilde{\rho}} \right)_{div} \approx \kappa \, \nabla_z \widetilde{\rho}, \tag{7.96}$$

where only the divergent part of the eddy correlation is considered. The importance of the $\overline{\mathbf{u}}_z \overline{\phi} / \partial_z \widetilde{\rho}$ term has not been examined in the literature. Instead, for example, Treguier et al. (1997) and Roberts and Marshall (2000) ignore this term, thence equating the Gent et al. (1995) parameterization with the traditional horizontal downgradient diffusion of potential density $(\overline{\mathbf{u}' \, \rho'})_{div} = -\kappa \, \nabla_z \widetilde{\rho}$. Again, such is *not correct* when the variance term $\overline{\mathbf{u}}_z \overline{\phi} / \partial_z \widetilde{\rho}$ is important. Careful examination of eddying z-coordinate models is necessary to clarify the importance of the variance term.

7.5.3 Effective meridional volume transport

It is often of interest to compute the net transport of volume across a portion of the ocean. In particular, meridonal-overturning streamfunctions allow us to visualize and quantify the zonally averaged circulation occuring in a closed basin or over the full globe. The quasi-Stokes transport provides a transport in addition to that from the resolved scale Eulerian mean transport. The parameterization of Gent et al. (1995) leads to a straightforward computation of the quasi-Stokes contribution. To see this, we write the net meridional transport across a basin at a particular depth in the form

$$\begin{aligned} \mathcal{T}(y,z,t) &= \int \mathrm{d}x\, \overline{V}^{\#}(y,z,t) \\ &= \int \mathrm{d}x\, (\overline{V} - \kappa\, S_y). \end{aligned} \tag{7.97}$$

The parameterized quasi-Stokes transport adds a contribution that scales linearly with basin size, isopyncal slope, and diffusivity

$$\mathcal{T}^{qs} \sim L\, S\, \kappa. \tag{7.98}$$

As an example, let $\kappa = 10^3 \mathrm{m}^2\,\mathrm{s}^{-1}$, $S = 10^{-3}$, and $L = 10^7$ m, which yields $\mathcal{T} \approx$ 10Sv. Such transport can represent a nontrivial addition to that from the resolved scale velocity field.

7.5.4 GM90 and diffusion of isopycnal thickness

Recall the ensemble mean thickness equation (7.22) derived in Section 7.3.2

$$\partial_t\, \overline{h}^{\rho} + \nabla_\rho \cdot (\overline{h}^{\rho}\, \widehat{\mathbf{u}}) = 0, \tag{7.99}$$

where $\widehat{\mathbf{u}} = \overline{\mathbf{u}}^{\rho} + \mathbf{u}^b$ provides an effective transport velocity for the ensemble mean thickness $\overline{h}^{\rho}$. Isopycnal correlations of horizontal velocity and thickness define the bolus velocity via $\mathbf{u}^b\, \overline{h}^{\rho} = \overline{h'\, \mathbf{u}'}^{\rho}$. Now consider a downgradient diffusive closure for this correlation

$$\overline{h}^{\rho}\, \mathbf{u}^b = \overline{h'\, \mathbf{u}'}^{\rho} = -\mathbf{K} \cdot \nabla_\rho\, \overline{h}^{\rho} \tag{7.100}$$

with $\mathbf{K}$ a symmetric 2×2 diffusion tensor. The mean thickness equation thus takes the form of an advection-diffusion equation

$$\partial_t\, \overline{h}^{\rho} + \nabla_\rho \cdot (\overline{h}^{\rho} \overline{\mathbf{u}}^{\rho}) = \nabla_\rho \cdot (\mathbf{K} \cdot \nabla_\rho\, \overline{h}^{\rho}). \tag{7.101}$$

To make a connection between the thickness diffusion closure (7.100) and the Gent and McWilliams (1990) closure discussed in Section 7.5.2, note that the mean thickness is related to the vertical gradient of the modified mean density

$$\overline{h}^{\rho} = (\partial_z\, \widetilde{\rho})^{-1} \tag{7.102}$$

where the right hand side is evaluated at a depth given by the mean of the isopycnal $z = \overline{z}^{\rho}$. Correspondingly, with $\nabla_\rho = \nabla_z + \mathbf{S}\, \partial_z$, again where $z = \overline{z}^{\rho}$ and $\mathbf{S}$ is the

slope of the modified mean density field $\widetilde{\rho}$ (see equation (7.47)), we have

$$\begin{aligned}
\nabla_\rho \ln \overline{h}^\rho &= -\overline{h}^\rho \, \nabla_\rho \, (\overline{h}^\rho)^{-1} \\
&= -(\partial_z \, \widetilde{\rho})^{-1} \, (\nabla_z + \mathbf{S} \, \partial_z) \, \partial_z \, \widetilde{\rho} \\
&= -\frac{\partial_z \, (\nabla_z \, \widetilde{\rho})}{\partial_z \, \widetilde{\rho}} + \frac{\partial_{zz} \, \widetilde{\rho} \, \nabla_z \widetilde{\rho}}{(\partial_z \widetilde{\rho})^2} \\
&= \partial_z \mathbf{S}.
\end{aligned} \tag{7.103}$$

Consequently, the bolus velocity takes the form

$$\mathbf{u}^b = -\mathbf{K} \cdot \nabla_\rho \, \ln \overline{h}^\rho = -\mathbf{K} \cdot \partial_z \mathbf{S}. \tag{7.104}$$

For the special case where $\mathbf{K}$ is independent of depth and proportional to the 2×2 identity matrix, then

$$\mathbf{u}^b = -\partial_z \, (\kappa \, \mathbf{S}) = \mathbf{u}^* \tag{7.105}$$

where we identified the horizontal component of the Gent and McWilliams (1990) velocity $\mathbf{u}^*$ from equation (7.93). Again, this identity holds only for the special case of a vertically independent diffusivity tensor proportional to the identity. The relevance of a depth independent diffusivity has been questioned by recent research, such as Killworth (1997) and Smith and Vallis (2002). For these cases, placement of the vertical derivative is crucial.

We note that there has been no identification of the vertical velocity component $w^* = \nabla_z \cdot (\kappa \mathbf{S})$ from Gent and McWilliams (1990) with the vertical component of the bolus velocity. Indeed, there is no need to explicitly compute the vertical component of the bolus velocity since the thickness equation (7.99) only cares about the horizontal transport. However, for z-models, the full three-dimensional velocity is needed. Details of these points are provided by McDougall and McIntosh (2001).

CHAPTER
EIGHT

Concerning unresolved physics

Contents

8.1 Represented dynamics and parameterized physics 188
8.1.1 Resolved and SGS advective transport 188
8.1.2 Information loss and the closure problem 189
8.2 Lateral (neutral) and vertical processes 190
8.2.1 Neutral directions 190
8.2.2 Neutral and dianeutral physical processes 190
8.3 Basic mechanisms for dianeutral transport 191
8.3.1 Mechanical and buoyant mixing 191
8.3.2 The nonlinear equation of state 192
8.3.3 Double diffusion 193
8.3.4 Summary of the processes 193
8.4 Dianeutral transport in models 193
8.4.1 Dianeutral velocity component 193
8.4.2 Dianeutral advection of buoyancy 194
8.4.3 The tracer equation with a diffusive closure 194
8.4.4 Evolution of locally referenced potential density 196
8.5 Numerically induced spurious dianeutral transport 198

A fundamental problem in computational fluid dynamics is how to parameterize processes that live at space-time scales not explicitly represented by the computational grid. Due to nonlinear interactions between scales, details of these subgrid-scale processes (SGS) are often critical to the structure of the larger scales explicitly represented. Hence, SGS processes cannot be ignored. Instead, they must be parameterized.

The purpose of this chapter is to provide an overview of the issues involved in SGS parameterizations. This discussion builds on that given in Chapters 6 and 7, and furthermore helps to introduce the material in Parts IV and V which focus on parameterizations of lateral (i.e., neutral) transport processes. We also highlight here some issues related to small-scale dianeutral processes. Such processes affect

the vertical stratification of properties in the ocean, and their parameterization in ocean models is critical for physically realistic simulations.

8.1 Represented dynamics and parameterized physics

A process that is not resolved by a numerical model grid constitutes part of the model's *subgrid-scale* (SGS). As we are concerned in this book with physics, unresolved processes are said to comprise the *SGS physics*. In contrast, space-time scales that are resolved by the model grid are termed *model dynamics*.

This terminology is somewhat loose. For example, resolved dynamical processes are often not fully resolved in a sense satisfactory to a rigorous numerical analyst. Likewise, unresolved processes are often not fully subgrid-scale. Furthermore, even if there are plenty of grid points to in principle resolve a dynamical process, the model's discrete representation may not be faithful to the continuum dynamics due to inaccurate or inappropriate numerical techniques. Given these caveats, the loose distinction between *resolved dynamics* and *SGS physics* serves our purposes for this chapter.

8.1.1 Resolved and SGS advective transport

An averaged description of nonlinear advective transport appearing in both the momentum and the tracer equations introduces SGS transport terms whose form is not prescribed by knowledge of just the resolved scales. Averaged advective transport serves as the canonical example of the split between resolved dynamics and SGS physics. Mathematically, we see this split by considering the average of advective transport for an arbitrary field ψ

$$\langle \mathbf{v}\, \psi \rangle = \langle \mathbf{v} \rangle \, \langle \psi \rangle + \langle \mathbf{v}' \, \psi' \rangle . \tag{8.1}$$

In this equation, the angle brackets denote an average operator, primes represent deviations from the average, and $\mathbf{v}$ is the velocity field. The first term, $\langle \mathbf{v} \rangle \, \langle \psi \rangle$, is explicitly represented in an ocean model via a choice for a numerical advection scheme. That is, this term is cast onto a numerical lattice and constitutes part of the model's resolved advective transport. The integrity of this representation of resolved advection is very dependent on the sophistication of the numerical advection scheme. Such matters are crucial to the model simulation, yet are not part of the discussion here (see Durran (1999) for a thorough and pedagogical presentation of advection).

Given knowledge only of the mean fields $\langle \mathbf{v} \rangle$ and $\langle \psi \rangle$, the correlation term $\langle \mathbf{v}' \, \psi' \rangle$ must be parameterized in order to account for its effects on the mean fields. Discussions in Chapters 6 and 7 provide specific examples of such SGS transport terms. Notably, via a choice in how one interprets the averaged fields and/or applies the averaging operator, we are able in both chapters to reduce the number of SGS terms appearing in the averaged primitive equations, especially those appearing in the kinematic equations. Nonetheless, the advective transport term $\langle \mathbf{v}' \, \psi' \rangle$ cannot generally be eliminated via reinterpreting the resolved fields. Instead, it leads to a SGS transport whose effects on the mean flow is often nontrivial, especially in regions of the ocean where unresolved turbulent transport occurs. Hence,

ignoring this term, or prescribing a closure in an *ad hoc* or cavalier manner, is generally not an acceptable option.

8.1.2 Information loss and the closure problem

Providing a systematic means to rationally parameterize SGS transport processes constitutes one of the fundamental unresolved problems in classical physics. This is the *turbulence closure problem*. Notably, there is not just one closure problem. Instead, there is a closure problem for each of the many unresolved physical processes. For example, the closure problem for unresolved mesoscale eddies is distinct from that for unresolved breaking internal waves. Details of the unresolved processes generally prompt the kinematic framework for posing the closure problem as well as the dynamical assumptions that lead to a choice for the form of the closure.

The closure problem results from a need to average over scales that are unresolved or unmeasured by the coarse-grained observer. Hence, the averaged equations contain less information than the unaveraged equations. This situation reflects the limits inherent in modeling and/or observing. Information loss also occurs when formulating a thermodynamic description of matter from statistical mechanics (e.g., Jaynes (1957a)). From a statistical perspective, the two problems share much in common. Indeed, some effort has been focused on the use of information theory and its associated principle of maximum entropy to motivate the form of turbulence closure in certain fluid systems (see e.g., Kraichnan and Montgomery (1980), Holloway (1992), Salmon (1998), and Chavanis (2002)). These approaches generally lead to a SGS term that relaxes the system to a state of maximum entropy.

For simple physical systems, such as a Brownian particle (Reif (1965)) whereby we consider the averaged motion of a particle within a fluid affected by random molecular bombardments, entropy maximization leads to a diffusive SGS term in the mean-field equations. In fluid closures, following the Brownian particle example, one may wish to diffuse all quantities that are materially conserved in ideal flow, such as material and thermodynamic tracers. It may also be appropriate to diffuse dynamical tracers that are materially conserved, such as potential vorticity (PV) in ideal flows (see Section 3.7). In contrast, it is not generally appropriate to diffuse linear momentum since it is not materially conserved due to pressure forces. As discussed in Part V of this book, linear momentum diffusion is nonetheless often used in ocean modeling due to numerical constraints.

Tracer diffusion is both numerically necessary and physically reasonable, so long as the diffusion respects the huge separation in magnitude between neutral and dianeutral processes (see Section 8.2). Additionally, PV diffusion is the preferred approach for many theories of mesoscale eddy closures (see Section 7.1 for more discussion). Notably, arguments leading to Brownian motion diffusive closures of material and thermodynamic tracers must be extended when considering a dynamically active tracer such as PV. At present, there is no general theory leading to the identification of PV diffusion with entropy maximization within the primitive equations. The significance of this theoretical hole depends on what one regards as more fundamental, PV diffusion or entropy maximization.

8.2 Lateral (neutral) and vertical processes

As mentioned above, ocean turbulence associated with small scale mixing tends to be three-dimensional. As an approximation, this mixing is typically assumed isotropic. At the larger scales (order many kilometres to hundreds of kilometres), transport exhibits a significant amount of anisotropy, with lateral processes far more efficient than vertical processes. In this section, we describe why such is the case, and make more precise what oceanographers mean by lateral.

8.2.1 Neutral directions

Consider a fluid stably stratified in the vertical yet without any horizontal stratification (i.e., flat isopycnals and so zero baroclinicity). Adiabatic motion of a parcel in the vertical exposes the parcel to a restoring force due to its being surrounded by fluid with different buoyancy $-\rho\, g/\rho_o$. When displaced a small amount in the vertical, the parcel undergoes linear oscillations about its initial equilibrium position with a frequency proportional to the vertical density gradient (e.g., Gill (1982)). However, adiabatic motion in the horizontal does not alter buoyancy, and so horizontal motion encounters no restorative force. That is, the horizontal defines a direction of neutral buoyancy, or *neutral direction*, for horizontal density surfaces.

For non-horizontal isopycnals, neutral directions remain those directions that a parcel can adiabatically move without altering its buoyancy. McDougall (1987a) formalized this definition and showed us how to be mindful to locally define the potential density when the equation of state is nonlinear. The result of his analysis indicates that neutral directions are tangent to the locally referenced potential density surface. Therefore, the adiabatic interchange of two fluid parcels along a neutral direction does not alter the buoyancy of either parcel, nor does it alter the gravitational potential energy of the combined two fluid parcel system. Locally referenced potential density can be considered a generalization of buoyancy (with a factor of $-g/\rho_o$ added), and so we often use the terms synonymously in the following.

This discussion indicates how stratification acts as an effective barrier to motion across neutral directions. A consequence of this barrier is that many important transport processes tend to spread properties much more efficiently along rather than across neutral directions. Hence, neutral directions define what oceanographers mean by "lateral." Measurements indicate that on the large-scales (order hundreds of kilometres), the associated neutral-dianeutral anisotropy seen in tracer transport can be as high as 10^8 in the ocean interior, with smaller anisotropies in regions of strong dianeutral mixing such as within boundary layers or above rough topography.

8.2.2 Neutral and dianeutral physical processes

Because of the anisotropy in transport processes described above, oceanographers find it useful to classify SGS physical processes according to whether they predominantly act to transport properties in neutral directions or dianeutral directions. The processes associated with neutral directed transport (both advective and diffusive) are dominated by mesoscale eddies, and their effects on the lateral distribution of

oceanic properties require parameterization in present-day ocean climate models. As introduced in Chapter 7, kinematic and dynamic issues related to understanding and parameterizing these eddies constitutes a large focus of current research. Parameterizations of such *neutral physics* processes are discussed in Part IV.

Even though the dominant transport of ocean properties is along, and not across, neutral directions, those processes contributing to the material modification of buoyancy are crucial to establishing the oceanic stratification and vertical distribution of oceanic properties such as heat, salt, and nutrients. Depending on the model class, certain of these *dianeutral* transport processes constitute the dieneutral SGS physics in an ocean model. Research on understanding and parameterizing processes affecting dianeutral transport is quite active. This book is remiss at not devoting a full set of chapters to this active field. To partially remedy this absence, we provide a terse discussion in Sections 8.3 and 8.4. A far more thorough compendium can be found in various contributions to Chassignet and Verron (1998), the thorough review of small-scale mixing in the book by Kantha and Clayson (2000b), and the review articles by Holloway (1989) and Toole and McDougall (2001).

In closing this section, we comment on the situation in atmospheric modeling, where SGS physics generally refers only to physical processes materially affecting buoyancy, such as convection and boundary layer processes. This situation is related to the larger Rossby radius in the atmosphere (some 10 times larger than the ocean), which sets the scale for quasi-geostrophic turbulent processes (i.e., synoptic weather) that dominanat the processes associated with neutral physics. This larger spatial scale is reasonably well resolved by current global atmospheric models, thus alleviating the need to parameterize their effects.

8.3 Basic mechanisms for dianeutral transport

Small-scale (order millimeters to many tens to hundreds of meters) turbulence leads to dianeutral mixing of fluid parcels. Such mixing is typically parameterized as Fickian diffusion of the material constituents and thermodynamic properties of the fluid (i.e., dianeutral diffusion of tracers as discussed in Section 8.4). Dianeutral tracer diffusion leads to dianeutral buoyancy diffusion. If the equation of state is linear, then dianeutral buoyancy diffusion is the only contributor to dianeutral buoyancy transport. In the presence of a nonlinear equation of state for seawater, processes parameterized as neutral diffusion of potential temperature and salinity lead to an additional contribution to dianeutral buoyancy transport which is itself non-diffusive in nature. The purpose of this section is to introduce some processes leading to dianeutral transport, with Section 8.4 following-up with a more mathematical discussion.

8.3.1 Mechanical and buoyant mixing

As noted in Section 1.2 of Kantha and Clayson (2000b), there are two basic methods whereby turbulent mixing is generated and maintained in the ocean: mixing via shear instabilities (mechamical mixing) and mixing via gravitational instabilities (buoyant mixing). Shear instabilities result from the differential motion of adjacent fluid parcels. An example of such include the passage of internal waves through a

quiescent background, in which wave breaking and/or shear instabilities prompt mixing (see, e.g., Polzin et al. (1997), Toole (1998), and Chapter 6 of Kantha and Clayson (2000b)). This mixing can affect the stratification of deep waters in the World Ocean. Another example is the wind induced mixing occuring in the surface ocean due to breaking surface waves. This mixing is critical for transferring momentum from the atmosphere into the ocean interior. A third example involves the strong vertical shears within the equatorial current system. Here, shears can induce unstable waves that break and thus mix properties vertically and are thought to be essential for the maintenance of the equatorial currents.

Gravitational instabilities occur when buoyancy forcing causes a fluid parcel to become gravitationally unstable. An important example of such occurs when buoyancy is lost to the atmosphere as a winter storm passes over the ocean at high latitudes (see, e.g., Send and Käse (1998), Large (1998), Marshall and Shutts (1999), and Chapter 2 of Kantha and Clayson (2000b)). This process is associated with deep convection and is critical for the formation of water masses in the World Ocean. Another source of gravitational instability is the heating of deep waters by geothermal effects, thus inducing a rising column of fluid from the bottom to its level of neutral buoyancy.

If shear and/or buoyancy forcing are maintained, flow instabilities evolve into a fully developed three-dimensional turbulent state. Such turbulence is characterized by random behaviour where fluid parcels readily mix their properties. Nonlinear interactions afforded by advection of momentum lead to a cascade of kinetic energy to ever smaller scales. Ultimately, kinetic energy introduced by forcing at the relatively large scales is dissipated by molecular viscosity at the Kolmogorov scale, which is the scale where the effects on kinetic energy due to molecular viscosity are comparable to advection. That is, where the Reynolds number defined in terms of the molecular viscosity is on the order of unity. Notably, as three-dimensional turbulence leads to the dissipation of kinetic energy, the associated mixing leads to an increase in potential energy along with small levels of heating.* This behaviour is in contrast to quasi-geostrophic turbulence, whereby potential energy contained in sloped isopycnal surfaces is transferred to kinetic energy via the mechanisms of baroclinic instability (see Section 7.1 for more discussion).

8.3.2 The nonlinear equation of state

Besides mixing associated with mechanical and buoyant mechanisms, *cabbeling*, *thermobaricity*, and the weaker *halobaricity* lead to a nontrivial dianeutral transport when two fluid parcels mix with equal density yet different potential temperature and salinity. As seen in Sections 8.4.4 and 12.1.7, these effects are associated with nonlinearities in the equation of state and lead to the material evolution of buoyancy that is distinctly non-diffusive. McDougall (1987b) provides a more thorough discussion of these effects, whose importance in certain regions of the ocean can rival those from more traditional mixing processes.

*As noted in Section 4.4.2, heating due to frictional dissipation by molecular viscosity is very small in the ocean.

8.3.3 Double diffusion

Dianeutral buoyancy transport also arises when a multi-component fluid, such as seawater, is gravitationally stable yet possesses a nontrivial vertical gradient of its constituents. For seawater, such occurs when vertical temperature and salinity gradients have the same sign, and so contribute oppositely to the vertical density gradient. Due to the differing rates of molecular diffusion for temperature and salinity, mixing can ensue whereby potential energy is decreased. Notably, the resulting density flux is up-gradient, although the temperature and salinity fluxes are each down-gradient. Schmitt (1994), Schmitt (1998), Chapter 7 of Kantha and Clayson (2000b), and Toole and McDougall (2001) provide reviews of such *double diffusive* processes.

8.3.4 Summary of the processes

There is a plethora (a zoo!) of physical processes leading to dianeutral transport throughout the World Ocean. Broadly, large levels of mixing occur in the upper planetary boundary layer (the surface mixed layer) due to the effects of mechanical and buoyant forcing from the atmosphere, sea-ice, and rivers. Large levels also occur in selected regions of the ocean bottom boundary layer, where cascading overflows of heavy fluid enter regions of lighter fluid. The strong flows can setup Kelvin-Helmholz instabilities and/or hydraulic jumps that render a significant level of entrainment with surrounding water as the overflow descends to a level of neutral buoyancy (see, e.g., Price and Baringer (1994) and Beckmann (1998)). Additionally, breaking internal waves scattered from rough topography is now thought to provide for nontrivial levels of mixing extending thousands of meters upward into the fluid column. Incorporating these processes, and others such as cabbeling, thermobaricity, double-diffusion, breaking surface waves, etc., into ocean climate models requires a synergistic relationship between ocean climate modelers, ocean process modelers, and ocean observationalists.

8.4 Dianeutral transport in models

The purpose of this section is to provide some mathematical statements in support the discussion in Sections 8.2 and 8.3. In so doing, we hope to clarify some of the main ideas.

8.4.1 Dianeutral velocity component

In Section 5.7 we introduced the dianeutral velocity component

$$w^{(\rho)} = z_{,\rho} \frac{\mathrm{d}\rho}{\mathrm{d}t} \tag{8.2}$$

with ρ symbolizing the locally referenced potential density, or buoyancy (sans the $-g/\rho_o$ factor). This equation is simply a definition of the dianeutral velocity component $w^{(\rho)}$. In particular, we are not yet obliged to average over the micro-scale fluctuations that lead to a "fuzziness" in surfaces of constant potential density. However, if we wish to use isopycnal coordinates (or more generally, neutral density

coordinates), we must provide such an average, which necessarily leads to a coarsening of the field equations over small-scale mixing processes (see, e.g., deSzoeke and Bennett (1993) and Davis (1994a,b)). We assume in the following that this averaging has been done, which allows us to write the material time derivative in isopycnal coordinates as

$$\frac{\mathrm{d}}{\mathrm{d}t} = \left(\frac{\partial}{\partial t}\right)_\rho + \mathbf{u}\cdot\nabla_\rho + w^{(\rho)}\left(\frac{\partial}{\partial z}\right). \tag{8.3}$$

This expression prompts us to interpret $w^{(\rho)}$ as a dianeutral velocity component measuring the rate at which fluid parcels penetrate a potential density surface (see Section 5.7 for kinematic details leading to this interpretation). When $w^{(\rho)}$ vanishes, the dynamics simplifies since the spatial component of the material time derivative reduces from three to two dimensions. Some kinematic aspects of such quasi two dimensional dynamics form the focus of Chapter 7. We are concerned in the following with physical processes yielding a nontrivial $w^{(\rho)}$, thus leading to the material evolution of buoyancy.

8.4.2 Dianeutral advection of buoyancy

With the dianeutral velocity component $w^{(\rho)}$ nonzero, buoyancy is materially altered via

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = w^{(\rho)}\rho_{,z}. \tag{8.4}$$

This equation follows by definition from equation (8.2). It affords an interpretation of the material evolution of buoyancy according to dianeutral advection. That is, all forms of dianeutral transport of buoyancy are summarized by their effects on $w^{(\rho)}$, and so this perspective highlights the need to parameterize processes such as those mentioned in Section 8.3 that lead to a nonzero $w^{(\rho)}$, as these processes are key to providing a sound simulation of water mass transformations.

Expanding the material time derivative in equation (8.4) leads to the equivalent statement

$$(\partial_t + \mathbf{u}\cdot\nabla_z + w\,\partial_z)\rho = w^{(\rho)}\rho_{,z}. \tag{8.5}$$

Consequently, dianeutral advection $w^{(\rho)}\rho_{,z}$ encompasses in one term what four terms describe in a z-coordinate perspective. This result highlights the fundamental distinction between the vertical velocity component w and the dianeutral velocity component $w^{(\rho)}$. This distinction was also noted from a kinematic perspective in Section 5.7. Hence, vertical advection is fundamentally distinct from dianeutral advection, simply because (see equation (5.71)) $w^{(\rho)} = w - (\partial_t + \mathbf{u}\cdot\nabla_\rho)\,z \neq w$.

8.4.3 The tracer equation with a diffusive closure

Ocean climate models traditionally time step salinity and potential temperature (or the potential enthalpy variable introduced in Section 4.7.2). The evolution of density is then diagnosed. The reason models time step tracers is because small-scale mixing in the ocean can typically be parameterized for tracers as downgradient diffusion. However, as we see here and as mentioned in Section 8.3.2, dianeutral

transport of buoyancy cannot generally be parameterized as diffusion when the equation of state is nonlinear. We expose these points in this section.

To get started, let us write the tracer equation in the form

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -\nabla \cdot \mathbf{F}. \tag{8.6}$$

To arrive at this equation, we must provide a coarsened description so that small scale turbulent mixing processes have been averaged over. The convergence of the flux $\mathbf{F}$ parameterizes the SGS mixing processes.* This coarsening renders, in particular, a smooth buoyancy field thus allowing us to exploit an isopycnal description when convenient.

With small-scale mixing in the ocean associated with three-dimensional turbulence, it is common to assume that a diffusive parameterization is appropriate. We argued such in Section 8.1.2, since tracers are conserved properties and so are analogous to the Brownian particle. This assumption allows us to write the tracer equation (8.6) as

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \nabla \cdot (\mathbf{K} \cdot \nabla C) \tag{8.7}$$

where $\mathbf{K}$ is a second order diffusion tensor (positive semi-definite and symmetric). Mathematical properties of the diffusion tensor are discussed in Section 11.4, with further focus on the neutral diffusion tensor in Section 12.1. For present purposes, we note that components to this tensor can generally be written in a projection operator form (see equation (12.13))

$$K^{mn} = A_I(\delta^{mn} - \hat{\boldsymbol{\rho}}^m \hat{\boldsymbol{\rho}}^n) + A_D \hat{\boldsymbol{\rho}}^m \hat{\boldsymbol{\rho}}^n. \tag{8.8}$$

In this equation, A_I and A_D are non-negative diffusivities associated with parameterized tracer mixing along and across neutral directions, and $\hat{\boldsymbol{\rho}}$ is a unit direction perpendicular to a neutral direction. For diffusion parameterizing three-dimensional isotropic turbulence, it makes sense to set $A_I = A_D$. However, as indicated in our discussion of neutral directions in Section 8.2.1, it is useful to anticipate an anisotropy between along and across neutral directions for processes affected by buoyancy, which generally leads to $A_I > A_D$.

Due to large inhomogenities in unresolved dianeutral processes, the diffusivity A_D exhibits a widely varying space-time dependence. For example, in the quiescent ocean interior and pycnocline, measurements indicate that $A_D \approx 10^{-5}\,\mathrm{m}^2\,\mathrm{s}^{-1}$ (e.g., Ledwell et al. (1993), Toole et al. (1994), and Kunze and Sanford (1996)), whereas A_D is enhanced one to three orders of magnitude above rough topography due to breaking internal waves (Polzin et al. (1997)) or near rapid changes in bottom topographic slope (Toole et al. (1994), Toole et al. (1997), Polzin et al. (1996)). The diffusivity is also very large in various boundary layers, such as the surface and bottom mixed layers. Rationalizing and parameterizing various regimes of turbulent transport are central areas of research.

Averaging over small-scale motions so that neutral directions are well defined leads to neutral slopes somewhat smaller than 1/100 in the bulk of the ocean. Consequently, it is common to approximate the dianeutral diffusive contribution to the

*Large et al. (1994) introduced to the ocean modeling community a non-local source term to parameterize non-diffusive dianeutral transport processes, especially those occurring in the ocean's upper planetary boundary layer. We are not considering such effects in the present discussion.

diffusion tensor (8.8) as

$$A_D \hat{\boldsymbol{\rho}}^m \hat{\boldsymbol{\rho}}^n \approx A_D \hat{\mathbf{z}}\,\hat{\mathbf{z}}. \tag{8.9}$$

As shown in Section 12.1.4, these two expressions differ to within terms going as $A_D\, S^2$, which is quite small for slopes $S \leq 1/100$. Hence, the use of vertical diffusion to represent dianeutral diffusion is very accurate where neutral slopes are small, which includes most regions outside of boundary layers. A similar approximation for the along-slope diffusive fluxes, also detailed in Section 12.1.4, allows for

$$\nabla \cdot (\mathbf{K} \cdot \nabla C) \approx \nabla_\rho \cdot (A_I\, \nabla_\rho\, C) + \partial_z\, (A_D\, C_{,z}), \tag{8.10}$$

where neglected neutral and dianeutral terms are proportional to $(A_I, A_D)\, S^2$, respectively, and

$$\nabla_\rho = \nabla_z + \mathbf{S}\, \partial_z \tag{8.11}$$

is the lateral gradient along a neutral direction (see Section 5.5.3 for a derivation), with $\mathbf{S}$ the slope of the neutral direction given by $\mathbf{S} = \nabla_\rho z$. Consequently, the tracer equation under a diffusive closure takes the form

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \partial_z\, (A_D\, C_{,z}) + \nabla_\rho \cdot (A_I\, \nabla_\rho\, C). \tag{8.12}$$

8.4.4 Evolution of locally referenced potential density

With potential temperature and salinity satisfying an evolution equation of the form (8.6) with a diffusive closure for the small-scale mixing processes, we can deduce the form for the evolution of locally referenced potential density

$$\begin{aligned} \frac{\mathrm{d}\rho}{\mathrm{d}t} &= \rho_{,\theta}\, \frac{\mathrm{d}\theta}{\mathrm{d}t} + \rho_{,s}\, \frac{\mathrm{d}s}{\mathrm{d}t} \\ &= -\nabla \cdot (\rho_{,\theta}\, \mathbf{F}^{(\theta)} + \rho_{,s}\, \mathbf{F}^{(s)}) + \mathbf{F}^{(\theta)} \cdot \nabla \rho_{,\theta} + \mathbf{F}^{(s)} \cdot \nabla \rho_{,s} \end{aligned} \tag{8.13}$$

where $\rho_{,\theta} = \partial\rho/\partial\theta$ and $\rho_{,s} = \partial\rho/\partial s$ are the thermal and saline partial derivatives of the locally referenced potential density. The absence of a $\rho_{,p}\, \mathrm{d}p/\mathrm{d}t$ contribution to $\mathrm{d}\rho/\mathrm{d}t$ is due to our focus on locally referenced potential density, rather than *in situ* density. Even so, $\rho_{,\theta}$ and $\rho_{,s}$ are affected by pressure deviations when considering their spatial gradients

$$\nabla \rho_{,\theta} = \rho_{,\theta\theta} \nabla\theta + \rho_{,\theta s} \nabla s + \rho_{,\theta p} \nabla p \tag{8.14}$$

$$\nabla \rho_{,s} = \rho_{,s\theta} \nabla\theta + \rho_{,ss} \nabla s + \rho_{,sp} \nabla p. \tag{8.15}$$

The convergence $-\nabla \cdot (\rho_{,\theta}\, \mathbf{F}^{(\theta)} + \rho_{,s}\, \mathbf{F}^{(s)})$ in equation (8.13) represents the convergence of the diffusive flux of locally referenced potential density

$$\mathbf{F}^{(\rho)} = \rho_{,\theta}\, \mathbf{F}^{(\theta)} + \rho_{,s}\, \mathbf{F}^{(s)}. \tag{8.16}$$

Notably, the contribution to $\mathbf{F}^{(\rho)}$ from diffusion of potential temperature and salinity along neutral directions vanishes, as it must by definition. That is, one cannot diffuse buoyancy along a surface of constant buoyancy. This basic point is emphasized in Section 12.1.6, where we note that a numerical implementation of neutral diffusion in a z-model must respect this property in order to remain numerically

stable. Hence, when considering small-scale mixing parameterized as diffusion, the flux of buoyancy reduces to just a vertical diffusive contribution

$$\mathbf{F}^{(\rho)} = -A_D \, \rho_{,z} \, \hat{\mathbf{z}}. \tag{8.17}$$

The term $\mathbf{F}^{(\theta)} \cdot \nabla \rho_{,\theta} + \mathbf{F}^{(s)} \cdot \nabla \rho_{,s}$ appearing in equation (8.13) does not take the form of a flux convergence. Hence, it is best thought of as a source term affecting buoyancy. It is nontrivial only when the second partial derivatives of density do not all vanish, and so they account for the nonlinear equation of state processes introduced in Section 8.3.2. Manipulations in Section 12.1.7 illustrate how to identify the source terms arising from cabbeling, thermobaricity, and halobaricity, which are associated with temperature and salinity diffusion along neutral surfaces. A more complete presentation is given by McDougall (1987a,b).

In summary, small scale mixing parameterized as tracer diffusion leads to the material evolution of buoyancy

$$\begin{aligned} \frac{\mathrm{d}\rho}{\mathrm{d}t} &= w^{(\rho)} \, \rho_{,z} \\ &= \partial_z \left(A_D \, \rho_{,z} \right) + \mathbf{F}^{(\theta)} \cdot \nabla \rho_{,\theta} + \mathbf{F}^{(s)} \cdot \nabla \rho_{,s}. \end{aligned} \tag{8.18}$$

The first term on the right hand side represents the effects on buoyancy from dianeutral diffusion of potential temperature and salinity. The second term represents the effects due to the nonlinear equation of state from seawater. These two terms can be on the same order of magnitude in certain parts of the ocean (e.g., McDougall (1987a,b)). Nonetheless, many discussions in the literature ignore the nonlinear equation of state terms, and so the reader should be mindful of this limitation in those discussions.

Equation (8.18) provides a useful venue to highlight how certain ocean models provide a more or less faithful numerical representation of these processes. For this purpose, we focus on the distinctions between isopycnal models and z-models, with sigma-models sharing much in common with z-models in the present discussion (see Section 5.2 for a presentation of these models). For the dianeutral diffusive process, z-models trivially incorporate this via linear vertical diffusion of potential temperature and salinity, and they have no problem allowing for arbitrarily large vertical diffusivities, so long as the time stepping is handled implicitly. In contrast, diapycnal diffusion is a nonlinear process in isopycnal models, and details of how to stably handle large diffusivities have only recently been clarified by Hallberg (2000). Z-models incorporate effects from the nonlinear equation of state simply by using a realistic equation for seawater density, such as the one recently proposed by McDougall et al. (2002b). In contrast, these effects are nontrivial to incorporate into isopycnal models, although successful methods are now available such as those from Sun et al. (1999). Finally, the adiabatic stirring effects arising from advective processes (appearing when the material time derivative is represented in its Eulerian form) are more cleanly handled by isopycnal models, thus accounting for their use in many idealized studies of the adiabatic ocean circulation. We have more to say in these regards in Sections 8.5 and 11.1.

8.5 Numerically induced spurious dianeutral transport

The huge anisotropy between neutral and dianeutral mixing processes in the ocean places a heavy burden on the numerical integrity of an ocean model. In particular, one cannot be too cavalier about how both resolved and parameterized transport is realized in ocean models, since doing so risks not properly respecting the anisotropy. Until the past decade or so, physical levels of anisotropy in the ocean interior were not respected by the most common class of ocean climate models, the z-models which form the focus of later chapters in this book. However, recent advances documented in Part IV have tremendously improved the integrity of coarse resolution models.

Nonetheless, potential problems remain for refined models admitting mesoscale eddies. As documented by Griffies et al. (2000b), since mesoscale eddies transfer variance to the grid scale, this variance must be dissipated. Otherwise, eddying simulations will degenerate into a sea of grid noise. Unfortunately, common methods for absorbing this variance introduce unphysically large levels of spurious dianeutral mixing, and this mixing can swamp the levels introduced by the physically based SGS parameterizations. A possible means for resolving the spurious mixing problem is mentioned in Section 11.1, where it is argued that the same neutral physics operators physically motivated for coarse resolution models also serve as useful numerical closure operators at the refined scales.

Additional problems with spuriuos mixing occur in z-models next to its step-like topography. As noted by Beckmann (1998), much of the emphasis in simulating overflow processes in z-models is focused in providing a more suitable representational framework, rather than in parameterizing physical processes. That is, although there are physical reasons to expect enhanced mixing near the ocean bottom, z-models can spuriously provide far more than the real ocean. Hence, consistent with our discussion in Section 5.2, the z-model is cumbersome for the overflow problem, whereas the sigma-model and isopycnal-model have been found to be far more suitable.

As noted in Section 5.2, isopycnal models using a linear equation of state trivially respect the large anisotropy between neutral and dianeutral processes. This point is highlighted in the review article by Bleck (1998). However, isopycnal models employing a realistic nonlinear equation of state must also contend with potentially unphysical levels of diapycnal advection. In this case, when isopycnal models advectively transport temperature and salinity along isopycnal surfaces, numerical isopycnal mixing associated with the advection schemes can introduce unphysically large amounts of cabbeling (Eric Chassignet, personal communication 2002). This issue is also relevant in z-models. It therefore appears that all model classes must contend with some level of spurious dianeutral processes. Algorithm developers must seek more refined methods which reduce the spurious transport to physically acceptable, if not trivial, levels.

Part III

SEMI-DISCRETE Z-EQUATIONS

SEMI-DISCRETE Z-EQUATIONS

Given the continuum version of the mean-field ocean equations, we here cast these equations onto a discrete lattice. What lattice should we choose? How should one discretize the various terms appearing in the equations, in both space and time? How can we best ensure that the numerical fluid looks like the ocean? These are questions which ocean modelers continually ask, and there are few unambiguous answers.

In this and the remaining parts of this book, we expose our choice to use a fixed Eulerian lattice appropriate for a z-model. By making this choice we recognize that it introduces many egregious problems. The next decade will undoubtedly see research into alternatives, and many promising ones exist today. Nonetheless, z-models are the most mature and popular choice for those aiming to simulate the global ocean climate. This in itself warrants a careful rationalization of their basic equations.

This part of the book aims to perform the first steps towards obtaining a fully discretized set of equations appropriate for a z-model. The method used is to integrate the continuum equations over a discrete model grid cell. Formally, this approach is termed a *finite volume* approach. We are far from rigorous, however, in that we choose to interpret our discrete variables in a manner that accords with our chosen numerical schemes, some of which are finite-volume based, some of which are not. Nonetheless, the finite volume approach is an effective means to derive the semi-discrete finite differenced equations, and to expose some of the important points about boundary conditions.

This part of the book considers in sequence the semi-discrete equations for mass, tracers, and momentum. We present algorithms for solving these equations numerically. In particular, the momentum equation chapter presents a particularly popular means to time step the equations of motion forward, that being the *split-explicit* free surface method. Alternative methods exist, with this one providing a pedagogical example of issues arising in many of the approaches.

CHAPTER
NINE

Semi-discrete mass and tracer budgets

Contents

9.1 Summary of the continuous model equations 203
9.2 Tracer and mass/volume compatibility 204
9.3 Mass budget for a grid cell 205
9.3.1 Diagnosing vertical velocity components 205
9.3.2 Advection velocity components 207
9.3.3 Specifying the density time tendency 207
9.4 Mass budget for a discrete fluid column 207
9.5 Tracer budget for a grid cell 209
9.5.1 Vertical integration over a grid cell 210
9.5.2 Vertical tracer flux at the ocean surface 211
9.5.3 Tracer budgets for surface grid cells 211
9.5.4 Tracer concentration in fresh water 212
9.5.4.1 Neutral tracers 212
9.5.4.2 Salt 212
9.6 Fluxes for turbulence mixed layer schemes 213
9.7 Flux plus restore boundary conditions 214

This chapter discusses the semi-discrete mass and tracer budgets appropriate for a z-coordinate ocean model with a free surface. Particular attention is paid to details of the surface grid cell, where the undulating free surface allows the volume of these cells to fluctuate in time. Discretization in time is assumed to occur using a leap-frog scheme as commonly done in z-models such as Griffies et al. (2003).

9.1 Summary of the continuous model equations

The equations considered in this chapter represent discrete realizations of the continuum model equations for conservation of tracer and mass presented in Section

6.8. The continuous non-Boussinesq mass and tracer equations are

$$\rho_{,t} + \rho_o \nabla \cdot \mathbf{v} = 0 \tag{9.1}$$

$$(D\,\overline{\rho}^z)_{,t} = -\rho_o \nabla \cdot \mathbf{U} + \rho_o\, q_w \tag{9.2}$$

$$(\rho\, C)_{,t} + \rho_o \nabla \cdot (\mathbf{v}\, C) = -\rho_o \nabla \cdot \mathbf{F} + \rho\, \mathcal{S}. \tag{9.3}$$

Finite domains have the associated surface and bottom kinematic boundary conditons

$$\rho\, \eta_{,t} + \rho_o\, \mathbf{u} \cdot \nabla \eta = \rho_o\, w + \rho_o\, q_w \qquad \text{at } z = \eta \tag{9.4}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H. \tag{9.5}$$

In these equations,

$$\rho\, \mathbf{v}^\rho = \rho_o\, \mathbf{v} \tag{9.6}$$

is the linear momentum density of a fluid parcel, and

$$D = H + \eta \tag{9.7}$$

is the thickness of a column of fluid extending from the surface at $z = \eta$ to the bottom at $z = -H$. The Boussinesq equations are recovered by setting $\rho \to \rho_o$

$$\nabla \cdot \mathbf{v} = 0 \tag{9.8}$$

$$\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w \tag{9.9}$$

$$C_{,t} + \nabla \cdot (\mathbf{v}\, C) = -\nabla \cdot \mathbf{F} + \mathcal{S} \tag{9.10}$$

with surface and bottom kinematic boundary conditions

$$\eta_{,t} + \mathbf{u} \cdot \nabla \eta = w + q_w \qquad \text{at } z = \eta \tag{9.11}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H. \tag{9.12}$$

The interpretation of the model fields as ensemble means is given by Table 6.1. For models not resolving the mesoscale, the subgrid-scale closure flux $\mathbf{F}$ is interpreted according to the isentropic ensemble analysis given in Chapter 7.

9.2 Tracer and mass/volume compatibility

The tracer equation (9.3) represents a combination of mass and tracer conservation. In the special case where the tracer is set to a uniform constant, the tracer equation reduces to the mass continuity equation (9.1). Similarly, for a Boussinesq fluid, the tracer equation (9.10) reduces to volume continuity (9.8). Such compatibility must also be maintained over vertical columns of fluid, where the column integrated mass conservation equation (9.2) must be compatible with the column integrated non-Boussinesq tracer budget. Likewise, for a Boussinesq fluid, the column integrated volume conservation equation (9.9) must be compatible with the column integrated Boussinesq tracer budget.

The importance of maintaining such compatibility conditions was emphasized by Griffies et al. (2001) when describing the implementation of a free surface method for a Boussinesq fluid in a z-coordinate model. Time dependent thicknesses for

the surface cells called for extra care in the discretizations. Compatibility between the tracer and mass/volume budgets means that, for example, an initially uniform tracer concentration will not change in the absence of sources, even if the surface height changes. Griffies et al. (2001) referred to this compatibility condition as a *local conservation* property, to distinguish it from global conservation which provides for constancy of the globally integrated tracer content in the absence of external sources.

Compatibility between tracer and mass/volume budgets means that the time stepping scheme used for the tracer and surface height must be the same. In the case of MOM4, this means that the leap-frog discrete tracer equation leads to a leap-frog discrete surface height equation. This is the reason we employ a "big-leap-frog" for updating the surface height in Section 10.5 (equation (10.71)). It also means that mass continuity must be updated using a leap-frog scheme, as discussed in Section 9.3.

9.3 Mass budget for a grid cell

We start our development by vertically integrating volume and mass conservation over the thickness of a grid cell. The same approach is used in Section 9.5 for developing the discrete tracer budget and Section 10.2 for the discrete momentum budget. Mass conservation, although involving a time tendency, is implemented in a diagnostic manner to determine vertical advective velocities. For the Boussinesq fluid, similar diagnostic relations determine the vertical velocity, although the density time tendency term is absent.

In the following, we refer to a schematic of surface cells given in Figure 9.1. This figure is modified slightly from the free surface paper of Griffies et al. (2001). Additionally, we focus on the non-Boussinesq budget, with the Boussinesq budget trivially recovered by setting appropriate density factors to the constant reference value ρ_o.

9.3.1 Diagnosing vertical velocity components

For Boussinesq fluids, the vertical velocity component is diagnosed from vertically integrating the non-divergence condition $\partial_z w = -\nabla \cdot \mathbf{u}$. As the continuity equation is first order, it requires a single boundary condition for its integration. There are two equivalent approaches: integrating from the top downward to the level of interest, and integrating from the bottom upwards. Integration downwards has been the common approach in many B-grid z-models since this allows for a useful diagnostic whereby the diagnosed bottom velocity on a tracer cell must vanish, since this cell has a flat surface (see Pacanowski and Griffies (1999) for details). Non-zero values help to identify coding errors. To diagnose the vertical velocity component in a mass conserving ocean model, it is necessary to integrate mass conservation $\rho_{,t} + \rho_o \nabla \cdot \mathbf{u} + \rho_o \, \partial_z w = 0$ vertically over a grid cell. Again there are two approaches: integrating downwards or integrating upwards, with the downwards approach again preferred.

Let us now consider the details. Integration of mass conservation over a cell

interior to the ocean ($k > 1$), whose volume is constant in time, yields

$$
\begin{aligned}
0 &= \int_{z_k}^{z_{k-1}} dz\,(\rho_{,t} + \rho_o\,\nabla\cdot\mathbf{v}) \\
&= \partial_t \int_{z_k}^{z_{k-1}} dz\,\rho + \rho_o\,\nabla\cdot\int_{z_k}^{z_{k-1}} dz\,\mathbf{u} + \rho_o\,(w_{z_{k-1}} - w_{z_k}) \\
&\rightarrow \partial_t\,(h\,\rho)_{z_k} + \rho_o\,\nabla\cdot(h\,\mathbf{u})_{z_k} + \rho_o\,(w_{z_{k-1}} - w_{z_k}),
\end{aligned}
\tag{9.13}
$$

where $h_k = z_{k-1} - z_k$ is the thickness of the tracer cell (the cell on which density is naturally defined). The continuous to discrete correspondence relation in the last step can be made by interpreting the discrete model variables as averages of the continuous variables over the grid cell. Solving for w_{z_k} leads to the iterative relation for vertical advective velocity passing across the lower boundary of the $k'th$ interior grid cell

$$
\rho_o\,w_{z_k} = \rho_o\,w_{z_{k-1}} + h_{z_k}\,\partial_t\,\rho_{z_k} + \rho_o\,\nabla\cdot(h\,\mathbf{u})_{z_k} \tag{9.14}
$$

The time tendency term is absent for the Boussinesq fluid, where mass conservation over the cell reduces to volume conservation.

Vertical integration over a surface ocean grid cell ($k = 1$) leads to

$$
\begin{aligned}
\int_{z_1}^{\eta} dz\,(\rho_{,t} + \rho_o\,\nabla\cdot\mathbf{v}) &= \partial_t \int_{z_1}^{\eta} dz\,\rho + \rho_o\,\nabla\cdot\int_{z_1}^{\eta} dz\,\mathbf{u} - \rho_o\,w_{z_1} - (\rho\,\eta_{,t} - \rho_o\,\widehat{\mathbf{N}}\cdot\mathbf{v})_{z_0} \\
&\rightarrow \partial_t\,(h\,\rho)_{z_1} + \rho_o\,\nabla\cdot(h\,\mathbf{u})_{z_1} - \rho_o\,w_{z_1} - \rho_o\,q_w,
\end{aligned}
\tag{9.15}
$$

where $z_0 = \eta$,

$$
\widehat{\mathbf{N}} = (-\nabla\eta, 1) \tag{9.16}
$$

is an outward pointing vector at the ocean surface, and the surface kinematic boundary condition (9.4) was used. To maintain self-consistency with the discrete mass budget, we time step density instead of thickness weighted density, thus leading to the vertical advective velocity passing upwards into the surface grid cell

$$
\rho_o\,w_{z_1} = -\rho_o\,q_w + h_{z_1}\,\partial_t\,\rho_{z_1} + \rho_{z_1}\,\eta_{,t} + \rho_o\,\nabla\cdot(h\,\mathbf{u})_{z_1}. \tag{9.17}
$$

For the Boussinesq case, the $\partial_t\,\rho_{z_1}$ term is dropped and all density factors are set to ρ_o. The iterative equation (9.14) can be used for all vertical cells by defining the vertical advective velocity w_{z_0} according to

$$
\rho_o\,w_{z_0} = -\rho_o\,q_w + \rho_{z_1}\,\eta_{,t} \tag{9.18}
$$

for the non-Boussinesq case, and

$$
w_{z_0} = -\nabla\cdot\mathbf{U} \tag{9.19}
$$

for the Boussinesq case.

9.3.2 Advection velocity components

Equations (9.14) and (9.17) are semi-discrete. A full discretization requires averaging over the grid cell. Doing so results in all velocity components being averaged onto the sides and top faces of the grid cell. One then has a full specification of the convergence of mass onto the cell, which drives time tendencies in the grid cell's density.

When averaged onto cell faces, the velocity components are known as *advective velocities*, since they are used for advecting tracers (Section 9.5) or momenta. On the B-grid, a detailed form of the advective velocity components for tracers can be prescribed according to the needs of energetic consistency as detailed in Griffies et al. (2003). Furtheremore, the relation between advective velocities for tracers and momentum can be specified according to a liner mapping, as discussed in Pacanowski and Griffies (1999) and Griffies et al. (2003). Notably, the same discretization of the advective velocities can be used whether the discrete fluid is Boussinesq or non-Boussinesq.

9.3.3 Specifying the density time tendency

As the mass continuity equation is implemented in a diagnostic mode, we must specify the time tendency for the density. A prescription successfully used in MOM4 is given by the leap-frog scheme

$$\partial_t \rho \approx \frac{\rho^{(e)}(\tau + \Delta\tau) - \rho^{(e)}(\tau - \Delta\tau)}{2\,\Delta\tau}, \tag{9.20}$$

where $\rho^{(e)}$ is a density field set according to linear extrapolation (Greatbatch et al. (2001))

$$\rho^{(e)}(\tau + \Delta\tau) \equiv \rho(\tau) + [\rho(\tau) - \rho(\tau - \Delta\tau)] = 2\,\rho(\tau) - \rho(\tau - \Delta\tau). \tag{9.21}$$

Although it adds to the model memory requirements, it is useful to explicitly carry the density variable $\rho^{(e)}$ in the model since it is used very often with the non-Boussinesq formulation.

9.4 Mass budget for a discrete fluid column

As seen in Section 9.3.1, the discrete fluid conserves mass by construction, since the vertical velocity components are derived by integrating mass conservation over a grid cell. Additionally, it is important to maintain mass conservation over a column of fluid extending from the ocean surface to the bottom. As in the continuum case discussed in Section 3.3.3, this constraint leads to an expression for the time evolution of the surface height. We now pursue a similar analysis for the discrete ocean fluid.

For this purpose, recall from Section 9.3.1 that the discrete vertical velocity components take the form

$$\rho_o\, w_{z_k} = \rho_o\, w_{z_{k-1}} + h_{z_k}\, \partial_t\, \rho_{z_k} + \rho_o\, \nabla \cdot (h\, \mathbf{u})_{z_k} \qquad k > 1 \tag{9.22}$$

$$\rho_o\, w_{z_1} = -\rho_o\, q_w + h_{z_1}\, \partial_t\, \rho_{z_1} + \rho_{z_1}\, \eta_{,t} + \rho_o\, \nabla \cdot (h\, \mathbf{u})_{z_1} \qquad k = 1. \tag{9.23}$$

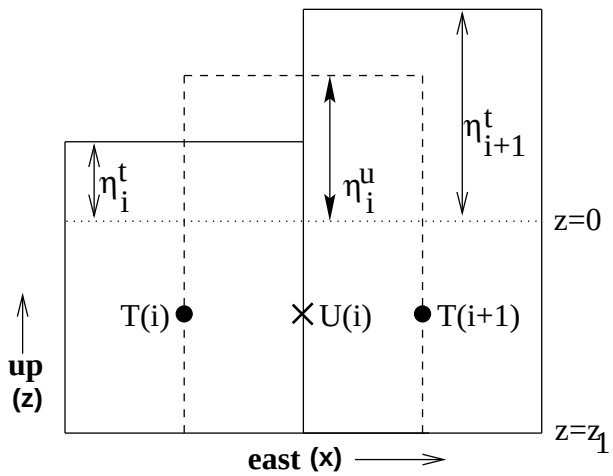

Figure 9.1: Schematic of rectangular surface tracer and velocity cells with a free surface z-coordinate model. This figure is modified slightly from the free surface paper of Griffies et al. (2001). Tracer points are denoted by a solid dot and the tracer cells are enclosed by solid lines; a velocity point is denoted by an "x" and is enclosed by dashed lines. These points have fixed vertical position $z = z_1/2 < 0$, *regardless* of the value for the free surface height. The thickness of a tracer cell is $h^t = -z_1 + \eta^t$, where η^t is the prognostic surface height determined through volume conservation. The thickness of a velocity cell is $h^u = -z_1 + \eta^u$, where η^u is determined by a conservative average of the surrounding η^t values. Thicknesses of the cells are assumed to be positive, so that $-z_1 + \eta^t > 0$. In models without an explicit representation of tides, this constraint in practice means that $|z_1|$ must be greater than roughly 2m. For models with tides, $|z_1|$ may need to be somewhat larger, depending on the tidal range considered. Additionally, when coupling to sea-ice, the weight of the sea-ice will depress the ocean surface by an amount roughly equivalent to the ice thickness. This process thus places another constraint on the thickness of the top model grid cell, or on the thickness of the sea-ice. In either of these cases (tides or sea ice), the limitations of the model cell thickness places an awkward constraint on the model's ability to run with fine vertical resolution. This represents a fundamental limitation of the z-coordinate models.

A vertical sum of these velocities over the depth of a discrete column with N_k vertical grid cells leads to an expression for the vertical velocity at the bottom of the bottom-most grid cell

$$\rho_o \, w_{z_{N_k}} = -\rho_o \, q_w + \rho_o \, \nabla \cdot \mathbf{U} + \rho_{z_1} \, \eta_{,t} + \sum_{k=1}^{N_k} h_{z_k} \, \partial_t \, \rho_{z_k} \tag{9.24}$$

where $\mathbf{U} = \sum_{k=1}^{N_k} h_{z_k} \, \mathbf{u}_{z_k}$ is the vertically integrated horizontal velocity. For the B-grid used in MOM4, the land-sea boundary on the bottom of a column of tracer cells is assumed flat (see Section 22.3.3 of Pacanowski and Griffies (1999) for more discussion). Hence, the bottom kinematic boundary condition leads to

$$w_{z_{N_k}} = 0 \qquad \text{at the bottom of a bottom-most tracer cell.} \tag{9.25}$$

This boundary condition is maintained for a discrete mass conserving fluid so long as the surface height at the top of the tracer cell column satisfies

$$\rho_{z_1} \, \eta_{,t} = -\rho_o \, \nabla \cdot \mathbf{U} + \rho_o \, q_w - \sum_{k=1}^{N_k} h_{z_k} \, \partial_t \, \rho_{z_k}. \tag{9.26}$$

This expression for mass balance of a fluid column is directly analogous to the continuum result given by equation (3.38). Additionally, by noting that interior vertical cell thicknesses are constant in time, we have

$$\partial_t \left(\sum_{k=1}^{N_k} h_{z_k} \, \rho_{z_k} \right) = -\rho_o \, \nabla \cdot \mathbf{U} + \rho_o \, q_w, \tag{9.27}$$

which is directly analogous to the continuum mass balance given by equation (3.37). For a Boussinesq fluid we have

$$\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w \tag{9.28}$$

which is the same form as the continuum volume balance given by equation (3.22). The discrete form of these budgets provides the means to time step the surface height. Details of this time stepping are pursued in Section 10.5. Finally, before closing this subsection, we note that the expression (9.26) for $\eta_{,t}$ leads the relation (9.18) for $\rho_o \, w_{z_0}$ to take the form

$$w_{z_0} = -\nabla \cdot \mathbf{U} - \rho_o^{-1} \sum_{k=1}^{N_k} h_{z_k} \, \partial_t \, \rho_{z_k}, \tag{9.29}$$

which reduces to

$$w_{z_0} = -\nabla \cdot \mathbf{U} \tag{9.30}$$

in the Boussinesq limit.

9.5 Tracer budget for a grid cell

We now develop the budget for tracer mass per volume within a discrete model grid cell. As for the discrete mass conservation budget, we vertically integrate the continuous tracer budget over the depth of a model grid cell. Special care is taken when considering the surface boundary conditions.

9.5.1 Vertical integration over a grid cell

We integrate the continuous tracer equation $(\rho\, C)_{,t} + \rho_o\, \nabla \cdot (\mathbf{v}\, C) = -\rho_o\, \nabla \cdot \mathbf{F} + \rho\, \mathcal{S}$ vertically over the depth of a model tracer grid cell. Firstly, the source term is interpreted as

$$\int_{z_k}^{z_{k-1}} dz\, \rho\, \mathcal{S} \rightarrow (h\, \rho\, \mathcal{S})_{z_k}. \tag{9.31}$$

For an interior grid cell, the material transport term integrates to

$$\int_{z_k}^{z_{k-1}} dz\, [\, (\rho\, C)_{,t} + \rho_o\, \nabla \cdot (\mathbf{v}\, C)] \\ \rightarrow \partial_t (h\, \rho\, C)_{z_k} + \rho_o\, \nabla \cdot (h\, \mathbf{u}\, C)_{z_k} + \rho_o\, (w\, C)_{z_{k-1}} - \rho_o\, (w\, C)_{z_k}. \tag{9.32}$$

The advective tracer fluxes are computed on the sides of the cell, with their convergence contributing to the time tendency of the tracer value within the cell.

Equation (9.32) is semi-discrete, with a full discretization of the thickness weighted advective fluxes depending on details of the chosen algorithm. Additionally, the thickness weighting of the horizontal fluxes can be implemented according to Adcroft et al. (1997) or Pacanowski and Gnanadesikan (1998) to account for the generally different adjacent cell thicknesses. Vertical integration of the SGS flux term over an interior grid cell leads to

$$\int_{z_k}^{z_{k-1}} dz\, \nabla \cdot \mathbf{F} \rightarrow \nabla \cdot (h\, \mathbf{F}^h)_{z_k} + (F^z_{z_{k-1}} - F^z_{z_k}), \tag{9.33}$$

where the three-dimensional SGS tracer flux is given by $\mathbf{F} = (\mathbf{F}^h, F^z)$. Thickness weighting of these fluxes in a discrete model is also described by Pacanowski and Gnanadesikan (1998).

For a surface model grid cell, vertical integration leads to

$$\int_{z_1}^{\eta} dz\, [\, (\rho\, C)_{,t} + \rho_o\, \nabla \cdot (\mathbf{v}\, C)] = \\ \partial_t \int_{z_1}^{\eta} dz\, (\rho\, C) + \rho_o \nabla \cdot \int_{z_1}^{\eta} dz\, (\mathbf{u}\, C) - \rho_o\, (w\, C)_{z_1} - C_{z_0}\, (\rho\, \eta_{,t} - \rho_o\, \widehat{\mathbf{N}} \cdot \mathbf{v})_{z_0} \\ \rightarrow \partial_t (h\, \rho\, C)_{z_1} + \rho_o\, \nabla \cdot (h\, \mathbf{u}\, C)_{z_1} - \rho_o\, (w\, C)_{z_1} - \rho_o\, q_w\, C_{z_0}, \tag{9.34}$$

where the surface kinematic boundary condition (9.4) was used. Similar considerations for the divergence of the SGS flux lead to

$$\int_{z_1}^{z_0} dz\, \nabla \cdot \mathbf{F} \rightarrow \nabla \cdot (h\, \mathbf{F}^h)_{z_1} - F^z_{z_1} + (\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0}. \tag{9.35}$$

9.5.2 Vertical tracer flux at the ocean surface

The tracer flux is generally comprised of an advective and a SGS turbulent contribution

$$\mathbf{F}_C = \mathbf{v}\, C + \mathbf{F}. \tag{9.36}$$

In particular, the tracer flux crossing the ocean surface is given by

$$(\mathbf{F}_C \cdot \widehat{\mathbf{N}})_{z_0} = -q_w\, C_{z_0} + (\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0}. \tag{9.37}$$

This flux has a positive sign when tracer leaves the ocean.

In contrast to an interior vertical flux, ocean-only information is not sufficient to estimate the individual terms in the surface tracer flux (9.37). Instead, as for the surface momentum flux discussed in Section 10.2.4, the system is closed by prescribing a total tracer flux crossing the ocean surface via a boundary layer model or parameterization. Boundary layer models present the ocean with a total tracer flux Q_C, which is the tracer flux crossing the ocean surface from other component models, such as the atmosphere, river, and sea ice models. Q_C can also be specified from data, or determined via damping to some specified tracer concentration.

The flux Q_C generally has a contribution from parameterized turbulence as well as a tracer flux with fresh water,

$$Q_C = -q_w\, C_w + Q_C^{turb}, \tag{9.38}$$

where C_w is the tracer concentration in the fresh water. Hence, assuming a continuous flux at the ocean surface $Q_C = (\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0}$ leads to

$$-q_w\, C_{z_0} + (\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = -q_w\, C_w + Q_C^{turb}. \tag{9.39}$$

In the absence of fresh water forcing,

$$(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = Q_C^{turb} \qquad \text{if } q_w = 0. \tag{9.40}$$

However, in general it is not always appropriate to equate the individual terms in equation (9.39). In particular, C_w and C_{z_0} are not always the same. Salt provides the most striking example, where salinity of fresh water, s_w, is near zero yet the salinity on the ocean side of the ocean surface, s_{z_0}, is not generally zero. However, as seen below, other tracers, termed *neutral tracers*, do allow for equating the individual terms.

9.5.3 Tracer budgets for surface grid cells

Given the above prescription for the ocean surface tracer flux, the tracer budget within a surface grid cell takes the form

$$\begin{aligned} h_{z_1}\, \partial_t (\rho\, C)_{z_1} = -\rho_o\, \nabla \cdot (h\, \mathbf{u}\, C + h\, \mathbf{F}^h)_{z_1} + \rho_o\, (w\, C)_{z_1} + (\rho_o\, F^z)_{z_1} - \eta_{,t}\, (\rho\, C)_{z_1} \\ + \rho_o\, q_w\, C_w - \rho_o\, Q_C^{turb} + (h\, \rho\, \mathcal{S})_{z_1}. \end{aligned} \tag{9.41}$$

We separated out the time tendency for the surface height as we are interested in time stepping the density weighted tracer $\rho\, C$ instead of the thickness and density

weighted tracer $h\,\rho\,C$. The discrete value for the time tendency $\eta_{,t}$ is given by the leap-frog expression (10.71) for the non-Boussinesq fluid, and (10.72) for the Boussinesq fluid. Note that boundary condition information which closes the surface cell tracer budget is knowledge of the tracer concentration in fresh water, C_w, as well as the parameterized turbulent tracer flux Q_C^{turb}. Neither the tracer concentration at the ocean surface, C_{z_0}, nor the turbulent flux $(\mathbf{F}\cdot\widehat{\mathbf{N}})_{z_0}$ are explicitly needed.

The budget for a Boussinesq fluid simplifies by cancelling all density factors and using the surface height equation $\eta_{,t} = -\nabla\cdot\mathbf{U} + q_w$ to yield

$$h_{z_1}\,\partial_t C_{z_1} = -\nabla\cdot(h\,\mathbf{u}\,C + h\,\mathbf{F}^h)_{z_1} + (w\,C)_{z_1} + F^z_{z_1} \\ + q_w\,(C_w - C_{z_1}) + C_{z_1}\,\nabla\cdot\mathbf{U} - Q_C^{turb} + (h\,\mathcal{S})_{z_1}. \quad (9.42)$$

Hence, if the tracer concentration in water equals that in the surface cell, the fresh water contribution to the time tendency drops out. Note, however, than the influence of fresh water is still felt through its influence on the divergence $\nabla\cdot\mathbf{U}$.

9.5.4 Tracer concentration in fresh water

To compute the surface tracer budget, it is necessary to specify the tracer concentration C_w in the fresh water. We consider here two general classes of tracers.

9.5.4.1 Neutral tracers

Most climate models do not carry information regarding the tracer concentration in their rivers or atmospheric component models. Furthermore, the tracer concentration C_w is often quite close to the tracer concentration within the adjacent ocean cell. Hence, for many purposes, it is appropriate to assume

$$C_w \approx C_{z_0} \approx C_{z_1}. \quad (9.43)$$

Such tracers are called here *neutral tracers* since they do not add or subtract from the tracer concentration already present in the ocean cell, although they do alter the total tracer content. Notably, neutral tracers allow for the identification of individual terms in the surface flux equation (9.39)

$$(\mathbf{F}\cdot\widehat{\mathbf{N}})_{z_0} = Q_C^{turb}, \quad (9.44)$$

even when there is fresh water forcing.

9.5.4.2 Salt

Due to a large hydration energy for salt, the air-sea interface effectively acts as an impenetrable barrier to salt transport. Therefore, neglecting the formation of sea spray and the interchange of salt with sea ice, and in the absence of salt entering through the solid earth boundaries, total ocean salt remains constant. In turn, the salt mass per unit volume $\rho\,s$ within a grid cell, where s is salinity (a dimensionless concentration), changes only through advective and turbulent fluxes from the interior ocean, as well as time tendencies in the volume of the grid cell.

By assuming salt does not pass the air-sea interface, the total flux of salt passing across the ocean surface vanishes

$$Q_s = (\mathbf{F}_s \cdot \widehat{\mathbf{N}})_{z_0} = 0, \tag{9.45}$$

thus leading to the following budget for the mass per volume of salt in a surface grid cell

$$h_{z_1}\, \partial_t (\rho\, s)_{z_1} = -\rho_o\, \nabla \cdot (h\, \mathbf{u}\, s + h\, \mathbf{F}^h)_{z_1} + \rho_o\, (w\, s)_{z_1} + (\rho_o\, F^z)_{z_1} - \eta_{,t}\, (\rho\, s)_{z_1} + (h\, \rho\, \mathcal{S})_{z_1}. \tag{9.46}$$

As the salt flux is comprised of parameterized turbulent and advective terms, a zero surface salt flux implies a balance between the two contributions. On the ocean side of the interface, this balance leads to

$$(\mathbf{F}_s \cdot \widehat{\mathbf{N}})_{z_0} = -q_w\, s_{z_0} + (\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = 0. \tag{9.47}$$

That is, the turbulence flux of salt on the ocean side of the ocean surface is proportional to the flux of salt associated with fresh water. It is this turbulence salt flux which is to be used for computing buoyancy fluxes in ocean mixed layer schemes (Section 9.6). Further discussion of this balance for the Boussinesq fluid can be found in Huang (1993), where the turbulence flux was called an *anti-advective flux*.

On the atmosphere/river side, a zero total salt flux leads to the balance

$$Q_s = -q_w\, s_w + Q_s^{turb} = 0. \tag{9.48}$$

Assuming the salinity of fresh water vanishes then leads to

$$Q_s^{turb} = 0. \tag{9.49}$$

This result makes sense when recalling that a fluid with uniformly zero salinity, such as a fresh water river, trivially has no net transport of salinity, whether turbulent or otherwise.

These considerations for salt flux across the air-sea interface are modified when considering salt flux between the ocean and sea ice, since sea ice generally has a non-zero salinity $s_w \neq 0$. Many modern sea-ice models carry salinity content of the ice. For this case, ocean salt content changes when sea-ice melts or forms.

9.6 Fluxes for turbulence mixed layer schemes

Ocean mixed layer schemes such as that of Large et al. (1994) require the computation of the oceanic turbulence buoyancy flux at the ocean surface. Hence, we find it useful to summarize expressions for the turbulence flux of potential temperature and salinity $(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0}$. Section 9.5.2 noted that continuity of the tracer flux across the air-sea interface leads to (equation (9.39))

$$-q_w\, C_{z_0} + (\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = -q_w\, C_w + Q_C^{turb}, \tag{9.50}$$

which renders

$$(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = Q_C^{turb} + q_w\, (C_{z_0} - C_w). \tag{9.51}$$

Some special cases are now noted.

Recall from Section 9.5.4.1 that neutral tracers have $C_{z_0} = C_w$. Potential temperature is a reasonable approximation to a neutral tracer. In this case,

$$(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = Q_\theta^{turb} \qquad \text{for neutral tracers with } C_{z_0} = C_w, \tag{9.52}$$

as already noted in Section 9.5.4.1. This result is familiar from rigid lid models.

For ocean regions free of sea-ice, recall from Section 9.5.4.2 that the total salt flux vanishes at the ocean surface. Hence,

$$(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = q_w \, s_{z_0} \qquad \text{for tracers with zero air-sea flux such as salt.} \tag{9.53}$$

In the absence of fresh water flux, the turbulence flux of salt vanishes

$$(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = 0 \qquad \text{for salt if } q_w = 0. \tag{9.54}$$

With non-zero fresh water fluxes, we may wish to set salinity at the ocean surface using a centered approximation $s_{z_0} \approx (s_{z_1} + s_w)/2 = s_{z_1}/2$. However, it is arguable that a better estimate is $s_{z_0} \approx s_{z_1}$ for regions where fresh water fluxes are associated just with evaporation and precipitation. In this case, the turbulence salt flux is given by

$$(\mathbf{F} \cdot \widehat{\mathbf{N}})_{z_0} = q_w \, s_{z_1} \qquad \text{for salt if } s_{z_0} \approx s_{z_1}. \tag{9.55}$$

9.7 Flux plus restore boundary conditions

It is common to run ocean models with a flux provided from data or another component model, plus restoring to ocean data. The form for such a flux is given by

$$Q_C = (\gamma_C \, h_{z_1}) \, (C_{z_1} - C^{(data)}) + Q_C^{turb,0} - q_w \, C_w, \tag{9.56}$$

where $Q_C^{turb,0}$ is the turbulence flux from data or another model. The coefficient γ_C is an inverse restoring time so that the product $\gamma_C \, h_{z_1}$ is a velocity scale. This velocity scale is often called the *piston velocity*

$$V_{\text{piston}} = \gamma_C \, h_{z_1} \tag{9.57}$$

with a larger V_{piston} resulting in a stronger restoring. Note that with a free surface, the piston velocity changes according to the thickness of the time dependent top model grid cell h_{z_1} (see Figure 10.1). For less ambiguity in interpretation, the modeler may prefer to use the time independent top cell thickness (defined for $\eta = 0$) instead, so to have a time independent piston velocity.

The flux $V_{\text{piston}} \, (C_{z_1} - C^{(data)})$ is positive when $C_{z_1} > C^{(data)}$, indicating a transfer of tracer out of the model through the top, thus acting to damp the ocean surface tracer back towards $C^{(data)}$. This restoring flux is considered part of the turbulent flux, as opposed to part of the advective fresh water flux. That is, the total turbulent flux is given by

$$Q_C^{turb} = V_{\text{piston}} \, (C_{z_1} - C^{(data)}) + Q_C^{turb,0}. \tag{9.58}$$

The case of salinity restoring is somewhat distinct. There are two approaches. First, the restoring flux $\rho_{z_1} V_{\text{piston}} (s_{z_1} - s^{(data)})$ can be considered an actual flux of salt. Yet as mentioned in Section 9.5.4.2, salt does not generally cross the ocean surface. Instead, the ocean salt content is reasonably close to constant, at least for purposes of climate simulations. Hence, to maintain a constant salt content and to provide a local restoring for the salinity field, we can eliminate the salt flux in favor of an implied local fresh water flux via the equality

$$V_{\text{piston}} (s_{z_1} - s^{(data)}) = q_w \, s_{z_1}, \tag{9.59}$$

which leads to the fresh water flux

$$q_w = V_{\text{piston}} (1 - s^{(data)}/s_{z_1}). \tag{9.60}$$

As a check, we see that if the ocean surface salinity s_{z_1} is larger than the salinity from the data, there is a positive input of fresh water to the ocean, thus damping the surface salinity back towards that of the data. Note that there is no *a priori* reason that this fresh water flux will integrate to zero over the globe. Hence, its use will generally lead to a drift in ocean volume.

CHAPTER
TEN

Momentum equation solution methods

Contents

10.1 Summary of the continuous model equations 218
10.2 Momentum budget for a grid cell 219
10.2.1 Material time derivative 219
10.2.2 Horizontal pressure gradient 221
10.2.3 Friction due to vertical shears 221
10.2.4 Special considerations for boundary cells 222
10.2.4.1 Vertical flux of horizontal momentum through the surface 222
10.2.4.2 Vertical flux of horizontal momentum through the bottom 223
10.2.4.3 Hydrostatic pressure in the surface cell 224
10.3 Strategy for time stepping 225
10.3.1 Acoustic modes and the quasi-non-Boussinesq approximation 225
10.3.2 Computing *in situ* density 226
10.3.3 Barotropic or external gravity waves 226
10.3.4 General form for the momentum equation 227
10.4 Basics of the time stepping algorithm 228
10.4.1 Goals 228
10.4.2 Baroclinic/barotropic split of the velocity 229
10.5 Time stepping the barotropic system 230
10.6 Updating the baroclinic velocity 232
10.7 Summarizing a model time step 233
10.7.1 Advection velocities 233
10.7.2 Advective and diffusive fluxes 233
10.7.3 Tracers 234
10.7.4 Momentum 234
10.8 Rigid lid streamfunction method 235
10.8.1 Volume transport streamfunction 235

10.8.2 Time stepping the streamfunction 236
10.8.3 Closing comments 236

The purpose of this chapter is to illustrate a solution method for the momentum equations discretized in a free surface z-model. The splitting between slow three dimensional dynamics and fast two dimensional dynamics is motivated and detailed. The algorithm that we detail is known as a *split-explicit* free surface method and we illustrate this method using a leap-frog time stepping scheme. Our discussion shares much in common Killworth et al. (1991) and Griffies et al. (2001). An alternative approach solves the fast dynamical system implicitly in time, with the paper by Campin et al. (2003) documenting a novel scheme possessing desirable conservation properties only approximately maintained by the scheme described here. Care is taken in the following to distinguish between Boussinesq and non-Boussinesq equations. Additionally, the rigid lid method for Boussinesq fluids is presented and critiqued.

It is noted that we present here details of a commonly used time stepping scheme, but do not provide a numerical analysis of its stability. For that purpose, the reader is referred to the numerical analysis book of Durran (1999) which presents a thorough analysis of various time stepping schemes of use for geophysical fluid modeling.

10.1 Summary of the continuous model equations

The equations considered in this chapter represent discrete realizations of the continuum model momentum and kinematic equations presented in Section 6.8. For completeness, we repeat them here assuming hydrostatic balance. The non-Boussinesq hydrostatic equations are

$$\rho_{,t} + \rho_o \nabla \cdot \mathbf{v} = 0 \tag{10.1}$$

$$(D\,\overline{\rho}^z)_{,t} = -\rho_o \nabla \cdot \mathbf{U} + \rho_o\, q_w \tag{10.2}$$

$$\mathbf{u}_{,t} + \nabla \cdot (\mathbf{v}\,\mathbf{u}^\rho) + \mathcal{M}\,\hat{\mathbf{z}} \wedge \mathbf{v}^\rho = -f\,\hat{\mathbf{z}} \wedge \mathbf{v} - \nabla(p/\rho_o) + \mathbf{F}^{(\mathbf{u})} \tag{10.3}$$

$$p_{,z} = -\rho\, g \tag{10.4}$$

$$(\rho\, C)_{,t} + \rho_o \nabla \cdot (\mathbf{v}\, C) = -\rho_o \nabla \cdot \mathbf{F} + \rho\, \mathcal{S}, \tag{10.5}$$

with surface and bottom kinematic boundary conditions

$$\rho\, \eta_{,t} + \rho_o\, \mathbf{u} \cdot \nabla \eta = \rho_o\, w + \rho_o\, q_w \qquad \text{at } z = \eta \tag{10.6}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H. \tag{10.7}$$

In these expressions,

$$\rho\, \mathbf{v}^\rho = \rho_o\, \mathbf{v} \tag{10.8}$$

is the linear momentum density of a fluid parcel, and $\rho_o\, q_w$ is the linear momentum density of fresh water entering the ocean surface. The Boussinesq hydrostatic

equations are recovered by setting $\rho \to \rho_o$, except when multiplying gravity

$$\nabla \cdot \mathbf{v} = 0 \tag{10.9}$$

$$\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w \tag{10.10}$$

$$\mathbf{u}_{,t} + \nabla \cdot (\mathbf{v}\,\mathbf{u}) + \mathcal{M}\,\hat{\mathbf{z}} \wedge \mathbf{v} = -f\,\hat{\mathbf{z}} \wedge \mathbf{v} - \nabla(p/\rho_o) + \mathbf{F}^{(\mathbf{u})} \tag{10.11}$$

$$p_{,z} = -\rho\, g \tag{10.12}$$

$$C_{,t} + \nabla \cdot (\mathbf{v}\, C) = -\nabla \cdot \mathbf{F} + \mathcal{S}, \tag{10.13}$$

with surface and bottom kinematic boundary conditions

$$\eta_{,t} + \mathbf{u} \cdot \nabla \eta = w + q_w \qquad \text{at } z = \eta \tag{10.14}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H. \tag{10.15}$$

Our interpretation of the model fields is summarized in Table 6.1.

10.2 Momentum budget for a grid cell

Before developing methods for time stepping the ocean model equations, we provide a semi-discrete form of the momentum equation. To do so, the continuum model equations are integrated over the volume of a grid cell. In particular, we focus on vertical integration since there are important points to highlight regarding the surface and bottom boundary conditions. Horizontal integration follows similarly, with boundary conditions on the sides simpler than the surface and bottom. Where a choice is needed about horizontal placement of discrete fields, the B-grid is used. The result of these manipulations is a consistent discretization of the continuous equations. Some details are left unspecified, such as the precise form of advective and diffusive fluxes.

10.2.1 Material time derivative

To start, we integrate the material time derivative

$$\rho\,\frac{D\mathbf{u}^\rho}{Dt} = (\rho\,\mathbf{u}^\rho)_{,t} + \nabla \cdot (\mathbf{v}^\rho\,\rho\,\mathbf{u}^\rho) + \rho\,\mathcal{M}\,\hat{\mathbf{z}} \wedge \mathbf{u}^\rho = \rho_o\,(\mathbf{u}_{,t} + \nabla \cdot (\mathbf{v}\,\mathbf{u}^\rho) + \mathcal{M}\,\hat{\mathbf{z}} \wedge \mathbf{u}^\rho) \tag{10.16}$$

over the depth of a grid cell. For the metric term, we prescribe the correspondence

$$\int_{z_k}^{z_{k-1}} dz\,\mathcal{M}\,\hat{\mathbf{z}} \wedge \mathbf{u}^\rho \to h_k\,(\mathcal{M}\,\hat{\mathbf{z}} \wedge \mathbf{u}^\rho)_k \tag{10.17}$$

where $h_k = z_{k-1} - z_k$ is the thickness of the velocity cell, and k labels the vertical cell number. The correspondence (10.17) is afforded by assuming that correlation terms, which arise from integrating the nonlinear product over the grid cell, are absorbed into the SGS friction operator $\mathbf{F}^{\mathbf{u}}$.

Vertical integration of the time and space derivative operators over the surface model grid cell (Figure 10.1) leads to

$$\begin{aligned}\int_{z_1}^{\eta} dz\,[\mathbf{u}_{,t} + \nabla\cdot(\mathbf{v}\,\mathbf{u}^{\rho})] &= \partial_t\left(\int_{z_1}^{\eta} dz\,\mathbf{u}\right) + \nabla\cdot\left(\int_{z_1}^{\eta} dz\,\mathbf{u}\,\mathbf{u}^{\rho}\right)\\ &\quad - (w\,\mathbf{u}^{\rho})_{z_1} + [\mathbf{u}^{\rho}\,(w - \mathbf{u}\cdot\nabla\eta) - \mathbf{u}\,\eta_{,t}]_{z=\eta}\\ &= \partial_t\left(\int_{z_1}^{\eta} dz\,\mathbf{u}\right) + \nabla\cdot\left(\int_{z_1}^{\eta} dz\,\mathbf{u}\,\mathbf{u}^{\rho}\right) - (w\,\mathbf{u}^{\rho})_{z_1} - (\mathbf{u}^{\rho})_{z=\eta}\,q_w, \qquad (10.18)\end{aligned}$$

where $z_0 = \eta$, and we made use of the surface kinematic boundary condition $\rho\,\eta_{,t} + \rho_o\,\mathbf{u}\cdot\nabla\eta = \rho_o\,w + \rho_o\,q_w$. For Boussinesq models, $\mathbf{u}^{\rho} = \mathbf{u}$. In the discrete model, we make the correspondence

$$\partial_t\left(\int_{z_1}^{\eta} dz\,\mathbf{u}\right) + \nabla\cdot\left(\int_{z_1}^{\eta} dz\,\mathbf{u}\,\mathbf{u}^{\rho}\right) \rightarrow \partial_t(h\,\mathbf{u})_{z_1} + \nabla\cdot(h\,\mathbf{u}\,\mathbf{u}^{\rho})_{z_1}, \qquad (10.19)$$

thus leading to

$$\begin{aligned}&\int_{z_1}^{\eta} dz\,[\mathbf{u}_{,t} + \nabla\cdot(\mathbf{v}\,\mathbf{u}^{\rho})]\\ &\rightarrow \partial_t(h\,\mathbf{u})_{z_1} + \nabla\cdot(h\,\mathbf{u}\,\mathbf{u}^{\rho})_{z_1} - (w\,\mathbf{u}^{\rho})_{z_1} - (\mathbf{u}^{\rho})_{z_0}\,q_w,\end{aligned} \qquad (10.20)$$

where

$$h_1 = -z_1 + \eta > 0 \qquad (10.21)$$

is the time dependent thickness of the surface velocity cell. The $k > 1$ interior grid cells, which have time independent thicknesses, have the correspondence

$$\int_{z_k}^{z_{k-1}} dz\,[\mathbf{u}_{,t} + \nabla\cdot(\mathbf{v}\,\mathbf{u}^{\rho})] \rightarrow \partial_t(h\,\mathbf{u})_{z_k} + \nabla\cdot(h\,\mathbf{u}\,\mathbf{u}^{\rho})_{z_k} - (w\,\mathbf{u}^{\rho})_{z_k} + (w\,\mathbf{u}^{\rho})_{z_{k-1}}. \qquad (10.22)$$

The thickness weighted horizontal fluxes can be computed as in Adcroft et al. (1997) or Pacanowski and Gnanadesikan (1998) to account for the generally different adjacent cell thicknesses.

Equations (10.20) and (10.22) represent semi-discrete forms of the continuous expression. A complete spatial discretization requires the computation of *advective velocities* $\mathbf{v}_{adv}$ on the respective sides of the velocity cell. Advective fluxes $\mathbf{v}\,\mathbf{u}^{\rho}$ are then constructed through multiplying the advective velocities by an approximate value of $\mathbf{u}^{\rho}$ at the cell faces. Model forms for the momentum advective fluxes are given in Griffies et al. (2003), where discrete energetic balances are presented. These forms are constructed using second order centered expressions, as in Bryan (1969). For our purposes, schematic semi-discrete expressions, such as equations (10.20) and (10.22), are sufficient.

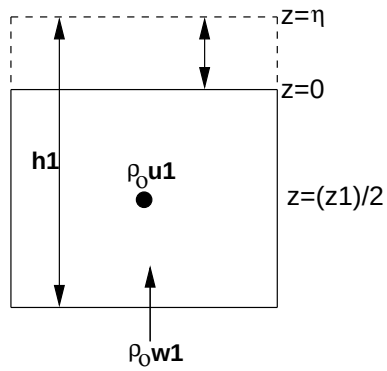

Figure 10.1: A grid cell used for computing the surface cell momentum budget. The cell's upper and lower boundaries are set by $z_0 = \eta$ and $z_1 < 0$, and the position of the grid point is $z_1/2$. See also Figure 9.1 for more details.

10.2.2 Horizontal pressure gradient

Vertical integration of the horizontal pressure gradient over the depth of a grid cell leads to the correspondence

$$\int_{z_k}^{z_{k-1}} dz\, \nabla p \rightarrow h_k\, \nabla p_k, \tag{10.23}$$

where p_k is the hydrostatic pressure at the depth of the velocity point. Note that to maintain energetic consistency, pressure is computed in a discrete z-model as the hydrostatic pressure *at the depth* of the velocity point, rather than as a *depth average* over the velocity cell (see Griffies et al. (2003)).

10.2.3 Friction due to vertical shears

The form of friction appropriate for an ocean model on the sphere using generalized horizontal coordinates is derived in Chapter 15. It is comprised of two terms,

$$\mathbf{F}^{\mathbf{u}} = \mathbf{F}^{\mathbf{u}}_{horz} + \mathbf{F}^{\mathbf{u}}_{vert}, \tag{10.24}$$

where $\mathbf{F}^{\mathbf{u}}_{horz}$ arises from horizontal shears in the fluid, and $\mathbf{F}^{\mathbf{u}}_{vert}$ arises from vertical shears. To develop the full discrete equations, we find it useful to detail here the contribution from vertical shears when integrated over a model cell. In general, $\mathbf{F}^{\mathbf{u}}_{vert}$ takes the form

$$\mathbf{F}^{\mathbf{u}}_{vert} = (\kappa\, \mathbf{u}^{\rho}_{,z})_{,z} \tag{10.25}$$

where κ is a non-negative vertical viscosity arising from unresolved small-scale turbulence. As indicated here, it is important to compute friction using the velocity $\mathbf{u}^{\rho}$ instead of $\mathbf{u}$, where again $\rho\, \mathbf{u}^{\rho} = \rho_o\, \mathbf{u}$. The reason is that it is $\mathbf{u}^{\rho}$ which sets the kinetic energy, and friction is defined so that it directly affects kinetic energy. Likewise, it is $\mathbf{u}^{\rho}$ which should be used to determine the Richardson number for use in turbulence closure schemes. The distinction between $\mathbf{u}^{\rho}$ and $\mathbf{u}$ is not made

for the Boussinesq fluids, where $\mathbf{u}^\rho = \mathbf{u}$, and the factors of density are replaced by ρ_o, which then cancel out from the friction vector.

Vertical integration of $\mathbf{F}^{\mathbf{u}}_{vert}$ over the depth of a model grid cell leads to

$$\int_{z_k}^{z_{k-1}} dz\, \mathbf{F}^{\mathbf{u}}_{vert} = (\kappa\, \mathbf{u}^\rho_{,z})_{z_{k-1}} - (\kappa\, \mathbf{u}^\rho_{,z})_{z_k}. \tag{10.26}$$

Differences in the diffusive fluxes across a cell lead to a net vertical transfer of horizontal turbulent momentum into or out of the cell.

10.2.4 Special considerations for boundary cells

Written as a prognostic equation for thickness weighted horizontal velocity, the momentum budget vertically integrated over a model grid cell becomes

$$\begin{aligned}(\partial_t + f\,\hat{\mathbf{z}}\wedge)\,(h\,\mathbf{u})_{z_k} &= -(h\,\mathcal{M}\,\hat{\mathbf{z}}\,\wedge\,\mathbf{u}^\rho)_{z_k} - \nabla\cdot(h\,\mathbf{u}\,\mathbf{u}^\rho)_{z_k} - h_k\,\nabla p_k/\rho_o + (h\,\mathbf{F}^{\mathbf{u}}_{horz})_{z_k} \\ &\quad - [(w\,\mathbf{u}^\rho)_{z_{k-1}} - (w\,\mathbf{u}^\rho)_{z_k}] + [(\kappa\,\mathbf{u}^\rho_{,z})_{z_{k-1}} - (\kappa\,\mathbf{u}^\rho_{,z})_{z_k}].\end{aligned} \tag{10.27}$$

The surface and bottom grid cells require some added consideration, and that is the focus of this subsection.

10.2.4.1 Vertical flux of horizontal momentum through the surface

Based on integrating the model equations over a grid cell, the vertical flux of horizontal momentum passing through the bottom of grid cell k is given by

$$\mathbf{F}^z_{z_k} = (\mathbf{u}^\rho\, w)_{z_k} - (\kappa\, \mathbf{u}_{,z})_{z_k}. \tag{10.28}$$

The first term represents the vertical advective flux of horizontal momentum, and the second term is the vertical diffusive flux of horizontal momentum that parameterizes SGS processes. The signs are such that horizontal momentum of the k grid cell is gained if $w > 0$ or if velocity decreases upwards. The vertical flux of horizontal momentum at the surface $z_{k=0} = \eta$ takes the form

$$\mathbf{F}^z_{z_0} = -(\mathbf{u}^\rho)_{z_0}\, q_w - (\kappa\, \mathbf{u}^\rho_{,z})_{z_0}. \tag{10.29}$$

As with the ocean interior, the surface flux is comprised of two contributions: an advective flux, in this case due to fresh water crossing the ocean surface, and a turbulent flux parameterizing unresolved momentum mixing processes, such as those associated with atmospheric winds. The opposite sign on the fresh water term accounts for the convention that $q_w > 0$ for fresh water entering the ocean (moving downward), whereas $w > 0$ for parcels moving upward.

At the ocean surface, the vertical flux of horizontal momentum is given in terms of the generally unknown surface values for the horizontal velocity $(\mathbf{u}^\rho)_{z_0}$ and its vertical shear $(\mathbf{u}^\rho_{,z})_{z_0}$. Unlike interfaces in the ocean interior, it is unclear how to approximate these terms since there is no ocean above the surface at $z_0 = \eta$. An alternative approach seeks not to approximate the terms individually. Rather, it assumes a continuous total momentum flux across ocean surface. This approach now places the responsibility of garnering the momentum flux onto a boundary

layer model. Such models depend on the multifaceted interactions between wind, waves, sea ice, rivers, etc.

To procede, we assume this boundary layer model has produced a value for the total vertical flux of horizontal momentum, and this flux is written $-(\mathbf{u}^\rho_w \, \rho_o \, q_w + \boldsymbol{\tau}^{winds})$, with $\boldsymbol{\tau}^{wind}$ a stress imparted by wind (and sea ice) on the ocean surface. Continuity across the ocean surface renders

$$-\mathbf{F}^z_{z_0} = (\mathbf{u}^\rho)_{z_0} \, q_w + (\kappa \, \mathbf{u}^\rho_{,z})_{z_0} \equiv \mathbf{u}^\rho_w \, q_w + \rho_o^{-1} \, \boldsymbol{\tau}^{winds}. \tag{10.30}$$

When there is no surface fresh water flux, there is a one-to-one relation between the stress $\boldsymbol{\tau}^{winds}$ and the diffusive momentum flux at the surface

$$\rho_o \, (\kappa \, \mathbf{u}^\rho_{,z})_{z_0} = \boldsymbol{\tau}^{winds} \quad \text{if } q_w = 0. \tag{10.31}$$

In general, although the total momentum flux is continuous at the ocean surface, the two components of the flux need not be. For example, the fresh water velocity $\mathbf{u}^\rho_w$ is generally different from the ocean surface velocity $\mathbf{u}^\rho_{z_0}$. Such may be important when coupling an ocean model to a detailed river model. Nonetheless, most present-day climate models (i.e., models circa 2003) assume the fresh water velocity is equal to the horizontal velocity $\mathbf{u}_{k=1}$ in the top model grid cell.

We are interested in developing a scheme for time stepping velocity $\mathbf{u}$, and not the thickness weighted velocity $h \, \mathbf{u}$. Hence, the surface cell $k = 1$ has a velocity equation given by

$$\begin{aligned} h_{z_1} \, (\partial_t + f \, \hat{\mathbf{z}} \wedge) \, \mathbf{u}_{z_1} &= -\mathbf{u}_{z_1} \, \eta_{,t} - (h \, \mathcal{M} \, \hat{\mathbf{z}} \wedge \mathbf{u}^\rho)_{z_1} - \nabla \cdot (h \, \mathbf{u} \, \mathbf{u}^\rho)_{z_1} - (h \, \nabla p)_{z_1} / \rho_o \\ &\quad + (h \, \mathbf{F}^{\mathbf{u}}_{horz})_{z_1} - [(w \, \mathbf{u}^\rho)_{z_0} - (w \, \mathbf{u}^\rho)_{z_1}] + [(\kappa \, \mathbf{u}^\rho_{,z})_{z_0} - (\kappa \, \mathbf{u}^\rho_{,z})_{z_1}]. \end{aligned} \tag{10.32}$$

In particular, combining the $\mathbf{u}_{z_1} \, \eta_{,t}$ term with the vertical flux of horizontal momentum at the ocean surface leads to

$$\begin{aligned} -\mathbf{u}_{z_1} \, \eta_{,t} - \mathbf{F}^z_{z_0} &= -\mathbf{u}_{z_1} \, \eta_{,t} + \rho_o^{-1} \, (\rho_o \, q_w \, \mathbf{u}^\rho_w + \boldsymbol{\tau}^{winds}) \\ &= \rho_o^{-1} \left(-\mathbf{u}^\rho_{z_1} \, \rho_{z_1} \, \eta_{,t} + \rho_o \, q_w \, \mathbf{u}^\rho_w + \boldsymbol{\tau}^{winds} \right), \end{aligned} \tag{10.33}$$

where we set $\rho_o \, \mathbf{u}_{z_1} = \mathbf{u}^\rho_{z_1} \, \rho_{z_1}$. For a Boussinesq fluid, we set density to ρ_o and use the column integrated volume conservation $\eta_{,t} = -\nabla \cdot \mathbf{U} + q_w$ to find

$$-\mathbf{u}_{z_1} \, \eta_{,t} + \rho_o^{-1} \, (\rho_o \, q_w \, \mathbf{u}^\rho_w + \boldsymbol{\tau}^{winds}) \rightarrow \mathbf{u}_{z_1} \, \nabla \cdot \mathbf{U} + q_w \, (\mathbf{u}^\rho_w - \mathbf{u}_{z_1}) + \boldsymbol{\tau}^{winds} / \rho_o. \tag{10.34}$$

Hence, for the special case of equal fresh water and surface ocean velocities, $\mathbf{u}^\rho_w = \mathbf{u}^\rho_{z_1}$, the fresh water flux drops out as an explicit contribution to the Boussinesq momentum budget. It nonetheless remains involved in the budget through its effects on the convergence $-\nabla \cdot \mathbf{U}$ within the surface height equation.

10.2.4.2 Vertical flux of horizontal momentum through the bottom

Considerations analogous to those relevant at the ocean surface also apply at the ocean bottom. Yet at the bottom there is no advective momentum flux from the solid earth to the ocean. Furthermore, the bottom grid cell is typically assumed to have time independent thickness (though this may not be the case with a bottom

boundary layer model as described by Killworth and Edwards (1999)). So the only contribution to the vertical flux of horizontal momentum arises from the SGS friction. For $k = Nk$ (the bottom-most grid cell in a column), this friction is commonly written

$$-\rho_o \, \mathbf{F}^z_{z_{Nk}} = \rho_o \left(\kappa \, \mathbf{u}^\rho_{,z} \right)_{z_k} = \boldsymbol{\tau}^{bottom}, \tag{10.35}$$

where $\boldsymbol{\tau}^{bottom}$ is the bottom stress vector (with units of pressure). The bottom stress is often associated with currents or tides causing fluid to move over rough small scale topography. However, some SGS interactions between topography and mesoscale eddies act to increase the kinetic energy, as described by Holloway (1992).

10.2.4.3 Hydrostatic pressure in the surface cell

In Section 3.8.1 we considered an approximate expression for the continuum hydrostatic pressure at an arbitrary depth

$$p = p_a + p_s + p_b, \tag{10.36}$$

where

$$p_b = g \int_z^0 \rho \, dz \tag{10.37}$$

is known as the *baroclinic* or *internal* pressure field,

$$p_s = g \, \eta \, \rho(z = 0) \tag{10.38}$$

is known as the *surface pressure* field, and p_a is the pressure at $z = \eta$ from the overlying atmosphere or sea ice. The baroclinic pressure is that pressure at depth $z < 0$ arising from ocean fluid contained between $z = 0$ and $z < 0$. The surface pressure is that pressure at $z = 0$ arising from the ocean fluid between $z = 0$ and $z = \eta$, assuming density in this layer is vertically uniform with value $\rho(z = 0)$. The surface pressure is negative if $\eta < 0$.

Given the linear relation between pressure and density, it is natural to define the discrete pressure field on the density/tracer points. For a B-grid arrangement of variables, the offset in the horizontal between velocity and tracer/density points means that the hydrostatic pressure at the face of a velocity cell must be computed via a spatial average. The gradient of this pressure across the velocity cell determines the horizontal pressure gradient force acting on the fluid in the cell.

The particular case of a surface cell is worth highlighting. Here, pressure at the western face of a surface velocity cell and at the depth of the velocity point (see Figure 9.1) is given by

$$p = g \, \overline{\rho \left(|z_1|/2 + \eta \right)}^y + p_a, \tag{10.39}$$

where $\overline{(\ )}^y$ is a meridional grid average and η is the surface height on the tracer point. A similar expression is used to estimate pressure on the meridional face of the velocity cell. The first piece of this pressure,

$$p_b \equiv \left(g \, |z_1|/2 \right) \overline{\rho}^y, \tag{10.40}$$

is a discretization of the baroclinic pressure in the surface cell. This is the pressure due to fluid between the resting surface $z = 0$ and the velocity point at $z = |z_1|/2$.

Note that z_1 has been assumed independent of horizontal position, hence its removal from the meridional average operator. Such is the case for surface grid cells in z-coordinate models. The second contribution

$$p_s \equiv g\,\overline{\rho}\,\overline{\eta}^y \tag{10.41}$$

is a discretization of the surface pressure onto the western face of the velocity cell. Again, this is the hydrostatic pressure at $z = 0$ associated with fluid between $z = 0$ and $z = \eta$. Gradients in the surface pressure arise from those in the density weighted free surface height.

For Boussinesq models, it is common to approximate the surface pressure on the western face of the velocity cell as $p_s \approx g\,\rho_o\,\overline{\eta}^y$, rather than use the hydrostatic form $g\,\overline{\rho}\,\overline{\eta}^y$. With this approximation, surface pressure gradients arise *solely* from gradients in the free surface height. To maintain self-consistency with the hydrostatic baroclinic pressure field, while incurring only trivial computational expense, we recommend the hydrostatic form $p_s = g\,\overline{\rho}\,\overline{\eta}^y$ even when using the Boussinesq option.

10.3 Strategy for time stepping

Our goal is to develop a strategy for time stepping the momentum budget over the course of a discrete time step of length $\Delta\tau$. This goal is complicated by the presence of acoustic waves in the non-Boussinesq fluid, and external or barotropic gravity waves in either the non-Boussinesq or Boussinesq fluid. Resolving these fast modes requires $\Delta\tau$ to be too small for use in climate studies, where simulations are desired on the order of centuries to millenia. Hence, approximate approaches are needed.

10.3.1 Acoustic modes and the quasi-non-Boussinesq approximation

Acoustic waves are three-dimensional fluctuations in the pressure field (e.g., Gill (1982), Apel (1987)). They travel at roughly $1500\,\mathrm{m\,s^{-1}}$. However, there is no evidence that resolving acoustic waves is essential for the physical integrity of ocean climate models. Hence, we are not motivated to resolve these waves for our purposes.

As mentioned in Section 2.4.1, use of the hydrostatic approximation in primitive equation ocean climate models acts to filter all acoustic modes except the Lamb wave. We can be assured of filtering the Lamb wave by either taking the Boussinesq approximation or by assuming density appearing in mass continuity is a function of a time independent pressure (Section 4.8.4). The latter method is described in Durran (1999), where he calls the resulting fluid *pseudo-incompressible*, as well as Greatbatch et al. (2001). We prefer the term *quasi-non-Boussinesq*. This approximation, however, has *not* been found necessary in the non-Boussinesq form of the code described by Griffies et al. (2003) (MOM4). Instead, the Lamb mode in practice has been found to be weak and so the model is stable even when using the full density.

10.3.2 Computing *in situ* density

We are here more precise regarding how density is evaluated. Recall from Section 4.8 that density can be written as a function of potential temperature, salinity, and pressure

$$\rho = \rho(\theta, s, p). \tag{10.42}$$

Potential temperature and salinity evolve according to their respective prognostic tracer equations (see Section 10.1). Pressure, however, is determined diagnostically by vertically integrating the hydrostatic balance (e.g., Section 10.2.4.3), and so pressure is itself a function of density.

As discussed in Griffies et al. (2001), a simple way to break the pressure-density loop is to approximate density at a model time step τ according to

$$\rho(\tau) = \rho(\theta(\tau), s(\tau), p(\tau - \Delta\tau)). \tag{10.43}$$

A slightly more accurate approximation is to estimate the pressure $p(\tau)$ according to a linear extrapolation using $p(\tau - \Delta\tau)$ and $p(\tau - 2\Delta\tau)$

$$p(\tau) \approx (p(\tau - \Delta\tau) - p(\tau - 2\Delta\tau)) + p(\tau - \Delta\tau). \tag{10.44}$$

Alternatively, we may wish to consider an iterative method as in Dewar et al. (1998), but these approaches tend to be time consuming. Whatever method is used, the resulting density is a function of a time dependent pressure field, as it should be. This density is then used for computing pressure $p(\tau)$ whose gradient sets the horizontal pressure gradient.

For purposes of computing mass continuity, a quasi-non-Boussinesq fluid uses the alternative density field

$$\rho(\tau) = \rho(\theta(\tau), s(\tau), -\rho_o\, g\, z), \tag{10.45}$$

where $-\rho_o\, g\, z$ is the hydrostatic pressure associated with a fluid of constant density ρ_o. Note that this is the form for density suggested by Bryan and Cox (1972) and used by many z-coordinate models until the middle 1990's (e.g., Cox (1984), Pacanowski et al. (1991) (MOM1), and Pacanowski (1995) (MOM2)). As desired, time tendencies of ρ do not have contributions from $p_{,t}$, and so all acoustic modes are eliminated. As stated above, this approximation has not been found necessary in practice for the non-Boussinesq form of MOM4. Hence, the pressure dependence of density as computed in equation (10.43) is recommended.

10.3.3 Barotropic or external gravity waves

In both non-Boussinesq or Boussinesq fluids, external waves are roughly 100 times the speed of the next fastest internal wave or advective signal (see discussion in Section 3.8). Their propagation causes fluctuations in the ocean surface height as the waves propagate through the fluid columns. Fortunately, these external or barotropic gravity waves are nearly two-dimensional in structure, and so they are largely represented by the simpler dynamics of the vertically integrated fluid column.

Hence, if we can separate or split the fast vertically integrated dynamics and kinematics from the slow and more complicated vertically dependent dynamics

and kinematics, we stand the chance of substantially increasing the model's efficiency. Providing details for this split forms the bulk of the remainder of this chapter.

10.3.4 General form for the momentum equation

Recall the equation we wish to time step is the semi-discrete momentum equation (10.27), written here as

$$h_{z_k}\,(\partial_t + f\,\hat{\mathbf{z}}\wedge)\,\mathbf{u}_{z_k} = -\delta_{k1}\,\mathbf{u}_{z_k}\,\partial_t\,\eta - (h\,\mathcal{M}\,\hat{\mathbf{z}}\wedge\mathbf{u}^\rho)_{z_k} - \nabla\cdot(h\,\mathbf{u}\,\mathbf{u}^\rho)_{z_k} - (h\,\nabla p)_{z_k}/\rho_o \\ - [(w\,\mathbf{u}^\rho)_{z_{k-1}} - (\kappa\,\mathbf{u}^\rho_{,z})_{z_{k-1}}] + [(w\,\mathbf{u}^\rho)_{z_k} - (\kappa\,\mathbf{u}^\rho_{,z})_{z_k}] + (h\,\mathbf{F}^{\mathbf{u}}_{horz})_{z_k}, \quad (10.46)$$

where δ_{k1} is the Kronecher delta function which is unity for $k = 1$ and zero otherwise. To develop a strategy for time stepping this equation, we find it convenient to write it in the shorthand form

$$h_{z_k}\,(\partial_t + f\,\hat{\mathbf{z}}\wedge)\,\mathbf{u}_{z_k} = -h_{z_k}\,\nabla\,p_s/\rho_o + (h\,\tilde{\mathbf{G}})_{z_k}, \quad (10.47)$$

where

$$(h\,\tilde{\mathbf{G}})_{z_k} = (h\,\mathbf{G})_{z_k} - \delta_{k1}\,\mathbf{u}_{z_k}\,\partial_t\eta \quad (10.48)$$

and

$$(h\,\mathbf{G})_{z_k} = -(h\,\mathcal{M}\,\hat{\mathbf{z}}\wedge\mathbf{u}^\rho)_{z_k} - \nabla\cdot(h\,\mathbf{u}\,\mathbf{u}^\rho)_{z_k} - (h\,\nabla p_b)_{z_k}/\rho_o \\ - [(w\,\mathbf{u}^\rho)_{z_{k-1}} - (\kappa\,\mathbf{u}^\rho_{,z})_{z_{k-1}}] + [(w\,\mathbf{u}^\rho)_{z_k} - (\kappa\,\mathbf{u}^\rho_{,z})_{z_k}] + (h\,\mathbf{F}^{\mathbf{u}}_{horz})_{z_k}. \quad (10.49)$$

$\mathbf{G}_k$ incorporates the baroclinic pressure gradient, advection, and friction. Notably, the velocity cell thicknesses h used for computing the various terms in $\mathbf{G}_k$ are taken at baroclinic time τ, as are the other inviscid contributions such as pressure and advection as appropriate for a leap-frog scheme. In contrast, the velocity cell thicknesses are computed at time $\tau - \Delta\tau$ for the disspative terms, except for vertical dissipation which is generally computed implicitly in time to allow for rapidly changing vertical processes.

For the vertically integrated column dynamics, we time step the vertically integrated velocity

$$\mathbf{U} = \sum_k h_k\,\mathbf{u}_k, \quad (10.50)$$

instead of the depth averaged velocity

$$\overline{\mathbf{u}}^z = \mathbf{U}/D \quad (10.51)$$

where

$$D = \sum_k h_k = H + \eta \quad (10.52)$$

is the time dependent total depth of ocean fluid at a particular horizontal position. The evolution of $\mathbf{U}$ takes the form

$$\partial_t\mathbf{U} = \mathbf{u}_1\,\partial_t\eta + \sum_k h_k\,\partial_t\mathbf{u}_k \\ = -f\,\hat{\mathbf{z}}\wedge\mathbf{U} - D\,\nabla\,p_s/\rho_o + \mathbf{G}, \quad (10.53)$$

where equation (10.47) was used for $\partial_t \mathbf{u}_k$, and the vertically integrated forcing is given by

$$\mathbf{G} = \sum_k h_k \, \mathbf{G}_k \tag{10.54}$$

with $\mathbf{G}_k$ defined in equation (10.49). With $\mathbf{G} = 0$, the system reduces to the momentum equation with inertia-gravity wave solutions. The gravity waves have speeds of

$$c_g = \sqrt{g\,H}, \tag{10.55}$$

or about $100 - 200 \mathrm{m\,s^{-1}}$ in the deep ocean. These barotropic gravity waves determine the discrete time step to be used when temporally discretizing the depth integrated dynamics of fluid columns.

10.4 Basics of the time stepping algorithm

The focus in this subsection is on time and depth discretization, with Figure 10.2 summarizing the following split-explicit algorithm. Discrete baroclinic times and time steps are denoted by the Greek τ and $\Delta\tau$, respectively, whereas the much smaller barotropic analogs use the Latin t and Δt. As stated in the introduction, alternative schemes exist, such as the scheme recently proposed by Campin et al. (2003). Our discussion documents a commonly used approach whose general algorithm has been used by many modelers. It therefore serves our pedagogical purposes.

10.4.1 Goals

Allowing the grid cell thicknesses to evolve introduces a fundamentally new element to the traditional algorithms relevant for constant cells in z-coordinate models (e.g., Bryan (1969), Semtner (1974), Cox (1984), Killworth et al. (1991), and Dukowicz and Smith (1994)). However, since we impose positive cell thicknesses for all cells, including the surface, modifications to the constant cell approach should be relatively modest. That is a goal of the algorithm.

More precisely, the main goals guiding the choices made in the following momentum equation algorithm are the following:

- Upon setting the baroclinic and barotropic time steps equal, the algorithm reduces the split baroclinic/barotropic system to a non-split system. Besides allowing the model to be used for detailed studies of tidal motions, running an un-split algorithm is a powerful means of checking code integrity. Using a leap-frog for the baroclinic and tracer portions necessitates a leap-frog for the barotropic portion of the algorithm.

- The model should be stable upon taking tracer time steps longer than baroclinic time steps for purposes of spinning-up coarse model integrations.

- The surface height should be smooth to within the limitations inherent in the presence of the B-grid splitting mode (e.g., Killworth et al. (1991), Pacanowski and Griffies (1999), Griffies et al. (2003)).

- Upon setting the baroclinic and tracer time steps equal, there should be exact conservation of total tracer content.
- For all tracer time steps, there should be exact local tracer conservation. That is, the tracer and mass/volume budgets must be compatible over each model grid cell.
- The algorithm should involve local calculations in order to minimize the issues of costly communication across parallel computers.

As described in Griffies et al. (2001) and Griffies et al. (2003), use of time filtering in the three time step leap-frog scheme precludes the model from having an exact discrete global tracer conservation property. All other properties are satisfied with the scheme documented in the following. Campin et al. (2003) discuss conservation properties for free surface algorithms with differing time stepping approaches. They show that moving away from the leap-frog allows schemes to conserve tracer both locally and globally.

10.4.2 Baroclinic/barotropic split of the velocity

As with the approach of Bryan (1969), Semtner (1974), and Cox (1984), the velocity at an arbitrary depth level k and baroclinic time τ' is split into two components

$$\mathbf{u}_k(\tau') = B_{km}(\tau)\,\mathbf{u}_m(\tau') + (\delta_{km} - B_{km}(\tau))\,\mathbf{u}_m(\tau'). \tag{10.56}$$

This equation is an identity valid for any baroclinic times τ' and τ and any operator $B_{km}(\tau)$. In Griffies et al. (2001), $\tau' = \tau + \Delta\tau$ was chosen since the baroclinic system was solved prior to the barotropic. However, it is more convenient to solve the barotropic system first as in Griffies et al. (2003) (MOM4), and so we here set $\tau' = \tau$, thus leading to

$$\mathbf{u}_k(\tau) = B_{km}(\tau)\,\mathbf{u}_m(\tau) + (\delta_{km} - B_{km}(\tau))\,\mathbf{u}_m(\tau) \equiv \widehat{\mathbf{u}}_k(\tau) + \overline{\mathbf{u}}^z(\tau). \tag{10.57}$$

The utility of this velocity field split relies on the form of the *baroclinicity* operator

$$B_{km}(\tau) = \delta_{km} - D(\tau)^{-1}\,h^u_m(\tau), \tag{10.58}$$

where δ_{km} is the Kronecker delta, summation over the repeated vertical level index m is implied, and

$$D(\tau) = \sum_k h^u_k(\tau) = H + \eta^u(\tau) \tag{10.59}$$

is the ocean depth at baroclinic time τ over a column of velocity points, with H the time independent resting ocean depth. In contrast to the rigid lid case, where all cells have time independent thicknesses, the baroclinicity operator used here is based on the instantaneous distribution of cell thicknesses at time τ. However, just as in a rigid lid model, the baroclinicity operator eliminates the depth independent part of a field, and leaves an approximation to the baroclinic portion. Hence, $\widehat{\mathbf{u}}_k$ is an approximation to the baroclinic velocity field, and $\overline{\mathbf{u}}^z$ is an approximation to the barotropic.

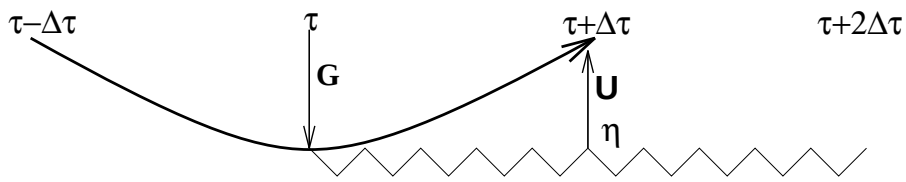

Figure 10.2: Schematic of the split-explicit time stepping scheme for the barotropic and baroclinic modes. Time increases to the right. Baroclinic time steps are denoted by $\tau - \Delta\tau, \tau, \tau + \Delta\tau$, and $\tau + 2\Delta\tau$. The curved line represents a baroclinic leap-frog time step, and the smaller barotropic time steps $N\,\Delta t = 2\,\Delta\tau$ are denoted by the zig-zag line. The vertically integrated forcing $\mathbf{G}(\tau)$ (equation (10.53)) computed at baroclinic time step τ represents the interaction between the barotropic and baroclinic motions. While keeping $\mathbf{G}(\tau)$ and the ocean depth $D(\tau)$ fixed, as well as the tracer, density, and fresh water forcing fields fixed, a leap-frog integration carries the surface height and vertically integrated velocity from τ to $\tau + 2\Delta\tau$ using N barotropic time steps of length Δt. Time averaging the barotropic fields over the $N + 1$ time steps (endpoints included) centers the vertically integrated velocity at the baroclinic time step $\tau + \Delta\tau$. A baroclinic leap-frog time step carries the surface height to $\tau + \Delta\tau$ using the convergence of the time averaged vertically integrated velocity taken from baroclinic time step τ.

If the baroclinic/barotropic split introduced in equation (10.57) is successful, the baroclinic velocity field $\widehat{\mathbf{u}}_k(\tau)$ evolves on a slow time scale $\Delta\tau$ and the barotropic velocity $\overline{\mathbf{u}}^z(\tau)$ evolves on the fast time scale

$$\Delta t = (2/N)\,\Delta\tau, \tag{10.60}$$

with N determined by the ratio of external to internal gravity wave speeds. The overall method therefore proceeds by separately updating $\widehat{\mathbf{u}}_k(\tau)$ and $\overline{\mathbf{u}}^z(\tau)$ by exploiting the time scale split. Upon doing so, the right hand side of the identity (10.57) is specified, hence allowing for an update of the full velocity field. The following sections provide details.

10.5 Time stepping the barotropic system

Let us assume knowledge of the full velocity and tracer fields at baroclinic time τ. Hence, we can update the surface height and vertically integrated velocity with a leap-frog scheme using the small barotropic time step Δt. For the non-Boussinesq model, mass conservation over a column of discrete fluid (see equation (9.26) in Section 9.4) leads to the time discretized equation

$$\begin{aligned}\rho_{z_1}(\tau)\,[\eta^{(b)}(\tau, t_{n+1}) - \eta^{(b)}(\tau, t_{n-1})] \\ = 2\,\Delta t\,[-\rho_o\,\nabla\cdot\mathbf{U}^{(b)}(\tau, t_n) + \rho_o\,q_w(\tau) - \sum_{k=1}^{N_k} h_k^t(\tau)\,\partial_\tau\rho_{z_k}(\tau)]\end{aligned} \tag{10.61}$$

$$\mathbf{U}^{(b)}(\tau, t_{n+1}) - \mathbf{U}^{(b)}(\tau, t_{n-1}) = 2\Delta t\,[-f\,\hat{\mathbf{z}} \wedge \mathbf{U}^{(b)}(\tau, t_n) - D(\tau)\,\nabla\,\tilde{p}_s^{(b)}(\tau, t_n) + \mathbf{G}(\tau)]. \quad (10.62)$$

For the Boussinesq fluid, the surface height evolution takes on the simpler form

$$\eta^{(b)}(\tau, t_{n+1}) - \eta^{(b)}(\tau, t_{n-1}) = 2\,\Delta t\,[-\nabla \cdot \mathbf{U}^{(b)}(\tau, t_n) + q_w(\tau)]. \quad (10.63)$$

In these equations, a raised (b) denotes values of surface height and vertically integrated velocity updated with the small barotropic time steps. The τ time label on $\eta^{(b)}$ and $\mathbf{U}^{(b)}$ denotes the baroclinic time at which certain variables are held for the duration of the barotropic time stepping over a single cycle. The variables held fixed are the vertically integrated forcing $\mathbf{G}(\tau)$, the tracer and density fields, the fresh water flux $q_w(\tau)$ and fresh water density $\rho_o(\tau)$, thickness of a grid cell $h_k^t(\tau)$, and total thickness of an ocean column $D(\tau)$. The time τ is also the time that sets the barotropic time steps via

$$t_n = \tau + n\,\Delta t \quad (10.64)$$

with n an integer. The density scaled surface pressure is evaluated via

$$\tilde{p}_s^{(b)}(\tau, t_n) = g\,\eta^{(b)}(\tau, t_n)\,\rho(\tau)_{k=1}/\rho_o. \quad (10.65)$$

To get started, we assume the following initial conditions

$$\eta^{(b)}(\tau, t_{n=-1}) = \eta^{(b)}(\tau, t_{n=0}) = \overline{\eta^{(b)}}(\tau), \quad (10.66)$$

with

$$\overline{\eta^{(b)}}(\tau) = \frac{1}{N+1}\sum_{n=0}^{N} \eta^{(b)}(\tau - \Delta\tau, t_n) \quad (10.67)$$

the time averaged surface height taken from the previous barotropic cycle. Likewise,

$$\mathbf{U}^{(b)}(\tau, t_{n=-1}) = \mathbf{U}^{(b)}(\tau, t_{n=0}) = \overline{\mathbf{U}^{(b)}}(\tau), \quad (10.68)$$

with

$$\overline{\mathbf{U}^{(b)}}(\tau) = \frac{1}{N+1}\sum_{n=0}^{N} \mathbf{U}^{(b)}(\tau - \Delta\tau, t_n) \quad (10.69)$$

the time averaged vertically integrated transport. Notably, we find no need to introduce a Robert-Asselin time filter in the barotropic portion of the integration. Additionally, because of the identification of the fields at $t_{n=-1}$ with those at $t_{n=0}$, the very first time step of a barotropic cycle is a forward Euler step instead of a leap-frog step.

Upon reaching $t_{n=N} = \tau + 2\Delta\tau$, the vertically integrated velocity is time averaged to produce the updated vertically integrated velocity at baroclinic time $\tau + \Delta\tau$

$$\mathbf{U}(\tau + \Delta\tau) \equiv \overline{\mathbf{U}^{(b)}}(\tau + \Delta\tau) = \frac{1}{N+1}\sum_{n=0}^{N} \mathbf{U}(\tau, t_n). \quad (10.70)$$

The surface height on the new baroclinic time step is then determined via a baroclinic leap-frog using the following form

$$\rho_{z_1}(\tau)\,(\eta(\tau + \Delta\tau) - \eta^F(\tau - \Delta\tau)) = 2\,\Delta\tau\,[-\rho_o\,\nabla \cdot \mathbf{U}(\tau) + \rho_o\,q_w(\tau) - \sum_{k=1}^{N_k} h_k^t(\tau)\,\partial_\tau \rho_{z_k}(\tau)] \quad (10.71)$$

for the non-Boussinesq fluid, and

$$\eta(\tau + \Delta\tau) - \eta^F(\tau - \Delta\tau) = 2\,\Delta\tau\,[-\nabla \cdot \mathbf{U}(\tau) + q_w(\tau)] \tag{10.72}$$

for the Boussinesq fluid. The use of this "big-leap-frog" scheme for the surface height ensures compatibility between the mass/volume budgets and the tracer budgets. More discussion of this point is provided in Chapter 9 (see in particular Section 9.2).

In general, some form of time filter is needed to maintain integrity of the surface height field due to the leap-frog splitting mode (e.g.,Haltiner and Williams (1980)) in equations (10.71) and (10.72). We have tried various forms of such filtering, with the following method discussed in Griffies et al. (2001) chosen due to its stability and reasonably good maintenance of tracer conservation properties (see Griffies et al. (2003))

$$\eta^F(\tau) = \overline{\eta^{(b)}}(\tau). \tag{10.73}$$

Another approach tried was

$$\eta^F(\tau) = \eta(\tau) + (\alpha/2)\,[\overline{\eta^{(b)}}(\tau + \Delta\tau) + \overline{\eta^{(b)}}(\tau - \Delta\tau) - 2\,\eta(\tau)], \tag{10.74}$$

which is useful since it isolates all the time filtering aspects into the term multiplied by α. This isolation allows for an easy check that global tracer conservation is exact when eliminating tracer and surface height time filtering. However, in the general case with a non-zero α, the filter (10.73) was found to be more conservative for certain tests, and so is recommended.

10.6 Updating the baroclinic velocity

By construction, evolution of the velocity field $\widehat{\mathbf{u}}_k(\tau)$ is unaffected by vertically independent forces, such as those from surface pressure gradients. Therefore, it is sufficient to update the "primed" velocity

$$\mathbf{u}'_k(\tau + \Delta\tau) = \mathbf{u}^R_k(\tau - \Delta\tau) + 2\Delta\tau\,[-f\,\hat{\mathbf{z}} \wedge \mathbf{u}_k(\tau) + \tilde{\mathbf{G}}_k(\tau)], \tag{10.75}$$

which represents a temporal discretization of the full momentum equation (10.47), yet without the surface pressure gradient, thus allowing for stable use of the longer baroclinic time step. The lagged velocity

$$\mathbf{u}^R_k(\tau - \Delta\tau) = \mathbf{u}_k(\tau - \Delta\tau) + (\alpha/2)\,(\mathbf{u}_k(\tau) - 2\,\mathbf{u}_k(\tau - \Delta\tau) + \mathbf{u}^R_k(\tau - 2\Delta\tau)) \tag{10.76}$$

is a Robert time filtered version of the full velocity field. A weak form of such filtering, with $\alpha = 0.01$, has been found sufficient in applications to suppress splitting between the two leap-frog branches in the baroclinic system. It is also useful to dampen fast dynamics that may partially leak through the baroclinicity operator due to the generally imperfect separation between the slow and fast dynamics.

The baroclinic piece of the primed velocity $\mathbf{u}'_k(\tau + \Delta\tau)$ yields

$$\widehat{\mathbf{u}}_k(\tau + \Delta\tau) = B_{km}(\tau + \Delta\tau)\,\mathbf{u}'_m(\tau + \Delta\tau), \tag{10.77}$$

as can be shown by adding the missing surface pressure gradient to equation (10.75), where it is annihilated by application of B_{km}.

To complete an update of the full velocity field, we add the updated baroclinic velocity to the updated barotropic velocity

$$\mathbf{u}_k(\tau + \Delta\tau) = \widehat{\mathbf{u}}_k(\tau + \Delta\tau) + \mathbf{U}(\tau + \Delta\tau)/D(\tau + \Delta\tau). \tag{10.78}$$

Knowledge of the updated velocity, along with updated tracer $T(\tau + \Delta\tau)$, allows for construction of the vertically integrated forcing $\mathbf{G}(\tau + \Delta\tau)$, and hence movement onto the next time step. This completes a full split-explicit time step cycle.

10.7 Summarizing a model time step

We have now detailed a method whereby the momentum equations can be efficiently stepped forward in time. The method, known as a *split-explicit* time stepping approach, where "split" refers to the separation between the slow and fast dynamics, and "explicit" refers to the use of an explicit in time scheme. This algorithm is also known as an *explicit free surface* method. It has been found to be very useful for ocean climate simulations. Because the previous discussion required many formulae and may be considered somewhat opaque on first encounter, we summarize here, without equations, a model time step using this algorithm as realized in the MOM4 code.

10.7.1 Advection velocities

After initializing the model fields and boundary forcing, the advection velocities are computed in order to construct advective fluxes of tracer and momentum. For stepping forward the model fields to time $\tau + \Delta\tau$, the advection velocities are constructed using the prognostic velocity field $\mathbf{v}$ at baroclinic time step τ. The horizontal advective velocities for tracers are specified according to the needs of energetic consistency (see Griffies et al. (2003) for a discussion). Horizontal advective velocities for momentum are then diagnosed from the tracer velocities according to a conservative remapping scheme. Both of these points are detailed in Griffies et al. (2003).

Vertical advective velocities for both the tracer and horizontal momentum fields are diagnosed based on the requirements of mass or volume conservation over a grid cell. For the Boussinesq fluid, this diagnosis requires only a calculation of the convergence of the horizontal advective velocities onto the grid cell. For the non-Boussinesq fluid, it is also necessary to incorporate the time tendency of the density field within the cell. This time tendency is approximated by an extrapolation method detailed in the Section 9.3, where we discussed the mass budget for a cell.

10.7.2 Advective and diffusive fluxes

The advective velocities are used to calculate advective tracer and momentum fluxes. Advective momentum fluxes are computed using the traditional second order scheme as implied by energetic consistency (see Griffies et al. (2003)). Advective tracer fluxes are computed using a tracer advection scheme, such as one of those documented in Pacanowski and Griffies (1999) (see also Durran (1999) for a pedagogical

review). Additionally, it is necessary to compute fluxes arising from the parameterization of SGS processes. Discussion of the lateral SGS tracer fluxes is given in Chapters 7 and 11–14. Horizontal SGS momentum fluxes are discussed in Chapters 15–17. Vertical fluxes also arise from diabatic processes, such as those dominant in the surface and bottom boundary layers. The mixed layer schemes of Pacanowski and Philander (1981) and Large et al. (1994) are commonly used in ocean climate models such as MOM4 for computing vertical mixing coefficients.

For stability reasons, advective fluxes are computed on time step τ whereas lateral diffusive fluxes are on time step $\tau - \Delta\tau$. That is, the inviscid dynamics is time stepped with a leap-frog whereas the dissipative dynamics is time stepped with a single forward step. Vertical diffusive fluxes are also computed on time step $\tau - \Delta\tau$, unless they are computed implicitly in time to allow for large vertical mixing coefficients without incurring problems with time stepping instabilities.

10.7.3 Tracers

Tracers evolve according to advection, diffusion, convection, and boundary forcing. For coarse mesh models, the tracer equation can be time stepped with much longer steps than baroclinic time steps since the tracer transport mechanisms are far slower than either the inertial period or baroclinic gravity wave speed, both of which affect our ability to increase the baroclinic time step. Long tracer time steps help tremendously in the quest to reach equilibrium in the deep ocean, which can take many thousands of model years. Details of the transient evolution, however, are corrupted by using unequal time steps. The papers by Bryan (1984), Killworth et al. (1984), and Danabasoglu et al. (1996) detail these issues.

There are two practical reasons to expect that future ocean climate models will not stretch their tracer time step substantially larger than their baroclinic time step. First, as the grid mesh is refined, advective speeds become comparable to the baroclinic gravity wave speeds. Hence, time step constraints for the tracer and baroclinic velocity become more comparable. In this case, stretching the tracer time step longer than the baroclinic time step becomes less feasible.

As with the linear momentum, MOM4 time steps tracers using a leap-frog scheme with a Robert-Asselin filter to suppress the computational mode. Details are provided in Chapter 9. Updating tracers before momentum provides for the updated hydrostatic pressure $p(\tau + \Delta\tau)$. Having the updated pressure allows the leap-frog time stepped momentum equations to exploit the *pressure gradient averaging* approach of Brown and Campana (1978). Pressure gradient averaging is a favored approach of the Los Alamos Parallel Ocean Program (POP), as it can damp baroclinic gravity waves, thus allowing for an increase in model time step when gravity waves set the most stringent CFL constraint.

10.7.4 Momentum

After tracers are updated, momentum can be time stepped according to the split-explicit scheme described earlier in this chapter. Again, one method for speeding up this part of the model is to employ the Brown and Campana (1978) pressure gradient averaging approach for both the baroclinic and barotropic sub-systems. However, pressure gradient averaging may not be useful if details of the gravity

wave propagation are of interest. Tests should be run to compare the simulation with and without pressure gradient averaging to garner confidence in the results.

At the end of the momentum time stepping, we have the updated velocity fields as well as the updated ocean depth. This information is sufficient to then start the next time step.

10.8 Rigid lid streamfunction method

The rigid lid streamfunction method was described in Bryan (1969) for use in Boussinesq ocean models. This method has been remarkably versatile for z-coordinate ocean climate models. Hence, we are motivated to provide here some comments on this method, even though it is steadily becoming obsolete.

10.8.1 Volume transport streamfunction

The basic assumption underlying the rigid lid method is that the time tendency of the ocean surface height is zero. Setting $\eta_{,t} = 0$ eliminates the fast barotropic gravity waves, and this is the key motivation for this assumption.

For the Boussinesq fluid, setting $\eta_{,t} = 0$ leads to a balance between the divergence of horizontal transport and surface fresh water forcing

$$\nabla \cdot \mathbf{U} = q_w. \tag{10.79}$$

Although Huang (1993) described a rigid lid method allowing for this balance, most rigid lid models set the surface fresh water flux to zero, thus rendering $\mathbf{U}$ divergence-free

$$\nabla \cdot \mathbf{U} = 0. \tag{10.80}$$

Hence, the rigid lid Boussinesq model is fully volume conserving: all grid cells conserve volume and there are no fluxes of volume across the boundaries. Correspondingly, the vertical advective velocity (Section 9.3.1) vanishes at the surface. Herein lies the first major limiting aspect of the rigid lid streamfunction method: the effects of surface fresh water forcing must be incorporated into the model in an indirect, and often unphysical, manner. In particular, the effects on salinity are imposed via virtual salt fluxes across the ocean surface (see Huang (1993) and Griffies et al. (2001) for more discussion).

Since $\nabla \cdot \mathbf{U} = 0$, we can specify $\mathbf{U}$ by a single scalar *barotropic streamfunction*

$$\mathbf{U} = H\,\overline{\mathbf{u}}^z = \nabla \wedge (\hat{\mathbf{z}}\psi) = -\hat{\mathbf{z}} \wedge \nabla\psi. \tag{10.81}$$

Note that in this expression, it is assumed that the depth of ocean fluid is equal to the depth of a resting fluid H. However, this assumption is not strictly true since $\eta_{,t} = 0$ does not imply $\eta = 0$. Indeed, there is a non-zero surface or *lid* pressure $p_{lid} = \rho_o\, g\, \eta$ which, although not needed for the streamfunction method, is computed via the *rigid lid surface pressure* method of Smith et al. (1992), Dukowicz et al. (1993), and Pinardi et al. (1995).

The barotropic streamfunction is specified to within a constant, and so only differences are physically relevant. In particular, Stokes' Theorem implies that the

vertically integrated advective transport between two points is given by

$$T_{ab} = \int_a^b dl\ \hat{\mathbf{n}} \cdot \int_{-H}^0 dz\ \mathbf{u} = \psi_b - \psi_a, \tag{10.82}$$

where dl is the line element along any path connecting the points a and b, and $\hat{\mathbf{n}}$ is a unit vector pointing perpendicular to the path in a rightward direction when facing the direction of integration. As written, T_{ab} has units of $volume/time$, and so it represents a volume transport. Therefore, the difference between the barotropic streamfunction at two points represents the vertically integrated volume transport between the two points. It is for this reason that the barotropic streamfunction is sometimes called the *volume transport streamfunction*. Note that Bryan (1969) defined the barotropic streamfunction with an extra factor of the Boussinesq reference density ρ_o, such that his mass transport streamfunction has the dimensions of $mass/time$ rather than volume per time. The difference is trivial for a Boussinesq fluid.

10.8.2 Time stepping the streamfunction

To make use of the streamfunction ψ for time stepping the dynamics, we need to develop a prognostic equation for ψ. Such is formed by taking the curl of the momentum equation describing the dynamics of fluid columns (Section 3.8), and so forming the barotropic vorticity equation. This equation involves an elliptic operator that must be inverted in order to specify the updated transport $\mathbf{U}(\tau + \Delta\tau)$ given the updated streamfunction $\psi(\tau + \Delta\tau)$. Inverting the elliptic operator is quite difficult when forcing the ocean with realistic time dependent fluxes and using realistic geometry with multiple islands and bottom topography with realistic steepness. Herein lies the second major problem with the rigid lid streamfunction method: it leads to a difficult elliptic problem when running realistic ocean climate simulations. Furthermore, the elliptic problem involves specification of Dirichlet boundary conditions (the *island integrals*), and such non-local boundary conditions are very inefficient to solve on parallel computer architectures.

10.8.3 Closing comments

Given the limitations of the rigid lid streamfunction documented here, most ocean climate modelers are moving towards the less restrictive free surface methods. We documented a split-explicit approach in this chapter, and it notably does not involve the inversion of an elliptic operator. In contrast, the split-implicit approaches (e.g., Dukowicz and Smith (1994) and Campin et al. (2003)) involve an elliptic operator that must be inverted. In general, elliptic operator inversions are less efficient on parallel computer architectures than split-explicit algorithms, with Griffies et al. (2001) providing examples where this is the case.

Part IV

Z-MODEL NEUTRAL PHYSICS

Z-model Neutral Physics

One of the most important realizations of ocean climate modelers in the past two decades is the importance of closure terms in the tracer equation for coarsely resolved z-model simulations. The early approach whereby one uses a diffusion operator oriented according to geopotential surfaces will, over time, greatly compromise the simulation's physical integrity due to spuriously large mixing between simulated water masses. Given the highly ideal fluid dynamics of the ocean interior, most of the tracer transport by mesoscale eddies occurs along a locally referenced potential density direction, otherwise known as a *neutral direction*. Changing the orientation of the diffusion operator from horizontal-vertical to neutral-vertical greatly improves the simulation's realism. Besides a symmetric diffusive aspect to the sub-grid scale (SGS) tracer transport, mesoscale eddies stir tracers in a reversible manner, and this stirring corresponds to an anti-symmetric component to the SGS tracer transport tensor. Adding this *skew-diffusive* component again improves the simulations.

The combination of neutral diffusion and skew-diffusion, and various other related schemes whose structure depends on neutral directions, are generally termed *neutral physics* schemes in this book. It is interesting to note that many isopyncal models have been forced, through numerical constraints, to employ similar SGS operators to those neutral physics schemes now ubiquitous in z-models.

With the extremely small levels of irreversible mixing of potential density classes within much of the ocean interior, it is crucial that ocean climate models respect these levels in both its explicit physical closure schemes as well as the numerical realizations of all tracer transport operators. Unfortunately, refining grid resolution so that mesoscale eddies are admitted can actually *increase* the problems with spurious mixing in z-models. The reason is that an explicitly resolved quasi-geostrophic turbulent flow cascades tracer variance towards the grid scale, and this variance must be dissipated without incurring unphysically large levels of spurious diapycnal mixing. Consequently, just as for the coarsely resolved models, refined models are highly dependent on their SGS operators, even when the operators simply aim to satisfy the needs of *numerical closure*. A cavalier choice for how to close the equations is not sensible.

To maintain a small level of spurious mixing, one possible route for eddying ocean models is to continue using neutral physics operators introduced for physical reasons at the coarse resolution. At the fine resolution, these operators take on a role as a numerical closure by dissipating variance without admitting large spurious levels of mixing. This perspective motivates our emphasis on numerically realizing neutral physics operators with the highest degree of physical and numerical integrity since they are conjectured here to be appropriate for all model resolutions. Furthermore, because of the strong sensitivity of simulations to details of the neutral physics scheme, we provide a thorough treatment of all facets in this part of the book.

In regards to discretizing the neutral diffusion operator, it is useful to highlight a mathematical connection between this operator and the friction operator discussed in Part V. Namely, both operators are self-adjoint. Hence, they correspond to the functional derivative of a sign definite functional. Having a functional formulation provides for a "higher principle" that can be used when discretizing these opera-

tors. This principle is akin to an energy principle which is often used to guide the discretization of the inviscid part of the model equations. For the neutral physics and friction operators, the principle states that the neutral diffusion and friction operators should dissipate, respectively, variance and energy for all admitted flow features, and the skew diffusion operator should keep variance unchanged. This principle helps us to organize the many steps encountered when deriving the discrete operators, and thus it aids in our quest to remove arbitrary steps often made when casting the continuum equations onto the lattice.

The chapters in this part of the book aim to bring together the ideas, both physical and numerical, comprising neutral physics in z-models. There are many details and choices. Some of these issues are well known, yet others are often hidden. Care is given to expose many of the issues in hopes that the reader will either come to understand what may appear to be an arbitrary choice, or to find fault with some choice and to thus find a better approach.

CHAPTER
ELEVEN

Neutral transport: Part I

Contents

11.1 The general utility of neutral physics operators **241**
11.2 Summary of the mass and tracer budgets **243**
11.3 Compatibility in the mean field budgets **244**
11.4 The SGS transport tensor **244**
11.4.1 Evolution of tracer variance 245
11.4.2 Diffusive and skew-diffusive tracer fluxes 245
11.5 Advection and skewsion **246**
11.6 Neutral tracer fluxes **247**
11.6.1 Specifying the neutral directions 247
11.6.2 Geometric relations for stably stratified density layers . . . 248
11.6.3 Geometric relations for arbitrarily stratified density layers 249

The purpose of this chapter is to expose some of the basic physical and mathematical aspects of tracer transport associated with neutral physics in z-models. Additionally, we argue here that neutral physics operators, though motivated from the mesoscale parameterization problem, are of use for all z-coordinate model resolutions. Hence, we work here within the context of the model equations derived in Chapter 6 resulting from an average over the small scales of motion.

11.1 The general utility of neutral physics operators

As discussed in Chapter 7, the ocean is very close to an ideal fluid away from boundaries. By this we mean that tracer transport in the ocean interior, dominated by mesoscale eddies, occurs predominantly along neutral directions (Section 11.6), and the mass of fluid living between two isopycnals remains close to constant. Regardless the ocean model mesh size, it is crucial to respect this property within the model's tracer transport operators.

When running coarse mesh models, there are important sub-grid scale (SGS) fluxes that must be parameterized. Ocean climate model simulations are very sensitive to the form of these parameterizations. The work of Solomon (1971), Redi (1982), Olbers et al. (1985), McDougall and Church (1986), McDougall (1987a), Gent and McWilliams (1990), Gent et al. (1995), Griffies et al. (1998), Griffies (1998), and others stress the importance of neutral operators, whereby diffusion and skew diffusion is oriented according to the neutral directions. The use of such operators in coarse z-models has greatly improved their physical integrity (see Griffies et al. (2000a) for a review).

Models of refined grid mesh that admit time dependent quasi-geostrophic eddies tranform tracer variance towards the grid scale. Hence, such models require SGS operators or implicit dissipation within an advection scheme to absorb tracer variance at the grid scale. Otherwise, the simulation degenerates into a sea of grid noise. Coupled to this need to dissipate variance, we must also maintain physically very small levels of diapycnal mixing in the ocean interior. Roberts and Marshall (1998) and Griffies et al. (2000b) showed how these two requirements are often at odds with one another in z-coordinate ocean models. Notably, relying on horizontally oriented SGS operators or dissipative advection schemes to absorb grid scale variance can introduce unphysically large levels of diapycnal mixing, thus compromising the physical integrity of the solution, especially those run for climate purposes.

Hence, alternatives must be considered. We contend that the *same* neutral operators employed at coarse resolution, motivated from such ideas as those introduced in Chapter 7, also provide a suitable *numerical closure* for fine resolutions when run with suitably reduced diffusivities. Hence, *even for eddying models, neutral physics schemes represent good z-coordinate modeling practice*. Stated alternatively, whatever may arise from *the ultimate* in eddy parameterization scheme(s), it is conjectured that the basics of the neutral diffusion scheme of Redi (1982) and the skew diffusion scheme of Gent and McWilliams (1990) and Gent et al. (1995) will remain of primary importance for ocean climate models *of all resolutions*–they are the least common denominator.

It is worth highlighting that isopycnal models *must* run with some means to smooth potential density surfaces, or equivalently to smooth density interface height. Otherwise, the layers can, and will, develop negative thicknesses. Since interface height smoothing is directly analogous to skew-diffusion of potential density*, *all* isopycnal models use some form of Gent and McWilliams (1990). When transporting tracers, they likewise diffuse tracers along isopycnal directions, in a manner analogous to that proposed by Solomon (1971) and Redi (1982).

In the context of the above conjecture regarding the universal utility of neutral physics schemes in z-models, we speculate on the possible utility of schemes that prompt one to insert eddy-stirring effects into the momentum equation as an enhanced vertical friction (see, e.g., Greatbatch and Lamb (1990) and Greatbatch (1998)), rather than the commonly implemented eddy-induced advection or skew-sion applied directly to the tracer equation. The vertical friction schemes are abia-

*Note that for flat bottom rigid lid models, interface height smoothing is equivalent to thickness diffusion. However, in general it is crucial to note the difference (see, e.g., Holloway (1997)), with interface height smoothing the general method used in both isopycnal and z-models using the Gent and McWilliams (1990) scheme.

batic, by construction, since they only affect the momentum equation. This property has its advantages numerically since we need not worry about spurious diapycnal mixing commonly associated with numerical truncation errors in the SGS tracer operator. However, the effects of vertical momentum friction on density are only available indirectly via geostrophy. Consequently, it is unclear whether vertical friction will prove sufficient to suppress grid-scale power in the density field associated with dispersive advection schemes and/or a quasi-geostrophic cascade. Presently, no published studies document the ability of enhanced vertical friction to satisfy the needs of both physical closure and numerical closure in global ocean models.

11.2 Summary of the mass and tracer budgets

Mass and tracer budgets for unaveraged fluid parcels were derived in Chapters 3 and 4

$$\rho_{,t} + \rho_o \, \nabla \cdot \tilde{\mathbf{v}} = 0 \tag{11.1}$$

$$(\rho \, C)_{,t} + \rho_o \, \nabla \cdot (\tilde{\mathbf{v}} \, C) = 0, \tag{11.2}$$

where sources and molecular tracer diffusion are ignored for brevity, and

$$\rho_o \, \tilde{\mathbf{v}} = \rho \, \mathbf{v} \tag{11.3}$$

is the linear momentum per volume of a fluid parcel. Compatibility between mass and tracer budgets manifests by having the tracer budget reduce to the mass budget in the case when the tracer field is everywhere constant.

In chapter 6 we derived the ensemble mean primitive equations, with the ensemble mean mass and tracer budgets

$$\langle \rho \rangle_{,t} + \rho_o \, \nabla \cdot \langle \tilde{\mathbf{v}} \rangle = 0 \tag{11.4}$$

$$(\langle \rho \rangle \, \langle C \rangle^{\rho})_{,t} + \rho_o \, \nabla \cdot (\langle \tilde{\mathbf{v}} \rangle \, \langle C \rangle^{\rho}) = -\rho_o \, \nabla \cdot \langle \tilde{\mathbf{F}}_{sgs} \rangle \tag{11.5}$$

where the SGS tracer flux is given by

$$\rho_o \, \tilde{\mathbf{F}}_{sgs} = \rho \, C'_{\rho} \, \mathbf{v}'_{\rho} \tag{11.6}$$

and interior sources of tracer are again omitted for brevity. To reach the averaged tracer budget, we introduced the density weighted ensemble mean fields $\langle \rho \rangle \langle C \rangle^{\rho} = \langle \rho \, C \rangle$ and the fluctuation $C'_{\rho} = C - \langle C \rangle^{\rho}$ that satisfies $\langle \rho \, C'_{\rho} \rangle = 0$. Similar definitions apply to the density weighted velocity field.

For our concerns, the transport arising from $\rho \, C'_{\rho} \, \mathbf{v}'_{\rho}$ dominates that from molecular diffusion, so molecular diffusion is assumed to be incorporated into $\tilde{\mathbf{F}}_{sgs}$ without loss of generality. For those models not resolving mesoscale eddies, the piece of $\tilde{\mathbf{F}}_{sgs}$ associated with mesoscale eddies is interpreted in a manner introduced in Chapter 7 and further described by McIntosh and McDougall (1996), McDougall and McIntosh (2001), and Greatbatch and McDougall (2002). For those models that do resolve mesoscale eddies, we argued in Section 11.1 that the forms motivated by the eddy closure problem are also of relevance for these refined models. We therefore base our discussions in this chapter on this hypothesis, thus allowing for a unified treatment of the neutral tracer operators for all grid resolutions.

11.3 Compatibility in the mean field budgets

As noted above, compatibility between mass and tracer budgets for unaveraged fluid parcels manifests by having the tracer budget reduce to the mass budget in the case when the tracer field is everywhere constant. We insist on an analogous compatibility for the mean fluid so that the mean fluid retains mass and tracer conservation regardless the SGS fluxes. Doing so constrains the form of the SGS flux.

Compatibility for the ensemble averaged mass and tracer budgets is satisfied if $\nabla \cdot \langle \tilde{\mathbf{F}}_{sgs} \rangle = 0$ when $\langle C \rangle^{\rho}$ is everywhere constant. In the general case of a nontrivial mean tracer field, compatibility is maintained when the SGS tracer flux takes the form

$$\langle \tilde{\mathbf{F}}_{sgs} \rangle = \nabla \wedge \mathbf{B} - \mathbf{J} \cdot \nabla \langle C \rangle^{\rho}, \tag{11.7}$$

where $\mathbf{B}$ is an arbitrary vector field, and $\mathbf{J}$ is a second order *tracer transport tensor*. In general, the transport tensor is a function of the fluid flow and tracer field. Terms of higher order in the tracer derivative also satisfy compatibility, yet they are dropped for simplicity.

Since the divergence of $\nabla \wedge \mathbf{B}$ vanishes, it does not contribute to the time tendency of density weighted tracer. The freedom to add or subtract a curl to the tracer flux is exploited in the following. For present purposes it is dropped, thus bringing the m'th component of the SGS flux to the form

$$\langle \tilde{F}^{m}_{sgs} \rangle = -J^{mn} \, \partial_n \langle C \rangle^{\rho}, \tag{11.8}$$

where J^{mn} are components to the tracer transport tensor $\mathbf{J}$, and the summation convention is assumed, in which case repeated indices are summed over their range.

Given these results, the mapping defined by Table 6.1 leads to the mass and tracer equations to be discretized by the ocean model

$$\rho_{,t} + \rho_o \nabla \cdot \mathbf{v} = 0 \tag{11.9}$$

$$(\rho \, C)_{,t} + \rho_o \, \nabla \cdot (\mathbf{v} \, C) = -\rho_o \, \nabla \cdot \mathbf{F} \tag{11.10}$$

where

$$F^m = -J^{mn} \, C_{,n} \tag{11.11}$$

is the model's version of the SGS tracer flux. Notably, both the resolved and SGS tracer transport operators take the same mathematical form for the non-Boussinesq and Boussinesq fluids. The remainder of this chapter discusses mathematical and physical properties of this tracer flux.

11.4 The SGS transport tensor

We establish some general properties of the second order tracer transport tensor $\mathbf{J}$, decomposed as

$$J^{mn} = (J^{mn} + J^{nm})/2 + (J^{mn} - J^{nm})/2 = K^{mn} + A^{mn}, \tag{11.12}$$

where K^{mn} is the symmetric part, and A^{mn} is the anti-symmetric part. The tracer equation therefore takes the form

$$(\rho \, C)_{,t} + \rho_o \cdot (\mathbf{v} \, C) = \rho_o \, \nabla \cdot [(\mathbf{K} + \mathbf{A}) \cdot \nabla C] \tag{11.13}$$

As seen in the following, the transport properties of the symmetric and anti-symmetric tensors are quite distinct.

11.4.1 Evolution of tracer variance

We start by examining the behaviour of tracer vaiance, defined as

$$\mathcal{V} = M^{-1} \int (\rho \, dV) \, C^2 - \left(M^{-1} \int (\rho \, dV) \, C \right)^2 , \tag{11.14}$$

where

$$M = \int (\rho \, dV) \tag{11.15}$$

is the ocean fluid mass. Our concern here is how tracer variance evolves from the effects of SGS fluxes. Hence, it is sufficient to consider the case of an ocean with constant fluid mass and constant tracer mass. Tracer variance thus evolves according to evolution of the first term in equation (11.14), prompting us to drop the second term and write

$$\mathcal{V} = M^{-1} \int dA \int_{-H}^{\eta} dz \, \rho \, C^2 . \tag{11.16}$$

Use of mass conservation in the form $\rho_{,t} + \rho_o \, \nabla \cdot \mathbf{v} = 0$ leads to the evolution equation

$$(\rho \, C^2)_{,t} + \rho_o \, \nabla \cdot (\mathbf{v} \, C^2) = -2 \, \rho_o \, C \, \nabla \cdot \mathbf{F} . \tag{11.17}$$

Use of the surface and bottom kinematic boundary conditions

$$\rho \, \eta_{,t} + \rho_o \, \mathbf{u} \cdot \nabla \eta = \rho_o \, w + \rho_w \, q_w \qquad \text{at } z = \eta \tag{11.18}$$

$$\mathbf{u} \cdot \nabla H + w = 0 \qquad \text{at } z = -H \tag{11.19}$$

and manipulations similar to those performed in the calculation of global energy budgets in Section 4.5 lead to

$$M \, \mathcal{V}_{,t} = \int_{z=\eta} dA \, (\rho_w \, q_w \, C^2 - \rho_o \, C \, \mathbf{F} \cdot \hat{\mathbf{N}}) + \rho_o \int dA \int_{-H}^{\eta} dz \, \nabla C \cdot \mathbf{F} \tag{11.20}$$

where $\hat{\mathbf{N}} = (-\nabla \eta, 1)$ is an orientation vector at the ocean surface, and we assumed there is no SGS flux of tracer at the ocean bottom.

11.4.2 Diffusive and skew-diffusive tracer fluxes

In addition to the affects from surface terms, tracer variance in equation (11.20) changes according to the global integral of

$$\nabla C \cdot \mathbf{F} = -(A^{mn} + K^{mn}) \, C_{,n} \, C_{,m} = -K^{mn} \, C_{,n} \, C_{,m} \tag{11.21}$$

where $A^{mn} \, C_{,n} \, C_{,m} = 0$ due to the anti-symmetry of A^{mn} and the symmetry of $C_{,n} \, C_{,m}$. A particular form of the symmetric transport tensor of use in ocean modeling is one that is positive semi-definite, for which we have

$$\nabla \, C \cdot \mathbf{K} \cdot \nabla \, C = C_{,m} \, K^{mn} \, C_{,n} \geq 0 . \tag{11.22}$$

Transport tensors satisfying this property are known as *diffusion tensors*. The associated *diffusive* flux is directed down the gradient of the tracer

$$\nabla C \cdot \mathbf{F} \leq 0. \tag{11.23}$$

Such fluxes are dissipative since within the ocean interior, they reduce the value of $\int (\rho \, dV) \, C^{2n}$, with n a positive integer. Symmetric negative definite SGS tracer transport tensors lead to unstable simulations, unless they are suitably balanced by a companion dissipation operator. We only consider positive semi-definite symmetric transport tensors in the following.

The anti-symmetric transport tensor identically dropped out from the tendency for tracer variance. The reason is that its associated tracer flux $F^m = -A^{mn} \, C_{,n}$ is directed perpendicular to the tracer gradient

$$\mathbf{F} \cdot \nabla C = -A^{mn} \, C_{,n} \, C_{,m} = 0. \tag{11.24}$$

Tracer fluxes with this orientation are denoted *skew* fluxes in the following. Skew fluxes do not alter the value of $\int (\rho \, dV) \, C^{2n}$. As such, they are akin to advection, which acts to stir the tracer without mixing.

11.5 Advection and skewsion

There are two complementary ways of interpreting the transport process associated with an anti-symmetric tensor A^{mn}: either via the convergence of an advective flux, known as *advection*, or the convergence of a skew-flux, called here *skew-diffusion* or *skewsion*. These issues were presented in Section 7.2. Because of the importance of understanding these two perspectives in the formulation of neutral physics operators, and motivated by a desire to keep parts of this book effectively independent of one another, we consider here an abbreviated version of the discussion given more completely in Section 7.2.

To start, we consider an arbitrary three-dimensional divergence-free velocity field $\nabla \cdot \mathbf{V} = 0$. The divergence-free condition represents a diagnostic relation, or constraint, that reduces the functional degrees of freedom for the velocity field from three to two. This constraint can be satisfied identically by introducing a vector streamfunction

$$\mathbf{V} = \nabla \wedge \boldsymbol{\Upsilon}. \tag{11.25}$$

The vector streamfunction $\boldsymbol{\Upsilon}$ is not completely specified by this relation, since the equally valid streamfunction

$$\boldsymbol{\Upsilon}' = \boldsymbol{\Upsilon} + \nabla \lambda \tag{11.26}$$

corresponds to the same velocity field $\mathbf{V}$. The arbitrary scalar function λ is known as a *gauge function*, and the freedom to modify the vector streamfunction through the addition of λ is known as *gauge freedom*.

The *advective tracer flux*

$$\mathbf{F}_{(a)} = \mathbf{V} \, C = (\nabla \wedge \boldsymbol{\Upsilon}) \, C \tag{11.27}$$

can be related to a skew flux

$$\mathbf{F}_{(s)} = -\nabla C \wedge \boldsymbol{\Upsilon} = \mathbf{F}_{(a)} - \mathbf{F}_{(r)} \tag{11.28}$$

through exploitation of the identity

$$\mathbf{V}\, C = (\nabla \wedge \boldsymbol{\Upsilon})\, C = -\nabla C \wedge \boldsymbol{\Upsilon} + \nabla \wedge (C\, \boldsymbol{\Upsilon}). \tag{11.29}$$

$\mathbf{F}_{(s)}$ qualifies as a skew flux since it is directed perpendicular to the tracer gradient $\mathbf{F}_{(s)} \cdot \nabla C = 0$. Since the *rotational flux*

$$\mathbf{F}_{(r)} = \nabla \wedge (C\, \boldsymbol{\Upsilon}) \tag{11.30}$$

has a vanishing divergence, the divergence of the skew and advective fluxes is identical

$$\nabla \cdot \mathbf{F}_{(s)} = \nabla \cdot \mathbf{F}_{(a)}. \tag{11.31}$$

Hence, if these fluxes enter into the evolution of a tracer, one may choose to use the skew flux or advective flux in describing the evolution. That is,

$$\begin{aligned}(\rho\, C)_{,t} + \rho_o\, \nabla \cdot (\mathbf{v}\, C) &= -\rho_o\, \nabla \cdot (\mathbf{V}\, C) \\ &= \rho_o\, \nabla \cdot (\nabla C \wedge \boldsymbol{\Upsilon}).\end{aligned} \tag{11.32}$$

We make contact with results in Section 11.4.2 by introducing the anti-symmetric transport tensor

$$F^m_{(s)} = -\epsilon^{mnp}\, C_{,n}\, \Upsilon_p = -A^{mn}\, C_{,n} \tag{11.33}$$

where A^{mn} represents a reorganization of the vector streamfunction

$$A^{mn} = \epsilon^{mnp}\, \Upsilon_p = \begin{pmatrix} 0 & \Upsilon_3 & -\Upsilon_2 \\ -\Upsilon_3 & 0 & \Upsilon_1 \\ \Upsilon_2 & -\Upsilon_1 & 0 \end{pmatrix}, \tag{11.34}$$

and ϵ^{mnp} is the totally anti-symmetric Levi-Civita tensor defined in Section 18.12. For our applications, we exploit gauge freedom by specifying a *vertical gauge* in which $\Upsilon_3 = 0$.

11.6 Neutral tracer fluxes

In the formulation of neutral physics transport operators in a z-model, it is useful to understand both how neutral directions are determined, and how the geometry of fluxes aligned according to these directions relate to the fixed x, y, z directions. In particular, we are frequently concerned with rotation from the z-coordinate frame to neutral directions.

11.6.1 Specifying the neutral directions

As discussed in Section 8.2.1, neutral directions represent directions in which the adiabatic displacement of a fluid parcel experiences no resistance from buoyancy forces. That is, neutral directions are parallel to lines of constant buoyancy. McDougall (1987a) systematically defined these directions and noted how they can differ from the certain potential density surfaces in many parts of the ocean.

In a fluid with a linear equation of state, buoyancy and density are linearly proportional. In this case, neutral directions are tangent to the density surface, and so are orthogonal to the local normal given by

$$\hat{\mathbf{n}} = \hat{\boldsymbol{\rho}} = \frac{\nabla \rho}{|\nabla \rho|}. \tag{11.35}$$

In a fluid with a nonlinear equation of state $\rho = \rho(\theta, s, p)$, such as seawater, pressure effects make it necessary to carefully specify how one computes density gradients to determine neutral directions. In particular, it is necessary to perform the gradient operation in a manner that removes pressure effects, just as done when checking for gravitational stability. That is, with pressure held fixed (i.e., locally referenced),

$$\nabla \rho = \nabla \theta \left(\frac{\partial \rho}{\partial \theta}\right)_{s,p} + \nabla s \left(\frac{\partial \rho}{\partial s}\right)_{\theta,p} = \rho\,(-\alpha\,\nabla \theta + \beta\,\nabla s). \tag{11.36}$$

In these relations, θ is the potential temperature, s is the salinity, p is the pressure, and partial derivatives of density are taken with the specified variables held fixed. The functions

$$\alpha = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial \theta}\right)_{s,p} \tag{11.37}$$

and

$$\beta = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial s}\right)_{\theta,p} \tag{11.38}$$

are thermal expansion and saline contraction coefficients. These functions can typically be determined analytically from an empirical expression for the equation of state, and such is useful to do in a numerical model. Note that the local normal is perpendicular to potential temperature surfaces when salinity is uniform.

As articulated by McDougall (1987a), neutral *surfaces* represent an accumulation of tangents to locally referenced potential density surfaces. However, McDougall and Jackett (1988) showed that due to pressure effects, neutral surfaces possess a helical topology. This nontrivial topology precludes neutral surfaces from being a useful means to partition the vertical direction in ocean models. Instead, it is possible only to unambiguously determine the locally defined neutral direction at each point in the ocean. An approximation to such directions is realized in z-models. The helical nature of neutral surfaces leads potential density ocean models* to have isopycnal surfaces deviating from neutral.

11.6.2 Geometric relations for stably stratified density layers

Figure 11.1 shows a region on a locally referenced potential density surface. At an arbitrary point along this surface, the unit normal $\hat{\mathbf{n}}$ and tangent $\hat{\mathbf{t}}$ vectors are shown, along with the angle δ between the tangent and the horizontal axis, or equivalently between the normal and the vertical. The following considerations assume that the local tangent has some nonzero component in the horizontal directions, which is the case for a stably stratified fluid.

*Such models are based on a single globally defined potential density coordinate to partition the vertical.

In the two-dimensional x-z plane, an arbitrary vector aligned with the local tangent takes the form

$$\mathbf{F} = |\mathbf{F}|\,\hat{\mathbf{t}} = F^{(x)}\,(\hat{\mathbf{x}} + S\,\hat{\mathbf{z}}), \tag{11.39}$$

where

$$F^{(x)} = |\mathbf{F}|\,\cos\delta \tag{11.40}$$

is the horizontal component of the vector, and

$$S = \tan\delta = z_{,x} = -\rho_{,x}/\rho_{,z} \tag{11.41}$$

is the slope of the neutral direction with respect to the horizontal. In three-dimensions, this slope takes the form

$$\mathbf{S} = \nabla_\rho z = -\left(\frac{\nabla_z \rho}{\rho_{,z}}\right) = -\left(\frac{-\alpha\,\nabla_z\theta + \beta\,\nabla_z s}{-\alpha\,\theta_{,z} + \beta\,s_{,z}}\right), \tag{11.42}$$

where ∇_ρ represents the horizontal gradient taken along the neutral direction, and ∇_z is the horizontal gradient along a constant z surface.

Given these general results, an arbitrary flux oriented parallel to the neutral direction at an arbitrary point in the ocean has horizontal $\mathbf{F}^{(h)}$ and vertical $F^{(z)}$ components related via

$$F^{(z)} = \mathbf{S}\cdot\mathbf{F}^{(h)}. \tag{11.43}$$

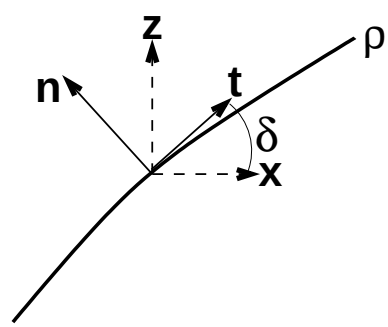

Figure 11.1: Geometry of a locally referenced potential density surface $\rho = const$. At any point along the surface, a local normal $\hat{\mathbf{n}} = \hat{\boldsymbol{\rho}}$ and tangent $\hat{\mathbf{t}}$ can be defined, with δ being the angle of the tangent from the horizontal, and the angle of the normal from the vertical. At the pressure where the locally referenced potential density surface is defined, the local tangent defines the neutral direction.

11.6.3 Geometric relations for arbitrarily stratified density layers

The previous geometric analysis, restricted to stably stratified density layers, is sufficient for large-scale ocean modeling. Nonetheless, in order to understand the geometric basis for the Redi diffusion tensor (Redi (1982)), we proceed in the following more general manner where neutral directions can be arbitrarily oriented.

For this purpose, we consider the symmetric diffusion tensor representing along and across neutral mixing given in the projection operator form

$$K^{mn} = A_I(\delta^{mn} - \hat{\rho}^m\hat{\rho}^n) + A_D\hat{\rho}^m\hat{\rho}^n, \tag{11.44}$$

where $\hat{\rho}^m$ are components to the local normal $\hat{\boldsymbol{\rho}}$ to the locally referenced potential density surface. The isoneutral and dianeutral diffusivities A_I and A_D are non-negative and can in general be functions of space-time. Although not written such in her paper, this is indeed the Redi (1982) diffusion tensor, as noted by Olbers et al. (1985) and Griffies et al. (1998).

The diffusion flux $F^m = -K^{mn}\, C_{,n}$ resulting from the Redi diffusion tensor takes the form

$$\mathbf{F} = -A_I[\nabla C - \hat{\boldsymbol{\rho}}\,(\hat{\boldsymbol{\rho}} \cdot \nabla C)] - A_D\, \hat{\boldsymbol{\rho}}\,(\hat{\boldsymbol{\rho}} \cdot \nabla C). \tag{11.45}$$

The first term, $\nabla C - \hat{\boldsymbol{\rho}}\,(\hat{\boldsymbol{\rho}} \cdot \nabla C)$, is perpendicular to the local normal $\hat{\boldsymbol{\rho}}$, and so it represents that part of the tracer field oriented parallel to the neutral direction. The second term, $\hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}} \cdot \nabla C)$, is parallel to the local normal, and so it represents that part of the tracer field perpendicular to the neutral direction. This result renders an intuitive geometric interpretation for the Redi diffusion tensor.

CHAPTER
TWELV

Neutral transport: Part II

Contents

12.1 Neutral diffusion **252**
12.1.1 Two bases of unit directions 253
12.1.2 Diffusion in the neutral frame 254
12.1.3 Diffusion in the z-level frame 254
12.1.4 Small angle approximation 255
12.1.4.1 Small angle diffusion tensor 255
12.1.4.2 Small angle diffusive fluxes 256
12.1.4.3 Small angle tensor in the neutral frame 256
12.1.5 Errors with horizontal diffusion in the ocean interior . . . 257
12.1.6 Balance between active tracer neutral diffusive fluxes . . . 257
12.1.6.1 Balance with the full diffusion tensor 258
12.1.6.2 Balance with the small angle neutral diffusion tensor 258
12.1.7 Cabbeling, thermobaricity, and halobaricity 259
12.1.7.1 Cabbeling 260
12.1.7.2 Thermobaricity and halobaricity 260
12.2 Gent-McWilliams stirring **261**
12.2.1 The GM potential density skew flux 261
12.2.2 GM skewsion and small angle neutral diffusion 263
12.2.2.1 Flux arising when $A_I = \kappa$ 263
12.2.2.2 Effective horizontal and vertical diffusivities . . . 264
12.3 A summary of the neutral physics fluxes **264**
12.4 Flow dependent diffusivities **265**
12.4.1 Horizontal and vertical dependent diffusivities 265
12.4.2 Length and time scales for Held and Larichev 266
12.4.3 Length and time scales for Visbeck et al. (1997) 267
12.4.4 Length scale given by the first baroclinic Rossby radius . . 267
12.4.5 The thermal wind Richardson number 268
12.4.6 Rotational independence of the time scale 268

12.4.7 One means to provide a regularized diffusivity 269
12.4.7.1 Regularization 270
12.4.7.2 A value for the tuning parameter 271
12.4.8 A recommended approach 271
12.4.8.1 Length scale 272
12.4.8.2 Velocity scale 272
12.5 Biharmonic operators 272
12.5.1 Tracer variance 273
12.5.2 Horizontal biharmonic mixing 274
12.5.3 Neutral biharmonic mixing 275
12.5.4 Roberts and Marshall's biharmonic stirring operator . . . 276
12.5.4.1 Effects on potential energy of the RM98 operator 276
12.5.4.2 Flux for a particular density profile 278
12.5.5 Gent's biharmonic stirring operator 278
12.5.6 Neutral biharmonic filtering 279
12.5.7 Biharmonic skew filtering operator 281
12.5.8 A note on iterative Laplacian skewsion processes 281

The purpose of this chapter is to follow-up on the discussion in Chapter 11. Here, we further detail mathematical and physical properties of the neutral diffusion scheme of Redi (1982) and the skew-diffusion scheme of Gent and McWilliams (1990), Gent et al. (1995), and Griffies (1998). We also introduce some notions relevant for determining the value for the diffusivities used to set the magnitude of the neutral physics fluxes. We close this chapter with a discussion of biharmonic tracer operators, and argue for their lack of utility for ocean climate modeling.

12.1 Neutral diffusion

From the standpoint of tracer transport via mesoscale eddies, a natural set of coordinates are those defined with respect to the neutral directions. These coordinates define what is termed here the *neutral frame*. Diffusive transport in this frame is assumed to be diagonal in the formulation of Redi (1982), with diffusion along the neutral directions on the order of 10^8 times greater than dianeutral diffusion.

The simplest and most elegant way to derive the neutral diffusion tensor is through the projection operator form given by equation (11.44). The diffusion tensor is written down by inspection based on assumptions of the strength and orientation of the diffusive fluxes. Nonetheless, it is useful for many purposes to present the rotational approach used by Redi (1982) to derive this tensor. That is the purpose of this section. In addition, we highlight a key balance between the neutral diffusive fluxes of active tracers. Maintaining this balance has proven to be crucial for realizing a numerically stable neutral diffusion scheme (Griffies et al. (1998)).

12.1.1 Two bases of unit directions

As in Sections 19.12.3 and 19.11.1, we find it useful to use *physical tensor components* to represent tensors. Physical components each have the same dimensions, and they represent the components to a tensor as projected onto a basis of dimensionless unit directions $\hat{\mathbf{e}}_m$, where $m = 1, 2, 3$ for the three-dimensional spaces considered here.

There are two sets of unit directions defining the coordinate frames of reference considered here. The first one, termed here the *z-level* frame, is defined by the dimensionless unit directions

$$\hat{\mathbf{e}}_1 = \hat{\mathbf{x}} \tag{12.1}$$

$$\hat{\mathbf{e}}_2 = \hat{\mathbf{y}} \tag{12.2}$$

$$\hat{\mathbf{e}}_3 = \hat{\mathbf{z}}, \tag{12.3}$$

where $\hat{\mathbf{z}}$ is the usual vertical direction taken anti-parallel to gravity, whereas $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ represent generalized orthogonal horizontal directions. For example, with spherical coordinates, $\hat{\mathbf{x}}$ is in the longitudinal direction and $\hat{\mathbf{y}}$ is in the latitudinal direction. More general directions are available in MOM4 through the use of generalized orthogonal coordinates (see Section 19.12 or 19.11).

The second reference frame is defined by the dimensionless unit directions

$$\hat{\mathbf{e}}_{\overline{1}} = \frac{\hat{\mathbf{z}} \wedge \nabla\rho}{|\hat{\mathbf{z}} \wedge \nabla\rho|} \tag{12.4}$$

$$\hat{\mathbf{e}}_{\overline{2}} = \hat{\mathbf{e}}_{\overline{3}} \wedge \hat{\mathbf{e}}_{\overline{1}}, \tag{12.5}$$

$$\hat{\mathbf{e}}_{\overline{3}} = \hat{\boldsymbol{\rho}} = \frac{\nabla\rho}{|\nabla\rho|}, \tag{12.6}$$

where $\nabla\rho$ is computed as in equation (11.36) where pressure effects are removed. These three directions define the *neutral reference frame* determined by the fluctuating geometry of a locally referenced isopycnal surface. A similar reference frame was discussed in Section 5.4 in the chapter on generalized vertical coordinates.

Both sets of basis directions define orthonormal reference frames. Hence, the transformation between the two frames is provided by a rotation matrix, which is an orthogonal matrix (unit determinant and inverse given by the transpose). This transformation can be interpreted in terms of Euler angles as discussed by Redi (1982).

As a tensor equation, the transformation between reference frames is written

$$\hat{\mathbf{e}}_{\overline{m}} = \Lambda^{m}{}_{\overline{m}}\, \hat{\mathbf{e}}_m, \tag{12.7}$$

where the summation convention is employed. The components of the transformation matrix are directional cosines determined via

$$\Lambda^{m}{}_{\overline{m}} = \hat{\mathbf{e}}^m \cdot \hat{\mathbf{e}}_{\overline{m}}, \tag{12.8}$$

where unit directions do not distinguish between raised and lowered indices. Substituting the explicit forms for the two bases, and organizing components of the

transformation into a matrix-vector form, leads to

$$(\hat{\mathbf{e}}_{\bar{1}} \;\; \hat{\mathbf{e}}_{\bar{2}} \;\; \hat{\mathbf{e}}_{\bar{3}}) = (\hat{\mathbf{e}}_1 \;\; \hat{\mathbf{e}}_2 \;\; \hat{\mathbf{e}}_3) \begin{pmatrix} \frac{S_y}{S} & \frac{S_x}{S\sqrt{1+S^2}} & -\frac{S_x}{\sqrt{1+S^2}} \\ -\frac{S_x}{S} & \frac{S_y}{S\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} \\ 0 & \frac{S}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} \tag{12.9}$$

where

$$\mathbf{S} = \nabla_\rho z = -z_{,\rho} \nabla_z \rho = (S_x, S_y, 0) = (-\rho_{,x}/\rho_{,z}, -\rho_{,y}/\rho_{,z}, 0) \tag{12.10}$$

is the neutral slope vector (see also equation (11.42) and Figure 11.1) with magnitude S.

12.1.2 Diffusion in the neutral frame

In the neutral frame, the gradient of a tracer is written

$$\nabla C = \hat{\mathbf{e}}_{\bar{1}} \partial_{\hat{\mathbf{e}}_{\bar{1}}} C + \hat{\mathbf{e}}_{\bar{2}} \partial_{\hat{\mathbf{e}}_{\bar{2}}} C + \hat{\mathbf{e}}_{\bar{3}} \partial_{\hat{\mathbf{e}}_{\bar{3}}} C, \tag{12.11}$$

where derivatives are taken along the three orthogonal directions. Following Redi (1982), downgradient diffusion in this frame is diagonally oriented, in which case the diffusion tensor takes the form

$$K^{\overline{mn}} = \begin{pmatrix} A_I & 0 & 0 \\ 0 & A_I & 0 \\ 0 & 0 & A_D \end{pmatrix}, \tag{12.12}$$

which can also be written in terms of projection operators

$$K^{\overline{mn}} = A_I(\delta^{\overline{mn}} - \hat{\boldsymbol{\rho}}^{\overline{m}} \hat{\boldsymbol{\rho}}^{\overline{n}}) + A_D \hat{\boldsymbol{\rho}}^{\overline{m}} \hat{\boldsymbol{\rho}}^{\overline{n}}, \tag{12.13}$$

as noted in Section 11.6.3. Note that in the neutral frame, $\hat{\boldsymbol{\rho}} = (0, 0, 1)$. The diffusion operator is thus written

$$R(C) = \partial_{\hat{\mathbf{e}}_{\bar{1}}}(A_I \partial_{\hat{\mathbf{e}}_{\bar{1}}} C) + \partial_{\hat{\mathbf{e}}_{\bar{2}}}(A_I \partial_{\hat{\mathbf{e}}_{\bar{2}}} C) + \partial_{\hat{\mathbf{e}}_{\bar{3}}}(A_D \partial_{\hat{\mathbf{e}}_{\bar{3}}} C). \tag{12.14}$$

12.1.3 Diffusion in the z-level frame

In order to describe the neutral diffusion process in the z-level frame of reference, the components of the diffusion tensor $\mathbf{K}$ must be transformed. The simplest approach is to note that $\hat{\boldsymbol{\rho}} = \nabla\rho/|\nabla\rho|$ in the z-level frame, and to substitute this into the projection operator form of the tensor given by equation (12.13).

Another approach, employed by Redi (1982), is to use the familiar rules of tensor analysis (e.g., Aris (1962) or Chapter 18) to transform the representation of the second order diffusion tensor from the neutral frame to z-level frame. In this way, the diagonal components $K^{\overline{mn}}$ of the diffusion tensor in the neutral frame are transformed to the z-level frame through

$$K^{mn} = \Lambda^m{}_{\overline{m}} K^{\overline{mn}} \Lambda^n{}_{\overline{n}}, \tag{12.15}$$

where the transformation matrix Λ has components given by equation (12.9).

Written as matrices with Λ having components $\Lambda^n{}_{\overline{n}}$, this transformation takes the form $K = \Lambda \overline{K} \Lambda^T$, where $\overline{K}$ has the diagonal form given in equation (12.12). Performing the matrix multiplication

$$K^{mn} = \begin{pmatrix} \frac{S_y}{S} & \frac{S_x}{S\sqrt{1+S^2}} & -\frac{S_x}{\sqrt{1+S^2}} \\ -\frac{S_x}{S} & \frac{S_y}{S\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} \\ 0 & \frac{S}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} \begin{pmatrix} A_I & 0 & 0 \\ 0 & A_I & 0 \\ 0 & 0 & A_D \end{pmatrix} \begin{pmatrix} \frac{S_y}{S} & -\frac{S_x}{S} & 0 \\ \frac{S_x}{S\sqrt{1+S^2}} & \frac{S_y}{S\sqrt{1+S^2}} & \frac{S}{\sqrt{1+S^2}} \\ -\frac{S_x}{\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} \tag{12.16}$$

yields the diffusion tensor in the z-level system given by Redi (1982)

$$K^{mn} = \frac{A_I}{(1+S^2)} \begin{pmatrix} 1 + \frac{\rho_{,y}^2 + \epsilon \rho_{,x}^2}{\rho_{,z}^2} & (\epsilon - 1)\frac{\rho_{,x}\rho_{,y}}{\rho_{,z}^2} & (\epsilon - 1)\frac{\rho_{,x}}{\rho_{,z}} \\ (\epsilon - 1)\frac{\rho_{,x}\rho_{,y}}{\rho_{,z}^2} & 1 + \frac{\rho_{,x}^2 + \epsilon \rho_{,y}^2}{\rho_{,z}^2} & (\epsilon - 1)\frac{\rho_{,y}}{\rho_{,z}} \\ (\epsilon - 1)\frac{\rho_{,x}}{\rho_{,z}} & (\epsilon - 1)\frac{\rho_{,y}}{\rho_{,z}} & \epsilon + S^2 \end{pmatrix}, \tag{12.17}$$

which can also be written

$$K^{mn} = \frac{A_I}{(1+S^2)} \begin{pmatrix} 1 + S_y^2 + \epsilon S_x^2 & (\epsilon - 1) S_x S_y & (1-\epsilon) S_x \\ (\epsilon - 1) S_x S_y & 1 + S_x^2 + \epsilon S_y^2 & (1-\epsilon) S_y \\ (1-\epsilon) S_x & (1-\epsilon) S_y & \epsilon + S^2 \end{pmatrix}. \tag{12.18}$$

The ratio of the dianeutral to isoneutral diffusivities is generally quite small in the ocean interior

$$\epsilon = A_D / A_I \approx 10^{-8}. \tag{12.19}$$

12.1.4 Small angle approximation

The bulk of the mesoscale eddying ocean maintains neutral directions no more than 5% from the horizontal. Hence, it is quite accurate to take the small angle approximation. This section details this approximation.

12.1.4.1 Small angle diffusion tensor

Gent and McWilliams (1990) first wrote down the following small angle diffusion tensor

$$K^{mn}_{\text{small}} = A_I \begin{pmatrix} 1 & 0 & -\frac{\rho_{,x}}{\rho_{,z}} \\ 0 & 1 & -\frac{\rho_{,y}}{\rho_{,z}} \\ -\frac{\rho_{,x}}{\rho_{,z}} & -\frac{\rho_{,y}}{\rho_{,z}} & \epsilon + S^2 \end{pmatrix} = A_I \begin{pmatrix} 1 & 0 & S_x \\ 0 & 1 & S_y \\ S_x & S_y & \epsilon + S^2 \end{pmatrix}. \tag{12.20}$$

Note we retained the S^2 term in the $(3,3)$ element since, although small, it is potentially larger than the small ratio $\epsilon \approx 10^{-8}$. Cox (1987) originally retained the $(1,2) = (2,1)$ elements equal to $-S_x S_y$. In the small angle approximation, however, these terms are negligible. Additionally, and most crucially, by retaining these terms, the incorrect small angle tensor of Cox (1987) diffused buoyancy whereas the full tensor does not (see Section 12.1.6). Hence, it is important to use the physically consistent form of the small angle tensor which drops the $(1,2)$ and $(2,1)$ elements.

12.1.4.2 Small angle diffusive fluxes

The horizontal and vertical components to the small angle diffusive flux are given by

$$\mathbf{F}^{(h)} = -A_I \, \nabla_\rho C \tag{12.21}$$

$$F^{(z)} = -A_I \, \mathbf{S} \cdot \nabla_\rho C - A_D \, C_{,z} \tag{12.22}$$

where

$$\nabla_\rho = \nabla_z + \mathbf{S} \, \partial_z \tag{12.23}$$

is the neutral oriented gradient operator as derived in Section 5.5.3. The neutral portion of this flux satisfies $F^{(z)} = \mathbf{S} \cdot \mathbf{F}^{(h)}$. As discussed in Section 11.6.2, this relation is satisfied by an arbitrary vector oriented along the neutral direction in a region of stable stratification.

The dianeutral diffusive term reduces under the small slope approximation to just vertical diffusive flux. As seen in the following, this approximation is quite accurate since it neglects a term proportional to $A_D \, S^2$, which is very small.

12.1.4.3 Small angle tensor in the neutral frame

Given the small angle representation of the diffusion tensor in the z-level frame, what is the representation of this tensor in the neutral frame? This tensor cannot be the diagonal form given in equation (12.12) since that form transformed into the full angle representation of equation (12.17). Using the full transformation back to the neutral frame, $K^{\overline{mn}}_{\text{small}} = \Lambda^{\overline{m}}{}_m K^{mn}_{\text{small}} \Lambda^{\overline{n}}{}_n$, or in matrix form

$$K^{\overline{mn}}_{\text{small}} = A_I \begin{pmatrix} \frac{S_y}{S} & -\frac{S_x}{S} & 0 \\ \frac{S_x}{S\sqrt{1+S^2}} & \frac{S_y}{S\sqrt{1+S^2}} & \frac{S}{\sqrt{1+S^2}} \\ -\frac{S_x}{\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} \times \tag{12.24}$$

$$\begin{pmatrix} 1 & 0 & (1-\epsilon)S_x \\ 0 & 1 & (1-\epsilon)S_y \\ (1-\epsilon)S_x & (1-\epsilon)S_y & \epsilon + S^2 \end{pmatrix} \begin{pmatrix} \frac{S_y}{S} & \frac{S_x}{S\sqrt{1+S^2}} & -\frac{S_x}{\sqrt{1+S^2}} \\ -\frac{S_x}{S} & \frac{S_y}{S\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} \\ 0 & \frac{S}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} \tag{12.25}$$

yields the orthonormal isopycnal frame representation of the small angle approximated tensor*

$$K^{\overline{mn}}_{\text{small}} = \begin{pmatrix} A_I & 0 & 0 \\ 0 & A_I(1+S^2) & 0 \\ 0 & 0 & A_D \end{pmatrix} + \frac{A_D S^2}{1+S^2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & S \\ 0 & S & 1 \end{pmatrix}. \tag{12.26}$$

The small angle approximation thus adds a small amount of along neutral mixing (the $A_I \, S^2$ addition to the $(2,2)$ element) relative to the full neutral diffusion tensor. Additionally, there is a contribution proportional to $A_D S^2$. This term means that by

*Note the rotation need not transform the $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{y}}$ directional symmetry present in the (x, y, z) form of the small angle mixing tensor into a $\hat{\mathbf{e}}_{\overline{1}} \leftrightarrow \hat{\mathbf{e}}_{\overline{2}}$ symmetry in the $(\hat{\mathbf{e}}_{\overline{1}}, \hat{\mathbf{e}}_{\overline{2}}, \hat{\mathbf{e}}_{\overline{3}})$ form. The (x, y) coordinate symmetry, however, is preserved.

approximating the dianeutral diffusive flux as a vertical diffusive flux, we are led to a difference on the order of $A_D\, S^2$, which is again trivially small. In conclusion, the small slope approximation leads to a neutral and dianeutral diffusive fluxes that differ from the full slope by terms on the order $(A_I, A_D)\, S^2$, respectively, and so it is a very accurate approximation for purposes of ocean climate modeling. Stated alternatively, the small slope tensor is consistent with the small angle approximation to within terms on the order of S^2.

12.1.5 Errors with horizontal diffusion in the ocean interior

Given the rotation formalism developed here, it is possible to consider errors made when using horizontal diffusion instead of neutral diffusion. For this purpose, let **I** represent the horizontal-vertical diffusion tensor, which is diagonal in the z-level frame. Transforming this tensor to the neutral frame

$$I^{\overline{mn}} = \begin{pmatrix} \frac{S_y}{S} & -\frac{S_x}{S} & 0 \\ \frac{S_x}{S\sqrt{1+S^2}} & \frac{S_y}{S\sqrt{1+S^2}} & \frac{S}{\sqrt{1+S^2}} \\ -\frac{S_x}{\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix} \begin{pmatrix} A_H & 0 & 0 \\ 0 & A_H & 0 \\ 0 & 0 & A_V \end{pmatrix} \times \begin{pmatrix} \frac{S_y}{S} & \frac{S_x}{S\sqrt{1+S^2}} & -\frac{S_x}{\sqrt{1+S^2}} \\ -\frac{S_x}{S} & \frac{S_y}{S\sqrt{1+S^2}} & -\frac{S_y}{\sqrt{1+S^2}} \\ 0 & \frac{S}{\sqrt{1+S^2}} & \frac{1}{\sqrt{1+S^2}} \end{pmatrix}, \tag{12.27}$$

yields

$$I^{\overline{mn}} = \frac{A_H}{1+S^2} \begin{pmatrix} 1+S^2 & 0 & 0 \\ 0 & 1+\tilde{\epsilon}S^2 & -S(1-\tilde{\epsilon}) \\ 0 & -S(1-\tilde{\epsilon}) & \tilde{\epsilon}+S^2 \end{pmatrix}, \tag{12.28}$$

where $\tilde{\epsilon} = A_V/A_H$ is the ratio of the vertical to horizontal diffusion coefficient. Therefore, diffusion with **I** in the z-level frame introduces first order in slope errors in the off diagonal terms, whereas the diagonal terms contain second order in slope errors. The error in the $(3,3)$ component, however, is the most relevant as it represents an added and potentially huge source of dianeutral diffusion. For example, with the usual diffusivity ratio $\tilde{\epsilon} \approx 10^{-8}$, modest slopes larger than $\sqrt{10^{-8}} = 10^{-4}$ are sufficient to add spurious dianeutral mixing through the S^2 term which is on the same order as $\tilde{\epsilon}$. It is for this reason that horizontal diffusion is incompatible with the hypothesis that diffusive mixing occurs predominantly along the neutral directions.

12.1.6 Balance between active tracer neutral diffusive fluxes

Neutral diffusion does not affect the buoyancy field. That is, it does not diffuse locally referenced potential density. However, neutral diffusion does separately diffuse the active tracers potential temperature and salinity along the neutral directions. As shown by equation (11.36), the neutral directions are themselves functions of the active tracers. Therefore, preserving the neutral directions while diffusing the active tracers requires a balance between the active tracer neutral diffusive fluxes.

12.1.6.1 Balance with the full diffusion tensor

The balance of active tracer neutral diffusion fluxes is a trivial consequence of the definition (11.44) of the neutral diffusion tensor, and the relation (11.36) for the local normal to the neutral direction. That is, with the neutral diffusive fluxes for potential temperature and salinity given by (recall equation (11.45))

$$\mathbf{F}(\theta) = -A_I(\nabla\theta - \hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}}\cdot\nabla\theta)) \tag{12.29}$$

$$\mathbf{F}(s) = -A_I(\nabla s - \hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}}\cdot\nabla s)), \tag{12.30}$$

we have

$$\begin{aligned}\rho(-\alpha\,\mathbf{F}(\theta) + \beta\,\mathbf{F}(s)) &= \left(\frac{\partial\rho}{\partial\theta}\right)_{s,p}\mathbf{F}(\theta) + \left(\frac{\partial\rho}{\partial s}\right)_{\theta,p}\mathbf{F}(s)\\ &= -A_I\,\rho_{,\theta}\,(\nabla\theta - \hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}}\cdot\nabla\theta)) - A_I\,\rho_{,s}\,(\nabla s - \hat{\boldsymbol{\rho}}(\hat{\boldsymbol{\rho}}\cdot\nabla s))\\ &= -A_I\,(\rho_{,\theta}\,\nabla\theta + \rho_{,s}\,\nabla s) + A_I\,\hat{\boldsymbol{\rho}}\,(\hat{\boldsymbol{\rho}}\cdot(\rho_{,\theta}\,\nabla\theta + \rho_{,s}\,\nabla s))\\ &= 0,\end{aligned} \tag{12.31}$$

where substitution of equation (11.36) for $\hat{\boldsymbol{\rho}}$ led to the cancelation. Maintaining the balance

$$\alpha\,\mathbf{F}(\theta) = \beta\,\mathbf{F}(s) \tag{12.32}$$

on the lattice is crucial for realizing a stable discrete neutral diffusion operator (Griffies et al. (1998)).

12.1.6.2 Balance with the small angle neutral diffusion tensor

Maintenance of the balance (12.32) is also afforded by small slope neutral diffusive fluxes. In this case, it is sufficient to show that the horizontal fluxes maintain a balance, since the vertical flux component is related to the horizontal through equation (11.43): $F^{(z)} = \mathbf{S}\cdot\mathbf{F}^{(h)}$. Equation (12.21) exhibits the horizontal neutral diffusive fluxes in the small slope limit

$$\mathbf{F}^{(h)}(\theta) = -A_I\,\nabla_\rho\theta \tag{12.33}$$

$$\mathbf{F}^{(h)}(s) = -A_I\,\nabla_\rho s, \tag{12.34}$$

where $\nabla_\rho = \nabla_z + \mathbf{S}\,\partial_z$ is the neutrally oriented gradient operator. Hence,

$$\rho_{,\theta}\,\mathbf{F}^{(h)}(\theta) + \rho_{,s}\,\mathbf{F}^{(h)}(s) = -A_I\,(\rho_{,\theta}\,\nabla_\rho\theta + \rho_{,s}\,\nabla_\rho s) = 0, \tag{12.35}$$

where we substituted $\nabla_\rho = \nabla_z + \mathbf{S}\partial_z$ for the derivative operator and relation (11.42) for the slope vector $\mathbf{S}$. Failure to maintain this balance proved to be the key reason that the Cox (1987) numerical realization of neutral diffusion was unstable (Griffies et al. (1998)).

12.1.7 Cabbeling, thermobaricity, and halobaricity

When diffusing active tracers with a nonlinear equation of state, locally referenced potential density is affected by three non-flux forms of irreversible mixing known as cabbeling, thermobaricity, and halobaricity (McDougall (1987b)). In this subsection, we present the mathematical form of such mixing as associated with neutral diffusion. An alternative formulation of the following, in much more detail, can be found in McDougall (1991).

Under the effects of neutral diffusion, the time tendency for locally referenced potential density is given by

$$\begin{aligned}\rho_{,t} &= \rho(-\alpha\,\theta_{,t} + \beta\,s_{,t}) \\ &= \rho[\alpha\nabla\cdot\mathbf{F}_I(\theta) - \beta\nabla\cdot\mathbf{F}_I(s)] \\ &= -\nabla(\rho\alpha)\cdot\mathbf{F_I}(\theta) + \nabla(\rho\beta)\cdot\mathbf{F}_I(s),\end{aligned} \tag{12.36}$$

where the active tracer balance $\alpha\mathbf{F}_I(\theta) = \beta\mathbf{F}_I(s)$ was used to reach the last equality. The forcing terms, which vanish for a linear equation of state and which cannot be written as the divergence of a flux, represent cabbeling, thermobaricity, and halobaricity. These processes provide irreversible, non-diffusive forms of mixing (McDougall (1987b)). Note for the special case of a single active tracer, neutral directions are aligned with iso-tracer surfaces, thus providing for a zero neutral diffusive flux of the single active tracer and therefore an absence of cabbeling, thermobaricity, and halobaricity.

It is useful to isolate the mathematical forms of the irreversible mixing, which can be done via the identities

$$\nabla s\cdot\mathbf{F}_I(\theta) = \nabla\theta\cdot\mathbf{F}_I(s), \tag{12.37}$$

$$\nabla(-\rho\alpha) = \rho_{,\theta\theta}\nabla\theta + \rho_{,\theta s}\nabla s + \rho_{,\theta\,p}\nabla p, \tag{12.38}$$

and likewise for $\nabla(\rho\beta)$. Note the presence of pressure gradients. These terms represent the effects of probing different pressure surfaces, and hence different potential density surfaces; i.e., these are the thermobaric terms. Probing different pressure surfaces is necessary when computing the spatial gradients of the thermal and saline expansion coefficients*. With these substitutions, the evolution of locally referenced potential density takes the form

$$\begin{aligned}\rho_{,t} &= \rho_{,\theta\theta}\nabla\theta\cdot\mathbf{F}_I(\theta) + \rho_{,ss}\nabla s\cdot\mathbf{F}_I(s) + 2\rho_{,\theta s}\nabla s\cdot\mathbf{F}_I(\theta) + \nabla p\cdot(\rho_{,\theta\,p}\mathbf{F}_I(\theta) + \rho_{,s\,p}\mathbf{F}_I(s)) \\ &= \nabla\theta\cdot\mathbf{F}_I(\theta)[\rho_{,\theta\theta} - 2\rho_{,\theta s}(\rho_{,\theta}/\rho_{,s}) + \rho_{,ss}(\rho_{,\theta}/\rho_{,s})^2] + \nabla p\cdot(\rho_{,\theta\,p}\mathbf{F}_I(\theta) + \rho_{,s\,p}\mathbf{F}_I(s)).\end{aligned} \tag{12.39}$$

The first part of this expression can be written as the product of two quadradic forms by introducing a vector

$$\mathbf{W} = (1, \alpha/\beta) \tag{12.40}$$

and a metric

$$\rho_{,ab} = \partial_a\partial_b\rho, \tag{12.41}$$

*Trevor McDougall kindly emphasized this point.

where the labels a, b represent the two tracer fields θ, s. This definition renders

$$\rho_{,t} = \nabla\theta \cdot \mathbf{F}_I(\theta) * \mathbf{W} \cdot \mathbf{W} + (\rho_{,\theta\, p}\mathbf{F}_I(\theta) + \rho_{,s\, p}\mathbf{F}_I(s)) \cdot \nabla p \tag{12.42}$$

where the inner product

$$\begin{aligned} \mathbf{W} \cdot \mathbf{W} &= \rho_{,ab} W^a\, W^b \\ &= \rho_{,\theta\theta} - 2\rho_{,s\theta}\, (\rho_{,\theta}/\rho_{,s}) + \rho_{,ss}\, (\rho_{,\theta}/\rho_{,s})^2 \\ &= -\alpha_{,\theta} - 2(\alpha/\beta)\alpha_{,s} + (\alpha/\beta)^2\beta_{,s} \end{aligned} \tag{12.43}$$

represents the squared length of the vector $\mathbf{W}$ on the curved potential density surface characterized locally by the metric $\rho_{,ab}$.

12.1.7.1 Cabbeling

The term

$$\text{cabbeling} = \nabla\theta \cdot \mathbf{F}_I(\theta) \times \mathbf{W} \cdot \mathbf{W} \tag{12.44}$$

represents the effects from cabbeling. As written here, cabbeling is seen to be the product of a piece associated with the downgradient neutral diffusive flux of potential temperature ($\nabla\theta \cdot \mathbf{F}_I(\theta) \leq 0$), and a piece associated with the local geometric properties intrinsic to the potential density surface ($\mathbf{W} \cdot \mathbf{W}$). In the ocean, the total or Gaussian curvature $\det(\rho_{,ab})(1 + \rho_{,\theta}^2 + \rho_{,s}^2)^{-1}$ is negative. This negative curvature renders $\mathbf{W} \cdot \mathbf{W} \leq 0$, which, when combined with downgradient neutral diffusion of potential temperature, always results in a non-negative tendency for ρ and a consequent downward dianeutral advection (McDougall (1987b)).

If the term $\nabla\theta \cdot \mathbf{F}_I(\theta)$ is sign-indefinite, then the contribution of cabbeling to the locally referenced potential density is also sign-indefinite. Sign-indefiniteness is not desirable for a sub-grid scale process associated with mixing, since mixing in the ocean generally produces a sign-definite cabbeling term. Hence, we should avoid the use of continuum mixing operators that are not down-gradient. As discussed in Section 12.5, biharmonic tracer operators are mixing operators that have sign-indefinite tracer fluxes. This is a key reason for eschewing these operators in realistic ocean climate simulations. The other reason is that the biharmonic tracer operators do not lend any advantage to the simulation beyond the Laplacian tracer operators. The crucial advantage to more scale selective operators comes from the use of biharmonic momentum friction, not biharmonic tracer diffusion.

12.1.7.2 Thermobaricity and halobaricity

The terms proportional to the pressure gradient

$$\text{thermobaricity} + \text{halobaricity} = [\rho_{,\theta\, p}\mathbf{F}_I(\theta) + \rho_{,s\, p}\mathbf{F}_I(s)] \cdot \nabla p \tag{12.45}$$

represent the effects from thermobaricity and halobaricity. These terms depend on the pressure dependence of the equation of state for seawater. Note that the thermobaric term $\rho_{,\theta\, p}\mathbf{F}_I(\theta) \cdot \nabla p$ dominates the halobaric term $\rho_{,s\, p}\mathbf{F}_I(s) \cdot \nabla p$ (McDougall (1987b)). In contrast to cabbeling, thermobaricity and halobaricity do not provide a sign definite source for locally referenced potential density.

12.2 Gent-McWilliams stirring

The SGS stirring suggested by Gent and McWilliams (1990) is based on assuming (i) SGS mesoscale eddies locally provide an adiabatic sink of available potential energy (APE) (ii) SGS mesoscale eddies act on all tracers in the same fashion. Alternative proposals continue to be the subject of intense research (see Section 11 of Griffies et al. (2000a) for a brief review with numerous references). However, this section, and the current implementation in MOM4, maintain the basic GM assumptions.

As a stirring parameterization, the GM scheme can be represented either via an advective or skew-diffusive form. Griffies (1998) showed how the skew approach is far more convenient numerically than the advective approach. This section aims to highlight some basic properties of the GM scheme.

12.2.1 The GM potential density skew flux

The simplest GM skew flux acts on potential density in an adiabatic Boussinesq ocean. Hence, it is useful to consider the equation for the large-scale potential density ρ in a way that respects the GM assumptions. For this purpose, consider the equation for the potential density field

$$\rho_{,t} = -\nabla \cdot (\mathbf{v}\,\rho) - \nabla \cdot \mathbf{F}, \tag{12.46}$$

where the first term on the right hand side represents advective stirring via the divergence-free velocity explicitly represented in the numerical model, and the second term represents the parameterized effects of the SGS baroclinic eddies. The form of the flux $\mathbf{F}$ is now to be determined according to the GM assumptions.

Our arguments here are aimed at reducing the derivation of the GM90 skew-flux to a concise geometric and local energetic argument. For this purpose, we note that a flux aligned parallel to the potential density surfaces will cause no mixing of potential density classes. Hence, such a flux will result in an adiabatic SGS operator.* Geometry discussed in Section 11.6.2 leads to the general form of this flux

$$\mathbf{F} = \mathbf{F}^{(h)} + \hat{\mathbf{z}}\,\mathbf{S} \cdot \mathbf{F}^{(h)}, \tag{12.47}$$

where

$$\mathbf{S} = -\left(\frac{\nabla_z \rho}{\rho_{,z}}\right) \tag{12.48}$$

is the slope of the potential density surface. Hence, aligning the flux along the potential density surface reduces the three unknown flux components down to two. Additionally, the relation

$$\mathbf{F} \cdot \nabla \rho = 0 \tag{12.49}$$

defines $\mathbf{F}$ as a potential density skew-flux. We re-emphasize here a geometric point brought out in Section 11.5, where we saw that every skew flux has a corresponding advective flux $\mathbf{F}_{(a)} = \mathbf{v}^* \rho$, whose form differs from the skew flux by a total curl.

*Our choice to orient the flux parallel to the potential density surfaces is indeed a sufficient condition for producing an adiabatic transport operator, though not necessary. There are other physically relevant transport fluxes that have a non-zero cross-isopycnal component, such as the advective flux $\mathbf{v}\,\rho$.

Notably, the component of the non-divergent advection velocity $\mathbf{v}^*$ aligned *perpendicular* to potential density isolines creates local changes in potential density; i.e., $\rho_{,t} = -\mathbf{v}^* \cdot \nabla\rho$. The advective flux, which leads to an adiabatic stirring operator, therefore provides an important counter-example to our assumption above that the adiabatic SGS flux needs to be aligned parallel to density isolines. We return to the advective form of GM later in this section.

We now consider the constraints on the flux imposed by assuming the operator locally and adiabatically reduces available potential energy (APE) of the flow. A state of zero APE is one with no baroclinicity: isopycnals are flat. The potential energy of the adiabatic Boussinesq system is given by

$$P = \int_V dV\, \rho\, z. \tag{12.50}$$

The tendency associated with the unknown GM flux field is given by

$$\begin{aligned} P_{,t} &= g \int_V dV\, z\, \rho_{,t} \\ &= -g \int_V dV\, (z\, \nabla \cdot \mathbf{F}) \\ &= -g \int_{-H}^{0} dz\, (z\, \partial_z F^{(z)}) \\ &= g \int_{-H}^{0} dz\, F^{(z)}, \end{aligned} \tag{12.51}$$

where $F^{(z)}$ is the vertical component to the GM-flux. Note that we assumed a flat bottom domain $H(\mathbf{x}_h) = H$ and neglected effects of the free ocean surface $z = \eta(\mathbf{x}_h, t)$. These assumptions are based on our desire to focus on baroclinic physics in the ocean interior isolated from bottom and surface effects. In order to provide a local APE sink, we require

$$F^{(z)} \leq 0, \tag{12.52}$$

where zero occurs when the isopycnals are flat. It is sufficient to construct the vertical flux component using only the potential density field itself. For a stably stratified fluid in which $\rho_{,z} < 0$, the following form provides a local APE sink required by the GM assumptions

$$F^{(z)} = \kappa\, S^2\, \rho_{,z} \tag{12.53}$$

where κ is a positive diffusivity that sets the strength of the flux, and S^2 is the squared isopycnal slope. Consequently, the horizontal flux component reduces to a downgradient diffusive flux

$$\mathbf{F}^{(h)} = -\kappa \nabla_z \rho. \tag{12.54}$$

In summary, the GM skew-flux for potential density takes the form

$$\mathbf{F} = -\kappa\, (\nabla_z \rho - \hat{\mathbf{z}}\, S^2\, \rho_{,z}). \tag{12.55}$$

Notably, when breaking the three-dimensional GM skew-flux into its horizontal and vertical components, the horizontal flux is downgradient whereas the vertical flux is upgradient. There is no instability associated with the upgradient vertical

flux, since it is the convergence of the full three-dimensional flux that creates the evolution of the potential density field.

The advective GM flux $\mathbf{v}^* \rho$ was presented in Gent et al. (1995), where the divergence-free GM *eddy-induced* velocity takes the form

$$\mathbf{v}^* = \nabla \wedge \boldsymbol{\Upsilon}_{gm} = -\partial_z (\kappa \mathbf{S}) + \hat{\mathbf{z}} \nabla_z \cdot (\kappa \mathbf{S}), \tag{12.56}$$

where the vector streamfunction is

$$\boldsymbol{\Upsilon}_{gm} = \hat{\mathbf{z}} \wedge \kappa \mathbf{S}. \tag{12.57}$$

It turns out that the advective form of the GM parameterization is significantly more cumbersome to numerically implement than the skew-flux form. The following section provides some motivation for this point, with Griffies (1998) presenting more details with numerical examples.

12.2.2 GM skewsion and small angle neutral diffusion

GM assumed that the same SGS eddy-induced velocity $\mathbf{v}^*$ acts to transport all tracers, in addition to the model's resolved scale velocity field $\mathbf{v}$. In this case, Griffies (1998) showed that for an arbitrary tracer, the GM skew flux takes the form

$$\mathbf{F} = \kappa \left(\mathbf{S}\, C_{,z} - \hat{\mathbf{z}}\, \mathbf{S} \cdot \nabla_z C \right). \tag{12.58}$$

Since $\mathbf{F} \cdot \nabla C = 0$, the GM skew flux is oriented parallel to surfaces of constant tracer.

In addition to GM stirring, we need to consider neutral diffusion where the small angle slope approximation is relevant for parameterizing mesoscale ocean dynamics. Hence, the combination of GM stirring and neutral diffusive mixing leads to the tracer flux

$$\mathbf{F} = -A_I\, \nabla_z C + (\kappa - A_I)\, \mathbf{S}\, C_{,z} - \hat{\mathbf{z}} \left((A_I + \kappa)\, \mathbf{S} \cdot \nabla_z C + A_I\, S^2\, C_{,z} \right). \tag{12.59}$$

This flux can be written in terms of the following neutral physics mixing tensor

$$\begin{pmatrix} F^{(x)} \\ F^{(y)} \\ F^{(z)} \end{pmatrix} = \begin{pmatrix} A_I & 0 & (A_I - \kappa)\, S_x \\ 0 & A_I & (A_I - \kappa)\, S_y \\ (A_I + \kappa)\, S_x & (A_I + \kappa)\, S_y & A_I\, S^2 \end{pmatrix} \begin{pmatrix} C_{,x} \\ C_{,y} \\ C_{,z} \end{pmatrix}. \tag{12.60}$$

12.2.2.1 Flux arising when $A_I = \kappa$

Making the common assumption that the diffusivities associated with SGS stirring and diffusion are equal leads to the very simple result

$$\mathbf{F} = -\kappa\, \nabla_z C - \hat{\mathbf{z}}\, \kappa \left(2\, \mathbf{S} \cdot \nabla_z C + S^2\, C_{,z} \right), \tag{12.61}$$

in which the 2×2 horizontal mixing tensor is diagonal. Hence, the horizontal tracer flux is the same as that which arises from downgradient horizontal diffusion. The simplicity of the horizontal flux component reduces the total computational cost of these two SGS schemes to less than the cost of either one alone. This is the key practical result of the Griffies (1998) paper, and it is the central reason for preferring the skew-diffusive form of GM over the advective form. Additional benefits arise since the discretized skew flux is less prone to discretization errors relative to the advective flux.

12.2.2.2 Effective horizontal and vertical diffusivities

An additional point worth raising concerns the diagnoses of an effective horizontal/vertical diffusivity from a z-model run with neutral diffusion plus GM skewsion. Those used to projecting their tracer fluxes along lines of constant geopotential may find the computation of such effective diffusivities of interest. In the special case of $A_I = \kappa$, this diagnoses becomes quite straightforward, as we have

$$\mathbf{F} = -\kappa \, \nabla_z C - K_{eff}^{(z)} \, C_{,z}. \tag{12.62}$$

That is, the effective horizontal diffusivity is simply κ, as just mentioned. Although the effective vertical diffusivity $K_{eff}^{(z)}$ is more complicated, it can be written in the reasonably elegant form

$$K_{eff}^{(z)} = \kappa \, \mathbf{S} \cdot (\mathbf{S} - 2 \, \mathbf{S}_C) \tag{12.63}$$

where $\mathbf{S}_C$ defines the slope of tracer isolines

$$\mathbf{S}_C = -\left(\frac{\nabla_z C}{C_{,z}} \right). \tag{12.64}$$

As defined, $K_{eff}^{(z)}$ can be either positive or negative, depending on the relative orientation of the neutral directions and tracer isolines. In particular, when the tracer is aligned with the neutral direction,

$$K_{eff}^{(z)} = -\kappa \, S^2 \tag{12.65}$$

which is the same negative effective diffusivity seen when considering GM skewsion acting on neutral density (i.e., equation (12.53)). It is notable that a formulation of GM according to the advective flux does not afford such a straightforward interpretation/summary of neutral diffusion plus GM skewsion as that provided here. This is another key point raised by Griffies (1998).

12.3 A summary of the neutral physics fluxes

We provide here a brief summary of the tracer fluxes introduced in the previous sections. For all cases, the mixing operator $R(C)$, with units tracer/time, is given by the convergence of the flux

$$R(C) = -\nabla \cdot \mathbf{F}(C). \tag{12.66}$$

- The small angle neutral Laplacian diffusive flux for an arbitrary tracer is given by

$$\mathbf{F} = -A \left(\nabla_\rho + \hat{\mathbf{z}} \, \mathbf{S} \cdot \nabla_\rho \right) C, \tag{12.67}$$

  where

$$\begin{aligned} \mathbf{S} &= \nabla_\rho z \\ &= -\left(\frac{\nabla_z \rho}{\rho_{,z}} \right) \\ &= -\left(\frac{-\alpha \, \nabla_z \theta + \beta \, \nabla_z s}{-\alpha \, \theta_{,z} + \beta \, s_{,z}} \right) \end{aligned} \tag{12.68}$$

is the neutral slope and

$$\nabla_\rho = \nabla_z + \mathbf{S}\,\partial_z \tag{12.69}$$

is the neutral gradient operator (see Section 5.5.3 for a derivation). Notably, the small angle neutral Laplacian diffusive flux for locally referenced potential density vanishes

$$\mathbf{F}(\rho) = 0. \tag{12.70}$$

- The Laplacian GM skew-flux for an arbitrary tracer is given by

$$\mathbf{F} = -\kappa\,(-\mathbf{S}\,\partial_z + \hat{\mathbf{z}}\,\mathbf{S}\cdot\nabla_z)\,C, \tag{12.71}$$

and the skew-flux for locally referenced potential density ρ is given by

$$\mathbf{F} = -\kappa\,(\nabla_z + \hat{\mathbf{z}}\,\mathbf{S}\cdot\nabla_z)\,\rho = -\kappa\,(\nabla_z - \hat{\mathbf{z}}\,S^2\,\partial_z)\,\rho. \tag{12.72}$$

12.4 Flow dependent diffusivities

Most global ocean climate models run in the 1990's used neutral diffusive and GM skew-diffusive mixing coefficients of equal value independent of the local flow properties: $A = \kappa$. A notable exception is when the diffusivities are tapered in regions of steep neutral direction slopes and the zero eddy flux condition is applied at the ocean surface (Gent et al. (1995), Treguier et al. (1997), Killworth (1997), Griffies et al. (1998), and McDougall and McIntosh (2001). Recent studies, however, aim to provide a diffusivity within the ocean interior in terms of length and time scales set by local properties of the model's resolved fields.

Currently, there are no strong arguments suggesting that it is best to use different diffusivities for the GM and neutral diffusion scheme, even when those diffusivities are functions of the flow. Indeed, Dukowicz and Smith (1997) provide arguments in favor of the same diffusivities. In practice, this choice has been made based on simplicity, although equal diffusivities is not a universal choice (e.g., Gordon et al. (2000)).

12.4.1 Horizontal and vertical dependent diffusivities

The papers by Held and Larichev (1996) and Visbeck et al. (1997) each propose methods for computing a diffusivity in terms of a length and time scale. Properties of the resolved flow that are local in the horizontal yet nonlocal in the vertical determine these length and time scales. Killworth (1997), following on a suggestion from Treguier et al. (1997), presents a theory based on linear baroclinic instability in which the diffusivity obtains both a horizontal and vertical dependence. Notably, closures based on diffusing potential vorticity downgradient *must* incorporate a nontrivial vertical dependence to their diffusivities in order to maintain basic kinematic balances (e.g.,Marshall and Shutts (1981), Treguier et al. (1997), Killworth (1997)). Results from the high resolution channel model study by Treguier (1999) are consistent with Killworth's ideas. Smith and Vallis (2002) also provide a vertical-horizontal dependent diffusivity for potential vorticity, with their theory based on scalings from fully developed quasi-geostrophic turbulence.

Even though there are an increasing number of idealized numerical studies, there remains a dearth of large-scale coarse resolution studies to investigate the utility of flow dependent diffusivities. In particular, there are no published tests of coarse global models using potential vorticity based closures. Instead, only closures based on the Gent and McWilliams (1990) ideas have been tested with non-constant diffusivities. A notable account of non-constant diffusivities run in a global model is that of Wright (1997), in which he compared a suite of constant coefficient GM90 experiments to a single run with the Visbeck et al. (1997) diffusivity. In these tests the flow dependence was given only to the GM diffusivity, with the neutral diffusion coefficient held constant. Tests were run with the Hadley Centre's 1.25° ocean model, and the results were generally favorable towards the Visbeck et al. (1997) diffusivity, hence prompting the Hadley Centre to employ this scheme in their climate simulations (Gordon et al. (2000)). Unpublished studies in coarser simulations (e.g., coarser than 2°) suggest that non-constant diffusivities may not be so advantageous at this and coarser resolutions (A. Adcroft, personal communication 2000; M. Roberts, personal communication 1998; report of the WOCE/CLIVAR Workshop on Ocean Modelling for Climate Studies, WOCE Report No. 165/99). The idea is that there needs to be a nontrivial structure to the resolved model flow for inhomogeneities in the diagnosed diffusivity to be of utility.

Additional research is ongoing to test the utility of vertically dependent diffusivities for large-scale modeling. However, we know of no published study exhibiting a significant improvement associated with the added complexity of a vertically dependent diffusivity. Notably, allowing for vertical dependence necessitates some added structure to the neutral tracer operators compared to the depth-independent form. At present most publicly supported ocean climate models only provide the framework for employing flow dependent diffusivities that are independent of depth. Hence, in the remainder of this section we limit our discussion to depth independent diffusivities.

12.4.2 Length and time scales for Held and Larichev

Both Held and Larichev (1996) and Visbeck et al. (1997) develop diagnostic expressions for a time scale T and length scale L according to properties of the model's resolved flow field. Thereafter, a diffusivity is computed via

$$\kappa = L^2/T. \tag{12.73}$$

The squared inverse time scale (squared growth rate) suggested by Held and Larichev (1996) is given by

$$T^{-2} = \frac{f^2}{D} \int_{-D_b}^{-D_t} Ri^{-1}\, dz. \tag{12.74}$$

In this expression, f is the Coriolis parameter and Ri is the Richardson number determined by large-scale flow features. This growth rate corresponds to that of a growing Eady wave and the time scale T is on the order of tens to hundreds of days. The integration depth range, $D = D_b - D_t$ was suggested by Treguier et al. (1997) to be $D_t = 100$m and $D_b = 2000$m. This depth range is also taken in the Hadley Centre ocean model in which they implement the Visbeck et al. (1997) scheme. This

range is intended to discount the unstratified mixed layer where the growth rate becomes unbounded. However, it is a bit arbitrary. Sensitivity of the results to this range has not been documented.

The length scale suggested by Held and Larichev (1996) is related to the effective Rhines' scale

$$L^{-1} = \beta_{eff}\, T. \tag{12.75}$$

The effective beta-parameter β_{eff} incorporates both planetary vorticity gradients and gradients in the large-scale topography

$$\beta_{eff} = H|\nabla_z(f/H)|. \tag{12.76}$$

The resulting diffusivity is given by

$$\kappa = L^2\, T^{-1} = (\beta_{eff}^2\, T^3)^{-1}. \tag{12.77}$$

12.4.3 Length and time scales for Visbeck et al. (1997)

The time scale suggested by Visbeck et al. (1997) is given by

$$T^{-1} = \frac{f}{D} \int_{-D_b}^{-D_t} Ri^{-1/2}\, dz. \tag{12.78}$$

The square of this expression is not identical to the squared inverse time scale (12.74) from the Held and Larichev (1996) scheme. However, in both approaches the time scale is determined by scaling arguments. Hence, differences are not fundamentally significant and so can be absorbed into an overall tuning parameter.

The length scale L in the Visbeck et al. (1997) scheme is determined by the regional structure of the baroclinicity–the "width of the baroclinic zone." Details of how to compute this length scale are somewhat ambiguous for a three-dimensional flow. No clear algorithm was described in the Visbeck et al. (1997) paper as they considered only two-dimensional flow. The Hadley Centre model of Gordon et al. (2000) employs a search algorithm described in Pacanowski and Griffies (1999) (Malcolm Roberts, personal communication 1997).

12.4.4 Length scale given by the first baroclinic Rossby radius

Instead of the Rhines' scale or width of the baroclinic zone, we prefer a length scale set according to the first baroclinic Rossby radius. The relevance of this length scale for eddy parameterization was suggested by theoretical studies of Stone (1972), satellite based data analysis by Stammer (1997b), and eddying ocean model studies by Bryan et al. (1999).

Outside an equatorial band of roughly $\pm 5°$, the mth-mode Rossby radius takes the form

$$\lambda_m = \frac{c_m}{|f|}, \tag{12.79}$$

where c_m is the phase speed of mth-mode ($m \geq 1$) gravity waves in a non-rotating, continuously stratified flat bottom ocean (Gill (1982)). As shown in a thorough study by Chelton et al. (1998), the Rossby radius is largely dominated by the inverse

Coriolis dependence (see their Figure 6). Within an equatorial band of $\pm 5^\circ$, the Rossby radius is given by

$$\lambda_m = \left(\frac{c_m}{2\,\beta} \right)^{1/2}, \tag{12.80}$$

where β is the planetary beta-parameter. A useful approximate gravity wave phase speed can be obtained by the WKB form described by Chelton et al. (1998)

$$c_m \approx \frac{1}{m\,\pi} \int_{-H}^{0} N\,dz, \tag{12.81}$$

where N is the buoyancy frequency. Notably, the vertical integral is over the full depth range. Unlike the Eady growth rate calculation, the vertical integration can be easily extended to unstratified regions since the integrand simply vanishes there.

12.4.5 The thermal wind Richardson number

Fundamental to the mesoscale eddy closure theories is the assumption that the mesoscale eddy field is quasi-geostrophic. Consequently, the Richardson number, Ri, determining the time scale is the large-scale Richardson number based on vertical shears under thermal-wind balance with the buoyancy field. This assumption renders

$$\begin{aligned} Ri &= \frac{N^2}{U_{,z}^2 + V_{,z}^2} \\ &= \frac{-g\,\rho_{,z}/\rho_o}{(g/f\rho_o)^2(\rho_{,x}^2 + \rho_{,y}^2)} \\ &= -\left(\frac{f^2\,\rho_o}{g\,\rho_{,z}} \right) S^{-2} \\ &= (f/NS)^2 \end{aligned} \tag{12.82}$$

where

$$S^2 = (\rho_{,x}^2 + \rho_{,y}^2)/\rho_{,z}^2 \tag{12.83}$$

is the squared neutral direction slope vector and

$$N^2 = -g\,\rho_{,z}/\rho_o = -(g/\rho_o)\,(\rho_{,\theta}\,\theta_{,z} + \rho_{,s}\,s_{,z}) \tag{12.84}$$

is the squared buoyancy frequency based on the vertical gradient of locally referenced potential density.

12.4.6 Rotational independence of the time scale

The thermal wind Richardson number (12.82) brings the squared inverse time scale written according to the Held and Larichev form (12.74) to

$$T^{-2} = \frac{1}{D} \int_{-D_b}^{-D_t} N^2\,S^2\,dz = \overline{(N\,S)^2}, \tag{12.85}$$

where an over-bar represents a vertically averaged quantity with the average taken over the chosen depth range. Likewise, the Visbeck et al. (1997) form of the inverse time scale (12.78) is

$$T^{-1} = \frac{1}{D} \int_{-D_b}^{-D_t} N\, S\, dz = \overline{N\, S}. \tag{12.86}$$

As mentioned in Section 12.4.3, any difference between $(\overline{N\, S})^2$ and $\overline{(N\, S)^2}$ is not fundamentally significant and so can be ascribed to an overall dimensionless scaling factor.

Note how the explicit dependence on the Coriolis parameter f has canceled from the time scale. The source of this cancellation is the use of thermal wind balance for computing the Richardson number. Consequently, the time scale is an explicit function only of the vertically averaged horizontal and vertical stratification. Relatedly, the explicit cancelation of the Coriolis parameter allows for the time scale calculation to be naively applied globally, including at the equator where geostrophy is irrelevant.

The inverse time scale, or the growth rate, vanishes when the vertically integrated horizontal stratification vanishes; i.e., when there is zero baroclinicity. Hence, the diffusivity vanishes when the neutral directions are flat, as one would expect for a configuration absent available potential energy. Conversely, the growth rate becomes large when the neutral direction slope becomes large, which is expected for regions of large available potential energy.

For testing the validity of the code, it is useful to estimate a maximum expected Eady growth rate. For this purpose, assume the buoyancy frequency to be on the order of f and the slopes to be a constant of $1/100$ throughout the column. This case leads to a maximum growth rate on the order of $T^{-1} = f/100$. Note that in unstratified regions, the growth rate is actually unbounded. This behavior is undesirable numerically and signals a limitation of the quasi-geostrophic theory to extend into highly unstratified regions. Regularization methods proposed in Section 12.4.7 and 12.4.8 aim to numerically eliminate these singularities. Regardless the details, the result of these regularization methods should be a growth rate no larger than the maximum estimate given here.

12.4.7 One means to provide a regularized diffusivity

In regions of zero vertical stratification, the Eady growth rate given in Section 12.4.6 becomes unbounded. It is therefore necessary to provide a regularization method for the numerical model. A method suggested by Treguier et al. (1997), and used in the Hadley Centre and Pacanowski and Griffies (1999) implementations of these schemes, is to restrict the vertical integration to regions where the vertical stratification is nontrivial. This approach, as mentioned in Section 12.4.2, leads to the somewhat *ad hoc* specification of two depths D_b and D_t. This section presents another method arguably just as valid. It aims to maintain consistency with the Rossby radius calculation of Chelton et al. (1998) discussed in Section 12.4.4, where the vertical integration is over the full ocean depth.

12.4.7.1 Regularization

To motivate a regularization method handling unstratified regions, write the first mode gravity wave phase speed in the form

$$\pi\, c_1 = \int_{-H}^{0} N\, dz = H\, \overline{N}. \tag{12.87}$$

Hence, the first baroclinic Rossby radii for the equatorial and off-equatorial regions take the form

$$L_{<|5^\circ|} = \left(\frac{H\, \overline{N}}{2\, \pi\, \beta}\right)^{1/2} \tag{12.88}$$

$$L_{>|5^\circ|} = \frac{H\, \overline{N}}{\pi |f|}, \tag{12.89}$$

and the corresponding diffusivities $\kappa = L^2\, T^{-1}$ are

$$\kappa_{<|5^\circ|} = \left(\frac{H}{2\, \pi\, \beta}\right) (\overline{N})\, \overline{N\, S} \tag{12.90}$$

$$\kappa_{>|5^\circ|} = \left(\frac{H}{\pi\, f}\right)^2 (\overline{N})^2\, \overline{N\, S}, \tag{12.91}$$

where the inverse time scale (12.86) of Visbeck et al. (1997) was used. The numerical problem is that $\overline{N\, S}$ becomes infinite if any piece of a vertical column is unstratified. One means to resolve this problem is to write

$$\overline{N\, S} = \overline{\left(\frac{N^2\, S}{N}\right)} = \gamma \left(\frac{\overline{N^2\, S}}{\overline{N}}\right). \tag{12.92}$$

If we assume the dimensionless parameter γ is a constant, we are led to the regularized diffusivities

$$\kappa_{<|5^\circ|} = \gamma \left(\frac{H}{2\, \pi\, \beta}\right) \overline{N^2\, S} \tag{12.93}$$

$$\kappa_{>|5^\circ|} = \gamma \left(\frac{H}{\pi\, f}\right)^2 \overline{N}\, \overline{N^2\, S}. \tag{12.94}$$

Notably, the diffusivity within the equatorial band is only a function of the horizontal stratification via $\overline{N^2\, S}$, whereas the diffusivity outside the band is a function of both horizontal and vertical stratification via $\overline{N}\, \overline{N^2\, S}$. In practice, the dimensionless parameter γ will be used as a tuning parameter. Additionally, it is necessary to set a lower and upper limit on the diffusivity for purposes of maintaining numerical stability.

Regularization methods, such as this one and the preferred method described in Section 12.4.8, remedy a numerical problem associated with an incomplete physical theory. In the present case, the regularization method numerically remedies a limitation of quasi-geostrophic theory to extend into vertically unstratified regions. One goal of the numericist designing regularization methods is to maintain numerical stability without egregiously compromising the physical integrity of the algorithm. In general, we counsel strong skepticism be given any regularization scheme since they are often based on numerical considerations moreso than physical considerations.

12.4.7.2 A value for the tuning parameter

A useful means of gauging the size of the tuning parameter is to consider a linear equation of state $\rho = \rho_o\,(1 - \alpha\theta)$ and a depth independent neutral direction slope. In this case, the equatorial-band diffusivity becomes

$$\begin{aligned}\kappa_{<|5^\circ|} &= \gamma \left(\frac{H}{2\,\pi\,\beta}\right) \overline{S\,N^2} \\ &= \gamma \left(\frac{S\,H}{2\,\pi\,\beta}\right) \overline{N^2} \\ &= \gamma \left(\frac{\alpha\,g\,S\,H}{2\,\pi\,\beta}\right) (\theta(z=0) - \theta(z=-H)). \end{aligned} \tag{12.95}$$

With $\beta = 2.28 \times 10^{-11}(\mathrm{m\,s})^{-1}$ at the equator, $g = 9.8\,\mathrm{m\,s^{-2}}$, $\alpha = 2 \times 10^{-4}(^\circ K)^{-1}$, and $\theta(z=0) - \theta(z=-H) = 20^\circ K$, we have

$$\kappa_{<|5^\circ|} \approx S\,\gamma \left(3 \times 10^8 \mathrm{m^2\,s^{-1}}\right). \tag{12.96}$$

If the depth averaged slope is $S = 10^{-3}$, then

$$\kappa_{<|5^\circ|} \approx \gamma \left(3 \times 10^5 \mathrm{m^2\,s^{-1}}\right). \tag{12.97}$$

Based on experience with ocean models run with constant neutral physics diffusivities, where values are on the order $10^2 - 10^3 \mathrm{m^2\,s^{-1}}$, one is motivated to set the tuning parameter to

$$\gamma = 0.01, \tag{12.98}$$

which is consistent with the dimensionless scaling parameter suggested by Visbeck et al. (1997). This choice then brings the diffusivity to

$$\kappa_{<|5^\circ|} \approx 3 \times 10^3 \mathrm{m^2\,s^{-1}}. \tag{12.99}$$

12.4.8 A recommended approach

A cumbersome aspect of the flow dependent diffusivities presented thus far is that they do not explicitly depend on the grid scale. Hence, there is no tendency for their values to decrease as the grid is refined. Indeed, for refined grids, the Eady growth rate increases since the model allows more vigorous mean flows with stronger gradients and slopes. One thus has the undesirable situation of larger flow dependent diffusivities for refined grids.

A diffusivity decreasing upon grid refinement is useful since (1) refined grids allow for more active mean flows that one does not wish to overly diffuse, (2) relatedly, with enough resolution, the model transitions to a regime where the parameterized quasi-geostrophic eddying flow becomes explicitly represented, thus motivating the reduction of the SGS operator's contribution to the time tendency. It is notable that frictional dissipation via the Smagorinsky viscosity scheme (Smagorinsky (1993) and Griffies and Hallberg (2000)) provides an example where the SGS operator naturally becomes less important as resolution is refined. This is the goal of the diffusivity proposed in this section.

12.4.8.1 Length scale

For purposes of quasi-geostrophic eddy parameterizations, coarseness of a model grid is based on how the grid size compares to the first baroclinic Rossby radius of deformation. If the grid is coarse, then an appropriate length scale for computing the diffusivity is given by the Rossby radius as discussed in Section 12.4.4. For refined models where the Rossby radius is larger than the grid, which is now quite common at least for models with refined tropical resolution, then usage of the Rossby radius to determine the diffusivity will produce an overly large value. Hence, it is proposed that the length scale be given by the following smooth function of the Rossby radius and the grid scale

$$L = \frac{2\,\Delta s\,\lambda_1}{\Delta s + \lambda_1}, \tag{12.100}$$

where

$$\Delta s = \frac{2\,\Delta x\,\Delta y}{\Delta x + \Delta y}. \tag{12.101}$$

The ratios provide for a smoother length scale than the alternative usage of minimum functions. In particular, (a) when $\Delta s << \lambda_1$, then $L \approx \Delta s$, (b) when $\Delta s = \lambda_1$, then $L = \lambda_1$, and (c) when $\Delta s >> \lambda_1$, then $L \approx \lambda_1$.

12.4.8.2 Velocity scale

To remedy the numerical problem with an increasing Eady growth rate for refined grids, we propose to use the following velocity scale

$$V = \frac{2\,V_m\,V_{qg}}{V_m + V_{qg}}. \tag{12.102}$$

The velocity scale V_m is a constant motivated from numerical considerations* and is a tuning parameter, with common values on the order

$$V_m = 0.05\,\mathrm{m\,s^{-1}}. \tag{12.103}$$

The velocity scale

$$V_{qg} = L\,T^{-1} \tag{12.104}$$

is motivated from the quasi-geostrophic theories using the Eady growth rate (12.85) to compute T^{-1}, and the length scale is taken from equation (12.100). Instead of limiting the integration to be over an *a priori* specified depth range, as suggested by Treguier et al. (1997), we prefer to integrate over the full depth range. Yet contributions from unstratified regions should be limited by the maximum slope allowed for numerical stability, which is typically around 0.01.

12.5 Biharmonic operators

Biharmonic tracer operators dissipate tracer variance. As with the biharmonic friction operators discussed in Section 15.9, they concentrate dissipation at the smaller

*Such a velocity scale was suggested by Eric Chassignet in personal communication 1999.

scales than Laplacian operators, thus allowing the larger scales to be more advectively dominant. It is for this reason that biharmonic operators are typically preferred for eddying simulations, where it is desirable to allow the flow to exhibit its natural tendency towards hydrodynamic instability.

In practice, the general ideas mentioned above have not proven to be relevant for tracers. That is, for the purpose of enhancing the advectively dominant aspects of the flow, it has been found unnecessary to employ biharmonic tracer operators, so long as biharmonic friction is used. The main point is that dissipation of flow features is dominated by friction, thus warranting biharmonic friction when aiming to allow for a broader range of advectively dominant flow. The use of either a biharmonic or Laplacian tracer operator is not so crucial, so long as the tracer diffusivity is appropriately tuned.

The above conclusion came to the author as a surprise, and only after having considered the pros and cons of various biharmonic tracer operators described in this section. Although just the traditional horizontal biharmonic tracer operator is supported in MOM4, we present the following analysis of some rather interesting, albeit not so useful, biharmonic tracer operators. Points relevant for implementating these operators are also made for those wishing to numerically test the operators. Note that the discussion in this section is limited to Boussinesq fluids.

An additional fundamental reason for avoiding the use of biharmonic tracer operators is that they produce sign-indefinite tracer fluxes. When running models with more than one active tracer, such fluxes generally produce a sign-indefinite cabbeling term, which is unphysical as we saw in Section 12.1.7. This result points to a fundamental problem with many of the biharmonic operators discussed in this section, in that they are not readily generalized in a satisfying manner to the realistic case of more than one active tracer.

Before starting the analysis, we mention the common practice in the isopycnal modeling community. For isopycnal models, biharmonic friction is commonly used in eddying simulations (e.g., Griffies and Hallberg (2000)). The transport of tracers, however, is typically implemented via a positive-definite dissipative advection scheme, with zero added explicit diffusion. Many models use dissipative advection schemes that have truncation errors appearing as Laplacian isopycnal diffusion. In this case, the isopycnal simulations are a direct analog of what we contend is relevant for z-coordinate eddying simulations.

12.5.1 Tracer variance

As in Section 11.4.1, we find occasion both in this section and Section 14.1 to consider the effects on tracer variance of certain operators. For our purposes with Boussinesq fluids, tracer variance is given by

$$\mathcal{V} = V^{-1} \int dV \, C^2 - V^{-1} \left(\int dV \, C \right)^2 , \tag{12.105}$$

where $V = \int dV$ is the volume of seawater in the domain. In a conservative ocean system without sources and sinks, the total tracer mass $\int (\rho_o \, dV) \, C$ and total volume are constant with time. However, the total squared tracer can change due to transport effects. Therefore, changes in tracer variance are equivalent to changes in

integrated squared tracer $\int dV\ C^2$, and so for our purposes the two are considered the same.

12.5.2 Horizontal biharmonic mixing

Let us start with the traditional horizontal biharmonic operator. This operator is an iteration on the horizontal Laplacian operator, and the horizontal biharmonic tracer flux is given by

$$\mathbf{F}_B = \sqrt{B}\,\nabla_z L \tag{12.106}$$

where

$$L = \nabla_z \cdot (\sqrt{B}\,\nabla_z C) \tag{12.107}$$

is the horizontal Laplacian operator acting on the tracer, and $B > 0$ is a biharmonic mixing coefficient with units $(\text{length})^4\,(\text{time})^{-1}$. The convergence of the biharmonic flux $\mathbf{F}_B$ yields the horizontal biharmonic mixing operator.

$$R_B = -\nabla \cdot \mathbf{F}_B. \tag{12.108}$$

We illustrate the biharmonic operator's variance dissipation property via the following manipulations

$$\begin{aligned}
\frac{1}{2}\,\partial_t \int dV\, C^2 &= -\int dV\, C\,\nabla_z \cdot \mathbf{F}_B \\
&= -\int dV\,\nabla_z \cdot (C\,\mathbf{F}_B) + \int dV\,\nabla_z C \cdot \mathbf{F}_B \\
&= \int dV\,\sqrt{B}\,\nabla_z C \cdot \nabla_z L \\
&= \int dV\,\nabla_z \cdot (\sqrt{B}\,L\,\nabla_z C) - \int dV\, L^2 \\
&= -\int dV\, L^2 \le 0.
\end{aligned} \tag{12.109}$$

To reach this result, we assumed that the horizontal Laplacian and biharmonic tracer fluxes satisfy no-flux side boundary conditions

$$\hat{\mathbf{n}} \cdot \sqrt{B}\,\nabla_z C = 0 \tag{12.110}$$

$$\hat{\mathbf{n}} \cdot \mathbf{F}_B = 0. \tag{12.111}$$

We also ignored free surface boundary effects. Even if the tracer gradients do not vanish at the boundaries, the flux can still vanish if the diffusivity is assumed to vanish there. Either way, this boundary condition is easily ensured via masks placed on the numerical model's flux components.

Notably, without the square root on the biharmonic mixing coefficient, tracer variance is not generally dissipated when using a spatially dependent diffusivity since it is no longer possible to form a perfect square. This point was discussed by Griffies and Hallberg (2000) in the context of biharmonic friction.

The orientation of the biharmonic flux relative to the tracer field

$$\mathbf{F}_B \cdot \nabla_z C = \sqrt{B}\,\nabla_z L \cdot \nabla_z C, \tag{12.112}$$

is not sign definite. Hence, the biharmonic flux is not generally down the tracer gradient. It therefore cannot be considered a downgradient diffusive flux, whereas the Laplacian analog is strictly downgradient. As discussed in Section 12.1.7, such sign-indefinite mixing flux leads to a sign-indefinite cabbeling term, which is unphysical. This is a fundamental reason for avoiding the use of any biharmonic tracer operator in ocean simulations.

In MOM4, horizontal biharmonic mixing is trivial to implement. The only detail concerns the boundary conditions. As mentioned above, variance reduction is ensured by setting no-flux conditions on the biharmonic flux just as for the Laplacian flux.

12.5.3 Neutral biharmonic mixing

We now extend the Laplacian neutral diffusion of Solomon (1971) and Redi (1982) into the biharmonic realm, thus leading to a neutral biharmonic mixing operator. This operator represents a direct analog of the horizontal biharmonic mixing operator. It also suffers from the same sign-indefinite nature of the local tracer flux, thus leading to unphysical cabbeling processes.

As with horizontal biharmonic mixing, we employ an iterative approach to construct the neutral biharmonic mixing operator. Hence, the m'th component of the neutral biharmonic flux is given by

$$F_B^m = K^{mn}\, L_{,n} \tag{12.113}$$

where

$$L = (K^{pq}\, C_{,q})_{,p} \tag{12.114}$$

is the Laplacian neutral diffusion operator. As a vector, the biharmonic flux is written

$$\mathbf{F}_B = \sqrt{B}\,(\nabla_\rho + \hat{\mathbf{z}}\,\mathbf{S}\cdot\nabla_\rho)\,L, \tag{12.115}$$

and the biharmonic operator is computed as the convergence of the neutral biharmonic flux

$$R_B = -\nabla\cdot\mathbf{F}_B. \tag{12.116}$$

As for horizontal biharmonic mixing in Section 12.5.2, we apply the diffusivity symmetrically at each stage of the calculation with a square-root of the full biharmonic diffusivity. This approach ensures variance reduction with non-constant diffusivities, as seen via the following manipulations

$$\begin{aligned}
\frac{1}{2}\,\partial_t \int dV\, C^2 &= -\int dV\, C\,\nabla_z\cdot\mathbf{F}_B \\
&= -\int dV\,\nabla_z\cdot(C\,\mathbf{F}_B) + \int dV\,\nabla_z C\cdot\mathbf{F}_B \\
&= \int dV\, K^{mn}\, C_{,m}\, L_{,n} \\
&= \int dV\,(L\,K^{mn}\, C_{,m})_{,n} - \int dV\, L^2 \\
&= -\int dV\, L^2 \le 0.
\end{aligned} \tag{12.117}$$

To reach this result, we assumed both the Laplacian and biharmonic tracer fluxes vanish at the side boundaries

$$\hat{\mathbf{n}}_m \, K^{mn} \, C_{,n} = 0 \tag{12.118}$$

$$\hat{\mathbf{n}}_m \, K^{mn} \, L_{,n} = 0. \tag{12.119}$$

As for horizontal biharmonic mixing, these no-flux boundary conditions are easily established in the numerical model via masks. Furthermore, as for the Laplacian neutral diffusion case, the neutral biharmonic flux vanishes for tracers aligned with the neutral directions.

The numerical algorithm proceeds as for the horizontal biharmonic operator. First, the full neutral Laplacian diffusion operator L is computed. For the second iteration, L is treated as the new "tracer" and the neutral biharmonic operator then constructed. Notably, the diagonal term in the vertical biharmonic flux contributes to the time evolution in an implicit manner, just as used when performing Laplacian neutral diffusion (see Section 14.8.3 for details).

12.5.4 Roberts and Marshall's biharmonic stirring operator

Roberts and Marshall (1998) (RM98) considered the following divergence-free velocity

$$\mathbf{u}^* = \nabla \wedge \boldsymbol{\Upsilon} = \partial_z \, \nabla_z^2 (B \, \mathbf{S}) - \hat{\mathbf{z}} \, \nabla_z \cdot \nabla_z^2 (B \, \mathbf{S}), \tag{12.120}$$

where the vector streamfunction is

$$\boldsymbol{\Upsilon} = -\hat{\mathbf{z}} \wedge \nabla_z^2 (B \, \mathbf{S}) \tag{12.121}$$

with $B > 0$ a biharmonic mixing coefficient. The corresponding skew flux is given by

$$\mathbf{F} = -C_{,z} \, \nabla_z^2 \, (B \, \mathbf{S}) + \hat{\mathbf{z}} \, \nabla_z C \cdot \nabla_z^2 \, (B \, \mathbf{S}). \tag{12.122}$$

Dropping the horizontal Laplacian ∇_z^2, and setting $B \to -\kappa$, recovers the Laplacian GM90 scheme of Section 12.2.

Enhanced scale selectivity with the operator is afforded to the buoyancy field. Indeed, RM98 showed that for buoyancy, this operator is quite effective at dissipating grid scale variance while reducing spurious diapycnal mixing. For other tracers, the operator remains second order and so is not as scale selective as either the horizontal or neutral biharmonic mixing operators.

RM98 termed their operator a "biharmonic GM" operator since it represents a straightforward generalization of the usual "Laplacian GM" operator. Yet, as shown by RM98 and in the following, their biharmonic operator does not generally dissipate potential energy, whereas the Laplacian GM operator does. Therefore, the RM98 operator is perhaps better considered another one of many possible neutral biharmonic operators.

12.5.4.1 Effects on potential energy of the RM98 operator

Consider how the RM98 operator affects the potential energy for the case when density is a linear function of potential temperature. A similar discussion was given in the RM98 paper, where they employed the advective flux formulation rather

than the skew-flux formulation. Focusing just on the biharmonic operator, the time tendency for potential energy is given by

$$\begin{aligned} P_{,t} &= g \int dV \; z \, \rho_{,t} \\ &= -g \int dV \; z \, \nabla \cdot \mathbf{F} \\ &= g \int dV \; F^{(z)}, \end{aligned} \tag{12.123}$$

where the no-normal flux condition at the sides was assumed. Note that for simplicity, a rigid lid was also assumed, although this assumption has no bearing on the effects the $g \int dV \; F^{(z)}$ term has on the evolution of the total potential energy. Introducing the vertical component to the RM98 skew-flux leads to

$$\begin{aligned} \int dV \; F^{(z)} &= - \int dV \; B \, \rho_{,z} \, \mathbf{S} \cdot \nabla_z^2 \mathbf{S} \\ &= - \int dV \; \nabla_z \cdot (B \, \rho_{,z} \, S_m \, \nabla_z S_m) + \int dV \; \nabla_z (B \, \rho_{,z} \, S_m) \cdot \nabla_z S_m, \end{aligned} \tag{12.124}$$

where $m = 1, 2$ is summed. Assuming $B \, \rho_{,z} \, S_i$ vanishes next to the lateral ocean boundaries allows for the total derivative term to be dropped. This term may be set to zero via model masks, and it may formally be associated with a vanishing diffusivity at this boundary. The result is

$$\int dV \; F^{(z)} = (1/2) \int dV \; \nabla_z \, (B \, \rho_{,z}) \cdot \nabla_z S^2 + \int dV \; B \, \rho_{,z} \, (\partial_m \mathbf{S} \cdot \partial_m \mathbf{S}), \tag{12.125}$$

where $S^2 = \mathbf{S} \cdot \mathbf{S}$, and $m = 1, 2$ is summed in the second term over the horizontal spatial dimensions. The second term is non-positive in stably stratified fluids, for which $\rho_{,z} \leq 0$. It therefore represents a potential energy sink. The first term, however, is sign indefinite even when B is a constant. For the special case of $B \, \rho_{,z}$ independent of horizontal position, the first term vanishes and so potential energy is reduced. For example, the special density profile $\rho = \rho_o(z) + \rho_1 \, \cos(\mathbf{p} \cdot \mathbf{x})$, with $\mathbf{p}$ a horizontal wave vector, has $B \, \rho_{,z}$ horizontally constant if B is constant, and so the potential energy is reduced. In the slightly more general case of constant B and with $\rho_{,z} = \partial_z \rho_o + \partial_z \rho_1$, where $|\partial_z \rho_o| >> |\partial_z \rho_1|$, the first term in the expression for potential energy is nonzero, but subdominant to the second term. So potential energy is again reduced in this case. It is unclear what happens for more general profiles.

It might be speculated that the inability to prove that the potential energy is generally reduced may indicate that the RM98 operator is unstable. However, if numerically implemented according to the triad approach of Griffies et al. (1998), the discretized RM98 skew-flux preserves density variance. As variance growth is typically associated with linearly unstable numerical schemes, any potential instability of the RM98 scheme will likely be nonlinear. So far, no such instabilities have been encountered. Rather, the operator appears to be quite stable in a wide suite of both coarse and fine models (M. Roberts, personal communication 1998). Hence, it is unclear how important it is to provide a potential energy sink with the stirring operators.

12.5.4.2 Flux for a particular density profile

Recall from Section 12.2.1 that in the special case of a linear equation of state, the GM90 density skew-flux takes the especially simple form

$$\mathbf{F}^{(h)}(\rho) = -\kappa\, \nabla_z \rho \tag{12.126}$$

$$F^{(z)}(\rho) = \kappa\, S^2 \rho_{,z}. \tag{12.127}$$

With a stable density profile, $\rho_{,z} < 0$, thus leading to a negative vertical skew-flux component. In general, the horizontal GM90 skew-flux components are directed down the density gradient, and the vertical component is upgradient. With an always upgradient flux of density, the GM90 scheme always decreases the potential energy in a stably stratified fluid.

As just seen, this potential energy dissipation property is not generally respected by the RM98 operator. However, it is useful to highlight a case where these properties are shared, such as the case where density is given by

$$\rho = \rho_o(z) + \rho_1\, \cos(\mathbf{p}\cdot\mathbf{x}), \tag{12.128}$$

with $\rho_o(z)$ a stable vertical profile, ρ_1 an amplitude function, and $\mathbf{p} = (p_x, p_y, 0)$ a horizontal wave-vector. The slope vector for this profile is given by

$$\mathbf{S} = \frac{\rho_1\, \mathbf{p}\, \sin(\mathbf{p}\cdot\mathbf{x})}{\rho_o'(z)}, \tag{12.129}$$

and the horizontal Laplacian is

$$\nabla_z^2 \mathbf{S} = -p^2\, \mathbf{S}, \tag{12.130}$$

where $p^2 = \mathbf{p}\cdot\mathbf{p}$. The RM98 skew-flux therefore takes the form

$$\mathbf{F}^{(h)} = (B\, p^2)\, \mathbf{S}\, C_{,z} \tag{12.131}$$

$$F^{(z)} = -(B\, p^2)\, \mathbf{S}\cdot\nabla_z C. \tag{12.132}$$

The RM98 skew-flux of density, linearly dependent on temperature, is given by

$$\mathbf{F}^{(h)}(\rho) = -(B\, p^2)\, \nabla_z \rho \tag{12.133}$$

$$F^{(z)}(\rho) = (B\, p^2)\, S^2\, \rho_{,z}. \tag{12.134}$$

Hence, for this profile, the horizontal RM98 density skew-flux components are down the horizontal density gradient, and the vertical skew-flux component is up the vertical density gradient. Potential energy is therefore dissipated. The effective diffusivity is scale-dependent, with small scales, or large p^2, acted on with the largest effective diffusivity.

12.5.5 Gent's biharmonic stirring operator

As reported in Roberts and Marshall (1998), P. Gent suggested the alternative velocity

$$\mathbf{u}^* = \nabla \wedge \boldsymbol{\Upsilon} = -\partial_z\, \nabla_z M + \hat{\mathbf{z}}\, \nabla_z^2\, M, \tag{12.135}$$

with vector streamfunction

$$\boldsymbol{\Upsilon} = \hat{\mathbf{z}} \wedge \nabla_z M \tag{12.136}$$

and

$$M = B \left(\frac{\nabla_z^2 \rho}{\rho_{,z}} \right). \tag{12.137}$$

The corresponding skew flux is given by

$$\mathbf{F} = C_{,z} \, \nabla_z \, M - \hat{\mathbf{z}} \, \nabla_z C \cdot \nabla_z \, M. \tag{12.138}$$

In contrast to the RM98 operator, the Gent operator provides a sign definite sink of potential energy for any given density profile, as seen since the integrated vertical skew-flux component is negative semi-definite in stably stratified water

$$\begin{aligned} \int dV \; F^{(z)} &= - \int dV \; \nabla_z \left(B \, \frac{\nabla_z^2 \rho}{\rho_{,z}} \right) \cdot \nabla_z \rho \\ &= \int dV \; \nabla_z \cdot (B \, \mathbf{S} \, \nabla_z^2 \rho) + \int dV \; (B/\rho_{,z}) \; \left(\nabla_z^2 \rho \right)^2 . \end{aligned} \tag{12.139}$$

The first term can be dropped when assuming $B \, \mathbf{S} \, \nabla_z^2 \rho$ vanishes at horizontal boundaries. The second term is non-positive in stably stratified water, and so represents a potential energy sink for any density profile. Notably, the spectral representation of this operator when acting on the density profile $\rho = \rho_0(z) + \rho_1 \, \cos(\mathbf{k} \cdot \mathbf{x})$ is the same as the RM98 operator given by equations (12.131) and (12.132).

The Gent biharmonic operator cannot be discretized using the proven technology developed by Griffies et al. (1998), Griffies (1998) (see chapters 13 and 14 for details). The reason is that the slope vector $\mathbf{S}$ is not a fundamental piece of the Gent operator, whereas it is fundamental for all other neutral physics schemes.

12.5.6 Neutral biharmonic filtering

The neutral biharmonic filtering operator is designed with the following points in mind:

1. Neutral biharmonic mixing does not act on buoyancy, except through effects due to the nonlinear equation of state. Hence, an alternative means for scale-selective buoyancy dissipation should be considered.

2. Horizontal biharmonic mixing provides a cheap, scale-selective dissipation mechanism.

3. However, the horizontal orientation of the horizontal biharmonic flux introduces too much diapycnal mixing, and so it is necessary to re-orient the flux according to the neutral directions.

These points motivate us to consider the following tracer flux

$$\mathbf{F} = \sqrt{B} \, (\nabla_z + \hat{\mathbf{z}} \, \mathbf{S} \cdot \nabla_z) \, L, \tag{12.140}$$

with

$$L = \nabla_z \cdot (\sqrt{B} \, \nabla_z C) \tag{12.141}$$

the horizontal Laplacian diffusion operator. The convergence of the horizontal flux components $\mathbf{F}^{(h)} = \sqrt{B}\,\nabla_z L$ yields the familiar horizontal biharmonic operator, yet the vertical flux component $F^{(z)} = \mathbf{S}\cdot\mathbf{F}^{(h)}$ is chosen to orient the three-dimensional flux vector parallel to the neutral direction (see Figure 11.1). That is,

$$\mathbf{F}\cdot\nabla\rho = 0, \tag{12.142}$$

which is analogous to the neutral diffusive fluxes. However, in contrast to the neutral diffusive fluxes, $\mathbf{F}(\rho) \neq 0$. Instead, $\mathbf{F}(\rho)$ is a skew flux and so has a corresponding advection velocity (see equation (12.146)).

When acting on the tracer profile

$$C = C_o(z) + C_1(z)\,\cos(\mathbf{p}\cdot\mathbf{x}) \tag{12.143}$$

with constant biharmonic coefficient B, the flux (12.140) takes the form

$$\mathbf{F} = -B\,p^2\,(\nabla_z + \hat{\mathbf{z}}\,\mathbf{S}\cdot\nabla_z)C. \tag{12.144}$$

The horizontal flux components are downgradient and scale-selective, yet the orientation of the vertical flux depends on the relative orientation of the tracer and the neutral slopes. When $C = \rho$,

$$\mathbf{F} = -B\,p^2\,(\nabla_z\rho - \hat{\mathbf{z}}\,S^2\rho_{,z}). \tag{12.145}$$

Since the vertical skew-flux component is upgradient, this profile has its potential energy dissipated. Note the slight generalization of this result compared to that of the GM98 operator, where potential energy is dissipated only when $\rho_1(z) \to \rho_1$. However, as with the RM98 operator, an arbitrary density profile is not guaranteed to have its potential energy dissipated.

It is useful to highlight the different perspective taken here as compared to RM98 and Gent considered in Sections 12.5.4 and 12.5.5. Both RM98 and Gent focused on the advective form of GM90, which led them to their higher order eddy-induced advection velocity. That is, they both proposed a stirring process. In contrast, the neutral biharmonic filter only stirs density (assuming a linear equation of state), in which case one can associate the following advection velocity

$$\mathbf{u}^* = \nabla\wedge\boldsymbol{\Upsilon} = -\partial_z\left(\sqrt{B}\,\frac{\nabla_z L}{\rho_{,z}}\right) + \hat{\mathbf{z}}\,\nabla_z\cdot\left(\sqrt{B}\,\frac{\nabla_z L}{\rho_{,z}}\right), \tag{12.146}$$

where the vector streamfunction is

$$\boldsymbol{\Upsilon} = \hat{\mathbf{z}}\wedge\left(\sqrt{B}\,\frac{\nabla_z L}{\rho_{,z}}\right). \tag{12.147}$$

For other tracers, however, the new operator produces mixing fluxes oriented along the neutral direction. Hence, there is no corresponding advection velocity.

Numerical discretization of the horizontal flux components follows from the usual methods for horizontal biharmonic mixing. Notably, the horizontal fluxes are applied for all values of the neutral direction slope. The vertical flux follows the triad approach described in Chapter 14, with tapering applied in steep sloped regions as described in Section 13.1.5.

12.5.7 Biharmonic skew filtering operator

Another biharmonic operator, closely related to the neutral biharmonic filtering operator given in Section 12.5.6, is termed here a *biharmonic skew filtering operator*. The tracer flux for this operator is given by

$$\mathbf{F} = \sqrt{B}\,(\nabla_z + \hat{\mathbf{z}}\,\mathbf{S}(C)\cdot\nabla_z)\,L. \tag{12.148}$$

Importantly, the slope here is that of the particular tracer

$$\mathbf{S}(C) = -\left(\frac{\nabla_z C}{C_{,z}}\right). \tag{12.149}$$

In contrast, the neutral filter flux (12.140) is constructed with the neutral slope $\mathbf{S}(C)$ for all tracers. Both fluxes are the same in the special case of a linear equation of state with $C = \rho$. However, in general the fluxes are distinct.

When density is a function of both temperature and salinity, separately stirring temperature and salinity with the biharmonic skew filtering operator generally mixes density. Hence, biharmonic skew filtering is not useful for active tracers. Yet it may be a useful means to suppress noise in passive tracer fields. However, the approach may become quite computationally burdensome when adding many passive tracers since the cost of computing successively more tracer slopes is nontrivial. Instead, one may be more motivated to consider neutral biharmonic mixing discussed in Section 12.5.3 for passive tracers.

12.5.8 A note on iterative Laplacian skewsion processes

As noted in Section 12.5.2, the horizontal biharmonic flux is not generally oriented down the horizontal tracer gradient; likewise for the neutral biharmonic mixing operator. It is for this reason that these biharmonic operators are termed "mixing" operators rather than "diffusive" operators. Even so, they satisfy the usual global constraint of reducing tracer variance, hence their utility as dissipation operators. Furthermore, the neutral biharmonic mixing operator remains true to the needs of fluxing tracers along the neutral directions. That is, the operator vanishes when a tracer is aligned with the neutral direction.

The three biharmonic operators acting on buoyancy each provide for an adiabatic stirring of buoyancy. However, neither completely represents a straightforward iteration of a Laplacian stirring operator. There is good reason for such: iteration of a Laplacian skewsion process does not generally lead to a biharmonic skewsion process. Hence, there is ambiguity concerning the choice of an appropriate skewed biharmonic operator.

CHAPTER

THIRTEEN

Neutral physics near boundaries

Contents

13.1 Linear numerical stability for neutral diffusion 284

13.1.1 Linear stability analysis for Laplacian neutral diffusion . . 284

13.1.2 Linear stability analysis for biharmonic neutral mixing . . 285

13.1.3 Slope clipping the neutral diffusive fluxes 285

13.1.4 Quadratic tapering the neutral diffusive fluxes 286

13.1.5 Exponential tapering the neutral diffusive fluxes 287

13.2 Linear stability for GM stirring 287

13.3 Neutral physics near boundaries 288

13.3.1 General considerations 289

13.3.2 Surface boundary condition for fluxes from neutral physics 290

13.3.3 A kinematic argument for tapering near a boundary 290

13.3.4 Summary and recommendations 291

13.3.4.1 Review of the mathematical form for the fluxes . 292

13.3.4.2 Suggestions for neutral diffusion 293

13.3.4.3 Suggestions for GM 295

An implementation of neutral physics in an ocean climate model must address the question of what to do when neutral slopes steepen, as they will in realistic simulations. Numerical stability warrants some form of tapering towards a non-neutral oriented scheme. Are there physical reasons to do so as well? We address this, and other related, questions in this chapter. That is, we consider issues related to how neutral physical processes are handled in regions near boundaries and/or where neutral slopes are steep, such as the surface boundary layer. Simulations are sensitive to details in these regions, and much of this discussion is at the level of active research. The reader should therefore consider this material with a healthy bit of skepticism.

13.1 Linear numerical stability for neutral diffusion

There are various methods to preserve linear numerical stability of neutral diffusion in regions of steep neutral slopes. One method is to employ the full diffusion tensor (Section 12.1). This tensor has finite values regardless the value of the neutral slopes. Hence, its discretization generally requires no artificial numerical regularization, so long as a reasonable time step condition is maintained (see Appendix C of Griffies et al. (1998) for discussion). However, the full Redi tensor contains terms that are very small throughout the bulk of the ocean where slopes are modest, and these terms are awkward and expensive to compute. Additionally, as argued in later sections in this Chapter, there is little physical motivation to maintain the integrity of neutrally oriented diffusion in regions where the neutral slopes are steep. That is, we are not motivated physically to use the full Redi tensor since the small angle approximation (Section 12.1) is sufficient for all ocean climate related purposes using z-models.

Hence, we are motivated to use the small slope neutral diffusion tensor. This tensor, however, is not bounded as the neutral slopes increase, and so a regularization method is needed for simulations where the slopes can be vertical. Relatedly, when the slope of the neutral direction is large, it is possible for the neutral diffusive flux to project substantially onto the vertical direction. Because the vertical grid spacing is generally much smaller than the horizontal, care is needed to preserve numerical stability in regions of steep neutral slopes. Consequently, one must prescribe a sensible regularization method for ocean climate models, where the neutral slope generally can become vertical.

One approach for maintaining numerical stability is to solve the neutral diffusion operator implicitly, as already done for most models for their vertical physics schemes. However, since neutral diffusion has lateral flux components, an implicit solution method involves the whole horizontal grid. Additionally, the possibility of inifinite slopes, and correspondingly infinite fluxes, makes even an implicit scheme numerically sensitive. Such an approach is therefore not practical in a realistic climate model and so is not pursued. The remainder of this section presents other options. The discussion represents in part an abbreviated form of that given in Appendix C of Griffies et al. (1998).

13.1.1 Linear stability analysis for Laplacian neutral diffusion

The linear stability constraints described in Section 16.1.2 for un-rotated diffusion are extended here to the case of rotated diffusion. It has been found sufficient in practice to assume that stability for rotated diffusion is ensured if stability for each of the three directions separately holds. Less stringent constraints may result from a more general analysis, but such is not pursued here.

The linear numerical constraint from the diffusion equation, as discussed in Section 16.1.2 and Cox (1987), indicates that an explicit numerical scheme with a $2\Delta t$ forward time step is stable if the grid CFL number satisfies

$$\frac{|K^{mn}|\Delta t}{\Delta x_m \Delta x_n} \leq \frac{1}{4}, \tag{13.1}$$

where K^{mn} are components of the diffusion tensor (12.20), and the inequality must

hold for all combinations of $m, n = x, y, z$. Assuming a geophysically relevant vertical to horizontal aspect ratio for the grid ($\Delta z/\Delta x \leq 1/100$), the two dimensional horizontal sub-system is stable when the diffusion equation in the horizontal is stable. In general, satisfying this stability constraint in the horizontal is trivial and so is not considered further. Solving the vertical K^{zz} diagonal piece implicitly, as done by Cox (1987), points to the K^{xz} and K^{yz} cross terms as setting the most restrictive constraint. From these terms, the small angle neutral diffusion equation is linearly stable when, for each grid cell,

$$|S_a| \leq \frac{\Delta a \, \Delta z}{4 A_I \Delta t} \equiv \delta, \tag{13.2}$$

where Δa is either Δx or Δy. The parameter δ represents the maximum allowable slope that can be used before some prescription must be employed to ensure numerical stability. For many large-scale ocean model configurations, this slope check parameter is roughly $1/100$.

As discussed in Section 12.1.4, neutral slopes are typically smaller than $1/100$ for the interior ocean away from mixed layers and convective regions. For slopes larger than $1/100$, the motivation to remain "true" to the neutrally directed diffusion process remains less physically compelling. The reason is that in such large sloped regions, the ocean admits many other leading processes, such as three-dimensional boundary layer turbulence which generally mixes across neutral directions as well as along. Hence, restricting neutral diffusion to act in an unregularized manner only for slopes less than $1/100$ is not physically restrictive. Importantly, although we are not guided so much from physical principles regarding regularization scheme details, these details generally do matter, largely because they affect the upper ocean where boundary layer processes are crucial to establishing the ocean climate model's physical integrity. Hence, great care must be exercised, and it is furthermore important to test solution sensitivity to the various regularization approaches.

13.1.2 Linear stability analysis for biharmonic neutral mixing

Although argued to be of little physical or numerical utility, we list here the linear constraint for biharmonic neutral mixing discussed in Section 12.5.3. Following the previous subsection and building from the horizontal biharmonic analysis of Section 16.1.2 leads to the constraint

$$|S_a| \leq \frac{(\Delta a)^3 \, \Delta z}{64 \, B \Delta t} \equiv \delta. \tag{13.3}$$

Again, δ represents the maximum allowable slope which can be used before some prescription must be employed to ensure numerical stability. As for Laplacian neutral diffusion, for many large-scale ocean model configurations δ is roughly $1/100$.

13.1.3 Slope clipping the neutral diffusive fluxes

The first approach to maintaining numerical stability for small slope neutral diffusion is the *slope clipping* scheme of Cox (1987). Slope clipping limits the slope along which neutral diffusion occurs. For example, when the x-slope satisfies $|S_x| > S_{clip}$,

with S_{clip} some maximum slope usually taken as a fraction of the grid parameter δ given in equation (13.2), then

$$\rho_{,z} \rightarrow (\rho_{,z})_{clip}, \tag{13.4}$$

where

$$S_{clip} = \frac{|\rho_{,x}|}{(\rho_{,z})_{clip}}. \tag{13.5}$$

The same sort of clipping is applied independently to the y-component of the slope.

Limiting the slopes along which the fluxes are directed introduces significant false dianeutral fluxes in those regions where the actual slopes exceed the clipped slope. As mentioned in Section 12.1.4, one may expect increased cross-neutral fluxes in regions of steep slopes where there is the potential for enhanced energy release at all scales. Therefore, one may guess that slope clipping is a reasonable parameterization of these effects. However, the amount of uncontrolled fluxes can be egregiously large and unphysical.

An idealized example should convince the reader that slope clipping is not appropriate. Consider the case where density is a linear function only of potential temperature, and there are no vertical or meridional gradients. If maintaining a true neutral nature to the diffusive flux, each component of the temperature flux should vanish. However, as mentioned above, the clipped flux does not vanish. In particular, the off-diagonal piece of the vertical neutral flux of temperature becomes

$$F^{z\ clip} = A_I |\theta_{,x}|\ S_{clip} > 0. \tag{13.6}$$

The diagonal piece $S^2\,\theta_{,z}$ is handled via an implicit time stepping scheme. Focusing on the clipped off-diagonal term here, consider the typical case where $A_I = 10^3 \mathrm{m}^2\,\mathrm{sec}^{-1}$, $S_{clip} = 0.01$, $\theta_{,x} = 10^{-5}\ {}^\circ K\,\mathrm{m}^{-1}$. In this case, the clipped piece contributes to a vertical flux of potential temperature with magnitude $10^{-4}\,{}^\circ K\,\mathrm{m}\,\mathrm{sec}^{-1}$. When multiplied by $\rho_o C_p \approx 4 \times 10^6 J/(m^3\ {}^\circ K)$, the clipped flux contributes a heat flux of roughly 400 $\mathrm{W\,m}^{-2}$. This is a substantial unphysical heat flux. It has been seen to affect the solution integrity especially in convective regions of models run with this scheme. This result motivates other, more physically satisfying, means to satisfy linear stability constraints in regions of steep neutral slopes.

13.1.4 Quadratic tapering the neutral diffusive fluxes

Two methods aim to maintain numerical stability via a rescaling or tapering of the along neutral diffusivity. The diffusive flux remains oriented along the neutral direction, yet its magnitude is scaled smaller according to the steepness of the slope, once the slope reaches above some *a priori* set maximum slope.

The prescription of Gerdes et al. (1991) was the first method aiming to preserve the neutral orientation of the flux regardless the slopes. It does so via a quadratical rescaling of the diffusivity to a smaller value. That is, if the slope satisfies $|S| > S_{max}$, where S_{max} is the maximum slope, then the corresponding diffusivity is rescaled as

$$A_I \rightarrow A_I \left(\frac{S_{max}}{S}\right)^2. \tag{13.7}$$

A quadratic rescaling is suggested by the quadratic term $1 + S^2$ in the full slope neutral diffusion tensor given in equation (12.18).

Notably, as for the full slope tensor, the Gerdes et al. (1991) taper does not remove all components of the neutral diffusion tensor as the slope goes vertical. Instead, it leaves the diagonal $(3, 3)$ term saturated at the value

$$A_I \, S^2 \to A_I \, S^2_{max}, \tag{13.8}$$

whereas all other components taper to zero. With maximum slope of $1/100$ in each of the two horizontal directions, and diffusivity of $A_I = 10^3 \, \mathrm{m}^2 \, \mathrm{s}^{-1}$,

$$A_I \, S^2_{max} = 0.2 \mathrm{m}^2 \, \mathrm{s}^{-1}. \tag{13.9}$$

This is a large vertical diffusivity. In some cases, its effects will be much smaller than those from convective adjustment. However, this diffusivity is comparable to that diagnosed from some mixed layer schemes. This confusion of vertical mixing effects motivates us to look for yet another approach that allows for a clean separation.

13.1.5 Exponential tapering the neutral diffusive fluxes

Danabasoglu and McWilliams (1995) introduced an alternative rescaling in which the quadratic factor of Gerdes et al. (1991) is changed to a hyperbolic tangent

$$A_I \to A_I \, T_{\mathrm{tanh}} \tag{13.10}$$

with

$$T_{\mathrm{tanh}} = \left(1 + \tanh\left(\frac{S_{max} - |S|}{S_d}\right)\right) \tag{13.11}$$

the exponentially rapid taper function where S_d determines the width of the transition region. Since the hyperbolic tangent rapidly moves between its limiting values, it is possible to always taper the diffusivity instead of introducing the "if-test" check necessary for the quadratic tapering. Such continuous tapering is familiar from numerical implementations of physical processes such as vertical mixing.

This form of tapering ensures that all components of the neutral mixing tensor are rapidly scaled to zero as the slope becomes larger than S_{max}. Notably, even the $(3, 3)$ diagonal term is exponentially tapered to zero, and so this method does not suffer from the problem of the Gerdes et al. (1991) scheme identified in Section 13.1.4.

13.2 Linear stability for GM stirring

Skew fluxes from GM remain numerically stable so long as the corresponding eddy-induced velocity components maintain the same CFL constraint that the Eulerian velocities maintain. For illustrative purposes, let us assume the maximum CFL velocity is $1 \mathrm{m} \, \mathrm{s}^{-1}$. The scaling relation

$$U^* \sim \frac{\kappa \, S}{\Delta z}, \tag{13.12}$$

with $\kappa = 10^3\,\mathrm{m}^2\,\mathrm{s}^{-1}$ and $\Delta z = 10\,\mathrm{m}$, indicates that the slope S can be no larger than 0.01 in order for U^* to be less than $1\mathrm{m}\,\mathrm{s}^{-1}$. As discussed in Section 13.1, this slope corresponds to that typically setting the largest slope allowable for stable small slope neutral diffusion. Hence, it is practical to provide a numerically stable GM scheme if the GM flux is computed as if the slope times diffusivity cannot get any larger than the maximum value

$$\kappa \mathbf{S} \rightarrow (\kappa\,\mathbf{S})_{max} \tag{13.13}$$

in the regions where $S > S_{max}$.

If we wish to maintain a nontrivial GM flux even when the neutral slopes are vertical, care must be exercised when introducing numerical artifices commonly used to eliminate computational overflows. A useful method for this purpose is to compute the neutral slope as

$$\mathbf{S} = -\left(\frac{\nabla_z \rho}{\rho_{,z} - \epsilon}\right) = -\left(\frac{-\alpha\,\nabla_z \theta + \beta\,\nabla_z s}{-\alpha\,\theta_{,z} + \beta\,s_{,z} - \epsilon}\right). \tag{13.14}$$

The small number $\epsilon > 0$ is placed in the denominator to prevent computational overflows when there is no vertical stratification. Crucially, its contribution is subtracted from the denominator in order to provide the correct sign of the slope when $\rho_{,z} \rightarrow 0^-$. The example given by Figure 13.1 illustrates this point. In this figure, the x-projection of the neutral slope is positive, $S_x > 0$. Notably, the slope becomes vertical when moving towards the eastern-most isopycnal surface region since $\rho_{,z} \rightarrow 0^-$. The subtraction of ϵ in the denominator is necessary to have the computed slope go to $+\infty$. If one instead used a $+\epsilon$ in the denominator, the computed slope would go to $-\infty$ and so yield unphysical and unstable behavior.

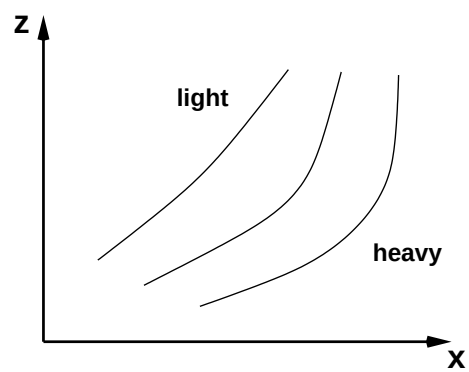

Figure 13.1: An example of steeply sloping isopycnals, where the slope in the x-direction, $S_x = -\rho_{,x}/\rho_{,z}$ is large and positive, especially near the surface in the east. The numerical slope computation must maintain this positive sign to compute the correct sign for the GM fluxes.

13.3 Neutral physics near boundaries

What to do with the neutral physics schemes when entering boundary regions, which are often where steep neutral slopes occur, remains a research question.

Should all neutral physics processes be turned off via some tapering scheme? Or should some of the parameterizations remain, such as GM? Should there be horizontal diffusion in the boundary layer, and if so, then what diffusivity should be used? Should neutral diffusivities remain the same as the skew-diffusivities regardless the neutral slope? How sensitive are the solutions to the choices?

This section provides some recommended answers to these questions. Most of the issues are physically based, although it is difficult to cleanly separate physical from numerical in this region, since we know from the previous discussion in this chapter that numerical considerations are imposed on us in most boundary regions. Given the research nature of these topics, a healthy skepticism is suggested. Throughout the discussion, we focus on the surface boundary layer, although we provide some comment on the bottom boundary layer.

13.3.1 General considerations

Surface and bottom boundary layers are typically where three-dimensional turbulent processes occur (e.g., Large et al. (1994)). Notably, the surface boundary layer is where strong diabatic processes associated with air-sea, ice-sea, and river-sea interactions occur. These physical considerations temper attempts to maintain the "purity" of neutral diffusion in such regions, where indeed such purity is more difficult to numerically maintain relative to the ocean interior. Instead, Treguier et al. (1997) note that the lateral mixing effects from eddies in the surface boundary layer should be parameterized via a *horizontal* diffusion, which is in fact diapycnal in regions where the neutral slopes are steeply sloped. These ideas suggest reducing neutral diffusion in the surface boundary layer to horizontal diffusion.

Haine and Marshall (1998) point out that mesoscale eddies retain an adiabatic slumping effect in surface boundary layer. Similarly, Send and Marshall (1995) and Visbeck et al. (1996) cite similar effects in regions of deep convection. Maintaining nonzero GM in these regions can provide an effective parameterization of these processes, and we provide a numerical method in the following for doing so.

These general considerations, and the discussion in this section, lead to the following recommendations for how to handle neutral physics in the surface boundary region.

- For neutral diffusion, replace the small angle neutral flux components with horizontal diffusive flux components in regions of steep neutral slopes and/or within some distance of boundaries.

- Maintain a non-zero GM skew-flux in the surface boundary region where neutral slopes are large. Do so by linearly tapering $\kappa\,\mathbf{S}$ to zero over the depth of the boundary layer. This approach leads to a constant horizontal eddy-induced velocity throughout the boundary layer, and a vertical eddy-induced velocity that is linearly tapered to zero as the ocean surface is approached.

- In regions with shallow neutral slopes that are near the surface boundaries, taper the neutral flux components to parameterize the truncation of the vertical eddy transport due to the presence of a boundary.

- With nonzero GM fluxes in the boundary layer such as recommended above, it is essential to ensure GM always acts on a stably stratified column, thus

necessitating the use of a full convection scheme such as that from Rahmstorf (1993). When combined with a large vertical diffusivity time stepped implicitly, the need to employ convective adjustment is reduced. This combined use of vertical diffusion plus convective adjustment may alleviate some of the unphysical behaviour sometimes characteristic of convective adjustment schemes (e.g., Cessi (1996)).

13.3.2 Surface boundary condition for fluxes from neutral physics

The tracer flux entering through the ocean surface arises from coupled processes; i.e., air-sea, river-sea, or ice-sea interations. The tracer flux entering through the ocean bottom is associated with geothermal and/or other geological effects. Through both interfaces there is zero tracer flux associated with neutral physics. Such can be understood from the quasi-Stokes transport ideas discussed in Section 7.3.5.1, as well as Section 8 of McDougall and McIntosh (2001). Analogously, we can understand these no-flux conditions as a statement of tracer conservation, whereby all boundary fluxes of tracer are associated with transport from or to other components in the climate system and are represented by physical processes not part of the neutral physics schemes.

Figure 13.2 illustrates these points for a surface model grid cell. Notably, because there is always a zero vertical flux at the ocean surface from neutral physics processes, there is a nontrivial jump in the vertical flux across the surface grid cell. Hence, as the surface grid cell thickness is refined, the vertical divergence of the vertical neutral physics tracer flux generally increases across the cell.

This numerical problem announces the need for some careful tapering of the neutral flux as the surface is approached. Such is important regardless the neutral slope, yet is most apparent in simulations where neutral slopes are nearly vertical in the upper ocean, as within the boundary layer. In this case, tracer fluxes from neutral physical processes, either from neutral diffusion or GM skew diffusion, can be large and so can exhibit an especially large vertical gradient across the surface grid cell. In either case, a large surface-trapped divergence reflects a physically-based need to taper the neutral processes towards zero as the boundary is approached. We now explain this reasoning.

13.3.3 A kinematic argument for tapering near a boundary

Recall from our discussion in Section 7.3.6 that an approximation for the vertical displacement of a fluid parcel living on an undulating potential density surface is given by equation (7.55), repeated here for convenience

$$\xi = -\rho'/\partial_z \rho. \tag{13.15}$$

In this equation, ρ is the potential density field explicitly carried by the numerical model, and ρ' is an eddy fluctuation. Now assume* that the fluctuation ρ' is proportional to a mixing length times the horizontal density gradient $\nabla_z \rho$. For mesoscale eddying motions, Smith and Vallis (2002) argued that the first baroclinic Rossby

*Thanks to Geoff Vallis (2002) for making this suggestion.

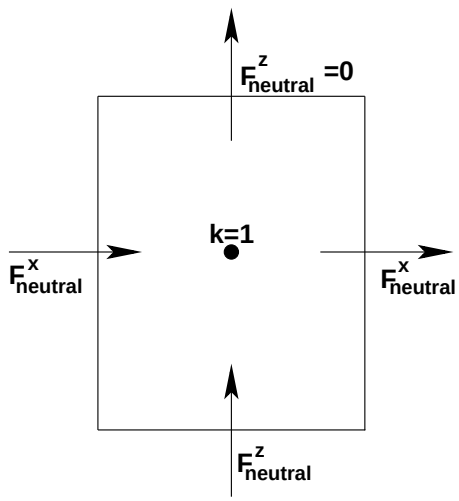

Figure 13.2: Details of how neutral physics is handled over the surface model grid cell. There is a zero neutral physics tracer flux entering the top of this cell, yet there is generally a nonzero flux crossing the bottom. Horizontal fluxes cross the sides of the cell, as for interior cells away from boundaries. The inverse picture is realized for tracer cells next to the ocean bottom, where there is zero vertical tracer flux from neutral physics processes entering through the bottom face of the bottom cells.

radius R_1 provides a useful mixing length. In this case, the vertical displacement takes the form

$$\xi \sim -R_1 \, |\nabla_z \rho| / \partial_z \rho = R_1 |\mathbf{S}|, \tag{13.16}$$

with $|\mathbf{S}|$ the magnitude of the neutral slope.

An analogous argument given in Appendix B of Large et al. (1997) and Section 8 of McDougall and McIntosh (2001) suggests that in the absence of a boundary, a typical neutrally directed displacement of a parcel is proportional to the first baroclinic Rossby radius R_1. For typical neutral slopes less than $1/100$, this lateral displacement is roughly the same as the horizontal displacement, and the corresponding vertical displacement is given by $\xi \sim R_1 \, |\mathbf{S}|$, which is the same result as derived above. Figure 13.3 illustrates this argument.

When a fluid parcel nears a boundary, if the vertical displacement $\xi \sim R_1 \, |\mathbf{S}|$ is greater than the distance to the boundary, the density surface associated with that eddy is truncated or clipped by the boundary (see Treguier et al. (1997), Held and Schneider (1999), and McDougall and McIntosh (2001) for discussions). In turn, the quasi-Stokes transport associated with this eddy is truncated. Consequently, we are obliged to reflect this truncation in the parameterized neutral physics fluxes.

13.3.4 Summary and recommendations

We have identified two reasons for tapering neutral physics schemes. First, there are numerical constraints discussed in Sections 13.1 and 13.2 that preclude our maintaining unregularized neutral physical processes in regions of steep neutral slopes. Second, there are kinematic reasons for tapering near boundaries where

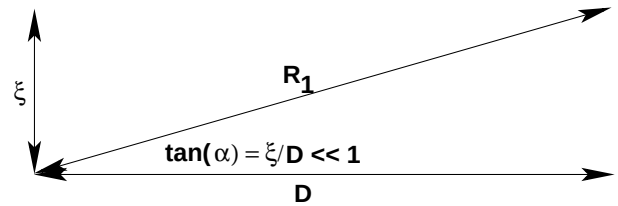

Figure 13.3: A fluid parcel within a mesoscale eddy undergoes both a horizontal and vertical displacement. The scale of the along neutral displacement is on the order of the first baroclinic Rossby radius R_1. Let the angle of the neutral direction α from the vertical be such that the neutral direction is shallow sloped: $|\mathbf{S}| = \tan\alpha << 1$. In this case, the corresponding vertical displacement is given by $\xi = D\,|\mathbf{S}| \approx R_1\,|\mathbf{S}|$.

undulating density surfaces can outcrop or incrop, and so their effects on transport are truncated.

There are two general methods to taper neutral physical processes. First, one may taper towards zero all neutral physics processes as the boundary and/or steep neutral slope region is approached. This is the approach favored by Large et al. (1997) and McDougall and McIntosh (2001). Alternatively, one may taper to zero only those pieces of the neutral physics schemes that affect a vertical transport. That is, one may continue to provide a horizontal downgradient diffusion and horizontal eddy-induced velocity transport. The vertical eddy-induced velocity must continue to be tapered towards zero as it reaches the boundary.

The second approach, whereby a nontrivial parameterized horizontal advective-diffusive transport is maintained near boundaries and/or in steep neutral slope regions, is promoted in the following. In particular, we are motivated by discussions in Treguier et al. (1997) and Held and Schneider (1999) which note that although eddies within the surface boundary layer can no longer transport vertically across the boundary, in the mean they affect a horizontal transport parameterizable as horizontal diffusion. Additional arguments for maintaining a nontrivial horizontal eddy-induced transport in the surface boundary layer are provided by the studies of Send and Marshall (1995), Visbeck et al. (1996), and Haine and Marshall (1998).

13.3.4.1 Review of the mathematical form for the fluxes

For use in the following, recall from Section 12.2.2 that the horizontal and vertical flux components arising from small-angle neutral diffusion, GM skewsion, plus vertical diffusion take the form

$$\begin{pmatrix} F^{(x)} \\ F^{(y)} \\ F^{(z)} \end{pmatrix} = \begin{pmatrix} A_I & 0 & (A_I - \kappa)\, S_x \\ 0 & A_I & (A_I - \kappa)\, S_y \\ (A_I + \kappa)\, S_x & (A_I + \kappa)\, S_y & A_D + A_I\, S^2 \end{pmatrix} \begin{pmatrix} C_{,x} \\ C_{,y} \\ C_{,z} \end{pmatrix} \tag{13.17}$$

which leads to the untapered flux components

$$\mathbf{F}^{(h)} = -A_I\, \nabla_z C - (A_I - \kappa)\, \mathbf{S}\, C_{,z} \tag{13.18}$$

$$F^{(z)} = -(A_D + A_I\, S^2)\, C_{,z} - (A_I + \kappa)\, \mathbf{S} \cdot \nabla_z C. \tag{13.19}$$

In these equations, A_I is the neutral diffusivity, A_D is the vertical diffusivity, and κ is the GM diffusivity. The first terms in the respective flux components are termed "diagonal", whereas the second terms are "off-diagonal".

13.3.4.2 Suggestions for neutral diffusion

Following the discussion in Section 13.1, our recommended tapering method for steep neutral slope regions employs the exponential taper of Danabasoglu and McWilliams (1995), yet does not taper the horizontal diagonal diffusive terms. The diagonal vertical piece is exponentially tapered as well in order to remove all vertical flux contributions to neutral diffusion in such regions. In this way, the diffusive flux components in steep neutral sloped regions reduce to those from horizontal-vertical diffusion

$$|\mathbf{S}| >> S_{max} \text{ and partial exp taper} \Rightarrow \begin{cases} \mathbf{F}^{(h)} & = -A_I \, \nabla_z C \\ F^{(z)} & = -A_D \, C_{,z}. \end{cases} \tag{13.20}$$

Figure 13.4 illustrates this approach. The integrity of the horizontal neutral diffusive flux components are "broken" in the steep sloped regions, thus compromising the neutral orientation of the flux. Yet as argued in Section 13.3.1 and Treguier et al. (1997), neutral orientation of the diffusive fluxes is not necessarily relevant in such boundary layer regions. Notably, this approach does not lead to the numerical problems identified in Section 13.1.3 associated with slope clipping, since the difficulties there arose from a clipped form of the off-diagonal diffusion terms, which are here properly tapered to zero.

Additional tapering is suggested in the near-boundary regions according to the arguments in Section 13.3.3. That is, we taper in these regions even if the neutral slope is shallow. In particular, we must ensure that there is a finite vertical divergence across a surface or bottom boundary cell from neutral diffusive fluxes. Since this tapering is meant to continuously truncate the neutral processes as the boundary is reached, an exponential tapering is arguably too rapid. Instead, we prefer the nearly linear tapering suggested by Large et al. (1997). In this case, to handle the surface boundary, we compute at each grid point the dimensionless ratio of depth of the grid point d to slope scaled Rossby radius

$$r = \frac{d}{R_1 \, |\mathbf{S}|}, \tag{13.21}$$

and then compute the boundary tapered slope times diffusivity

$$A_I \, S \rightarrow (A_I \, S) \, T_{\text{sine}} \tag{13.22}$$

where

$$T_{\text{sine}} = [1 + \sin \pi \, (r - 1/2)]/2 \tag{13.23}$$

defines a sine-taper function. Hence, in shallow sloped regions that are within a distance $R_1 \, |\mathbf{S}|$ of the ocean surface, we truncate the neutral flux components to

$$|\mathbf{S}| \leq S_{max} \text{ and depth} < R_1 \, |\mathbf{S}| \tag{13.24}$$

$$\Rightarrow \begin{cases} \mathbf{F}^{(h)} & = -A_I \, \nabla_z C - T_{\text{sine}} \, (A_I \, \mathbf{S}) \, C_{,z} \\ F^{(z)} & = -A_D \, C_{,z} - T_{\text{sine}} \, (A_I \, \mathbf{S}) \cdot (\mathbf{S} \, C_{,z} + \nabla_z C). \end{cases} \tag{13.25}$$

The first baroclinic Rossby radius is given by $R_1 = c/f$, with $f = 2\,\Omega\,\sin\phi$ the Coriolis parameter and Large et al. (1997) suggesting the value $c = 2\mathrm{m\,s}^{-1}$ as an approximate first baroclinic gravity wave speed. To handle the singularity at the equator, Large et al. (1997) restrict R_1 to the range 15km $\leq R_1 \leq$ 100km. An example is illustrative. Let the slope over the upper ocean be the modest value $S = 1/1000$ and let $R_1 = 100$km, thus yielding $R_1\,|\mathbf{S}| = \mathbf{100}$m. If the model grid cells are 10m thick over this part of the ocean, then for the $k = 1$ grid cell, $r = 0.1$ and $(A_I\,S)_{k=1} \approx .025\,(A_I\,S)$. For the $k = 9$ cell, $r = 0.1$ and $(A_I\,S)_{k=9} \approx .975\,(A_I\,S)$.

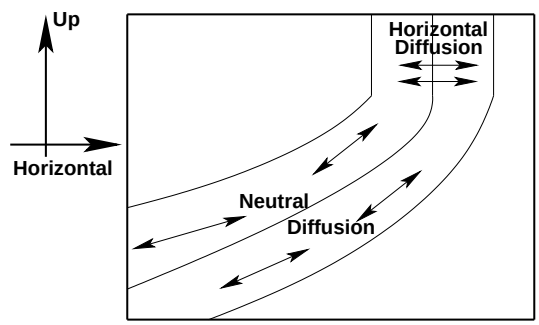

Figure 13.4: The recommended method for orienting the tracer fluxes from the neutral diffusion operator. In the ocean interior, where neutral slopes are modest, the neutral diffusion fluxes are oriented along neutral directions. In the regions where neutral slopes are steep, such as in the mixed layer, the diagonal horizontal pieces of the neutral diffusion operator are maintained, whereas the off-diagonal terms and vertical flux component is tapered to zero. Combined with the vertical diffusion operator, the resulting tracer diffusion is oriented in a horizontal-vertical manner in regions of steep neutral slopes.

In summary, we recommend the following tapering algorithm for neutral diffusion:

- For all neutral slopes regardless the distance of the grid point from the ocean surface, the diagonal horizontal diffusive terms are maintained without tapering.

- If the slope is steeper than S_{max}, then the exponential taper (13.11) is applied to all except the diagonal horizontal terms.

- If the slope is shallow ($|\mathbf{S}| < S_{max}$) and the grid cell is deeper than $R_1\,|\mathbf{S}|$ from the surface, then no tapering is applied.

- If the slope is shallow yet the grid cell is shallower than $R_1\,|\mathbf{S}|$ from the surface, than the sine taper (13.23) is applied to all except the horizontal diagonal terms.

- For the bottom, much remains to be studied. A reasonable and popular approach is to transition to an along-topography diffusion scheme such as that proposed by and Beckmann and Döscher (1997) and Döscher and Beckmann

(1999). More sophisticated bottom boundary layer schemes are at the cutting-edge of current research, with comments and references provided in the Griffies et al. (2000a) review paper.

13.3.4.3 Suggestions for GM

As described by Send and Marshall (1995), Visbeck et al. (1996), and Haine and Marshall (1998), baroclinic eddies are active near deep convective regions as well as more shallow surface boundary layers, both of which have steep neutral slopes. Consequently, mesoscale eddies provide a tendency to adiabatically flux density laterally and vertically as they extract the large store of available potential energy in the strongly baroclinic regions. One may thus choose to maintain a nontrivial GM flux in these regions to parameterize this process in coarse resolution models.

To motivate the form of this flux, we focus here on issues critical to the surface boundary layer. In particular, we consider a region of the upper ocean with vertical neutral directions, and ignore for now the near-boundary discussion in Section 13.3.3. If we set the product of GM diffusivity and slope to its saturated value $(\kappa \mathbf{S})_{max}$ within a column of cells with steeply sloping neutral directions, as allowed by numerical stability (see Section 13.2), then there is zero horizontal eddy-induced velocity (equation (12.56)) since $\mathbf{u}^* = -\partial_z (\kappa \mathbf{S})_{max} = 0$. The vertical velocity component $w^* = \nabla_z \cdot (\kappa \mathbf{S})_{max}$ is vertically constant and generally nonzero.

However, since all vertical components of the neutral physics fluxes vanish at the ocean top (see Figure 13.2), as noted by Gent et al. (1995), the vertical eddy-induced velocity must transition from a vertically constant value at the bottom of the surface grid cell to zero at the grid cell's top. This rapid transition induces a large horizontal velocity $\mathbf{u}^*$ within the surface grid cell needed to satisfy continuity. As the thickness of the surface grid cell becomes smaller via grid cell refinement or free surface height undulations (see Figure 10.1), the transition region becomes thinner and so induces a larger jump in eddy-induced velocity across the surface grid cell, ultimately becoming a delta function in the limit similar to that discussed by Bretherton (1966), Killworth (1997), and Killworth (2001). The left panel of Figure 13.5 illustrates this point.

A numerically robust approach capturing the horizontal eddy-induced transport is to let the product $\kappa \mathbf{S}$ linearly move from its saturated value $(\kappa \mathbf{S})_{max}$ at the base of the boundary layer, to zero at the top of the surface grid cell. This approach was suggested by Treguier et al. (1997) and Greatbatch and Li (2000). In this way, the horizontal eddy-velocity $\mathbf{u}^*$ is vertically constant throughout the boundary layer with a magnitude inversely proportional to the boundary layer depth. The vertical velocity w^* also smoothly goes to zero at the surface. The right panel of Figure 13.5 illustrates this choice. Mathematically, this choice leads to the GM flux in steep sloped regions

$$|\mathbf{S}| >> S_{max} \Rightarrow \begin{cases} \mathbf{F}^{(h)} & = (\kappa \mathbf{S})_{\text{linear taper}} \, C_{,z} \\ F^{(z)} & = -(\kappa \mathbf{S})_{\text{linear taper}} \cdot \nabla_z C. \end{cases} \tag{13.26}$$

The subscript "linear taper" symbolizes linear tapering discussed above. Note that because we have the maximum value of $(\kappa \mathbf{S})$ at the base of the boundary layer, it is

here that restratification tendencies from GM are strongest. Finally, the vector character of $(\kappa\,\mathbf{S})_{\text{linear taper}}$ signifies that its sign is taken from the sign of the actual slope. A regularization method for computing slopes whose values approach infinity is given in Section 13.2. Figure 13.6 illustrates the tendency of the corresponding eddy-induced streamfunction.

Now considering the notions discussed in Section 13.3.3, for shallow sloped regions within a distance $R_1\,|\mathbf{S}|$ of the ocean surface, yet outside of the regions where the slopes are steeper than S_{max}, we truncate the GM flux components according to

$$|\mathbf{S}| < S_{max} \;\text{ and }\; \text{depth} < R_1\,|\mathbf{S}| \Rightarrow \begin{cases} \mathbf{F}^{(h)} & = (\kappa\,\mathbf{S})\,T_{\text{sine}}\,C_{,z} \\ F^{(z)} & = -(\kappa\,\mathbf{S})\,T_{\text{sine}}\cdot\nabla_z C \end{cases} \tag{13.27}$$

where T_{sine} is the sine-taper function defined by equation (13.23).

In summary, we recommend the following tapering algorithm for GM skew-sion:

- If a grid point is within the surface boundary layer, defined as that region starting at the ocean surface where neutral slopes are steeper than S_{max}, then let $\kappa\,\mathbf{S}$ linearly move from its saturated value $(\kappa\,\mathbf{S})_{max}$ at the base of the boundary layer, to zero at the top of the surface grid cell. This approach ensures a vertically constant horizontal eddy-induced velocity within the boundary layer, and a vertical eddy-induced velocity that linearly decays to zero at the ocean surface.

- If a grid point is outside the boundary layer, which means the slope is shallow ($|\mathbf{S}| < S_{max}$), yet the grid point is shallower than the penetration depth $R_1\,|\mathbf{S}|$, then the sine taper (13.23) is applied to all pieces of the GM flux.

- If the slope at a grid point is shallow ($|\mathbf{S}| < S_{max}$) and the depth is greater than $R_1\,|\mathbf{S}|$, then no tapering is applied.

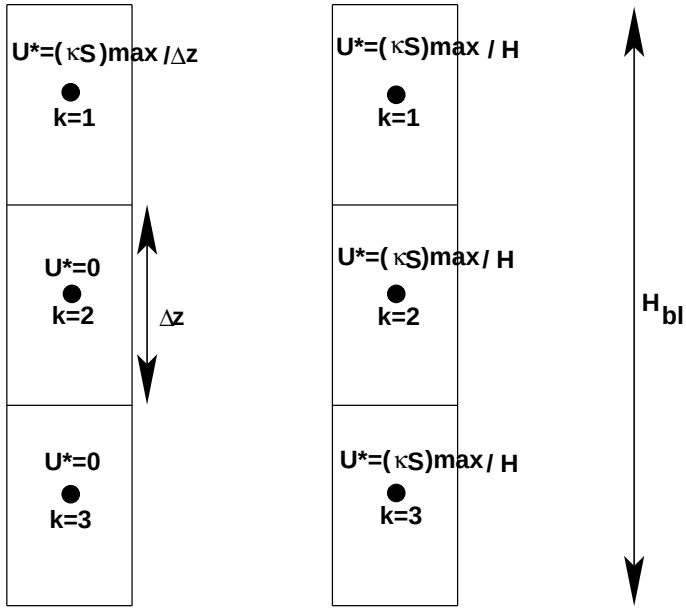

Figure 13.5: Horizontal eddy-induced velocity resulting from two different treatments of GM in the regions of steep neutral slopes. The right option is preferred since it leads to a well defined limit as the vertical grid spacing is refined, and provides for a constant horizontal eddy-velocity throughout the boundary layer instead of a large velocity just in the surface cell.

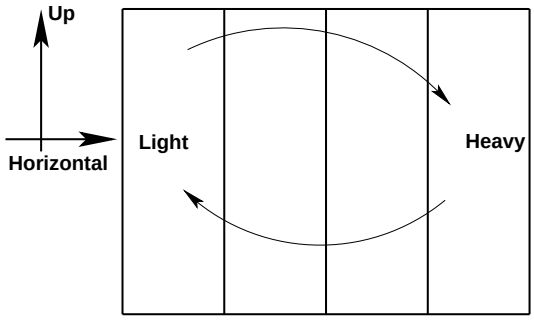

Figure 13.6: The recommended method for handling GM fluxes in regions where the neutral slope is steep. Shown is a schematic case where the GM flux is nonzero and oriented in a manner to maintain a nontrivial eddy-induced or quasi-Stokes streamfunction which moves light water over heavy water, thus slumping the isopycnals and reducing baroclinicity.

CHAPTER
FOURTEEN

Neutral physics discretization

Contents

14.1 Foundations for discrete neutral physics 300
14.1.1 Dissipation of tracer variance 300
14.1.2 Functional formalism 301
14.1.2.1 Dissipation functional 301
14.1.2.2 Functional derivative 301
14.1.2.3 Approach on the discrete lattice 303
14.1.3 Neutral directions 303
14.1.4 Active and passive tracers 303
14.1.5 Dissipation functional for neutral diffusion 304
14.1.5.1 Full Redi diffusion tensor 304
14.1.5.2 Small angle diffusion tensor 304
14.2 Introduction to the discretization 305
14.2.1 Summary and Caveats 305
14.2.2 General procedure for discretization 306
14.2.3 Some conventions 306
14.3 A one-dimensional warm-up 307
14.4 Elements of the discrete functional 308
14.4.1 Grid partitioning for the small angle functional 308
14.4.2 Vertical grid dimensions with partial cells 309
14.4.3 Horizontal tracer gradients and the minimum thickness rule 310
14.4.4 Quarter cell volumes 311
14.4.5 Tracer gradients within the quarter cells 312
14.4.6 Schematic form of the discretized functional 313
14.5 Triads and some notation 314
14.5.1 Tracer and density triads 314
14.5.2 Reference points for computing density derivatives 314
14.5.3 Notation for neutral direction slopes and diffusivities . . . 315
14.5.4 Notation for the quarter-cell dimensions 315
14.6 The discrete diffusion operator 316

14.6.1 The discretized x-z dissipation functional 316
14.6.2 Derivative of the x-z dissipation functional 316
14.6.3 Discretized diffusion operator 320
14.7 Diffusive flux components 320
14.7.1 Continuum considerations 320
14.7.2 Discrete fluxes 321
14.7.3 Identifying diagonal components to the diffusion tensor . 322
14.7.4 Stencils for small angle flux components 323
14.8 Some further issues of numerical implementation 323
14.8.1 Gent-McWilliams skew-flux plus neutral diffusive flux . . 323
14.8.2 Discretization of the GM velocity for diagnostics 323
14.8.3 Computing the vertical flux convergence 326

The purpose of this chapter is to derive the discrete small angle neutral diffusion operator. This discretization is also of use for discrete Gent and McWilliams (1990) skewsion according to Griffies (1998). As all details focus on fields defined at the tracer point, the discussion is generic to z-models with arbitrary grid arrangements.

The presentation introduces grid labels on fields living on a discrete lattice. In particular, discrete tracer fields in this chapter are denoted by the symbol $T_{i,j,k}$, which is the notation used in the GFDL Modular Ocean Model (MOM). The symbol C was used in other chapters for tracer concentration. Explicit sums and averages are also used in most places in this chapter rather than choosing amongst the more elegant alternatives. We choose this more explicit notation since it leads to less ambiguity for the reader. Furthermore, such is ultimately required by the numericist when translating an algorithm into computer code. Experience has shown that having the algorithm written nearly in its "full glory" in LaTeXreduces the chances for bugs to penetrate the computer code. Such is especially crucial for complicated schemes such as the one documented in this chapter. Hence, for pedogogical and "quality control" reasons, we prefer the explicit grid-label notation. Recognizing that this approach can lead to somewhat lengthy equations, we use a relatively small font for the remainder of this chapter to allow for visually more compact expressions.

14.1 Foundations for discrete neutral physics

The purpose of this section is to set the foundations for discretization of the neutral diffusion and skew-diffusion operators.

14.1.1 Dissipation of tracer variance

Diffusive mixing reduces tracer variance. Such can be seen by taking the time derivative of the variance, assuming a constant volume domain devoid of sources and boundary fluxes, and focusing on the effects due to mixing by a symmetric and positive semi-definite tensor

(recall definition of tracer variance from Section 12.5.1)

$$
\begin{aligned}
(V/2)\,\mathcal{V}_{,t} &= \int dV\, C\, C_{,t} \\
&= \int dV\, C\, (K^{mn}\, C_{,n})_{,m} \\
&= \int dV\, (C\, K^{mn}\, C_{,n})_{,m} - \int dV\, C_{,m}\, K^{mn}\, C_{,n} \\
&= -\int dV\, \nabla \cdot (C\, \mathbf{F}) + \int dV\, \nabla C \cdot \mathbf{F},
\end{aligned} \tag{14.1}
$$

where

$$F^m = -K^{mn}\, C_{,n} \tag{14.2}$$

is the diffusive tracer flux. The total derivative term can be dropped when there are no diffusive fluxes normal to the domain boundaries, as assumed here. The second term is negative semi-definite since diffusive fluxes are down the tracer gradient, or equivalently the diffusion tensor is positive semi-definite

$$\nabla C \cdot \mathbf{F} = -C_{,m}\, K^{mn}\, C_{,n} \leq 0. \tag{14.3}$$

14.1.2 Functional formalism

The diffusion operator, when acting on a passive tracer, is a linear self-adjoint operator. As such, it has an associated negative semi-definite functional (e.g., Courant and Hilbert (1953, 1962)). For example, the Laplacian operator $\nabla^2 C$ is identified with the functional derivative $\nabla^2 C = \delta\mathcal{F}/\delta C$, where $\mathcal{F} \equiv -(1/2)\int |\nabla C|^2 dV$ is the associated functional. We derive this result in the general case in this section.

14.1.2.1 Dissipation functional

For a general diffusion operator, the *dissipation functional* is written

$$\mathcal{F} = \int d\mathbf{x}\, \mathcal{L}, \tag{14.4}$$

where the non-positive integrand is given by

$$2\,\mathcal{L} = \mathbf{F} \cdot \nabla C = -C_{,m}\, K^{mn}\, C_{,n} \tag{14.5}$$

and it is convenient to write the volume element as $dV = d\mathbf{x}$. The name *dissipation* is attached to the functional $\mathcal{F}$ due to its relation to the dissipation of tracer variance described in the previous section

$$\mathcal{F} = \frac{V}{4}\, \mathcal{V}_{,t}. \tag{14.6}$$

Hence, properties of a discrete realization of the functional $\mathcal{F}$ are shared by evolution of the discrete tracer variance.

14.1.2.2 Functional derivative

We now illustrate the connection between the functional derivative of $\mathcal{F}$ and the diffusion operator

$$R(C) = -\nabla \cdot \mathbf{F}(C). \tag{14.7}$$

For this purpose, consider functional variations of the tracer field δC, thus leading to variations in the functional given by

$$\delta\mathcal{F} = \int d\mathbf{x} \left(\frac{\delta\mathcal{L}}{\delta C}\, \delta C + \frac{\delta\mathcal{L}}{\delta C_{,m}}\, \delta C_{,m} \right). \tag{14.8}$$

Integration by parts leads to

$$\delta\mathcal{F} = \int d\mathbf{x} \left(\frac{\delta\mathcal{L}}{\delta C}\,\delta C + \partial_m \left(\frac{\delta\mathcal{L}}{\delta C_{,m}}\,\delta C \right) - \partial_m \left(\frac{\delta\mathcal{L}}{\delta C_{,m}} \right) \delta C \right). \tag{14.9}$$

The second term is a total derivative that integrates to a surface term. Assuming the tracer field satisfies *natural boundary conditions*, whereby

$$\hat{\mathbf{n}} \cdot \frac{\delta\mathcal{L}}{\delta\nabla C} = 0 \tag{14.10}$$

with $\hat{\mathbf{n}}$ an outward normal at boundaries, allows for the variation to be written

$$\delta\mathcal{F} = \int d\mathbf{x} \left(\frac{\delta\mathcal{L}}{\delta C} - \partial_m \left(\frac{\delta\mathcal{L}}{\delta C_{,m}} \right) \right) \delta C. \tag{14.11}$$

Consequently, the functional derivative is given by

$$(d\mathbf{y})^{-1} \frac{\delta\mathcal{F}}{\delta C(\mathbf{y})} = \frac{\delta\mathcal{L}}{\delta C} - \partial_m \left(\frac{\delta\mathcal{L}}{\delta C_{,m}} \right), \tag{14.12}$$

where $d\mathbf{y}$ is the volume element at the field point $\mathbf{y}$. To reach the last step, it was necessary to use the identity

$$\frac{\delta C(\mathbf{x})}{\delta C(\mathbf{y})} = d\mathbf{y}\,\delta(\mathbf{x}-\mathbf{y}), \tag{14.13}$$

where $\delta(\mathbf{x}-\mathbf{y})$ is the Dirac delta-function satisfying

$$\int d\mathbf{y}\,\delta(\mathbf{x}-\mathbf{y}) = 1, \tag{14.14}$$

so long as the integration range includes the singular point $\mathbf{x} = \mathbf{y}$. Note that by definition, the delta-function has physical dimensions of inverse volume L^{-3}

We now reintroduce the specific form of the diffusion integrand

$$2\,\mathcal{L} = -C_{,m}\,K^{mn}\,C_{,n}, \tag{14.15}$$

thus leading to

$$(d\mathbf{y})^{-1} \frac{\delta\mathcal{F}}{\delta C(\mathbf{y})} = -\partial_m \left(\frac{\delta\mathcal{L}}{\delta C_{,m}} \right) = -\partial_m \left(K^{mn}\,C_{,n} \right). \tag{14.16}$$

The right hand side is identified as the diffusion operator, thus leading to the general result

$$(d\mathbf{y})^{-1} \frac{\delta\mathcal{F}}{\delta C(\mathbf{y})} = R(C(\mathbf{y})). \tag{14.17}$$

Natural boundary conditions are maintained by tracer fluxes satisfying the usual no-normal flux boundary condition

$$\hat{\mathbf{n}} \cdot \mathbf{F} = 0, \tag{14.18}$$

with

$$\frac{\delta\mathcal{L}}{\delta\nabla C} = \mathbf{F}(C). \tag{14.19}$$

Thus completes the connection between the dissipation functional, the diffusion fluxes, and the diffusion operator.

14.1.2.3 Approach on the discrete lattice

On the discrete lattice, not every consistent* discrete diffusion operator corresponds to a negative semi-definite discrete functional. Therefore, a consistent numerical diffusion operator does not necessarily possess the dissipative properties of the continuum operator. For the Laplacian operator in an isotropic media, it is trivial to produce a dissipative numerical operator. In the anisotropic case, such as neutral diffusion, it is nontrivial. Indeed, the original discretization of the neutral diffusion operator in the GFDL model (Cox (1987)) is numerically consistent but not always dissipative. Griffies et al. (1998) provide examples.

The traditional method of discretization in ocean modeling is to first discretize a particular operator and then to *hope* that it satisfies certain global properties present in the continuum. The alternative approach used here is to first discretize the dissipation functional, which is a global object of interest. Thereafter, the discrete version of the functional derivative is taken, which then produces the discrete diffusion operator local to a grid point. This approach ensures that the discretized diffusion operator dissipates tracer variance.

14.1.3 Neutral directions

Within the functional framework, provision is made for a discretization of the diffusive fluxes that are aligned according to a *self-consistent* approximation to the neutral directions. Self-consistency means that there is a zero neutral diffusive flux of locally referenced potential density. As shown in Section 12.1.6, a zero neutral flux of locally referenced potential density implies a balance between the neutral direction diffusive flux of potential temperature and salinity

$$\alpha \mathbf{F}(\theta) = \beta \mathbf{F}(s), \tag{14.20}$$

where

$$\alpha = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial \theta}\right)_{s,p} \tag{14.21}$$

and

$$\beta = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial s}\right)_{\theta,p} \tag{14.22}$$

are the thermal expansion and saline contracion coefficients introduced in Section 11.6.1. In order to ensure this balance in a z-coordinate ocean model, it is sufficient to compute the density gradients in terms of the active tracer gradients and the thermal and saline coefficients. Special care must be taken when choosing the reference points for evaluating these gradients, with details given in Section 14.5.2. Without a self-consistent discretization that guarantees a zero flux of locally referenced potential density, the neutral diffusion operator can produce grid noise even if it ensures variance does not increase (Griffies et al. (1998)).

14.1.4 Active and passive tracers

When diffusing passive tracers, the neutral diffusion operator is linear since the diffusion tensor is independent of the passive tracer fields. However, when diffusing the active tracers potential temperature and salinity, the neutral diffusion operator is nonlinear since the diffusion tensor is itself a function of the active tracers. Since the functional formalism assumes the diffusion operator to be a linear self-adjoint operator, one may question the validity of using this formalism for the active tracers.

However, there is no problem applying the formalism since the diffusion operator is structurally the same for both the active and passive tracers. What is done is to use the functional machinery assuming the tracer is passive. Yet when it comes time to discretize

*Consistent here means that the discretization reduces to the correct continuum operator as the grid size goes to zero.

the neutral directions, the understanding of how to properly align the slopes so that the diffusion operator is self-consistent (i.e., so that it does not flux locally referenced potential density along the neutral directions) is incorporated.

14.1.5 Dissipation functional for neutral diffusion

Here we record the dissipative functional in the continuum for neutral diffusion using both the full Redi diffusion tensor and the small slope approximation. Note that only the small slope diffusion tensor is discretized in z-coordinate ocean climate models.

14.1.5.1 Full Redi diffusion tensor

As discussed in Sections 11.6.3 and 12.1, it is possible to write the Redi diffusion tensor representing along and across neutral diffusion in the projection operator form

$$\begin{aligned} K^{mn} &= A_I(\delta^{mn} - \hat{\rho}^m\hat{\rho}^n) + A_D\hat{\rho}^m\hat{\rho}^n \\ &= (A_I - A_D)(\delta^{mn} - \hat{\rho}^m\hat{\rho}^n) + A_D\delta^{mn} \end{aligned} \tag{14.23}$$

with δ^{mn} the Kronecker delta and

$$\hat{\boldsymbol{\rho}} = \frac{\nabla\rho}{|\nabla\rho|} = \frac{-\alpha\,\nabla\theta + \beta\,\nabla s}{|-\alpha\,\nabla\theta + \beta\,\nabla s|} \tag{14.24}$$

the unit vector normal to the neutral direction. The diffusivities A_I and A_D are non-negative and can in general be functions of space-time. As shown in Section 11.6.3, the neutral piece of the diffusion tensor projects a vector onto the local neutral direction, and the dianeutral piece projects a vector into the local normal direction.

Given this diffusion tensor, the functional for neutral diffusion is written

$$2\,\mathcal{F} = -\int dV\ (A_I - A_D)\frac{|\nabla\rho\wedge\nabla C|^2}{|\nabla\rho|^2} - \int dV\ A_D|\nabla C|^2, \tag{14.25}$$

or more explicitly,

$$\begin{aligned} 2\,\mathcal{F} = &-\int dV\ (A_I - A_D)\frac{(C_{,x}\rho_{,y} - C_{,y}\rho_{,x})^2 + (C_{,y}\rho_{,z} - C_{,z}\rho_{,y})^2 + (C_{,z}\rho_{,x} - C_{,x}\rho_{,z})^2}{\rho_{,x}^2 + \rho_{,y}^2 + \rho_{,z}^2} \\ &-\int dV\ A_D(C_{,x}^2 + C_{,y}^2 + C_{,z}^2). \end{aligned} \tag{14.26}$$

Hence, the diffusive flux is given by

$$\mathbf{F} = \frac{\delta\mathcal{L}}{\delta\,\nabla C} = -(A_I - A_D)(\hat{\boldsymbol{\rho}}\wedge\nabla C)\wedge\hat{\boldsymbol{\rho}} - A_D\nabla C. \tag{14.27}$$

The first term vanishes when the tracer is parallel to the neutral directions. For most oceanographic cases, the difference $A_I - A_D \approx A_I$ to many orders of accuracy. Note the presence of $A_D C_{,x}$ and $A_D C_{,y}$ in the full tensor. In the case of very steep neutral directions, these terms become relevant. However, as discussed in Section 12.1.4, such regions possess mostly three-dimensionally turbulent processes, where a neutral formalism in less physically relevant.

14.1.5.2 Small angle diffusion tensor

As discussed in Section 12.1.4, the small slope approximation amounts to taking the limit of $|\rho_{,x}|, |\rho_{,y}| << |\rho_{,z}|$, and dropping terms of order $slope * (A_D/A_I)$, with $slope$ the small

neutral slope. The resulting functional is

$$
\begin{aligned}
2\,\mathcal{F}^{small} &= -\int dV\; A_I \frac{(C_{,x}\rho_{,z} - C_{,z}\rho_{,x})^2 + (C_{,y}\rho_{,z} - C_{,z}\rho_{,y})^2}{\rho_{,z}^2} - \int dV\; A_D (C_{,z})^2 \\
&= -\int dV\; A_I\; (\nabla_z C + \mathbf{S}\, C_{,z})^2 - \int dV\; A_D (C_{,z})^2 \qquad (14.28)
\end{aligned}
$$

where $\mathbf{S}$ is the neutral slope vector. The corresponding diffusive flux is given by

$$
\mathbf{F} = \frac{\delta \mathcal{L}}{\delta\, \nabla C} = -\left(\frac{A_I}{\rho_{,z}}\right)\, (\nabla \rho \wedge \nabla C) \wedge \hat{\mathbf{z}} - \hat{\mathbf{z}}\, A_D\, C_{,z} \qquad (14.29)
$$

which has components

$$
\mathbf{F}^{(h)} = -A_I\, (\nabla_z C + \mathbf{S} C_{,z}) \qquad (14.30)
$$

$$
F^z = \mathbf{S} \cdot \mathbf{F}^{(h)} - A_D C_{,z} = -A_I (\mathbf{S} \cdot \nabla_z C + S^2\, C_{,z}) - A_D C_{,z} \qquad (14.31)
$$

As for the full tensor, the A_I term in these expressions vanishes when the tracer is parallel to the neutral directions.

14.2 Introduction to the discretization

The remainder of this chapter focuses on the small angle neutral diffusion operator, and the associated skew fluxes arising from Gent and McWilliams (1990) (GM in the following). Pacanowski and Griffies (1999) also presented a discretization of the full neutral diffusion tensor. For reasons stated earlier (Section 12.1.4), the small angle diffusion tensor is sufficient for all ocean-climate modeling purposes with z-models, and so discretization of the full tensor of Redi (1982) is not discussed here.

14.2.1 Summary and Caveats

As documented by Griffies et al. (1998), the discrete neutral diffusion operator described in this chapter does the following:

1. It produces a zero horizontal flux of locally referenced potential density as embodied by the balance given in equation (12.32). The Cox (1987) scheme did not respect this property, and this problem led to the nonlinear numerical instability of that scheme.
2. It reduces tracer variance, and produces downgradient oriented tracer fluxes along neutral directions when considering a particular finite volume region of the discrete lattice.
3. It computes a second order approximation to the neutral slopes.
4. It requires zero background horizontal diffusion to remain numerically stable; the Cox (1987) scheme blew-up without this diffusion.

Some limitations of the discrete neutral diffusion operator include the following:

1. Because part of the vertical flux is computed explicitly in time and part implicitly, the balance between the vertical component of the neutral diffusive salinity and temperature fluxes (equation (12.32)) is slightly broken. This problem has proven to be minor in realistic settings, though it is useful to keep in mind.
2. Because it only guarantees downgradient oriented tracer fluxes over finite volumes, rather than individual tracer cells, tracers can move outside their physically constrained range. Consequently, the scheme does not guarantee positive-definiteness

of tracers. This problem appears most pernicious for passive tracers in steep sloped regions. The temperature and salinity, being constrained by the need to produce a zero flux of locally referenced potential density, appear less problematic, though they too can move outside physical bounds. In order to do the pure neutral diffusion problem for all slopes, it appears necessary to employ a positive-definite neutral diffusion scheme. Proposals are discussed by Beckers et al. (2000).

As argued in Chapter 13 (see especially Section 13.3), it is physically justifiable to apply horizontal fluxes in regions where neutral slopes are steep. In this case, the numerical issues appearing in such regions become less problematic.

3. The discrete neutral diffusion scheme requires no horizontal background diffusion previously required with the Cox (1987) scheme. Unfortunately, removing horizontal diffusion exposes the simulations to problems with dispersive advection schemes. These problems are most apparent next to topography in regions where the tracers are aligned with neutral directions, and so diffusive fluxes are weak. This "Peclet grid noise" problem is fundamental to dispersive advection schemes, and it further prompts the use of horizontal diffusion next to boundary layers. Another means for removing these problems is to use the sigma diffusion scheme of Beckmann and Döscher (1997) and Döscher and Beckmann (1999), which is a rudimentary scheme aiming to move dense water down slopes.

14.2.2 General procedure for discretization

Motivated from the discussion in Section 14.1.2, discretization of the diffusion operator at a particular grid point is derived from the functional derivative of the discretized neutral diffusion functional $\mathcal{F}[T]$. On a discrete lattice, the functional derivative becomes a partial derivative

$$R(T)_{i,j,k} = \frac{1}{V_{T_{i,j,k}}} \frac{\partial \mathcal{F}[T_{i,j,k}]}{\partial T_{i,j,k}}, \tag{14.32}$$

where

$$V_{T_{i,j,k}} = dxt_{i,j}\, dyt_{i,j}\, dht_{i,j,k} \tag{14.33}$$

is the volume of the tracer cell (T-cell). This result corresponds to a discrete version of the continuum relation given by equation (14.17).

The general procedure for discretization is to identify those pieces of the discretized functional that contain contributions from the discretized tracer value $T_{i,j,k}$. These are the pieces possessing nonzero functional derivatives, and so contribute to the diffusion operator for this T-cell. An enumeration of these pieces depends on the particular discretization of the functional. Most models choose second order numerics. Higher order numerics do not appear to be motivated for an operator aiming to dissipate via mixing. One overriding principle used to guide this discretization is to recover the familiar discretization of the Laplacian in the case of zero neutral slopes, where neutral diffusion reduces to horizontal diffusion.

All details about the discretization are made at the level of the functional. These details include the particular grid choice and the choice for reference points to be used in approximating the neutral directions.

14.2.3 Some conventions

To reduce clutter, no reference is made to the time step. As it is necessary to lag the time step for numerical stability of the diffusion equation (e.g., Haltiner and Williams (1980)), a time step of $\tau - 1$ is assumed throughout.

It is sufficient to work through details for the two-dimensional x-z plane. Generalization to the three dimensional case is straightforward and will be made at the end.

14.3 A one-dimensional warm-up

In order to illustrate the general framework provided by the functional approach, it is useful to consider the case of one-dimensional diffusion along the zonal direction. Neutral diffusion does not exist in one-dimension, so this example cannot illustrate any of the subtle issues related to discretization of the neutral directions. The issues of partial vertical cells are also avoided here and will be addressed later.

Figure 14.1 shows the zonal grid. The grid is partitioned into four subcells called *quarter cells* in the following, with each possessing a generally different diffusivity $A(n)$ and length $V(n)$ (generalized to volume in the three-dimensional case). The quarter cells each contain a contribution from the central tracer point T_i upon discretizing the functional with second order numerics. That is, we consider a second order discretization of the functional

$$\mathcal{F} = -\frac{1}{2}\sum_i \sum_n A(n)V(n)(\delta_x T(n))^2, \tag{14.34}$$

where the n sum is over the subcells relevant for each T-cell. The four terms containing a contribution from T_i are given by

$$\begin{aligned} 2\mathcal{F}[T_i] = -A(1)V(1)(\delta_x T_{i-1})^2 - A(2)V(2)(\delta_x T_{i-1})^2 \\ - A(3)V(3)(\delta_x T_i)^2 - A(4)V(4)(\delta_x T_i)^2, \end{aligned} \tag{14.35}$$

where the tracer derivative is assumed to be the same for the two subcells 1 and 2 and the two subcells 3 and 4. The "volumes" of the subcells are

$$V(1) = dte_{i-1} \tag{14.36}$$
$$V(2) = dtw_i \tag{14.37}$$
$$V(3) = dte_i \tag{14.38}$$
$$V(4) = dtw_{i+1} \tag{14.39}$$

and the derivatives are

$$\delta_x T_{i-1} = \frac{T_i - T_{i-1}}{dxte_{i-1}} \tag{14.40}$$
$$\delta_x T_i = \frac{T_{i+1} - T_i}{dxte_i}. \tag{14.41}$$

Taking the derivative of the functional with respect to T_i leads to

$$\begin{aligned} -2\frac{\partial \mathcal{F}}{\partial T_i} &= \frac{\delta_x T_{i-1}}{dxte_{i-1}}\left(A(1)\,V(1) + A(2)\,V(2)\right) - \frac{\delta_x T_i}{dxte_i}\left(A(3)\,V(3) + A(4)\,V(4)\right) \\ &= \frac{\delta_x T_{i-1}}{dxte_{i-1}}\left(A(1)\,dte_{i-1} + A(2)\,dtw_i\right) - \frac{\delta_x T_i}{dxte_i}\left(A(3)\,dte_i + A(4)\,dtw_{i+1}\right). \end{aligned} \tag{14.42}$$

Defining a distance weighted average diffusivity

$$A(1)\,dte_{i-1} + A(2)\,dtw_i = 2\left(\overline{dtx\,A}^x\right)_{i-1} \tag{14.43}$$
$$A(3)\,dte_i + A(4)\,dtw_{i+1} = 2\left(\overline{dtx\,A}^x\right)_i \tag{14.44}$$

leads to

$$\frac{\partial \mathcal{F}}{\partial T_i} = dxt_i\,\delta_x\left(\frac{\left(\overline{dtx\,A}^x\right)_{i-1}}{dxte_{i-1}}\,\delta_x T_{i-1}\right) \tag{14.45}$$

and to the diffusion operator

$$R(T)_i = \frac{1}{dxt_i}\frac{\partial \mathcal{F}}{\partial T_i} = \delta_x \left(\frac{\left(\overline{dtx\, A}^{x}\right)_{i-1}}{dxte_{i-1}} \delta_x T_{i-1} \right). \tag{14.46}$$

Note that for a constant grid with a constant diffusivity, this discretization reduces to the standard 3-point discrete one-dimensional Laplacian.

The main points to be taken from this example are the following:

1. We assumed that the tracer gradient is the same across the two adjacent subcells 1 & 2 and 3 & 4, respectively.

2. The grid weighted average operator naturally appeared.

3. We introduced a diffusion coefficient associated with a quarter tracer cell as opposed to a full tracer cell.

The assumption about tracer gradients across subcells is necessary to have the functionally derived discretization reduce to the traditional 3-point Laplacian when the grid and diffusivity are constant.

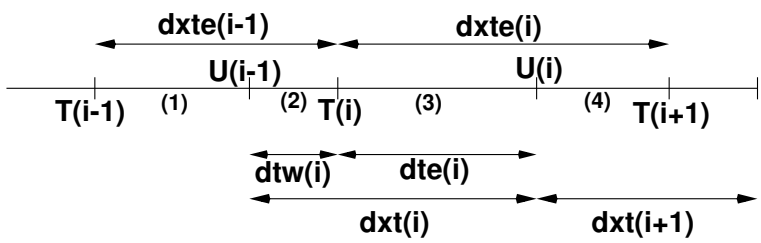

Figure 14.1: One dimensional zonal grid with quarter cells (1), (2), (3), (4). The quarter cells are bounded by their adjacent velocity and tracer points: quarter cell (1) is bounded by T_{i-1} and U_{i-1}, quarter cell (2) by U_{i-1} and T_i, quarter cell (3) by T_i and U_i, and quarter cell (4) by T_{i+1} and U_{i+1}. The dimension of a tracer cell is dxt_i, velocity points live on the boundary of tracer cells, the distance between tracer points T_i and T_{i+1} is $dxte_i$, the distance from the tracer point T_i to the western boundary of the tracer cell is dtw_i, and the distance to the eastern boundary is dte_i.

14.4 Elements of the discrete functional

This section details the basic elements of the discrete neutral diffusion functional. It is here that we consider issues regarding the three-dimensional lattice, including the bottom partial cells of Adcroft et al. (1997) and Pacanowski and Gnanadesikan (1998) commonly used in z-models to represent topography.

14.4.1 Grid partitioning for the small angle functional

Use of the small angle functional given by Equation (14.28) motivates a discretization which projects separately onto the two planar slices x-z and y-z. Discretization of the dianeutral

term $A_D(T_{,z})^2$ can be done using the one-dimensional warm-up formalism just discussed, and so is ignored in the following.*

Figure 14.2 shows the projection for the x-z plane. Within a plane, the central T-cell is partitioned into 4 generally non-equal squares, with one corner being the T-point and the other corners being the corners of the T-cell. Shown here are cells next to the partial cell bottom.

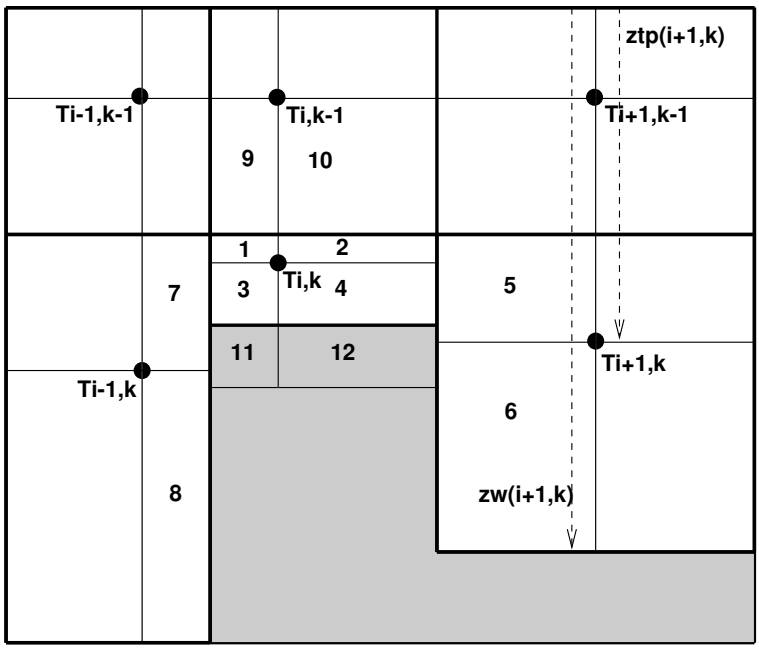

Figure 14.2: The x-z plane for the case with partial bottom cells. T-cells are surrounded by dark solid lines. Rock is shaded gray. Thin and thick solid lines define the quarter cells. The vertical position of a T-point within a T-cell maintains the same ratio for all the cells, whether full or partial. 12 quarter cells are shown, and these correspond to the cells where $T_{i.j,k}$ contributes to the functional. Note that quarter-cells 11 and 12 are in rock for this particular example. The vertical distances from the ocean surface to the bottom of the T-cells (zw) and to the T-cell center (ztp) are shown by lines with arrows. Note that the field $zw_{i,j,k}$ is actually one-dimensional in MOM4, but its full three-dimensional structure can be deduced from other information in the model.

14.4.2 Vertical grid dimensions with partial cells

The following fields define the thickness of partial cells

$$dht_{i,j,k} = zw_{i,j,k} - zw_{i,j,k-1} \tag{14.47}$$

$$dhwt_{i,j,k-1} = ztp_{i,j,k} - ztp_{i,j,k-1}. \tag{14.48}$$

*Although providing a useful discretization of the one-dimensional diffusion equation, the functional approach from Section 14.3 is not commonly employed in z-models for vertical diffusion. The reason is that vertical diffusivities are often computed on the cell-faces and are not considered equal across a quarter cell.

The field $zw_{i,j,k}$ defines the vertical distance from the ocean surface to the bottom of T-cell $T_{i,j,k}$. The field $ztp_{i,j,k}$ defines the vertical distance from the ocean surface to the T-cell point $T_{i,j,k}$. The grid dimension $dht_{i,j,k}$ measures the vertical thickness of the T-cell $T_{i,j,k}$, whereas $dhwt_{i,j,k-1}$ measures the vertical distance between the grid points $T_{i,j,k}$ and $T_{i,j,k-1}$.

Inspection of Figure 14.2 indicates that in order to construct the volume of the quarter cells, it is necessary to take vertical differences such as $zw_{i,j,k} - ztp_{i,j,k}$ (vertical dimension of quarter-cells 3 and 4), and $zt_{i,j,k} - zw_{i,k-1,j}$ (vertical dimension of quarter-cells 1 and 2). It will therefore prove useful to define the following field

$$\delta^{(i,j,k1)}_{(i,j,k2)} = (-1)^{k1-k2}(zw_{i,j,k1} - zt_{i,j,k2}). \tag{14.49}$$

Explicitly, for example,

$$\delta^{i,j,k-1+kr}_{i,j,k} = \begin{cases} (zt_{i,j,k} - zw_{i,j,k-1}) & \text{if } kr = 0 \\ (zw_{i,j,k} - zt_{i,j,k}) & \text{if } kr = 1 \end{cases} \tag{14.50}$$

and

$$\delta^{i,j,k}_{i,j,k+kr} = \begin{cases} (zw_{i,j,k} - zt_{i,j,k}) & \text{if } kr = 0 \\ (zt_{i,j,k+1} - zw_{i,j,k}) & \text{if } kr = 1. \end{cases} \tag{14.51}$$

Note that in Pacanowski and Griffies (1999), an extra multiplicative factor of 2 was used to facilitate easier comparison with full bottom cell results given in Griffies et al. (1998). This factor cancels out later and so is not exposed here.

14.4.3 Horizontal tracer gradients and the minimum thickness rule

With partial bottom cells, tracers in horizontally adjacent grid cells generally live at different depths. Before taking horizontal gradients between these tracers, linear interpolation is used to approximate the deeper tracer as if it actually lived at the depth of the shallower tracer point. For example, when taking a horizontal derivative between points $T_{i,j,k}$ and $T_{i+1,j,k}$ shown in Figure 14.2, the value of $T_{i+1,j,k}$ is approximated by a linear interpolation to the same depth as $T_{i,j,k}$. The method and motivation for this interpolation is described in Section 26.2.3 of Pacanowski and Griffies (1999) as well as the partial cell paper of Pacanowski and Gnanadesikan (1998).

The interpolation generally is an interpolation to the *minimum* depth between two laterally adjacent tracer cells. Hence, it effectively reduces the vertical spacing of quarter cell 5 to that of quarter cell 2. A similar consideration applies between quarter cells 7 and 1, for which the vertical spacing for quarter cell 7 is effectively the same as the smaller vertical spacing in quarter cell 1. With this prescription for determining the horizontal gradients and grid cell volumes, the discrete neutral diffusion scheme is compatible with the horizontal diffusion scheme described in Section 26.2.3 of Pacanowski and Griffies (1999) as well as the partial cell paper of Pacanowski and Gnanadesikan (1998).

In general, the above *minimum thickness rule* implies that it is sufficient to only explicitly denote the vertical height of the quarter cells living in the eastern and northern halves, say, of a particular T-cell. Such is the case for full vertical cells as well, in which the vertical height is a function only of the vertical grid position. Therefore, the only difference between the volume specification for the partial cells and the full cells is that for the partial cells, it is necessary to introduce the minimum operation.

With this *minimum thickness rule*, the vertical spacing relevant for defining the x-z plane quarter cell volumes is written

$$\Delta^{(i,j,k1)}_{(i,j,k2)} = \min\left(\delta^{(i,j,k1)}_{(i,j,k2)}, \delta^{(i+1,j,k1)}_{(i+1,j,k2)}\right). \tag{14.52}$$

Likewise, the vertical spacing relevant for defining the y-z plane quarter cell volumes is written

$$\Delta^{(i,j,k1)}_{(i,j,k2)} = \min\left(\delta^{(i,j,k1)}_{(i,j,k2)}, \delta^{(i,j+1,k1)}_{(i,j+1,k2)}\right). \tag{14.53}$$

Explicitly, the vertical spacings for the 12 x-z plane quarter cells in Figure 14.2 are given by (dropping the j label for brevity)

$$\begin{aligned}
\delta(1) &= (zt_{i,k} - zw_{i,k-1}) = \delta_{i,k}^{i,k-1} \\
\delta(2) &= (zt_{i,k} - zw_{i,k-1}) = \delta_{i,k}^{i,k-1} \\
\delta(3) &= (zw_{i,k} - zt_{i,k}) = \delta_{i,k}^{i,k} \\
\delta(4) &= (zw_{i,k} - zt_{i,k}) = \delta_{i,k}^{i,k} \\
\delta(5) &= (zt_{i+1,k} - zw_{i+1,k-1}) = \delta_{i+1,k}^{i+1,k-1} \\
\delta(6) &= (zw_{i+1,k} - zt_{i+1,k}) = \delta_{i+1,k}^{i+1,k} \\
\delta(7) &= (zt_{i-1,k} - zw_{i-1,k-1}) = \delta_{i-1,k}^{i-1,k-1} \\
\delta(8) &= (zw_{i-1,k} - zt_{i-1,k}) = \delta_{i-1,k}^{i-1,k} \\
\delta(9) &= (zw_{i,k-1} - zt_{i,k-1}) = \delta_{i,k-1}^{i,k-1} \\
\delta(10) &= (zw_{i,k-1} - zt_{i,k-1}) = \delta_{i,k-1}^{i,k-1} \\
\delta(11) &= (zt_{i,k+1} - zw_{i,k}) = \delta_{i,k+1}^{i,k} \\
\delta(12) &= (zt_{i,k+1} - zw_{i,k}) = \delta_{i,k+1}^{i,k},
\end{aligned} \tag{14.54}$$

whereas the vertical spacing used to compute the volumes of these quarter cells is given by

$$\begin{aligned}
\Delta(1) &= \min\left(\delta_{i-1,k}^{i-1,k-1}, \delta_{i,k}^{i,k-1}\right) = \Delta_{(i-1,k)}^{(i-1,k-1)} \\
\Delta(2) &= \min\left(\delta_{i,k}^{i,k-1}, \delta_{i+1,k}^{i+1,k-1}\right) = \Delta_{(i,k)}^{(i,k-1)} \\
\Delta(3) &= \min\left(\delta_{i-1,k}^{i-1,k}, \delta_{i,k}^{i,k}\right) = \Delta_{(i-1,k)}^{(i-1,k)} \\
\Delta(4) &= \min\left(\delta_{i,k}^{i,k}, \delta_{i+1,k}^{i+1,k}\right) = \Delta_{(i,k)}^{(i,k)} \\
\Delta(5) &= \min\left(\delta_{i,k}^{i,k-1}, \delta_{i+1,k}^{i+1,k-1}\right) = \Delta_{(i,k)}^{(i,k-1)} \\
\Delta(6) &= \min\left(\delta_{i,k}^{i,k}, \delta_{i+1,k}^{i+1,k}\right) = \Delta_{(i,k)}^{(i,k)} \\
\Delta(7) &= \min\left(\delta_{i-1,k}^{i-1,k-1}, \delta_{i,k}^{i,k-1}\right) = \Delta_{(i-1,k)}^{(i-1,k-1)} \\
\Delta(8) &= \min\left(\delta_{i-1,k}^{i-1,k}, \delta_{i,k}^{i,k}\right) = \Delta_{(i-1,k)}^{(i-1,k)} \\
\Delta(9) &= \min\left(\delta_{i-1,k-1}^{i-1,k-1}, \delta_{i,k-1}^{i,k-1}\right) = \Delta_{(i-1,k-1)}^{(i-1,k-1)} \\
\Delta(10) &= \min\left(\delta_{i,k-1}^{i,k-1}, \delta_{i+1,k-1}^{i+1,k-1}\right) = \Delta_{(i,k-1)}^{(i,k-1)} \\
\Delta(11) &= \min\left(\delta_{i-1,k+1}^{i-1,k}, \delta_{i,k+1}^{i,k}\right) = \Delta_{(i-1,k+1)}^{(i-1,k)} \\
\Delta(12) &= \min\left(\delta_{i,k+1}^{i,k}, \delta_{i+1,k+1}^{i+1,k}\right) = \Delta_{(i,k+1)}^{(i,k)}.
\end{aligned} \tag{14.55}$$

14.4.4 Quarter cell volumes

In order to discretize the functional, it is necessary to construct the volumes $V(n)$ of the various subcells which partition the grid into the quarter-cells shown in Figure 14.2. Referring

to Figure 14.2 yields volumes for the 12 quarter cells

$$
\begin{aligned}
V(1) &= dtw_{i,j}\, dyt_{i,j}\, \Delta_{(i-1,k)}^{(i-1,k-1)} \\
V(2) &= dte_{i,j}\, dyt_{i,j}\, \Delta_{(i,k)}^{(i,k-1)} \\
V(3) &= dtw_{i,j}\, dyt_{i,j}\, \Delta_{(i-1,k)}^{(i-1,k)} \\
V(4) &= dte_{i,j}\, dyt_{i,j}\, \Delta_{(i,k)}^{(i,k)} \\
V(5) &= dtw_{i+1,j}\, dyt_{i+1,j}\, \Delta_{(i,k)}^{(i,k-1)} \\
V(6) &= dtw_{i+1,j}\, dyt_{i+1,j}\, \Delta_{(i,k)}^{(i,k)} \\
V(7) &= dte_{i-1,j}\, dyt_{i-1,j}\, \Delta_{(i-1,k)}^{(i-1,k-1)} \\
V(8) &= dte_{i-1,j}\, dyt_{i-1,j}\, \Delta_{(i-1,k)}^{(i-1,k)} \\
V(9) &= dtw_{i,j}\, dyt_{i,j}\, \Delta_{(i-1,k-1)}^{(i-1,k-1)} \\
V(10) &= dte_{i,j}\, dyt_{i,j}\, \Delta_{(i,k-1)}^{(i,k-1)} \\
V(11) &= dtw_{i,j}\, dyt_{i,j}\, \Delta_{(i-1,k+1)}^{(i-1,k)} \\
V(12) &= dte_{i,j}\, dyt_{i,j}\, \Delta_{(i,k+1)}^{(i,k)}.
\end{aligned}
\tag{14.56}
$$

The j label has been omitted in the Δ expressions for purposes of brevity. The distance $dtw_{i,j}$ represents the distance from the point $T_{i,j,k}$ to the western boundary of the tracer cell, whereas $dte_{i,j}$ is the distance to the eastern boundary. All grid distances are in meters, and have the appropriate non-Euclidean metric or stretching functions absorbed according to the discussion in Section 19.12.4.

14.4.5 Tracer gradients within the quarter cells

Tracer and density gradients are required within the quarter cells. The second order accurate difference operators are given here. There are further details regarding the reference points to be used in determining the density gradient, and these reference point issues are discussed in Section 14.5.2. The difference operators approximating derivatives are written

$$
\begin{aligned}
\delta_x T_{i,j,k} &= \frac{T_{i+1,k,j} - T_{i,j,k}}{dxte_{i,j}} \\
\delta_y T_{i,j,k} &= \frac{T_{i,j+1,k} - T_{i,j,k}}{dytn_{i,j}} \\
\delta_z T_{i,j,k} &= \frac{T_{i,j,k} - T_{i,j,k+1}}{dhwt_{i,j,k}}.
\end{aligned}
\tag{14.57}
$$

The distance $dxte_{i,j}$ is the zonal distance between $T_{i+1,k,j}$ and $T_{i,j,k}$, the distance $dytn_{i,j}$ is the meridional distance between $T_{i,j+1,k}$ and $T_{i,j,k}$, and the distance $dhwt_{i,j,k}$ is the vertical distance between $T_{i,j,k}$ and $T_{i,j,k+1}$.

The difference operators are located at the T-cell faces and provide second order accurate approximations to the continuous derivatives. For partial cells, horizontal gradients are computed according to the minimum thickness rule and the associated linear interpolation described in Section 14.4.3. For brevity, such interpolation is assumed in the notation here; i.e., each horizontal derivative next to the bottom is computed using this interpolation algorithm.

Suppressing the j-index for brevity leads to the tracer gradient in the 12 quarter cells

$$
\begin{aligned}
\nabla T(1) &= \hat{\mathbf{x}}\,\delta_x T_{i-1,k} + \hat{\mathbf{y}}\,\delta_y T_{i,k} + \hat{\mathbf{z}}\,\delta_z T_{i,k-1} \\
\nabla T(2) &= \hat{\mathbf{x}}\,\delta_x T_{i,k} + \hat{\mathbf{y}}\,\delta_y T_{i,k} + \hat{\mathbf{z}}\,\delta_z T_{i,k-1} \\
\nabla T(3) &= \hat{\mathbf{x}}\,\delta_x T_{i-1,k} + \hat{\mathbf{y}}\,\delta_y T_{i,k} + \hat{\mathbf{z}}\,\delta_z T_{i,k} \\
\nabla T(4) &= \hat{\mathbf{x}}\,\delta_x T_{i,k} + \hat{\mathbf{y}}\,\delta_y T_{i,k} + \hat{\mathbf{z}}\,\delta_z T_{i,k} \\
\nabla T(5) &= \hat{\mathbf{x}}\,\delta_x T_{i,k} + \hat{\mathbf{y}}\,\delta_y T_{i+1,k} + \hat{\mathbf{z}}\,\delta_z T_{i+1,k-1} \\
\nabla T(6) &= \hat{\mathbf{x}}\,\delta_x T_{i,k} + \hat{\mathbf{y}}\,\delta_y T_{i+1,k} + \hat{\mathbf{z}}\,\delta_z T_{i+1,k} \\
\nabla T(7) &= \hat{\mathbf{x}}\,\delta_x T_{i-1,k} + \hat{\mathbf{y}}\,\delta_y T_{i-1,k} + \hat{\mathbf{z}}\,\delta_z T_{i-1,k-1} \\
\nabla T(8) &= \hat{\mathbf{x}}\,\delta_x T_{i-1,k} + \hat{\mathbf{y}}\,\delta_y T_{i-1,k} + \hat{\mathbf{z}}\,\delta_z T_{i-1,k} \\
\nabla T(9) &= \hat{\mathbf{x}}\,\delta_x T_{i-1,k-1} + \hat{\mathbf{y}}\,\delta_y T_{i,k-1} + \hat{\mathbf{z}}\,\delta_z T_{i,k-1} \\
\nabla T(10) &= \hat{\mathbf{x}}\,\delta_x T_{i,k-1} + \hat{\mathbf{y}}\,\delta_y T_{i,k-1} + \hat{\mathbf{z}}\,\delta_z T_{i,k-1} \\
\nabla T(11) &= \hat{\mathbf{x}}\,\delta_x T_{i-1,k+1} + \hat{\mathbf{y}}\,\delta_y T_{i,k+1} + \hat{\mathbf{z}}\,\delta_z T_{i,k} \\
\nabla T(12) &= \hat{\mathbf{x}}\,\delta_x T_{i,k+1} + \hat{\mathbf{y}}\,\delta_y T_{i,k+1} + \hat{\mathbf{z}}\,\delta_z T_{i,k}
\end{aligned}
\tag{14.58}
$$

and the functional derivative of these x-z plane gradients is given by

$$
\begin{aligned}
\frac{\partial \nabla T(1)}{\partial T_{i,j}} &= \frac{\hat{\mathbf{x}}}{dxte_{i-1,j}} - \frac{\hat{\mathbf{y}}}{dytn_{i,j}} - \frac{\hat{\mathbf{z}}}{dhwt_{i,j,k-1}} \\
\frac{\partial \nabla T(2)}{\partial T_{i,k}} &= -\frac{\hat{\mathbf{x}}}{dxte_{i,j}} - \frac{\hat{\mathbf{y}}}{dytn_{i,j}} - \frac{\hat{\mathbf{z}}}{dhwt_{i,k-1,j}} \\
\frac{\partial \nabla T(3)}{\partial T_{i,k}} &= \frac{\hat{\mathbf{x}}}{dxte_{i-1,j}} - \frac{\hat{\mathbf{y}}}{dytn_{i,j}} + \frac{\hat{\mathbf{z}}}{dhwt_{i,k,j}} \\
\frac{\partial \nabla T(4)}{\partial T_{i,k}} &= -\frac{\hat{\mathbf{x}}}{dxte_{i,j}} - \frac{\hat{\mathbf{y}}}{dytn_{i,j}} + \frac{\hat{\mathbf{z}}}{dhwt_{i,k,j}} \\
\frac{\partial \nabla T(5)}{\partial T_{i,k}} &= -\frac{\hat{\mathbf{x}}}{dxte_{i,j}} \\
\frac{\partial \nabla T(6)}{\partial T_{i,k}} &= -\frac{\hat{\mathbf{x}}}{dxte_{i,j}} \\
\frac{\partial \nabla T(7)}{\partial T_{i,k}} &= \frac{\hat{\mathbf{x}}}{dxte_{i-1,j}} \\
\frac{\partial \nabla T(8)}{\partial T_{i,k}} &= \frac{\hat{\mathbf{x}}}{dxte_{i-1,j}} \\
\frac{\partial \nabla T(9)}{\partial T_{i,k}} &= -\frac{\hat{\mathbf{z}}}{dhwt_{i,k-1,j}} \\
\frac{\partial \nabla T(10)}{\partial T_{i,k}} &= -\frac{\hat{\mathbf{z}}}{dhwt_{i,k-1,j}} \\
\frac{\partial \nabla T(11)}{\partial T_{i,k}} &= \frac{\hat{\mathbf{z}}}{dhwt_{i,k,j}} \\
\frac{\partial \nabla T(12)}{\partial T_{i,k}} &= \frac{\hat{\mathbf{z}}}{dhwt_{i,k,j}}.
\end{aligned}
\tag{14.59}
$$

14.4.6 Schematic form of the discretized functional

For each of the two vertical planes x-z and y-z, there are 12 components to the functional that contain contributions from the grid tracer value $T_{i,j,k}$. These 12 components correspond

to the 12 quarter cells shown in Figure 14.2. For example, the x-z plane functional can be written (see equation (14.28) for continuum version)

$$\mathcal{F}^{(x-z)} = -\frac{1}{2}\sum_{i,k}\sum_{n=1}^{12} A(n)\,V(n)\,[\delta_x T(n) + S_{(x)}(n)\,\delta_z T(n)]^2 \equiv \sum_{i,k} L_{i,k} \tag{14.60}$$

where the discretized quadratic form

$$L_{i,k} = -\frac{1}{2}\sum_{n=1}^{12} A(n)\,V(n)\,[\delta_x T(n) + S_{(x)}(n)\,\delta_z T(n)]^2 \equiv \sum_{n=1}^{12} L_{i,k}^{(n)} \tag{14.61}$$

consists of 12 non-positive contributions, with $A(n)$ the non-negative diffusion coefficient for each of the different quarter cells.

14.5 Triads and some notation

This section details the methods used to compute the density gradients and the corresponding neutral directions. In the process, we encounter the *triad* structure of the elements comprising the discrete functional.

14.5.1 Tracer and density triads

Consider the component of the x-z plane functional arising from the first subcell (see Figure 14.2), which has the form

$$L_{i,k}^{(1)} = -\left(\frac{A(1)\,V(1)}{2}\right)\left(\delta_x T_{i-1,k} - \delta_z T_{i,k-1}\,\frac{\delta_x \rho_{i-1,k}}{\delta_z \rho_{i,k-1}}\right)^2. \tag{14.62}$$

This term has contributions from the tracer and density at the three grid points: $T_{i-1,k}$, $T_{i,k}$, and $T_{i,k-1}$. These three points form a triangle, or *triad*, in the x-z plane. All pieces of the functional are discretized into these triad groups. That is, *triads are the fundamental combinations of tracer and density points within the discrete functional*. For each triad there is a unique quarter cell volume and diffusivity.

14.5.2 Reference points for computing density derivatives

Horizontal and vertical derivatives of density are needed to compute neutral slopes (Section 11.6.1). Hence, the density derivatives appearing in the discrete neutral diffusion functional are discretized as finite differences of locally referenced potential density. A self-consistent choice for the reference point must be chosen. That is, we need to choose a temperature, salinity, and pressure to reference the horizontal and vertical density derivatives. Given the prominence of the triad structure in the discrete functional, we choose the triad corner as reference point. Alternatives could be considered, but in practice they likely will lead to minimal differences from the triad corner. With the corner as the reference point, the triad forming $L_{i,k}^{(1)}$ uses the T-cell point $T_{i,j,k}$ as the reference. This choice is valid whether the triad is wholly in a full vertical cell region, or has a part within a partial cell. Other triads use similarly defined reference points.

With the triad corner as reference point, each of the three discrete densities appearing in a triad employ the *same* thermal and saline partial derivatives referenced to this single point. Therefore, the density derivatives appearing in $L_{i,k}^{(1)}$ are discretized as

$$\delta_x \rho_{i-1,j,k} = (\rho_{,\theta})_{i,j,k}\,\delta_x \theta_{i-1,j,k} + (\rho_{,s})_{i,j,k}\,\delta_x s_{i-1,j,k} \equiv \delta_x \rho_{i-1,j,k}^{(i,j,k)} \tag{14.63}$$

$$\delta_z \rho_{i,j,k-1} = (\rho_{,\theta})_{i,j,k}\,\delta_z \theta_{i,j,k-1} + (\rho_{,s})_{i,j,k}\,\delta_y s_{i,j,k-1} \equiv \delta_z \rho_{i,j,k-1}^{(i,j,k)}. \tag{14.64}$$

In these expressions, the thermal and saline partial derivatives are

$$(\rho_{,\theta})_{i,j,k} = (\partial\rho/\partial\theta)\left(\theta_{i,j,k}, s_{i,j,k}, p_{i,j,k}\right) \tag{14.65}$$

$$(\rho_{,s})_{i,j,k} = (\partial\rho/\partial s)\left(\theta_{i,j,k}, s_{i,j,k}, p_{i,j,k}\right), \tag{14.66}$$

where the arguments of the density partial derivatives indicate the value of the potential temperature, salinity, and pressure reference for use in their computation. The derivatives $\rho_{,\theta}$ and $\rho_{,s}$ are computed analytically and tabulated along with the equation of state (see Griffies et al. (2003)). The superscripts introduced on the density gradients in equations (14.63) and (14.64) allow for a compact notation that exposes information about the reference point. Parentheses are included to help distinguish it from subscripts referring to the usual grid points.

In an early derivation of the discretized diffusion operator, issues of reference points were ignored until the very end of the derivation. At that point, reasonable choices were made for choosing the reference points, which consisted of referencing on the various sides of the T-cells consistent with the location of the diffusive fluxes. However, it was soon realized that the numerical constraint of defining a dissipative diffusion operator was not satisfied by this choice. That is, such choices led to the possibility of a sign indefinite discrete neutral diffusion functional, and this allowed for the corresponding discrete neutral diffusion operator to increase variance. This point emphasizes the importance of addressing all discretization details when discretizing the functional in order to ensure a dissipative discrete neutral diffusion operator.

14.5.3 Notation for neutral direction slopes and diffusivities

The discretized neutral direction slopes appear throughout the functional, so it is useful to devise a notation to help reduce the clutter. For this purpose, let

$$Sx^{(i,k)}_{(i1,k1|i2,k2)} \equiv -\left(\frac{\delta_x \rho^{(i,k)}_{i1,k1}}{\delta_z \rho^{(i,k)}_{i2,k2}}\right) \tag{14.67}$$

Given this notation and the methods of regularization discussed in Section 13.3, it is also useful to introduce an associated diffusivity with notation

$$A(n) \to A^{(i,k)}_{(i1,k1|i2,k2)}. \tag{14.68}$$

The regularization of $A^{(i,k)}_{(i1,k1|i2,k2)}$ is determined by the value of $Sx^{(i,k)}_{(i1,k1|i2,k2)}$ according to one of the schemes discussed in Section 13.3.

14.5.4 Notation for the quarter-cell dimensions

In the derivative of the x-z functional, we write for the horizontal area of quarter cells living in adjacent tracer cells

$$(dtwe\, dyt)^{(ip)}_{i,j} = \begin{cases} dtw_{i+1,j}\, dyt_{i+1,j} & \text{if } ip = 1 \\ dte_{i,j}\, dyt_{i,j} & \text{if } ip = 0 \end{cases} \tag{14.69}$$

When considering the y-z functional, we require the analogous notation

$$(dxt\, dtsn)^{(jq)}_{i,j} = \begin{cases} dxt_{i,j+1}\, dts_{i,j+1} & \text{if } jq = 1 \\ dxt_{i,j}\, dtn_{i,j} & \text{if } jq = 0. \end{cases} \tag{14.70}$$

It is also useful to introduce the following notation for the zonal dimensions of quarter cells within a tracer cell

$$dtew_{i,j}^{(ip)} = \begin{cases} dte_{i,j} & \text{if } ip = 1 \\ dtw_{i,j} & \text{if } ip = 0 \end{cases} \tag{14.71}$$

and meridional dimensions

$$dtns_{i,j}^{(jq)} = \begin{cases} dtn_{i,j} & \text{if } jq = 1 \\ dts_{i,j} & \text{if } jq = 0. \end{cases} \tag{14.72}$$

14.6 The discrete diffusion operator

This section details the construction of the discrete diffusion operator as derived from the functional derivative of the discrete diffusion functional.

14.6.1 The discretized x-z dissipation functional

The different components of the neutral diffusion functional have now been considered, and so we now piece things together. We focus on the x-z piece, as y-z plane results are symmetric. The x-z plane quadratic form contains contributions from the T-point $T_{i,j,k}$ in the following 12 terms (j index is suppressed)

$$\begin{aligned}
L_{i,k}^{(1)} &= -\frac{1}{2}\, V(1)\, A_{(i-1,k|i,k-1)}^{(i,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i,k-1)}^{(i,k)}\, \delta_z T_{i,k-1}\right)^2 \\
L_{i,k}^{(2)} &= -\frac{1}{2}\, V(2)\, A_{(i,k|i,k-1)}^{(i,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i,k-1)}^{(i,k)}\, \delta_z T_{i,k-1}\right)^2 \\
L_{i,k}^{(3)} &= -\frac{1}{2}\, V(3)\, A_{(i-1,k|i,k)}^{(i,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i,k)}^{(i,k)}\, \delta_z T_{i,k}\right)^2 \\
L_{i,k}^{(4)} &= -\frac{1}{2}\, V(4)\, A_{(i,k|i,k)}^{(i,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i,k)}^{(i,k)}\, \delta_z T_{i,k}\right)^2 \\
L_{i,k}^{(5)} &= -\frac{1}{2}\, V(5)\, A_{(i,k|i+1,k-1)}^{(i+1,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i+1,k-1)}^{(i+1,k)}\, \delta_z T_{i+1,k-1}\right)^2 \\
L_{i,k}^{(6)} &= -\frac{1}{2}\, V(6)\, A_{(i,k|i+1,k)}^{(i+1,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i+1,k)}^{(i+1,k)}\, \delta_z T_{i+1,k}\right)^2 \\
L_{i,k}^{(7)} &= -\frac{1}{2}\, V(7)\, A_{(i-1,k|i-1,k-1)}^{(i-1,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i-1,k-1)}^{(i-1,k)}\, \delta_z T_{i-1,k-1}\right)^2 \\
L_{i,k}^{(8)} &= -\frac{1}{2}\, V(8)\, A_{(i-1,k|i-1,k)}^{(i-1,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i-1,k)}^{(i-1,k)}\, \delta_z T_{i-1,k}\right)^2 \\
L_{i,k}^{(9)} &= -\frac{1}{2}\, V(9)\, A_{(i-1,k-1|i,k-1)}^{(i,k-1)} \left(\delta_x T_{i-1,k-1} + Sx_{(i-1,k-1|i,k-1)}^{(i,k-1)}\, \delta_z T_{i,k-1}\right)^2 \\
L_{i,k}^{(10)} &= -\frac{1}{2}\, V(10)\, A_{(i,k-1|i,k-1)}^{(i,k-1)} \left(\delta_x T_{i,k-1} + Sx_{(i,k-1|i,k-1)}^{(i,k-1)}\, \delta_z T_{i,k-1}\right)^2 \\
L_{i,k}^{(11)} &= -\frac{1}{2}\, V(11)\, A_{(i-1,k+1|i,k)}^{(i,k+1)} \left(\delta_x T_{i-1,k+1} + Sx_{(i-1,k+1|i,k)}^{(i,k+1)}\, \delta_z T_{i,k}\right)^2 \\
L_{i,k}^{(12)} &= -\frac{1}{2}\, V(12)\, A_{(i,k+1|i,k)}^{(i,k+1)} \left(\delta_x T_{i,k+1} + Sx_{(i,k+1|i,k)}^{(i,k+1)}\, \delta_z T_{i,k}\right)^2 .
\end{aligned} \tag{14.73}$$

14.6.2 Derivative of the x-z dissipation functional

To construct the diffusion operator, it is necessary to take the discrete functional derivative of $\mathcal{F}$ according to equation (14.32). When taking this derivative, all terms in $\mathcal{F}$ independent of $T_{i,j,k}$ drop out. Hence, it is only necessary to consider that part of $\mathcal{F}$ including contributions from $T_{i,j,k}$. This fact motivated the focus on just the 12 quarter cells shown in Figure

14.2. Taking the derivative of the 12 contributions to the functional in the x-z plane yields

$$\frac{\partial L_{i,k}^{(1)}}{\partial T_{i,j,k}} = -V(1)\, A_{(i-1,k|i,k-1)}^{(i,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i,k-1)}^{(i,k)}\, \delta_z T_{i,k-1}\right) \left(\frac{1}{dxte_{i-1,j}} - \frac{Sx_{(i-1,k|i,k-1)}^{(i,k)}}{dhwt_{i,k-1}}\right)$$

$$\frac{\partial L_{i,k}^{(2)}}{\partial T_{i,j,k}} = -V(2)\, A_{(i,k|i,k-1)}^{(i,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i,k-1)}^{(i,k)}\, \delta_z T_{i,k-1}\right) \left(-\frac{1}{dxte_{i,j}} - \frac{Sx_{(i,k|i,k-1)}^{(i,k)}}{dhwt_{i,k-1}}\right)$$

$$\frac{\partial L_{i,k}^{(3)}}{\partial T_{i,j,k}} = -V(3)\, A_{(i-1,k|i,k)}^{(i,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i,k)}^{(i,k)}\, \delta_z T_{i,k}\right) \left(\frac{1}{dxte_{i-1,j}} + \frac{Sx_{(i-1,k|i,k)}^{(i,k)}}{dhwt_{i,k}}\right)$$

$$\frac{\partial L_{i,k}^{(4)}}{\partial T_{i,j,k}} = -V(4)\, A_{(i,k|i,k)}^{(i,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i,k)}^{(i,k)}\, \delta_z T_{i,k}\right) \left(-\frac{1}{dxte_{i,j}} + \frac{Sx_{(i,k|i,k)}^{(i,k)}}{dhwt_{i,k}}\right)$$

$$\frac{\partial L_{i,k}^{(5)}}{\partial T_{i,j,k}} = -V(5)\, A_{(i,k|i+1,k-1)}^{(i+1,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i+1,k-1)}^{(i+1,k)}\, \delta_z T_{i+1,k-1}\right) \left(-\frac{1}{dxte_{i,j}}\right)$$

$$\frac{\partial L_{i,k}^{(6)}}{\partial T_{i,j,k}} = -V(6)\, A_{(i,k|i+1,k)}^{(i+1,k)} \left(\delta_x T_{i,k} + Sx_{(i,k|i+1,k)}^{(i+1,k)}\, \delta_z T_{i+1,k}\right) \left(-\frac{1}{dxte_{i,j}}\right)$$

$$\frac{\partial L_{i,k}^{(7)}}{\partial T_{i,j,k}} = -V(7)\, A_{(i-1,k|i-1,k-1)}^{(i-1,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i-1,k-1)}^{(i-1,k)}\, \delta_z T_{i-1,k-1}\right) \left(\frac{1}{dxte_{i-1,j}}\right)$$

$$\frac{\partial L_{i,k}^{(8)}}{\partial T_{i,j,k}} = -V(8)\, A_{(i-1,k|i-1,k)}^{(i-1,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i-1,k)}^{(i-1,k)}\, \delta_z T_{i-1,k}\right) \left(\frac{1}{dxte_{i-1,j}}\right)$$

$$\frac{\partial L_{i,k}^{(9)}}{\partial T_{i,j,k}} = -V(9)\, A_{(i-1,k-1|i,k-1)}^{(i,k-1)} \left(\delta_x T_{i-1,k-1} + Sx_{(i-1,k-1|i,k-1)}^{(i,k-1)}\, \delta_z T_{i,k-1}\right) \left(-\frac{Sx_{(i-1,k-1|i,k-1)}^{(i,k-1)}}{dhwt_{i,k-1}}\right)$$

$$\frac{\partial L_{i,k}^{(10)}}{\partial T_{i,j,k}} = -V(10)\, A_{(i,k-1|i,k-1)}^{(i,k-1)} \left(\delta_x T_{i,k-1} + Sx_{(i,k-1|i,k-1)}^{(i,k-1)}\, \delta_z T_{i,k-1}\right) \left(-\frac{Sx_{(i,k-1|i,k-1)}^{(i,k-1)}}{dhwt_{i,k-1}}\right)$$

$$\frac{\partial L_{i,k}^{(11)}}{\partial T_{i,j,k}} = -V(11)\, A_{(i-1,k+1|i,k)}^{(i,k+1)} \left(\delta_x T_{i-1,k+1} + Sx_{(i-1,k+1|i,k)}^{(i,k+1)}\, \delta_z T_{i,k}\right) \left(\frac{Sx_{(i-1,k+1|i,k)}^{(i,k+1)}}{dhwt_{i,k}}\right)$$

$$\frac{\partial L_{i,k}^{(12)}}{\partial T_{i,j,k}} = -V(12)\, A_{(i,k+1|i,k)}^{(i,k+1)} \left(\delta_x T_{i,k+1} + Sx_{(i,k+1|i,k)}^{(i,k+1)}\, \delta_z T_{i,k}\right) \left(\frac{Sx_{(i,k+1|i,k)}^{(i,k+1)}}{dhwt_{i,k}}\right)$$

To organize results more compactly, rearrange terms to identify components to the discrete diffusive flux. For this purpose, consider the three terms from quarter cells $1 + 7 + 9$ and

insert expressions for the volumes given in Section 14.4.4:

$$
\begin{aligned}
&\frac{1}{\partial T_{i,j,k}}\left(L_{i,k}^{(1)}+L_{i,k}^{(7)}+L_{i,k}^{(9)}\right)=-\left(\frac{\Delta_{(i-1,k)}^{(i-1,k-1)}}{dxte_{i-1,j}}\right)\times\\
&\Big(A_{(i-1,k|i,k-1)}^{(i,k)}\,dtw_{i,j}\,dyt_{i,j}\,(\delta_x T_{i-1,k}+Sx_{(i-1,k|i,k-1)}^{(i,k)}\,\delta_z T_{i,k-1})\\
&+A_{(i-1,k|i-1,k-1)}^{(i-1,k)}\,dte_{i-1,j}\,dyt_{i-1,j}\,(\delta_x T_{i-1,k}+Sx_{(i-1,k|i-1,k-1)}^{(i-1,k)}\,\delta_z T_{i-1,k-1})\Big)\\
&+\left(\frac{dtw_{i,j}\,dyt_{i,j}}{dhwt_{i,k-1}}\right)\times\\
&\Big(\Delta_{(i-1,k)}^{(i-1,k-1)}\,(A\ Sx)_{(i-1,k|i,k-1)}^{(i,k)}(\delta_x T_{i-1,k}+Sx_{(i-1,k|i,k-1)}^{(i,k)}\,\delta_z T_{i,k-1})\\
&+\Delta_{(i-1,k-1)}^{(i-1,k-1)}\,(A\ Sx)_{(i-1,k-1|i,k-1)}^{(i,k-1)}(\delta_x T_{i-1,k-1}+Sx_{(i-1,k-1|i,k-1)}^{(i,k-1)}\,\delta_z T_{i,k-1})\Big)
\end{aligned}
\tag{14.74}
$$

To clean up this expression, introduce notation for the quarter-cell area from Section 14.5.4 to render

$$
\begin{aligned}
&\frac{1}{\partial T_{i,j,k}}\left(L_{i,k}^{(1)}+L_{i,k}^{(7)}+L_{i,k}^{(9)}\right)=-\left(\frac{\Delta_{(i-1,k)}^{(i-1,k-1)}}{dxte_{i-1,j}}\right)\sum_{ip=0}^{1}(dtwe\,dyt)_{(i-1,j)}^{(ip)}\,A_{(i-1,k|i-1+ip,k-1)}^{(i-1+ip,k)}\times\\
&\left(\delta_x T_{i-1,k}+Sx_{(i-1,k|i-1+ip,k-1)}^{(i-1+ip,k)}\,\delta_z T_{i-1+ip,k-1}\right)\\
&+\left(\frac{dtw_{i,j}\,dyt_{i,j}}{dhwt_{i,k-1}}\right)\sum_{kr=0}^{1}\Delta_{(i-1,k-1+kr)}^{(i-1,k-1)}\,(A\ Sx)_{(i-1,k-1+kr|i,k-1)}^{(i,k-1+kr)}\times\\
&\left(\delta_x T_{i-1,k-1+kr}+Sx_{(i-1,k-1+kr|i,k-1)}^{(i,k-1+kr)}\,\delta_z T_{i,k-1})\right).
\end{aligned}
\tag{14.75}
$$

Results for the other three combinations of three quarter cells in the x-z plane follow similarly. Thus,

$$
\begin{aligned}
&\frac{1}{\partial T_{i,j,k}}\left(L_{i,k}^{(2)}+L_{i,k}^{(5)}+L_{i,k}^{(10)}\right)=\left(\frac{\Delta_{(i,k)}^{(i,k-1)}}{dxte_{i,j}}\right)\sum_{ip=0}^{1}(dtwe\,dyt)_{i,j}^{(ip)}\,A_{(i,k|i+ip,k-1)}^{(i+ip,k)}\times\\
&\left(\delta_x T_{i,k}+Sx_{(i,k|i+ip,k-1)}^{(i+ip,k)}\,\delta_z T_{i+ip,k-1}\right)\\
&+\left(\frac{dte_{i,j}\,dyt_{i,j}}{dhwt_{i,k-1}}\right)\sum_{kr=0}^{1}\Delta_{(i,k-1+kr)}^{(i,k-1)}\,(A\ Sx)_{(i,k-1+kr|i,k-1)}^{(i,k-1+kr)}\times\\
&\left(\delta_x T_{i,k-1+kr}+Sx_{(i,k-1+kr|i,k-1)}^{(i,k-1+kr)}\,\delta_z T_{i,k-1})\right)
\end{aligned}
\tag{14.76}
$$

$$
\begin{aligned}
&\frac{1}{\partial T_{i,j,k}}\left(L_{i,k}^{(3)}+L_{i,k}^{(8)}+L_{i,k}^{(11)}\right)=-\left(\frac{\Delta_{(i-1,k)}^{(i-1,k)}}{dxte_{i-1,j}}\right)\sum_{ip=0}^{1}(dtwe\,dyt)_{i-1,j}^{(ip)}\,A_{(i-1,k|i-1+ip,k)}^{(i-1+ip,k)}\times\\
&\left(\delta_x T_{i-1,k}+Sx_{(i-1,k|i-1+ip,k)}^{(i-1+ip,k)}\,\delta_z T_{i-1+ip,k}\right)\\
&-\left(\frac{dtw_{i,j}\,dyt_{i,j}}{dhwt_{i,k}}\right)\sum_{kr=0}^{1}\Delta_{(i-1,k+kr)}^{(i-1,k)}\,(A\;Sx)_{(i-1,k+kr|i,k)}^{(i,k+kr)}\times\\
&\left(\delta_x T_{i-1,k+kr}+Sx_{(i-1,k+kr|i,k)}^{(i,k+kr)}\,\delta_z T_{i,k}\right)
\end{aligned}
\tag{14.77}
$$

and finally

$$
\begin{aligned}
&\frac{1}{\partial T_{i,j,k}}\left(L_{i,k}^{(4)}+L_{i,k}^{(6)}+L_{i,k}^{(12)}\right)=\left(\frac{\Delta_{(i,k)}^{(i,k)}}{dxte_{i,j}}\right)\sum_{ip=0}^{1}(dtwe\,dyt)_{i,j}^{(ip)}\,A_{(i,k|i+ip,k)}^{(i+ip,k)}\times\\
&\left(\delta_x T_{i,k}+\delta_z T_{i+ip,k}\,Sx_{(i,k|i+ip,k)}^{(i+ip,k)}\right)\\
&-\left(\frac{dte_{i,j}\,dyt_{i,j}}{dhwt_{i,k}}\right)\sum_{kr=0}^{1}\Delta_{(i,k+kr)}^{(i,k)}\,(A\;Sx)_{(i,k+kr|i,k)}^{(i,k+kr)}\times\\
&\left(\delta_x T_{i,k+kr}+Sx_{(i,k+kr|i,k)}^{(i,k+kr)}\,\delta_z T_{i,k}\right).
\end{aligned}
\tag{14.78}
$$

Bringing the 12 pieces together leads to the functional derivative

$$
\begin{aligned}
&\frac{\partial\mathcal{F}^{(x-z)}}{\partial T_{i,j,k}}=\\
&\left(\frac{1}{dxte_{i,j}}\right)\sum_{kr=0}^{1}\Delta_{(i,k)}^{(i,k-1+kr)}\sum_{ip=0}^{1}(dtwe\,dyt)_{i,j}^{(ip)}\,A_{(i,k|i+ip,k-1+kr)}^{(i+ip,k)}\times\\
&\left(\delta_x T_{i,k}+Sx_{(i,k|i+ip,k-1+kr)}^{(i+ip,k)}\,\delta_z T_{i+ip,k-1+kr}\right)\\
&-\left(\frac{1}{dxte_{i-1,j}}\right)\sum_{kr=0}^{1}\Delta_{(i-1,k)}^{(i-1,k-1+kr)}\sum_{ip=0}^{1}(dtwe\,dyt)_{i-1,j}^{(ip)}\,A_{(i-1,k|i-1+ip,k-1+kr)}^{(i-1+ip,k)}\times\\
&\left(\delta_x T_{i-1,k}+Sx_{(i-1,k|i-1+ip,k-1+kr)}^{(i-1+ip,k)}\,\delta_z T_{i-1+ip,k-1+kr}\right)\\
&+\left(\frac{dyt_{i,j}}{dhwt_{i,k-1}}\right)\sum_{ip=0}^{1}dtwe_{i,j}^{(ip)}\sum_{kr=0}^{1}\Delta_{(i-1+ip,k-1+kr)}^{(i-1+ip,k-1)}\times\\
&(A\;Sx)_{(i-1+ip,k-1+kr|i,k-1)}^{(i,k-1+kr)}\left(\delta_x T_{i-1+ip,k-1+kr}+Sx_{(i-1+ip,k-1+kr|i,k-1)}^{(i,k-1+kr)}\,\delta_z T_{i,k-1}\right)\\
&-\left(\frac{dyt_{i,j}}{dhwt_{i,k}}\right)\sum_{ip=0}^{1}dtwe_{i,j}^{(ip)}\sum_{kr=0}^{1}\Delta_{(i-1+ip,k+kr)}^{(i-1+ip,k)}\\
&(A\;Sx)_{(i-1+ip,k+kr|i,k)}^{(i,k+kr)}\left(\delta_x T_{i-1+ip,k+kr}+Sx_{(i-1+ip,k+kr|i,k)}^{(i,k+kr)}\,\delta_z T_{i,k}\right)
\end{aligned}
$$

14.6.3 Discretized diffusion operator

Based on manipulations in the previous subsection, and generalizations to three-dimensional geometry, we identify the discretized neutral diffusion operator

$$
\begin{aligned}
R[T_{i,j,k}] &= \frac{1}{V_{T_{i,j,k}}} \frac{\partial \mathcal{F}[T_{i,j,k}]}{\partial T_{i,j,k}} \\
&= \left(\frac{1}{dyt_{i,j}\, dht_{i,j,k}} \right) \delta_x \left(\frac{1}{dxte_{i-1,j}} \sum_{kr=0}^{1} \Delta_{(i-1,k)}^{(i-1,k-1+kr)} \sum_{ip=0}^{1} (dtwe\, dyt)_{i-1,j}^{(ip)} \right. \\
&\quad \left. A_{(i-1,k|i-1+ip,k-1+kr)}^{(i-1+ip,k)} \left(\delta_x T_{i-1,k} + Sx_{(i-1,k|i-1+ip,k-1+kr)}^{(i-1+ip,k)} \delta_z T_{i-1+ip,k-1+kr} \right) \right) \\
&\quad + \left(\frac{1}{dxt_{i,j}\, dht_{i,j,k}} \right) \delta_y \left(\frac{1}{dytn_{i,j-1}} \sum_{kr=0}^{1} \Delta_{(j-1,k)}^{(j-1,k-1+kr)} \sum_{jq=0}^{1} (dxt\, dtwe)_{i,j-1}^{(jq)} \right. \\
&\quad \left. A_{(j-1,k|j-1+jq,k-1+kr)}^{(j-1+jq,k)} \left(\delta_y T_{j-1,k} + Sy_{(j-1,k|j-1+jq,k-1+kr)}^{(j-1+jq,k)} \delta_z T_{j-1+jq,k-1+kr} \right) \right) \\
&\quad + \left(\frac{1}{dxt_{i,j}} \right) \delta_z \left(\frac{1}{dhwt_{i,k-1}} \sum_{ip=0}^{1} (dtew)_{i,j}^{(ip)} \sum_{kr=0}^{1} \Delta_{(i-1+ip,k-1+kr)}^{(i-1+ip,k-1)} (A\, Sx)_{(i-1+ip,k-1+kr|i,k-1)}^{(i,k-1+kr)} \right. \\
&\quad \left. \left(\delta_x T_{i-1+ip,k-1+kr} + Sx_{(i-1+ip,k-1+kr|i,k-1)}^{(i,k-1+kr)} \delta_z T_{i,k-1} \right) \right) \\
&\quad + \left(\frac{1}{dyt_{i,j}} \right) \delta_z \left(\frac{1}{dhwt_{i,k-1}} \sum_{jq=0}^{1} (dtns)_{i,j}^{(jq)} \sum_{kr=0}^{1} \Delta_{(j-1+jq,k-1+kr)}^{(j-1+jq,k-1)} (A\, Sy)_{(j-1+jq,k-1+kr|j,k-1)}^{(j,k-1+kr)} \right. \\
&\quad \left. \left(\delta_y T_{j-1+jq,k-1+kr} + Sy_{(j-1+jq,k-1+kr|j,k-1)}^{(j,k-1+kr)} \delta_z T_{j,k-1} \right) \right) .
\end{aligned}
\tag{14.79}
$$

14.7 Diffusive flux components

Although equation (14.79) provides the desired discretized diffusion operator, it is useful to identify the three diffusive flux components since ocean models are typically coded in terms of flux components defined at the tracer cell faces. Implementing no-flux boundary conditions is also easier when identifying fluxes.

14.7.1 Continuum considerations

Given generalized horizontal coordinates and partial cells, we require some care when defining the flux components. For this purpose, recall that in the continuum the diffusion operator is

$$R(T) = -\nabla \cdot \mathbf{F} \tag{14.80}$$

where

$$\mathbf{F} = \hat{\mathbf{x}}\, F^{(x)} + \hat{\mathbf{y}}\, F^{(y)} + \hat{\mathbf{z}}\, F^{(z)} \tag{14.81}$$

is the flux vector expressed in terms of physical vector components (see Section 19.12.3 for a discussion of physical tensor components). In generalized orthogonal horizontal coordinates, Section 19.11.4 notes that the divergence of a vector is written

$$\nabla \cdot \mathbf{F} = (dy)^{-1} \left(dy\, F^{(x)} \right)_{,x} + (dx)^{-1} \left(dx\, F^{(y)} \right)_{,y} + F_{,z}^{(z)}, \tag{14.82}$$

where the squared line element is given by

$$(ds)^2 = (h_1\, d\xi^1)^2 + (h_2\, d\xi^2)^2 + (dz)^2 = (dx)^2 + (dy)^2 + (dz)^2, \tag{14.83}$$

with h_1 and h_2 the stretching functions and (ξ^1, ξ^2) the generalized orthogonal coordinates. For spherical longitude-latitude coordinates $(\xi^1, \xi^2) = (\lambda, \phi)$, the stretching functions are $(h_1, h_2) = (R, R\,\cos\phi)$ where R is the earth's radius. For more on generalized orthogonal coordinates, see Section or 19.12.2 for a summary, or the discussions in Chapters 18 and 19 for a complete treatment.

When the vertical grid spacing is a function of horizontal and vertical position, as for the case with partial bottom cells, the divergence takes the more general form

$$\begin{aligned}\nabla\cdot\mathbf{F} &= (dy\,dz)^{-1}\,(dy\,dz\,F^{(x)})_{,x} + (dx\,dz)^{-1}\,(dx\,dz\,F^{(y)})_{,y} + F^{(z)}_{,z}\\ &\equiv (dy\,dz)^{-1}\,(F^{(x)}_{area})_{,x} + (dx\,dz)^{-1}\,(F^{(y)}_{area})_{,y} + F^{(z)}_{,z}.\end{aligned} \tag{14.84}$$

The area weighted flux components $F^{(x)}_{area}$ and $F^{(y)}_{area}$ have units $(L^2 \times \text{tracer} \times L/t)$.

14.7.2 Discrete fluxes

Given the continuum notions, we can identify the discrete versions of the diffusive flux. The flux components can be identified directly from equations (14.79) and (14.84)

$$R[T]_{i,j,k} = -\left(\frac{\delta_x F^{(x)}_{i-1,j,k}}{dyt_{i,j}\,dht_{i,j,k}} + \frac{\delta_y F^{(y)}_{i,j-1,k}}{dxt_{i,j}\,dht_{i,j,k}} + \delta_z F^{(z)}_{i,k-1,j}\right), \tag{14.85}$$

where the subscript "*area*" was dropped from the horizontal fluxes. The area weighted zonal flux component defined at the east face of T-cell $T_{i,j,k}$ is

$$\begin{aligned}-F^{(x)}_{i,j,k} = \frac{1}{dxte_{i,j}} \sum_{kr=0}^{1} \Delta^{(i,k-1+kr)}_{(i,k)} \sum_{ip=0}^{1} (dtwe\,dyt)^{(ip)}_{i,j} \times\\ A^{(i+ip,k)}_{(i,k|i+ip,k-1+kr)} \left(\delta_x T_{i,k} + Sx^{(i+ip,k)}_{(i,k|i+ip,k-1+kr)}\,\delta_z T_{i+ip,k-1+kr}\right).\end{aligned} \tag{14.86}$$

The area weighted meridional flux component defined at the north face of T-cell $T_{i,j,k}$ is

$$\begin{aligned}-F^{(y)}_{i,j,k} = \frac{1}{dytn_{i,j}} \sum_{kr=0}^{1} \Delta^{(j,k-1+kr)}_{(j,k)} \sum_{jq=0}^{1} (dxt\,dtsn)^{(jq)}_{i,j} \times\\ A^{(j+jq,k)}_{(j,k|j+jq,k-1+kr)} \left(\delta_y T_{j,k} + Sy^{(j+jq,k)}_{(j,k|j+jq,k-1+kr)}\,\delta_z T_{j+jq,k-1+kr}\right).\end{aligned} \tag{14.87}$$

The vertical flux component defined at the bottom face of T-cell $T_{i,j,k}$ is

$$\begin{aligned}-F^{(z)}_{i,j,k} = &\left(\frac{1}{dxt_{i,j}\,dhwt_{i,j,k}}\right) \sum_{ip=0}^{1} dtew^{(ip)}_{i,j} \sum_{kr=0}^{1} \Delta^{(i-1+ip,k)}_{(i-1+ip,k+kr)} (A\,Sx)^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)} \times\\ &\left(\delta_x T_{i-1+ip,k+kr} + Sx^{(i,k+kr)}_{(i-1+ip,k+kr|i,k-1)}\,\delta_z T_{i,k}\right)\\ &+ \left(\frac{1}{dyt_{i,j}\,dhwt_{i,j,k}}\right) \sum_{jq=0}^{1} dtns^{(jq)}_{i,j} \sum_{kr=0}^{1} \Delta^{(j-1+jq,k)}_{(j-1+jq,k+kr)} (A\,Sy)^{(j,k+kr)}_{(j-1+jq,k+kr|j,k)} \times\\ &\left(\delta_y T_{j-1+jq,k+kr} + Sy^{(j,k+kr)}_{(j-1+jq,k+kr|j,k)}\,\delta_z T_{j,k}\right).\end{aligned} \tag{14.88}$$

14.7.3 Identifying diagonal components to the diffusion tensor

The off-diagonal terms of the small angle diffusion tensor (12.20) are not separable from the tracers. However, the diagonal terms are separable. It is useful to explicitly identify the non-negative diagonal elements since they can be computed once per time step and used repeatedly for all tracers.

From the area weighted zonal flux component (14.86), we identify the area weighted diagonal tensor component

$$K^{11}_{i,j,k} = \frac{1}{dxte_{i,j}} \sum_{kr=0}^{1} \Delta^{(i,k-1+kr)}_{(i,k)} \sum_{ip=0}^{1} (dtwe\, dyt)^{(ip)}_{i,j} A^{(i+ip,k)}_{(i,k|i+ip,k-1+kr)} \tag{14.89}$$

which then leads to the form for the area weighted zonal flux component

$$\begin{aligned} -F^{x}_{i,j,k} = K^{11}_{i,j,k}\, \delta_x T_{i,j,k} + \frac{1}{dxte_{i,j}} \sum_{kr=0}^{1} \Delta^{(i,k-1+kr)}_{(i,k)} \times \\ \sum_{ip=0}^{1} (dtwe\, dyt)^{(ip)}_{i,j}\, (Sx\, A)^{(i+ip,k)}_{(i,k|i+ip,k-1+kr)}\, \delta_z T_{i+ip,k-1+kr}. \end{aligned} \tag{14.90}$$

Likewise, from the area weighted meridional flux component (14.87), we have the area weighted diagonal tensor component

$$K^{22}_{i,j,k} = \frac{1}{dytn_{i,j}} \sum_{kr=0}^{1} \Delta^{(j,k-1+kr)}_{(j,k)} \sum_{jq=0}^{1} (dxt\, dtsn)^{(jq)}_{i,j} A^{(j+jq,k)}_{(j,k|j+jq,k-1+kr)} \tag{14.91}$$

which then leads to the area weighted meridional flux component

$$\begin{aligned} -F^{y}_{i,j,k} = K^{22}_{i,j,k}\, \delta_y T_{i,j,k} + \frac{1}{dytn_{i,j}} \sum_{kr=0}^{1} \Delta^{(j,k-1+kr)}_{(j,k)} \times \\ \sum_{jq=0}^{1} (dxt\, dtsn)^{(jq)}_{i,j}\, (Sy\, A)^{(j+jq,k)}_{(j,k|j+jq,k-1+kr)}\, \delta_z T_{j+jq,k-1+kr} \end{aligned} \tag{14.92}$$

From the vertical flux component (14.88), we have

$$\begin{aligned} K^{33}_{i,j,k} = \left(\frac{1}{dxt_{i,j}\, dhwt_{i,j,k}}\right) \sum_{ip=0}^{1} (dtew)^{(ip)}_{i,j} \sum_{kr=0}^{1} \Delta^{(i-1+ip,k)}_{(i-1+ip,k+kr)} ((Sx)^2\, A)^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)} \\ + \left(\frac{1}{dyt_{i,j}\, dhwt_{i,j,k}}\right) \sum_{jq=0}^{1} (dtns)^{(jq)}_{i,j} \times \sum_{kr=0}^{1} \Delta^{(j-1+jq,k)}_{(j-1+jq,k+kr)} ((Sy)^2 A)^{(j,k+kr)}_{(j-1+jq,k+kr|j,k)} \end{aligned} \tag{14.93}$$

which then leads to the form for the vertical flux component

$$\begin{aligned} &- F^{z}_{i,j,k} = K^{33}_{i,j,k}\, \delta_z T_{i,j,k} \\ &+ \left(\frac{1}{dxt_{i,j}\, dhwt_{i,j,k}}\right) \sum_{ip=0}^{1} (dtew)^{(ip)}_{i,j} \sum_{kr=0}^{1} \Delta^{(i-1+ip,k)}_{(i-1+ip,k+kr)} (A\, Sx)^{(i,k+kr)}_{(i-1+ip,k+kr|i,k)}\, \delta_x T_{i-1+ip,k+kr} \\ &+ \left(\frac{1}{dyt_{i,j}\, dhwt_{i,j,k}}\right) \sum_{jq=0}^{1} (dtns)^{(jq)}_{i,j} \sum_{kr=0}^{1} \Delta^{(j-1+jq,k)}_{(j-1+jq,k+kr)} (A\, Sy)^{(j,k+kr)}_{(j-1+jq,k+kr|j,k)}\, \delta_y T_{j-1+jq,k+kr}. \end{aligned} \tag{14.94}$$

For the discussion of Section 14.8.3, we refer to the $K^{33}\, \delta_z T_{i,j,k}$ term as the *diagonal* portion of the vertical flux component, and the remaining terms represent the *off-diagonal* portion.

For diagnostic purposes, it is often useful to map the diagonal tensor components K^{11}, K^{22}, K^{33}. As defined above, K^{11} and K^{22} represent area weighted versions of their continuum analogs. However, it is relevant to map a field with dimensions of diffusivity in order to compare with the input diffusivity A. For this purpose, one may wish to map

$$\frac{K^{11}_{i,j,k}}{dyt_{i,j}\, dht_{i,j,k}} \tag{14.95}$$

$$\frac{K^{22}_{i,j,k}}{dxt_{i,j}\, dht_{i,j,k}} \tag{14.96}$$

whose dimensions are those of a diffusivity. On a uniform grid in Cartesian space using full vertical cells and neutral slopes less than the critical slope, these fields reduce to the value of the neutral diffusivity A. On a sphere using spherical coordinates, full vertical cells, and slopes less than the critical slope, K^{22} is generally less than A, whereas K^{11} equals A. On a sphere with generalized orthogonal coordinates and/or next to a partial cell bottom, the reduction generally does not occur for either field. This lack of reduction to the input diffusivity is not a problem with the diffusion algorithm, rather it is a limitation of the diagnostic.

14.7.4 Stencils for small angle flux components

A figure is useful to garner insight into the stencil deduced from the functional discretization. Figure 14.3 provides such a stencil for the zonal component to the small angle neutral flux $F^x_{i,k}$. Each triad is weighted by the smallest vertical distance consistent with the *minimum thickness rule* discussed in Section 14.4.3. A similar stencil holds for the meridional component to the neutral diffusive flux.

For the vertical component (Figure 14.4) to the diffusive flux, a set of four triads are used, each of which is rotated by 90 degrees relative to the triads shown for the zonal component to the flux. The weighting for each triad is a combination of the zonal grid spacing and a ratio of the *minimum thickness rule* spacing, normalized by the relevant vertical T-cell distance. As a result of the rotation, the Δ weighting for each of the four triads is generally different.

14.8 Some further issues of numerical implementation

This section presents some general comments and details regarding the implementation of the neutral diffusion scheme.

14.8.1 Gent-McWilliams skew-flux plus neutral diffusive flux

As noted in Section 12.2.2, discretization of the small angle neutral diffusion tensor also provides a discretization for the Gent-McWilliams skew diffusion process. All that is needed is to make appropriate changes to the diffusivities according to the discussions in Chapter 13.

14.8.2 Discretization of the GM velocity for diagnostics

It is sometimes useful to compute the eddy-induced velocity (equation 12.56)) for diagnostic purposes. In early realizations of the Gent and McWilliams (1990) scheme, such diagnostics were used to compute the meridional-vertical transport streamfunction. However, a simpler and more direct method is available as described in Section 7.5.3.

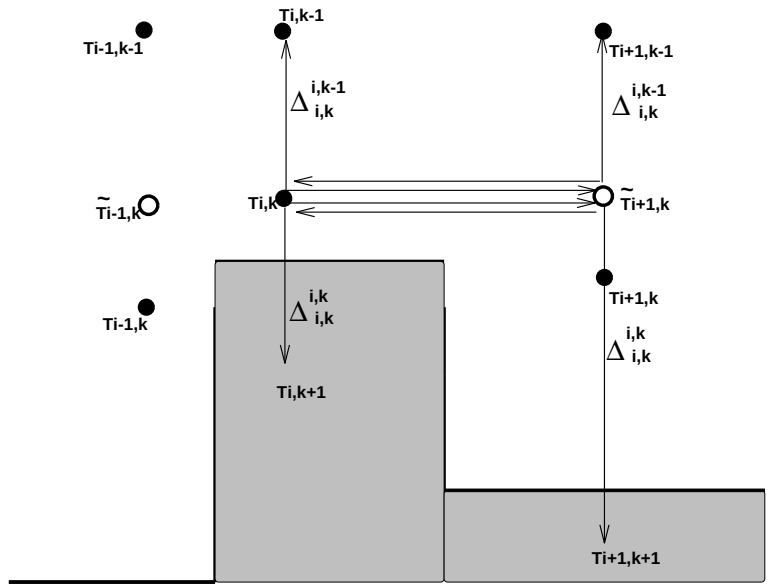

Figure 14.3: Stencil for computing the zonal component to the neutral diffusive flux $F^x_{i,k}$ (equation (14.86)). This flux component is located at the east face of T-cell $T_{i,k}$. Shown are four density triads, each weighted by the relevant vertical spacing factor Δ. For the two triads extending upwards, the weighting is $\Delta^{(i,k-1)}_{(i,k)}$, whereas the two triads extending downwards are weighted by $\Delta^{(i,k)}_{(i,k)}$. The tracer points with the open circles denote tracer values which are determined through linear interpolation. For the two triads extending into rock (shaded regions), they are zeroed out in the code since they should not contribute to the diffusion operator.

The following discussion illustrates one choice for discretizing the three-dimensional GM eddy-induced velocity. Additionally, it highlights the ambiguity inherent in the discretization, thus further motivating the skew-flux approach. That is, q discretization of GM skewsion relies on the functional formalism of Section 14.1. In contrast, we have found no general principle to guide discretization of GM advection. More than one approach has been tried in the literature (e.g., see Danabasoglu and McWilliams (1995) and Weaver and Eby (1997)). The form chosen here uses unbroken triad-based neutral slopes. In particular, averaging operators are not applied to the numerator and denominator of the discretized form of $\nabla_z \rho / \rho_{,z}$. Use of unbroken triads means that the neutral slopes computed for the diffusion calculation can be employed for diagnosis of the eddy-induced velocity.

For computing the zonal eddy-induced velocity, defined at the zonal face of the tracer cell $T_{i,j,k}$, we take the difference between diffusivity weighted slopes averaged so that they live at the bottom-zonal corners of a tracer cell. Thus, we define the following diffusivity weighted slope

$$4\,(\kappa\, Sx)_{i,k} = \\ \kappa_{i,k}\, Sx^{(i,k)}_{(i,k|i,k)} + \kappa_{i+1,k}\, Sx^{(i+1,k)}_{(i,k|i+1,k)} + \kappa_{i,k+1}\, Sx^{(i,k+1)}_{(i,k+1|i,k)} + \kappa_{i+1,k+1}\, Sx^{(i+1,k+1)}_{(i,k+1|i+1,k)}, \quad (14.97)$$

which represents a zonal and vertical average of the four diffusivity weighted triad-slopes surrounding the bottom-zonal corner of the tracer cell $T_{i,k-1}$. Note the irrelevant meridional

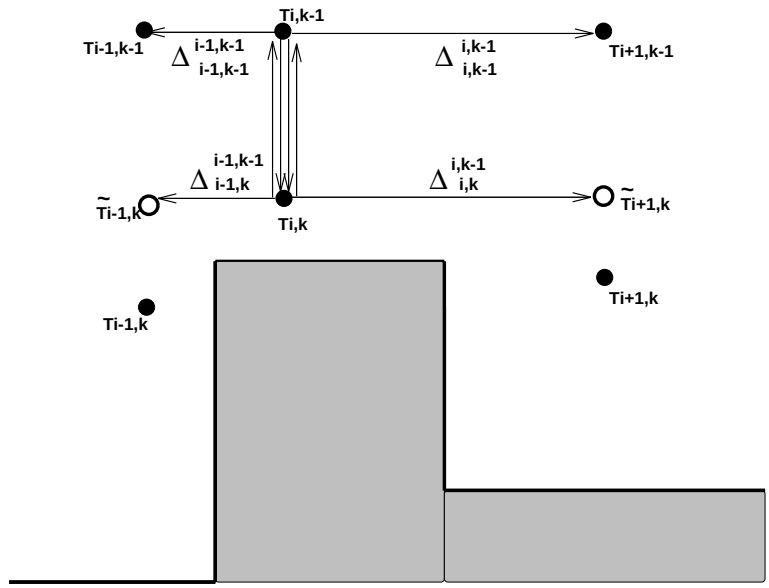

Figure 14.4: Stencil for computing the vertical component neutral diffusive flux $F^z_{i,k-1}$ (equation (14.88)), which is located at the bottom face of T-cell $T_{i,k-1}$. Shown are four density triads, each weighted by the relevant vertical spacing factor Δ. Note the rotation of the triads relative to those used to construct the zonal flux in Figure 14.3. Each of the four triads in this figure employ a generally different vertical weighting factor Δ.

label j was dropped. Likewise, we define the diffusivity weighted meridional slope

$$4\,(\kappa\, Sy)_{j,k} = \\ \kappa_{j,k}\, Sy^{(j,k)}_{(j,k|j,k)} + \kappa_{j+1,k}\, Sy^{(j+1,k)}_{(j,k|j+1,k)} + \kappa_{j,k+1}\, Sy^{(j,k+1)}_{(j,k+1|j,k)} + \kappa_{j+1,k+1}\, Sy^{(j+1,k+1)}_{(j,k+1|j+1,k)}. \quad (14.98)$$

Given these slopes, the zonal eddy-induced advection velocity, defined at the zonal face of the tracer cell $T_{i,j,k}$, is given by

$$u^*_{i,k} = \frac{(\kappa\, Sx)_{i,k-1} - (\kappa\, Sx)_{i,k}}{dht_{i,k}}, \quad (14.99)$$

and the meridional eddy-induced advection velocity, defined at the meridional face of the tracer cell $T_{i,j,k}$, is given by

$$v^*_{j,k} = \frac{(\kappa\, Sy)_{j,k-1} - (\kappa\, Sy)_{j,k}}{dht_{j,k}}. \quad (14.100)$$

The vertical component w^* of the eddy-induced advection velocity is obtained by vertically integrating the convergence of the horizontal eddy-advection velocities, as done to compute the resolved scale vertical advection velocity. Note that traditionally there is a zero vertical eddy-advection velocity at the top and bottom of the ocean. This boundary condition on the velocity effectively places a boundary condition on the diffusivity κ (e.g., see discussion in Gent et al. (1995) and Treguier et al. (1997)).

14.8.3 Computing the vertical flux convergence

The vertical flux convergence $-\delta_z F^{(z)}$ presents a special case numerically due to the small grid spacing in the vertical relative to the horizontal. As noted by Cox (1987), it is numerically efficient to split the off-diagonal portions of $F^{(z)}$ from the diagonal K^{33} portion. The K^{33} term presents the most severe limitation on the time step, and so it is useful to time step this term implicitly, along with the usual vertical diffusion term. That is, it is common to have the following portion of the diffusion equation time stepped implicitly

$$\partial_t T = \delta_z \left[(A_D + K^{33})\, \delta_z T \right]. \tag{14.101}$$

The off-diagonal terms in $F^{(z)}$ present a less severe time step constraint, and so they are time stepped explicitly. Indeed, it is not feasible to time step the off-diagonal components implicitly since to do so would involve inversion of a matrix the size of the model.

The above approach is useful for maintaining numerical stability. However, because of the time split of the vertical flux component $F^{(z)}$, it is generally not possible to rigorously maintain the balance within the vertical flux component of active tracers according to the discussion in Section 12.1.6. That is, the continuum relation

$$\alpha\, F^{(z)}(\theta) = \beta\, F^{(z)}(s) \tag{14.102}$$

is not fully satisfied by the discretized flux, so long as the K^{33} piece is time stepped implicitly. As discussed in Griffies et al. (1998), this limitation of the scheme has not proven to be a problem in realistic simulations. Even so, it has been found useful to reduce the imbalance in the vertical flux component by splitting K^{33} into explicit and implicit pieces

$$K^{33} = K^{33}_{explicit} + K^{33}_{implicit}, \tag{14.103}$$

where

$$K^{33}_{explicit} \leq (\Delta z)^2 / (4\, \Delta t). \tag{14.104}$$

The explicit piece is combined with the rest of the vertical flux for explicit time stepping, whereas the implicit piece is split off for implicit time stepping. For parts of the simulation where $K^{33} \leq (\Delta z)^2/(4\,\Delta t)$, $K^{33}_{implicit}$ is set to zero. Otherwise, its value is reduced to its minimum available for linear stability of the explicitly time stepped vertical diffusion equation.

Part V

MODEL HORIZONTAL FRICTION

Model Horizontal Friction

Frictional dissipation of kinetic energy in the ocean is tiny (Section 4.4.2), yet kinetic energy dissipation in ocean models is huge. Why the distinction? Ocean models require friction for numerical reasons, and these reasons are of no concern to the real ocean. It is therefore a goal of most ocean climate modelers to reduce friction levels to their lowest where we both retain a numerically stable and noise-free solution, and allow the flow to have a nontrivial Reynolds number.

There is no well established theory for how to close the momentum equations used in ocean climate models. Instead, various "tricks" are employed. So long as the methods maintain certain desired symmetries, they are just as warranted as any other approach. This part of the book attempts to document a variety of approaches. Much of the material transcends details of a particular discrete ocean model lattice. Hence, it should be of interest to modelers of all flavors.

The formulation of friction in ocean models follows from standard methods in continuum mechanics. Many of the mathematical tools developed in Part VI are useful here. In particular, these tools help to maintain some basic requirements of symmetry on the sphere, whose absence would represent a disrespect of the continuum system's symmetries. But the sophistication of the mathematics involved in formulating the frictional operator should not lead the reader to assume this is physically rigorous material. Instead, it is best regarded as sophisticated engineering with the practical goal, again, of maintaining small levels of friction without incurring unexceptable levels of grid noise.

CHAPTER
FIFTEEN

Horizontal friction

Contents

15.1 Boussinesq and non-Boussinesq friction **332**
15.2 Introduction and general framework **332**
15.3 Properties of the stress tensor **333**
15.3.1 The strain tensor 334
15.3.2 Relating strain to stress 336
15.3.3 Angular momentum and symmetry of the stress tensor . . 336
15.3.4 Quasi hydrostatic approximation 337
15.3.5 Trace-free frictional stress 338
15.4 Properties of the viscosity tensor **339**
15.4.1 Hooke's law 339
15.4.2 Angular momentum 339
15.4.3 Kinetic energy dissipation 340
15.4.4 Trace-free transverse stress 340
15.5 Transverse isotropy **341**
15.5.1 Viscosity tensor 341
15.5.2 Stress tensor 342
15.5.3 The friction vector 343
15.5.4 The case of nonconstant viscosity 343
15.6 Transverse anisotropy **344**
15.6.1 Viscosity tensor 344
15.6.2 Stress tensor 345
15.7 Friction in generalized orthogonal coordinates **347**
15.7.1 Deformation rates 347
15.7.2 Stress tensor components 348
15.7.3 The friction vector 349
15.8 Dissipation functional **350**
15.8.1 Motivation for the functional approach 350
15.8.2 Defining the functional 350
15.8.3 Performing the functional derivative 351

15.8.4 The connection to friction 353
15.9 Biharmonic friction 354
15.9.1 General formulation 354
15.9.2 Effects on kinetic energy 355
15.10 Some mathematical details 356
15.10.1 Salient results from Chapters 18 and 19 356
15.10.2 Details of the deformation rate derivation 357
15.10.3 Details of the friction derivation 357

The purpose of this chapter is to describe the fundamentals of horizontal momentum friction used in many ocean climate models. Discussions are limited to the continuum. Chapter 16 provides guidelines for choosing the viscosity setting the strength of the model's friction, and Chapter 17 presents discretization methods appropriate for a B-grid ocean model.

15.1 Boussinesq and non-Boussinesq friction

Throughout this chapter as well as Chapter 17, we focus on establishing friction within the context of the unaveraged equations of motion derived in Chapter 3. Such is strictly a mis-representation of the unaveraged equations, since they employ dissipation operators due to the tiny molecular friction, not the qhuge friction necessary for ocean climate models of focus here. Nonetheless, we take this approach largely for purposes of tidiness. When transferring the present results to model fields, we follow the mapping given by Table 6.1. In particular, $\mathbf{v} \to \mathbf{v}^\rho$ for the non-Boussinesq fluid, and $\mathbf{v} = \mathbf{v}^\rho$ for the Boussinesq fluid. Notably, this mapping means that it is the velocity $\mathbf{v}^\rho = \mathbf{v}\,(\rho_o/\rho)$ that is used to construct the friction operator in the non-Boussinesq model, *not* the velocity $\mathbf{v}$. In addition, the friction operator in the model represents a discrete version of the averaged momentum flux

$$\langle \tilde{\mathbf{F}}^{(\mathbf{u})}_{sgs} \rangle \to \mathbf{F}^{(\mathbf{u})}_{\text{model}}. \tag{15.1}$$

Hence, when considering model implementations, factors of density reduce to their Boussinesq value ρ_o, which simplifies the numerical treatment in the non-Boussinesq case.

15.2 Introduction and general framework

The physical framework used here for deriving the form of the momentum friction operator is the following:

- The fluid motion preserves the hydrostatic balance.* This assumption is part of the Traditional Approximation (see Section 3.1) fundamental to large scale ocean climate modeling, whereby the ocean is considered a shallow layer of fluid moving on a rotating sphere.

*Williams (1972) generalizes to non-hydrostatic motion.

- Horizontal kinetic energy, which dominates the vertical kinetic energy in hydrostatic flows, is dissipated by friction.

- Friction does not introduce interior sources or sinks of angular momentum. Hence, friction vanishes when the fluid undergoes a solid body rotation. On a plane, friction also vanishes under uniform translations.

- Stresses associated with friction in the interior of the fluid are directly proportional to the local strain acting on fluid parcels. This assumption follows the generalized Hooke's Law approach commonly used in elasticity theory (e.g., Aris (1962), Landau and Lifshitz (1986), and Smagorinsky (1993)).

It is common to additionally assume that frictional stresses maintain transverse isotropy in which gravity picks out the only special direction (e.g., Smagorinsky (1993), Pacanowski and Griffies (1999), and Griffies and Hallberg (2000)). However, Smith and McWilliams (2002) suggested that anisotropic transverse stresses are more appropriate. The freedom afforded from transverse anisotropy allows one, in particular, to prescribe a friction force that is a function of the flow direction and/or an *a priori* specified direction. Such has been found to be of use for ocean climate modeling, especially when aiming to increase the strength of simulated equatorial undercurrents (Large et al. (2001)).

For simplicity, the analysis in much of this chapter uses Cartesian coordinates and Cartesian tensors. Generalizations are straightforward so long as the formulation respects the tensor rules summarized in Section 19.1 and 19.12. Section 15.7 details the friction operator appropriate for a generalized orthogonal coordinate ocean model.

15.3 Properties of the stress tensor

As we saw in Section 3.4.2, forces acting on an element of a continuous media can be classified into two categories. *External or body* forces act throughout the media, with forces from gravity, Coriolis, and electromagnetism common examples. *Internal or contact* forces act on an element of volume through its bounding surface, with forces from pressure or other stresses common examples. The balance between all forces with acceleration leads, through Newton's Second Law of Mechanics, to the equations of motion. Furthermore, if all torques acting on a fluid parcel arise from macroscopic forces, which is the case for Newtonian fluids such as seawater, then fluid parcels conserve angular momentum, and so should ocean model solutions. Notably, however, the presence of meridional boundaries precludes angular momentum conservation in the World Ocean, except in regions of the Southern Ocean devoid of meridional barriers. More is said regarding angular momentum in Sections 3.9 and 3.10.

As introduced in Section 3.4.4, the stresses acting within a continuous media can be organized into a second order stress tensor with generally 3×3 independent elements. The divergence of these stresses gives rise to the internal forces acting in the media. A proper account of the angular momentum budget implies that the stress tensor is symmetric, which brings the number of independent stress elements down to six. It is useful to note that symmetry of the stress tensor is equivalent to

Cauchy's reciprocal theorem (see Section 3.4 of this book or Section 5.13 of Aris (1962)), which says that each of two stresses at a point has an equal projection on the normal to the surface on which the other acts.

15.3.1 The strain tensor

Consider two infinitesimally close fluid parcels with material coordinates ζ^a and $\zeta^a + d\zeta^a$, where $a = 1, 2, 3$. The components

$$(u^1, u^2, u^3) = \frac{\mathrm{d}}{\mathrm{d}t}\,(x^1, x^2, x^3) \tag{15.2}$$

of the velocity for these two parcels differ by the increment

$$\begin{aligned} \mathrm{d}u^m &= \frac{\partial u^m}{\partial \zeta^a}\,\mathrm{d}\zeta^a \\ &= \frac{\partial u^m}{\partial \zeta^a}\,\frac{\partial \zeta^a}{\partial x^n}\,\mathrm{d}x^n \\ &= \frac{\partial u^m}{\partial x^n}\,\mathrm{d}x^n \\ &= u^m_{,n}\,\mathrm{d}x^n, \end{aligned} \tag{15.3}$$

where the summation convention is used in which repeated indices are summed over their range. The velocity derivatives

$$u^m_{,n} = \frac{\partial u^m}{\partial x^n} \tag{15.4}$$

form the components to a second order tensor. In order to attach physical significance to this tensor, it is useful to separately consider its symmetric and anti-symmetric components*, which are written

$$u_{m,n} = \Omega_{mn} + e_{mn}, \tag{15.5}$$

where

$$2\,\Omega_{mn} = u_{m,n} - u_{n,m} \tag{15.6}$$

$$2\,e_{mn} = u_{m,n} + u_{n,m}, \tag{15.7}$$

and $u_m = g_{mn}\,u^n$ are the covariant components to the velocity vector. Note that the components to the metric tensor g_{mn} are equal to the Kronecker symbol δ_{mn} when working with Cartesian coordinates.† In curvilinear coordinates, the partial derivatives appearing in Ω_{mn} and e_{mn} generalize to covariant derivatives. We consider this generalization in Section 15.7.1, where the comma notation for partial derivative introduced in equation (15.4) is replaced by a semi-colon for covariant derivative. Chapters 18 and 19 provide full discussion of covariant derivatives.

*A similar decomposition is considered in Section 11.4 where we discuss the sub-grid scale tracer transport tensor.

†More general metric tensors are described in Chapters 18, 19, and 5. See also the summary discussion in Section 19.12.

The anti-symmetric piece of the velocity derivative tensor is related to the vorticity through

$$\begin{aligned} 2\,\omega^m &= -\epsilon^{mnp}\,\Omega_{np} \\ &= -\epsilon^{mnp}\,u_{n,p}, \end{aligned} \tag{15.8}$$

where ϵ^{mnp} is the Levi-Civita symbol. In Cartesian vector notation, this result takes the form

$$\boldsymbol{\omega} = \frac{1}{2}\,\nabla \wedge \mathbf{u}. \tag{15.9}$$

Standard results from fluid mechanics establish the connection between vorticity and rigid body rotation of a fluid parcel.* If the motion is completely rigid, which means that it consists of a translation plus a rotation, then the symmetric part of the velocity derivative tensor vanishes. Consequently, the symmetric tensor e_{mn} is called the *deformation* or *rate of strain* tensor since it represents deviations from rigid body motion.

To provide a further interpretation of the strain tensor, consider the squared arc-distance $(\mathrm{d}s)^2 = \mathrm{d}s^2$ between two infinitesimally close material parcels of fluid

$$\begin{aligned} \mathrm{d}s^2 &= g_{mn}\,\mathrm{d}x^m\,\mathrm{d}x^n \\ &= g_{mn}\,\frac{\partial x^m}{\partial \zeta^a}\frac{\partial x^n}{\partial \zeta^b}\,\mathrm{d}\zeta^a\,\mathrm{d}\zeta^b. \end{aligned} \tag{15.10}$$

The material time derivative of this distance is given by

$$\frac{\mathrm{d}(\mathrm{d}s^2)}{\mathrm{d}t} = g_{mn}\left(\frac{\partial u^m}{\partial \zeta^a}\frac{\partial x^n}{\partial \zeta^b} + \frac{\partial x^m}{\partial \zeta^a}\frac{\partial u^n}{\partial \zeta^b}\right)\mathrm{d}\zeta^a\,\mathrm{d}\zeta^b \tag{15.11}$$

where $\mathrm{d}\zeta^a/\mathrm{d}t = 0$ since ζ^a are material coordinates. Use of the chain rule in the forms

$$\frac{\partial u^m}{\partial \zeta^a}\,\mathrm{d}\zeta^a = u^m_{\ ,n}\,\mathrm{d}x^n \tag{15.12}$$

$$\frac{\partial x^m}{\partial \zeta^a}\,\mathrm{d}\zeta^a = \mathrm{d}x^m \tag{15.13}$$

renders

$$\frac{\mathrm{d}(\mathrm{d}s^2)}{\mathrm{d}t} = 2\,e_{mn}\,\mathrm{d}x^m\,\mathrm{d}x^n, \tag{15.14}$$

or equivalently

$$\frac{1}{\mathrm{d}s}\,\frac{\mathrm{d}(\mathrm{d}s)}{\mathrm{d}t} = e_{mn}\,\frac{\mathrm{d}x^m}{\mathrm{d}s}\,\frac{\mathrm{d}x^n}{\mathrm{d}s}. \tag{15.15}$$

Now $\mathrm{d}x^m/\mathrm{d}s$ is a component of the unit vector pointing from one fluid parcel to the other. Hence, equation (15.15) says that the rate of change of the infinitesimal distance separating the two parcels, as a fraction of the distance, is related to the relative position of the parcels through the strain tensor. Section 4.42 of Aris (1962), amongst others, provides further elaboration.

*See also Section 3.9.2 for a discussion of solid-body motion of a particle on a rotating sphere.

15.3.2 Relating strain to stress

The divergence of the stress tensor yields the internal forces acting on a fluid parcel. Stress is related to strain, where strain arises from the kinematics of parcel deformations. In elasticity theory, the relation between stress and strain is typically assumed to follow some form of Hooke's law. In its simplest form, this "law" linearly relates the stress to the strain.* In fluid dynamics, it is common to also assume a stress-strain relation in the form of Hooke's law. The details of this relation often depend quite strongly on the properties of the fluid as well as the flow state. Such dependencies can generally make the fluid's stress-strain relation nonlinear.

When the fluid is static and exposed only to buoyancy forces, it remains in hydrostatic balance, and the only form of stress on a fluid parcel is due to the pressure (Section 3.6). Hence, the stress tensor for such a state takes the form

$$T^{mn} = -p\,\delta^{mn} \tag{15.16}$$

where p is the pressure and δ^{mn} is the Kronecker delta. Chapter 1 of Salmon (1998) provides some comments on the implicit identification of this pressure with thermodynamic pressure.

When the fluid undergoes deformations, further stresses bring the stress tensor to the more general form

$$T^{mn} = -p\,\delta^{mn} + \widetilde{\tau}^{mn}. \tag{15.17}$$

For purposes of formulating friction to be used in an ocean model, the divergence of $\widetilde{\tau}^{mn}$ is associated here with dissipative stresses in the fluid, thus motivating the name *frictional stress tensor*. For a Newtonian fluid, the frictional stress tensor can be written

$$\widetilde{\tau}^{mn} = \rho\, C^{mnpq}\, e_{pq}. \tag{15.18}$$

In general, this relation between stress and strain is of the form of Hooke's law, where the components C^{mnpq} of the fourth-order kinematic *viscosity tensor* can depend on the state of the fluid. Assuming this form for the internal stresses, the essential problem with subgrid scale parameterization of momentum fluxes reduces to determining appropriate forms for C^{mnpq}.

15.3.3 Angular momentum and symmetry of the stress tensor

As mentioned in Section 15.3.1, the symmetric strain tensor e_{mn} vanishes for motion consisting of rigid rotation plus uniform translation. In such cases, the generalized Hooke's law (15.18) says that the stress tensor T^{mn} reduces to $-p\,\delta^{mn}$ since $\widetilde{\tau}^{mn}$ vanishes. The purpose of this section is to provide some further details regarding these ideas and their connection to conservation of angular momentum.

The continuum form of Newton's law is given by

$$\rho\,\frac{\mathrm{d}u^m}{\mathrm{d}t} = \rho\, f^m + T^{mn}_{\ \ ,n} \tag{15.19}$$

where ρ is the mass density, and f^i are components to external or body forces such as those arising from gravity and the Coriolis force. $T^{mn}_{\ \ ,n}$ is the divergence of the

*A similar relation between dynamics and kinematics forms the basis of general relativity, where stresses associated with matter/energy are directly proportional to space-time curvature.

stress tensor, where T^{mn} is written in the form (15.17) which incorporates the pressure. A component of the angular momentum for a fluid parcel is given by

$$L_m = \epsilon_{mnp}\, x^n\, u^p\, \rho\, \mathrm{d}V, \tag{15.20}$$

where $\rho\, \mathrm{d}V$ is the mass of the infinitesimal parcel, and ϵ_{mnp} is the totally antisymmetric Levi-Civita symbol (Section 18.12). The material time derivative of the parcel's angular momentum is given by

$$\frac{\mathrm{d}L_m}{\mathrm{d}t} = \epsilon_{mnp}\, x^n\, \frac{\mathrm{d}u^p}{\mathrm{d}t}\, \rho\, \mathrm{d}V, \tag{15.21}$$

where $\mathrm{d}(\rho\, \mathrm{d}V)/\mathrm{d}t = 0$ follows from mass conservation. For a Boussinesq fluid, ρ appears as the constant ρ_o, and $\mathrm{d}(\mathrm{d}V)/\mathrm{d}t = 0$ then follows from volume conservation. Substituting Newton's law into this expression leads to

$$\frac{\mathrm{d}L_m}{\mathrm{d}t} = \epsilon_{mnp}\, (x^n\, \rho\, f^p + x^n\, T^{pq}_{,q})\, \mathrm{d}V. \tag{15.22}$$

The first term accounts for torques placed on the parcel from external forces. The second term arises from torques on the fluid from internal stresses. To further interpret the second term, consider the budget for total angular momentum of the fluid, which is obtained by integrating over the fluid volume

$$\int \frac{\mathrm{d}L_m}{\mathrm{d}t} = \int \epsilon_{mnp}\, (x^n\, \rho\, f^p + x^n\, T^{pq}_{,q})\, \mathrm{d}V. \tag{15.23}$$

Now integrate by parts on the stress tensor term to find

$$\int \epsilon_{mnp}\, x^n\, T^{pq}_{,q}\, \mathrm{d}V = \int \epsilon_{mnp}\, [\partial_q\, (x^n\, T^{pq}) - T^{pn}]\, \mathrm{d}V. \tag{15.24}$$

The first term integrates to a boundary contribution, which is non-vanishing for cases in which there are torques arising from boundary stresses. The second term is a volume contribution and it picks out the term $\epsilon_{mnp}\, \widetilde{\tau}^{pn}$, since $\epsilon_{mnp}\, \delta^{pn} = 0$. For Newtonian fluids, such as ocean water, the internal torques are balanced and so there is no net contribution to angular momentum from internal stresses. This case can be ensured if the frictional stress tensor is symmetric

$$\widetilde{\tau}^{mn} = \widetilde{\tau}^{nm}, \tag{15.25}$$

thus rendering $\epsilon_{mnp}\, \widetilde{\tau}^{pn} = 0$. This result provides the reasoning for our considering only symmetric stress tensors in the following.

15.3.4 Quasi hydrostatic approximation

As pointed out by Smith and McWilliams (2002), there is some ambiguity regarding use of the quasi-hydrostatic approximation in so far as it reduces the form of the stress tensor. We follow their pragmatic approach whereby it is noted that our concern is to derive friction for a fluid satisfying the Traditional Approximation (i.e., the hydrostatic primitive equations), where friction is applied explicitly only to the horizontal momentum. Friction due to deformations of the vertical velocity are

ignored, since the vertical momentum equation is reduced to the hydrostatic balance. Such forms here the *quasi-hydrostatic* approximation, where the "quasi" label signifies that the fluid is not static, although the hydrostatic balance is maintained.

These considerations suggest that we split the frictional stress tensor into two sub-tensors

$$\widetilde{\tau}^{mn} = \tau^{mn}_{vert} + \tau^{mn}. \tag{15.26}$$

The first part of the stress tensor, τ^{mn}_{vert}, is assumed to arise only from vertical deformations of the horizontal velocity field. It can be written as

$$\tau^{mn}_{vert} = 2\,\rho\,\kappa\,\delta^{n3}\,e^{m3} \approx \rho\,\kappa\,\delta^{n3}\,\partial_3\,u^m \tag{15.27}$$

with κ a non-negative viscosity, and $m = 1, 2$. The second tensor, τ^{mn}, embodies the stress arising from horizontal deformations of the horizontal velocity field. Consequently, the labels $m, n = 1, 2$ range only over the transverse coordinates, with τ^{mn} set to zero if one of the indices equals 3. It is symmetric, according to the requirements of angular momentum conservation discussed in Section 15.3.3. Hence, the general form of the frictional stress tensor is given by

$$\widetilde{\tau}^{mn} = \begin{pmatrix} \tau^{11} & \tau^{12} & \rho\,\kappa\,u_{,z} \\ \tau^{12} & \tau^{22} & \rho\,\kappa\,v_{,z} \\ \rho\,\kappa\,u_{,z} & \rho\,\kappa\,v_{,z} & 0 \end{pmatrix}. \tag{15.28}$$

Note that $\widetilde{\tau}^{33} = 0$ is consistent with our use of the hydrostatic approximation, which reduces the vertical momentum equation to the inviscid hydrostatic balance.

15.3.5 Trace-free frictional stress

The frictional stress tensor under consideration here is a deviatoric stress tensor (e.g., Smagorinsky (1993), Salmon (1998)), which is defined to have a zero trace

$$\delta_{mp}\,\widetilde{\tau}^{mp} = \widetilde{\tau}^1_1 + \widetilde{\tau}^2_2 + \widetilde{\tau}^3_3 = 0, \tag{15.29}$$

where we lowered one index on the stress tensor components to yield a coordinate invariant expression for the trace.* Pressure incorporates the trace part of the general stress tensor, as seen in equation (15.17). As mentioned in Section 15.3.4, the quasi-hydrostatic approximation allows us to ignore stresses due to deformations in the vertical velocity field, and so we can set $\widetilde{\tau}^{33} = 0$ without loss of generality. If we make this approximation, the trace-free and symmetric conditions lead to the transverse stress tensor

$$\tau^{mn} = \begin{pmatrix} \tau^{11} & \tau^{12} \\ \tau^{12} & -\tau^{11} \end{pmatrix}. \tag{15.30}$$

Specification of this tensor, and the corresponding friction vector, is the focus of the rest of this chapter.

*Again, with Cartesian tensors, there is no difference in the index placement.

15.4 Properties of the viscosity tensor

For three-dimensional systems, the viscosity tensor C^{mnpq} generally contains 81 degrees of freedom ($81 = 3 \times 3 \times 3 \times 3$). However, the properties just described, and others to follow, greatly reduce this number. In particular, our focus on τ^{mn}, for $m, n = 1, 2$, allows us to start with $16 = 2 \times 2 \times 2 \times 2$ degrees of freedom. Hence, in the following discussion all labels are assumed to run over just the transverse coordinates $1, 2$.

15.4.1 Hooke's law

The Hooke's Law form for stress

$$\tau^{mn} = \rho\, C^{mnpq}\, e_{pq} \tag{15.31}$$

indicates that the only relevant forms of the viscosity tensor are those satisfying

$$C^{mnpq} = C^{mnqp} \tag{15.32}$$

since the strain tensor e_{mn} is symmetric. This constraint reduces the degrees of freedom in the viscosity tensor to $2 \times 2 \times 3 = 12$. The 3 arises from the $2 + 1 = 3$ degrees of freedom in a symmetric 2×2 matrix. We can therefore write the general viscosity tensor

$$C^{mnpq} = \begin{pmatrix} C^{1111} & C^{1112} & C^{1121} & C^{1122} \\ C^{1211} & C^{1212} & C^{1221} & C^{1222} \\ C^{2111} & C^{2112} & C^{2121} & C^{2122} \\ C^{2211} & C^{2212} & C^{2221} & C^{2222} \end{pmatrix} \tag{15.33}$$

in the restricted form

$$C^{mnpq} = \begin{pmatrix} C^{1111} & C^{1112} & C^{1112} & C^{1122} \\ C^{1211} & C^{1212} & C^{1212} & C^{1222} \\ C^{2111} & C^{2112} & C^{2112} & C^{2122} \\ C^{2211} & C^{2212} & C^{2212} & C^{2222} \end{pmatrix}. \tag{15.34}$$

15.4.2 Angular momentum

Assuming a symmetric stress tensor brings about the following symmetry on the viscosity tensor

$$C^{mnpq} = C^{nmpq}. \tag{15.35}$$

As such, the total degrees of freedom become $3 \times 3 = 9$, and the viscosity tensor takes the form

$$C^{mnpq} = \begin{pmatrix} C^{1111} & C^{1112} & C^{1112} & C^{1122} \\ C^{1211} & C^{1212} & C^{1212} & C^{1222} \\ C^{1211} & C^{1212} & C^{1212} & C^{1222} \\ C^{2211} & C^{2212} & C^{2212} & C^{2222} \end{pmatrix}. \tag{15.36}$$

15.4.3 Kinetic energy dissipation

The budget for horizontal kinetic energy per volume of a fluid parcel is given by

$$\begin{aligned}\frac{\rho}{2}\,\frac{\mathrm{d}(\delta_{mn}\,u^m\,u^n)}{\mathrm{d}t} &= \rho\,\delta_{mn}\,u^m\,f^n + \delta_{mn}\,u^m\,T^{np}_{,p} \\ &= \rho\,u_m\,f^m + \left(u_n\,T^{np}\right)_{,p} - u_{n,p}\,T^{np} \\ &= \rho\,u_m\,f^m + \left(u_n\,T^{np}\right)_{,p} - e_{np}\,T^{np} \\ &= \rho\,u_m\,f^m + \left(u_n\,T^{np}\right)_{,p} + p\,u^n{}_{,n} - e_{np}\,\tau^{np},\end{aligned} \tag{15.37}$$

where we assumed Cartesian coordinates for simplicity (see Section 4.4.2 for the general case). The first term on the right hand side arises from work done by external forces. The second term, when integrated over the fluid domain, accounts for work done at boundaries by the stresses. The third term arises from pressure work against changes in the parcel's volume. This term vanishes for a volume conserving fluid. The fourth term is present throughout the fluid domain, and it can be written

$$e_{mn}\,\tau^{mn} = \rho\,e_{mn}\,C^{mnpq}\,e_{pq}. \tag{15.38}$$

In general, this term is sign-indefinite. However, we insist that the frictional stress tensor manifests dissipative friction at each point in the fluid, thus necessitating

$$e_{mn}\,C^{mnpq}\,e_{pq} \geq 0. \tag{15.39}$$

Since the strain tensor e_{mn} is symmetric, this constraint is satisfied if

$$C^{mnpq} = C^{pqmn}. \tag{15.40}$$

This constraint brings the number of degrees of freedom in the viscosity tensor down to $6 = 3 + 2 + 1$, which is the number of degrees of freedom in a 3×3 symmetric matrix. The resulting viscosity tensor takes the form

$$C^{mnpq} = \begin{pmatrix} C^{1111} & C^{1112} & C^{1112} & C^{1122} \\ C^{1112} & C^{1212} & C^{1212} & C^{1222} \\ C^{1112} & C^{1212} & C^{1212} & C^{1222} \\ C^{1122} & C^{1222} & C^{1222} & C^{2222} \end{pmatrix}. \tag{15.41}$$

15.4.4 Trace-free transverse stress

As discussed in Section 15.3.5, we are considering trace-free transerve stress tensors. Hence,

$$\begin{aligned}&\tau^{11} + \tau^{22} = \\ &\left(C^{1111} + C^{1122}\right)\,e_{11} + 2\,\left(C^{1112} + C^{2212}\right)\,e_{12} + \left(C^{1122} + C^{2222}\right)\,e_{22} = 0.\end{aligned} \tag{15.42}$$

Given the general independence of the three strain tensor components, we satisfy this constraint for all cases only when the three terms individually vanish

$$C^{1122} = -C^{1111} \tag{15.43}$$

$$C^{2212} = -C^{1112} \tag{15.44}$$

$$C^{1122} = -C^{2222}. \tag{15.45}$$

There are now three degrees of freedom in the viscosity tensor

$$C^{mnpq} = \begin{pmatrix} \alpha & \gamma & \gamma & -\alpha \\ \gamma & \delta & \delta & -\gamma \\ \gamma & \delta & \delta & -\gamma \\ -\alpha & -\gamma & -\gamma & \alpha \end{pmatrix} \tag{15.46}$$

where we introduced the short-hand notation for the three viscosities

$$C^{1111} = \alpha \tag{15.47}$$

$$C^{1112} = \gamma \tag{15.48}$$

$$C^{1212} = \delta. \tag{15.49}$$

15.5 Transverse isotropy

Gravity breaks three-dimensional isotropy down to transverse isotropy about the local vertical direction $\hat{\mathbf{z}}$. Before considering the more general anisotropic case in Section 15.6, we first consider the isotropic case. This case is presently chosen by most ocean climate models (e.g., Griffies and Hallberg (2000)).

15.5.1 Viscosity tensor

Transverse, or axial, isotropy means two things. First, the physical system remains invariant under arbitrary rotations about the $\hat{\mathbf{z}}$ direction. Second, the physical system remains invariant under the transformation $z \to -z$, and $x \to y$, $y \to x$, which is a transformation between two right handed coordinate systems, with the vertical pointing up and down, respectively. Since friction is directly proportional to viscosity, and since the physical system is directly affected by friction, transverse isotropy for the physical system means that the viscosity tensor must also manifest this symmetry.

The transformation matrix for rotation about the vertical takes the form

$$\Lambda^{\overline{m}}_{m} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{15.50}$$

and the transformation matrix between right handed coordinate systems takes the form

$$\Lambda^{\overline{m}}_{m} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{15.51}$$

Under an arbitrary transformation, the fourth order viscosity tensor transforms as

$$C^{\overline{m}\,\overline{n}\,\overline{p}\,\overline{q}} = \Lambda^{\overline{m}}_{m}\,\Lambda^{\overline{n}}_{n}\,\Lambda^{\overline{p}}_{p}\,\Lambda^{\overline{q}}_{q}\,C^{mnpq}. \tag{15.52}$$

Chapters 5, 18, and 19 provide discussions of such transformation rules, and Section 19.12 provides a summary of certain salient points. Note that restricting attention to the two-dimensional transverse subspace allows us to focus on the two-dimensional portion of these transformation matrices. In general, transverse isotropy imposes the constraint on the viscosity tensor

$$\begin{aligned} C^{\overline{m}\,\overline{n}\,\overline{p}\,\overline{q}} &\equiv C^{mnpq} \\ &= \Lambda^{\overline{m}}_{m}\,\Lambda^{\overline{n}}_{n}\,\Lambda^{\overline{p}}_{p}\,\Lambda^{\overline{q}}_{q}\,C^{mnpq}, \end{aligned} \tag{15.53}$$

where $\Lambda^{\overline{m}}_{m}$ is one of the given transformation matrices.

Determining the relations between the elements of C^{mnpq} requires enumeration of the possibilities. For example, with a rotation angle of $\pi/2$ about $\hat{\mathbf{z}}$, rotational symmetry implies

$$C^{\overline{1}\,\overline{2}\,\overline{2}\,\overline{2}} \equiv C^{1222} = -C^{2111}, \tag{15.54}$$

which is a constraint also imposed by the trace-free condition discussed in Section 15.4.4. However, the transformation between two right handed coordinate systems implies

$$C^{1222} = C^{2111}. \tag{15.55}$$

These two results are satisfied only if

$$C^{1222} = C^{2111} = 0 \Rightarrow \gamma = 0. \tag{15.56}$$

Likewise, insisting on isotropy when rotating by $\pi/4$ implies

$$C^{1111} = (C^{1111} + C^{1122} + 2\,C^{1212})/2, \tag{15.57}$$

or

$$C^{1212} = (C^{1111} - C^{1122})/2 \Rightarrow \delta = \alpha. \tag{15.58}$$

Consequently, the transverse isotropic viscosity tensor takes the simple form

$$C^{mnpq} = \alpha \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}, \tag{15.59}$$

thus manifesting only a single viscous degree of freedom.

15.5.2 Stress tensor

The transverse stress tensor satisfying transverse isotropy is given by

$$\tau^{mn} = \rho\,A \begin{pmatrix} (e_{11} - e_{22}) & 2\,e_{12} \\ 2\,e_{12} & (e_{22} - e_{11}) \end{pmatrix} = \rho\,A\,(2\,e^{mn} - g^{mn}\,e^{p}_{p}), \tag{15.60}$$

where we introduced $\alpha = A$ as this is the common notation for viscosity in the isotropic case (e.g., Pacanowski and Griffies (1999); Griffies and Hallberg (2000)). Notice how the stress is a function only of the two strain terms $e_{11} - e_{22}$ and $2\,e_{12}$,

instead of the three components e_{11}, e_{22}, e_{12}. For this reason, it is common to define the tension and shearing rate of strain, also known as the *deformation rates*, as

$$e_T = e_{11} - e_{22} = u_{1,1} - u_{2,2} \tag{15.61}$$

$$e_S = 2\, e_{12} = u_{1,2} + u_{2,1}. \tag{15.62}$$

The deformation rates have dimensions of inverse time. Their forms in generalized orthogonal coordinates are presented in Section 15.7.1.

The second expression in equation (15.60) is written in a covariant form, with g^{mn} the components to the inverse metric tensor (Section 19.12). We also introduced the trace of the strain tensor

$$g_{mn}\, e^{mn} = e^n_{\ n} = e^1_{\ 1} + e^2_{\ 2} \tag{15.63}$$

which is a scalar and so invariant under transformations of coordinates. The covariant form of the stress tensor (15.60) is valid for all coordinate systems. We consider the case of general orthogonal coordinates in Section 15.7.

15.5.3 The friction vector

The friction vector is given by the covariant divergence of the frictional stress tensor. In Cartesian coordinates, the covariant divergence is computed via the usual partial derivative operator

$$\rho\, F^m = \widetilde{\tau}^{mn}_{\ ,n}. \tag{15.64}$$

Performing the divergence using the transverse stress tensor in equation (15.60), as part of the full frictional stress tensor $\widetilde{\tau}^{mn}$ given by (15.28), leads to the friction components

$$\rho\, F^1 = \nabla_z \cdot (\rho\, A\, \nabla_z\, u^1) + \hat{\mathbf{z}} \cdot \nabla_z\, u^2 \wedge \nabla_z\, (\rho\, A) + [\rho\, \kappa\, (u^1_{,z})]_{,z} \tag{15.65}$$

$$\rho\, F^2 = \nabla_z \cdot (\rho\, A\, \nabla_z\, u^2) - \hat{\mathbf{z}} \cdot \nabla_z\, u^1 \wedge \nabla_z\, (\rho\, A) + [\rho\, \kappa\, (u^2_{,z})]_{,z} \tag{15.66}$$

$$\rho\, F^3 = 0. \tag{15.67}$$

In these expressions, the horizontal divergence operator $\nabla_z = (\partial_1, \partial_2, 0)$ was introduced, $z = \xi^3$ is the vertical coordinate, and A is the viscosity associated with horizontal deformations. The factors of density cancel out trivially upon making the Boussinesq approximation. Note that when making the quasi-hydrostatic approximation (as in the Traditional Approximation), the vertical friction F^3 is ignored so that the vertical momentum equation reduces to the inviscid hydrostatic equation. The extra cross-product terms appearing in the transverse friction vanish when using a constant viscosity. Their importance when using a spatially nonconstant viscosity is briefly highlighted in the next section.

15.5.4 The case of nonconstant viscosity

It now quite common for global ocean modelers to employ a nonconstant viscosity for various numerical reasons. As emphasized by Wajsowicz (1993), some implementations of the corresponding friction vector often ignore the importance of formulating friction as the divergence of a symmetric stress tensor. Namely, what is

sometimes done is to take the friction appropriate for a constant viscosity Boussinesq fluid

$$F^1_{const} = \nabla_z \cdot (A\, \nabla_z\, u^1) + [\kappa\, (u^1_{,z})]_{,z} \tag{15.68}$$
$$F^2_{const} = \nabla_z \cdot (A\, \nabla_z\, u^2), + [\kappa\, (u^2_{,z})]_{,z} \tag{15.69}$$

and allow A to be nonconstant. That is, the cross-product terms derived above are dropped. Focusing on the two-dimensional transverse sub-space, dropping the cross-product terms amounts to employing the non-symmetric stress tensor

$$\tau^{mn}_{NS} = \rho_o\, A \begin{pmatrix} u^1_{,1} & u^1_{,2} \\ u^2_{,1} & u^2_{,2} \end{pmatrix}. \tag{15.70}$$

It is easy to show that the chosen friction dissipates kinetic energy since it is written as a Laplacian. However, for a fluid in uniform rotation

$$\mathbf{u} = \mathbf{\Omega} \wedge \mathbf{x}, \tag{15.71}$$

where $\mathbf{u} = (u^1, u^2, 0)$, $\mathbf{x} = (x^1, x^2, 0)$, and $\mathbf{\Omega}$ is spatially constant, the horizontal friction vector takes the form

$$\mathbf{F}^{(h)} = -\nabla \wedge (A\, \mathbf{\Omega}), \tag{15.72}$$

and it vanishes only when A is a constant. As such, by using friction derived from a non-symmetric stress tensor and with a non-constant viscosity, a uniformly rotating fluid feels a nonzero stress. Conversely, such a stress tensor can introduce uniform rotation; i.e., it can act as an internal source or sink of angular momentum. Unless one has a physical reason for doing so, such viscosity dependent sources of angular momentum should be avoided.

15.6 Transverse anisotropy

Smith and McWilliams (2002) argued that friction should be determined according to the directions of the current flow. For example, in a zonal current, such as that along the equatorial region, friction arising from shears in the direction of the current ($u_{,x}$) should be preferentially larger than friction from shears perpendicular to the current ($u_{,y}$). They presented some compelling arguments that suggest this anisotropy is relevant for a physically based closure of the momentum equation. More pragmatically, both Smith and McWilliams (2002) and Large et al. (2001) noted that numerical constraints allow for anisotropy, and exploiting such freedom enhances the simulation's integrity.

15.6.1 Viscosity tensor

Following Smith and McWilliams (2002), we start with a viscosity tensor of the form

$$\begin{aligned} C^{mnpq} = {} & P\, \delta^{mn}\, \delta^{pq} + Q\, (\delta^{mp}\, \delta^{nq} + \delta^{mq}\, \delta^{np}) + R\, (\delta^{mn}\, s^{pq} + s^{mn}\, \delta^{pq}) \\ & + S\, (\delta^{mp}\, s^{nq} + \delta^{mq}\, s^{np} + s^{mp}\, \delta^{nq} + s^{mq}\, \delta^{np}) + T\, s^{mn}\, s^{pq} \end{aligned} \tag{15.73}$$

where P, Q, R, S, T are viscosities and s^{mn} is a symmetric tensor defined as an outer product

$$s^{mn} = \hat{s}^m \, \hat{s}^n. \tag{15.74}$$

Smith and McWilliams (2002) suggest orienting viscosity according to the instantaneous velocity field, in which case

$$\hat{\mathbf{s}} = \frac{\mathbf{u}}{|\mathbf{u}|}. \tag{15.75}$$

In contrast, Large et al. (2001) suggest orienting the viscosity according to the local grid lines. For example, with an eastward orientation using spherical coordinates,

$$\hat{\mathbf{s}} = (1, 0). \tag{15.76}$$

This orientation has been found useful to increase the strength of the predominantly zonal equatorial currents in the Pacific (Large et al. (2001)). In contrast, for time dependent eddying flows, the dynamic orientation may spread friction over a broader domain in the time mean, and so lead to larger effective friction than for the case with a static orientation. In general, tests are necessary to determine what is best for the particular application.

Regardless the choice for the *orientation tensor* s^{mn}, we assume that its components satisfy the two identities

$$s^{11} + s^{22} = 1 \tag{15.77}$$

$$s^{11} \, s^{22} = s^{12} \, s^{12}. \tag{15.78}$$

These identities are used to simplify the form of the stress tensor in Section 15.6.2. Additionally, C^{mnpq} as written in equation (15.73) satisfies the trace-free property from Section 15.4 if we impose

$$2\,(P + Q) + R = 0 \tag{15.79}$$

$$R + 2\,S + T/2 = 0. \tag{15.80}$$

These relations are assumed in the following.

15.6.2 Stress tensor

The transverse stress tensor $\tau^{mn} = \rho \, C^{mnpq} \, e_{pq}$ is given by

$$\tau^{mn} = \rho \begin{pmatrix} C^{1111}\,(e_{11} - e_{22}) + 2\,C^{1112}\,e_{12} & C^{1112}\,(e_{11} - e_{22}) + 2\,C^{1212}\,e_{12} \\ C^{1112}\,(e_{11} - e_{22}) + 2\,C^{1212}\,e_{12} & -C^{1111}\,(e_{11} - e_{22}) - 2\,C^{1112}\,e_{12} \end{pmatrix} \tag{15.81}$$

where the viscosity tensor components are

$$C^{1111} = (P + 2\,Q) + 2\,(R + 2\,S + (T/2)\,s^{11})\,s^{11} = (P + 2\,Q) - T\,s^{11}\,s^{22} \tag{15.82}$$

$$C^{1112} = (R + 2\,S + T\,s^{11})\,s^{12} = (T/2)\,s^{12}\,(s^{11} - s^{22}) \tag{15.83}$$

$$C^{1212} = Q + S\,(s^{11} + s^{22}) + T\,s^{12}\,s^{12} = (P + 2\,Q - T/4) + T\,s^{12}\,s^{12}. \tag{15.84}$$

Since the combination of viscosities $P + 2\,Q$ appears throughout the stress tensor, there are only two viscous degrees of freedom instead of the three present in Section

15.4.4. The choice for the tensor s^{mn} as the outer product $s^{mn} = \hat{s}^m \hat{s}^n$ eliminates one of the three degrees of freedom.

Aiming to maintain compatibility with the notation of Smith and McWilliams (2002), let us introduce the new viscosities

$$A = P + 2\,Q \tag{15.85}$$
$$D = T/2 \tag{15.86}$$

thus leading to the stress tensor components

$$\rho^{-1}\,\tau^{11} = (A - 2\,D\,s^{11}\,s^{22})\,(e_{11} - e_{22}) + 2\,D\,s^{12}\,(s^{11} - s^{22})\,e_{12} \tag{15.87}$$
$$\rho^{-1}\,\tau^{12} = D\,s^{12}\,(s^{11} - s^{22})\,(e_{11} - e_{22}) + 2\,(A - D/2 + 2\,D\,s^{12}\,s^{12})\,e_{12} \tag{15.88}$$
$$\tau^{22} = -\tau^{11}. \tag{15.89}$$

Using $s^{11} + s^{22} = 1$ and $s^{11}\,s^{22} = s^{12}\,s^{12}$ renders

$$\rho^{-1}\,\tau^{mn} = A \begin{pmatrix} e_T & e_S \\ e_S & -e_T \end{pmatrix} + D\,\Delta \begin{pmatrix} 2\,s^{12} & -(s^{11} - s^{22}) \\ -(s^{11} - s^{22}) & -2\,s^{12} \end{pmatrix} \tag{15.90}$$

where and $e_T = e_{11} - e_{22}$ and $e_S = 2\,e_{12}$ are the deformation rates introduced in Section 15.5.2, and

$$2\,\Delta = e_S\,(s^{11} - s^{22}) - 2\,e_T\,s^{12}. \tag{15.91}$$

Note that the isotropic results from Section 15.5 are recovered upon setting $D = 0$.

To help understand the nature of the new anisotropic contribution to the stress tensor

$$\tau^{mn}_{aniso} = \rho\,D\,\Delta \begin{pmatrix} 2\,s^{12} & -(s^{11} - s^{22}) \\ -(s^{11} - s^{22}) & -2\,s^{12} \end{pmatrix}, \tag{15.92}$$

it is useful to consider both its orientation and how it affects kinetic energy evolution. First, let us write the orientation vector as

$$\hat{\mathbf{s}} = (\cos\theta, \sin\theta) \tag{15.93}$$

to render

$$\tau^{mn}_{aniso} = \rho\,D\,\Delta \begin{pmatrix} \sin 2\theta & -\cos 2\theta \\ -\cos 2\theta & -\sin 2\theta \end{pmatrix} \tag{15.94}$$

and

$$2\,\Delta = e_S\,\cos 2\theta - e_T\,\sin 2\theta. \tag{15.95}$$

In the special case of zonal orientation (15.76), $\theta = 0$, whereas for the general flow dependent case (15.75), $\mathbf{s} = \mathbf{u}/|\mathbf{u}|$ is oriented according to the horizontal velocity vector. The anisotropic stress tensor is proportional to a rotation matrix

$$R^{mn}(\hat{\mathbf{u}}) = \begin{pmatrix} \sin 2\theta & -\cos 2\theta \\ -\cos 2\theta & -\sin 2\theta \end{pmatrix}, \tag{15.96}$$

and the corresponding friction is oriented according to this rotation matrix.*

*A rotation matrix is characterized by a unit determinant and inverse equal to its transpose.

The contribution from friction to kinetic energy evolution (Section 15.4.3) takes the form

$$-e_{mn}\,\tau^{mn} = -\rho\,A\,(e_T^2 + e_S^2) + 2\,\rho\,D\,\Delta^2. \tag{15.97}$$

Hence, the anisotropic stress reduces the magnitude of frictional dissipation of kinetic energy. Because Δ is proportional to the scalar dissipation term $e_{mn}\,\tau^{mn}$ when $A = 0$, Δ is also a scalar and so takes the same value regardless the coordinate system. Taking the special case $\hat{\mathbf{s}} = (1, 0)$ and $2\,\Delta = e_S$ leads to

$$-e_{mn}\,\tau^{mn} = -\rho\,A\,(e_T^2 + e_S^2) + \rho\,(D/2)\,e_S^2 \qquad \text{for } \hat{\mathbf{s}} = (1, 0). \tag{15.98}$$

Setting the anisotropic viscosity according to

$$D < 2\,A \tag{15.99}$$

ensures kinetic energy is dissipated by friction.

The above considerations allow us to determine a covariant form of the stress tensor. Summarizing the previous results, we first note that the isotropic contribution takes the covariant form $\rho\,A\,(2\,e^{mn} - g^{mn}\,e^p_p)$, as shown by equation (15.60). Second, the multiplier Δ is proportional to the scalar $e_{mn}\,\tau^{mn}$ when $A = 0$, and so is itself a scalar. Third, the rotation matrix $R(\hat{\mathbf{s}})$ is determined by an orientation vector, such as the local flow direction, which is coordinate invariant. Hence, a covariant form of the stress tensor is given by

$$\tau^{mn} = \rho\,A\,(2\,e^{mn} - g^{mn}\,e^p_p) + \rho\,D\,\Delta\,R^{mn}(\hat{\mathbf{s}}). \tag{15.100}$$

We make use of this result in Section 15.7 when deriving the friction given by the covariant divergence of the stress tensor.

15.7 Friction in generalized orthogonal coordinates

We now use the tensor analysis of Chapters 18 and 19, as summarized in Section 19.12, to derive the Laplacian frictional operator in generalized orthogonal coordinates. This section can be easily skipped for those uninterested in general tensor analysis.

15.7.1 Deformation rates

As noted in Section 15.5.2 and 15.6.2, the frictional stress tensor is a function of two independent combinations of strain tensor elements. To determine the forms of these *deformation rates* appropriate for general coordinates, let us reconsider the local effects on kinetic energy of the isotropic frictional stress tensor

$$\begin{aligned} e_{mn}\,\tau^{mn}_{iso} &= \rho\,A\,(2\,e^{mn}\,e_{mn} - g^{mn}\,e_{mn}\,e^p_p) \\ &= \rho\,A\,[(e^1_{\ 1} - e^2_{\ 2})^2 + 4\,e^1_{\ 2}\,e^2_{\ 1}]. \end{aligned} \tag{15.101}$$

We are thus motivated to introduce the deformation due to tension

$$e_T = e^1_{\ 1} - e^2_{\ 2} = e^x_{\ x} - e^y_{\ y} \tag{15.102}$$

and shears

$$e_S = 2\,(h_1/h_2)\,e^1_{\ 2} = 2\,(h_2/h_1)\,e^2_{\ 1} = 2\,e^x_{\ y} = 2\,e^y_{\ x} \tag{15.103}$$

so that

$$-e_{mn}\,\tau^{mn} = -\rho\,A\,[(e_T)^2 + (e_S)^2] + \rho\,(D/2)\,(e_S\,\cos 2\theta - e_T\,\sin 2\theta)^2. \tag{15.104}$$

Note that the Cartesian symbols x and y label physical components to tensors.* These deformation rates reduce to the forms given in Section 15.5.2 when using Cartesian coordinates on a flat manifold.

To express the deformation rates in terms of the velocity fields, we note that the strain tensor components are generally defined in terms of covariant derivatives of the velocity field (e.g., Aris (1962)). From the manipulations presented in Section 15.10.2, we are led to the expressions valid for orthogonal transverse coordinates

$$e_T = (\mathrm{d}y)\,(u/\mathrm{d}y)_{,x} - (\mathrm{d}x)\,(v/\mathrm{d}x)_{,y} \tag{15.105}$$

$$e_S = (\mathrm{d}x)\,(u/\mathrm{d}x)_{,y} + (\mathrm{d}y)\,(v/\mathrm{d}y)_{,x} \tag{15.106}$$

These forms reduce to the Cartesian case of Section 15.3.1 when the generalized orthogonal grid increments $(\mathrm{d}x, \mathrm{d}y) = (h_1\,\mathrm{d}\xi^1, h_2\,\mathrm{d}\xi^2)$ can be removed from the partial derivatives, as in the case of Cartesian coordinates on a flat space.

15.7.2 Stress tensor components

We now express the covariant form of the stress tensor (Section 15.6.2)

$$\tau^{mn} = \rho\,A\,(2\,e^{mn} - g^{mn}\,e^p_{\ p}) + \rho\,D\,\Delta\,R^{mn}(\hat{\mathbf{s}}) \tag{15.107}$$

in terms of the deformation rates. The $(1,1)$ element is given by

$$\begin{aligned}\rho^{-1}\,\tau^{11} &= A\,(2\,e^{11} - g^{11}\,e^1_{\ 1} - g^{11}\,e^2_{\ 2}) + D\,\Delta\,R^{11}\\ &= g^{11}\,[A\,(e^1_{\ 1} - e^2_{\ 2}) + D\,\Delta\,R^1_{\ 1}],\end{aligned} \tag{15.108}$$

thus leading to the the physical component of the stress tensor

$$\tau^{xx} = h_1\,h_1\,\tau^{11} = \rho\,(A\,e_T + D\,\Delta\,R^x_{\ x}) \tag{15.109}$$

where the Cartesian symbols x and y label physical components to tensors. Likewise,

$$\tau^{yy} = h_2\,h_2\,\tau^{22} = \rho\,(-A\,e_T + D\,\Delta\,R^y_{\ y}). \tag{15.110}$$

For the off-diagonal term, we have

$$\begin{aligned}\rho^{-1}\,\tau^{12} &= 2\,A\,e^{12} + D\,\Delta\,R^{12}\\ &= 2\,(h_1\,h_2)^{-1}\,A\,e^x_{\ y} + (h_2\,h_2)^{-1}\,D\,\Delta\,R^1_{\ 2}\end{aligned} \tag{15.111}$$

*As discussed in Section 19.11.1, physical components of tensors have dimensions that are physically meaningful, whereas tensorial components of tensors can have arbitrary dimensions depending on the choice of coordinates. Physical components do not transform in the convenient manner that tensorial components do. Hence, it is best to work with the more abstract tensorial components for mathematical manipulations. When reaching a point where one aims to clarify the dimensions and physical content of a mathematical statement, and to clean up expressions prior to discretization, then it is relevant to introduce physical tensor components.

thus leading to

$$\tau^{xy} = \rho\,(A\,e_S + D\,\Delta\,R^x_y). \tag{15.112}$$

In summary, we have the physical component version of the stress tensor given by

$$\tau^{(m)(n)} = \begin{pmatrix} \tau^{xx} & \tau^{xy} \\ \tau^{xy} & -\tau^{xx} \end{pmatrix} = \rho \begin{pmatrix} (A\,e_T + D\,\Delta\,R^x_x) & (A\,e_S + D\,\Delta\,R^x_y) \\ (A\,e_S + D\,\Delta\,R^y_x) & (-A\,e_T + D\,\Delta\,R^y_y) \end{pmatrix} \tag{15.113}$$

and the rotation matrix

$$R^{(m)}_{(n)} = \begin{pmatrix} \sin 2\theta & -\cos 2\theta \\ -\cos 2\theta & -\sin 2\theta \end{pmatrix}. \tag{15.114}$$

The parentheses notation is used to distinguish physical components of tensors, which use the Cartesian labels x and y, from the more abstract tensorial components, which use the numerals 1 and 2 (see Section 19.11.1 and footnote following equation (15.104)). Note that when using the flow dependent orientation as in Smith and McWilliams (2002), the orientation of the physical velocity components

$$\mathbf{u} = (u, v) = (h_1\,u^1, h_2\,u^2) \tag{15.115}$$

serves to orient the anisotropic portion of the stress tensor according to equation (15.74), rather than the dimensionally ambiguous tensorial components (u^1, u^2). In spherical coordinates

$$(u, v) = R\,(\cos\phi\,\mathrm{d}\lambda/\mathrm{d}t, \mathrm{d}\phi/\mathrm{d}t) \tag{15.116}$$

are the familiar zonal and meridional velocity components, with $\mathrm{d}/\mathrm{d}t$ the material time derivative.

15.7.3 The friction vector

The friction arising from horizontal and vertical deformations is given by the covariant divergence of the frictional stress tensor. From Section 19.3, we can write the covariant divergence as

$$\begin{aligned} \rho\,F^m &= \widetilde{\tau}^{mn}_{;n} \\ &= \widetilde{\tau}^{mn}_{,n} + \Gamma^m_{nc}\,\widetilde{\tau}^{nc} + \Gamma^n_{nc}\,\widetilde{\tau}^{mc} \end{aligned} \tag{15.117}$$

where $\widetilde{\tau}^{mn}$ is the frictional stress tensor, whose tranverse portion is given by τ^{mn} (Section 15.3.4). For implementation in an ocean model, it is necessary to massage equation (15.117). We provide details of these steps in Section 15.10.3. The final results are the physical friction components

$$\rho\,F^x = (\mathrm{d}y)^{-2}\,[(\mathrm{d}y)^2\,\tau^{xx}]_{,x} + (\mathrm{d}x)^{-2}\,[(\mathrm{d}x)^2\,\tau^{xy}]_{,y} + (\rho\,\kappa\,u_{,z})_{,z} \tag{15.118}$$

$$\rho\,F^y = (\mathrm{d}x)^{-2}\,[(\mathrm{d}x)^2\,\tau^{yy}]_{,y} + (\mathrm{d}y)^{-2}\,[(\mathrm{d}y)^2\,\tau^{xy}]_{,x} + (\rho\,\kappa\,v_{,z})_{,z}, \tag{15.119}$$

where

$$\tau^{xx} = -\tau^{yy}. \tag{15.120}$$

For Boussinesq fluids, the factors of density can be canceled on both sides of the expressions for friction, since ρ is formally replaced by the constant reference density ρ_o.

15.8 Dissipation functional

The purpose of this section is to introduce the dissipation functional corresponding to the friction operator. This functional is fundamental to the method used to discretize friction. The formalism is analogous to how the tracer variance is related to the diffusion operator (see Griffies et al. (1998) and Section 14.1). A discussion of the dissipation functional was given by Griffies and Hallberg (2000), Dukowicz and Baumgardner (2002), and Smith and McWilliams (2002).

15.8.1 Motivation for the functional approach

There are many ways to discretize an operator. One approach is to take the continuum form and provide reasonable second order discretizations, with averaging used whenever needed to place quantities in their proper grid position. This was the common approach taken for friction in the past, with the continuum formulas of Bryan (1969) and Wajsowicz (1993), or Smagorinsky (1963, 1993), the starting point. This direct approach is not always satisfying since it may produce a discrete operator with undesirable properties. An example of the problems with naive discretizations is described by Griffies et al. (1998) for the neutral diffusion operator.

After having some success with the reformulation of neutral diffusion, in which the diffusion operator was shown to be equivalent to the functional derivative of the tracer variance, it was decided to attempt a similar approach for friction. As shown in Section 15.8.2, the friction operator is equivalent to the functional derivative of the local kinetic energy dissipation. Both of these results follow from the self-adjointness of the diffusion and friction operators. Through this connection, the approach is to first discretize the appropriate functional. Thereafter, the functional derivative of the discrete functional is used to derive the discretized friction operator.

It is notable that discretization of the friction operator via first discretizing the functional with second order numerics leads to the use of velocity triads. These triads result because of the B-grid in which both velocity components are at a single point. For second order discretization of the neutral diffusion operator, density triads are fundamental (Griffies et al. (1998)). Triads emerge for diffusion since all quantities of interest live on the tracer grid. Hence, discrete neutral diffusion, derived from a functional, contains triads on any grid.

15.8.2 Defining the functional

As noted in Section 15.5.2 and 15.6.2, the effects on local kinetic energy dissipation from horizontal deformations in the fluid take the form

$$-e_{mn}\,\tau^{mn} = -\rho\,A\,(e_T^2 + e_S^2) + 2\,\rho\,D\,\Delta^2, \tag{15.121}$$

where

$$2\,\Delta = e_S\,\cos 2\theta - e_T\,\sin 2\theta. \tag{15.122}$$

The angle θ sets the orientation of the friction. The first term in equation (15.121) is associated with transverse isotropic stresses, whereas the second term arises from anisotropic stresses. Ensuring a sign-definite dissipation of kinetic energy requires

$$D < 2\,A. \tag{15.123}$$

Integration over the full ocean domain leads to the positive semi-definite *dissipation functional*

$$\mathcal{S} \equiv \int \mathrm{d}V \, e_{mn} \, \tau^{mn} = \int \rho \, \mathrm{d}V \, [A \, (e_T^2 + e_S^2) - 2 \, D \, \Delta^2] \tag{15.124}$$

where

$$\mathrm{d}V = \sqrt{\mathcal{G}} \, \mathrm{d}\xi^1 \, \mathrm{d}\xi^2 \, \mathrm{d}z \tag{15.125}$$

is the invariant volume element written using generalized horizontal coordinates (Section 18.11), and $\sqrt{\mathcal{G}} = h_1 \, h_2 \, h_3$ is the square-root of the metric determinant (Section 19.12). With the Traditional Approximation (Section 3.1), the vertical stretching funtion $h_3 = 1$.

As shown in this section, the functional derivative $\delta\mathcal{S}/\delta u^a$ is proportional to the friction $g_{ab} \, F^b$. The connection between a dissipation functional and the friction operator is afforded through the self-adjointness of the friction operator. This result leads to a numerical discretization of friction that is ensured to dissipate kinetic energy on the discrete lattice. The same approach for the tracer diffusion operator is detailed in Section 14.1.2. Novelty is afforded in this section by the introduction of vector velocities in the operator, instead of scalar tracers.

Two assumptions are needed to make the friction operator self-adjoint: (1) The viscosity is functionally independent of the velocity field. (2) The flow satisfies *natural boundary conditions*, of which no-slip is one. The Smagorinsky viscosity does not satisfy the first assumption. Nonetheless, the functional approach leads to a discretization inside of which one can employ the Smagorinsky viscosity. A similar assumption was used to discretize the neutral diffusion operator when diffusing active tracers, where in that case the operator is a nonlinear function of the active tracers.

15.8.3 Performing the functional derivative

In this subsection, we perform the functional derivative, thus providing details to the above statements. For this purpose, let us write the dissipation functional as

$$\mathcal{S} = \int \mathrm{d}V \, \mathcal{L} \tag{15.126}$$

with

$$\mathcal{L} = e_{mn} \, \tau^{mn} = \rho \, A \, (e_T^2 + e_S^2) - 2 \, \rho \, D \, \Delta^2. \tag{15.127}$$

If we now consider a variation in the tensorial components to the velocity field

$$u^a \to u^a + \delta u^a, \tag{15.128}$$

with the underlying space-time geometry held fixed, then the dissipation functional $\mathcal{S}$ changes according to

$$\delta \, \mathcal{S} = \int \mathrm{d}V \, \delta\mathcal{L}. \tag{15.129}$$

Since $\mathcal{L} = \mathcal{L}[u^a, u^a_{,b}]$ is a functional of the velocity and its partial derivatives, its variation leads to

$$\begin{aligned}\delta\,\mathcal{S} &= \int \sqrt{\mathcal{G}}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}z\,\left[\frac{\delta\,\mathcal{L}}{\delta u^a}\,\delta u^a + \frac{\delta\,\mathcal{L}}{\delta u^a_{,b}}\,\delta u^a_{,b}\right] \\ &= \int \mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}z\,\left[\sqrt{\mathcal{G}}\,\frac{\delta\,\mathcal{L}}{\delta u^a}\,\delta u^a + \partial_b\left(\sqrt{\mathcal{G}}\,\frac{\delta\,\mathcal{L}}{\delta u^a_{,b}}\,\delta u^a\right) - \partial_b\left(\sqrt{\mathcal{G}}\,\frac{\delta\,\mathcal{L}}{\delta u^a_{,b}}\right)\,\delta u^a\right],\end{aligned} \tag{15.130}$$

where an integration by parts has been performed. The total derivative reduces to a surface term, which vanishes when either

$$\delta u^a = 0 \tag{15.131}$$

on all solid boundaries, or

$$\hat{n}_b\,(\delta\,\mathcal{L}/\delta u^a_{,b})\,\delta u^a = 0, \tag{15.132}$$

where $\hat{n}_b$ are components to the outward normal at the boundaries. These two conditions define the natural boundary conditions mentioned above. If velocity, and hence its variation, satisfies the no-slip condition, then $\delta u^a = 0$ on all boundaries and the total derivative can be dropped. More general boundary conditions can be derived from the second type of natural boundary condition, yet they are not considered here since all B-grid ocean climate models employ no-slip on the side boundaries.

With natural boundary conditions, variation of the dissipation functional takes the form

$$\delta\,\mathcal{S} = \int \sqrt{\mathcal{G}}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}z\,\left[\frac{\delta\,\mathcal{L}}{\delta u^a} - \mathcal{G}^{-1/2}\,\partial_b\left(\sqrt{\mathcal{G}}\,\frac{\delta\,\mathcal{L}}{\delta u^a_{,b}}\right)\right]\,\delta u^a, \tag{15.133}$$

thus leading to the functional derivative

$$(\mathrm{d}V(\mathbf{y}))^{-1}\,\frac{\delta\,\mathcal{S}}{\delta u^a} = \frac{\delta\,\mathcal{L}}{\delta u^a} - \mathcal{G}^{-1/2}\,\partial_b\left(\sqrt{\mathcal{G}}\,\frac{\delta\,\mathcal{L}}{\delta u^a_{,b}}\right), \tag{15.134}$$

where $\mathrm{d}V(\mathbf{y})$ is the volume element evaluated at the field point $\mathbf{y}$. To reach this result, it was necessary to use the identity

$$\frac{\delta u^a(\mathbf{x})}{\delta u^b(\mathbf{y})} = \mathrm{d}V(\mathbf{y})\,\delta^a_b\,\delta(\mathbf{x}-\mathbf{y}), \tag{15.135}$$

where $\delta(\mathbf{x}-\mathbf{y})$ is the Dirac delta-function. The delta-function has physical dimensions of inverse volume L^{-3}. Hence,

$$\int \mathrm{d}V(\mathbf{y})\,\delta(\mathbf{x}-\mathbf{y}) = 1, \tag{15.136}$$

so long as integration is over a domain containing the singular point $\mathbf{x} = \mathbf{y}$; otherwise, the integral vanishes. The derivation of equation (15.134) should be familiar to those having studied the Euler-Lagrange equations used in classical mechanics (e.g., Marion and Thornton (1988)).

15.8.4 The connection to friction

Now that the functional derivative of the dissipation $\mathcal{S}$ has been computed, it remains to prove its connection to friction. For this purpose, we write a general variation of $\mathcal{L}$ as

$$\begin{aligned}\delta\mathcal{L} &= 2\,\rho\,[\delta e_T\,(A\,e_T + D\,\Delta\,\sin 2\theta) + \delta e_S\,(A\,e_S - D\,\Delta\,\cos 2\theta)] \\ &= 2\,(\tau^{xx}\,\delta e_T + \tau^{xy}\,\delta e_S)\end{aligned} \tag{15.137}$$

where the second step introduced the expressions for the stress tensor components given in Section 15.7.2. Now recall the discussion in Section 15.10.2, where it was shown that the horizontal tension can be written

$$e_T = u^1{}_{,1} - u^2{}_{,2} + u^m\,\partial_m\,\ln(h_1/h_2) \tag{15.138}$$

and the horizontal shearing strain is

$$e_S = \frac{h_1}{h_2}\,u^1{}_{,2} + \frac{h_2}{h_1}\,u^2{}_{,1}. \tag{15.139}$$

Consequently, the deformation rates have functional derivatives given by

$$\frac{\delta e_T}{\delta u^a} = \delta^m_a\,\partial_m\,\ln(h_1/h_2) \tag{15.140}$$

$$\frac{\delta e_T}{\delta u^a{}_{,b}} = \delta^b_1\,\delta^1_a - \delta^b_2\,\delta^2_a \tag{15.141}$$

$$\frac{\delta e_S}{\delta u^a} = 0 \tag{15.142}$$

$$\frac{\delta e_S}{\delta u^a{}_{,b}} = (h_1/h_2)\,\delta^b_2\,\delta^1_a + (h_2/h_1)\,\delta^b_1\,\delta^2_a \tag{15.143}$$

thus leading to

$$\rho^{-1}\,\frac{\delta\mathcal{L}}{\delta u^a} = 2\,\tau^{xx}\,\delta^m_a\,\partial_m\,\ln(h_1/h_2) \tag{15.144}$$

$$\rho^{-1}\,\frac{\delta\mathcal{L}}{\delta u^a_{,b}} = 2\,\tau^{xx}\,(\delta^b_1\,\delta^1_a - \delta^b_2\,\delta^2_a) + 2\,\tau^{xy}\,\left(\frac{h_1}{h_2}\,\delta^1_a\,\delta^b_2 + \frac{h_2}{h_1}\,\delta^2_a\,\delta^b_1\right). \tag{15.145}$$

For $a = 1$, these results yield

$$\begin{aligned}-\frac{1}{2\,\mathrm{d}V}\,\frac{\delta\,\mathcal{S}}{\delta u^1} &= \tau^{xx}\,\partial_1\,\ln(h_1/h_2) + \frac{1}{h_1\,h_2}\,\partial_b\left(h_1\,h_2\,\delta^b_1\,\tau^{xx} + (h_1/h_2)\,\delta^b_2\,\tau^{xy}\right) \\ &= h_1\,\left((\mathrm{d}y)^{-2}\,[(\mathrm{d}y)^2\,\tau^{xx}]_{,x} + (\mathrm{d}x)^{-2}\,[(\mathrm{d}x)^2\,\tau^{xy}]_{,y}\right) \\ &= \rho\,h_1\,F^x.\end{aligned} \tag{15.146}$$

The intermediate step utilized manipulations described in Section 15.10, and the last step identified the friction vector (sans the piece arising from vertical shears) given in Section 15.7.3. Hence, in general,

$$-\frac{1}{2\,\mathrm{d}V}\,\frac{\delta\mathcal{S}}{\delta u^{(a)}} = \rho\,F^{(a)} \tag{15.147}$$

where the parentheses indicate the physical tensor components (see Section 19.11.1 for discussion of physical tensor components). This result establishes the desired connection between the functional derivative of the dissipation functional, and the friction operator. The discrete analog of this result is employed in Section 17.3 to derive a discrete version of the friction operator.

15.9 Biharmonic friction

The previous derivations were all concerned with second order, or *Laplacian*, friction. It is often useful to consider another method of dissipating momentum through use of a fourth order, or *biharmonic*, friction. As reviewed by Griffies and Hallberg (2000), biharmonic friction acts more strongly on the small scales than the Laplacian friction, and less strongly on the large scales. Each property is desirable, especially when aiming to realize some form of a quasi-geostrophic turbulent cascade in which enstrophy cascades to the small scales and energy to the large scales. Biharmonic friction is therefore commonly used in eddying ocean simulations.

The goal of this section is to derive the appropriate form of the biharmonic operator that both dissipates kinetic energy yet does not introduce spurious sources of angular momentum. As with the previous derivations, some work is necessary in order to realize these properties on a sphere with a generally non-constant viscosity.

Recall that the quasi-hydrostatic approximation allowed for the separation of the transverse or horizontal subspace from the vertical subspace. Consequently, friction arising from the vertical gradient of the vertical strain, $(\kappa \mathbf{u}_{,z})_{,z}$, was isolated from gradients in the horizontal deformations. As a result, the following focuses solely on deriving the biharmonic operator for shears arising in the two-dimensional transverse subspace. Hence, all labels in this section only run over the transverse space $m, n, p = 1, 2$.

The utility of an anisotropic biharmonic friction operator has not been established. It is trivial to implement an anisotropic biharmonic friction, once the formalism for the corresponding Laplacian operator is established. For brevity, the following focuses on isotropic biharmonic friction.

15.9.1 General formulation

The general formulation of biharmonic friction is a straightforward extension of the Laplacian friction given in the previous sections. What is done is to basically iterate twice on the Laplacian approach. More precisely, the components F_B^m of the biharmonic friction vector are derived from the covariant divergence*

$$\rho\, F_B^m = \Theta^{mn}_{;n}, \tag{15.148}$$

where

$$\Theta^{mn} = -\rho\, B\, (2\, E^{mn} - g^{mn}\, E^p_p), \tag{15.149}$$

$B > 0$ has units of $L^2\, t^{-1/2}$, and ρ is set to the constant reference density ρ_o when making the Boussinesq approximation. As shown in the next subsection, use of

*Recall that all labels in this section run over $m, n, p = 1, 2$.

the "square-root" biharmonic viscosity is prompted by the desire to dissipate kinetic energy. This detail only matters for cases with a non-constant viscosity. Θ^{mn} has the same form as the stress tensor used for the isotropic second order friction discussed in Section 15.7.2, except with a minus sign. The components of the symmetric "strain" tensor are given by

$$E_{mn} = \frac{1}{2}\,(F_{m;n} + F_{n;m}). \tag{15.150}$$

F^m are components to the friction vector determined through the covariant divergence of the second order frictional stress tensor

$$\begin{aligned} \rho\, F^m &= \tau^{mn}_{;n} \\ &= [B\,\rho\,(2\,e^{mn} - g^{mn}\,e^p_p)]_{;n} \end{aligned} \tag{15.151}$$

as derived in the previous sections, where the only difference is that the viscosity used for computing the stress tensor τ^{mn} is now set to B, and the dimensions on F^m are $L\,t^{-3/2}$.

This approach ensures that the biharmonic friction is derived from the divergence of a symmetric tensor Θ^{mn}, hence ensuring a proper angular momentum budget. Additionally, the computational work necessary to compute the Laplacian friction is directly employed for the biharmonic friction. Finally, as shown in the next subsection, this form for biharmonic friction also dissipates kinetic energy.

15.9.2 Effects on kinetic energy

The manipulations necessary to show that the biharmonic friction dissipates kinetic energy are analogous to those used for second order friction in Section 15.7. As with that discussion, the relevant contribution from horizontal biharmonic friction is given by $\mathrm{d}V\,u_m\,\Theta^{mn}_{;n}$. Assuming either no-slip or no-normal Θ stress at the boundaries brings this expression to the form

$$\begin{aligned} \int \mathrm{d}V\,u_m\,\Theta^{mn}_{;n} &= -\int \mathrm{d}V\,\Theta^{mn}\,e_{mn} \\ &= \int \mathrm{d}V\,\rho\,B\,[2\,E^{mn}\,e_{mn} - e^n_n\,E^m_m]. \end{aligned} \tag{15.152}$$

For the product of traces, one has

$$\begin{aligned} \int \mathrm{d}V\,\rho\,B\,e^n_n\,E^m_m &= \int \mathrm{d}V\,\rho\,B\,e^n_n\,F^m_{;m} \\ &= \int \mathrm{d}V\,[(\rho\,B\,e^n_n\,F^m)_{;m} - (\rho\,B\,e^n_n)_{;m}\,F^m]. \end{aligned} \tag{15.153}$$

The first term reduces to a boundary contribution, which vanishes if we assume the second order friction F^m vanishes on the boundaries. For the contraction of the two strain tensors, one has

$$\begin{aligned} 2\int \mathrm{d}V\,\rho\,B\,e^{mn}\,E_{mn} &= 2\int \mathrm{d}V\,\rho\,B\,e^{mn}\,F_{m\,;n} \\ &= -2\int \mathrm{d}V\,F_m\,(\rho\,B\,e^{mn})_{;n}, \end{aligned} \tag{15.154}$$

where again the boundary term $(\rho\, B\, e^{mn}\, F_m)_{;n}$ was assumed to vanish. Combining the two contributions leads to

$$\begin{aligned}\int \mathrm{d}V\, u_m\, \Theta^{mn}_{;n} &= -\int \mathrm{d}V\, F_m\, [2\, B\, \rho\, e^{mn} - g^{mn}\, B\, \rho\, e^k_k]_{;n} \\ &= -\int \mathrm{d}V\, \rho\, F_m\, F^m, \end{aligned} \tag{15.155}$$

which is non-positive, thus ensuring that kinetic energy is dissipated by the friction for non-trivial flows.

If the viscosity B is distributed non-symmetrically, then kinetic energy is guaranteed to be dissipated by the biharmonic friction only for the special case of constant viscosity. That is, in Cartesian coordinates, the operator $\nabla_z \cdot B\, \nabla_z\, (\nabla_z \cdot B \nabla_z \psi)$ is dissipative for all $B > 0$, whereas $\nabla_z \cdot B^2\, \nabla_z\, (\nabla_z^2 \psi)$ or $\nabla_z^2\, (\nabla_z \cdot B^2 \nabla_z \psi)$ can be proven to be dissipative only for constant B.

15.10 Some mathematical details

The purpose of this section is to present some of the steps in the derivation of the deformation rates and Laplacian friction operator. The material may be easily skipped for those uninterested in mathematical details. Recognizing that many of the equations are lengthy, we use a smaller font for this section to allow for more compact expressions.

15.10.1 Salient results from Chapters 18 and 19

The infinitesimal distance between two points using orthogonal coordinates on the sphere is given by

$$\begin{aligned}(\mathrm{d}s)^2 &= g_{mn}\, \mathrm{d}\xi^m\, \mathrm{d}\xi^n \\ &= (h_1\, \mathrm{d}\xi^1)^2 + (h_2\, \mathrm{d}\xi^2)^2 + (h_3\, \mathrm{d}\xi^3)^2 \\ &= (\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2, \end{aligned} \tag{15.156}$$

where we introduced the physical displacements

$$(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = (h_1\, \mathrm{d}\xi^1, h_2\, \mathrm{d}\xi^2, h_3\, \mathrm{d}\xi^3) \tag{15.157}$$

whose dimensions are length. Within the Traditional Approximation (Section 3.1), $h_3 = 1$ and the transverse stretching functions h_1, h_2 are dependent only on the transverse coordinates (ξ^1, ξ^2). In the special case of geographical spherical coordinates, $(\xi^1, \xi^2) = (\lambda, \phi)$ are the dimensionless longitude and latitude, and $(h_1, h_2) = R\, (\cos\phi, 1)$ are the stretching functions with R the earth's radius.

The determinant of the metric tensor g_{mn} is positive definite. Using orthogonal coordinates, the determinant is given by

$$\mathcal{G} = g_{11}\, g_{22}\, g_{33} = (h_1\, h_2\, h_3)^2, \tag{15.158}$$

where $h_3 = 1$ with the Traditional Approximation. The metric determinant appears in certain differential formulae, such as the invariant volume element (Section 18.11).

We will also find use in this section for the physical components to the transverse partial derivative operator

$$(\partial_x, \partial_y) = (h_1^{-1}\, \partial_1, h_2^{-1}\, \partial_2), \tag{15.159}$$

and the physical velocity components

$$(u, v) = (h_1\, u^1, h_2\, u^2). \tag{15.160}$$

In spherical coordinates

$$(u, v) = R\,(\cos\phi\,\mathrm{d}\lambda/\mathrm{d}t, \mathrm{d}\phi/\mathrm{d}t) \tag{15.161}$$

are the familiar zonal and meridional velocity components, with $\mathrm{d}/\mathrm{d}t$ the material time derivative. Likewise,

$$(\partial_x, \partial_y) = ((R\,\cos\phi)^{-1}\,\partial_\lambda, (R)^{-1}\,\partial_\phi) \tag{15.162}$$

are the familiar partial derivative operators. A full discussion of physical tensor components using orthogonal coordinates is provided in Section 19.12.3. See also the summary discussion in Section 19.12.

15.10.2 Details of the deformation rate derivation

The methods described in Chapters 18 and 19, as summarized above, lead to

$$\begin{aligned}
e_T &= e^1_{\ 1} - e^2_{\ 2} \\
&= u^1_{\ ;1} - u^2_{\ ;2} \\
&= u^1_{\ ,1} - u^2_{\ ,2} + u^m\,(\Gamma^1_{1m} - \Gamma^2_{2m}) \\
&= u^1_{\ ,1} - u^2_{\ ,2} + \frac{1}{2}\,u^m\,\partial_m\,\ln(g_{11}/g_{22}) \\
&= (h_2/h_1)\,\left(u^1\,(h_1/h_2)\right)_{,1} - (h_1/h_2)\,\left(u^2\,(h_2/h_1)\right)_{,2} \\
&= h_2\,(u/h_2)_{,x} - h_1\,(v/h_1)_{,y} \\
&= (\mathrm{d}y)\,(u/\mathrm{d}y)_{,x} - (\mathrm{d}x)\,(v/\mathrm{d}x)_{,y}
\end{aligned} \tag{15.163}$$

where we eliminated the stretching functions h_1 and h_2 in the last step via the properties articulated in Section 19.11.2. Additionally, Γ^p_{mn} are the Christoffel symbols defined in Section 19.2.2. They can be written in terms of the metric tensor as (Section 19.4)

$$\Gamma^p_{mn} = \frac{1}{2}\,g^{pq}(g_{qm,n} + g_{qn,m} - g_{mn,q}). \tag{15.164}$$

Proceeding in a similar manner leads to

$$\begin{aligned}
(h_1/h_2)\,e_S &= 2\,e^2_{\ 1} \\
&= u^2_{\ ;1} + g^{22}\,g_{11}\,u^1_{\ ;2} \\
&= u^2_{\ ,1} + \Gamma^2_{1m}\,u^m + g^{22}\,g_{11}\,(u^1_{\ ,2} + \Gamma^1_{2m}\,u^m) \\
&= u^2_{\ ,1} + \frac{1}{2}\,g^{2d}\,(g_{1d,m} + g_{md,1} - g_{1m,d})\,u^m + g^{22}\,g_{11}\,u^1_{\ ,2} \\
&+ \frac{1}{2}\,g^{22}\,g_{11}\,g^{1d}\,(g_{2d,m} + g_{md,2} - g_{2m,d})\,u^m \\
&= g^{22}\,(g_{11}\,u^1_{\ ,2} + g_{22}\,u^2_{\ ,1}),
\end{aligned} \tag{15.165}$$

which brings the horizontal shearing strain to

$$\begin{aligned}
e_S &= h_1\,(u/h_1)_{,y} + h_2\,(v/h_2)_{,x} \\
&= (\mathrm{d}x)\,(u/\mathrm{d}x)_{,y} + (\mathrm{d}y)\,(v/\mathrm{d}y)_{,x}
\end{aligned} \tag{15.166}$$

15.10.3 Details of the friction derivation

To deduce the form of the friction vector appropriate for an ocean model, it is necessary to massage equation (15.117). For this purpose, we start from the equivalent expression

$$\rho\,F^m = (\sqrt{\mathcal{G}})^{-1}\,\left(\sqrt{\mathcal{G}}\,\tau^{mn}\right)_{,n} + \Gamma^m_{ab}\,\tau^{ab}, \tag{15.167}$$

which is valid for any metric. Its derivation is given in Section 19.6. We now exploit the diagonal form of the metric tensor. First, the contraction $\Gamma^n_{ab}\,\tau^{ab}$ is given by

$$\begin{aligned}\Gamma^m_{ab}\,\tau^{ab} &= \frac{1}{2}\,g^{md}\,(g_{ad,b}+g_{bd,a}-g_{ab,d})\,\tau^{ab}\\ &= g^{mm}\,(g_{am,b}-\frac{1}{2}\,g_{ab,m})\,\tau^{ab}\\ &= g^{mm}\,g_{mm,b}\,\tau^{mb}-\frac{1}{2}\,g^{mm}\,g_{ab,m}\,\tau^{ab}\end{aligned} \tag{15.168}$$

where there is no sum on the m label. Plugging this result into equation (15.167) for the friction vector yields

$$\rho\,F^m = (g_{mm}\,\sqrt{\mathcal{G}})^{-1}\,\left(g_{mm}\,\sqrt{\mathcal{G}}\,\tau^{mn}\right)_{,n}-\frac{1}{2}\,g^{mm}\,g_{ab,m}\,\tau^{ab}. \tag{15.169}$$

Now recall that the transverse stress tensor considered here has a zero trace (Section 15.3.5), which in general coordinates is given by the invariant expression

$$g_{mn}\,\tau^{mn} = \tau^m_m = \tau^1_1+\tau^2_2 = 0. \tag{15.170}$$

Consequently, the generalized friction component with $m=1$ is given by

$$\begin{aligned}\rho\,F^1 &= (g_{11}\,\sqrt{\mathcal{G}})^{-1}\,\left(g_{11}\,\sqrt{\mathcal{G}}\,\tau^{1n}\right)_{,n}-\frac{1}{2}\,g^{11}\,g_{ab,1}\,\tau^{ab}\\ &= \tau^{1n}_{,n}+\tau^{1n}\,\partial_n\,\ln(g_{11}\,\sqrt{\mathcal{G}})-\tau^{11}\,\partial_1\,\ln\sqrt{g_{11}}-\frac{1}{2}\,\tau^{22}\,g^{11}\,g_{22,1}\\ &= \tau^{11}_{,1}+\tau^{11}\,\partial_1\,\ln(g_{11}\,\sqrt{\mathcal{G}})+(g_{11}\,\sqrt{\mathcal{G}})^{-1}(g_{11}\,\sqrt{\mathcal{G}}\,\tau^{12}\,)_{,2}-\tau^{11}\,\partial_1\,\ln\sqrt{g_{11}}+\tau^{11}\,\partial_1\,\ln\sqrt{g_{22}}\\ &= \tau^{11}_{,1}+\tau^{11}\,\partial_1\,\ln\mathcal{G}+(g_{11}\,\sqrt{\mathcal{G}})^{-1}(g_{11}\,\sqrt{\mathcal{G}}\,\tau^{12}\,)_{,2}\\ &= \mathcal{G}^{-1}\,(\mathcal{G}\,\tau^{11})_{,1}+(g_{11}\,\sqrt{\mathcal{G}})^{-1}(g_{11}\,\sqrt{\mathcal{G}}\,\tau^{12}\,)_{,2}\\ &= [h_1\,(\mathrm{d}y)^2]^{-1}\,[(\mathrm{d}y)^2\,h_1\,h_1\,\tau^{11}]_{,x}+[h_1\,(\mathrm{d}x)^2]^{-1}\,[(\mathrm{d}x)^2\,h_1\,h_2\,\tau^{12}]_{,y}\end{aligned} \tag{15.171}$$

Multiplying by $h_1 = \sqrt{g_{11}}$ determines the physical component to the generalized zonal friction

$$\rho\,F^x = (\mathrm{d}y)^{-2}\,[(\mathrm{d}y)^2\,\tau^{xx}]_{,x}+(\mathrm{d}x)^{-2}\,[(\mathrm{d}x)^2\,\tau^{xy}]_{,y}+(\rho\,\kappa\,u_{,z})_{,z} \tag{15.172}$$

where we reintroduced the contribution due to vertical strains, and the physical components to the friction and stress tensor are given by

$$F^x = h_1\,F^1 \tag{15.173}$$

$$\tau^{xx} = h_1\,h_1\,\tau^{11} \tag{15.174}$$

$$\tau^{xy} = h_1\,h_2\,\tau^{12}. \tag{15.175}$$

Letting $x \leftrightarrow y$ leads to the generalized meridional friction component

$$\rho\,F^y = (\mathrm{d}x)^{-2}\,[(\mathrm{d}x)^2\,\tau^{yy}]_{,y}+(\mathrm{d}y)^{-2}\,[(\mathrm{d}y)^2\,\tau^{xy}]_{,x}+(\rho\,\kappa\,v_{,z})_{,z}. \tag{15.176}$$

CHAPTER
SIXTEEN

Horizontal viscosity

Contents

16.1 Stability and resolution considerations 359
16.1.1 Balance between advection and diffusion 360
16.1.2 Linear stability of the diffusion equation 361
16.1.2.1 Laplacian mixing 361
16.1.2.2 Biharmonic mixing 363
16.1.3 Western boundary currents 363
16.1.4 Computational modes on the B-grid 364
16.1.5 Considerations for viscosity on the sphere 364
16.1.6 Comparing Laplacian and biharmonic mixing 365
16.2 Smagorinsky viscosity 367
16.2.1 General ideas and motivation 367
16.2.2 The dimensionless Smagorinsky scaling parameter 368
16.2.3 Choosing the scaling coefficient 369
16.2.4 What about tracers? 370
16.2.5 Biharmonic Smagorinsky friction 371
16.3 Background viscosity 371
16.4 Viscosities for anisotropic friction 372

The purpose of this chapter is to discuss how to choose a viscosity for use in the friction schemes. For simplicity, we consider only the Boussinesq expressions here, though the ideas are relevant whether using Boussinesq or non-Boussinesq equations.

16.1 Stability and resolution considerations

The purpose of this section is to review some of the basic numerical considerations that constrain the viscosity.

16.1.1 Balance between advection and diffusion

The representation of advection in a numerical model is nontrivial. In practice, most concern has gone into providing some options for advecting tracers. Advection of momentum is still commonly done with the centered difference scheme used by Bryan (1969). Centered advection of momentum provides for an energetically consistent scheme (see Bryan (1969) and Griffies et al. (2003)), which in turn eliminates the nonlinear instability of Phillips (1949).

The purpose of this section is to summarize the issues concerning the use of centered differenced advection. Most notably, it will be shown that some amount of mixing is needed in order to eliminate a computational mode. The arguments are taken from Appendix B of Bryan et al. (1975).

Consider the steady state balance between one-dimensional advection and diffusion

$$U\,\psi_{,x} = A\,\psi_{,xx} \tag{16.1}$$

where U is a constant advection velocity, ψ is any field that is advected, including velocity and tracers, and A is a Laplacian viscosity or diffusivity. Using centered differences in space on a constant grid of size Δ, the finite difference counterpart to this equation takes the form

$$(R-2)\,\psi_{i+1} + 4\,\psi_i - (R+2)\,\psi_{i-1} = 0, \tag{16.2}$$

where

$$R = \frac{U\,\Delta}{A}. \tag{16.3}$$

When ψ is one of the velocity components, R is the grid Reynolds number, where the qualifier "grid" is used since the length scale is the grid scale. When ψ is a tracer, R is the grid Peclet number. In either case, R measures the ratio of advection to diffusion.

A constant solves the finite difference equation (16.2). In addition, a power solves it

$$\psi_i = C\,\xi^i, \tag{16.4}$$

where C is a constant and the i superscript on the right-hand side represents a power. Substituting this function into the finite difference equation yields the quadratic equation

$$(R-2)\,\xi^2 + 4\,\xi - (R+2) = 0. \tag{16.5}$$

The two real roots to this equation are

$$\xi = 1 \tag{16.6}$$

$$\xi = \frac{2+R}{2-R}. \tag{16.7}$$

The first root is a constant as already mentioned. The second root is more relevant. With $\psi_i = C\,\xi^i$, if ξ^i is negative, then ψ_i oscillates in space with a wave length

given by the grid size. Such behaviour is not physical and so should be avoided. Ensuring that $R < 2$, or

$$\frac{U\,\Delta}{A} < 2 \tag{16.8}$$

keeps the second root positive and so eliminates the unphysical behaviour. Therefore, with centered difference discretization of advection, it is necessary to also include a nonzero amount of viscosity or diffusivity sufficient to maintain this inequality.

For the case of tracers, various discretizations other than centered differences can either reduce the need to include diffusivity, or will introduce diffusivity in regions where the Peclet number does not satisfy the above constraint. For momentum, this constraint is always relevant since many ocean climate models still use centered differences for momentum advection.

Experience has shown that if the Reynolds/Peclet number constraint is not satisfied, then the model does not necessarily blow-up. Instead, the solution may have a tendency to slowly degrade via the introduction of grid scale noise.

16.1.2 Linear stability of the diffusion equation

Convergence of the meridions with spherical coordinates makes it possible that a horizontal mixing coefficient appropriate for the mid-latitudes is too large in the high latitudes. In order to derive conservative constraints, it is sufficient to consider the situation in Cartesian coordinates.

16.1.2.1 Laplacian mixing

Consider two dimensional Laplacian mixing in Cartesian coordinates

$$\psi_{,t} = A\,(\psi_{,xx} + \psi_{,yy}) \tag{16.9}$$

where A is the Laplacian viscosity. With a uniform grid, the discrete form of this equation is given by

$$\psi_{i,j}^{n+1} = (1 - 2\sigma_x - 2\sigma_y)\,\psi_{i,j}^{n-1} + \sigma_x\,(\psi_{i+1,j}^{n-1} + \psi_{i-1,j}^{n-1}) + \sigma_y\,(\psi_{i,j+1}^{n-1} + \psi_{i,j-1}^{n-1}) \tag{16.10}$$

where

$$\sigma_x = A\left(\frac{2\,\Delta t}{(\Delta x)^2}\right) \tag{16.11}$$

$$\sigma_y = A\left(\frac{2\,\Delta t}{(\Delta y)^2}\right). \tag{16.12}$$

In time, this equation has the form of a *forward discretization* with a time step $2\Delta t$. As shown in Section (5-9) of Haltiner and Williams (1980), such forward time stepping is necessary for stability of the diffusion equation. Note that the splitting incurred by the leapfrog scheme can be removed by a Robert time filter.

A von Neumann stability analysis is sufficient for the present purposes. This method determines the constraints necessary for numerical stability of an arbitrary grid wave taking the form

$$\psi_{i,j}^{n} = \mathcal{B}^n\, e^{i\,\mu\,x_i}\, e^{i\,\nu\,y_j}, \tag{16.13}$$

where $x_i = i\,\Delta x$, $y_j = j\,\Delta y$, and μ and ν are arbitrary wavenumbers of dimension inverse length. With this wave ansatz, the finite difference form of the diffusion equation becomes

$$\mathcal{B}^{n+1} = \Omega\,\mathcal{B}^{n-1} \tag{16.14}$$

where the amplification factor is

$$\Omega = 1 - 4\,\sigma_x\,\sin^2(\mu\,x_i/2) - 4\,\sigma_y\,\sin^2(\nu\,y_j/2). \tag{16.15}$$

To ensure stability, $-1 < \Omega < 1$ is required. So long as σ_x and σ_y are positive, $\Omega < 1$ is trivial to satisfy. The opposite inequality requires

$$\sigma_x\,\sin^2(\mu\,x_i/2) + \sigma_y\,\sin^2(\nu\,y_j/2) < 1/2. \tag{16.16}$$

The most conservative form of this constraint occurs when the sine terms are unity, in which case one finds

$$A < \frac{1}{4\Delta t}\,\left((\Delta x)^{-2} + (\Delta y)^{-2}\right)^{-1}. \tag{16.17}$$

For one spatial dimension, this constraint implies

$$A^{one-dim} < \frac{(\Delta s)^2}{4\,\Delta t}, \tag{16.18}$$

where again $2\Delta t$ is the leap frog time step.

For two dimensions, again being conservative and so choosing $\Delta x = \Delta y = \Delta s$, yields the constraint

$$A^{two-dim} < \frac{(\Delta s)^2}{8\,\Delta t}. \tag{16.19}$$

A final conservative approximation is to take Δs as the minimum of Δx and Δy within this equation.

The constraint discussed here can be likened to the CFL constraint placed on the time step due to wave propagation. That is, a diffusive or viscous flux represents a transfer of information across the grid. If this transfer occurs too fast relative to the model's time step, then the numerical scheme will go unstable. Correspondingly, in one dimension, the constraint is less strong by a factor of $1/2$. The increased restriction placed on the two-dimensional problem is similar to the two dimensional CFL constraint, as discussed in Section (5-6-7) of Haltiner and Williams (1980). Namely, the effective size of the smallest grid wave is reduced through the addition of an extra dimension, and so the constraint is stronger.

As with the CFL condition, experience has shown that if the linear diffusion equation constraint is not satisfied, even mildly, then the model soon becomes wildly unstable. The instability can be removed by either reducing the time steps or reducing the mixing coefficient.

16.1.2.2 Biharmonic mixing

The same type of analysis for the two-dimensional Cartesian biharmonic equation

$$\psi_{,t} = -B\,(\psi_{,xxxx} + 2\,\psi_{,xxyy} + \psi_{,yyyy}) \tag{16.20}$$

where $B > 0$ is the biharmonic viscosity, leads to the biharmonic amplification factor

$$\Omega = 1 - 16\,[\sqrt{\sigma_x}\,\sin^2(\mu\,\Delta x/2) + \sqrt{\sigma_y}\,\sin^2(\nu\,\Delta y/2)]^2, \tag{16.21}$$

where now

$$\sigma_x = B\left(\frac{2\,\Delta t}{(\Delta x)^4}\right) \tag{16.22}$$

$$\sigma_y = B\left(\frac{2\,\Delta t}{(\Delta y)^4}\right). \tag{16.23}$$

Restricting $-1 < \Omega < 1$ implies

$$[\sqrt{\sigma_x}\,\sin^2(\mu\,\Delta x/2) + \sqrt{\sigma_y}\,\sin^2(\nu\,\Delta y/2)]^2 < 1/8. \tag{16.24}$$

This constraint is satisfied if the following more conservative constraint is satisfied

$$(\sqrt{\sigma_x} + \sqrt{\sigma_y})^2 < 1/8, \tag{16.25}$$

or

$$B < \frac{1}{16\,\Delta t}\left((\Delta x)^{-2} + (\Delta y)^{-2}\right)^{-2}. \tag{16.26}$$

The most conservative constraint is to set

$$B < \frac{(\Delta s)^4}{64\,\Delta t} \tag{16.27}$$

where Δs is the minimum grid spacing.

16.1.3 Western boundary currents

Models with meridional boundaries have boundary currents. The Munk layer (Munk (1950); Gill (1982)) is relevant for determining the width of the boundary layer in MOM. As discussed by Bryan et al. (1975), the model must resolve this layer with at least one grid point (optimally more than one grid point) in order to maintain numerical stability. Under-resolution of the Munk layer shows up most visibly in the vertically integrated velocity field (i.e., the barotropic streamfunction when using the rigid lid, or the free surface height with the free surface). In addition, the work of Griffies et al. (2000b) emphasized the importance of having at least two grid points in the Munk layer in order to minimize the level of spurious diapycnal mixing associated with tracer advection.

With N grid points within the Munk layer, the Laplacian viscosity must satisfy

$$A > \beta(N\,\Delta s\sqrt{3}/\pi)^3, \tag{16.28}$$

where $\beta = \partial_y f = 2.28 \times 10^{-11} (\mathrm{m\,s})^{-1} \cos\phi$ is the meridional gradient of the planetary vorticity. Hence, for Δs in meters

$$A(\mathrm{m}^2\,\mathrm{s}^{-1}) > 3.82 \times 10^{-12} (N\,\Delta s)^3 \cos\phi. \tag{16.29}$$

For example, with $\Delta s = 100 \times 10^3$m at the equator, having one grid cell within the Munk boundary layer requires a viscosity of $2.7 \times 10^3 \mathrm{m}^2\,\mathrm{s}^{-1}$, whereas with $\Delta s = 10 \times 10^3$m, the viscosity must be larger than only $2.7 \times 10^1 \mathrm{m}^2\,\mathrm{s}^{-1}$.

16.1.4 Computational modes on the B-grid

As discussed by Killworth et al. (1991), discretization of gravity waves on a B-grid can admit a stationary grid scale checkerboard pattern (see also Messinger (1973) and Janjić (1977) for atmospheric discussions). This pattern is associated with an unsuppressed grid splitting that can be initiated through grid scale forcing, such as topography. That is, the discrete equations admit a checkerboard null mode. This null mode is seen quite frequently in the surface height field, since this field is typically updated without dissipation. Methods to smooth this field are discussed in these references, as well as **The MOM3 Manual** of Pacanowski and Griffies (1999) and the free surface paper of Griffies et al. (2001).

In addition to noise in the surface height, the computational mode typically is seen when running with very low levels of viscosity. This mode manifests most clearly as zig-zag velocity vectors. For example, such zig-zag patterns are visible in quiescent regions of the model when using the Smagorinsky scheme with no background viscosity.

For some modelers, such as the author, any sign of such unresolved grid-scale power is unacceptable. Although the model may remain quite stable, enhanced power at the grid scale arises from accepting too much unresolved structure at the model's smallest scale. There is no justification for considering such structure physically relevant. Instead, it is a numerical artifact. The extent to which this power influences the larger scales depends on the situation. One egregious problem occurs over long integration periods when noisy velocity fields enhance the levels of spurious mixing of tracers (Griffies et al. (2000b)).

16.1.5 Considerations for viscosity on the sphere

On a sphere using spherical coordinates, the grid spacing in the zonal direction changes according to $\cos\phi$. From many numerical and physical perspectives, it is useful to employ square grids in which the latitudinal resolution is kept abreast with the converging meridions

$$\Delta\phi = \cos\phi\,\Delta\lambda. \tag{16.30}$$

In this way, a grid cell is roughly square with squares becoming smaller as one moves poleward.

Consistent with the desire to employ square grids, it might be useful to prescribe a momentum friction that damps a particular grid scale anomaly with the same time scale regardless of the position on the sphere. As shown in Section 16.1.6, the

damping times for a constant viscosity used for Laplacian and biharmonic friction in one Cartesian dimension is given by

$$\tau_{Lap} = (\Delta/2)^2/A \tag{16.31}$$

$$\tau_{Bih} = (\Delta/2)^4/B, \tag{16.32}$$

where Δ is the grid spacing and B is the biharmonic viscosity. Preserving the damping time as Δ changes on the sphere suggests letting A have a $\cos^2\phi$ dependence and B have a $\cos^4\phi$ dependence.

Besides providing for a constant damping time, a latitudinally dependent friction can be prescribed that relieves the time step constraint given in Section 16.1.2 that ensues when employing a constant viscosity over the extent of the sphere. That constraint becomes more restrictive on the size of $A\,\Delta t$ when moving towards the poles. Again, letting A have a $\cos^2\phi$ dependence relieves this constraint. Similar stability considerations with biharmonic friction leads to a biharmonic coefficient with $\cos^4\phi$ dependence.

The above considerations neglect the lower bound considerations given in Sections 16.1.1 and 16.1.3. Notably, if the viscosity gets too small, the flow becomes numerically unstable and/or noisy. Therefore, as a compromise, instead of a $\cos^2\phi$ dependence, the Laplacian viscosity is typically given a $\cos\phi$ dependence

$$A = A_o \cos\phi, \tag{16.33}$$

where A_o has no latitudinal dependence. This form for the viscosity is furthermore not carried all the way to the pole. In practice, the $\cos\phi$ dependence appears sufficient to alleviate the time step restrictions arising from friction. For biharmonic friction, the analogous viscosity is given by

$$B = B_o \cos^3\phi, \tag{16.34}$$

where B_o has no latitudinal dependence.

In summary, there are three main constraints placed on the viscosity. The Reynolds number and Munk boundary layer constraint provide a lower bound on viscosity, whereas the linear diffusion equation constraint provides an upper bound. In general, modelers hope to reduce the viscosity to its lowest value consistent with these constraints. The reason is that it allows the flow to become more advectively dominant, which typically leads to more realistic simulations. Unfortunately, that effort is often difficult to achieve, largely due to the Reynolds number and Munk constraints.

As proposed in Section 16.2, the Smagorinsky scheme provides the most general means of satisfying each of the above three constraints with only one adjustable dimensionless constant. This scheme, originally used in the GFDL ocean model by Rosati and Miyakoda (1988), is currently being employed more frequently in experiments at GFDL (see Griffies and Hallberg (2000) for a review).

16.1.6 Comparing Laplacian and biharmonic mixing

To understand the differences and similarities between Laplacian and biharmonic dissipation, it is sufficient to consider the following linear evolution equations in

one Cartesian space dimension (Semtner and Mintz (1977))

$$\psi_{,t} = A_l\, \psi_{,xx} \tag{16.35}$$
$$\psi_{,t} = -B\, \psi_{,xxxx} \tag{16.36}$$

where A is the Laplacian eddy coefficient and B is the biharmonic eddy coefficient. Discretizing the Laplacian operator on a Cartesian grid of constant size Δ yields

$$\begin{aligned} \psi_{,t} &= (A_l/\Delta^2)\,(\psi_{i+1} - 2\psi_i + \psi_{i-1}) \\ &= (A_l/\Delta^2)\left(D^{1/2} - D^{-1/2}\right)^2 \psi_i \end{aligned} \tag{16.37}$$

where

$$D^m \psi_i = \psi_{i+m} \tag{16.38}$$

is a linear shift operator. A similar discretization of the biharmonic operator leads to

$$\begin{aligned} \psi_{,t} &= -(B/\Delta^4)\,(\psi_{i+2} - 4\psi_{i+1} + 6\psi_i - 4\psi_{i-1} + \psi_{i-2}) \\ &= -(B/\Delta^4)\left(D^{1/2} - D^{-1/2}\right)^4 \psi_i \end{aligned} \tag{16.39}$$

To garner a sense of how the two forms of dissipation compare, consider a monochromatic grid wave

$$\psi = c(t)\, e^{j\,k\,x_i}, \tag{16.40}$$

where k is a wavenumber, $j = \sqrt{-1}$, and $x_i = i\,\Delta$. For such a wave,

$$(D^{1/2} - D^{-1/2})e^{j\,k\,x_i} = 2\,j\,\sin(k\,\Delta/2)\,e^{j\,k\,x_i} \tag{16.41}$$
$$(D^{1/2} - D^{-1/2})^2 e^{j\,k\,x_i} = -4\,\sin^2(k\,\Delta/2)\,e^{j\,k\,x_i}. \tag{16.42}$$
$$(D^{1/2} - D^{-1/2})^4 e^{j\,k\,x_i} = 16\,\sin^4(k\,\Delta/2)\,e^{j\,k\,x_i}. \tag{16.43}$$

As such, the evolution equations for this wave under the two forms of dissipation are given by

$$\frac{d\ln c(t)}{dt} = -A\left(\frac{\sin(k\,\Delta/2)}{(\Delta/2)}\right)^2 \tag{16.44}$$
$$\frac{d\ln c(t)}{dt} = -B\left(\frac{\sin(k\,\Delta/2)}{(\Delta/2)}\right)^4. \tag{16.45}$$

The solution to these equations is an exponential damping $c(t) = c(0)\,e^{-t/\tau}$, where the inverse damping times are given by

$$\tau_{Lap}^{-1}(k) = A\left(\frac{\sin(k\,\Delta/2)}{\Delta/2}\right)^2 \tag{16.46}$$
$$\tau_{Bih}^{-1}(k) = B\left(\frac{\sin(k\,\Delta/2)}{\Delta/2}\right)^4. \tag{16.47}$$

The damping times are equal whenever

$$A = B\left(\frac{\sin(k\,\Delta/2)}{\Delta/2}\right)^2. \tag{16.48}$$

For example, the smallest grid wave that can live on a discrete grid has size

$$\lambda_{smallest} = 2\Delta, \tag{16.49}$$

which means the wavenumber for this wave is

$$k_{largest} = \pi/\Delta. \tag{16.50}$$

In this case, $\sin(k\Delta/2) = 1$, and so the damping times for this wave are given by

$$\tau_{Lap} = (\Delta/2)^2/A \tag{16.51}$$

$$\tau_{Bih} = (\Delta/2)^4/B. \tag{16.52}$$

This is the strongest damping available from either form of dissipation; waves of smaller wavenumber, or longer wavelength, are less damped and so have larger τ. For the $k = \pi/\Delta$ grid wave, if the dissipation coefficients satisfy

$$B = (\Delta/2)^2\,A, \tag{16.53}$$

then the damping times are the same. For a typical mid-latitude eddy permitting model, let $\Delta = .25 \times \cos(\pi/6) \times 111 \times 1000 = 2.4 \times 10^4 m$. A biharmonic coefficient $B = 10^{11} m^4/sec$ leads to a damping time of 2.4days for the smallest grid scale waves. Relation (16.53) says that if $A = 7 \times 10^2 m^2/sec$, the Laplacian dissipation leads to the same damping time.

16.2 Smagorinsky viscosity

As considered in ocean models, the Smagorinsky scheme provides expressions for horizontal momentum viscosity. The viscosities are flow dependent and hence are are nonlinear. The ideas and history of the method are nicely summarized in Smagorinsky (1993). See also the paper by Griffies and Hallberg (2000) for a discussion focusing on ocean modeling.

16.2.1 General ideas and motivation

Smagorinsky (1963, 1993) proposed that the effective Laplacian viscosity due to unresolved scales should be proportional to the resolved horizontal deformation rate times the squared grid spacing. This scheme is a physically plausible parameterization of the effects of three-dimensional isotropic turbulence, and it has found much use in large-eddy simulations (e.g., see Galperin and Orszag (1993) for a compendium). For large-scale geophysical fluid simulations, however, it has little physical justification since the unresolved scales are dominated by quasi two-dimensional geostrophic turbulence. For this reason, Leith (1968, 1996) proposed an alternative approach based on two-dimensional turbulence. Leith's viscosity is

proportional to the horizontal gradient of the relative vorticity times the cubed grid spacing. This approach has found some use in atmospheric models (e.g., Boer and Shepherd (1983)), but it is not commonly used in ocean models.

The Smagorinsky scheme has found notable use in large-scale ocean models for pragmatic reasons (e.g., Blumberg and Mellor (1987); Rosati and Miyakoda (1988); Bleck and Boudra (1981); Bleck et al. (1992); Griffies and Hallberg (2000)). Firstly, it is more convenient to compute than the Leith viscosity due to the smaller required grid stencil, and because the deformation rate used to compute the Smagorinsky viscosity is furthermore needed to compute the stress tensor. Secondly, as with the Leith scheme, the Smagorinsky viscosity tailors the local dissipation to both the local flow state and local grid resolution with only a single, nondimensional adjustable parameter. If this parameter is properly chosen, the resulting viscosity ensures that the flow respects the relevant numerical stability properties, even when simulating multiple flow and grid regimes such as occur in realistic ocean simulations.

16.2.2 The dimensionless Smagorinsky scaling parameter

In the Smagorinsky scheme, one introduces an adjustable dimensionless parameter which is in practice used as a tuning parameter. It acts to set the overall magnitude of the viscosity. In an attempt to bring the definition of this parameter into line with that used by others, it is useful to present the following discussion, based largely on that given in Section 1.9 of Smagorinsky (1993).

Recall that locally kinetic energy is dissipated by isotropic transverse frictional stresses according to (see Section 15.7.1)

$$-\, e_{mn}\, \tau^{mn} = A\, \rho\, (e_T^2 + e_S^2) \tag{16.54}$$

where e_T and e_S are the transverse deformation rates due to transverse strains (Section 15.7.1). Smagorinsky chose his viscosity according to

$$A = (\Upsilon / k_m)^2\, |E| \tag{16.55}$$

where

$$|E| = \sqrt{e_T^2 + e_S^2} \tag{16.56}$$

is the total deformation rate whose units are inverse time. As written, the viscosity depends in a nonlinear manner on the flow, and so is often termed a *nonlinear viscosity*. The wavenumber k_m will be specified below. Υ is a dimensionless adjustable parameter. For certain flow regimes, it has been predicted by theories and determined by experiments. Originally, Smagorinsky (1963) was motivated to set $\Upsilon = 0.4$, which is the value expected from wall boundary layer turbulence with 0.4 the von Karman constant. However, this value for Υ is not necessarily relevant for an ocean model, where grid lengths less than a kilometer are rarely achieved. Consequently, there is little theoretical justification for choosing one particular value of Υ, and the modeler should consider tuning this parameter accordingly.

Notably, as emphasized by Griffies and Hallberg (2000), one practical advantage of the Smagorinsky approach is that the value for Υ is often appropriate across a broad range of grid sizes. That is, if simulating a flow with the Smagorinsky

scheme, changing to a higher resolution may typically be done without changing Υ, so long as similar flow regimes are present. Such is not the case in general when tuning the dimensionful viscosity. Hence, the Smagorinsky approach helps to reduce the burden on retuning the model upon changing resolution.

In MOM, k_m is defined to be the largest resolvable wavenumber local to the grid cell of interest. That is, for a given horizontal grid scale Δs, $2\Delta s$ represents the smallest explicitly resolvable wave length. The corresponding largest resolvable wavenumber is

$$k_m = (2\pi)/(2\Delta s) = \pi/\Delta s. \tag{16.57}$$

Hence, the viscosity is determined by

$$A = (\Upsilon \Delta s/\pi)^2 |E|. \tag{16.58}$$

This result differs by a factor of π from that employed by Smagorinsky (1963).

It is convenient to choose the grid spacing as

$$\Delta s \equiv \frac{2\,\Delta x\,\Delta y}{\Delta x + \Delta y}. \tag{16.59}$$

Δs reduces to Δx when the grid spacing is equal in the two directions, $2\,\Delta x$ when $\Delta x << \Delta y$, and an appropriate value in between when the grid spacing is roughly equal. This choice for Δs has continuous derivatives, in contrast to alternative choice of $\Delta s = \min(\Delta x, \Delta y)$.*

Note that when using the isotropic form of the stress tensor discussed in Section 15.5, it is important to employ a single transverse viscosity, not a separate one for each of the two horizontal directions as suggested by Rosati and Miyakoda (1988). The reason is that if two viscosities are introduced, then that would generally break the trace-free and symmetric properties of the stress tensor discussed in Section 15.3. A general way to consistently introduce two horizontal viscosities is described in Section 15.6, which follows the approach of Smith and McWilliams (2002). We discuss the choice of these two viscosities in Section 16.4.

16.2.3 Choosing the scaling coefficient

There are at least three things to consider when deciding what value to use for the scaling coefficient Υ. First, the viscosity used in a centered differenced momentum advection scheme must be large enough so that the cell Reynolds number satisfies the advection-diffusion constraint established in Section 16.1.1

$$R = U\Delta s/A < 2 \Rightarrow A > U\Delta s/2, \tag{16.60}$$

where U is a velocity scale and Δs the corresponding grid spacing. In order to get a rough value for Υ based on this constraint, let $|E| \approx U/\Delta s$, which means that the Smagorinsky viscosity is roughly $A \approx (\Upsilon/\pi)^2\, U\, \Delta s$. Forcing $R < 2$ implies that

$$\Upsilon > \pi/\sqrt{2} \approx 2.2. \tag{16.61}$$

*Thanks to Bob Hallberg for this point.

Actually, this estimate provides an upper bound on the viscosity realized with the scheme, since $\Delta U \approx U$ only in regions of strong horizontal shear. In other regions, the nonlinear viscosity is significantly smaller.

The second consideration involves the western boundary layer considered in Section 16.1.3. Again, with N grid points within the Munk layer and grid spacing Δs in meter, the viscosity must satisfy

$$A(\mathrm{m}^2\,\mathrm{s}^{-1}) > 3.82 \times 10^{-12} (N\,\Delta s)^3 \cos\phi. \tag{16.62}$$

In a 100×10^3m resolution model, typical velocities are $0.1\mathrm{m\,s}^{-1}$, and so the maximum Smagorinsky nonlinear viscosities are on the order of $\Upsilon^2 \times 10^3 \mathrm{m}^2\,\mathrm{s}^{-1}$. With $\Upsilon = 2.2$, the model might be safely resolving the Munk layer since the regions where the layer is formed is also where there are strong shears. A larger value of Υ, however, might be necessary and so tests are appropriate.

The Munk boundary layer considerations can be computed prior to any model computations since they do not involve flow properties. If the result of this calculation is that the needed Munk viscosity far excedes a reasonable estimate based on a chosen Υ, such as the 2.2 based on the cell Reynolds number, then there are two choices that can be made. The first is to run with a time-independent background viscosity set large enough to satisfy the Munk boundary layer condition. If this choice is made, there is little reason to employ the Smagorinsky scheme since the computed viscosities are dominated by the background. The second choice is to increase Υ so that the nonlinear viscosity satisfies the Munk boundary layer condition, at least for the cells within the boundary layer.

For constant viscosity, the main region in which the linear stability constraint

$$A < \frac{(\Delta s)^2}{4\,\Delta t} \tag{16.63}$$

becomes important is in regions of small grid spacing. For spherical grids, such regions typically are in the high latitudes where meridions converge. Because the Smagorinsky scheme has the proper grid stretching factors, it naturally produces small viscosities in the high latitudes. That is, Υ generally need not have any latitudinal dependence. This property is quite convenient for global models and any model with variable resolution.

16.2.4 What about tracers?

For the dissipation of tracer variance through horizontal diffusion, Smagorinsky (1963) assumed that a unit value for the horizontal turbulent Prandtl number*. That is, the horizontal viscosity and diffusivity are set equal. Such an approach has been followed by some ocean modelers who use the Smagorinsky scheme.

Since geostrophic turbulence is the dominant mixing/stirring process in the large-scale ocean, most modelers interested in the long term properties of ocean tracers have not used the Smagorinsky algorithm for computing diffusivities for horizontal tracer diffusion, isopyncal tracer diffusion, or thickness diffusion. Instead, alternative theories based on quasi-geostrophic turbulence are recommended. Such theories are discussed in Chapters 11–14.

*Recall that the Prandtl number is the ratio of the viscosity to diffusivity.

One further practical shortcoming of the Smagorinsky approach for horizontal tracer diffusion is that the horizontal Smagorinsky diffusivities tend to be largest near boundaries, where the shears are largest. Hence, spurious upwelling at the side boundaries caused by large horizontal diffusivities; i.e., the *Veronis effect* (Veronis (1973, 1975)) can actually be enhanced with Smagorinsky tracer diffusivities when used in geopotentially aligned horizontal diffusion operators.

For the above reasons, the Smagorinsky scheme is of use in ocean climate models *only* for use in determining the horizontal viscosity for dissipating kinetic energy through friction. It is not of use for determining tracer diffusivities at the grid resolutions of interest for ocean climate models.

16.2.5 Biharmonic Smagorinsky friction

If the Smagorinsky scheme is employed along with the Laplacian mixing, then linear stability suggests that the Smagorinsky viscosity should satisfy

$$A < \frac{(\Delta s)^2}{4\,\Delta t}, \tag{16.64}$$

where $\Delta s \equiv (2\,\Delta x\,\Delta y)/(\Delta x + \Delta y)$. A biharmonic viscosity should likewise satisfy

$$B < \frac{(\Delta s)^4}{32\,\Delta t}. \tag{16.65}$$

These results suggest setting the biharmonic viscosity from the Smagorinsky scheme according to

$$B = A\,(\Delta s)^2/8, \tag{16.66}$$

where A is the Laplacian Smagorinsky viscosity. This scheme has the advantages of the Laplacian Smagorinsky scheme while providing the enhanced scale selectivity of a biharmonic operator.

16.3 Background viscosity

The Smagorinsky viscosity suffers from one problem necessitating the addition of a nontrivial background viscosity. The problem is that in regions where the flow is weak, the Smagorinsky viscosity can be quite small. Such regions may be next to boundaries and so the boundary layer may go unresolved. In this case, the simulated velocity field becomes quite noisy. Increasing the non-dimensional Smagorinsky parameter Υ reduces the noise, but at the cost of over-dissipating other regions where the Smagorinsky viscosity is sufficient.

To alleviate this problem*, we consider a grid scale dependent viscosity to set a minimum allowable viscosity. The idea is to let the minimum Laplacian viscosity be

$$A_{\text{min}} = U_A\,\Delta s, \tag{16.67}$$

*Eric Chassignet from the University of Miami suggested this approach to the author in 1999.

where U_A is a constant velocity scale and

$$\Delta s = \frac{2\,\Delta x\,\Delta y}{\Delta x + \Delta y}. \tag{16.68}$$

The value $U_A = 0.25\mathrm{m\,s^{-1}}$ has been found effective for many purposes. When allowing for multiple regimes of flow, such as with a more highly resolved equatorial region, it may be necessary to provide an *a priori* spatial dependence to U.

For biharmonic viscosity,

$$B_{\mathrm{min}} = U_B\,(\Delta s)^3, \tag{16.69}$$

where $U_B = 0.05\mathrm{m\,s^{-1}}$ has been found to be useful for many applications with eddying flows.

When comparing experiments run with Laplacian versus biharmonic operators, it is often useful to use grid-scale dependent viscosities that damp grid-scale waves at the same rate. As seen in Section 16.1.6, if $B = (\Delta/2)^2\,A$ then the biharmonic operator damps a grid wave the same as the corresponding Laplacian operator. If each viscosity is set at its minimum given above, then

$$\begin{aligned} B_{\mathrm{min}} &= U_B\,(\Delta s)^3 && (16.70)\\ &= (\Delta s)^2\,A_{\mathrm{min}}/4 && (16.71)\\ &= (\Delta s)^3\,U_A/4, && (16.72) \end{aligned}$$

thus leading to the correspondence between the velocity scales

$$U_B = U_A/4. \tag{16.73}$$

16.4 Viscosities for anisotropic friction

The goal for choosing model viscosities is to satisfy the numerical constraints of Section 16.1 without overly dissipating the simulation. The anisotropic friction operator considered in Section 15.6 provides an added degree of freedom when specifying the transverse friction, and this extra degree of freedom can be exploited to reduce the overall model dissipation. We provide here some guidelines for choosing the isotropic viscosity A and anisotropic viscosity D. Similar considerations were discussed by Smith and McWilliams (2002).

There are two main numerical constraints from Section 16.1. First, frictional boundary layers must be resolved. Doing so requires enough friction be applied in the cross-stream direction. Second, friction must be large enough to satisfy the grid Reynolds number constraint. Doing so requires enough friction be applied in the along-stream direction.

To see how these requirements translate into setting the viscosities, let us consider the stress and friction associated with a flow aligned with the coordinate directions. For this purpose, we recall the stress tensor is generally given by (Section 15.7.2)

$$\tau^{(m)(n)} = \rho \begin{pmatrix} (A\,e_T + D\,\Delta\,\sin 2\theta) & (A\,e_S - D\,\Delta\,\cos 2\theta) \\ (A\,e_S - D\,\Delta\,\cos 2\theta) & (-A\,e_T - D\,\Delta\,\sin 2\theta) \end{pmatrix} \tag{16.74}$$

where the deformation rates are (Section 15.7.1)

$$e_T = (dy)\,(u/dy)_{,x} - (dx)\,(v/dx)_{,y} \tag{16.75}$$
$$e_S = (dx)\,(u/dx)_{,y} + (dy)\,(v/dy)_{,x} \tag{16.76}$$

the anisotropic term is (Section 15.6.2)

$$2\,\Delta = e_S\,\cos 2\theta - e_T\,\sin 2\theta, \tag{16.77}$$

and θ sets the orientation of the friction, with the flow dependent orientation of Smith and McWilliams (2002) given by

$$\mathbf{u} = |\mathbf{u}|\,(\cos\theta, \sin\theta). \tag{16.78}$$

Consider friction oriented so that $\theta = \pi/2$, which occurs for the flow dependent case when currents are moving in the generalized meridional direction, then $2\,\Delta = -e_S$ and the stress tensor is

$$\tau^{(m)(n)} = \rho \begin{pmatrix} A\,e_T & (A - D/2)\,e_S \\ (A - D/2)\,e_S & -A\,e_T \end{pmatrix}. \tag{16.79}$$

For simplicity, now assume Cartesian coordinates on a plane with a Boussinesq fluid and constant viscosities. Doing so leads to the friction vector

$$F^x = (A + D/2)\,u_{,xx} + (A - D/2)\,u_{,yy} - (D/2)\,\partial_x\,\nabla\cdot\mathbf{u} \tag{16.80}$$
$$F^y = (A + D/2)\,v_{,yy} + (A - D/2)\,v_{,xx} - (D/2)\,\partial_y\,\nabla\cdot\mathbf{u}. \tag{16.81}$$

With a small horizontal divergence $\nabla\cdot\mathbf{u}$, the last term in these expressions is subdominant to the first and second terms. For meridional flow away from boundaries, the along-stream viscosity

$$A_{along} = A + D/2 \tag{16.82}$$

must satisfy the grid Reynolds number constraint. For meridional flow aligned with a meridional boundary, the cross-stream viscosity

$$A_{cross} = A - D/2 \tag{16.83}$$

must satisfy the Munk boundary layer constraint. Finally, recall from Section 15.6.2 that $D < 2\,A$ is required to ensure that kinetic energy is dissipated by the friction operator (see equation (15.99)), which means that the cross-stream viscosity can never go negative.

CHAPTER
SEVENTEEN

Discrete B-grid horizontal friction

Contents

17.1 Comments on the notation **376**
17.2 Summary of the various formulations **376**
17.3 Horizontal friction discretization **377**
17.3.1 General form of the discrete dissipation functional 377
17.3.2 Grid cell distances and subcell volumes 378
17.3.3 Derivative operators 380
17.3.4 Tension and strain for the subcells 380
17.3.5 Functional derivative of e_T 382
17.3.6 Functional derivative of e_S 383
17.3.7 Tidy form for the discretized friction 383
17.3.8 Tension and strain in the quadrants 385
17.3.9 Comments 386
17.3.10 Discretized Smagorinsky viscosity 386
17.4 Laplacian plus metric form of isotropic friction **387**
17.4.1 Spherical form of second order friction 387
17.4.2 Zonal friction 388
17.4.3 Meridional friction 389

The purpose of this chapter is to formulate a discrete version of the horizontal momentum friction operators derived in Chapter 15. The basis for this discretization is the functional approach described in Section 15.8. This chapter will appeal to those interested in implementing friction in a B-grid ocean model. Notably, the formulation on C-grid models is simpler, as mentioned by Griffies and Hallberg (2000).

As with our discussion of the discrete neutral physics operator in Chapter 14, the presentation here introduces grid labels on fields living on a discrete lattice. Explicit sums and averages are also used in most places rather than choosing amongst the more elegant alternatives. We choose this more explicit notation since it eads to less ambiguity for the reader. Furthermore, such is ultimately required by the

numericist when translating an algorithm into computer code. Experience has shown that having the algorithm written nearly in its "full glory" in LaTeXreduces the chances for bugs to penetrate the computer code. Such is especially crucial for complicated schemes such as the one documented in this chapter as well as Chapter 14. Hence, for pedogogical and "quality control" reasons, we prefer the explicit grid-label notation. Recognizing that this approach can lead to somewhat lengthy equations, we use a relatively small font for the remainder of this chapter to allow for visually more compact expressions.

17.1 Comments on the notation

Chapter 15 established the form of friction within the context of the unaveraged equations of motion derived in Chapter 3. Such is strictly a mis-representation of the unaveraged equations, since they employ dissipation operators due to molecular friction, not the sub-grid-scale (SGS) friction of focus here. Nonetheless, we took this approach for purposes of tidiness. Now that we are transferring the results from Chapter 15 to model fields, we follow the mapping given by Table 6.1, whereby $\mathbf{v} \rightarrow \mathbf{v}^{\rho}$ for the non-Boussinesq fluid, and $\mathbf{v} = \mathbf{v}^{\rho}$ for the Boussinesq fluid. Notably, this mapping means that it is the velocity $\mathbf{v}^{\rho} = \mathbf{v}\,(\rho_o/\rho)$ that is used to construct the friction operator in the non-Boussinesq model, *not* the velocity $\mathbf{v}$. In addition, the friction operator in the model represents a discrete version of the SGS momentum flux

$$\langle \tilde{\mathbf{F}}^{(\mathbf{u})}_{sgs} \rangle \rightarrow \mathbf{F}^{(\mathbf{u})}_{\text{model}}. \tag{17.1}$$

When considering model implementations, factors of density reduce to their Boussinesq value ρ_o, which simplifies the numerical treatment in the non-Boussinesq case.

In the present chapter, in order to maintain a minimum set of labels, we are *not* careful to abide by the above notation for the velocity. Hence, as emphasized by Greatbatch et al. (2001), it is important to recognize that when const ructing friction within a non-Boussinesq model, we use $\mathbf{v}^{\rho}$ instead of $\mathbf{v}$ to ensure that kinetic energy is dissipated.

17.2 Summary of the various formulations

On a horizontal plane with second order numerics, the discrete second order Laplacian forms a 5-point stencil. This stencil precludes computational modes, and hence always acts to smooth a field. It is for this reason that many ocean models employ a Laplacian operator for dissipating momentum in order to satisfy various numerical stability requirements.

On a sphere and in the presence of non-homogeneous grids, boundaries, and various flow regimes, a Laplacian operator using a constant viscosity is often sub-optimal in the sense that it may over-dissipate the simulated flow. The paper by Griffies and Hallberg (2000) motivates the consideration of the Smagorinsky viscosity (Section 16.2) in either a second order (Laplacian) or fourth order (biharmonic) friction operator. Additional considerations necessary to preserve angular momentum on a sphere warrant added sophistication to the operator beyond that of the familiar Cartesian form of the Laplacian operator.

Many ocean models handle the added complications due to sphericity and non-constant viscosities in two basic ways. First, there are the formulae of Bryan (1969), Wajsowicz (1993), Smith et al. (1995), and Murray and Reason (2002). In this approach, friction is written as a "Laplacian plus frictional metric" form. The Laplacian piece is the spherical coordinate form of a Laplacian operator acting on a component of the velocity, as if the velocity component was a scalar field. The frictional metric piece is an added term accounting for the fact that velocity is a vector, not a scalar. The friction metric term arises for two general reasons: (1) the sphericity of the earth (Sections 3.10.2 and 3.10.3), and (2) non-constant

viscosities (Section 15.5.4). Without the frictional metric term, the corresponding stress tensor would not be symmetric, thus allowing friction to introduce spurious torques and thus compromising the angular momentum budgets. Bryan (1969) used a friction metric appropriate for constant viscosities, Wajsowicz (1993) showed how it is modified for use with non-constant viscosities, and Smith et al. (1995) and Murray and Reason (2002) generalized these results to arbitrary orthogonal coordinates.

The frictional metric term is cumbersome to discretize, especially with generalized orthogonal coordinates. Additionally, its discretization on a B-grid admits computational modes. Hence, the frictional metric is sometimes ignored, as in Cox (1984). If maintaining the metric term, the presence of the scalar Laplacian appears to render the computational modes harmless. Nonetheless, the formulation is not satisfying since ideally no part of the friction operator should admit computational modes. An alternative formulation is therefore desired.

A second way to discretize the friction operator is to apply the tensor formalism of Smagorinsky (1963, 1993) in the process of discretizing the Smagorinsky nonlinear viscosity method. This was the approach of Rosati and Miyakoda (1988). In the continuum, this approach is equivalent to Wajsowicz (1993) (see Section 17.4 for details). Yet there are differences on the discrete lattice. Furthermore, the tensor formalism is considerably more concise, and the ability to generalize to arbitrary orthogonal curvilinear coordinates is more elegant. It is for these reasons that we prefer this method here.

The following history of friction in the GFDL ocean code helps to summarize the previous discussion:

- Bryan (1969): Scalar Laplacian plus metric term for constant viscosity.
- Cox (1984): Omitted metric terms for constant viscosity.
- Rosati and Miyakoda (1988): Smagorinsky scheme in which friction, with a non-constant viscosity, was determined from derivatives of the stress tensor components.
- Pacanowski et al. (1991) (MOM1): Reintroduced metric terms for constant viscosity.
- Pacanowski (1995) and Pacanowski and Griffies (1999) (MOM2 and MOM3): Scalar Laplacian plus constant viscosity metric terms of Bryan (1969) and non-constant viscosity terms of Wajsowicz (1993).
- Griffies et al. (2003) (MOM4): Tensor formalism with general orthogonal coordinates discretized based on a functional.

17.3 Horizontal friction discretization

The purpose of this section is to discretize the horizontal friction operator. Much here follows the functional approach used for the neutral diffusion discretization detailed in Chapter 14. However, the present discussion is simpler because we need to only discretize a two dimensional (x-y plane) operator, and because there are no subtleties associated with neutral directions. The approach here is therefore a straightforward application of the functional formalism described in Section 15.8.

17.3.1 General form of the discrete dissipation functional

The discrete dissipation functional takes the form

$$\mathcal{S} = \sum_{i,j} \sum_{n=1}^{12} V(n) \left[A(n) \left(e_T^2(n) + e_S^2(n) \right) - 2\, D(n)\, \Delta^2(n) \right] \equiv \sum_{i,j} \mathcal{S}_{i,j}. \tag{17.2}$$

Note that this functional is relevant for both the Boussinesq and non-Boussinesq fluids due to the mapping (17.1) as derived in Section 6.6. Figure 17.1 illustrates the nearest neighbor stencil used for discretizing the dissipation functional using second order numerics. The summation $n = 1, 12$ arises from the 12 subcells to which the velocity point $U_{i,j}$ contributes when discretizing the functional. $V(n)$ are the volumes of each of the subcells, $A(n)$ are the viscosities, and $e_T(n)$, $e_S(n)$, and $\Delta(n)$ are the corresponding tensions, strains, and anisotropic contribution, respectively.

For a particular velocity grid cell $U_{i,j,k}$, the physical component of the friction vector acting at the velocity cell $U_{i,j,k}$ is given by the discrete functional derivative (see Section 15.8.4)

$$
\begin{aligned}
-F^{(b)}_{i,j} &= \frac{1}{2\,V^U_{i,j}} \frac{\partial \mathcal{S}_{i,j}}{\partial (u^{(b)})_{i,j}} \\
&= \frac{1}{V^U_{i,j}} \sum_{n=1}^{12} V(n) \times \\
&\left(A(n)\, e_T(n) \frac{\partial e_T(n)}{\partial (u^{(b)})_{i,j}} + A(n)\, e_S(n) \frac{\partial e_S(n)}{\partial (u^{(b)})_{i,j}} - 2\, D(n)\, \Delta(n) \frac{\partial \Delta(n)}{\partial (u^{(b)})_{i,j}} \right) \\
&= \frac{1}{\rho_o\, V^U_{i,j}} \sum_{n=1}^{12} V(n) \left(\frac{\partial e_T(n)}{\partial (u^{(b)})_{i,j}} \tau^{xx}(n) + \frac{\partial e_S(n)}{\partial (u^{(b)})_{i,j}} \tau^{xy}(n) \right) .
\end{aligned} \tag{17.3}
$$

In this equation, $b = 1, 2$ labels the generalized zonal and meridional directions, the discrete depth label k was dropped since all points are at the same depth level,

$$
\left((u^{(1)})_{i,j}, (u^{(2)})_{i,j} \right) = (u_{i,j}, v_{i,j}) \tag{17.4}
$$

are physical components of the velocity vector,

$$
V^U_{i,j,k} = dxu_{i,j}\, dyu_{i,j}\, dhu_{i,j,k} \tag{17.5}
$$

is the velocity cell volume,

$$
\tau^{xx}(n) = \rho_o \left[A(n)\, e_T(n) + D(n)\, \Delta(n)\, \sin 2\theta \right] \tag{17.6}
$$

$$
\tau^{xy}(n) = \rho_o \left[A(n)\, e_S(n) - D(n)\, \Delta(n)\, \cos 2\theta \right] \tag{17.7}
$$

are the stress tensor components, and the orientation angle θ determines the orientation of the friction according to the discussion in Section 15.6.1. In the remainder of this section, we perform the discrete functional derivatives and then manipulate the results into a tidy expression for the discrete friction operator.

17.3.2 Grid cell distances and subcell volumes

Enumerating the 12 quarter cell volumes constitutes an important part of the discretization. We assume knowledge of the distance along each of the four sides of a velocity and tracer cell, as well as the distance from the velocity and tracer points to the sides of their respective cells. Although nonuniform grids with general coordinates do not allow an exact (i.e., analytically exact) area calculation, the horizontal area of the quarter cells is well approximated using this information. Even so, for the purpose of discretizing friction, there is no reason to take pains to compute a very accurate quarter cell area. Indeed, our aim is to realize the dissipative property of friction with minimal computational expense. Hence, we make the simplification that the volume of a quarter cell is equal to one-quarter the volume of the corresponding velocity cell. In summary, we employ the following grid information (refer to Figure 17.1):

- The longitudinal spacing, in meters, between $U_{i,j}$ and $U_{i+1,j}$ is $dxue_{i,j}$. The latitudinal spacing, in meters, between $U_{i,j}$ and $U_{i,j+1}$ is $dyun_{i,j}$.

- The volume of a quarter cell is taken equal to one-quarter the volume of the velocity cell where the quarter cell lives. For example, quarter cells 1,2,3,4 live inside velocity cell $U_{i,j,k}$ and so have volume

$$(1/4)\, V^U_{i,j,k} = (1/4)\, dxu_{i,j}\, dyu_{i,j}\, dhu_{i,j,k}, \tag{17.8}$$

whereas quarter cells 5,6 live inside velocity cell $U_{i+1,j,k}$ and so have volume

$$(1/4)\, V^U_{i+1,j,k} = (1/4)\, dxu_{i+1,j}\, dyu_{i+1,j}\, dhu_{i+1,j,k}. \tag{17.9}$$

In summary, the volumes of the 12 quarter cells are taken to be

$$\begin{aligned}
&V(1) = V(2) = V(3) = V(4) = (1/4)\, V^U_{i,j,k} \\
&V(5) = V(6) = (1/4)\, V^U_{i+1,j,k} \\
&V(7) = V(8) = (1/4)\, V^U_{i-1,j,k} \\
&V(9) = V(10) = (1/4)\, V^U_{i,j+1,k} \\
&V(11) = V(12) = (1/4)\, V^U_{i,j-1,k}
\end{aligned} \tag{17.10}$$

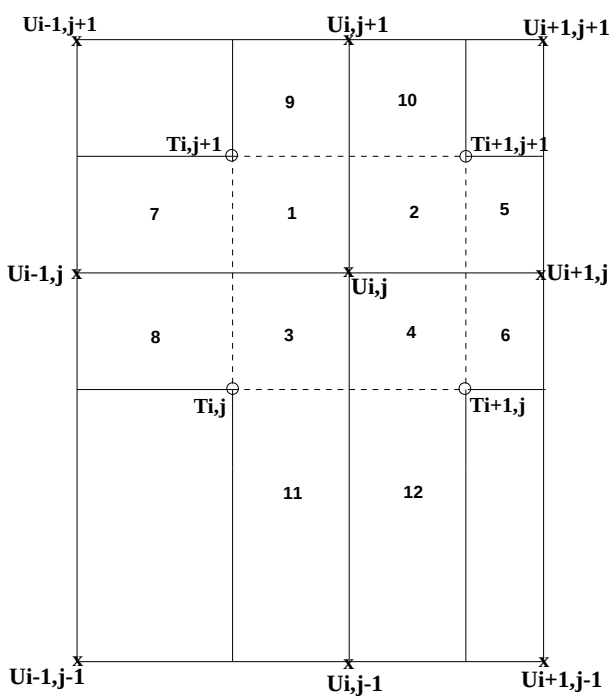

Figure 17.1: Stencil for the discrete horizontal friction functional. The 12 quarter cells each contain contributions to the functional from the central velocity point $U_{i,j}$.

17.3.3 Derivative operators

The horizontal finite difference derivative operators are given by

$$\delta_x u_{i,j} = \frac{u_{i+1,j} - u_{i,j}}{dxue_{i,j}} \tag{17.11}$$

$$\delta_y u_{i,j} = \frac{u_{i,j+1} - u_{i,j}}{dyun_{i,j}}. \tag{17.12}$$

The distances $dxue_{i,j}$ and $dyun_{i,j}$ represent the zonal and meridional distances between the velocity points. All metric stretching factors are absorbed into these grid distances according to the discussion in Section 19.12.4.

17.3.4 Tension and strain for the subcells

We must decide what form of the tension and strain to discretize. Our goal is to work with the discretization involving the least amount of grid factors in order to minimize computational expense. The first forms are more symmetric

$$e_T = dy\,(u/dy)_{,x} - dx\,(v/dx)_{,y} \tag{17.13}$$

$$e_S = dx\,(u/dx)_{,y} + dy\,(v/dy)_{,x} \tag{17.14}$$

and lead to a highly symmetric and compact form for the discretized friction operator (see below), directly analogous to the continuum form given by equations (15.118) and (15.119). The second forms isolate the Cartesian expression

$$e_T = u_{,x} - v_{,y} + v\,\partial_y\,\ln dx - u\,\partial_x\,\ln dy \tag{17.15}$$

$$e_S = u_{,y} + v_{,x} - u\,\partial_y\,\ln dx - v\,\partial_x\,\ln dy. \tag{17.16}$$

The second forms require less computation since the metric terms

$$(M_T)_{i,j} = -u_{i,j}\,(\partial_x\,\ln dy)_{i,j} + v_{i,j}\,(\partial_y\,\ln dx)_{i,j} \tag{17.17}$$

$$(M_S)_{i,j} = -u_{i,j}\,(\partial_y\,\ln dx)_{i,j} - v_{i,j}\,(\partial_x\,\ln dy)_{i,j} \tag{17.18}$$

are common to each of the four surrounding triads, and the following grid factors can be computed at the start of the integration

$$dh1dy_{i,j} = (\partial_y\,\ln dx)_{i,j} \tag{17.19}$$

$$dh2dx_{i,j} = (\partial_x\,\ln dy)_{i,j}. \tag{17.20}$$

Hence, we choose the first form of the deformation rates to develop the discrete friction, and the second form to evaluate the deformation rates within the discrete friction. For purposes of completeness, we display both discrete forms of the deformation rates here.

The first form for tension in the 12 subcells is

$$
\begin{aligned}
e_T(1) &= dyue_{i-1,j}\,\delta_x\,(u_{i-1,j}/dyu_{i-1,j}) - dxun_{i,j}\,\delta_y\,(v_{i,j}/dxu_{i,j}) \\
e_T(2) &= dyue_{i,j}\,\delta_x\,(u_{i,j}/dyu_{i,j}) - dxun_{i,j}\,\delta_y\,(v_{i,j}/dxu_{i,j}) \\
e_T(3) &= dyue_{i-1,j}\,\delta_x\,(u_{i-1,j}/dyu_{i-1,j}) - dxun_{i,j-1}\,\delta_y\,(v_{i,j-1}/dxu_{i,j-1}) \\
e_T(4) &= dyue_{i,j}\,\delta_x\,(u_{i,j}/dyu_{i,j}) - dxun_{i,j-1}\,\delta_y\,(v_{i,j-1}/dxu_{i,j-1}) \\
e_T(5) &= dyue_{i,j}\,\delta_x\,(u_{i,j}/dyu_{i,j}) - dxun_{i+1,j}\,\delta_y\,(v_{i+1,j}/dxu_{i+1,j}) \\
e_T(6) &= dyue_{i,j}\,\delta_x\,(u_{i,j}/dyu_{i,j}) - dxun_{i+1,j-1}\,\delta_y\,(v_{i+1,j-1}/dxu_{i+1,j-1}) \\
e_T(7) &= dyue_{i-1,j}\,\delta_x\,(u_{i-1,j}/dyu_{i-1,j}) - dxun_{i-1,j}\,\delta_y\,(v_{i-1,j}/dxu_{i-1,j}) \\
e_T(8) &= dyue_{i-1,j}\,\delta_x\,(u_{i-1,j}/dyu_{i-1,j}) - dxun_{i-1,j-1}\,\delta_y\,(v_{i-1,j-1}/dxu_{i-1,j-1}) \\
e_T(9) &= dyue_{i-1,j+1}\,\delta_x\,(u_{i-1,j+1}/dyu_{i-1,j+1}) - dxun_{i,j}\,\delta_y\,(v_{i,j}/dxu_{i,j}) \\
e_T(10) &= dyue_{i,j+1}\,\delta_x\,(u_{i,j+1}/dyu_{i,j+1}) - dxun_{i,j}\,\delta_y\,(v_{i,j}/dxu_{i,j}) \\
e_T(11) &= dyue_{i-1,j-1}\,\delta_x\,(u_{i-1,j-1}/dyu_{i-1,j-1}) - dxun_{i,j-1}\,\delta_y\,(v_{i,j-1}/dxu_{i,j-1}) \\
e_T(12) &= dyue_{i,j-1}\,\delta_x\,(u_{i,j-1}/dyu_{i,j-1}) - dxun_{i,j-1}\,\delta_y\,(v_{i,j-1}/dxu_{i,j-1})
\end{aligned}
\tag{17.21}
$$

and the first form for strain is

$$
\begin{aligned}
e_S(1) &= dxun_{i,j}\,\delta_y\,(u_{i,j}/dxu_{i,j}) + dyue_{i-1,j}\,\delta_x\,(v_{i-1,j}/dyu_{i-1,j}) \\
e_S(2) &= dxun_{i,j}\,\delta_y\,(u_{i,j}/dxu_{i,j}) + dyue_{i,j}\,\delta_x\,(v_{i,j}/dyu_{i,j}) \\
e_S(3) &= dxun_{i,j-1}\,\delta_y\,(u_{i,j-1}/dxu_{i,j-1}) + dyue_{i-1,j}\,\delta_x\,(v_{i-1,j}/dyu_{i-1,j}) \\
e_S(4) &= dxun_{i,j-1}\,\delta_y\,(u_{i,j-1}/dxu_{i,j-1}) + dyue_{i,j}\,\delta_x\,(v_{i,j}/dyu_{i,j}) \\
e_S(5) &= dxun_{i+1,j}\,\delta_y\,(u_{i+1,j}/dxu_{i+1,j}) + dyue_{i,j}\,\delta_x\,(v_{i,j}/dyu_{i,j}) \\
e_S(6) &= dxun_{i+1,j-1}\,\delta_y\,(u_{i+1,j-1}/dxu_{i+1,j-1}) + dyue_{i,j}\,\delta_x\,(v_{i,j}/dyu_{i,j}) \\
e_S(7) &= dxun_{i-1,j}\,\delta_y\,(u_{i-1,j}/dxu_{i-1,j}) + dyue_{i-1,j}\,\delta_x\,(v_{i-1,j}/dyu_{i-1,j}) \\
e_S(8) &= dxun_{i-1,j-1}\,\delta_y\,(u_{i-1,j-1}/dxu_{i-1,j-1}) + dyue_{i-1,j}\,\delta_x\,(v_{i-1,j}/dyu_{i-1,j}) \\
e_S(9) &= dxun_{i,j}\,\delta_y\,(u_{i,j}/dxu_{i,j}) + dyue_{i-1,j+1}\,\delta_x\,(v_{i-1,j+1}/dyu_{i-1,j+1}) \\
e_S(10) &= dxun_{i,j}\,\delta_y\,(u_{i,j}/dxu_{i,j}) + dyue_{i,j+1}\,\delta_x\,(v_{i,j+1}/dyu_{i,j+1}) \\
e_S(11) &= dxun_{i,j-1}\,\delta_y\,(u_{i,j-1}/dxu_{i,j-1}) + dyue_{i-1,j-1}\,\delta_x\,(v_{i-1,j-1}/dyu_{i-1,j-1}) \\
e_S(12) &= dxun_{i,j-1}\,\delta_y\,(u_{i,j-1}/dxu_{i,j-1}) + dyue_{i,j-1}\,\delta_x\,(v_{i,j-1}/dyu_{i,j-1}).
\end{aligned}
\tag{17.22}
$$

The second form for tension is

$$
\begin{aligned}
e_T(1) &= \delta_x\,u_{i-1,j} - \delta_y\,v_{i,j} + (M_T)_{i,j} \\
e_T(2) &= \delta_x\,u_{i,j} - \delta_y\,v_{i,j} + (M_T)_{i,j} \\
e_T(3) &= \delta_x\,u_{i-1,j} - \delta_y\,v_{i,j-1} + (M_T)_{i,j} \\
e_T(4) &= \delta_x\,u_{i,j} - \delta_y\,v_{i,j-1} + (M_T)_{i,j} \\
e_T(5) &= \delta_x\,u_{i,j} - \delta_y\,v_{i+1,j} + (M_T)_{i+1,j} \\
e_T(6) &= \delta_x\,u_{i,j} - \delta_y\,v_{i+1,j-1} + (M_T)_{i+1,j} \\
e_T(7) &= \delta_x\,u_{i-1,j} - \delta_y\,v_{i-1,j} + (M_T)_{i-1,j} \\
e_T(8) &= \delta_x\,u_{i-1,j} - \delta_y\,v_{i-1,j-1} + (M_T)_{i-1,j} \\
e_T(9) &= \delta_x\,u_{i-1,j+1} - \delta_y\,v_{i,j} + (M_T)_{i,j+1} \\
e_T(10) &= \delta_x\,u_{i,j+1} - \delta_y\,v_{i,j} + (M_T)_{i,j+1} \\
e_T(11) &= \delta_x\,u_{i-1,j-1} - \delta_y\,v_{i,j-1} + (M_T)_{i,j-1} \\
e_T(12) &= \delta_x\,u_{i,j-1} - \delta_y\,v_{i,j-1} + (M_T)_{i,j-1}
\end{aligned}
\tag{17.23}
$$

and the second form for strain is

$$
\begin{aligned}
e_S(1) &= \delta_y\, u_{i,j} + \delta_x\, v_{i-1,j} + (M_S)_{i,j} \\
e_S(2) &= \delta_y\, u_{i,j} + \delta_x\, v_{i,j} + (M_S)_{i,j} \\
e_S(3) &= \delta_y\, u_{i,j-1} + \delta_x\, v_{i-1,j} + (M_S)_{i,j} \\
e_S(4) &= \delta_y\, u_{i,j-1} + \delta_x\, v_{i,j} + (M_S)_{i,j} \\
e_S(5) &= \delta_y\, u_{i+1,j} + \delta_x\, v_{i,j} + (M_S)_{i+1,j} \\
e_S(6) &= \delta_y\, u_{i+1,j-1} + \delta_x\, v_{i,j} + (M_S)_{i+1,j} \\
e_S(7) &= \delta_y\, u_{i-1,j} + \delta_x\, v_{i-1,j} + (M_S)_{i-1,j} \\
e_S(8) &= \delta_y\, u_{i-1,j-1} + \delta_x\, v_{i-1,j} + (M_S)_{i-1,j} \\
e_S(9) &= \delta_y\, u_{i,j} + \delta_x\, v_{i-1,j+1} + (M_S)_{i,j+1} \\
e_S(10) &= \delta_y\, u_{i,j} + \delta_x\, v_{i,j+1} + (M_S)_{i,j+1} \\
e_S(11) &= \delta_y\, u_{i,j-1} + \delta_x\, v_{i-1,j-1} + (M_S)_{i,j-1} \\
e_S(12) &= \delta_y\, u_{i,j-1} + \delta_x\, v_{i,j-1} + (M_S)_{i,j-1}.
\end{aligned}
\tag{17.24}
$$

17.3.5 Functional derivative of e_T

Choosing to work with the discretized first form of tension (equation (17.13)) leads to the functional derivative of the tension within the 12 subcells

$$
\begin{aligned}
\frac{\partial e_T(1)}{\partial (u^{(b)})_{i,j}} &= \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\,\frac{\delta_b^1}{dyu_{i,j}} + \frac{dxun_{i,j}}{dyun_{i,j}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(2)}{\partial (u^{(b)})_{i,j}} &= -\frac{dyue_{i,j}}{dxue_{i,j}}\,\frac{\delta_b^1}{dyu_{i,j}} + \frac{dxun_{i,j}}{dyun_{i,j}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(3)}{\partial (u^{(b)})_{i,j}} &= \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\,\frac{\delta_b^1}{dyu_{i,j}} - \frac{dxun_{i,j-1}}{dyun_{i,j-1}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(4)}{\partial (u^{(b)})_{i,j}} &= -\frac{dyue_{i,j}}{dxue_{i,j}}\,\frac{\delta_b^1}{dyu_{i,j}} - \frac{dxun_{i,j-1}}{dyun_{i,j-1}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(5)}{\partial (u^{(b)})_{i,j}} &= -\frac{dyue_{i,j}}{dxue_{i,j}}\,\frac{\delta_b^1}{dyu_{i,j}} \\
\frac{\partial e_T(6)}{\partial (u^{(b)})_{i,j}} &= -\frac{dyue_{i,j}}{dxue_{i,j}}\,\frac{\delta_b^1}{dyu_{i,j}} \\
\frac{\partial e_T(7)}{\partial (u^{(b)})_{i,j}} &= \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\,\frac{\delta_b^1}{dyu_{i,j}} \\
\frac{\partial e_T(8)}{\partial (u^{(b)})_{i,j}} &= \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\,\frac{\delta_b^1}{dyu_{i,j}} \\
\frac{\partial e_T(9)}{\partial (u^{(b)})_{i,j}} &= \frac{dxun_{i,j}}{dyun_{i,j}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(10)}{\partial (u^{(b)})_{i,j}} &= \frac{dxun_{i,j}}{dyun_{i,j}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(11)}{\partial (u^{(b)})_{i,j}} &= -\frac{dxun_{i,j-1}}{dyun_{i,j-1}}\,\frac{\delta_b^2}{dxu_{i,j}} \\
\frac{\partial e_T(12)}{\partial (u^{(b)})_{i,j}} &= -\frac{dxun_{i,j-1}}{dyun_{i,j-1}}\,\frac{\delta_b^2}{dxu_{i,j}}.
\end{aligned}
\tag{17.25}
$$

17.3.6 Functional derivative of e_S

Choosing to work with the discretized first form of strain (equation (17.13)) leads to the functional derivative of the strain within the 12 subcells

$$
\begin{aligned}
\frac{\partial e_S(1)}{\partial (u^{(b)})_{i,j}} &= -\frac{dxun_{i,j}}{dyun_{i,j}}\frac{\delta_b^1}{dxu_{i,j}} + \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(2)}{\partial (u^{(b)})_{i,j}} &= -\frac{dxun_{i,j}}{dyun_{i,j}}\frac{\delta_b^1}{dxu_{i,j}} - \frac{dyue_{i,j}}{dxue_{i,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(3)}{\partial (u^{(b)})_{i,j}} &= \frac{dxun_{i,j-1}}{dyun_{i,j-1}}\frac{\delta_b^1}{dxu_{i,j}} + \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(4)}{\partial (u^{(b)})_{i,j}} &= \frac{dxun_{i,j-1}}{dyun_{i,j-1}}\frac{\delta_b^1}{dxu_{i,j}} - \frac{dyue_{i,j}}{dxue_{i,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(5)}{\partial (u^{(b)})_{i,j}} &= -\frac{dyue_{i,j}}{dxue_{i,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(6)}{\partial (u^{(b)})_{i,j}} &= -\frac{dyue_{i,j}}{dxue_{i,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(7)}{\partial (u^{(b)})_{i,j}} &= \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(8)}{\partial (u^{(b)})_{i,j}} &= \frac{dyue_{i-1,j}}{dxue_{i-1,j}}\frac{\delta_b^2}{dyu_{i,j}} \\
\frac{\partial e_S(9)}{\partial (u^{(b)})_{i,j}} &= -\frac{dxun_{i,j}}{dyun_{i,j}}\frac{\delta_b^1}{dxu_{i,j}} \\
\frac{\partial e_S(10)}{\partial (u^{(b)})_{i,j}} &= -\frac{dxun_{i,j}}{dyun_{i,j}}\frac{\delta_b^1}{dxu_{i,j}} \\
\frac{\partial e_S(11)}{\partial (u^{(b)})_{i,j}} &= \frac{dxun_{i,j-1}}{dyun_{i,j-1}}\frac{\delta_b^1}{dxu_{i,j}} \\
\frac{\partial e_S(12)}{\partial (u^{(b)})_{i,j}} &= \frac{dxun_{i,j-1}}{dyun_{i,j-1}}\frac{\delta_b^1}{dxu_{i,j}}.
\end{aligned}
\tag{17.26}
$$

17.3.7 Tidy form for the discretized friction

Let us focus first on the zonal friction with $b = 1$. Using the volumes enumerated in Section 17.3.2, and dropping the k label since it is the same for all terms, we have for the tension

$$
\begin{aligned}
&-\sum_{n=1}^{12} V(n)\,\tau^{xx}(n)\,\frac{\partial e_T(n)}{\partial (u^{(1)})_{i,j}} \\
&= \left(\frac{dyue_{i,j}}{4\,dxue_{i,j}\,dyu_{i,j}}\right)\Big([\tau^{xx}(5)+\tau^{xx}(6)]\,V^U_{i+1,j} + [\tau^{xx}(2)+\tau^{xx}(4)]\,V^U_{i,j}\Big) \\
&- \left(\frac{dyue_{i-1,j}}{4\,dxue_{i-1,j}\,dyu_{i-1,j}}\right)\Big([\tau^{xx}(1)+\tau^{xx}(3)]\,V^U_{i,j} + [\tau^{xx}(7)+\tau^{xx}(8)]\,V^U_{i-1,j}\Big).
\end{aligned}
\tag{17.27}
$$

Figure 17.1 indicates that a velocity point $U_{i,j}$ is associated with four triads, each of which is used to construct a tension and strain along with a viscosity. Such arrangement motivates

the following notation (see Figure 17.2):

$$\begin{aligned}
(1,1) &= \text{northeast triad} \\
(0,1) &= \text{northwest triad} \\
(0,0) &= \text{southwest triad} \\
(1,0) &= \text{southeast triad}
\end{aligned} \tag{17.28}$$

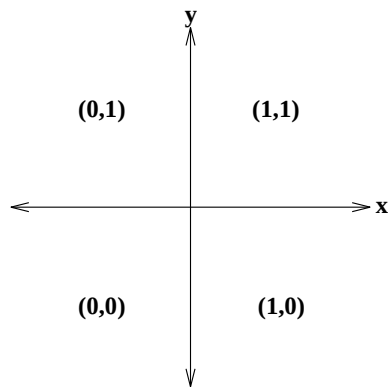

Figure 17.2: Notation for the quadrants surrounding a velocity point.

With this notation,

$$\begin{aligned}
-\sum_{n=1}^{12} V(n)\,\tau^{xx}(n)\,\frac{\partial e_T(n)}{\partial (u^{(1)})_{i,j}} &= \left(\frac{dyue_{i,j}}{4\,dxue_{i,j}\,dyu_{i,j}}\right)\sum_{ip=0}^{1} V^U_{i+ip,j}\sum_{jq=0}^{1}(\tau^{xx})^{(1-ip,jq)}_{(i+ip,j)} \\
&\quad -\left(\frac{dyue_{i-1,j}}{4\,dxue_{i-1,j}\,dyu_{i,j}}\right)\sum_{ip=0}^{1} V^U_{i-1+ip,j}\sum_{jq=0}^{1}(\tau^{xx})^{(1-ip,jq)}_{(i-1+ip,j)}.
\end{aligned} \tag{17.29}$$

The two terms on the right hand side are centered on the east and west faces, respectively, of the $U_{i,j}$ velocity cell. Consequently, we introduce the finite difference derivative operator to yield

$$-\sum_{n=1}^{12} V(n)\,\tau^{xx}(n)\,\frac{\partial e_T(n)}{\partial (u^{(1)})_{i,j}} = \left(\frac{dxu_{i,j}}{4\,dyu_{i,j}}\right)\delta_x\left[\frac{dyue_{i-1,j}}{dxue_{i-1,j}}\sum_{ip=0}^{1} V^U_{i-1+ip,j}\sum_{jq=0}^{1}(\tau^{xx})^{(1-ip,jq)}_{(i-1+ip,j)}\right]. \tag{17.30}$$

Dividing by the velocity cell volume $V^U_{i,j} = dxu_{i,j}\,dyu_{i,j}\,dhu_{i,j}$ leads to

$$\begin{aligned}
&-\left(\frac{1}{V^U_{i,j}}\right)\sum_{n=1}^{12} V(n)\,\tau^{xx}(n)\,\frac{\partial e_T(n)}{\partial (u^{(1)})_{i,j}} \\
&\qquad = \left(\frac{1}{4\,dhu_{i,j}\,(dyu_{i,j})^2}\right)\delta_x\left[\frac{dyue_{i-1,j}}{dxue_{i-1,j}}\sum_{ip=0}^{1} V^U_{i-1+ip,j}\sum_{jq=0}^{1}(\tau^{xx})^{(1-ip,jq)}_{(i-1+ip,j)}\right].
\end{aligned} \tag{17.31}$$

Similar manipulations apply for the strain terms, thus leading to

$$
\begin{aligned}
&-\left(\frac{1}{V^U_{i,j}}\right)\sum_{n=1}^{12} V(n)\,\tau^{xy}(n)\,\frac{\partial e_S(n)}{\partial (u^{(1)})_{i,j}} \\
&\qquad = \left(\frac{1}{4\,dhu_{i,j}\,(dxu_{i,j})^2}\right)\delta_y\left[\frac{dxun_{i,j-1}}{dyun_{i,j-1}}\sum_{jq=0}^{1} V^U_{i,j-1+jq}\sum_{ip=0}^{1}(\tau^{xy})^{(ip,1-jq)}_{(i,j-1+jq)}\right]. \qquad (17.32)
\end{aligned}
$$

Bringing the two pieces together leads to the zonal friction acting at the velocity cell $U_{i,j}$

$$
\begin{aligned}
\rho_o\,F^{(x)}_{i,j} &= \left(\frac{1}{4\,dhu_{i,j}\,(dyu_{i,j})^2}\right)\delta_x\left[\frac{dyue_{i-1,j}}{dxue_{i-1,j}}\sum_{ip=0}^{1} V^U_{i-1+ip,j}\sum_{jq=0}^{1}(\tau^{xx})^{(1-ip,jq)}_{(i-1+ip,j)}\right] \\
&+ \left(\frac{1}{4\,dhu_{i,j}\,(dxu_{i,j})^2}\right)\delta_y\left[\frac{dxun_{i,j-1}}{dyun_{i,j-1}}\sum_{jq=0}^{1} V^U_{i,j-1+jq}\sum_{ip=0}^{1}(\tau^{xy})^{(ip,1-jq)}_{(i,j-1+jq)}\right]. \qquad (17.33)
\end{aligned}
$$

By inspection, the meridional friction is given by

$$
\begin{aligned}
\rho_o\,F^{(y)}_{i,j} &= \left(\frac{1}{4\,dhu_{i,j}\,(dyu_{i,j})^2}\right)\delta_x\left[\frac{dyue_{i-1,j}}{dxue_{i-1,j}}\sum_{ip=0}^{1} V^U_{i-1+ip,j}\sum_{jq=0}^{1}(\tau^{xx})^{(1-ip,jq)}_{(i-1+ip,j)}\right] \\
&- \left(\frac{1}{4\,dhu_{i,j}\,(dxu_{i,j})^2}\right)\delta_y\left[\frac{dxun_{i,j-1}}{dyun_{i,j-1}}\sum_{jq=0}^{1} V^U_{i,j-1+jq}\sum_{ip=0}^{1}(\tau^{xx})^{(ip,1-jq)}_{(i,j-1+jq)}\right]. \qquad (17.34)
\end{aligned}
$$

Comparison with the continuum friction components given by equations (15.118) and (15.119) indicates that the discretization is consistent; i.e., the discrete friction reduces to the continuum friction as the grid size goes to zero.

17.3.8 Tension and strain in the quadrants

There are four tensions and strains corresponding to the four triads surrounding each velocity point. Referring to Figure 17.2, assuming the central point is $U_{i,j}$, we discretize the tensions and strains starting from the second form of the continuous tension and strain (equations (17.15) and (17.16)) to find

$$
\begin{aligned}
(e_T)_{i,j,(0,1)} &= \delta_x\,u_{i-1,j} - \delta_y\,v_{i,j} + (M_T)_{i,j} \\
(e_T)_{i,j,(1,1)} &= \delta_x\,u_{i,j} - \delta_y\,v_{i,j} + (M_T)_{i,j} \\
(e_T)_{i,j,(0,0)} &= \delta_x u_{i-1,j} - \delta_y\,v_{i,j-1} + (M_T)_{i,j} \\
(e_T)_{i,j,(1,0)} &= \delta_x\,u_{i,j} - \delta_y\,v_{i,j-1} + (M_T)_{i,j} \\
(e_S)_{i,j,(0,1)} &= \delta_y\,u_{i,j} + \delta_x\,v_{i-1,j} + (M_S)_{i,j} \\
(e_S)_{i,j,(1,1)} &= \delta_y\,u_{i,j} + \delta_x\,v_{i,j} + (M_S)_{i,j} \\
(e_S)_{i,j,(0,0)} &= \delta_y\,u_{i,j-1} + \delta_x\,v_{i-1,j} + (M_S)_{i,j} \\
(e_S)_{i,j,(1,0)} &= \delta_y\,u_{i,j-1} + \delta_x\,v_{i,j} + (M_S)_{i,j}
\end{aligned} \qquad (17.35)
$$

In general, the four tensions can be written

$$
(e_T)_{i,j,(ip,jq)} = \delta_x\,u_{i+ip-1,j} - \delta_y\,v_{i,j+jq-1} + (M_T)_{i,j} \qquad (17.36)
$$

and the four strains can be written

$$
(e_S)_{i,j,(ip,jq)} = \delta_y\,u_{i,j+jq-1} + \delta_x\,v_{i+ip-1,j} + (M_S)_{i,j} \qquad (17.37)
$$

where $ip = 0, 1$ and $jq = 0, 1$. Notably, the metric terms

$$(M_T)_{i,j} = -u_{i,j}\,(\partial_x \ln dy)_{i,j} + v_{i,j}\,(\partial_y \ln dx)_{i,j} \tag{17.38}$$

$$(M_S)_{i,j} = -u_{i,j}\,(\partial_y \ln dx)_{i,j} - v_{i,j}\,(\partial_x \ln dy)_{i,j} \tag{17.39}$$

are common to the four triads, and so only need be computed once per velocity point. Generally, the four tensions and strains are computed in the model and are then used to compute the friction operator. When the Smagorinsky viscosity scheme (Sections 16.2 and 17.3.10) is enabled, they are used to compute the Smagorinsky viscosity as well.

17.3.9 Comments

This section contains some specific comments.

- The tension and strain for ocean points next to land contain a contribution from a zero velocity living at the land-sea interface. In order to provide a full accounting of the generally strong shears next to no-slip walls, it is important to include such contributions rather than masking them out.
- In the special case of a uniform Cartesian grid, a constant isotropic viscosity, and a zero anisotropic viscosity, the functionally derived discrete friction operator reduces to the familiar 5-point discrete Laplacian. This operator has familiar dissipative computational properties.
- Practical experience has revealed problems with the discretization for bottom-most grid cells when these cells are thin partial cells that are surrounded by thicker partial cells. The problem is that contributions from surrounding thick cells are then normalized by the thin *dhu* of the central cell. To alleviate this problem, it is effective to use the traditional 5-point Laplacian operator for computing friction in the bottom-most grid cells.

17.3.10 Discretized Smagorinsky viscosity

The nonlinear viscosity coefficient is determined in terms of the deformation rate and grid spacing. Since e_T and e_S involve terms with derivatives in both horizontal directions, there needs to be an averaging performed to place them at a common grid position.

Pacanowski et al. (1991) (MOM1) defined both deformation rates at the north face of the U-cell. This is the natural position for the meridional derivative terms. To get the zonal derivative terms defined at the north face, it was necessary to average over the four zonal derivatives surrounding the north face. The problem with such "four point" averages on the B-grid is that they can introduce computational modes. Computational modes are not always problematical if there are other processes that can suppress the growth of the modes. The problem with the computational modes in the Smagorinsky scheme is that they allow nontrivial field configurations yielding a zero deformation rate. Hence, they produce a zero Smagorinsky viscosity and so are not dissipated. Furthermore, these modes represent grid scale waves, which are the waves an ideal implementation of the Smagorinsky scheme should dissipate the most. Therefore, it is not acceptable to allow these modes in the discretized Smagorinsky scheme. Another approach is necessary.

The functional discretization described in this chapter eliminates the computational modes. For each velocity triad, there is a corresponding Smagorinsky viscosity. In particular, referring to the deformation rates defined in equation (17.35), we have the corresponding Smagorinsky diffusivities

$$A_{i,j,(ip,jq)} = (\Upsilon \Delta s/\pi)^2 |E|_{i,j,(ip,jq)} \tag{17.40}$$

where $ip = 0, 1$, $jq = 0, 1$ are the triad labels, and

$$|E|_{i,j,(ip,jq)} = \sqrt{[(e_T)_{i,j,(ip,jq)}]^2 + [(e_S)_{i,j,(ip,jq)}]^2} \tag{17.41}$$

is the discrete total deformation rate. As mentioned earlier, one advantage of the functional approach over the "Laplacian plus metric" approach (see Section 17.4) is the exploitation of the deformation rates for computing both the Smagorinsky viscosity and the friction operator.

17.4 Laplacian plus metric form of isotropic friction

The purpose of this section is to illustrate the mathematical equivalence between the "Laplacian plus metric" form of friction and the stress tensor approach. Recall that we commented on the numerical disadvantages of the Laplacian plus metric approach in Section 17.2. We focus here on isotropic friction using spherical coordinates.

17.4.1 Spherical form of second order friction

In spherical coordinates, the metric takes the form

$$g_{ij} = \text{diag}\,(g_{11}, g_{22}, g_{33}) = \text{diag}\,(R^2 \cos^2\phi, R^2, 1) \tag{17.42}$$

and horizontal increments are

$$(dx, dy, dz) = (R\cos\phi\, d\lambda, R\, d\phi, dz). \tag{17.43}$$

Consequently, it is only g_{11} that has nonzero spatial dependence. The friction then can be written

$$F^x = \frac{\partial(A\, e_T)}{\partial x} + \frac{1}{\cos^2\phi}\frac{\partial(A\, \cos^2\phi\, e_S)}{\partial y} + (\kappa\, u_{,z})_{,z} \tag{17.44}$$

$$F^y = \frac{\partial(A\, e_S)}{\partial x} - \frac{1}{\cos^2\phi}\frac{\partial(A\, \cos^2\phi\, e_T)}{\partial y} + (\kappa\, v_{,z})_{,z} \tag{17.45}$$

where the deformation rates are

$$\begin{aligned} e_T &= u_{,x} - v_{,y} - (v/R)\, \tan\phi \\ &= (R\cos\phi)^{-1}\,(u_{,\lambda} - v_{,\phi}\, \cos\phi - v\, \sin\phi) \\ e_S &= v_{,x} + \cos\phi\,(u/\cos\phi)_{,y} \\ &= (R\cos\phi)^{-1}\,(v_{,\lambda} + u_{,\phi}\, \cos\phi + u\, \sin\phi). \end{aligned} \tag{17.46}$$

The terms associated with horizontal deformations

$$F^u = \frac{1}{R\, \cos\phi}\frac{\partial(A\, e_T)}{\partial\lambda} + \frac{1}{R\, \cos^2\phi}\frac{\partial(A\, \cos^2\phi\, e_S)}{\partial\phi} \tag{17.47}$$

$$F^v = \frac{1}{R\, \cos\phi}\frac{\partial(A\, e_S)}{\partial\lambda} - \frac{1}{R\, \cos^2\phi}\frac{\partial(A\, \cos^2\phi\, e_T)}{\partial\phi} \tag{17.48}$$

can be massaged into the form presented by Bryan (1969) and Wajsowicz (1993). That is the purpose of the remainder of this section.

17.4.2 Zonal friction

The lateral friction acting on the zonal velocity takes the expanded form

$$
\begin{aligned}
R\,F^u &= \frac{1}{\cos\phi}\,(e_T\,A_{,\lambda} + A\,\partial_\lambda\,e_T) \\
&+ \frac{1}{\cos^2\phi}\,(e_S\,A_{,\phi}\,\cos^2\phi + A\,\partial_\phi e_S\,\cos^2\phi - 2\,A\,e_S\,\cos\phi\sin\phi) \\
&= \frac{1}{\cos\phi}\,(e_T\,A_{,\lambda} + e_S\,A_{,\phi}\,\cos\phi) \\
&- \frac{2A}{R\cos\phi}(v_{,\lambda}\,\tan\phi + u_{,\phi}\,\sin\phi + u\,\sin\phi\,\tan\phi) \\
&+ \frac{A}{R\,\cos\phi}(u_{,\lambda\lambda}\,\sec\phi + u_{,\phi\phi}\cos\phi + u_{,\phi}\,\sin\phi + u\,\sec\phi) \\
&= \frac{1}{\cos\phi}\,(e_T\,A_{,\lambda} + e_S\,A_{,\phi}\,\cos\phi) \\
&+ \frac{A}{R}\left(u_{,\lambda\lambda}\,\sec^2\phi + u_{,\phi\phi} - u_{,\phi}\tan\phi + u\,(\sec^2\phi - 2\tan^2\phi) - 2v_{,\lambda}\sec^2\phi\sin\phi\right) \\
&= \frac{1}{\cos\phi}\,(e_T\,A_{,\lambda} + e_S\,A_{,\phi}\,\cos\phi) \\
&+ \frac{A}{R}\left(u_{,\lambda\lambda}\,\sec^2\phi + \sec\phi\,(u_{,\phi}\cos\phi)_{,\phi} + u(1-\tan^2\phi) - 2v_{,\lambda}\sec^2\phi\sin\phi\right), \qquad (17.49)
\end{aligned}
$$

which renders

$$
\begin{aligned}
F^u &= A\left(\nabla_z^2\,u + \frac{u(1-\tan^2\phi)}{R^2} - \frac{2v_{,\lambda}\,\sin\phi}{R^2\cos^2\phi}\right) \\
&+ \frac{1}{R\,\cos\phi}\,(e_T\,A_{,\lambda} + e_S\,A_{,\phi}\,\cos\phi). \qquad (17.50)
\end{aligned}
$$

The second bracketed term in this expression arises from the spatial dependence of the viscosity coefficient, and should be present for any non-constant viscosity coefficient model. These non-constant viscosity coefficient terms amount to those identified by Wajsowicz (1993).

It is useful to perform one final step in the formulation in order to bring the non-constant viscosity coefficient inside the Laplacian. For this purpose, the Laplacian and non-constant viscosity coefficient terms are expanded to yield

$$
\begin{aligned}
F^u &= \frac{A}{R^2\cos^2\phi}u_{,\lambda\lambda} + \frac{A}{R^2\cos\phi}(u_{,\phi}\cos\phi)_{,\phi} + \frac{A\,u\,(1-\tan^2\phi)}{R^2} - \frac{2\,A\,v_{,\lambda}\,\sin\phi}{R^2\cos^2\phi} \\
&+ \frac{A_\lambda}{R^2\cos^2\phi}(u_{,\lambda} - v_{,\phi}\,\cos\phi - v\,\sin\phi) + \frac{A_{,\phi}}{R^2\cos\phi}(v_{,\lambda} + u_{,\phi}\,\cos\phi + u\,\sin\phi) \\
&= \frac{1}{R^2\cos^2\phi}(A\,u_{,\lambda\lambda} + A_{,\lambda}\,u_{,\lambda}) + \left(\frac{1}{R^2\cos\phi}(A\,u_{,\phi}\cos\phi)_{,\phi} - A_{,\phi}\,u_{,\phi}R^{-2}\right) \\
&+ \frac{A\,u\,(1-\tan^2\phi)}{R^2} - \frac{2\,A\,v_{,\lambda}\,\sin\phi}{R^2\cos^2\phi} \\
&- \frac{A_{,\lambda}}{R^2\cos^2\phi}(v_{,\phi}\,\cos\phi + v\,\sin\phi) + R^{-2}\,A_\phi(v_{,\lambda}\sec\phi + u_{,\phi} + u\,\tan\phi) \\
&= \nabla_z\cdot(A\,\nabla_z u) + \frac{A\,u\,(1-\tan^2\phi)}{R^2} - \frac{2\,A\,v_{,\lambda}\,\sin\phi}{R^2\cos^2\phi} \\
&- \frac{A_{,\lambda}}{R^2\cos\phi}(v_{,\phi} + v\,\tan\phi) + \frac{A_{,\phi}}{R^2\cos\phi}(v_{,\lambda} + u\,\sin\phi). \qquad (17.51)
\end{aligned}
$$

This expression can be written as

$$\nabla_z \cdot (A\,\nabla_z u) + \text{old_metric}^u + \text{new_metric}^u, \tag{17.52}$$

where

$$\nabla_z \cdot (A\,\nabla_z u) = \frac{1}{R^2\cos^2\phi}(A u_{,\lambda})_{,\lambda} + \frac{1}{R^2\cos\phi}(A\,u_{,\phi}\cos\phi)_{,\phi} \tag{17.53}$$

is the horizontal Laplacian with the generally non-constant viscosity coefficient inserted. Note that this Laplacian is acting on the zonal velocity as if it was a scalar field. The term

$$\text{old_metric}^u = \frac{A\,u\,(1-\tan^2\phi)}{R^2} - \frac{2\,A\,v_{,\lambda}\,\sin\phi}{R^2\cos^2\phi} \tag{17.54}$$

is the metric term employed for constant horizontal viscosity coefficient (Bryan (1969)), and

$$\text{new_metric}^u = -\frac{A_{,\lambda}}{R^2\cos\phi}(v_{,\phi} + v\,\tan\phi) + \frac{A_{,\phi}}{R^2\cos\phi}(v_{,\lambda} + u\,\sin\phi) \tag{17.55}$$

is the metric term arising from spatial dependence in the viscosity coefficient (Wajsowicz (1993)).

17.4.3 Meridional friction

Repeating the exercise just performed for the zonal friction yields the following lines of algebra for the meridional friction

$$\begin{aligned}
R\,F^v &= \frac{1}{\cos\phi}\,(e_S\,A_{,\lambda} + A\,\partial_\lambda\,e_S) \\
&- \frac{1}{\cos^2\phi}\,(e_T\,A_{,\phi}\,\cos^2\phi + A\,\partial_\phi e_T\,\cos^2\phi - 2\,A\,e_T\,\cos\phi\sin\phi) \\
&= \frac{1}{\cos\phi}\,(e_S\,A_{,\lambda} - e_T\,A_{,\phi}\,\cos\phi) \\
&+ \frac{2A}{R\,\cos\phi}(u_{,\lambda}\tan\phi - v_{,\phi}\tan\phi\cos\phi - v\tan\phi\sin\phi) \\
&+ \frac{A}{R\,\cos\phi}\,(v_{,\lambda\lambda}\,\sec\phi + v_{,\phi\phi}\,\cos\phi + v_{,\phi}\,\sin\phi + v\,\sec\phi) \\
&= \frac{1}{\cos\phi}\,(e_S\,A_{,\lambda} - e_T\,A_{,\phi}\,\cos\phi) \\
&+ \frac{A}{R}\left(v_{,\lambda\lambda}\sec^2\phi + v_{,\phi\phi} - v_{,\phi}\tan\phi + v(\sec^2\phi - 2\tan^2\phi) + 2u_{,\lambda}\sec^2\phi\sin\phi\right) \\
&= \frac{1}{\cos\phi}\,(e_S\,A_{,\lambda} - e_T\,A_{,\phi}\,\cos\phi) \\
&+ \frac{A}{R}\left(v_{,\lambda\lambda}\sec^2\phi + \sec\phi\,(v_{,\phi}\cos\phi)_{,\phi} + v(1-\tan^2\phi) + 2u_{,\lambda}\sec^2\phi\sin\phi\right),
\end{aligned} \tag{17.56}$$

which renders

$$\begin{aligned}
F^v = A&\left(\nabla_z^2\,v + \frac{v(1-\tan^2\phi)}{R^2} + \frac{2u_{,\lambda}\,\sin\phi}{R^2\cos^2\phi}\right) \\
&+ \frac{1}{R\,\cos\phi}\,(e_S\,A_{,\lambda} - e_T\,A_{,\phi}\,\cos\phi).
\end{aligned} \tag{17.57}$$

Again, it is useful to expand this friction one more step in order to bring the viscosity coefficient inside the Laplacian. This manipulation yields

$$F^v = \nabla_z \cdot (A\,\nabla_z v) + \text{old_metric}^v + \text{new_metric}^v, \tag{17.58}$$

where

$$\text{old_metric}^v = \frac{A\,v\,(1-\tan^2\phi)}{R^2} + \frac{2\,A\,u_{,\lambda}\,\sin\phi}{R^2\cos^2\phi} \tag{17.59}$$

is the metric employed for constant horizontal viscosity coefficient, and

$$\text{new_metric}^v = \frac{A_{,\lambda}}{R^2\cos\phi}(u_{,\phi} + u\,\tan\phi) + \frac{A_{,\phi}}{R^2\cos\phi}(-u_{,\lambda} + v\,\sin\phi). \tag{17.60}$$

Part VI

TENSOR ANALYSIS

Tensor Analysis

The purpose of this part of the book is to familiarize the reader with the basic tools of general tensor analysis. An understanding of tensor analysis is necessary to formulate the equations of rotating fluid dynamics on a sphere. The presentation is aimed at the reader who has a solid foundation in undergraduate vector calculus. No exposure to Cartesian or general tensors is assumed.

The logical place for this material is at the start of this book. However, it is placed at the end to allow the pedestrian reader ample opportunity to skip this material. Indeed, most practioners of geophysical fluid dynamics have little use for tensor analysis, other than to note the appropriate formulae in their favorite text. Nonetheless, tensor analysis is important for those formulating and building modern ocean climate models, and so its presence in this book is appropriate.

CHAPTER

EIGHTEEN

Tensor analysis basics

Contents

18.1 Introduction **396**
18.2 Some practical motivation **397**
18.2.1 Generalized horizontal coordinates 397
18.2.2 Generalized vertical coordinates 398
18.3 Coordinates and vectors **398**
18.4 The metric and coordinate transformations **401**
18.4.1 Transformation between Cartesian and spherical 402
18.4.2 Metric for S^2 403
18.5 Transformations of a vector **403**
18.6 One-forms **404**
18.7 Metric as the map between vectors and one-forms **406**
18.8 Transformation of a one-form **406**
18.9 Arbitrary tensors and their transformations **407**
18.10 Tensorial properties of the gradient operator **408**
18.11 The invariant volume element **409**
18.11.1 Derivation of the invariant volume element 409
18.11.2 Invariant volume element in different coordinates 410
18.12 Determinants and the Levi-Civita symbol **411**
18.13 Surfaces embedded in Euclidean space **413**
18.13.1 Induced metric tensor for a surface 413
18.13.2 The invariant surface area element 414
18.13.3 Surface and space components of a tensor 414
18.13.4 Normals and tangents 415

The purpose of this chapter is to introduce basic notions of tensor analysis fundamental to the mathematical formulation of geophysical fluid mechanics on a sphere.

18.1 Introduction

Physics is independent of coordinates. Such a coordinate independent framework lies at the core of theoretical physics. However, the deduction of physical laws can at times be initiated through working within the framework of a particular set of coordinates. For example, it is often convenient to use Cartesian coordinates when describing physics on a flat manifold. Additionally, the implementation of physical laws for numerical computation *requires* a chosen set of coordinates. Hence, those working in geophysical fluid dynamics (GFD), and in particular numerical GFD, find it useful to be familiar with various sets of coordinates and the transformations of physical laws from one set of coordinates to another.

Differential geometry is an area of mathematics that provides tools for use in working with coordinates. Additionally, it helps to expose the underlying physical essence of a mathematical statement, rather than being encumbered with coordinate dependent details. Differential geometry is yet another name for calculus as defined on arbitrary smooth manifolds. It has its roots in classical vector analysis and calculus, and so many of its results are familiar to those having studied elementary calculus. Over time, the language and notation of differential geometry has been refined to the point where tensor indices are nearly absent in modern treatments. This *index free* approach is not taken here, since to do so would exclude a large number of readers. Rather, the treatment here employs tensorial notation, with a sprinkling of the modern language thrown in where clarity is manifest. The topic of this chapter is therefore basic tensor analysis from the perspective of elementary differential geometry.

Consequently, the purpose of this and the next chapter is to expose the reader to elements of calculus on manifolds, also known as tensor analysis, with particular attention placed on applications to geophysical fluid dynamics. In the process, many formulae proving useful in this book are derived and listed here.

As with any mathematical subject, there is a fair amount of notation and convention to absorb. Indeed, the rules and gymnastics involved can appear quite tedious. However, the rules are systematic and have sound motivation. A goal of this chapter and the next is to expose some of the reasoning leading to the formalism, and so to reduce the tedium and intimidation typically associated with tensor analysis.

Those readers interested in somewhat more complete treatments of differential geometry, with physical applications in mind, are encouraged to read one or more of the following books:

- Aris (1962): **Vectors, Tensors and the Basic Equations of Fluid Mechanics**. This book presents the basic ideas of tensor analysis within the context of fluid dynamics. The treatment is general and the notation is clean. Aris's approach is used throughout this book.

- Schutz (1980): **Geometrical Methods of Mathematical Physics** and Schutz (1985): **A First Course in General Relativity**. The first book provides a discussion of rather deep mathematical ideas central to much of modern physics, with some application to fluid mechanics. The second book is a very readable presentation of introductory general relativity. Much of the approach taken

in the relativity book is followed here, and it provides another of the central references for this chapter. Additionally, note that nearly all introductory general relativity books include a discussion of tensor analysis in a manner consistent with that given here.

- Flanders (1989): **Differential Forms with Applications to the Physical Sciences**. This book, first published in 1963, is an early work aiming towards the use of index-free notation in physical theories. Flanders provides some applications to fluid mechanics.

- Morse and Feshbach (1953): **Methods of Theoretical Physics**. This book, which comes in two volumes, is a classic. It is listed here since it articulates an older and more cumbersome formalism not employed in the present book. Nevertheless, certain authors still employ Morse and Feshbach as their primary reference for tensor analysis. Hence, it might be useful for some to become familiar with their notation.

18.2 Some practical motivation

It is useful to begin with some practical motivation for learning the methods of tensor analysis for use in GFD.

18.2.1 Generalized horizontal coordinates

When studying geophysical fluids on a sphere, the horizontal coordinates have traditionally been the spherical latitude-longitude coordinates (λ, ϕ), with $(\mathrm{d}x, \mathrm{d}y) = R\,(\cos\phi\,\mathrm{d}\lambda, \mathrm{d}\phi)$ being the infinitesimal distance along the two orthogonal coordinate axes, with R the sphere's radius. Specifically, the coordinate ϕ is latitude, which increases northward and is zero at the equator. λ is longitude, which increases eastward with zero defined at an arbitrary longitude (e.g., Greenwich, England) (see Figure 18.1). Spherical coordinates are locally orthogonal, thus greatly simplifying the representation of the governing equations of motion.

Generalizations of the spherical coordinates are possible, and are becoming quite popular in global ocean modeling. The central reason is to avoid the spherical coordinate singularity at the north pole. Generalized horizontal coordinates can be designed, after a bit of mathematical analysis, so that the polar singularity is moved over land instead of being in the ocean. As the Arctic circulation is quite important for global climate simulations, it is necessary to avoid coordinate artifacts, such as a polar singularity, in these models. Additionally, practical limitations of the time step available when the meridions converge greatly hinder the utility of spherical coordinates.

In the generalizations of spherical coordinates, ocean modelers have predominantly chosen not to sacrifice the simplifications inherent in orthogonal curvilinear coordinates. Section 19.11 presents many of the mathematical formulae needed to derive the fluid equations of motion with generalized orthogonal coordinates. Once the fluid equations are written with the generalized metric factors, one of the many sophisticated orthogonal coordinate choices (e.g., Murray (1996)) is available to the modeler in order to remove the polar singularity from the ocean domain.

The papers by Smith et al. (1995); Murray (1996); Madec and Imbard (1996) provide discussions of ocean modeling in generalized horizontal coordinates.

18.2.2 Generalized vertical coordinates

GFD consists of the study of rotating and stratified fluids. To describe stratified fluids, it is often useful to employ specialized vertical coordinates that are distinct from the usual vertical distance from the bottom topography. The mathematical issues of generalized vertical coordinates are discussed in Chapter 5. Many aspects of tensor analysis described here prove to be of use.

18.3 Coordinates and vectors

A surface possessing certain smoothness properties affords the usual rules of calculus, such as differentiation and integration; i.e., it is a *differential manifold*. Importantly, these manifolds are locally Euclidean, which means that they possess flat tangent planes. These are the sorts of surfaces of interest in this book. One familiar two dimensional example is the surface of a sphere S^2. Another is three-dimensional Euclidean space $\mathcal{R}^3$. A third is a constant density surface, commonly known as an *isopycnal* surface, in a stratified fluid. An isopycnal surface results from taking a suitable space-time or ensemble average to smooth out microscale fluctuations.

To specify the position of a point on the manifold, an observer living on the manifold may decide to set up a grid in some systematic fashion using a set of coordinates. Let the observer decide to call these coordinates ξ^a, where the raised index a can take on integer values up to M, the dimension of the manifold.* For example, a person living in $\mathcal{R}^3$ may decide to use Cartesian coordinates or spherical coordinates, depending on the application. Figure 18.1 illustrates their relation. In contrast, a person confined to live on the surface of a sphere most likely will decide to use the spherical angles (λ, ϕ), thus setting up the familiar latitude and longitude lines on S^2. A person confined to live on a general undulating surface, such as an isopycnal, may decide to use some general curvilinear coordinates. The analytical relation between these general coordinates and the more familiar Cartesian coordinates might be quite complicated, assuming such a relation even exists.

At any point on a differential manifold, the directions in which the coordinates increase define special directions. For an M-dimensional manifold, M directions form a basis for representing all vectors originating from this point. An arbitrary *first order tensor* can be written as the linear combination

$$\boldsymbol{\xi} = \sum_{a=1}^{M} \xi^a \, \vec{e}_a \equiv \xi^a \, \vec{e}_a, \tag{18.1}$$

where $\vec{e}_a$ are a set of M linearly independent vectors forming a basis for the manifold. A first order tensor represented in terms of a basis of vectors, such as $\boldsymbol{\xi}$, is often called a *vector* for brevity. Note the use of an arrow to denote a basis vector and a boldface to denote a general tensor. The second expression in equation

*This raised index should not be confused with ξ raised to the $a'th$ power. The distinction should be clear from the context.

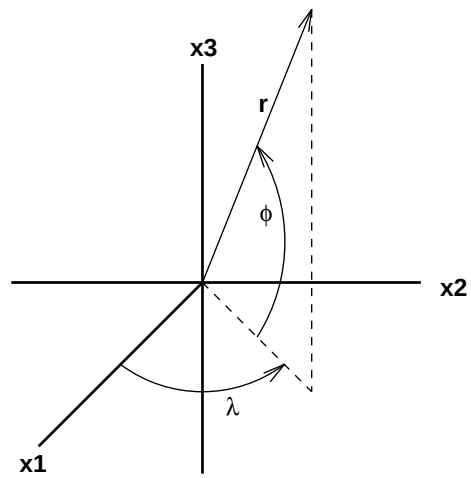

Figure 18.1: The position vector for a point in 3D Euclidean space can be represented in terms of many sets of coordinates. The simplest is the Cartesian triad (x^1, x^2, x^3). Another is the spherical coordinate system (λ, ϕ, r). In a geophysical context, the angular coordinate $0 \leq \lambda \leq 2\pi$ is the longitude, with positive values measured eastward from Greenwich, England. The angular coordinate ϕ is the latitude, with values $\phi = 0$ at the equator and $\phi = \pi/2(-\pi/2)$ at the north (south) poles. The radial distance r is measured here with respect to the center of the sphere. The coordinate transformation between Cartesian and spherical is given by $(x^1, x^2, x^3) = (r\,\cos\phi\cos\lambda, r\,\cos\phi\sin\lambda, r\,\sin\phi)$.

(18.1) defines the Einstein summation convention, for which repeated indices are summed when one index is raised and the other lowered. The choice to lower the index on the basis vectors $\vec{e}_a$ is motivated by this very useful convention, whose power will become apparent in subsequent discussions. To summarize, it is said that $\boldsymbol{\xi}$ is a first order tensor that has a *representation* given by the coordinates ξ^a in terms of the basis vectors $\vec{e}_a$.

Using the above notation, the components of a basis vector, as written in terms of another basis, take the form

$$\vec{e}_a = \Lambda^{\bar{b}}_{\ a}\,\vec{e}_{\bar{b}}, \tag{18.2}$$

where $\Lambda^{\bar{b}}_{\ a}$ is a number representing the components of the basis vector $\vec{e}_a$ as written in terms of another basis $\vec{e}_{\bar{b}}$. That is, $\Lambda^{\bar{b}}_{\ a}$ are the components to the transformation matrix between the bases. We have more to say about transformations in Section 18.4 and the following.

When the bases are identical, $\vec{e}_{\bar{a}} = \vec{e}_a$, the components of the transformation are simply given by the Kronecker symbol $\Lambda^{\bar{b}}_{\ a} = \delta^{\bar{b}}_{\ a}$, where

$$\delta^{\bar{b}}_{\ a} = \begin{cases} 1, \text{ if } \ a = \bar{b} \\ 0, \text{ if } \ a \neq \bar{b}. \end{cases} \tag{18.3}$$

The Kronecker symbol forms an invariant representation of the second order *unit* tensor. It is invariant since it takes the same numerical values when written in any coordinate basis. The issues of tensorial order are discussed later in this chapter.

Some examples are useful. First, in the Euclidean space $\mathcal{R}^3$, the following set of vectors, based on the spherical coordinates, are suitable as a basis for all points except the poles

$$\vec{e}_1 = r\,\cos\phi\,\hat{\boldsymbol{\lambda}} \tag{18.4}$$

$$\vec{e}_2 = r\,\hat{\boldsymbol{\phi}} \tag{18.5}$$

$$\vec{e}_3 = \hat{\mathbf{r}}. \tag{18.6}$$

The unit directions, denoted by hats, point in the longitudinal (eastward) direction ($\hat{\boldsymbol{\lambda}}$), latitudinal (northward) direction ($\hat{\boldsymbol{\phi}}$), and radial direction ($\hat{\mathbf{r}}$). At the poles, $\phi = \pm\pi/2$, and so the spherical basis is singular. Another set of basis vectors is the familiar Cartesian unit vectors

$$\vec{e}_1 = \hat{\mathbf{x}} \tag{18.7}$$

$$\vec{e}_2 = \hat{\mathbf{y}} \tag{18.8}$$

$$\vec{e}_3 = \hat{\mathbf{z}}. \tag{18.9}$$

Both the spherical and Cartesian basis vectors form mutually orthogonal sets when using the familiar Euclidean metric. Furthermore, the Cartesian set is normalized under the Euclidean metric, so that

$$\vec{e}_a \cdot \vec{e}_b = \delta_{ab}, \tag{18.10}$$

where δ_{ab} is yet another representation of the unit tensor to be discussed in the next section. The spherical basis is not normalized, however, and the motivation for not normalizing this basis will become apparent later.

For describing a general surface, such as an isopycnal surface, one may use the following orthonormal basis vectors

$$\vec{e}_1 = \frac{\hat{\mathbf{y}} \wedge \nabla\rho}{|\hat{\mathbf{y}} \wedge \nabla\rho|} \tag{18.11}$$

$$\vec{e}_2 = \vec{e}_3 \wedge \vec{e}_1 \tag{18.12}$$

$$\vec{e}_3 = \hat{\boldsymbol{\rho}}, \tag{18.13}$$

where the wedge symbol $\wedge$ represents the usual Cartesian vector cross product. The unit normal

$$\hat{\boldsymbol{\rho}} = \frac{\nabla\rho}{|\nabla\rho|} \tag{18.14}$$

points outward to the surfaces of constant density $\rho(x, y, z, t)$. The gradient has the following representation in terms of the Cartesian basis vectors

$$\nabla\rho = \hat{\mathbf{x}}\,\rho_{,x} + \hat{\mathbf{y}}\,\rho_{,y} + \hat{\mathbf{z}}\,\rho_{,z}, \tag{18.15}$$

where

$$\rho_{,a} = \partial_a\,\rho \tag{18.16}$$

is a useful notation further motivated in later discussions. Note that when the isopycnal surface has an outward normal $\hat{\boldsymbol{\rho}} = \hat{\mathbf{z}}$, which is parallel to the vertical, then the basis vectors reduce to $\vec{e}_1 \to \hat{\mathbf{x}}, \vec{e}_3 \to \hat{\mathbf{y}}, \vec{e}_3 \to \hat{\mathbf{z}}$. In other words, the surface is locally flat, or in the horizontal plane. In the general case, each of the vectors changes whenever the isopycnal surface evolves.

18.4 The metric and coordinate transformations

Upon specifying the position of a point with respect to some origin, it is of interest to determine how far this point is from some other point. Since the points are assumed to live on a smooth manifold, it is sufficient to consider the distance between two infinitesimally close points, and then use integration to find the distance between finitely displaced points. Let the coordinates for the two points be given by ξ^a and $\xi^a + \mathrm{d}\xi^a$, and let $\mathrm{d}\boldsymbol{\xi} = \mathrm{d}\xi^a\,\vec{e}_a$ be the infinitesimal vector pointing from one point to the other, where $\vec{e}_a$ are the basis vectors. If these points lived in the flat Euclidean $\mathcal{R}^3$ space, and the coordinates ξ^a are the usual Cartesian coordinates, then the squared distance between them is simply given by the familiar Euclidean norm

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}\boldsymbol{\xi}\cdot\mathrm{d}\boldsymbol{\xi}\\ &= \vec{e}_a\cdot\vec{e}_b\,\mathrm{d}\xi^a\,\mathrm{d}\xi^b\\ &= \delta_{ab}\,\mathrm{d}\xi^a\,\mathrm{d}\xi^b.\end{aligned}\tag{18.17}$$

In this expression, $(\mathrm{d}s)^2 \equiv \mathrm{d}s^2$ is the conventional way to write the infinitesimal squared length. The Kronecker symbol δ_{ab} is symmetric $\delta_{ab} = \delta_{ba}$, vanishes when $a \neq b$, and is unity when $a = b$. As defined by equation (18.17), δ_{ab} forms the Cartesian components to the *metric tensor* for Euclidean space. Note that the properties of δ_{ab} are also shared by $\delta^{\bar{a}}_{\ b}$ as defined by equation (18.3). These very simple properties of the Kronecker symbol, or the unit tensor, make much of the formalism of general tensor analysis in Euclidean space using Cartesian coordinates unnecessary. Yet when working on non-Euclidean spaces, such as spheres, or with general curvilinear coordinates, such as spherical coordinates, tensor analysis provides a very transparent and elegant framework.

As seen by equation (18.17), the representation of the metric tensor for Euclidean space using Cartesian coordiantes is δ_{ab}. To determine the representation in another set of coordinates, consider the transformation from Cartesian coordinates ξ^a to arbitrary coordinates $\xi^{\bar{a}}$. Use of the chain rule leads to the equivalent expression for the squared infinitesimal length

$$\begin{aligned}\mathrm{d}s^2 &= \delta_{ab}\left(\frac{\partial\xi^a}{\partial\xi^{\bar{a}}}\frac{\partial\xi^b}{\partial\xi^{\bar{b}}}\right)\mathrm{d}\xi^{\bar{a}}\,\mathrm{d}\xi^{\bar{b}}\\ &= \delta_{ab}\Lambda^a_{\ \bar{a}}\,\Lambda^b_{\ \bar{b}}\,\mathrm{d}\xi^{\bar{a}}\,\mathrm{d}\xi^{\bar{b}}\\ &= g_{\bar{a}\bar{b}}\,\mathrm{d}\xi^{\bar{a}}\,\mathrm{d}\xi^{\bar{b}},\end{aligned}\tag{18.18}$$

where

$$g_{\bar{a}\bar{b}} = \delta_{ab}\Lambda^a_{\ \bar{a}}\,\Lambda^b_{\ \bar{b}}\tag{18.19}$$

define the components to the metric tensor as represented by the new set of coordinates $\xi^{\bar{a}}$. Also introduced in these expressions are the elements to the transformation matrix

$$\Lambda^a_{\ \bar{a}} = \frac{\partial\xi^a}{\partial\xi^{\bar{a}}}.\tag{18.20}$$

This matrix, whose rows are denoted by a and columns by $\bar{a}$, is non-singular for well-defined coordinate transformations. Its determinant, called the *Jacobian of the transformation*, is therefore non-vanishing and single signed.

Once the metric is determined, the distance along a curve between two finitely separated points on the manifold is given by the integration

$$\begin{aligned} L &= \int \sqrt{\mathrm{d}s^2} \\ &= \int_{\varphi_1}^{\varphi_2} \left| g_{ab} \frac{\mathrm{d}\xi^a}{\mathrm{d}\varphi} \frac{\mathrm{d}\xi^b}{\mathrm{d}\varphi} \right|^{1/2} \mathrm{d}\varphi, \end{aligned} \tag{18.21}$$

where φ is a parameter specifying the curve, and $\varphi_{1,2}$ are the curve's endpoints. For example, if the curve is a path on a space-time manifold traversed by a particle, φ would be the time measured by a clock carried along by the particle (the *proper time*).

Before proceeding with the general discussion, it is useful to establish some explicit results with the two most familiar sets of coordinates used in GFD: Cartesian and spherical.

18.4.1 Transformation between Cartesian and spherical

Consider the transformation between Cartesian and spherical coordinates (see Figure 18.1)

$$(\xi^1, \xi^2, \xi^3) = (x^1, x^2, x^3) \tag{18.22}$$

$$(\xi^{\bar{1}}, \xi^{\bar{2}}, \xi^{\bar{3}}) = (\lambda, \phi, r). \tag{18.23}$$

These coordinates are related by the expressions

$$x^1 = r \cos\phi \cos\lambda \tag{18.24}$$

$$x^2 = r \cos\phi \sin\lambda \tag{18.25}$$

$$x^3 = r \sin\phi, \tag{18.26}$$

the transformation matrix is given by

$$\Lambda^a{}_{\bar{a}} = \begin{pmatrix} -r \cos\phi \sin\lambda & -r \sin\phi \cos\lambda & \cos\phi \cos\lambda \\ r \cos\phi \cos\lambda & -r \sin\phi \sin\lambda & \cos\phi \sin\lambda \\ 0 & r \cos\phi & \sin\phi \end{pmatrix}, \tag{18.27}$$

and the determinant of the transformation is

$$\det(\Lambda^a{}_{\bar{a}}) = r^2 \cos\phi. \tag{18.28}$$

This transformation matrix can be used, for example, to determine the transformation of the partial derivative operator

$$\partial_{\bar{a}} = \Lambda^a{}_{\bar{a}} \partial_a. \tag{18.29}$$

Organized as matrix-vector multiplication, this result takes the form

$$(\partial_\lambda, \partial_\phi, \partial_r) = (\partial_1, \partial_2, \partial_3) \begin{pmatrix} -r \cos\phi \sin\lambda & -r \sin\phi \cos\lambda & \cos\phi \cos\lambda \\ r \cos\phi \cos\lambda & -r \sin\phi \sin\lambda & \cos\phi \sin\lambda \\ 0 & r \cos\phi & \sin\phi \end{pmatrix}, \tag{18.30}$$

where the partial derivatives do not act on the components of the transformation matrix.

18.4.2 Metric for S^2

Consider the two-sphere S^2 with radius R. Use the angles $(\xi^1, \xi^2) = (\lambda, \phi)$ as the intrinsic coordinates, and the vectors $\vec{e}_1 = R\cos\phi\,\hat{\boldsymbol{\lambda}}$, $\vec{e}_2 = R\,\hat{\boldsymbol{\phi}}$ as the corresponding basis vectors. By embedding S^2 in the Euclidean space $\mathcal{R}^3$, we recover the geometric situation for classical GFD in which the sphere represents an idealization of the Earth embedded in the 3D Euclidean space of Newtonian physics. Embeddings of objects in more complicated space-times are studied in general relativity.

The metric for 3D Euclidean space is

$$g_{ab} = \delta_{ab} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{18.31}$$

where the labels a, b run from $1, 2, 3$. We will also find use to introduce labels α, β which run from $1, 2$. These Greek labels are used for coordinates $(\xi^1, \xi^2) = (\lambda, \phi)$ that intrinsic to the sphere.

Consider two very close points living on S^2, and let the Cartesian coordinates for these points differ by $\mathrm{d}x^a$. Since the sphere is assumed to be embedded in $\mathcal{R}^3$, the squared distance between the two points is determined through the Euclidean norm

$$\mathrm{d}s^2 = \delta_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b. \tag{18.32}$$

Now use the fact that the coordinates x^a for points on the sphere are functions of the sphere's intrinsic coordinates ξ^α. The chain rule then brings the squared distance to the form

$$\mathrm{d}s^2 = \delta_{ab}\left(\frac{\partial x^a}{\partial \xi^\alpha}\frac{\partial x^b}{\partial \xi^\beta}\right)\mathrm{d}\xi^\alpha\,\mathrm{d}\xi^\beta. \tag{18.33}$$

As such, we identify the metric for a sphere S^2 embedded in $\mathcal{R}^3$

$$\begin{aligned} g_{\alpha\beta} &= \delta_{ab}\left(\frac{\partial x^a}{\partial \xi^\alpha}\frac{\partial x^b}{\partial \xi^\beta}\right) \\ &= \mathrm{diag}(R^2\cos^2\phi, R^2) \\ &= \vec{e}_\alpha \cdot \vec{e}_\beta. \end{aligned} \tag{18.34}$$

The first and second columns of the transformation matrix (18.27) were used to deduce this result. The last expression is an elegant means to derive the metric for a surface embedded in Euclidean space.

18.5 Transformations of a vector

In differential geometry, a fundamental question one asks about an object is how it changes under coordinate transformations. The transformation properties provide valuable mathematical and physical information about the object. It turns out that many objects transform under ways familiar from elementary calculus, and they have been assigned the name *tensors*. The transformation properties determine the tensor's *order* or *rank*.

The simplest tensor is a *scalar*. This is a rank zero tensor, or equivalently a tensor of zeroth order. The defining property of a scalar is that it has the same coordinate

representation in any set of coordinates. In other words, scalars remain unchanged or invariant under coordinate transformations. Ocean tracers are examples of scalar fields (Section 4.2.1).

As mentioned in Section 18.3, a first order tensor represented in terms of a basis of vectors is often called a *vector* for brevity. Vectors are objects of physical interest whose coordinate representations are affected by coordinate transformations. Here we determine how a vector transforms under changes in coordinates, and hence determine its tensorial attributes. First, consider the representation of an arbitrary vector in terms of a particular set of basis vectors

$$\mathbf{F} = F^a \, \vec{e}_a. \tag{18.35}$$

Now consider a representation of the same vector in terms of another set of basis vectors

$$\mathbf{F} = F^{\bar{a}} \, \vec{e}_{\bar{a}}. \tag{18.36}$$

Note that the tensor **F** did not change, only its coordinate *representations* F^a and $F^{\bar{a}}$ changed. The tensor **F** is therefore abstracted from any particular coordinate representation and so can be ascribed physical significance. Its representation, however, depends on the arbitrary choice of coordinates which assign numerical values to its components according to these coordinates.

In general, the two sets of basis vectors differ by a transformation, thus allowing the second basis to be written in terms of the first through the relation

$$\mathbf{F} = F^{\bar{a}} \, \Lambda^a{}_{\bar{a}} \vec{e}_a, \tag{18.37}$$

where $\Lambda^a{}_{\bar{a}}$ are components to the transformation matrix relating the two sets of basis vectors. Equating the right hand sides of equations (18.35) and (18.37) implies that the two coordinate representations of **F** are related by

$$F^a = \Lambda^a{}_{\bar{a}} \, F^{\bar{a}}. \tag{18.38}$$

Similar arguments yield

$$F^{\bar{a}} = \Lambda^{\bar{a}}{}_a \, F^a, \tag{18.39}$$

where $\Lambda^{\bar{a}}{}_a$ are components to the inverse transformation Λ^{-1}. In the older tensor analysis literature, these transformation rules are used to define a *first order contravariant tensor*. These ideas and manipulations form the core of practical tensor analysis, and so will be used throughout the following.

18.6 One-forms

As discussed in Section 18.4, the metric tensor determines how to take the scalar product between two vectors. The result is the squared distance between the two vectors. More abstractly, the metric tensor takes two vectors as arguments, and yields the scalar distance

$$\mathbf{g}(\vec{F}, \vec{E}) = [\text{distance}(\vec{F}, \vec{E})]^2 \tag{18.40}$$

where we find it convenient in this section to use the arrow to denote a first order tensor expressed in terms of its vector basis. When supplied with only one vector, the metric produces an object known as a *one-form*

$$\widetilde{F} \equiv \mathbf{g}(\vec{F}, \quad), \tag{18.41}$$

where the tilde label is used to distinguish a one-form from a vector. In turn, a one-form takes a vector as argument and produces a scalar

$$\widetilde{F}(\vec{E}) = \mathbf{g}(\vec{F}, \vec{E}). \tag{18.42}$$

These definitions are symmetric, so that a vector acts on a one-form and also produces a scalar

$$\vec{E}(\widetilde{F}) = \mathbf{g}(\vec{E}, \vec{F}). \tag{18.43}$$

Since the metric tensor is symmetric, the ordering of its vector arguments is irrelevant, and so the result of a one-form acting on a vector is the same as that vector acting on the one-form

$$\widetilde{F}(\vec{E}) = \vec{E}(\widetilde{F}). \tag{18.44}$$

Just as for vectors, there are special one-forms defining a basis for all one-forms. Hence, any one-form can be written as a linear sum of the basis one-forms. For any basis of vectors, there is a corresponding basis of one-forms related through

$$\vec{e}_a(\widetilde{e}^b) = \delta^b_{\ a}. \tag{18.45}$$

For example, with the basis of vectors for the surface of the sphere of radius R,

$$\vec{e}_1 = R\,\cos\phi\,\hat{\boldsymbol{\lambda}} \tag{18.46}$$

$$\vec{e}_2 = R\,\hat{\boldsymbol{\phi}}\,, \tag{18.47}$$

the corresponding basis of one-forms is

$$\widetilde{e}^1 = (R\,\cos\phi)^{-1}\,\hat{\boldsymbol{\lambda}} \tag{18.48}$$

$$\widetilde{e}^2 = R^{-1}\,\hat{\boldsymbol{\phi}}\,. \tag{18.49}$$

Using the basis of one-forms, an arbitrary one-form is written as the linear sum

$$\widetilde{F} = F_a\,\widetilde{e}^a, \tag{18.50}$$

where F_a is the coordinate representation of the one-form $\widetilde{F}$. Using the orthonormality relation between the one-form and vector basis, it is possible to determine the coordinate components to a one-form through letting the one-form "eat" one of the basis vectors

$$\begin{aligned}\widetilde{F}(\vec{e}_b) &= F_a\,\widetilde{e}^a(\vec{e}_b)\\ &= F_a\,\delta^a_{\ b}\\ &= F_b.\end{aligned} \tag{18.51}$$

18.7 Metric as the map between vectors and one-forms

To each one-form, there is a corresponding vector related through the metric tensor. This result is easy to see through examining again the expression for the one-form components

$$\begin{aligned} F_a &= \widetilde{F}(\vec{e}_a) \\ &= \mathbf{g}(\vec{F}, \vec{e}_a) \\ &= \mathbf{g}(F^b\, \vec{e}_b, \vec{e}_a) \\ &= F^b \mathbf{g}(\vec{e}_b, \vec{e}_a) \\ &= F^b g_{ba}. \end{aligned} \tag{18.52}$$

The ability to take the component F^b outside the argument of the metric tensor follows since F^b is simply a number. In short, the component of a one-form F_a is related to the component of a vector F^b through the relation

$$F_a = g_{ab} F^b. \tag{18.53}$$

As a result, one finds that a one-form is a tensor of type $\binom{0}{1}$. Additionally, we see that the metric tensor maps one-forms to vectors. The previous arguments can be used to show that the inverse metric provides for the relation

$$F^a = g^{ab} F_b, \tag{18.54}$$

where g^{ab} are components of the inverse metric $\mathbf{g}^{-1}$. It is in this manner that a one-form and its corresponding vector are said to maintain a *duality* relationship, which is a generalization of the duality maintained between row vectors and column vectors in linear algebra.

These results establish the general manner in which to transform the representation of a tensor. Namely, the metric and its inverse transform the type or rank of a tensor through acting on the tensor's labels. For example, the inner product of two vectors can be written in either of the equivalent forms

$$\begin{aligned} \mathbf{g}(\vec{F}, \vec{E}) &= \widetilde{F}(\vec{E}) \\ &= F_a\, E^a \\ &= g_{ab} F^b\, E^a \\ &= g^{ab} F_a\, E_b. \end{aligned} \tag{18.55}$$

It is through this transformation that we can see why it is often irrelevant whether a tensor is considered in its one-form or vector representation, since they are easily mapped to one another via the metric tensor. The fundamental notion is the order of the tensor, where both one-forms and vectors are first order tensors. Further examples will be encountered in the following.

18.8 Transformation of a one-form

Positioning of the indices on the one-form components are dual to the indices on the vector components. As such, one suspects that the transformation of a one-form

occurs through the components to the inverse transformation used for vectors. Indeed, this result is easy to establish using the transformation properties of the metric, the mapping property between vectors and one-forms, and the transformation properties of a vector:

$$\begin{aligned} F_{\bar{a}} &= g_{\bar{a}\bar{b}}\, F^{\bar{b}} \\ &= g_{\bar{a}\bar{b}}\, \Lambda^{\bar{b}}{}_{b} F^{b} \\ &= \Lambda^{a}{}_{\bar{a}}\, \Lambda^{c}{}_{\bar{b}}\, g_{ac}\, \Lambda^{\bar{b}}{}_{b} F^{b} \\ &= \Lambda^{a}{}_{\bar{a}}\, \delta^{c}{}_{b}\, g_{ac} F^{b} \\ &= \Lambda^{a}{}_{\bar{a}}\, F_{a}. \end{aligned} \tag{18.56}$$

To establish this result, we used the relation

$$\Lambda^{c}{}_{\bar{b}}\, \Lambda^{\bar{b}}{}_{b} = \delta^{c}{}_{b}. \tag{18.57}$$

which follows since $\Lambda^{\bar{b}}{}_{b}$ are components to the matrix Λ^{-1} and $\Lambda^{b}{}_{\bar{b}}$ are components to the matrix Λ.

18.9 Arbitrary tensors and their transformations

The previous formalism provides the framework for determining how any tensor transforms under a coordinate change. For example, consider a *second order* tensor

$$\mathbf{T} = T^{ab}\, \vec{e}_a\, \vec{e}_b. \tag{18.58}$$

The numbers T^{ab} form the components to $\mathbf{T}$ as expanded in terms of the basis vectors $\vec{e}_a$. Considering $\mathbf{T}$ as an abstract object, we can also consider the representations

$$\mathbf{T} = T^{a}_{\ b}\, \vec{e}_a\, \tilde{e}^b, \tag{18.59}$$

or

$$\mathbf{T} = T_{ab}\, \tilde{e}^a\, \tilde{e}^b. \tag{18.60}$$

In classical tensor analysis, T^{ab} is termed the second order contravariant representation of $\mathbf{T}$; $T^{a}_{\ b}$ is the first order contravariant and first order covariant representation; and T_{ab} is the second order covariant representation. An easy way to remember the language is the mnemonic "co-lo", which associates <u>co</u>variant with <u>lo</u>wered indices. Modern tensor analysis avoids the use of these terms in favor of denoting the *type* of tensor as symbolized by $\binom{M}{N}$. T^{ab} is termed the $\binom{2}{0}$ representation; $T^{a}_{\ b}$ is the $\binom{1}{1}$ representation; and T_{ab} is the $\binom{0}{2}$ representation. Both languages are employed in the following, as it is useful to be familiar with these terms when reading the literature.

The transformation properties of the different representations of $\mathbf{T}$ are straightforward to determine. First, consider the transformation to another basis set. Recall that vectors and one-forms transform as

$$\vec{e}_a = \Lambda^{\bar{a}}{}_{a}\, \vec{e}_{\bar{a}} \tag{18.61}$$

$$\tilde{e}^a = \Lambda^{a}{}_{\bar{a}}\, \tilde{e}^{\bar{a}}. \tag{18.62}$$

Consequently,

$$\begin{aligned}\mathbf{T} &= T^{ab}\,\vec{e}_a\,\vec{e}_b \\ &= T^{ab}\,\Lambda^{\bar{a}}{}_a\,\Lambda^{\bar{b}}{}_b\,\vec{e}_{\bar{a}}\,\vec{e}_{\bar{b}} \\ &= T^{\bar{a}\bar{b}}\,\vec{e}_{\bar{a}}\,\vec{e}_{\bar{b}},\end{aligned} \tag{18.63}$$

where

$$T^{\bar{a}\bar{b}} = T^{ab}\,\Lambda^{\bar{a}}{}_a\,\Lambda^{\bar{b}}{}_b. \tag{18.64}$$

Similarly,

$$T^{\bar{a}}{}_{\bar{b}} = T^{a}{}_{b}\,\Lambda^{\bar{a}}{}_a\,\Lambda^{b}{}_{\bar{b}} \tag{18.65}$$

and

$$T_{\bar{a}\bar{b}} = T_{ab}\,\Lambda^{\bar{a}}{}_a\,\Lambda^{\bar{b}}{}_b. \tag{18.66}$$

Recall that in Section 18.4, we determined that the metric transforms under a change in coordinates as

$$g_{\bar{a}\bar{b}} = \Lambda^{a}{}_{\bar{a}}\,\Lambda^{b}{}_{\bar{b}}\,g_{ab}, \tag{18.67}$$

where this relation was established in the special case of $g_{ab} = \delta_{ab}$. In general, it is simple to show that any two representations of the metric are related by equation (18.67). As such, the metric transforms as second a order tensor, hence justifying its appellation as the *metric tensor*.

18.10 Tensorial properties of the gradient operator

The partial derivative operator

$$\partial_a = \frac{\partial}{\partial \xi^a} \tag{18.68}$$

naturally carries its index downstairs. As such, one might suspect that it forms the first order covariant components to the gradient tensor ∇. The chain rule

$$\begin{aligned}\partial_{\bar{a}} &= \frac{\partial}{\partial \xi^{\bar{a}}} \\ &= \frac{\partial \xi^a}{\partial \xi^{\bar{a}}}\,\frac{\partial}{\partial \xi^a} \\ &= \Lambda^{a}{}_{\bar{a}}\,\partial_a\end{aligned} \tag{18.69}$$

establishes that indeed, the partial derivative operator transforms as a first order covariant tensor, thus establishing that the gradient operator

$$\nabla = \tilde{e}^a\,\partial_a \tag{18.70}$$

is a first order tensor naturally represented as a one-form. Raising the label on the derivative gives

$$\partial^a = g^{ab}\,\partial_b, \tag{18.71}$$

which form the components to the contravariant or vector representation of the gradient operator

$$\nabla = \vec{e}_a\,\partial^a. \tag{18.72}$$

18.11 The invariant volume element

The volume element for Euclidean space, as written with Cartesian coordinates, takes the form

$$\mathrm{d}V = \mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3. \tag{18.73}$$

Volume elements arise, for example, when integrating over a finite volume, or for setting the volume of an infinitesimal fluid parcel. It is useful to generalize this result to arbitrary coordinates on an arbitrary manifold.

18.11.1 Derivation of the invariant volume element

To derive an expression for the volume element relevant for any manifold using any coordinates, begin with the familiar rules from elementary calculus relating the differential elements $\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3$ and $\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}}$ in two sets of coordinates

$$\begin{aligned}\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 &= \left(\frac{\partial(\xi^1,\xi^2,\xi^3)}{\partial(\xi^{\bar{1}},\xi^{\bar{2}},\xi^{\bar{3}})}\right)\,\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}}\\ &= \det(\Lambda^a{}_{\bar{a}})\,\mathrm{d}\xi^{\bar{1}}\,\mathrm{d}\xi^{\bar{2}}\,\mathrm{d}\xi^{\bar{3}}\end{aligned} \tag{18.74}$$

where $\det(\Lambda^a{}_{\bar{a}})$ is the Jacobian, or determinant of the transformation matrix. The transformation is well-defined so long as the Jacobian does not vanish. Note that the labels on the transformation matrix are maintained inside the determinant symbol to indicate between which sets of coordinates this tranformation matrix is referred. As such, it is possible to explicitly note the matching of indices on both sides if this equation. This notation is especially useful when more than one transformation is considered.

For certain cases, the Jacobian is tedious to compute. In these cases, the computation can be simplified by using the following relation between the Jacobian and the determinant of the metric. For this purpose, recall that the transformation of the metric (Section 18.9) is given by

$$g_{\bar{a}\bar{b}} = \Lambda^a{}_{\bar{a}}\Lambda^b{}_{\bar{b}}\,g_{ab}. \tag{18.75}$$

Written as a matrix equation, this result takes the form

$$(\bar{\mathbf{g}}) = (\Lambda^T)\,(\mathbf{g})\,(\Lambda) \tag{18.76}$$

where Λ^T is the matrix transposed. This equation is valid upon taking determinants of both sides. Now recall two properties of the determinant:

1. $\det(A\,B) = \det(A)\,\det(B)$, for any two matrices A and B. This result is proven in various linear algebra texts, such as page 203 of Noble and Daniel (1977).

2. $\det(\Lambda^T) = \det(\Lambda)$. This result is trivial to prove.

Consequently, re-exposing the labels on Λ and the metric,

$$\det(\Lambda^a{}_{\bar{a}}) = \left(\frac{\det(g_{\bar{a}\bar{b}})}{\det(g_{ab})}\right)^{1/2}, \tag{18.77}$$

where it has been assumed that the manifold has a positive definite metric, which is true for surfaces, such as the sphere, considered in this book. The determininat of the metric appears quite often in the following, and so it is useful to introduce the notation

$$\mathcal{G} = \det(g_{ab}) \tag{18.78}$$

$$\overline{\mathcal{G}} = \det(g_{\overline{a}\overline{b}}). \tag{18.79}$$

Note that most books use the symbol g rather than $\mathcal{G}$, which for our purposes is not useful since g is the acceleration of gravity at the earth's surface. In summary, we are left with the very simple relation

$$\sqrt{\mathcal{G}}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 = \sqrt{\overline{\mathcal{G}}}\,\mathrm{d}\xi^{\overline{1}}\,\mathrm{d}\xi^{\overline{2}}\,\mathrm{d}\xi^{\overline{3}}. \tag{18.80}$$

This expression leads us to conclude that the expression

$$\mathrm{d}V \equiv \sqrt{\mathcal{G}}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 \tag{18.81}$$

is invariant under coordinate transformations; i.e., it is a scalar. Hence, it forms the appropriate generalization to arbitrary coordinates of the differential volume element.

18.11.2 Invariant volume element in different coordinates

It is useful to consider some examples of how the volume element appears when writing it in various coordinates. First, as noted above, the volume element for three-dimensional Euclidean space using Cartesian coordinates (ξ^1, ξ^2, ξ^3) is $\mathrm{d}V = \mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3$. Spherical coordinates, $(\xi^1, \xi^2, \xi^3) = (\lambda, \phi, r)$ (Figure 18.1), with metric $g_{ab} = \mathrm{diag}(r^2\cos^2\phi, r^2, 1)$, have a volume element

$$\mathrm{d}V = \mathcal{G}^{1/2}\,\mathrm{d}r\,\mathrm{d}\lambda\,\mathrm{d}\phi = r^2\cos\phi\,\mathrm{d}r\,\mathrm{d}\lambda\,\mathrm{d}\phi. \tag{18.82}$$

The equality

$$\mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3 = r^2\cos\phi\,\mathrm{d}r\,\mathrm{d}\lambda \tag{18.83}$$

is familiar from elementary Calculus.

For an arbitrary nonsingular and orthogonal set of coordinates, defined such that the line element takes the diagonal form

$$\mathrm{d}s^2 = (h_1\,\mathrm{d}\xi^1)^2 + (h_2\,\mathrm{d}\xi^2)^2 + (h_3\,\mathrm{d}\xi^3)^2, \tag{18.84}$$

the invariant volume element takes the form

$$\mathrm{d}V = \mathcal{G}^{1/2}\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3 = h_1\,h_2\,h_3\,\mathrm{d}\xi^1\,\mathrm{d}\xi^2\,\mathrm{d}\xi^3. \tag{18.85}$$

These coordinates are often called *generalized orthogonal curvilinear coordinates*. They correspond to the orthogonal set of basis vectors

$$\vec{e}_a = h_a\,\hat{\mathbf{e}}_{(a)}, \tag{18.86}$$

where no summation is implied and $\hat{\mathbf{e}}_{(a)}$ are the dimensionless unit directions.* Hence, the functions h_a, which generally depend on space-time, determine the local "stretching" function in the three orthogonal unit directions $\hat{\mathbf{e}}_{(a)}$ for use in measuring distance. A common choice in ocean modeling is to set $h_3 = 1$ and then choose the horizontal stretching factors h_1 and h_2 so that the North Pole lives over land rather than ocean (Murray (1996); Madec and Imbard (1996); Smith et al. (1995)). In this way, the ocean domain does not contain the annoying coordinate singularity possessed by spherical coordinates at the North Pole. Much more is said about orthogonal coordinates in Section 19.11.

Finally, consider the case of isopycnal coordinates (x, y, ρ). In Section 5.5.2, we derive the transformation matrix between Cartesian and isopycnal coordinates. The Jacobian of that transformation is the specific thickness $z_{,\rho}$, which gives the expression for the volume element

$$\mathrm{d}V = z_{,\rho}\,\mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}\rho. \tag{18.87}$$

This expression can also be found through the formula $\mathrm{d}V = \mathcal{G}^{1/2}\,\mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}\rho$, where $\mathcal{G}$ is the determinant of the metric tensor for three dimensional Euclidean space as written using isopycnal coordinates. Using the transformation matrix $\Lambda^a{}_{\bar{a}}$ from Section 5.5.2, and the tensor transformation rules from Section 18.9, it is straightforward to derive the expression for the metric (see Section 5.6)

$$\begin{aligned} g_{\bar{a}\bar{b}} &= \Lambda^a{}_{\bar{a}}\,\Lambda^b{}_{\bar{b}}\delta_{ab} \\ &= \begin{pmatrix} 1+z_{,x}^2 & z_{,x}\,z_{,y} & z_{,x}\,z_{,\rho} \\ z_{,x}\,z_{,y} & 1+z_{,y}^2 & z_{,y}\,z_{,\rho} \\ z_{,x}\,z_{,\rho} & z_{,y}\,z_{,\rho} & z_{,\rho}^2 \end{pmatrix}, \end{aligned} \tag{18.88}$$

where δ_{ab} is the metric for Euclidean space written in terms of Cartesian coordinates. A simple calculation shows that indeed $\mathcal{G}^{1/2} = z_{,\rho}$.

18.12 Determinants and the Levi-Civita symbol

Consider a transformation matrix $\Lambda^a{}_{\bar{a}}$ between two sets of coordinates, and let the manifold be two dimensional. As already mentioned, a well defined transformation has a non-zero determinant

$$\det(\Lambda^a{}_{\bar{a}}) = \Lambda^1{}_{\bar{1}}\,\Lambda^2{}_{\bar{2}} - \Lambda^1{}_{\bar{2}}\,\Lambda^2{}_{\bar{1}}. \tag{18.89}$$

Determinants, especially those associated with transformations, are ubiquitous in tensor analysis. It turns out that for many manipulations it is useful to introduce the permutation or *Levi-Civita symbol*, which brings the determinant to the form

$$\det(\Lambda^a{}_{\bar{a}}) = \epsilon_{ab}\,\Lambda^a{}_{\bar{1}}\,\Lambda^b{}_{\bar{2}}, \tag{18.90}$$

where,

$$\epsilon_{ab} = \begin{cases} 0, & \text{if any two labels are the same} \\ 1, & \text{if } a, b \text{ is an even permutation of } 1, 2 \\ -1, & \text{if } a, b \text{ is an odd permutation of } 1, 2. \end{cases} \tag{18.91}$$

*As explained in Section 19.11, the index on the unit directions is enclosed in parentheses to advertise the fact that this label is not tensorial; i.e., the unit directions do not transform as tensors. Rather, the functions h_a carry the tensorial properties of the basis vectors $\vec{e}_a$.

This symbol is defined to have numerically the same values whether the labels are raised or lowered: $\epsilon^{ab} = \epsilon_{ab}$. As seen, the Levi-Civita symbol is totally anti-symmetric on its two labels. It has generalizations to any number of dimensions, for which one more label is added to the symbol, and one more number added to the permutation string. For example, in three dimensions,

$$\epsilon_{abc} = \begin{cases} 0, & \text{if any two labels are the same} \\ 1, & \text{if } a, b, c \text{ is an even permutation of } 1, 2, 3 \\ -1, & \text{if } a, b, c \text{ is an odd permutation of } 1, 2, 3. \end{cases} \tag{18.92}$$

Through its connection to the determinant, the Levi-Civita symbol is also useful for writing the curl of two vectors living in $\mathcal{R}^3$. For example, in Cartesian coordinates,

$$(\nabla \wedge \vec{A})^a = \epsilon^{abc} A_{c,b}. \tag{18.93}$$

Let us now see how the Levi-Civita symbol transforms. For this purpose, go back to two dimensions where things are simple to explicity verify. The central formula is

$$\epsilon_{ab} \Lambda^a{}_{\bar{a}} \Lambda^b{}_{\bar{b}} = \epsilon_{\bar{a}\bar{b}} \det(\Lambda^a{}_{\bar{a}}), \tag{18.94}$$

which follows directly from the definition of the determinant. This expression is valid so long as the transformed Levi-Civita symbol $\epsilon_{\bar{a}\bar{b}}$ is numerically identical to ϵ_{ab}. Additionally, we see that the Levi-Civita symbol *does not* transform as the components to a second order covariant tensor unless the determinant of the transformation is unity, such as for a simple rotation. Rather, in general, the covariant Levi-Civita symbol is said to transform as a second order tensor *density* of weight -1, where the weight is determined by the power of the determinant. In order to define a tensor of weight 0, consider the scaled Levi-Civita symbol

$$\varepsilon_{ab} = \mathcal{G}^{1/2} \epsilon_{ab}. \tag{18.95}$$

This scaling with the metric determinant is identical to that which was used to produce a coordinate invariant volume element in Section 18.11. These factors are indeed related, and in more complete discussions (e.g., Schutz (1980)), the connection is made explicit. For our purpuses, it is sufficient to note that ε_{ab} transforms as

$$\begin{aligned} \Lambda^a{}_{\bar{a}} \Lambda^b{}_{\bar{b}} \varepsilon_{ab} &= \Lambda^a{}_{\bar{a}} \Lambda^b{}_{\bar{b}} \mathcal{G}^{1/2} \epsilon_{ab} \\ &= \mathcal{G}^{1/2} \epsilon_{\bar{a}\bar{b}} \det(\Lambda^a{}_{\bar{a}}) \\ &= \overline{\mathcal{G}}^{1/2} \epsilon_{\bar{a}\bar{b}} \\ &= \varepsilon_{\bar{a}\bar{b}}, \end{aligned} \tag{18.96}$$

where equations (18.94) and (18.77) were employed. Therefore, ε_{ab} transforms as the components to a regular second order covariant tensor, without the extra determinant factors floating around. Likewise,

$$\varepsilon^{ab} = \mathcal{G}^{-1/2} \epsilon^{ab} \tag{18.97}$$

transforms as the components to a second order contravariant tensor. In subsequent development, ε_{ab} is referred to as the *covariant* Levi-Civita tensor.

Each of these expressions derived in two-dimensions are valid in any number of dimensions, as can be shown through induction. As hinted at above, one useful application of the third order tensor ε_{abc} is in defining the covariant curl operation (Section 19.8), which generalizes the usual curl from Cartesian vector analysis to curved manifolds.

18.13 Surfaces embedded in Euclidean space

In Chapter 5, we discuss some of the basic properties of surfaces through study of the metric for an isopycnal surface (see in particular Section 5.6). In this section, we systematize and extend that discussion. Note that this section is a bit outside the main line of development in this chapter and so can be easily skipped.

18.13.1 Induced metric tensor for a surface

Consider the line element in three-dimensional Euclidean space

$$\mathrm{d}s^2 = g_{ab}\,\mathrm{d}\xi^a\,\mathrm{d}\xi^b. \tag{18.98}$$

The components of the metric tensor g_{ab} are those for Euclidean space as written in terms of an arbitrary coordinate system $\xi^a = (\xi^1, \xi^2, \xi^3)$. If the line element lives on the surface of interest, then the chain rules renders

$$\mathrm{d}\xi^a = \left(\frac{\partial \xi^a}{\partial w^\alpha}\right)\,\mathrm{d}w^\alpha, \tag{18.99}$$

where $w^\alpha = (w^1, w^2)$ are coordinates intrinsic to the two-dimensional surface, and $\xi^a = \xi^a(w^\alpha)$ for points on the surface. In this section, Greek labels refer to coordinates intrinsic to the surface, and Latin labels refer to the Euclidean space. Equation (18.99) provides a transformation between infinitesimal displacements $\mathrm{d}w^\alpha$ on the surface, and the corresponding displacements $\mathrm{d}\xi^a$ in the space tangent to the surface. The numbers

$$\Lambda^a{}_\alpha = \left(\frac{\partial \xi^a}{\partial w^\alpha}\right) \tag{18.100}$$

form components to the transformation matrix which transforms between the surface and the embedding Euclidean space. It follows that the line element for points on the surface takes the form

$$\begin{aligned}\mathrm{d}s^2 &= g_{ab}\,\Lambda^a{}_\alpha\,\Lambda^b{}_\beta\,\mathrm{d}w^\alpha\,\mathrm{d}w^\beta \\ &\equiv G_{\alpha\beta}\,\mathrm{d}w^\alpha\,\mathrm{d}w^\beta,\end{aligned} \tag{18.101}$$

where $G_{\alpha\beta}$ are components to the metric tensor for the two-dimensional surface. This metric is said to be *induced* by the embedding of the surface in the three-dimensional Euclidean space. As an aside, we note that if the same surface is embedded in some other space, such as a non-Euclidean space, then the induced metric would generally differ. Equation (18.101) defining the metric is directly analogous to equation (18.67) expressing how components of the metric for Euclidean space transform under a coordinate change.

We use the above results implicitly when deriving the metric for an isopycnal surface in Section 5.6.1 (see also Section 18.11.2). For that case, the intrinsic coordinates are the usual (x, y) Cartesian coordinates of the point on the surface $\rho(x, y, z) = \rho_{const}$. The resulting surface metric $G_{\alpha\beta}$ is given by the non-diagonal 2×2 matrix in equation (5.55).

18.13.2 The invariant surface area element

The derivation of an invariant volume element given in Section 18.11 is general to any space dimension. For a two dimensional space, such as a surface embedded in Euclidean space, the invariant "volume" or area element takes the form

$$\mathrm{d}A = \sqrt{G}\,\mathrm{d}w^1\,\mathrm{d}w^2, \tag{18.102}$$

where $G = \det(G_{\alpha\beta})$ is the determinant of the surface metric tensor. This is the appropriate form of the surface area element for use in forming surface integrals. For example, if one wishes to integrate a field, such as temperature, over a particular isopycnal surface $\rho = \rho_1$ using the isopycnal coordinates (x, y, ρ), then the metric given by equation (5.55) indicates that the surface integral should be of the form

$$\int_{\rho=\rho_1} T(x, y, \rho)\,\mathrm{d}A = \int_{\rho=\rho_1} T(x, y, \rho)\,(1 + S_x^2 + S_y^2)^{1/2}\,\mathrm{d}x\,\mathrm{d}y. \tag{18.103}$$

For many geophysically relevant situations, the isopycnal slopes are no larger than 1/100. In this case, the squared slope terms can be safely dropped. This "small slope" approximation is fundamental to many numerical isopycnal ocean models.

18.13.3 Surface and space components of a tensor

Recall from Section 18.5 that a vector field $\vec{F}$ living in Euclidean space can be represented in terms of a basis of vectors that span Euclidean space

$$\vec{F} = F^a\,\vec{e}_a. \tag{18.104}$$

The components F^a are sometimes called the *space* components, as they are the components of $\vec{F}$ relative to the Euclidean space basis $\vec{e}_a$. For points on some surface, the chain-rule transformation (18.99) allows for $\vec{F}$ to also be written in the form

$$\begin{aligned}\vec{F} &= F^a\,\vec{e}_a \\ &= \Lambda^a{}_\alpha\,F^\alpha\,\vec{e}_a \\ &\equiv F^\alpha\,\vec{e}_\alpha,\end{aligned} \tag{18.105}$$

where

$$F^\alpha = \Lambda^\alpha{}_a\,F^a \tag{18.106}$$

defines the *surface* components of $\vec{F}$, and

$$\vec{e}_\alpha = \Lambda^a{}_\alpha\,\vec{e}_a \tag{18.107}$$

defines the surface basis vectors. As seen in the next subsection, the surface components F^α are the projection of the space components F^a onto the surface; i.e., they are tangent to the surface.

The covariant surface components F_α of the tensor are related to the contravariant components through the surface metric tensor

$$\begin{aligned} F_\alpha &= \Lambda^a_{\ \alpha}\, F_a \\ &= \Lambda^a_{\ \alpha}\, g_{ab}\, F^b \\ &= \Lambda^a_{\ \alpha}\, \Lambda^b_{\ \beta}\, g_{ab}\, F^\beta \\ &= G_{\alpha\beta}\, F^\beta. \end{aligned} \tag{18.108}$$

Consequently, the surface metric $G_{\alpha\beta}$ maps between surface vectors and surface one-forms, just as the space metric g_{ab} maps between space vectors and space one-forms.

18.13.4 Normals and tangents

Recall from Section 18.12 that the generalized form of a cross product between two vectors is conveniently written as

$$A_a = \varepsilon_{abc}\, B^b\, C^c \tag{18.109}$$

where $\varepsilon_{abc} = \mathcal{G}^{1/2}\, \epsilon_{abc}$ is the covariant form of the Levi-Civita symbol. A_a form the covariant components to a one-form $\widetilde{A} = A_a\, \widetilde{e}^a$. It is easy to show that $\widetilde{A}$ is orthogonal to either of the vectors $\vec{B} = B^b\, \vec{e}_b$ or $\vec{C} = C^c\, \vec{e}_c$. For example,

$$\begin{aligned} \widetilde{A}(\vec{B}) &= A_a\, B^a \\ &= \varepsilon_{abc}\, B^b\, C^c\, B^a \\ &= 0, \end{aligned} \tag{18.110}$$

which follows from the anti-symmetry of ε_{abc}.

If the two vectors $\vec{B}$ and $\vec{C}$ are not colinear, they define a two-dimensional surface that is tangent to these two vectors at each point in space. In this case, we can write the cross-product as

$$\begin{aligned} A_a &= \varepsilon_{abc}\, B^b\, C^c \\ &= \varepsilon_{abc}\, \Lambda^b_{\ \beta}\, \Lambda^c_{\ \gamma}\, B^\beta\, C^\gamma. \end{aligned} \tag{18.111}$$

The factor $\varepsilon_{abc}\, \Lambda^b_{\ \beta}\, \Lambda^c_{\ \gamma}$ is anti-symmetric under interchange of the surface tensor labels; viz.,

$$\begin{aligned} \varepsilon_{abc}\, \Lambda^b_{\ \beta}\, \Lambda^c_{\ \gamma} &= \varepsilon_{acb}\, \Lambda^c_{\ \beta}\, \Lambda^b_{\ \gamma} \\ &= -\varepsilon_{abc}\, \Lambda^b_{\ \gamma}\, \Lambda^c_{\ \beta}. \end{aligned} \tag{18.112}$$

Hence, we can introduce the covariant tensor N_a through

$$N_a\, \varepsilon_{\beta\gamma} = \varepsilon_{abc}\, \Lambda^b_{\ \beta}\, \Lambda^c_{\ \gamma}, \tag{18.113}$$

bringing the cross-product to the form

$$A_a = N_a\, \varepsilon_{\beta\gamma}\, B^\beta\, C^\gamma. \tag{18.114}$$

In this expression, $\varepsilon_{\beta\gamma} = \sqrt{G}\,\epsilon_{\beta\gamma}$ is the covariant Levi-Civita tensor on the surface. With $\varepsilon_{\beta\gamma}\,\varepsilon^{\beta\gamma} = 2$, it is possible to write

$$N_a = \frac{1}{2}\,\varepsilon_{abc}\,\varepsilon^{\beta\gamma}\,\Lambda^b{}_\beta\,\Lambda^c{}_\gamma. \tag{18.115}$$

By construction, the one-form $\widetilde{N} = N_a\,\widetilde{e}^a$ is orthogonal to all spatial vectors tangent to the surface. To see this property, note that

$$\begin{aligned}\widetilde{N}(\vec{B}) &= N_a\,B^a \\ &= N_a\,\Lambda^a{}_\mu\,B^\mu \\ &= \frac{1}{2}\,\varepsilon_{abc}\,\varepsilon^{\beta\gamma}\,\Lambda^b{}_\beta\,\Lambda^c{}_\gamma\,\Lambda^a{}_\mu\,B^\mu \\ &= 0,\end{aligned} \tag{18.116}$$

which follows because μ must equal either β or γ. Therefore, the surface components of an arbitrary one-form $F_\alpha = \Lambda^a{}_\alpha\,F_a$ can be thought of as the projection of the space components F_a onto the surface. For, if we write $F_a = \overline{F}_a + c\,N_a$, then

$$F_\alpha = \Lambda^a{}_\alpha\,(\overline{F}_a + c\,N_a) = \Lambda^a{}_\alpha\,\overline{F}_a, \tag{18.117}$$

which shows that the component normal to the surface has been destroyed. In this way, $F_\alpha = \Lambda^a{}_\alpha\,F_a$ represents the components of the one-form $\widetilde{F}$ which are tangent to the surface. With $c = N^b\,F_b$, the projection of an arbitrary covariant space tensor onto the surface is given by

$$\overline{F}_a = F_a - (N^b\,F_b)\,N_a. \tag{18.118}$$

In addition to describing a surface in terms of two non-colinear vectors, one often describes a surface in terms of the envelope of constant values of a scalar function of space. For example, the algebraic equation

$$\rho(\xi^a) = \rho_{const} \tag{18.119}$$

defines isopycnal surfaces. Following the ideas from Section 18.3, which are based on traditional vector analysis,

$$\begin{aligned}N_a &= \frac{\rho_{,a}}{|\rho_{,b}\,\rho^{,b}|^{1/2}} \\ &= \frac{\rho_{,a}}{|g^{bc}\,\rho_{,b}\,\rho_{,c}|^{1/2}}\end{aligned} \tag{18.120}$$

provides the components to a outward normal one-form for the surface. In this expression, g^{bc} are components to the inverse metric for the embedding space. The previous formalism follows with this definition of the normal one-form. In particular, the projection of an arbitrary one-form onto the surface is given by

$$\overline{F}_a = F_a - \frac{F^b\,\rho_{,b}\,\rho_{,a}}{g^{bc}\,\rho_{,b}\rho_{,c}}. \tag{18.121}$$

These are the components of $\widetilde{F}$ which are tangent to the surface.

CHAPTER
NINETEEN

Calculus on curved manifolds

Contents

19.1 Fundamental character of tensor equations **418**
19.2 Covariant differentiation **419**
19.2.1 The gradient one-form 420
19.2.2 Covariant derivative of a vector 420
19.2.3 Covariant derivative of a one-form 421
19.2.4 Covariant derivative of the metric 422
19.3 Covariant derivative of a second order tensor **422**
19.4 Christoffel symbols in terms of the metric **423**
19.5 Covariant divergence of a vector **423**
19.6 Covariant divergence of a second order tensor **424**
19.7 Covariant Laplacian of a scalar **425**
19.8 Covariant curl of a vector **425**
19.9 Covariant Laplacian of a vector **425**
19.10 Integral theorems **426**
19.11 Orthogonal curvilinear coordinates **426**
19.11.1 Physical components of tensors 428
19.11.2 Eliminating the stretching functions 430
19.11.3 Gradient one-form 430
19.11.4 Covariant divergence of a vector 431
19.11.5 Covariant divergence of an anti-symmetric tensor 431
19.11.6 Covariant Laplacian of a scalar 432
19.11.7 Covariant curl of a vector 433
19.12 Summary of some salient results **433**
19.12.1 Rules for tensor analysis on manifolds 433
19.12.2 Orthogonal coordinates 436
19.12.3 Physical tensor components with orthogonal coordinates . 437
19.12.4 Eliminating the stretching functions 438

The purpose of this chapter is to derive many of the formulas of vector calculus, yet allowing the underlying manifold to be curved. These formulae have application to deriving the equations for fluid dynamics on a sphere.

19.1 Fundamental character of tensor equations

Differential physical laws are most useful when written so that they remain form invariant under arbitrary coordinate transformations. In this way, the essence of the laws can be revealed, rather than being confused with potentially coordinate dependent artifacts. As such, the laws realize the postulate that physics does not care about coordinates. This fundamental postulate, which often goes by the name *covariance*, is at the root of modern physics.

Although physics does not care about coordinates, for many practical situations it is often convenient, if not necessary, to work in a specific coordinate system that is suited to the geometry. For example, when working in Euclidean space it is simplest for many purposes to use the familiar Cartesian coordinate system. And for numerical modeling, it is necessary to choose a particular set of coordinates in which to discretize the continuous equations.

After deriving a physical law in one set of coordinates, it is of interest to establish the form of the law in another set of coordinates. How does the differential equation transform? Does one have to re-derive the law from first principles working with the new set of coordinates? Fortunately, so long as the equations are written in a proper tensorial form, in which they exemplify covariance, then the equations are form invariant and so the derivation need not be repeated in each set of coordinates. In practice, an equation exemplifies covariance if all the tensor indices are properly matched on both sides of the equation and each derivative is covariant. The elegance allowed by this property is the key reason that tensor analysis is a ubiquitous tool in theoretical physics.

To motivate these ideas, note that the partial derivative of a field appears in most differential physical laws. For example, the divergence of velocity appears in the equation for mass conservation in a fluid. In Cartesian coordinates, this divergence takes the form

$$\begin{aligned} \nabla \cdot (v^a \, \vec{e}_a) &= \tilde{e}^a_{(m)} \, \partial_a \, (\vec{e}^{(m)}_b \, v^b) \\ &= v^a_{,a}. \end{aligned} \tag{19.1}$$

The intermediate expression is a rather cumbersome way of writing what is more often seen in the more concise final form. Yet the final form only arises since the basis vectors $\vec{e}_a$ are each constants in space-time. In more general coordinates, such as spherical coordinates, the basis vectors are not constants, and so the partial derivative operator picks up terms proportional to the derivative of the basis vectors. The result is a more complicated divergence. Additionally, it possesses clumsy terms depending on the nature of the chosen coordinates. Hence, the divergence as defined using partial derivatives does not remain form invariant under coordinate transformations. The notion of a covariant divergence described in Section 19.2 accounts for these extra terms so that the equations remain form invariant.

We now summarize some of the properties characterizing covariance (taken after page 153 of Schutz (1985)):

1. Certain manipulations of tensor components are called *permissible tensor operations* because they produce components of new tensors. The following are permissible operations:

   (a) Multiplication of a tensor by a scalar field produces a new tensor of the same type.

   (b) Addition of components of two tensors of the same type gives components of a new tensor of the same type. In particular, only tensors of the same type can be equal.

   (c) Multiplication of components of two tensors of arbitrary type gives components of a new tensor whose type is given by the sum of the types for the individual tensors. This operation is called the *outer product*.

   (d) Covariant differentiation (discussed in Section 19.2) of the components of a tensor of type $\binom{M}{N}$ gives the components to a tensor of type $\binom{M}{N+1}$.

   (e) Contraction on a pair of indices of the components of a tensor of type $\binom{M}{N}$ gives the components of a tensor of type $\binom{M-1}{N-1}$. For example, contraction of the components T^a_b of the second order tensor $\mathbf{T}$ produces the scalar $V = T^a_a$. Note that contraction is defined only between an upper and lower index.

2. If two tensors of the same type have equal components in a given basis, then they have equal components in all bases. Hence, they are identical. For example, if the components of the third rank tensors $\mathbf{A}$ and $\mathbf{B}$ satisfy $A^{ab}_c = B^{ab}_c$ in one set of coordinates, for all possible combinations of the indices a, b, c, then this equality holds in all sets of coordinates and so the tensors are identical. In particular, if all components of a tensor vanish in one set of coordinates, they vanish in all coordinates. This property is easy to prove through use of the transformation properties of tensors established in Section 18.9.

3. It follows from the previous properties that if an equation consists of tensors combined only by the permissible tensor operations, and if the equation is true in one basis, then it is true in any basis. If the equations involve covariant derivatives, then the equations remain form invariant under changes in coordinates.

19.2 Covariant differentiation

Differentiation of tensors on curved manifolds is a logical extension of the usual differentiation on flat manifolds. The name given to such operations is *covariant differentiation*. As mentioned earlier, covariant differentiation of a tensor of type $\binom{M}{N}$ gives the components to a tensor of type $\binom{M}{N+1}$. This result is exemplified below. In determining the form of the covariant derivatives of tensorial quantities, it is necessary to be aware that the basis vectors, basis one-forms, and coordinate representation of the tensor can each have nonzero partial derivatives.

19.2.1 The gradient one-form

Recall that a scalar function V can be thought of as a tensor of type $\binom{0}{0}$. As such, it maintains the same numerical value in any coordinate representation since it needs no basis vectors in order to represent it. When taking its partial derivative

$$\partial_a V = V_{,a} \tag{19.2}$$

the result is a type $\binom{0}{1}$ tensor, as follows since

$$\partial_{\bar{a}} V = \Lambda^a{}_{\bar{a}} V_{,a} \tag{19.3}$$

which defines the transformation rule of a type $\binom{0}{1}$ tensor. This result could be anticipated from the results in Section 18.10 where the transformation rules of the partial derivative were established. Written as a one-form, the partial derivatives of a scalar are components to the *gradient one-form*

$$\nabla V \equiv (\tilde{e}^a \, \partial_a) \, V, \tag{19.4}$$

where $\tilde{e}^a$ define a basis of one-forms.

19.2.2 Covariant derivative of a vector

Consider a vector field living on the manifold $\vec{F}(\xi^a)$. The partial derivative of this vector is given by

$$\partial_b \vec{F} = \partial_b (F^a \, \vec{e}_a). \tag{19.5}$$

To determine the coordinate components of the right hand side, account must be taken of the generally nontrivial dependence of the basis vectors on the coordinates ξ^a. As a result,

$$\begin{aligned} \partial_b \vec{F} &= (\partial_b \, F^a) \; \vec{e}_a + F^a \, \partial_b \, \vec{e}_a \\ &= \left(F^a_{\;,b} \, \vec{e}_a + F^a \, \Gamma^c_{ba} \, \vec{e}_c \right) \\ &= \left(F^a_{\;,b} + F^c \, \Gamma^a_{bc} \right) \; \vec{e}_a \end{aligned} \tag{19.6}$$

where

$$\partial_b \, \vec{e}_a = \Gamma^c_{ba} \, \vec{e}_c \tag{19.7}$$

defines the Christoffel symbols Γ^c_{ba}. The Christoffel symbols are the coordinate representations of the partial derivatives of the basis vectors. If the manifold is the flat Euclidean space, and the basis vectors are the Cartesian unit vectors, then all of the Christoffel symbols vanish. However, for Euclidean space described by non-Cartesian basis vectors, the Christoffel symbols do not vanish. This result indicates that the Christoffel symbols are *not* components to a tensor, since a tensor vanishing in one set of coordinates vanishes for all sets of coordinates (see Section 19.1). Contracting with the basis one-form $\tilde{e}^b$ yields the $\binom{1}{1}$ tensor

$$\tilde{e}^b \, \partial_b \vec{F} = \tilde{e}^b \; \left(F^a_{\;,b} + F^c \, \Gamma^a_{bc} \right) \; \vec{e}_a. \tag{19.8}$$

It is clear that this is a $\binom{1}{1}$ tensor because of the explicit presence of the basis one-form and basis vector. This tensor is known as the covariant derivative of the vector $\vec{F}$. It is typically written in one of the following manners

$$\begin{aligned} \tilde{e}^b\,\partial_b\vec{F} &= (\nabla_b\vec{F})^a\,\vec{e}_a && (19.9)\\ &= F^a_{\ ;b}\,\tilde{e}^b\,\vec{e}_a \\ &= (F^a_{\ ,b} + \Gamma^a_{bc}\,F^c)\,\tilde{e}^b\,\vec{e}_a. && (19.10) \end{aligned}$$

The semi-colon notation is used in the following. It naturally generalizes the comma notation used for the partial derivative. When all of the Christoffel symbols vanish, the components of the covariant derivative reduce to those of the partial derivative. Additionally, the results from Section 19.2.1 show that the covariant derivative of a scalar is identical to the partial derivative of that scalar

$$V_{;a} = V_{,a}. \tag{19.11}$$

This result is obvious, since the scalar field requires no basis vectors for its representation, and so there are no resulting Christoffel symbols.

19.2.3 Covariant derivative of a one-form

The results for the covariant derivative of a scalar and vector provide tools sufficient for determining the covariant derivatives of any arbitrary tensor. For example, the covariant derivative of a one-form can be found by considering the result of taking the covariant derivative on the inner product of the one-form and a vector

$$(E_a\,F^a)_{;b} = (E_a\,F^a)_{,b} \tag{19.12}$$

where equation (19.11) was used since $E_a\,F^a$ is a scalar. Expanding the partial derivative yields

$$\begin{aligned} (E_a\,F^a)_{;b} &= F^a\,\partial_b\,E_a + E_a\,\partial_b\,F^a \\ &= F^a\,\partial_b\,E_a + E_a(F^a_{\ ;b} - \Gamma^a_{bc}F^c) \\ &= F^a(\partial_b\,E_a - \Gamma^c_{ba}E_c) + E_a\,F^a_{\ ;b} \\ &\equiv F^a\,E_{a;b} + E_a\,F^a_{\ ;b}, && (19.13) \end{aligned}$$

where

$$E_{a;b} = \partial_b\,E_a - \Gamma^c_{ba}E_c \tag{19.14}$$

defines the components to the covariant derivative of the one form. Hence,

$$\partial_b\tilde{E} = E_{a;b}\,\tilde{e}^a. \tag{19.15}$$

These results establish that the covariant derivative of a one-form results in a tensor of type $\binom{0}{2}$.

19.2.4 Covariant derivative of the metric

As discussed in Section 19.1, one of the central reasons for employing tensor analysis in physics is that when the tensorial properties of an expression are established in one coordinate system, the same expression is also valid in an arbitrary coordinate system. Notably, when written in Cartesian coordinates, the covariant derivative of the metric for Euclidean space vanishes

$$g_{ab;c} = 0, \tag{19.16}$$

simply because the metric is the unit tensor δ_{ab}, and all the Christoffel symbols vanish. The previous results established the tensorial nature of the covariant derivative, consequently $g_{ab;c} = 0$ is a valid result for *all* coordinates. This result is often called the *metricity* condition. It represents a self-consistency condition required for the manifolds considered in this book.

19.3 Covariant derivative of a second order tensor

We are concerned in this book with manifolds that have zero *torsion*. For our purposes, a zero torsion manifold renders the Christoffel symbol symmetric on its lower two labels.

Given this property, we now compute the covariant derivative of an arbitrary tensor on torsionless manifolds. To illustrate the general ideas, consider the covariant derivative of a second order tensor. To start, note that the covariant derivative of the following scalar quantity is given by

$$\begin{aligned}
(T_{ab}\, F^a\, F^b)_{;c} &= (T_{ab}\, F^a\, F^b)_{,c} \\
&= T_{ab,c}\, F^a\, F^b + T_{ab}\, (F^a\, F^b)_{,c} \\
&= T_{ab,c}\, F^a\, F^b + T_{ab}\, (F^a_{,c}\, F^b + F^a\, F^b_{,c}) \\
&= T_{ab,c}\, F^a\, F^b + T_{ab}\, [(F^a_{;c} - \Gamma^a_{cd}\, F^d)\, F^b + F^a\, (F^b_{;c} - \Gamma^b_{cd}\, F^d)] \\
&= F^a\, F^b\, (T_{ab,c} - T_{db}\, \Gamma^d_{ca} - T_{ad}\, \Gamma^d_{cd}) + T_{ab}\, (F^a\, F^b)_{;c} \\
&= T_{ab\,;c}\, (F^a\, F^b) + T_{ab}\, (F^a\, F^b)_{;c}
\end{aligned} \tag{19.17}$$

where

$$T_{ab\,;c} = T_{ab,c} - \Gamma^d_{ca}\, T_{db} - \Gamma^d_{cb}\, T_{ad} \tag{19.18}$$

defines the components to the covariant derivative of the second order tensor T_{ab}. To reach this result, we used the symmetry $\Gamma^a_{bc} = \Gamma^a_{cb}$. This covariant derivative results in a tensor of type $\binom{0}{3}$. Similar considerations yield the expressions

$$T^{ab}_{;c} = T^{ab}_{,c} + \Gamma^a_{cd}\, T^{db} + \Gamma^b_{cd}\, T^{ad}, \tag{19.19}$$

which results in a tensor of type $\binom{2}{1}$, as well as

$$T^a_{b\,;c} = T^a_{b\,,c} + \Gamma^a_{cd}\, T^d_b - \Gamma^d_{bc}\, T^a_d \tag{19.20}$$

which results in a tensor of type $\binom{1}{2}$.

19.4 Christoffel symbols in terms of the metric

The vanishing covariant derivative of the metric, equation (19.16), combined with equation (19.18) leads to the identity

$$0 = g_{ab;c} = g_{ab,c} - \Gamma^d_{ca}\, g_{db} - \Gamma^d_{cb}\, g_{ad}, \tag{19.21}$$

which can be solved for the Christoffel symbols

$$\Gamma^c_{ab} = \frac{1}{2}\, g^{cd}(g_{da,b} + g_{db,a} - g_{ab,d}). \tag{19.22}$$

We now see explicitly that the Christoffel symbols are *symmetric* on their lower two indices

$$\Gamma^c_{ab} = \Gamma^c_{ba}. \tag{19.23}$$

19.5 Covariant divergence of a vector

There are many places in fluid mechanics where one needs to compute the covariant divergence of a vector. For example, the covariant divergence of the velocity field is proportional to the time rate of change of the density following a fluid parcel. The covariant convergence of a tracer flux contributes to the time tendency of the tracer. In general, the covariant divergence of the components to a vector results in a scalar

$$F^a_{;a} = F^a_{,a} + \Gamma^a_{ab} F^b. \tag{19.24}$$

Expression (19.22) for the Christoffel symbols yields for the contraction

$$\begin{aligned} \Gamma^a_{ab} &= \frac{1}{2}\, g^{ad}(g_{da,b} + g_{db,a} - g_{ab,d}) \\ &= \frac{1}{2}\, g^{ad} g_{ad,b} \end{aligned} \tag{19.25}$$

where symmetry of the metric tensor and its inverse was used.

The above expression for the covariant divergence is general, but it can be rendered in a more convenient form. For this purpose, employ the following relation, which follows for any symmetric positive definite matrix such as the metric tensor

$$\begin{aligned} \det(A) &= e^{\ln \det(A)} \\ &= e^{\ln(\Pi_i \Lambda_i)} \\ &= e^{\Sigma_i \ln \Lambda_i} \\ &= e^{Tr(\ln A)}. \end{aligned} \tag{19.26}$$

The first equality is a simple identity. The second equality relates the determinant of a matrix to the product of its eigenvalues. The third equality is a simple identity. The fourth equality relates the sum of the eigenvalues of a matrix to the trace of this matrix. Each of these identities are trivial to verify using a set of coordinates in

which the matrix is diagonal. For any symmetric and positive definite matrix, such a set of coordinates always exists. This result renders

$$\begin{aligned}\partial_c \ln \det(A) &= \partial_c [Tr(\ln A)] \\ &= Tr(\partial_c \ln A) \\ &= Tr(A^{-1}\partial_c A).\end{aligned} \tag{19.27}$$

With A equal to the metric and $\mathcal{G} = \det(g_{ab})$, this result yields

$$\partial_c \ln \mathcal{G} = g^{ab}\, g_{ab,c}, \tag{19.28}$$

which in turn renders for the contracted Christoffel symbol

$$\Gamma^a_{ac} = \partial_c \ln \sqrt{\mathcal{G}}. \tag{19.29}$$

This result brings the covariant divergence of a vector to the form

$$\begin{aligned}F^a_{;a} &= F^a_{,a} + F^a\, \partial_a \ln \sqrt{\mathcal{G}} \\ &= \frac{1}{\sqrt{\mathcal{G}}} \left(\sqrt{\mathcal{G}}\, F^a\right)_{,a}.\end{aligned} \tag{19.30}$$

This is a very convenient result and will be seen again in various contexts.

19.6 Covariant divergence of a second order tensor

In many applications, it is necessary to compute the covariant divergence of the components to a second order tensor T^{ab}. An important example includes the covariant divergence of the symmetric stress tensor, which typically parameterizes subgrid scale transfer of momentum (see Chapter 15). The results from Section 19.3 renders

$$T^{ab}_{;b} = T^{ab}_{,b} + \Gamma^a_{bd}\, T^{db} + \Gamma^b_{bd}\, T^{ad}, \tag{19.31}$$

To simplify this expression, it is useful to split the tensor components T^{ab} into their symmetric and anti-symmetric parts

$$S^{ab} = (T^{ab} + T^{ba})/2 \tag{19.32}$$

$$A^{ab} = (T^{ab} - T^{ba})/2. \tag{19.33}$$

Using equation (19.29), we find the covariant divergence of the symmetric components to the tensor is given by

$$S^{ab}_{;b} = \frac{1}{\sqrt{\mathcal{G}}} \left(\sqrt{\mathcal{G}}\, S^{ab}\right)_{,b} + \Gamma^a_{bd}\, S^{bd}. \tag{19.34}$$

The term $\Gamma^a_{bd}\, S^{bd}$ does not generally vanish since both Γ^a_{bd} and S^{bd} are symmetric under interchange of b, d. For the anti-symmetric components, however, the analogous term does vanish, hence leading to the covariant divergence of an anti-symmetric tensor

$$\begin{aligned}A^{ab}_{;b} &= A^{ab}_{,b} + \Gamma^b_{bd}\, A^{ad} \\ &= A^{ab}_{,b} + (\ln \sqrt{\mathcal{G}})_{,b}\, A^{ab} \\ &= \frac{1}{\sqrt{\mathcal{G}}} \left(\sqrt{\mathcal{G}}\, A^{ab}\right)_{,b}.\end{aligned} \tag{19.35}$$

This relation is analogous to the covariant divergence of a vector given by equation (19.30). Note that $F^a \equiv A^{ab}_{;b}$ form the components to a vector field which has zero covariant divergence

$$F^a_{;a} = A^{ab}_{;b;a} = 0. \tag{19.36}$$

This result follows trivially from the anti-symmetry of the components A^{ab} under interchange of a, b and the symmetry of the double covariant derivative under the same interchange. This result is the mathematical basis for the skew-flux description of tracer stirring (Section 7.2).

19.7 Covariant Laplacian of a scalar

Equation (19.11) showed that the covariant derivative of a scalar is the same as the partial derivative $V_{;a} = V_{,a}$. Recall also that $V_{,a}$ form the components to the gradient one-form $d\tilde{V} \equiv V_{,a}\,\tilde{e}^a$ (equation (19.4)). Raising the index on these components yields the components to a vector $g^{ab}\,V_{,a}$. Taking the covariant divergence of these components then yields a scalar field, which defines the covariant Laplacian of the scalar function V

$$(g^{ab}\,V_{,a})_{;b} = \frac{1}{\sqrt{\mathcal{G}}}\left(\sqrt{\mathcal{G}}\,g^{ab}\,V_{,a}\right)_{,b} \tag{19.37}$$

where the relation (19.30) for the covariant divergence of a vector was employed.

19.8 Covariant curl of a vector

As mentioned in Section 18.12, the scaled Levita-Civita symbol $\varepsilon_{ab} = \mathcal{G}^{1/2}\,\epsilon_{ab}$ is useful for generalizing the curl operation from Cartesian coordinates in Euclidean space to arbitrary coordinates on a curved manifold. As such, let

$$(\text{curl}\,\vec{F}) = \varepsilon^{abc}\,F_{c;b}\,\vec{e}_a \tag{19.38}$$

define the covariant curl vector. To simplify the curl, recall the expression (19.14) for the covariant derivative $F_{c;b} = F_{c,b} - \Gamma^a_{cb}\,F_a$. Conveniently, the contraction $\varepsilon^{abc}\,\Gamma^a_{cb}$ vanishes identically since $\varepsilon^{abc} = -\varepsilon^{acb}$ whereas $\Gamma^a_{cb} = \Gamma^a_{bc}$. Hence, we are left with the general expression for the covariant curl which involves just partial derivatives

$$(\text{curl}\,\vec{F}) = \varepsilon^{abc}\,F_{c,b}\,\vec{e}_a. \tag{19.39}$$

19.9 Covariant Laplacian of a vector

Recall from Section 19.2.2 that the covariant derivative of a vector is given by

$$\partial_a\vec{F} = F^b_{;a}\,\vec{e}_b, \tag{19.40}$$

where $F^b_{;a} = F^b_{,a} + \Gamma^b_{ad} F^d$ is the covariant derivative of the components to the vector. The covariant Laplacian of a vector is given by

$$\begin{aligned}
\partial^a \partial_a \vec{F} &= \partial^a \left(F^b_{;a} \vec{e}_b \right) \\
&= g^{ac} \partial_c \left(F^b_{;a} \vec{e}_b \right) \\
&= g^{ac} \left(F^d_{;a,c} + \Gamma^d_{cb} F^b_{;a} \right) \vec{e}_d \\
&= g^{ac} F^d_{;a;c} \vec{e}_d .
\end{aligned} \tag{19.41}$$

In general, the components $g^{ac} F^d_{;a;c}$ to the Laplacian are tedious to compute. Appendix 2 of Batchelor (1967) provides useful expressions for various orthogonal coordinates.

19.10 Integral theorems

The integral theorems from Cartesian vector analysis transform in a straightforward manner to arbitrary coordinates in arbitrary smooth spaces. The easiest way to prove the theorems in general is to invoke the ideas from Section 19.1 in which the integral theorems are written in a tensorially correct manner and then partial derivatives are changed to covariant derivatives. Gauss's Law, or the Divergence Theorem, provides a useful example of the approach. Namely, the volume integral of the divergence of a vector is given by

$$\begin{aligned}
\int dV \, F^a_{;a} &= \int d^3\xi \, \sqrt{\mathcal{G}} \, \frac{1}{\sqrt{\mathcal{G}}} \left(\sqrt{\mathcal{G}} \, F^a \right)_{,a} \\
&= \int dA \, F^a \, N_a
\end{aligned} \tag{19.42}$$

where the results from Section 18.11 were used for the invariant volume element, the covariant divergence was written as given in Section 19.5, N_a are components to the outward normal one-form for the boundary surface, $dA = d^2w \sqrt{\mathcal{G}}$ represents the invariant surface area element using surface coordinates (w^1, w^2), and the traditional form of Gauss's Law has been invoked. Other integral theorems are generalized in a similar manner.

19.11 Orthogonal curvilinear coordinates

The case of orthogonal curvilinear coordinates occurs frequently in applications. Consequently, it is useful to specialize some of the previous discussions to this case. Many of the following results can be found, amongst other places, at the end of chapter 1 in Morse and Feshbach (1953), various sections in Aris (1962), and Appendix 2 of Batchelor (1967).

Orthogonal coordinates are characterized by a diagonal metric tensor. Hence, the most general form of the metric written in orthogonal coordinates takes the form

$$g_{ab} = \mathrm{diag} \left((h_1)^2, (h_2)^2, (h_3)^2 \right) . \tag{19.43}$$

This metric is associated with the infinitesimal line element

$$\begin{aligned} ds^2 &= g_{ab}\, d\xi^a\, d\xi^b \\ &= (h_1\, d\xi^1)^2 + (h_2\, d\xi^2)^2 + (h_3\, d\xi^3)^2 \end{aligned} \tag{19.44}$$

(recall Section 18.11). It follows that the components to the inverse metric are given by

$$g^{ab} = \text{diag}\left((h_1)^{-2}, (h_2)^{-2}, (h_3)^{-2}\right). \tag{19.45}$$

For some purposes, the single label on the h_a can be thought of as a single covariant label since the squares of the h_a yield the components to the second order covariant metric tensor. With the metric as such, the square root of its determinant is given by

$$\sqrt{\mathcal{G}} = h_1\, h_2\, h_3. \tag{19.46}$$

Using orthogonal coordinates, the basis vectors take the form

$$\vec{e}_1 = h_1\, \hat{\mathbf{e}}_{(1)} \tag{19.47}$$

$$\vec{e}_2 = h_2\, \hat{\mathbf{e}}_{(2)} \tag{19.48}$$

$$\vec{e}_3 = h_3\, \hat{\mathbf{e}}_{(3)}, \tag{19.49}$$

where $\hat{\mathbf{e}}_{(a)}$ are dimensionless orthonormal *unit directions*. The index on the unit directions is enclosed in parentheses to advertise the fact that this label is not tensorial; i.e., the unit directions do not transform as tensors. Rather, the functions h_a carry the tensorial properties of the basis vectors $\vec{e}_a$. Note that the basis vectors are determined by writing the metric tensor in the form

$$g_{ab} = \vec{e}_a \cdot \vec{e}_b. \tag{19.50}$$

The corresponding basis one-forms are

$$\tilde{e}^1 = (h_1)^{-1}\, \hat{\mathbf{e}}^{(1)} \tag{19.51}$$

$$\tilde{e}^2 = (h_2)^{-1}\, \hat{\mathbf{e}}^{(2)} \tag{19.52}$$

$$\tilde{e}^3 = (h_3)^{-1}\, \hat{\mathbf{e}}^{(3)}, \tag{19.53}$$

where the unit directions are identical whether their indices are raised or lowered: $\hat{\mathbf{e}}^{(a)} = \hat{\mathbf{e}}_{(a)}$.

Spherical coordinates (λ, ϕ, r) on Euclidean $\mathcal{R}^3$ provides the canonical example

of a nontrivial set of orthogonal curvilinear coordinates

$$h_1 = r\,\cos\phi \tag{19.54}$$
$$h_2 = r \tag{19.55}$$
$$h_3 = 1 \tag{19.56}$$
$$\sqrt{\mathcal{G}} = r^2\,\cos\phi \tag{19.57}$$
$$\vec{e}_1 = r\,\cos\phi\,\hat{\boldsymbol{\lambda}} \tag{19.58}$$
$$\vec{e}_2 = r\,\hat{\boldsymbol{\phi}} \tag{19.59}$$
$$\vec{e}_3 = \hat{\mathbf{r}} \tag{19.60}$$
$$\tilde{e}^1 = (r\,\cos\phi)^{-1}\,\hat{\boldsymbol{\lambda}} \tag{19.61}$$
$$\tilde{e}^2 = r^{-1}\,\hat{\boldsymbol{\phi}} \tag{19.62}$$
$$\tilde{e}^3 = \hat{\mathbf{r}} \tag{19.63}$$

where the unit directions point in the radial, eastward, and northward directions, respectively (see Figure 18.1). In contrast, with Cartesian coordinates for $\mathcal{R}^3$, all the h_a are trivially set to unity.

19.11.1 Physical components of tensors

In terms of a locally orthogonal basis of unit directions, an arbitrary vector takes the form

$$\begin{aligned}\vec{F} &= F^a\,\vec{e}_a \\ &= (F^a\,h_a)\,\hat{\mathbf{e}}_{(a)} \\ &\equiv \mathcal{F}^{(a)}\,\hat{\mathbf{e}}_{(a)}.\end{aligned} \tag{19.64}$$

Note that there is no summation performed on the $F^a\,h_a$ product. Since the unit directions $\hat{\mathbf{e}}_{(a)}$ are dimensionless, the terms $\mathcal{F}^{(a)}$ each have the same physical dimensions. Consequently, $\mathcal{F}^{(a)}$ are termed the *physical components* to the vector $\vec{F}$ (see, e.g., Section 7.41 of Aris (1962) or Section 4.8 of Weinberg (1972)). The physical components, due to their dimensional consistency, are often the forms for which tensorial quantities appear in physical theories. They are also the forms most likely to appear in numerical models. As with the unit vectors, the index on the physical components is enclosed in parentheses since the physical components do not transform as components of a tensor. Rather, they transform as

$$\begin{aligned}\mathcal{F}^{(\bar{a})} &= F^{\bar{a}}\,h_{\bar{a}} \\ &= \Lambda^{\bar{a}}{}_a\,F^a\,h_{\bar{a}} \\ &= \frac{h_{\bar{a}}}{h_a}\,\Lambda^{\bar{a}}{}_a\,\mathcal{F}^{(a)}.\end{aligned} \tag{19.65}$$

The presence of the $h_{\bar{a}}/h_a$ factor precludes the physical components from transforming as a tensor. Another useful notation is to use the Cartesian symbols x and y to denote physical components, instead of the parentheses. For example,

$$\mathcal{F}^{(a)} = (\mathcal{F}^x, \mathcal{F}^y). \tag{19.66}$$

This notation will be used when the context is clear.

Expressed in terms of the orthogonal coordinates, an arbitrary one-form is written

$$\begin{aligned}\tilde{F} &= F_a\,\tilde{e}^a \\ &= (F_a/h_a)\,\hat{\mathbf{e}}^{(a)} \\ &= \mathcal{F}_{(a)}\,\hat{\mathbf{e}}^{(a)}.\end{aligned} \tag{19.67}$$

An important one-form operator is the gradient operator, to be considered in Section 19.11.3. We will find it useful to introduce the physical components to this operator via the use of physical components to the partial derivative operator

$$(\partial_x, \partial_y, \partial_z) = (\partial_{(1)}, \partial_{(2)}, \partial_{(3)}) = ((h_1)^{-1}\,\partial_1, (h_2)^{-1}\,\partial_2, \partial_z). \tag{19.68}$$

As for vectors, the fields $\mathcal{F}_{(a)}$ each have the same physical dimensions, and they are called the physical components to the one-form $\tilde{F}$. As with the physical components to the vectors, the physical components to the one-forms do not transform as tensors.

The above results for a vector and one-form generalize to an arbitrary tensor. For example, if we express a second order tensor in terms of its contravariant components, these components take the following form when using orthogonal coordinates.

$$\begin{aligned}\mathbf{T} &= T^{ab}\,\vec{e}_a\,\vec{e}_b \\ &= (h_a\,T^{ab}\,h_b)\,\hat{\mathbf{e}}_{(a)}\,\hat{\mathbf{e}}_{(b)} \\ &= \mathcal{T}^{(ab)}\,\hat{\mathbf{e}}_{(a)}\,\hat{\mathbf{e}}_{(b)}.\end{aligned} \tag{19.69}$$

The components

$$\mathcal{T}^{(ab)} = h_a\,T^{ab}\,h_b, \tag{19.70}$$

where there is no implied sum, are identified as the physical components of the contravariant tensor T^{ab}. Each of the physical components $\mathcal{T}^{(ab)}$ has the same dimensions. Likewise, the same second order tensor written in terms of its mixed second order components can be expressed in terms of the orthogonal coordinates as

$$\begin{aligned}\mathbf{T} &= T^a_b\,\vec{e}_a\,\tilde{e}^b \\ &= (h_a/h_b)\,T^a_b\,\hat{\mathbf{e}}_{(a)}\,\hat{\mathbf{e}}^{(b)} \\ &= \mathcal{T}^{(a)}_{(b)}\,\hat{\mathbf{e}}_{(a)}\,\hat{\mathbf{e}}^{(b)},\end{aligned} \tag{19.71}$$

where

$$\mathcal{T}^{(a)}_{(b)} = (h_a/h_b)\,T^a_b \tag{19.72}$$

are the physical components, each of which have the same dimensions, and again where there is no implied sum.

19.11.2 Eliminating the stretching functions

Orthogonal coordinates afford an added feature that is quite useful to exploit when discretizing the equations for an ocean model. Since the generalized coordinates are orthogonal, a coordinate increment in one direction is independent of the other direction:

$$\partial_1 \left(d\xi^2\right) = \partial_2 \left(d\xi^1\right) = 0. \tag{19.73}$$

Hence, for example,

$$h_2 \left(u/h_2\right)_{,x} = h_2 \, d\xi^2 \left(u/h_2 \, d\xi^2\right)_{,x} = dy \left(u/dy\right)_{,x}. \tag{19.74}$$

This property allows for all formulae to be written in terms of the physical displacements dx and dy, and so to eliminate stretching functions h_1 and h_2. This property is convenient since the physical displacements are ultimately what are relevant for the ocean model.

19.11.3 Gradient one-form

In orthogonal coordinates, the gradient one-form acting on a scalar is given by

$$\nabla V = \left(\tilde{e}^a \, \partial_a\right) V = \left(\hat{\mathbf{e}}^{(a)} \, \partial_{(a)}\right) V. \tag{19.75}$$

The second equality introduced the physical components to the gradient one-form operator

$$\nabla = \hat{\mathbf{e}}^{(a)} \, \partial_{(a)}, \tag{19.76}$$

with $\hat{\mathbf{e}}^{(a)}$ unit-directions, and physical components to the partial derivatives are

$$\partial_{(a)} = \left(h_{(a)}\right)^{-1} \partial_a, \tag{19.77}$$

with no implied sum on the right hand side. Often the physical components to the partial derivative are written $\partial_{(a)} = \left(\partial_x, \partial_y, \partial_z\right)$, yet with no implication that the space is flat. Indeed, the non-flat nature of the sphere manifests in the non-vanishing commutator

$$\left[\partial_x, \partial_y\right] = \partial_x \, \partial_y - \partial_y \, \partial_x = \left(\partial_x \, \ln \, dy\right) \partial_y - \left(\partial_y \, \ln \, dx\right) \partial_x. \tag{19.78}$$

It is useful to touch bases with spherical coordinates, where the gradient of a scalar is

$$\nabla V = V_{,\lambda} \left(R \, \cos \phi\right)^{-1} \hat{\boldsymbol{\lambda}} + V_{,\phi} \, R^{-1} \, \hat{\boldsymbol{\phi}} + \hat{\mathbf{z}} \, V_{,z} = V_{,x} \, \hat{\boldsymbol{\lambda}} + V_{,y} \, \hat{\boldsymbol{\phi}} + \hat{\mathbf{z}} \, V_{,z} \tag{19.79}$$

and the gradient operator is

$$\nabla = \hat{\boldsymbol{\lambda}} \, \left(R \, \cos \phi\right)^{-1} \partial_\lambda + \hat{\boldsymbol{\phi}} \, R^{-1} \, \partial_\phi + \hat{\mathbf{z}} \, \partial_z. \tag{19.80}$$

19.11.4 Covariant divergence of a vector

From Section 19.5, the expression for the covariant divergence of a vector

$$F^a_{;a} = \frac{1}{\sqrt{\mathcal{G}}} \left(\sqrt{\mathcal{G}}\, F^a\right)_{,a} \tag{19.81}$$

takes the form in orthogonal coordinates

$$\begin{aligned} F^a_{;a} &= \frac{1}{h_1\, h_2\, h_3} \left(h_1\, h_2\, h_3\, F^a\right)_{,a} \\ &= \frac{1}{h_1\, h_2\, h_3} \left\{ \left(h_2\, h_3\, (h_1\, F^1)\right)_{,1} + \left(h_1\, h_3\, (h_2\, F^2)\right)_{,2} + \left(h_1\, h_2\, (h_3\, F^3)\right)_{,3} \right\} \\ &= \frac{1}{h_1\, h_2\, h_3} \left\{ \left(h_2\, h_3\, \mathcal{F}^{(1)}\right)_{,1} + \left(h_1\, h_3\, \mathcal{F}^{(2)}\right)_{,2} + \left(h_1\, h_2\, \mathcal{F}^{(3)}\right)_{,3} \right\}. \end{aligned} \tag{19.82}$$

Eliminating the stretching functions in favor of the physical displacements, according to the properties noted in Section 19.11.2, leads to

$$F^a_{;a} = \nabla \cdot \mathbf{F} \tag{19.83}$$

$$= (dy\, dz)^{-1} \left(dy\, dz\, \mathcal{F}^{(x)}\right)_{,x} + (dx\, dz)^{-1} \left(dx\, dz\, \mathcal{F}^{(y)}\right)_{,y} + (dx\, dy)^{-1} \left(dx\, dy\, \mathcal{F}^{(z)}\right)_{,z} \tag{19.84}$$

where the first equality introduced the common notation used to represent the divergence of a vector. For spherical coordinates (λ, ϕ, r), with $(h_1, h_2, h_3) = (r \cos\phi, r, 1)$, the divergence becomes

$$\nabla \cdot \mathbf{F} = \frac{1}{r^2 \cos\phi} \left((r^2 \cos\phi\, F^\lambda)_{,\lambda} + (r^2 \cos\phi\, F^\phi)_{,\phi} + (r^2 \cos\phi\, F^r)_{,r} \right) \tag{19.85}$$

$$= (r \cos\phi)^{-1} \left((\mathcal{F}^\lambda)_{,\lambda} + (\mathcal{F}^\phi \cos\phi)_{,\phi} \right) + r^{-2} \left(r^2\, \mathcal{F}^r\right)_{,r}. \tag{19.86}$$

Note that with general vertical coordinates (Chapter 5), or even with partial cells (Adcroft et al. (1997) and Pacanowski and Gnanadesikan (1998)) used in many z-coordinate ocean models (Section 5.2.1), the vertical distance dz is a function of both the vertical and horizontal position. Hence, dz cannot be removed from the horizontal partial derivative in equation (19.84). This point is brought out in Section 14.7 where we discuss the divergence of tracer fluxes.

19.11.5 Covariant divergence of an anti-symmetric tensor

Employing the general results from Section 19.6, we find that for orthogonal coordinates, the covariant divergence of an anti-symmetric tensor is given by

$$\begin{aligned} A^{ab}_{;b} &= \frac{1}{\sqrt{\mathcal{G}}} \left(\sqrt{\mathcal{G}}\, A^{ab}\right)_{,b} \\ &= \frac{1}{h_1\, h_2\, h_3} \left(h_1\, h_2\, h_3\, A^{ab}\right)_{,b}. \end{aligned} \tag{19.87}$$

In particular,

$$F^1 \equiv A^{1b}_{;b} = \frac{1}{h_1\, h_2\, h_3}\, \left[(h_1\, h_2\, h_3\, A^{12})_{,2} + (h_1\, h_2\, h_3\, A^{13})_{,3}\right]$$
$$= \frac{1}{h_1\, h_2\, h_3}\, \left[(h_3\, \mathcal{A}^{(12)})_{,2} + (h_2\, \mathcal{A}^{(13)})_{,3}\right], \tag{19.88}$$

and likewise

$$F^2 \equiv A^{2b}_{;b} = \frac{1}{h_1\, h_2\, h_3}\, \left[(h_3\, \mathcal{A}^{(21)})_{,1} + (h_1\, \mathcal{A}^{(23)})_{,3}\right] \tag{19.89}$$

$$F^3 \equiv A^{3b}_{;b} = \frac{1}{h_1\, h_2\, h_3}\, \left[(h_2\, \mathcal{A}^{(31)})_{,1} + (h_1\, \mathcal{A}^{(32)})_{,2}\right]. \tag{19.90}$$

The physical components of the vector field F^a are therefore given by

$$\mathcal{F}^{(1)} = h_1\, F^1 = \frac{1}{h_2\, h_3}\, \left[(h_3\, \mathcal{A}^{(12)})_{,2} + (h_2\, \mathcal{A}^{(13)})_{,3}\right] \tag{19.91}$$

$$\mathcal{F}^{(2)} = h_2\, F^2 = \frac{1}{h_1\, h_3}\, \left[(h_3\, \mathcal{A}^{(21)})_{,1} + (h_1\, \mathcal{A}^{(23)})_{,3}\right] \tag{19.92}$$

$$\mathcal{F}^{(3)} = h_3\, F^3 = \frac{1}{h_1\, h_2}\, \left[(h_2\, \mathcal{A}^{(31)})_{,1} + (h_1\, \mathcal{A}^{(32)})_{,2}\right]. \tag{19.93}$$

Eliminating the stretching functions leads to

$$\mathcal{F}^{(x)} = (dz)^{-1}\, (dz\, \mathcal{A}^{(12)})_{,y} + (dy)^{-1}\, (dy\, \mathcal{A}^{(13)})_{,z} \tag{19.94}$$

$$\mathcal{F}^{(y)} = (dz)^{-1}\, (dz\, \mathcal{A}^{(21)})_{,x} + (dx)^{-1}\, (dx\, \mathcal{A}^{(23)})_{,z} \tag{19.95}$$

$$\mathcal{F}^{(z)} = (dy)^{-1}\, (dz\, \mathcal{A}^{(31)})_{,x} + (dx)^{-1}\, (dx\, \mathcal{A}^{(32)})_{,y}. \tag{19.96}$$

In terms of spherical coordinates (λ, ϕ, r), the physical components take the form

$$\mathcal{F}^{(\lambda)} = \frac{1}{r}\, \left[(r\, \mathcal{A}^{(\lambda r)})_{,r} + (\mathcal{A}^{(\lambda\phi)})_{,\phi}\right] \tag{19.97}$$

$$\mathcal{F}^{(\phi)} = \frac{1}{r\, \cos\phi}\, \left[(r\, \cos\phi\, \mathcal{A}^{(\phi r)})_{,r} + (\mathcal{A}^{(\phi\lambda)})_{,\phi}\right] \tag{19.98}$$

$$\mathcal{F}^{(r)} = \frac{1}{r^2\, \cos\phi}\, \left[(r\, \mathcal{A}^{(r\lambda)})_{,\lambda} + (r\, \cos\phi\, \mathcal{A}^{(r\phi)})_{,\phi}\right] \tag{19.99}$$

19.11.6 Covariant Laplacian of a scalar

In orthogonal coordinates, the covariant Laplacian of a scalar, derived in Section 19.7, takes the form

$$\frac{1}{\sqrt{\mathcal{G}}}\, \left(\sqrt{\mathcal{G}}\, g^{ab}\, V_{,a}\right)_{,b} = \frac{1}{h_1\, h_2\, h_3}\, \left\{\left(\frac{h_2\, h_3}{h_1}\, V_{,1}\right)_{,1} + \left(\frac{h_1\, h_3}{h_2}\, V_{,2}\right)_{,2} + \left(\frac{h_1\, h_2}{h_3}\, V_{,3}\right)_{,3}\right\}. \tag{19.100}$$

Eliminating the stretching functions leads to

$$\nabla^2 V = (dy\, dz)^{-1}\, (dy\, dz\, V_{,x})_{,x} + (dx\, dz)^{-1}\, (dx\, dz\, V_{,y})_{,y} + (dx\, dy)^{-1}\, (dx\, dy\, V_{,z})_{,z} \tag{19.101}$$

where we introduced the common notation $\nabla^2 V$ for the Laplacian of a scalar. In spherical coordinates

$$\nabla^2 V = \frac{1}{r^2}(r^2\, V_{,r})_{,r} + \frac{1}{r^2\, \cos^2\phi}\, (V_{,\lambda\lambda} + (V_{,\phi}\, \cos\phi)_{,\phi}) + V_{,zz}. \tag{19.102}$$

19.11.7 Covariant curl of a vector

From Section 19.8, the covariant curl takes the form

$$\nabla \wedge \vec{F} = \text{curl}\,\vec{F} = \varepsilon^{abc}\, F_{c,b}\, \vec{e}_a, \tag{19.103}$$

where the first expression is that commonly used in vector analysis. With orthogonal coordinates, the curl becomes

$$\begin{aligned}
\text{curl}\,\vec{F} &= \mathcal{G}^{-1/2}\, \epsilon^{abc}\, F_{c,b}\, \vec{e}_a \\
&= \frac{h_a\, \hat{e}_{(a)}}{h_1\, h_2\, h_3} \epsilon^{abc}\, (h_c\, \mathcal{F}_{(c)})_{,b} \\
&= \frac{1}{h_1\, h_2\, h_3} \begin{vmatrix} \hat{\mathbf{e}}_{(1)}\, h_1 & \hat{\mathbf{e}}_{(2)}\, h_2 & \hat{\mathbf{e}}_{(3)}\, h_3 \\ \partial_1 & \partial_2 & \partial_3 \\ h_1\, \mathcal{F}_{(1)} & h_2\, \mathcal{F}_{(2)} & h_3\, \mathcal{F}_{(3)} \end{vmatrix},
\end{aligned} \tag{19.104}$$

where the last expression is written in a determinantal form. Eliminating the stretching functions leads to

$$\begin{aligned}
\text{curl}\,\vec{F} = {} & \hat{\mathbf{e}}_{(1)} \left[(dz)^{-1}\, (dz\, \mathcal{F}_{(3)})_{,y} - (dy)^{-1}\, (dy\, \mathcal{F}_{(2)})_{,z} \right] \\
& + \hat{\mathbf{e}}_{(2)} \left[(dx)^{-1}\, (dx\, \mathcal{F}_{(1)})_{,z} - (dz)^{-1}\, (dz\, \mathcal{F}_{(3)})_{,x} \right] \\
& + \hat{\mathbf{e}}_{(3)} \left[(dy)^{-1}\, (dy\, \mathcal{F}_{(2)})_{,x} - (dx)^{-1}\, (dx\, \mathcal{F}_{(1)})_{,y} \right].
\end{aligned} \tag{19.105}$$

19.12 Summary of some salient results

Given the importance of the results of Chapters 18 and 19, and given the likelihood that most readers will avoid reading these chapters, it has been found useful to summarize here some of the main results to be used throughout this book.

19.12.1 Rules for tensor analysis on manifolds

The following rules and ideas are sufficient for applications in this book.

- **Conservation of indices**: Lower and upper tensor indices are balanced across equal signs.
- **Einstein summation convention**: Repeated indices are summed, unless otherwise noted.
- **Metric tensor**: The metric tensor provides a means to measure the distance between two points on a manifold:

  $$(ds)^2 = g_{mn}\, d\xi^m\, d\xi^n. \tag{19.106}$$

  In this expression, $(ds)^2$, often written ds^2, is the squared infinitesimal arc-length between the points, ξ^m is a general coordinate for a point, and $m = 1, 2, 3$ labels the coordinate (m is *not* a power).

The metric for spherical coordinates on a sphere is diagonal. With coordinates $(\xi^1, \xi^2, \xi^3) = (\lambda, \phi, r)$, where λ is longitude and ϕ latitude, the metric is

$$g_{mn} = \text{diag}(g_{11}, g_{22}, g_{33}) = \text{diag}((r\cos\phi)^2, r^2, 1). \tag{19.107}$$

The inverse metric components g^{mn} are also needed, and they are given by

$$g^{mn} = \text{diag}((r\ \cos\phi)^{-2}, r^{-2}, 1). \tag{19.108}$$

In Cartesian coordinates, the metric tensor is given by

$$g_{mn} = \delta_{mn} = \delta^{mn} = \delta^m_n, \tag{19.109}$$

where δ is the unit or Kronecker delta tensor. There is no distinction between raised and lowered indices in Cartesian coordinates, hence the ability to jettison the conservation of indices rule when working with Cartesian tensors. For curvilinear coordinates, however, conservation of indices is necessary.

- **Covariant and contravariant**: A lower label is often termed "covariant" and an upper label "contravariant." The mnemonic "co-low" assists in remembering the terminology.

  Covariant and contravariant tensors can be considered dual, where the transformation is through the metric tensor. For example, the covariant components to the velocity vector u_m are related to the contravariant components through

$$u_m = g_{mn} u^n. \tag{19.110}$$

  Some examples are useful. In Cartesian coordinates, the velocity vector takes the familiar form

$$(u^1, u^2, u^3) = (u_1, u_2, u_3) = \left(\frac{Dx}{Dt}, \frac{Dy}{Dt}, \frac{Dz}{Dt}\right), \tag{19.111}$$

  where again there is no distinction between covariant and contravariant for Cartesian tensors. In spherical coordinates, however, the contravariant velocity components are

$$(u^1, u^2, u^3) = \left(\frac{D\lambda}{Dt}, \frac{D\phi}{Dt}, \frac{Dr}{Dt}\right), \tag{19.112}$$

  whereas the covariant components $u_m = g_{mn}\, u^n$ are

$$(u_1, u_2, u_3) = \left((r\ \cos\phi)^2\, \frac{D\lambda}{Dt}, r^2\, \frac{D\phi}{Dt}, \frac{Dr}{Dt}\right). \tag{19.113}$$

- **Notation**: As the above indicates, for curvilinear tensor analysis the difference between a raised and lowered label is important. Additionally, in order to avoid confusion, partial derivatives are denoted with a comma:

$$u_{m,n} = \frac{\partial u_m}{\partial \xi^n}. \tag{19.114}$$

- **Covariant derivative**: In order to account for nonconstant basis vectors on a curved manifold, it is necessary to generalize partial derivatives to so-called covariant derivatives. In particular, the strain tensor (see Section 15.3.1) has components

$$2\, e_{mn} = u_{m;n} + u_{n;m} \tag{19.115}$$

where the comma has been generalized to a semi-colon.

For a "torsionless" manifold, such as a sphere, each component of the metric tensor has a vanishing covariant derivative

$$g_{mn;p} = 0. \tag{19.116}$$

This is a trivial property for Cartesian coordinates on a plane, in which case the metric is the constant unit tensor and the covariant derivative a partial derivative

$$\delta_{mn;p} = \delta_{mn,p} = 0. \tag{19.117}$$

However, for curvilinear coordinates $g_{mn;p} = 0$ is quite useful. For example, it provides for the convenient relation

$$\begin{aligned} u_{m;n} &= \left(g_{mp}\, u^p\right)_{;n} \\ &= g_{mp}\, u^p_{;n}. \end{aligned} \tag{19.118}$$

This result brings the strain tensor components to the form

$$2\, e_{mn} = g_{mp}\, u^p_{;n} + g_{np}\, u^p_{;m}. \tag{19.119}$$

In general, the covariant derivative of a vector on a torsionless manifold is given by

$$u^p_{;n} = u^p_{,n} + \Gamma^p_{mn}\, u^n, \tag{19.120}$$

where Γ^p_{mn} are components to the Christoffel symbol

$$\Gamma^p_{mn} = \frac{1}{2}\, g^{pq}\, (g_{qm,n} + g_{qn,m} - g_{mn,q}). \tag{19.121}$$

A more geometric means of understanding the Christoffel symbol is to note its components form the expansion coefficients of the partial derivative of the basis vectors for a manifold

$$\vec{e}_{a,b} = \Gamma^m_{ab}\, \vec{e}_m. \tag{19.122}$$

That is, the Christoffel symbol accounts for the nonzero changes in the basis vectors on a curved manifold. Note that it is symmetric on the lower two labels:

$$\Gamma^p_{mn} = \Gamma^p_{nm} \tag{19.123}$$

which is the defining property of torsionless manifolds.

- **Transformation rules**: Under a coordinate transformation

$$\xi^{\overline{m}} = \xi^{\overline{m}}(\xi^m), \tag{19.124}$$

tensors transform as, for example,

$$e_{\overline{mn}} = \Lambda^m_{\overline{m}} \Lambda^n_{\overline{n}} e_{mn}, \tag{19.125}$$

where the transformation matrix is given by the partial derivatives

$$\Lambda^m_{\overline{m}} = \frac{\partial \xi^m}{\partial \xi^{\overline{m}}}. \tag{19.126}$$

Sometimes it is useful to write the transformation matrix in traditional matrix form. The convention is that the index which is placed a bit closer to the Λ denotes the row (m in $\Lambda^m_{\overline{m}}$), and the one pushed away a bit is the column ($\overline{m}$ in $\Lambda^m_{\overline{m}}$). The inverse transformation of a tensor takes the form

$$e_{mn} = \Lambda^{\overline{m}}_m \Lambda^{\overline{n}}_n e_{\overline{mn}}, \tag{19.127}$$

where the inverse transformation matrix has components given by

$$\Lambda^{\overline{m}}_m = \frac{\partial \xi^{\overline{m}}}{\partial \xi^m}. \tag{19.128}$$

Transformations of tensors with arbitrary rank generalize with an extra factor of the transformation matrix corresponding to each tensor label. An example of how the fourth order viscosity tensor transforms is given in Section 15.5.1.

19.12.2 Orthogonal coordinates

The metric tensor for orthogonal coordinates is diagonal

$$g_{mn} = \mathrm{diag}(g_{11}, g_{22}, g_{33}), \tag{19.129}$$

where the components $g_{mn} = g_{mn}(t, \xi^1, \xi^2, \xi^3)$ are generally functions of space-time. The infinitesimal arc-length measuring the distance between any two closely spaced points is therefore given by the diagonal quadratic-form

$$\begin{aligned}(ds)^2 &= g_{11}(d\xi^1)^2 + g_{22}(d\xi^2)^2 + g_{33}(d\xi^3)^2 \\ &= (h_1\, d\xi^1)^2 + (h_2\, d\xi^2)^2 + (h_3\, d\xi^3)^2,\end{aligned} \tag{19.130}$$

where the positive metric or *stretching* functions $h_m = \sqrt{g_{mm}}$, with no implied sum, are often useful to introduce. Additionally, the relation between covariant and contravariant components of a tensor is given through a single multiplication. For example,

$$\begin{aligned}u_m &= g_{mn}\, u^n \\ &= g_{mm}\, u^m,\end{aligned} \tag{19.131}$$

relates the covariant velocity components u_m to the contravariant components u^m. Importantly, there is no sum on the m label in the last expression.

For the purposes of large-scale ocean modeling, it is usually sufficient to assume the simpler diagonal form of the metric

$$g_{mn} = \mathrm{diag}(g_{11}, g_{22}, 1), \tag{19.132}$$

where the nontrivial metric components are functions only of the horizontal position. This result follows from the Traditional Approximation (Section 3.1) and is made in this book. Another useful notation is to introduce the determinant of the metric tensor

$$\mathcal{G} = g_{11}\, g_{22}\, g_{33} \tag{19.133}$$

since it appears frequently. With the Traditional Approximation for which $g_{33} = 1$, the determinant is given by

$$\mathcal{G} = g_{11}\, g_{22} = (h_1\, h_2)^2. \tag{19.134}$$

19.12.3 Physical tensor components with orthogonal coordinates

In many applications, it is useful to introduce the *physical components* of a tensor (see Section 7.4 of Aris (1962) or 4.8 of Weinberg (1972) or Section 19.11.1 of this book). For example, the velocity field using spherical coordinates is often written

$$\begin{aligned}(u, v, w) &= \left(\sqrt{g_{\lambda\lambda}}\, u^\lambda, \sqrt{g_{\phi\phi}}\, u^\phi, \sqrt{g_{rr}}\, u^r\right) \\ &= \left(r\, \cos\phi\, \frac{D\lambda}{Dt}, r\, \frac{D\phi}{Dt}, \frac{Dr}{Dt}\right).\end{aligned} \tag{19.135}$$

Additionally, the infinitesimal displacements along the coordinate directions on the sphere are given by

$$(dx, dy, dz) = ((r\, \cos\phi)\, d\lambda, r\, d\phi, dr)\,. \tag{19.136}$$

More generally, for any orthogonal coordinate system, the physical components of the displacement are written

$$(dx, dy, dz) = (h_1\, d\xi^1,\, h_2\, d\xi^2,\, h_3\, d\xi^3). \tag{19.137}$$

Likewise, the physical components of the velocity are

$$(u, v, w) = (h_1\, u^1,\, h_2\, u^2,\, h_3\, u^3). \tag{19.138}$$

Consequently, for example,

$$u^1_{,1} = \sqrt{g_{11}}\, (u/\sqrt{g_{11}})_{,x} = h_1\, (u/h_1)_{,x} \tag{19.139}$$

Note that although the Cartesian notation x, y, z is used for convenience, the coordinates are generally curvilinear.

The key property of the physical components of a tensor is that each has the same dimensions; e.g., length for the physical displacement components, length/time for the physical velocity components, etc. Importantly, the physical components

are *not* components to a true tensor since the tensorial transformation rules are corrupted by the square root of the metric. Correspondingly, the physical components of the partial derivative operator do not generally commute

$$\left[\partial_x, \partial_y\right] = \partial_x \, \partial_y - \partial_y \, \partial_x = (\partial_x \, \ln \, dy) \, \partial_y - (\partial_y \, \ln \, dx) \, \partial_x, \tag{19.140}$$

which vanishes only when the horizontal geometry is flat instead of curved. Hence, as a rule of thumb it is best to perform mathematical manipulations with the tensorial components, and only after establishing the final result should the physical components be introduced.

19.12.4 Eliminating the stretching functions

Orthogonal coordinates afford an added feature that is quite useful to exploit when discretizing the equations. Since the generalized coordinates are orthogonal, a coordinate increment in one direction is independent of the other direction:

$$\partial_1 \left(d\xi^2\right) = \partial_2 \left(d\xi^1\right) = 0. \tag{19.141}$$

Hence, for example,

$$h_2 \left(u/h_2\right)_{,x} = h_2 \, d\xi^2 \left(u/h_2 \, d\xi^2\right)_{,x} = dy \left(u/dy\right)_{,x}. \tag{19.142}$$

This property allows for all formulae to be written in terms of the physical displacements dx and dy, and so to eliminate stretching functions h_1 and h_2. This property is convenient since the physical displacements are ultimately what are relevant for the ocean model.

BIBLIOGRAPHY

Adcroft, A., C. Hill, and J. Marshall, 1997: Representation of topography by shaved cells in a height coordinate ocean model. *Monthly Weather Review*, 125, 2293–2315.

Andrews, D. G., and M. E. McIntyre, 1978: An exact theory of nonlinear waves on a lagrangian-mean flow. *Journal of Fluid Mechanics*, 89, 609–646.

Apel, J. R., 1987: *Principles of Ocean Physics*. International Geophysics Series, Vol. 38, Academic Press, 634 pp.

Aris, R., 1962: *Vectors, Tensors and the Basic Equations of Fluid Mechanics*. Dover Publishing, 262 pp.

Bacon, S., and N. P. Fofonoff, 1996: Oceanic heat flux calculation. *Journal of Atmospheric and Oceanic Technology*, 13, 1327–1329.

Batchelor, G. K., 1967: *An Introduction to Fluid Dynamics*. Cambridge University Press, 615 pp.

Beckers, J. M., H. Burchard, E. Deleersnijder, and P. P. Mathieu, 2000: Numerical discretization of rotated diffusion operators in ocean models. *Monthly Weather Review*, 128, 2711–2733.

Beckmann, A., 1998: The representation of bottom boundary layer processes in numerical ocean circulation models. E. P. Chassignet and J. Verron, Eds., *Ocean Modeling and Parameterization*, NATO ASI Mathematical and Physical Sciences Series., Vol. 516, Kluwer, 135–154.

Beckmann, A., and R. Döscher, 1997: A method for improved representation of dense water spreading over topography in geopotential–coordinate models. *Journal of Physical Oceanography*, 27, 581–591.

Bleck, R., 1978: Finite difference equations in generalized vertical coordinates. part i: total energy conservation. *Contributions to Atmospheric Physics*, 51, 360–372.

Bleck, R., 1998: Ocean modeling in isopycnic coordinates. E. P. Chassignet and J. Verron, Eds., *Ocean Modeling and Parameterization*, NATO ASI Mathematical and Physical Sciences Series., Vol. 516, Kluwer, 423–448.

Bleck, R., and D. B. Boudra, 1981: Initial testing of a numerical ocean circulation model using a hybrid- (quasi-isopycnic) vertical coordinate. *Journal of Physical Oceanography*, 11, 755–770.

Bleck, R., C. Rooth, D. Hu, and L. T. Smith, 1992: Salinity-driven thermocline transients in a wind and thermohaline forced isopycnic coordinate model of the North Atlantic. *Journal of Physical Oceanography*, 22, 1486–1505.

Blumberg, A. F., and G. L. Mellor, 1987: A description of a three-dimensional coastal ocean circulation model. N. Heaps, Ed., *Three-Dimensional Coastal Ocean Models*, Vol. 4, American Geophysical Union.

Boer, G. J., and T. G. Shepherd, 1983: Large-scale two-dimensional turbulence in the atmosphere. *Journal of Atmospheric Sciences*, 40, 164–184.

Bokhove, O., 2000: On hydrostatic flows in isentropic coordinates. *Journal of Fluid Mechanics*, 402, 291–310.

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Quarterly Journal of the Royal Meteorological Society*, 92, 325–334.

Brown, J. A., and K. A. Campana, 1978: An economical time-differencing system for numerical weather prediction. *Monthly Weather Review*, 106, 1125–1136.

Bryan, K., 1969: A numerical method for the study of the circulation of the world ocean. *Journal Computational Physics*, 4, 347–376.

Bryan, K., 1984: Accelerating the convergence to equilibrium of ocean-climate models. *Journal of Physical Oceanography*, 14, 666–673.

Bryan, K., 1989: The design of numerical models of the ocean circulation. J. W. David L.T. Anderson, Ed., *Oceanic circulation models: combining data and dynamics*, Vol. 44, Kluwer Academic Publishers, 465–511.

Bryan, K., and M. D. Cox, 1972: An approximate equation of state for numerical models of the ocean circulation. *Journal of Physical Oceanography*, 4, 510–514.

Bryan, K., J. K. Dukowicz, and R. D. Smith, 1999: On the mixing coefficient in the parameterization of bolus velocity. *Journal of Physical Oceanography*, 29, 2442–2456.

Bryan, K., and L. J. Lewis, 1979: A water mass model of the world ocean. *Journal of Geophysical Research*, 84, 2503–2517.

Bryan, K., S. Manabe, and R. C. Pacanowski, 1975: A global ocean-atmosphere climate model. part ii. the oceanic circulation. *Journal of Physical Oceanography*, 5, 30–46.

Callen, H. B., 1985: *Thermodynamics and an introduction to thermostatics*. John Wiley and Sons.

Campin, J.-M., A. Adcroft, C. Hill, and J. Marshall, 2003: Conservation of properties in a free-surface model. *Ocean Modelling*.

Cessi, P., 1996: Grid-scale instability of convective adjustment schemes. *Journal of Marine Research*, 54, 407–420.

Chaikin, P. M., and T. C. Lubensky, 1995: *Principles of condensed matter physics*. Cambridge University Press.

Chandrasekhar, S., 1961: *Hydrodynamic and Hydromagnetic Stability*. Dover Publications.

Chassignet, E. P., and J. Verron, 1998: *Ocean Modeling and Parameterization*. Kluwer Academic Publishers,, 451 pp.

Chavanis, P.-H., 2002: Statistical mechanics of two-dimensional vortices and stellar systems. *http://xxx.lanl.gov/abs/cond-mat/0212223*.

Chelton, D. B., R. A. deSzoeke, M. G. Schlax, K. E. Naggar, and N. Siwertz, 1998: Geographical variability of the first baroclinic rossby radius of deformation. *Journal of Physical Oceanography*, 28, 433–460.

Courant, R., and D. Hilbert, 1953: *Methods of Mathematical Physics Volume I*. Wiley-Interscience, 560 pp.

Courant, R., and D. Hilbert, 1962: *Methods of Mathematical Physics Volume II: Partial Differential Equations*. Interscience publishers, 830 pp.

Cox, M. D., 1984: *A primitive equation, 3-dimensional model of the ocean*. NOAA/Geophysical Fluid Dynamics Laboratory.

Cox, M. D., 1987: Isopycnal diffusion in a z-coordinate ocean model. *Ocean Modelling*, 74, 1–5.

Cushman-Roisin, B., 1994: *Introduction to Geophysical Fluid Dynamics*. 1st ed., Prentice-Hall.

Danabasoglu, G., and J. C. McWilliams, 1995: Sensitivity of the global ocean circulation to parameterizations of mesoscale tracer transports. *Journal of Climate*, 8, 2967–2987.

Danabasoglu, G., J. C. McWilliams, and W. G. Large, 1996: Approach to equilibrium in accelerated global oceanic models. *Journal of Climate*, 9, 1092–1110.

Davis, R. E., 1994a: Diapycnal mixing in the ocean: equations for large-scale budgets. *Journal of Physical Oceanography*, 24, 777–800.

Davis, R. E., 1994b: Diapycnal mixing in the ocean: the Osborn-Cox model. *Journal of Physical Oceanography*, 24, 2560–2576.

DeGroot, S. R., and P. Mazur, 1984: *Non-Equilibrium Thermodynamics*. Dover Publications.

deSzoeke, R. A., and A. F. Bennett, 1993: Microstructure fluxes across density surfaces. *Journal of Physical Oceanography*, 23, 2254–2264.

deSzoeke, R. A., and R. M. Samelson, 2002: The duality between the boussinesq and non-boussinesq hydrostatic equations of motion. *Journal of Physical Oceanography*, 32, 2194–2203.

Dewar, W. K., Y. Hsueh, T. J. McDougall, and D. Yuan, 1998: Calculation of pressure in ocean simulations. *Journal of Physical Oceanography*, 28, 577–588.

Döscher, R., and A. Beckmann, 1999: Effects of a bottom boundary layer parameterization in a coarse-resolution model of the north atlantic ocean. *Journal of Atmospheric and Oceanic Technology*, 17, 698–707.

Dukowicz, J. K., and J. R. Baumgardner, 2002: Efficient discretization of consistent friction operators on a B-grid in general orthogonal coordinates on a sphere. *in preparation*.

Dukowicz, J. K., and R. D. Smith, 1994: Implicit free-surface method for the Bryan-Cox-Semtner ocean model. *Journal of Geophysical Research*, 99, 7991–8014.

Dukowicz, J. K., and R. D. Smith, 1997: Stochastic theory of compressible turbulent fluid transport. *Physics of Fluids*, 9, 3523–3529.

Dukowicz, J. K., R. D. Smith, and R. C. Malone, 1993: A reformulation and implementation of the Bryan-Cox-Semtner ocean model on the connection machine. *Journal of Atmospheric and Oceanic Technology*, 10, 195–208.

Durran, D. R., 1999: *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Springer Verlag.

Favre, A., 1965: équations des gaz turbulents compressibles i and ii. *Journel de Méchanic*, 4, 361–421.

Feistel, R., 1993: Equilibrium thermodynamics of seawater revisited. *Progress in Oceanography*, 31, 101–179.

Feistel, R., and E. Hagen, 1995: On the Gibbs thermodynamic potential of seawater. *Progress in Oceanography*, 36, 249–327.

Fetter, A. L., and J. D. Walecka, 1980: *Theoretical Mechanics of Particles and Continua*. McGraw-Hill Book Company.

Flanders, H., 1989: *Differential Forms with Applications to the Physical Sciences*. Dover Publications.

Fofonoff, N. P., 1962: Physical properties of seawater. M. N. Hill, Ed., *The Sea*, Vol. 1, Wiley-Interscience, 3–30.

Fox-Kemper, B., R. Ferrari, and J. Pedlosky, 2003: A note on the indeterminacy of rotationaal and divergent eddy fluxes. *Journal of Physical Oceanography*.

Galperin, B., and S. A. Orszag, 1993: *Large Eddy Simulation of Complex Engineering and Geophysical Flows*. Cambridge University Press.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, 20, 150–155.

Gent, P. R., J. Willebrand, T. J. McDougall, and J. C. McWilliams, 1995: Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, 25, 463–474.

Gerdes, R., C. Köberle, and J. Willebrand, 1991: The influence of numerical advection schemes on the results of ocean general circulation models. *Climate Dynamics*, 5, 211–226.

Gill, A., 1982: *Atmosphere-Ocean Dynamics*. International Geophysics Series, Vol. 30, Academic Press Inc.

Goosens, M., F. Mittelbach, and A. Samarin, 1994: *The LaTeXCompanion*. Addison-Wesley.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, T. C. J. J. M. Gregory, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16, 147–168.

Greatbatch, R. J., 1994: A note on the representation of steric sea level in models that conserve volume rather than mass. *Journal of Geophysical Research*, 99, 12 767–12 771.

Greatbatch, R. J., 1998: Exploring the relationship between eddy-induced transport velocity, vertical momentum transfer, and the isopycnal flux of potential vorticity. *Journal of Physical Oceanography*, 28, 422–432.

Greatbatch, R. J., and K. G. Lamb, 1990: On parameterizing vertical mixing of momentum in non-eddy resolving ocean models. *Journal of Physical Oceanography*, 20, 1634–1637.

Greatbatch, R. J., and G. Li, 2000: Alongslope mean flow and an associated upslope bolus flux of tracer in a parameterization of mesoscale turbulence. *Deep-Sea Research I*, 47, 709–735.

Greatbatch, R. J., Y. Lu, and Y. Cai, 2001: Relaxing the Boussinesq approximation in ocean circulation models. *Journal of Atmospheric and Oceanic Technology*, 18, 1911–1923.

Greatbatch, R. J., and T. J. McDougall, 2002: The non-boussinesq temporal-residual-mean. *Journal of Physical Oceanography*, submitted.

Gregg, M. C., 1984: Entropy generation in the ocean by small-scale mixing. *Journal of Physical Oceanography*, 14, 688–711.

Griffies, S. M., 1998: The Gent-McWilliams skew-flux. *Journal of Physical Oceanography*, 28, 831–841.

Griffies, S. M., and Coauthors, 2000a: Developments in ocean climate modelling. *Ocean Modelling*, 2, 123–192.

Griffies, S. M., and K. Bryan, 1997: Predictability of the North Atlantic multidecadal climate variability. *Science*, 275, 181–184.

Griffies, S. M., A. Gnanadesikan, R. C. Pacanowski, V. Larichev, J. K. Dukowicz, and R. D. Smith, 1998: Isoneutral diffusion in a z-coordinate ocean model. *Journal of Physical Oceanography*, 28, 805–830.

Griffies, S. M., and R. W. Hallberg, 2000: Biharmonic friction with a Smagorinsky viscosity for use in large-scale eddy-permitting ocean models. *Monthly Weather Review*, 128, 2935–2946.

Griffies, S. M., M. J. Harrison, R. C. Pacanowski, and A. Rosati, 2003: *A Technical Guide to MOM4*. NOAA/Geophysical Fluid Dynamics Laboratory.

Griffies, S. M., R. C. Pacanowski, and R. W. Hallberg, 2000b: Spurious diapycnal mixing associated with advection in a z-coordinate ocean model. *Monthly Weather Review*, 128, 538–564.

Griffies, S. M., R. C. Pacanowski, R. M. Schmidt, and V. Balaji, 2001: Tracer conservation with an explicit free surface method for z-coordinate ocean models. *Monthly Weather Review*, 129, 1081–1098.

Haidvogel, D. B., and A. Beckmann, 1999: *Numerical Ocean Circulation Modeling*. Imperial College Press, 318 pp.

Haine, T. W. N., and J. Marshall, 1998: Gravitational, symmetric, and baroclinic instability of the ocean mixed layer. *Journal of Physical Oceanography*, 28, 634–658.

Hallberg, R. W., 2000: Time integration of diapycnal diffusion and richardson number-dependent mixing in isopycnal coordinate ocean models. *Monthly Weather Review*, 128, 1402–1419.

Haltiner, G. T., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology*. John Wiley and Sons.

Haynes, P. H., and M. E. McIntyre, 1987: On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *Journal of Atmospheric Sciences*, 44, 828–841.

Haynes, P. H., and M. E. McIntyre, 1990: On the conservaiton and impermeability theorems for potential vorticity. *Journal of Atmospheric Sciences*, 47, 2021–2031.

Held, I. M., and V. D. Larichev, 1996: A scaling theory for horizontally homogeneous baroclinically unstable flow on a beta plane. *Journal of Atmospheric Sciences*, 53, 946–952.

Held, I. M., and T. Schneider, 1999: The surface branch of the zonally averaged mass transport circulation in the troposphere. *Journal of Atmospheric Sciences*, 56, 1688–1697.

Hinze, J. O., 1975: *Turbulence*. McGraw-Hill Publishers.

Holloway, G., 1989: Subgridscale representation. J. W. David L.T. Anderson, Ed., *Oceanic circulation models: combining data and dynamics*, Vol. 44, Kluwer Academic Publishers, 513–593.

Holloway, G., 1992: Representing topographic stress for large-scale ocean models. *Journal of Physical Oceanography*, 22, 1033–1046.

Holloway, G., 1997: Eddy transport of thickness and momentum in layer and level models. *Journal of Physical Oceanography*, 27, 1153–1157.

Holton, J. R., 1992: *An Introduction to Dynamic Meteorology*. Academic Press.

Huang, K., 1987: *Statistical Mechanics*. John Wiley and Sons.

Huang, R. X., 1993: Real freshwater flux as a natural boundary condition for the salinity balance and thermohaline circulation forced by evaporation and precipitation. *Journal of Physical Oceanography*, 23, 2428–2446.

Hughes, C. W., C. Wunsch, and V. Zlotnicki, 2000: Satellite peers through the oceans from space. *EOS*, 81, 68.

Jackson, J. D., 1975: *Classical Electrodynamics*. John Wiley and Sons.

Janjić, Z. I., 1977: Pressure gradient force and advection scheme used for forecasting with steep and small scale topography. *Contributions in Atmospheric Physics*, 50, 186–199.

Jaynes, E. T., 1957a: Information theory and statistical mechanics. *Physical Review*, 106, 620–630.

Jaynes, E. T., 1957b: Information theory and statistical mechanics. ii. *Physical Review*, 108, 171–190.

Kantha, L. H., and C. A. Clayson, 2000a: *Numerical Models of Oceans and Oceanic Processes*. Academic Press.

Kantha, L. H., and C. A. Clayson, 2000b: *Small Scale Processes in Geophysical Fluid Flows*. Academic Press.

Killworth, P. D., 1997: On the parameterization of eddy transfer part I: Theory. *Journal of Marine Research*, 55, 1171–1197.

Killworth, P. D., 2001: Boundary conditions on the quasi-Stokes velocities in parameterizations. *Journal of Physical Oceanography*, 31, 1132–1155.

Killworth, P. D., and N. Edwards, 1999: A turbulent bottom boundary layer code for use in numerical ocean models. *Journal of Physical Oceanography*, 29, 1221–1238.

Killworth, P. D., J. M. Smith, and A. E. Gill, 1984: Speeding up ocean circulation models. *Ocean Modelling*, 56, 1–5.

Killworth, P. D., D. Stainforth, D. J. Webb, and S. M. Paterson, 1991: The development of a free-surface Bryan-Cox-Semtner ocean mode. *Journal of Physical Oceanography*, 21, 1333–1348.

Kraichnan, R. H., and D. Montgomery, 1980: Two-dimensional turbulence. *Reports on Progress in Physics*, 43, 547–619.

Kunze, E., and T. B. Sanford, 1996: Abyssal mixing: where is is not. *Journal of Physical Oceanography*, 26, 2286–2296.

Kushner, P. J., and I. M. Held, 1999: Potential vorticity thickness fluxes and wave-mean flow interaction. *Journal of Atmospheric Sciences*, 56, 948–958.

Lamport, L., 1994: *LaTeX: A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley.

Landau, L. D., and E. M. Lifshitz, 1986: *Theory of Elasticity*. Pergamon Press.

Landau, L. D., and E. M. Lifshitz, 1987: *Fluid Mechanics*. Pergamon Press.

Large, W. G., 1998: Modeling and parameterizing the ocean planetary boundary layer. E. P. Chassignet and J. Verron, Eds., *Ocean Modeling and Parameterization*, NATO ASI Mathematical and Physical Sciences Series., Vol. 516, Kluwer, 81–120.

Large, W. G., G. Danabasoglu, S. C. Doney, and J. C. McWilliams, 1997: Sensitivity to surface forcing and boundary layer mixing in a global ocean model: annual-mean climatology. *Journal of Physical Oceanography*, 27, 2418–2447.

Large, W. G., G. Danabasoglu, J. C. McWilliams, P. R. Gent, and F. O. Bryan, 2001: Equatorial circulation of a global ocean climate model with anisotropic horizontal viscosity. *Journal of Physical Oceanography*, 31, 518–536.

Large, W. G., J. C. McWilliams, and S. C. Doney, 1994: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Reviews of Geophysics*, 32, 363–403.

Lau, N.-C., and J. M. Wallace, 1979: On the distribution of horizontal transports by transient eddies in the north hemisphere wintertime circulation. *Journal of Atmospheric Sciences*, 36, 1844–1861.

Ledwell, J. R., A. J. Watson, and C. S. Law, 1993: Evidence for slow mixing acroos the pycnocline from an open-ocean tracer-release experiment. *Nature*, 364, 701–703.

Leith, C. E., 1968: Diffusion approximation for two-dimensional turbulence. *Physics of Fluids*, 10, 1409–1416.

Leith, C. E., 1996: Stochastic models of chaotic systems. *Physica D*, 98, 481–491.

Lu, Y., 2001: Including non-Boussinesq effects in Boussinesq ocean circulation models. *Journal of Physical Oceanography*, 31, 1616–1622.

Madec, G., and M. Imbard, 1996: A global ocean mesh to overcome the north pole singularity. *CD*, 12, 381–388.

Marion, J. B., and S. T. Thornton, 1988: *Classical Dynamics of Particles and Systems*. Harcourt Brace Jovanovich.

Marshall, J., C. Hill, L. Perelman, and A. Adcroft, 1997: Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling. *Journal of Geophysical Research*, 102, 5733–5752.

Marshall, J., and G. Shutts, 1981: A note on rotational and divergent eddy fluxes. *Journal of Physical Oceanography*, 11, 1677–1680.

Marshall, J., and G. Shutts, 1999: Open-ocean convection: observations, theory, and models. *Reviews of Geophysics*, 37, 1–64.

McDougall, T. J., 1987a: Neutral surfaces. *Journal of Physical Oceanography*, 17, 1950–1967.

McDougall, T. J., 1987b: Thermobaricity, cabbeling, and water-mass conversion. *Journal of Geophysical Research*, 92, 5448–5464.

McDougall, T. J., 1991: Parameterizing mixing in inverse models. P. Müller and D. Henderson, Eds., *Dynamics of Oceanic Internal Gravity Waves: Proceedings of the sixth 'Aha Huliko 'a Hawaiian Winter Workshop*, 'Aha Huiliko'a, University of Hawaii at Manoa, 355–386.

McDougall, T. J., 1995: The influence of ocean mixing on the absolute velocity vector. *Journal of Physical Oceanography*, 25, 705–725.

McDougall, T. J., 2002: Potential enthalpy: a conservative oceanic variable for evaluating heat content and ocean heat fluxes. *Journal of Physical Oceanography*. Submitted.

McDougall, T. J., and J. A. Church, 1986: Pitfalls with numerical representations of isopycnal and diapycnal mixing. *Journal of Physical Oceanography*, 16, 196–199.

McDougall, T. J., and R. Feistel, 2002: What causes the adiabatic lapse rate. *Deep-Sea Research*. Submitted.

McDougall, T. J., and C. J. R. Garrett, 1992: Scalar conservation equations in a turbulent ocean. *Deep-Sea Research*, 11/12, 1953–1966.

McDougall, T. J., R. J. Greatbatch, and Y. Lu, 2002a: On conservation equations in oceanography: How accurate are Boussinesq ocean models? *Journal of Physical Oceanography*, 32, 1574–1584.

McDougall, T. J., and D. R. Jackett, 1988: On the helical nature of neutral trajectories in the ocean. *Progress in Oceanography*, 20, 153–183.

McDougall, T. J., D. R. Jackett, D. G. Wright, and R. Feistel, 2002b: Accurate and computationally efficient algorithms for potential temperature and density of seawater. *Journal of Atmospheric and Oceanic Technology*. Submitted.

McDougall, T. J., and P. C. McIntosh, 2001: The temporal-residual-mean velocity. part ii: isopycnal interpretation and the tracer and momentum equations. *Journal of Physical Oceanography*, 31, 1222–1246.

McIntosh, P. C., and T. J. McDougall, 1996: Isopycnal averaging and the residual mean circulation. *Journal of Physical Oceanography*, 26, 1655–1660.

Mellor, G. L., and T. Ezer, 1995: Sea level variations induced by heating and cooling: an evaluation of the boussinesq approximation in ocean models. *Journal of Geophysical Research*, 100, 20 565–20 577.

Messinger, F., 1973: A method for construction of second-order accurate difference schemes permitting no false two-grid-interval waves in the height field. *Tellus*, 25, 444–457.

Middleton, J. F., and J. W. Loder, 1989: Skew fluxes in polarized wave fields. *Journal of Physical Oceanography*, 19, 68–76.

Moritz, H., 2000: Geodetic reference system 1980. *Journal of Geodesy*, 74, 128–133.

Morse, P. M., and H. Feshbach, 1953: *Methods of Theoretical Physics Part I and II*. McGraw-Hill Book Company.

Müller, P., 1995: Ertel's potential vorticity theorem in physical oceanography. *Reviews of Geophysics*, 33, 67–97.

Munk, W. H., 1950: On the wind-driven ocean circulation. *J. Meteor.*, 7, 79–93.

Murray, R. J., 1996: Explicit generation of orthogonal grids for ocean models. *Journal Computational Physics*, 126, 251–273.

Murray, R. J., and C. J. C. Reason, 2002: A curvilinear version of the Bryan-Cox ocean model. *Journal Computational Physics*, 171, 1–46.

Noble, B., and J. W. Daniel, 1977: *Applied Linear Algebra*. Prentice-Hall.

O'Brien, J. J., 1986: *Advanced Physical Oceanographic Numerical Modelling*. D. Reidel Publishing Company,, 608 pp.

Olbers, D. J., M. Wenzel, and J. Willebrand, 1985: The inference of north atlantic circulation patterns from climatological hydrographic data. *Reviews of Geophysics*, 23, 313–356.

Osborn, T. R., and C. S. Cox, 1972: Oceanic fine structure. *Geophysical Fluid Dynamics*, 3, 321–345.

Pacanowski, R. C., 1995: *MOM2 Documentation, User's Guide, and Reference Manual*. NOAA/Geophysical Fluid Dynamics Laboratory.

Pacanowski, R. C., K. Dixon, and A. Rosati, 1991: *The GFDL Modular Ocean Model User Guide*. NOAA/Geophysical Fluid Dynamics Laboratory.

Pacanowski, R. C., and A. Gnanadesikan, 1998: Transient response in a z-level ocean model that resolves topography with partial-cells. *Monthly Weather Review*, 126, 3248–3270.

Pacanowski, R. C., and S. M. Griffies, 1999: *The MOM3 Manual*. NOAA/Geophysical Fluid Dynamics Laboratory.

Pacanowski, R. C., and G. Philander, 1981: Parametrization of vertical mixing in numerical models of the tropical ocean. *Journal of Physical Oceanography*, 11, 1442–1451.

Panton, R. L., 1996: *Incompressible Flow*. 2nd ed., John Wiley and Sons, 837 pp.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. 2nd ed., Springer-Verlag.

Pedlosky, J., 1996: *Ocean Circulation Theory*. Springer-Verlag.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*. American Institute of Physics.

Phillips, N. A., 1949: An example of a non-linear computational instability. *The Atmosphere and the Sea in Motion: Rossby Memorial Volume*, Rockefeller Institute Press, 501–504.

Pickard, G. L., and W. J. Emery, 1990: *Descriptive Physical Oceanography*. 5th ed., Pergamon Press, 320 pp.

Pinardi, N., A. Rosati, and R. C. Pacanowski, 1995: The sea surface pressure formulation of rigid lid models. implications for altimetric data assimilation studies. *Journal of Marine Systems*, 6, 109–119.

Plumb, R. A., 1979: Eddy fluxes of conserved quantities by small-amplitude waves. *Journal of Atmospheric Sciences*, 36, 1699–1704.

Plumb, R. A., and J. D. Mahlman, 1987: The zonally averaged transport characteristics of the gfdl general circulation/transport model. *Journal of Atmospheric Sciences*, 44, 298–327.

Polzin, K. L., K. Speer, J. M. Toole, and R. W. Schmitt, 1996: Intense mixing of antarcic bottom water in the equatorial atlantic. *Nature*, 380, 54–57.

Polzin, K. L., J. M. Toole, J. R. Ledwell, and R. W. Schmitt, 1997: Spatial variability of turbulent mixing in the abyssal ocean. *Science*, 276, 93–96.

Price, J. F., and M. O. Baringer, 1994: Outflows and deep water production by marginal seas. *Progress in Oceanography*, 33, 161–200.

Rahmstorf, S., 1993: A fast and complete convection scheme for ocean models. *Ocean Modelling*, 101, 9–11.

Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate. *Journal of Physical Oceanography*, 12, 1154–1158.

Reichl, L. E., 1987: *A Modern Course in Statistical Physics*. John Wiley and Sons.

Reif, F., 1965: *Fundamentals of statistical and thermal physics*. McGraw-Hill.

Rhines, P. B., 1982: Basic dynamics of the large-scale geostrophic circulation. *WHOI 1982 Summer Study Program*, Woods Hole Oceanographic Institute.

Roberts, M. J., and D. Marshall, 1998: Do we require adiabatic dissipation schemes in eddy-resolving ocean models? *Journal of Physical Oceanography*, 28, 2050–2063.

Roberts, M. J., and D. Marshall, 2000: On the validity of downgradient eddy closures in ocean models. *Journal of Geophysical Research*, 105, 28 613–28 628.

Rosati, A., and K. Miyakoda, 1988: A general circulation model for upper ocean simulation. *Journal of Physical Oceanography*, 18, 1601–1626.

Saffman, P. G., 1992: *Vortex Dynamics*. 1st ed., Cambridge University Press, 311 pp.

Salmon, R., 1988: Hamiltonian fluid mechanics. *Annual Review of Fluid Mechanics*, 20, 225–256.

Salmon, R., 1998: *Lectures on Geophysical Fluid Dynamics*. Oxford University Press.

Schmitt, R., 1998: Double-diffusive convection. E. P. Chassignet and J. Verron, Eds., *Ocean Modeling and Parameterization*, NATO ASI Mathematical and Physical Sciences Series., Vol. 516, Kluwer, 215–234.

Schmitt, R. W., 1994: Double diffusion in oceanography. *Annual Review of Fluid Mechanics*, 26, 255–285.

Schutz, B. F., 1980: *Geometrical Methods of Mathematical Physics*. Cambridge University Press.

Schutz, B. F., 1985: *A First Course in General Relativity*. Cambridge University Press.

Semtner, A. J., and Y. Mintz, 1977: Numerical simulation of the gulf stream and mid-ocean eddies. *Journal of Physical Oceanography*, 7, 208–230.

Semtner, A. J. J., 1974: An oceanic general circulation model with bottom topography. *Numerical Simulation of Weather and Climate*, Technical Report No. 9, UCLA Department of Meteorology.

Send, U., and R. Käse, 1998: Parameterization of processes in deep convection regimes. E. P. Chassignet and J. Verron, Eds., *Ocean Modeling and Parameterization*, NATO ASI Mathematical and Physical Sciences Series., Vol. 516, Kluwer, 191–214.

Send, U., and J. Marshall, 1995: Integral effects of deep convection. *Journal of Physical Oceanography*, 25, 855–872.

Siedler, G., J. Gould, and J. Church, 2001: *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*. Academic Press.

Smagorinsky, J., 1963: General circulation experiments with the primitive equations: I. the basic experiment. *Monthly Weather Review*, 91, 99–164.

Smagorinsky, J., 1993: Some historical remarks on the use of nonlinear viscosities. B. Galperin and S. A. Orszag, Eds., *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, Cambridge University Press, 3–36.

Smith, K. S., and G. K. Vallis, 2002: The scales and equilibration of midocean eddies: Forced-dissipative flow. *Journal of Physical Oceanography*, 32, 1699–1721.

Smith, R. D., 1999: The primitive equations in the stochastic theory of adiabatic stratified turbulence. *Journal of Physical Oceanography*, 29, 1865–1880.

Smith, R. D., J. K. Dukowicz, and R. C. Malone, 1992: Parallel ocean general circulation modeling. *Physica D*, 60, 38–61.

Smith, R. D., S. Kortas, and B. Meltz, 1995: Curvilinear coordinates for global ocean models. *Los Alamos preprint*, LA-UR-95-1146.

Smith, R. D., and J. C. McWilliams, 2002: Anisotropic horizonal viscosity for ocean models. *Ocean Modelling*, 5, 129–156.

Solomon, H., 1971: On the representation of isentropic mixing in ocean models. *Journal of Physical Oceanography*, 1, 233–234.

Stammer, D., 1997a: Global characteristics of ocean variability estimated from regional TOPEX/POSEIDON altimeter measurements. *Journal of Physical Oceanography*, 27, 1743–1769.

Stammer, D., 1997b: Global characteristics of ocean variability estimated from regional topex/poseidon altimeter measurements. *Journal of Physical Oceanography*, 27, 1743–1769.

Starr, V. P., 1945: A quasi-lagrangian system of hydrodynamical equations. *J. Meteor.*, 2, 227–237.

Stone, P., 1972: A simplified radiative-dynamical model for the static stability of rotating atmospheres. *Journal of Atmospheric Sciences*, 29, 405–418.

Sun, S., R. Bleck, C. Rooth, J. Dukowicz, E. Chassignet, and P. D. Killworth, 1999: Inclusion of thermobaricity in isopycnic-coordinate ocean models. *Journal of Physical Oceanography*, 29, 2719–2729.

Tomczak, M., and J. S. Godfrey, 1994: *Regional Oceanography: An Introduction*. Pergamon Press.

Toole, J., 1998: Turbulent mixing in the ocean. E. P. Chassignet and J. Verron, Eds., *Ocean Modeling and Parameterization*, NATO ASI Mathematical and Physical Sciences Series, Vol. 516, Kluwer Academic Publishers, 171–190.

Toole, J. M., and T. J. McDougall, 2001: Mixing and stirring in the ocean interior. G. Siedler, J. Gould, and J. Church, Eds., *Ocean Circulation and Climate: Observing and Modelling the Global Ocean*, Academic Press, 337–356.

Toole, J. M., K. L. Polzin, and R. W. Schmitt, 1994: Estimates of diapycnal mixing in the abyssal ocean. *Science*, 264, 1120–1123.

Toole, J. M., R. W. Schmitt, and K. L. Polzin, 1997: Near-boundary mixing above the flanks of a mid-latitude seamount. *Journal of Geophysical Research*, 102, 947–959.

Treguier, A. M., 1999: Evaluating eddy mixing coefficients from eddy-resolving ocean models: A case study. *Journal of Marine Research*, 57, 89–108.

Treguier, A. M., I. M. Held, and V. D. Larichev, 1997: On the parameterization of quasi-geostrophic eddies in primitive equation ocean models. *Journal of Physical Oceanography*, 27, 567–580.

Trenberth, K. E., 1992: Climate system modeling. K. E. Trenberth, Ed., *Climate System Modeling*, Cambridge University Press, 788.

Veronis, G., 1973: Large scale ocean circulation. *Advances in Applied Mechanics*, 13, 2–92.

Veronis, G., 1975: The role of models in tracer studies. *Numerical Models of Ocean Circulation*, National Academy of Sciences.

Visbeck, M., J. C. Marshall, T. Haine, and M. Spall, 1997: Specification of eddy transfer coefficients in coarse resolution ocean circulation models. *Journal of Physical Oceanography*, 27, 381–402.

Visbeck, M., J. C. Marshall, and H. Jones, 1996: On the dynamics of convective "chimneys" in the ocean. *Journal of Physical Oceanography*, 26, 1721–1734.

Wajsowicz, R. C., 1993: A consistent formulation of the anisotropic stress tensor for use in models of the large-scale ocean circulation. *Journal Computational Physics*, 105, 333–338.

Weaver, A. J., and M. Eby, 1997: On the numerical implementation of advection schemes for use in conjunction with various mixing parameterizations in the gfdl ocean model. *Journal of Physical Oceanography*, 27, 369–377.

Weinberg, S., 1972: *Gravitation and Cosmology*. John Wiley and Sons.

Williams, G. P., 1972: Friction term formulation and convective instability in a shallow atmosphere. *Journal of Atmospheric Sciences*, 29, 870–876.

Wright, D. K., 1997: A new eddy mixing parameterization in an ocean general circulation model. *International WOCE Newsletter*, 26, 27–29.

Wunsch, C., and V. Zlotnicki, 1999: Measuring ocean bottom pressure changes from space. report of a meeting. `http://puddle.mit.edu/~cwunsch/lndmtgwithfig/lndmtgwithfig.html`.

INDEX

Adcroft et al. (1997), 116, 117, 210, 220, 308, 431, 439
Andrews and McIntyre (1978), 171, 174, 439
Apel (1987), 5, 34, 85, 105, 225, 439
Aris (1962), 42, 53–55, 254, 333–335, 348, 396, 426, 428, 437, 439
Bacon and Fofonoff (1996), 106, 439
Batchelor (1967), 27, 103, 426, 439
Beckers et al. (2000), 306, 439
Beckmann and Döscher (1997), 294, 306, 439
Beckmann (1998), 193, 198, 439
Bleck and Boudra (1981), 368, 439
Bleck et al. (1992), 368, 439
Bleck (1978), 114, 122, 123, 126, 134, 439
Bleck (1998), 131, 198, 439
Blumberg and Mellor (1987), 368, 440
Boer and Shepherd (1983), 368, 440
Bokhove (2000), 94, 440
Bretherton (1966), 295, 440
Brown and Campana (1978), 234, 440
Bryan and Cox (1972), 108, 110, 226, 440
Bryan and Lewis (1979), 19, 440
Bryan et al. (1975), 360, 363, 440
Bryan et al. (1999), 181, 267, 440
Bryan (1969), 19, 37, 111, 115, 220, 228, 229, 235, 236, 350, 360, 376, 377, 387, 389, 440
Bryan (1984), 234, 440
Bryan (1989), 5, 440
Callen (1985), 88, 125, 440
Campin et al. (2003), 218, 228, 229, 236, 440
Cessi (1996), 290, 440
Chaikin and Lubensky (1995), 122, 440
Chandrasekhar (1961), 110, 441
Chassignet and Verron (1998), 5, 191, 441
Chavanis (2002), 189, 441
Chelton et al. (1998), 267–269, 441
Courant and Hilbert (1953), 301, 441
Courant and Hilbert (1962), 301, 441
Cox (1984), 19, 61, 115, 226, 228, 229, 377, 441
Cox (1987), 255, 258, 284, 285, 303, 305, 306, 326, 441
Cushman-Roisin (1994), 5, 441
Danabasoglu and McWilliams (1995), 287, 293, 324, 441
Danabasoglu et al. (1996), 234, 441
Davis (1994a), 98, 103, 194, 441
Davis (1994b), 194, 441
DeGroot and Mazur (1984), 27, 103, 441
Dewar et al. (1998), 111, 226, 441
Dukowicz and Baumgardner (2002), 350, 442
Dukowicz and Smith (1994), 228, 236, 442
Dukowicz and Smith (1997), 265, 442
Dukowicz et al. (1993), 235, 442
Durran (1999), 5, 35, 111, 162, 188, 218, 225, 233, 442
Favre (1965), 144, 442
Feistel and Hagen (1995), 108, 442
Feistel (1993), 108, 442
Fetter and Walecka (1980), 60, 74, 76, 442
Flanders (1989), 397, 442
Fofonoff (1962), 98, 106, 442
Fox-Kemper et al. (2003), 182, 442

Galperin and Orszag (1993), 367, 442
Gent and McWilliams (1990), 39, 96, 103, 110, 133, 136, 183–185, 242, 252, 255, 261, 266, 300, 305, 323, 442
Gent et al. (1995), 39, 183, 184, 242, 252, 263, 265, 295, 325, 442
Gerdes et al. (1991), 286, 287, 442
Gill (1982), 5, 17, 34, 35, 43, 51, 59, 61, 65, 66, 71, 73, 190, 225, 267, 363, 443
Goosens et al. (1994), 2, 443
Gordon et al. (2000), 265–267, 443
Greatbatch and Lamb (1990), 242, 443
Greatbatch and Li (2000), 295, 443
Greatbatch and McDougall (2002), 64, 146, 165, 243, 443
Greatbatch et al. (2001), 36, 52, 111, 161, 207, 225, 376, 443
Greatbatch (1994), 36, 52, 443
Greatbatch (1998), 242, 443
Gregg (1984), 98, 103, 104, 443
Griffies and Bryan (1997), 145, 443
Griffies and Hallberg (2000), 271, 273, 274, 333, 341, 342, 350, 354, 365, 367, 368, 375, 376, 443
Griffies et al. (1998), 39, 110, 242, 250, 252, 258, 265, 277, 279, 284, 303, 305, 310, 326, 350, 443
Griffies et al. (2000a), 19, 37, 64, 115, 117, 165, 242, 261, 295, 443
Griffies et al. (2000b), 165, 198, 242, 363, 364, 444
Griffies et al. (2001), 37, 72, 204, 205, 208, 218, 226, 229, 232, 235, 236, 364, 444
Griffies et al. (2003), 39, 203, 207, 220, 221, 225, 228, 229, 232, 233, 315, 360, 377, 444
Griffies (1998), 103, 183, 242, 252, 261, 263, 264, 279, 300, 443
Haidvogel and Beckmann (1999), 5, 444
Haine and Marshall (1998), 289, 292, 295, 444
Hallberg (2000), 197, 444
Haltiner and Williams (1980), 5, 35, 133, 162, 232, 306, 361, 362, 444
Haynes and McIntyre (1987), 68, 444
Haynes and McIntyre (1990), 68, 128, 444
Held and Larichev (1996), 265–267, 444
Held and Schneider (1999), 169, 291, 292, 444
Hinze (1975), 144, 444
Holloway (1989), 191, 444
Holloway (1992), 101, 189, 224, 444
Holloway (1997), 242, 444
Holton (1992), 79, 444
Huang (1987), 27, 86, 445
Huang (1993), 37, 213, 235, 445
Hughes et al. (2000), 36, 445
Jackson (1975), 61, 166, 445
Janjić (1977), 364, 445
Jaynes (1957a), 86, 189, 445
Jaynes (1957b), 86, 445
Kantha and Clayson (2000a), 5, 445
Kantha and Clayson (2000b), 5, 191–193, 445
Killworth and Edwards (1999), 224, 445
Killworth et al. (1984), 234, 445
Killworth et al. (1991), 71, 218, 228, 364, 445
Killworth (1997), 185, 265, 295, 445
Killworth (2001), 295, 445
Kraichnan and Montgomery (1980), 189, 445
Kunze and Sanford (1996), 195, 445
Kushner and Held (1999), 169–171, 445
Lamport (1994), 2, 445
Landau and Lifshitz (1986), 333, 446
Landau and Lifshitz (1987), 42, 98, 103, 446
Large et al. (1994), 195, 213, 234, 289, 446
Large et al. (1997), 291–294, 446
Large et al. (2001), 333, 344, 345, 446
Large (1998), 192, 446
Lau and Wallace (1979), 181, 446
Ledwell et al. (1993), 84, 195, 446
Leith (1968), 367, 446
Leith (1996), 367, 446
Lu (2001), 144, 446
Müller (1995), 66, 74, 448
Madec and Imbard (1996), 398, 411, 446

Marion and Thornton (1988), 59, 74, 352, 446
Marshall and Shutts (1981), 181, 265, 446
Marshall and Shutts (1999), 192, 446
Marshall et al. (1997), 34, 35, 45, 133, 446
McDougall and Church (1986), 39, 242, 447
McDougall and Feistel (2002), 92, 447
McDougall and Garrett (1992), 144, 447
McDougall and Jackett (1988), 248, 447
McDougall and McIntosh (2001), 164, 171, 172, 174, 177, 183, 185, 243, 265, 290–292, 447
McDougall et al. (2002a), 64, 144, 145, 155, 161, 447
McDougall et al. (2002b), 108, 197, 447
McDougall (1987a), 39, 102, 110, 190, 197, 242, 247, 248, 446
McDougall (1987b), 102, 192, 197, 259, 260, 447
McDougall (1991), 259, 447
McDougall (1995), 114, 122, 130, 447
McDougall (2002), 87, 95, 106–108, 447
McIntosh and McDougall (1996), 151, 164, 171, 183, 243, 447
Mellor and Ezer (1995), 36, 52, 447
Messinger (1973), 364, 447
Middleton and Loder (1989), 178, 447
Moritz (2000), 61, 448
Morse and Feshbach (1953), 397, 426, 448
Munk (1950), 363, 448
Murray and Reason (2002), 376, 377, 448
Murray (1996), 397, 398, 411, 448
Noble and Daniel (1977), 409, 448
O'Brien (1986), 5, 448
Olbers et al. (1985), 39, 242, 250, 448
Osborn and Cox (1972), 144, 448
Pacanowski and Gnanadesikan (1998), 116, 117, 210, 220, 308, 310, 431, 448
Pacanowski and Griffies (1999), 205, 207, 209, 228, 233, 267, 269, 305, 310, 333, 342, 364, 377, 448
Pacanowski and Philander (1981), 234, 448
Pacanowski et al. (1991), 226, 377, 386, 448
Pacanowski (1995), 226, 377, 448
Panton (1996), 182, 448
Pedlosky (1987), 5, 43, 66, 67, 70, 448
Pedlosky (1996), 5, 128, 448
Peixoto and Oort (1992), 5, 145, 448
Phillips (1949), 360, 448
Pickard and Emery (1990), 5, 449
Pinardi et al. (1995), 235, 449
Plumb and Mahlman (1987), 178, 449
Plumb (1979), 174, 178, 449
Polzin et al. (1996), 195, 449
Polzin et al. (1997), 192, 195, 449
Price and Baringer (1994), 193, 449
Rahmstorf (1993), 290, 449
Redi (1982), 39, 178, 180, 242, 249, 250, 252–255, 275, 305, 449
Reichl (1987), 86, 449
Reif (1965), 189, 449
Rhines (1982), 170, 449
Roberts and Marshall (1998), 39, 165, 242, 276, 278, 449
Roberts and Marshall (2000), 183, 449
Rosati and Miyakoda (1988), 365, 368, 369, 377, 449
Saffman (1992), 182, 449
Salmon (1988), 74, 449
Salmon (1998), 5, 27, 66, 74, 78, 189, 336, 338, 449
Schmitt (1994), 193, 450
Schmitt (1998), 193, 450
Schutz (1980), 32, 396, 412, 450
Schutz (1985), 396, 419, 450
Semtner and Mintz (1977), 366, 450
Semtner (1974), 228, 229, 450
Send and Marshall (1995), 289, 292, 295, 450
Send and Käse (1998), 192, 450
Siedler et al. (2001), 5, 450
Smagorinsky (1963), 350, 367–370, 377, 450
Smagorinsky (1993), 271, 333, 338, 350, 367, 368, 377, 450
Smith and McWilliams (2002), 333, 337, 344–346, 349, 350, 369, 372, 373,

450
Smith and Vallis (2002), 164, 185, 265, 290, 450
Smith et al. (1992), 235, 450
Smith et al. (1995), 376, 377, 398, 411, 450
Smith (1999), 144, 177, 450
Solomon (1971), 39, 242, 275, 450
Stammer (1997a), 164, 451
Stammer (1997b), 267, 451
Starr (1945), 114, 134, 451
Stone (1972), 267, 451
Sun et al. (1999), 110, 197, 451
Tomczak and Godfrey (1994), 5, 451
Toole and McDougall (2001), 191, 193, 451
Toole et al. (1994), 195, 451
Toole et al. (1997), 195, 451
Toole (1998), 192, 451
Treguier et al. (1997), 183, 265, 266, 269, 272, 289, 291–293, 295, 325, 451
Treguier (1999), 265, 451
Trenberth (1992), 5, 451
Veronis (1973), 42, 371, 451
Veronis (1975), 371, 451
Visbeck et al. (1996), 289, 292, 295, 452
Visbeck et al. (1997), 251, 265–267, 269–271, 451
Wajsowicz (1993), 343, 350, 376, 377, 387–389, 452
Weaver and Eby (1997), 324, 452
Weinberg (1972), 428, 437, 452
Williams (1972), 332, 452
Wright (1997), 266, 452
Wunsch and Zlotnicki (1999), 36, 452
deSzoeke and Bennett (1993), 141, 144, 145, 171, 172, 194, 441
deSzoeke and Samelson (2002), 133, 441
Döscher and Beckmann (1999), 294, 306, 442

acoustic modes, 109
  Boussinesq approximation, 35
  eliminating in model, 225–226
  hydrostatic approximation, 34
adiabatic, 103–104
advection, 31
advection velocity components
  introduction, 207, 233
angular momentum, 61–62, 73–81, 333, 336–337, 339, 343–344
anisotropic viscosity, 372–373
arrow notation, 398–399

B-grid computational modes, 364
background viscosity, 371–372
baroclinic instability, 25–27
baroclinic mode, 71
baroclinic pressure, 71–72
baroclinic zone, 267
baroclinic/barotropic split, 229–230
baroclinicity operator, 229
barotropic gravity waves, 226–227
barotropic mode, 71
barotropic mode time filtering, 232
barotropic streamfunction, 235–236
basis vectors
  cartesian, 399–400
  local isopycnal, 120–122, 400
  local surface, 120–122
  spherical, 399–400
Beta parameter, 363–364
big leap-frog, 205, 231–232
biharmonic diffusivity, 274–275
biharmonic tracer flux, 274–275
biharmonic tracer operators, 272–281
  Gent's operator, 278–279
  horizontal mixing, 274–275
  iterative skewsion, 281
  neutral filtering, 279–280
  neutral mixing, 275–276
  Roberts and Marshall, 276–278
  skew filtering, 281
biharmonic viscosity, 354–356
boldface notation, 53–54
bolus velocity, 169–170
bottom boundary layer, 114–115, 193
bottom kinematic boundary condition, 48
bottom pressure, 35–36, 65–66
boundary layer, 114–115
boundary layer model, 211
boundary waves, 25–27
Boussinesq approximation
  basic elements, 35–36
  for ensemble mean fields, 155

Brownian particle, 189
buoyant mixing, 191–192

cabbeling, 192, 260, 274–275
cabbeling(, 259
canonical momentum, 75–76
Cartesian tensors, 434
chain rule, 401
Christoffel symbols, 420–421, 423, 435
closure problems, 189
co-lo mnemonic, 407, 434
comma notation, 400, 434
commutator, 44, 430, 438
compatibility between mass and tracer, 85–86, 204–205
conservation of indices, 433–434
conservative temperature, 106–108
continuity equations, 45–46
continuum hypothesis, 27–28
contravariant tensor labels, 434
convection, 25–27
coordinate singularity at pole, 397–398
coordinates, 42–44
  cartesian and spherical, 43, 398, 402
  discussion, 398–400
  general orthogonal, 397–398, 410–411, 426–433
    infinitesimal distances, 43–44
    key formulae, 43–44
    metric tensor, 43–44
    partial derivatives, 43–44
Coriolis acceleration, 61–62
Coriolis parameter, 62
Courant-Friedricks-Levy condition, 35, 362
covariance of tensor equations, 418–419
covariant derivatives, 419–422, 434–435
  curl of a vector, 425
  divergence of a second order tensor, 424–425
  divergence of a vector, 423–424
  laplacian of a scalar, 425
  laplacian of a vector, 425–426
  of one-forms, 421
  of second order tensors, 422
  of the metric tensor, 422
  of vectors, 420–421
covariant tensor labels, 434

deformation rates, 342–343, 347–348, 357, 368
density
  expansion coefficients, 109, 248
  idealized equation of state, 110
  in situ, 108–111
  linear equation of state, 110
  modified mean, 151
  potential, 108–111
  quasi-non-Boussinesq, 110–111, 226
  reference for Boussinesq, 51
density reference points, 314–315
density weighted averages, 148–149, 154–155
depth dependent diffusivities, 265–266
determinant of metric, 409–410
determinants, 411–413, 423–424
dianeutral physics, 191
dianeutral velocity, 128–131
dianeutral/diapycnal processes, 189–193
diapycnal velocity, 128–131
diasurface velocity, 128–131
differential geometry, 396
differential manifold, 398
diffusion tensor, 245–246
diffusive closure, 189
diffusivities, 265–266
Dirac delta-function, 302, 352
discretization basics, 39–40, 162
dissipation of kinetic energy, 94–95
Divergence Theorem, 56, 426
double diffusion, 192–193
downgradient tracer diffusive flux, 245–246
duality, 406
dynamics, 53–64
  averaged for a parcel, 156–158
  Cauchy's Stress Principle, 54
  fluid columns, 71–73
  general principles, 53–56
  mapping to model variables, 159–161
  parcels, 56–64
  summary of averaged, 158–159
  vector invariant, 62–64

Eady growth rate, 266–267
earth angular velocity, 58–59
earth gravitational acceleration, 60–61
earth mass, 60
earth radius, 60
effective gravity, 59–61
effective transport velocity, 170
Einstein summation convention, 398–399, 433
energetics
  continuum, 84–111
  dissipation, 94–95
  fluid parcel, 92–97
  global ocean, 98–103
  internal, 87, 97–98
  parcel gravitational potential, 95–97
  parcel kinetic, 94–95
  total mechanical for a parcel, 97
ensemble averages
  details, 144–145
  motivation, 38–39
  opposed to space-time averages, 38
enthalpy, 106–108
equation of state, 108–109
Euler's Theorem, 89
Eulerian and Lagrangian mapping, 31–32
Eulerian description, 30
exact differential, 88
extensive variable, 87–89
external gravity waves, 226–227
external mode, 71

First Law of thermodynamics, 87–89
flow dependent diffusivities, 265–266, 272
fluxes, 233–234
  flux plus restore boundary condition, 214–215
  for surface mixed layers, 213–214
  oriented along neutral directions, 248–250
forces
  centrifugal, 59–61
  Coriolis, 61–62
  external/body, 54
  internal/contact, 54, 62
fresh water flux
  associated with salinity restoring, 214–215
  tracer concentration, 212–213
friction
  biharmonic friction, 354–356
  Cartesian horizontal friction, 343
  deformation rates, 347–348
  discretization details, 377–387
  discretization methods, 375–390
  dissipation functional, 350–354
  general laplacian friction, 347–349
  horizontal friction in continuum, 331–356
  Laplacian plus metric formulation, 387–390
  Laplacian versus biharmonic, 365–367
  nonconstant viscosity, 343–344
  quasi-hydrostatic approximation, 337–338
  vertical shears over grid cell, 221–222
friction in ocean models, 7
functional derivative, 301–302, 306, 350–352
functional formalism, 301–305, 350–354

gauge freedom, 166, 246
Gauss's Law, 56, 426
Gaussian curvature, 260
generalized vertical coordinates, 122–131
  transformation matrix, 123–124
  transformations, 124–125
Gent-McWilliams eddy velocity discretization, 323–325
Gent-McWilliams skew fluxes discretized, 323
Gent-McWilliams stirring, 261–264
geopotential coordinate ocean models, 115–116
geostrophic balance, 94
geostrophic turbulence, 25–27
global tracer conservation, 204–205
GM90 and thickness diffusion, 184–185
gradient operator, 408, 420

gravitational acceleration, 60–61
gravitational instability, 25–27, 191–192
gravitational potential energy, 95–97
grid Reynolds number, 360
grid wave length, 367
grid wave number, 367, 369

halobaricity, 192, 259–260
Hamiltonian function, 75–76
homogeneous functions, 89
Hooke's law, 336, 339
horizontal friction in ocean models, 7
horizontal shearing strain, 342–343, 348, 357
horizontal tension, 342–343, 347–348, 357
horizontal tracer diffusion, 257
hydrostatic pressure
  approximate, 71–72
  baroclinic pressure, 71–72
  basic elements, 34
  equations for, 64–66
  ocean surface grid cell, 224–225
  surface layer pressure, 71–72

inertial reference frames, 59
inexact differential, 88
information loss, 189
integral theorems, 426
integrating factor, 88
intensive variable, 88–89
internal energy, 87, 97–98
internal mode, 71
internal waves, 191–192, 195
interpreting model equations, 144–145
isentropic motion, 103–104, 109–110
isopycnal models, 110, 118, 248

Jacobian of transformation, 401, 409–410

Kelvin waves, 25–27
kinematics, 28–33, 45–53
  averaged for a finite domain, 149–153
  averaged for a parcel, 148–149
  basics of averaged, 145
  continuity relations, 32–33, 45–46
  Eulerian and Lagrangian mapping, 31–32
  Eulerian description, 30
  fixed space coordinates, 29–30
  Lagrangian description, 30
  material coordinates, 29–30
Kolmogorov scale, 192
Kronecker symbol, 399, 401, 434

Lagrangian description, 30
Lamb wave, 34, 111, 225
lapse rate, 90–92
lateral directions, 190
lateral physics, 189–191
Levi-Civita symbol, 32, 337, 411–413
linear momentum density, 51
local tracer conservation, 204–205

mass balance
  column, 50–51
  diagnosing the density time tendency, 207
  domains, 49–53
  generalized coordinates, 132
  mass and tracer compatibility, 204–205
  model grid cell, 205–207
  ocean, 51
  ocean column, 207–209
  parcels, 32–33, 33, 45–46
material tracers, 84–86
mechanical mixing, 191–192
mesoscale eddies, 164–166
metric tensor, 43–44, 401–402, 433–434, 436–437
  determinant, 409–410, 437
  for general surface, 127–128
  for surface in Euclidean space, 413–414
  for surface of sphere, 403
  induced, 413–414
  mapping vectors to one-forms, 406
  using generalized coordinates, 128
  vanishing covariant derivative, 422
metric terms
  advection, 58
  friction, 376–377, 387–390
  general discussion, 81–82

metricity, 422
mixed layer, 193
momentum balance
  fluxes through ocean surface, 222–225
  for a grid cell, 219–225
  general form of time stepping, 234–235
  split-explicit time stepping, 225–233
  time stepping summarized, 233
Monge gauge, 122
Munk boundary layer, 363–364, 370

natural boundary conditions, 301–302, 351–352
neutral diffusion fluxes discretized, 320–323
neutral diffusivities, 265–266
neutral directions, 110, 190
  specification, 247–248
neutral frame, 252
neutral physics, 190–191
  active tracer balance, 257–258
  diffusion in neutral frame, 254, 255
  diffusion in z-level frame, 254
  discretization details, 300–326
  effective diffusivity, 264
  errors with horizontal diffusion, 257
  functional formalism, 301–305
  Gent-McWilliams potential energy sink, 262
  Gent-McWilliams stirring, 261–264
  neutral diffusion plus Gent-McWilliams, 263–264
  relevance for all resolutions, 241–243
  small angle approximation, 255–257
neutral physics and boundaries, 288–296
neutral surface topology, 248
neutral tracers, 212
Newton gravitational constant, 60
Newton's Second Law, 56
Newton's Third Law, 54–55
no-slip boundary conditions, 351–352
non-hydrostatic modeling, 34
non-inertial reference frames, 59
nonlinear viscosity, 368
numerical dianeutral mixing, 197–198

ocean processes, 25–27
one-forms, 404–405
  covariant derivatives, 421
  outward normal, 54, 415–416
orthogonal coordinates, 426–433
  differential operators, 426–433
  summary of formulae, 436–438
outer product, 419

parameterization of eddies, 164–166
partial cells, 116, 210, 220
path dependent changes, 88
physical processes, 25–27
physical tensor components, 43–44, 428–429, 437–438
piston velocity, 214–215
planetary boundary layer, 114–115
planetary geostrophic equations, 69–71
potential enthalpy, 106–108
potential temperature, 105–108, 110
potential vorticity, 66–71, 87
potential vorticity diffusion, 189
potential vorticity mixing, 265–266
Prandtl number, 370
pressure gradient averaging, 234
pressure gradient errors, 117–118
pressure-work, 88
primitive equations
  averaged, 143–162
  model equations summarized, 218–219
  summary of unaveraged, 146–148
  unaveraged, 42–82
projected coordinates, 122–123

quasi-hydrostatic approximation, 337–338
quasi-non-Boussinesq approximation, 110–111, 225–226
quasi-Stokes transport, 172–174

Redi diffusion tensor, 255
reference points, 314–315
regularizing diffusivities, 269–270

representations, 404
resolved processes, 188–191
response functions, 91
Rhines scale, 267
Richardson number, 268
rigid lid approximation
  basic elements, 36–37
  details and critique, 235–236
Robert time filter, 231–232, 234, 361
Rossby radius of deformation, 267–268
Rossby waves, 25–27
rotational flux, 166–167, 246–247

salt in surface grid cell, 212–213
sea surface height, 35–36
  modified mean, 150–151
  steric effects, 35–36
  tendency for Boussinesq, 48
  tendency for modified mean, 152–153
  tendency for non-Boussinesq, 52
self-adjoint operators, 301, 350
semi-colon notation, 420–421, 434–435
sigma-coordinate ocean models, 116–118
skew-diffusive flux, 245–246
skewsion, 166–168, 246–247
slope clipping, 285–286
slope tapering, 286–287
  recommendations, 288–296
slope vector, 120
Smagorinsky viscosity, 367–372
  background, 371–372
  biharmonic, 371
  discretization, 386–387
  general ideas, 367–368
  scaling parameter, 368–370
small angle diffusion tensor, 255
solid body motion, 75, 344
sound speed, 109
space tensor components, 414
spurious cabbeling, 198
spurious dianeutral mixing, 197–198
stability
  advection and diffusion balance, 360–361
  biharmonic mixing, 363
  Gent-McWilliams stirring, 287–288
  Laplacian diffusion, 361–362
  recommended slope tapering, 288–296
  rotated mixing, 284–287
  slope clipping, 285–286
  slope tapering, 286–287
state function, 88
stencils for discrete neutral fluxes, 323
steric effects, 52
strain tensor, 334–335
stress tensor, 55–56, 62
  friction stresses, 333–337
  kinetic energy dissipation, 340
  symmetry and angular momentum, 336–337, 339
  trace-free, 338, 358
  trace-free constraint, 340–341
  transverse anisotropy, 344–347
  transverse isotropy, 341–344
stretching functions, 410–411, 436
summary of tensor analysis, 433–438
summation convention, 398–399, 433
surface boundary layer model, 222–223
surface kinematic boundary condition
  Boussinesq, 49
  for modified mean height, 152
  non-Boussinesq, 52–53
surface mass flux, 50
  modified mean, 152
surface pressure, 71–72
surface tensor components, 414
surface volume flux, 48
  modified mean, 152
surfaces in Euclidean space, 413–416
  area element, 414
  induced metric, 413–414
  normals and tangents, 415–416
  tensor components, 414–415
surfaces in the ocean, 119–120
symmetry and conservation laws, 79–82

tensor analysis motivated, 7–8
tensor density, 412
tensor transformations
  arbitrary rank, 407–408
  gradient operator, 408
  one-form, 406–407

scalar, 403–404
vector, 399, 403–404
tensors, 398–399
textbooks, 5
thermal wind equations, 70–71
thermal wind Richardson number, 268
thermo-hydrodynamics, 84–111
thermobaricity, 192, 259–260
thermodynamic tracers, 86–87, 105–108
thickness as Jacobian, 124
thickness diffusion, 184–185
thickness equation, 133
thickness of a layer, 119–120
thickness weighted fluxes, 210
torsionless manifold, 422, 435
trace of a tensor, 343
tracer and mass compatibility, 244
tracer balance
budget in surface grid cell, 211–212
flux through ocean surface, 211
model grid cell, 209–213
salt in surface grid cell, 212–213
tracer SGS transport flux, 244
tracer transport tensor, 244–246
tracer variance dissipation, 245
tracers
as scalar fields, 85
averaged tracer in a parcel, 153–156
general form of time stepping, 234
tendency for, 85–86
three general types, 84–87
tracer and mass compatibility, 204–205
types of tracers(, 84
traditional approximation
basic elements, 35
geometric restrictions, 42–45
point particle, 73–79
thin shell approximation, 42
transformation matrix, 253–254, 399, 401, 435–436
transport term, 31
transverse anisotropy, 344–347
transverse isotropy, 341–344
triad stencil structure
density and tracer, 314–315
velocity, 350, 380, 385–386
turbulence closure, 189

unit directions, 399–400, 410–411, 427–428
unit tensor, 399
unresolved physics, 188–191

vector invariant, 62–64
vector notation, 53–54
vector streamfunction, 166
vectors
covariant derivative, 420–421
discussion, 398–400
introduction, 398–399
transformations, 399
Veronis effect, 370–371
vertical coordinate choice, 114–118
generalized, 122–131
isopycnal-coordinates, 118
sigma-coordinates, 116–118
z-coordinates, 115–116
vertical gauge, 167
vertical physics, 189–191
virtual mass fluxes, 36, 96
virtual salt fluxes, 235
viscosity on a sphere, 364–365
viscosity tensor, 336
volume balance
column, 48
domains, 46–49
generalized coordinates, 133
ocean, 48–49
parcels, 33, 45–46
volume element, 409–411
cartesian coordinates, 410
general orthogonal coordinates, 410–411
isopycnal coordinates, 411
spherical coordinates, 410
volume rearrangement, 169–170
von Karman constant, 368
von Neumann stability, 361–362
vorticity, 62–71

wedge symbol for vector cross product, 400

z-coordinate ocean models, 115–116