

UNIVERSITY OF CALIFORNIA, SAN DIEGO

**Computation and Visualization of Geometric Partial Differential Equations**

A dissertation submitted in partial satisfaction of the  
requirements for the degree  
Doctor of Philosophy

in

Mathematics

by

Christopher L. Tiee

Committee in charge:

Professor Michael Holst, Chair  
Professor Bennett Chow  
Professor Ken Intrinsic  
Professor Xanthippi Markenscoff  
Professor Jeff Rabin

2015

Copyright  
Christopher L. Tiee, 2015  
All rights reserved.

The dissertation of Christopher L. Tsee is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

---

Chair

University of California, San Diego

2015

## DEDICATION

To my grandfathers, Henry Hung-yeh Tsee, Ph. D. and Jack Fulbeck,  
Ph. D. for their inspiration.



## EPIGRAPH

*L'étude approfondie de la nature est la source la plus féconde des découvertes mathématiques.* [Profound study of nature is the most fertile source of mathematical discoveries.]

—Joseph Fourier, *Theorie Analytique de la Chaleur*

## TABLE OF CONTENTS

Signature Page . . . . .	iii
Dedication . . . . .	iv
Epigraph . . . . .	v
Table of Contents . . . . .	vi
List of Figures . . . . .	ix
List of Tables . . . . .	xi
List of Supplementary Files . . . . .	xii
Acknowledgements . . . . .	xiii
Vita . . . . .	xv
Abstract of the Dissertation . . . . .	xvi
Introduction . . . . .	1
0.1 The Main Problem: its Motivation and Antecedents . . . . .	2
0.2 Part-by-Part Summary . . . . .	7
<b>I Background</b>	<b>10</b>
Chapter 1 Boundary Value Problems . . . . .	11
1.1 Differential Forms . . . . .	13
1.2 Integration of Differential Forms and Hodge Duality . . . . .	19
1.3 Sobolev Spaces of Differential Forms . . . . .	30
1.4 The Extended Trace Theorem . . . . .	36
1.5 Boundary Value Problems with the Hodge Laplacian . . . . .	42
1.6 The Hilbert Space Setting for Elliptic Problems . . . . .	56
1.6.1 Recasting in terms of Sobolev Spaces . . . . .	57
1.6.2 The General Elliptic Problem . . . . .	63
1.7 The Theory of Weak Solutions . . . . .	69
1.7.1 The Lax-Milgram Theorem . . . . .	71
1.7.2 Basic Existence Theorems . . . . .	72
1.8 Hilbert Complexes . . . . .	76
1.9 Evolutionary Partial Differential Equations . . . . .	87
1.9.1 Motivation: The Heat Equation . . . . .	88
1.9.2 Bochner Spaces . . . . .	90

Chapter 2	Numerical Methods . . . . .	96
2.1	The Finite Element Method . . . . .	97
2.1.1	The Rayleigh-Ritz Method . . . . .	98
2.1.2	The Galërkin Method . . . . .	102
2.2	Details of the Finite Element Method . . . . .	103
2.2.1	The Basis . . . . .	104
2.2.2	Shape Functions . . . . .	106
2.2.3	Computation of the Stiffness Matrix . . . . .	109
2.3	Adding Time Dependence . . . . .	110
2.4	Numerical Methods for Evolutionary Equations . . . . .	112
2.4.1	Euler Methods . . . . .	113
2.4.2	Other Methods . . . . .	117
2.5	Error Estimates for the Finite Element Method . . . . .	120
2.6	Discretization of Differential Forms . . . . .	126
2.6.1	Approximation in Hilbert Complexes . . . . .	128
2.6.2	Approximation with Variational Crimes . . . . .	130
2.6.3	Polynomial Spaces and Error Estimates for Differential Forms . . . . .	133
Chapter 3	Some Methods for Nonlinear Equations . . . . .	141
3.1	Overview . . . . .	142
3.2	Linearizing the Equation . . . . .	154
3.3	Adding Time Dependence . . . . .	156
3.4	Newton's Method . . . . .	157
3.4.1	Kantorovitch's Theorem . . . . .	160
3.4.2	Globalizing Newton's Method . . . . .	161

## **II Applications to Evolution Problems 164**

Chapter 4	Approximation of Parabolic Equations in Hilbert Complexes . . . . .	165
4.0	Abstract . . . . .	165
4.1	Introduction . . . . .	166
4.2	The Finite Element Exterior Calculus . . . . .	172
4.2.1	Hilbert Complexes . . . . .	172
4.2.2	Approximation of Hilbert Complexes . . . . .	177
4.2.3	Extension of Elliptic Error Estimates for a Nonzero Harmonic Part . . . . .	182
4.3	Abstract Evolution Problems . . . . .	194
4.3.1	Overview of Bochner Spaces and Abstract Evolution Problems . . . . .	195
4.3.2	Recasting the Problem as an Abstract Evolution Equation . . . . .	198
4.4	<i>A Priori</i> Error Estimates for the Abstract Parabolic Problem . . . . .	201

4.5	Parabolic Equations on Compact Riemannian Manifolds . . .	213
4.6	Numerical Experiments and Implementation Notes . . . . .	228
4.7	Conclusion and Future Directions . . . . .	230
4.8	Acknowledgements . . . . .	232
Chapter 5	Finite Element Methods for Ricci Flow on Surfaces . . . . .	233
5.1	Introduction . . . . .	233
5.2	Notation and Conventions . . . . .	234
5.3	The Ricci Flow on Surfaces . . . . .	238
5.4	Weak Form of the Equation . . . . .	242
5.5	Numerical Method . . . . .	245
5.6	A Numerical Experiment . . . . .	249
5.7	Conclusion and Future Work . . . . .	252
5.8	Acknowledgements . . . . .	256
<b>III Appendices</b>		<b>257</b>
Appendix A	Canonical Geometries . . . . .	258
A.1	Introduction to Spectral Geometry . . . . .	258
A.2	Solving Poisson's Equation . . . . .	261
A.3	Finding Dirichlet Green's Functions . . . . .	263
A.4	The Dirichlet Problem . . . . .	265
A.5	The Neumann Problem . . . . .	268
Appendix B	Examples of Green's Functions and Robin Masses . . . . .	272
B.1	In One Dimension . . . . .	272
B.2	Two-Dimensional Examples . . . . .	280
B.3	Two-Dimensional Example: The Hyperbolic Disk . . . . .	292
B.4	Derivations for Neumann Boundary Conditions . . . . .	298
B.5	The Finite Cylinder . . . . .	303
B.6	Domains with Holes in the Plane and the Bergman Metric . .	308
B.7	Conclusion and Future Work . . . . .	321
Bibliography	. . . . .	322
Index	. . . . .	330

## LIST OF FIGURES

Figure 1.1:	Vectors forming a parallelepiped. . . . .	15
Figure 1.2:	A nonorientable manifold: the Möbius strip and transition charts; the left and right edge are identified in opposite directions as indicated by the black arrows. The interior of the charts are indicated with the respective colored arrows and dashed curve boundaries. . . . .	22
Figure 1.3:	Demonstration of the cone condition and its violation: (1.3a): The cone condition. Note that the nontrivial cone fits in the corners (and of course, everywhere else) nicely, although it occasionally requires a rigid motion. (1.3b): . . . . .	31
Figure 1.4:	A 1-form $\omega$ (thin black level sets) whose hodge dual $\star\omega$ (gray field lines) has vanishing trace on the boundary $\partial U$ . This says the field lines of $\star\omega$ are tangent to $\partial U$ . . . . .	51
Figure 1.5:	A form and pseudoform in $\mathbb{R}^2$ dual to each other, with the two kinds of boundary conditions in the annulus $A = \{a < r < b\}$ . (1.5a): $d\theta$ , a harmonic form whose <i>Hodge dual</i> has vanishing trace on $\partial A$ . (“ $d\theta$ ” actually is a form determined by overlaps, $\theta \in (-\pi, \pi)$ ) and . . . . .	54
Figure 1.6:	Example of harmonic form on closed manifold (here, a torus). . . . .	55
Figure 1.7:	Two generators for the harmonic forms for $\mathring{\mathfrak{H}}^1(A)$ and $\mathfrak{H}^1(A)$ , where $A$ is the annulus $\{a < r < b\} \subseteq \mathbb{R}^2$ , reflecting the different kinds of boundary conditions. Note how different they are, but at the same time, how they are dual in some sense, one having level sets that . . . . .	81
Figure 2.1:	Example tent function constructed for the node $\frac{1}{2}$ ; where the nodes in the mesh are are $\frac{k}{4}$ , $k = 0, \dots, 6$ . . . . .	104
Figure 2.2:	The heat equation on a piecewise linear approximation of a sphere (3545 triangles). The solution is graphed in the normal direction of the sphere. The spatial discretization uses a surface finite element method detailed in Chapter 4 (based on [28]), and implemented . . . . .	117
Figure 2.3:	The wave equation on a piecewise linear approximation of a sphere (3545 triangles). The solution is graphed in the normal direction of the sphere. The spatial discretization uses a surface finite element method detailed in Chapter 4 (based on [28]), and implemented . . . . .	121
Figure 3.1:	Graphical illustration of for Newton’s Method on a function $f$ (the graph $y = f(x)$ is in blue). At each $x_i$ on the $x$ -axis, draw a vertical line (dashed red in the above) to the point $(x_i, f(x_i))$ . From that point, draw a tangent line (in red). Then $x_{i+1}$ is the intersection . . . . .	158

Figure 4.1:	A curve $M$ with a triangulation (blue polygonal curve $M_h$ ) within a tubular neighborhood $U$ of $M$ . Some normal vectors $\nu$ are drawn, in red; the distance function $\delta$ is measured along this normal. The intersection $x$ of the normal with $M_h$ defines a mapping $a$ from $x$ . . .	216
Figure 4.2:	Approximation of a quarter unit circle (black) with a segment (blue) and quadratic Lagrange interpolation for the normal projection (red). Even though the underlying triangulation is the same (and thus also the mesh size), notice how much better the quadratic . . .	217
Figure 4.3:	Hodge heat equation for $k = 2$ in a square modeled as a $100 \times 100$ mesh, using the mixed finite element method given above. Initial data is given as the (discontinuous) characteristic function of a C-shaped set in the square. The timestepping method is given . . .	230
Figure 5.1:	Embedded spheres for the metrics $e^{2u}g$ at time steps 1, 50, 150, and 300 (the timestep $\Delta t$ is $1/72000$ ). This is a picture of the <i>true</i> geometry, using the embedding equations (5.6.4)-(5.6.5). As one can see, the geometry near the equator dissipates faster than that . . .	253
Figure B.1:	Graph of the Green's function for a few values of $y$ , along with the Robin mass. . . . .	274
Figure B.2:	Full graph of the Green's function in two variables. . . . .	275
Figure B.3:	Transformation $f_w$ for $w \approx -0.6$ given by its action on a polar grid. . . . .	282
Figure B.4:	Visualizing the effects of the conformal mapping $f_w$ on the disk, distorting the reference image (B.4a), Bubi. . . . .	284

LIST OF TABLES

Table 1.1: Examples of *cis*- and *trans*-oriented submanifolds  $S$  in  $\mathbb{R}^3$ . Notice the duality of “arrow”-like orientations (orientation via one vector) and “clock-face” orientations (orientation via two vectors), and signs vs. “corkscrews.” See also [35]. . . . . 24

Table 1.2: Orienting the boundary of *cis*- and *trans*-oriented manifolds. . . . . 27

## LIST OF SUPPLEMENTARY FILES

- heat-demo-basic.mov:** This gives an animation of the graph of the solution to the heat equation in a square modeled as a  $100 \times 100$  mesh, using the weak form of the Laplacian on functions for the spatial discretization via a finite element method (see Chapter 2). The timestepping method is given by the backward Euler discretization, with timestep  $\Delta t = 5 \times 10^{-5}$ . Each timestep generates a frame, and the movie runs at 60 frames per second.
- heat-demo-hodge.mov:** The Hodge heat equation for 2-forms in a square modeled as a  $100 \times 100$  mesh, using the mixed finite element method given in Chapter 4 for the spatial discretization. Initial data is given as the (discontinuous) characteristic function of a C-shaped set in the square. The timestepping method is given by the backward Euler discretization, with timestep  $\Delta t = 5 \times 10^{-5}$ . Each timestep generates a frame, and the movie runs at 60 frames per second. See also Figure 4.3.
- heat-on-sphere.mpg:** The heat equation on a piecewise linear approximation of a sphere (3545 triangles). The solution is graphed in the normal direction of the sphere. The spatial discretization uses a surface finite element method detailed in Chapter 4 (based on [28]), and implemented using a modification of FETK [31], and the timestepping scheme is backward Euler. The timestep  $\Delta t$  is  $1/216000$ , and this movie runs at 60 frames per second.
- ricci-flow-on-sphere.mov:** Embedded, piecewise linearly approximated spheres (3545 triangles) for the metrics  $e^{2u}g$  evolving under Ricci flow. The spatial discretization uses a surface finite element method detailed in Chapter 5 (based on [28]), and implemented using a modification of FETK [31], and the timestepping scheme is backward Euler. The timestep  $\Delta t$  is  $1/72000$ . This is a picture of the *true* geometry, using the embedding equations (5.6.4)-(5.6.5). The geometry near the equator dissipates faster than that near the poles, because the value of  $u$  is concentrated over a smaller area, and the factor  $e^{-2u}$  slows the rate of diffusion. Also see Figure 5.1. This movie runs at 60 frames per second.
- waves-on-sphere.mpg:** The wave equation on a piecewise linear approximation of a sphere (3545 triangles). The solution is graphed in the normal direction of the sphere. The spatial discretization uses a surface finite element method, and implemented using a modification of FETK [31], and the timestepping scheme is symplectic Euler. This movie runs at 60 frames per second.



## ACKNOWLEDGEMENTS

I would first like to thank my advisor, Michael Holst, for introducing me to an exciting and interesting field that fits well with my interests. Working at an interesting intersection of topology, geometry, real analysis, and numerical analysis has certainly enriched my understanding and deepened my appreciation for all those fields. I appreciate that he was patient enough to give me the freedom to explore and build background, motivation, and intuition. In addition, I thank him for helping me realize on numerous occasions that I am more capable than I think, and for understanding and accommodating the rather unusual and convoluted path that I have taken.

I thank my committee members for their insights and suggestions on the writing of a thesis, and my previous advisor, Kate Okikiolu, for helping bridge the gap between coursework and research in the fields of analysis and partial differential equations. We had many interesting discussions about analysis, and it is upon her work that much of the material in the appendix is based.

I also thank various other faculty working in numerical analysis, in particular, Melvin Leok and Randy Bank, who let me sit in their classes, helped broaden my perspective, and provided interesting ideas and discussions. I thank the geometric analysis group for introducing me to the modern study of partial differential equations, and showing that despite feeling good about real analysis, I still had a lot to learn. I thank them also for giving me the opportunity to study at MSRI for two quarters.

Throughout this long journey, many friends and colleagues have come and gone. Special thanks is given to ones who have remained. But even for the ones who have gone, I thank them, because every one of them has had something to teach me and some inspiration to give. Students, in particular, have often proved to be some of the best teachers, so I thank them as well, for helping me keep it real (and occasionally, complex).

Many places have accommodated me in both my travels and simply a need for office space apart from home and the department. I would especially like to thank Peet's Coffee and Tea, and their wonderful baristas, for putting up with me staying hours at a time. There, I have met many inspirational people from the neighborhood, who have also provided support and insight. I thank MSRI for its hospitality on two separate occasions. And, for the occasions I did decide to stay in, I thank my current and former roommates for interesting late-night mathematical, scientific, and philosophical discussions. They have undoubtedly seen the most human side of me on a day-to-day basis and have had to put up with all sorts of varying stress levels.

Finally, I would like to thank my family, Mom, Dad, and Charlise (and new additions Scott and Theo Fernando), for their support, incredible patience, and having faith in me, even when I'd lost it in myself. I have dedicated this to my grandfathers, the two Ph.D.'s in my family who unfortunately did not get to see me finish this project and follow in their footsteps (although it is a little ironic that they both were Ph.D.'s in the humanities).

Chapter 4, in full, is currently being prepared for submission for publication. The material may appear as M. Holst and C. Tiede, *Approximation of Parabolic Equations in Hilbert Complexes*. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is currently being prepared for submission for publication. The material may appear as M. Holst and C. Tiede, *Finite Elements for the Ricci Flow on Surfaces*. The dissertation author was the primary investigator and author of this paper.

## VITA

2004	B. S. in Mathematics and Computer Science <i>cum laude</i> , University of California, Los Angeles
2004-2009	Graduate Teaching Assistant, University of California, San Diego
2006	M. A. in Mathematics, University of California, San Diego
2008	C. Phil. in Mathematics, University of California, San Diego
2015	Ph. D. in Mathematics, University of California, San Diego

## PUBLICATIONS

M. Holst and C. Tiee. *Approximation of Parabolic Equations in Hilbert Complexes*. In preparation.

M. Holst and C. Tiee. *Finite Element Methods for The Ricci Flow on Surfaces*. In preparation.

ABSTRACT OF THE DISSERTATION

**Computation and Visualization of Geometric Partial Differential Equations**

by

Christopher L. Tiee

Doctor of Philosophy in Mathematics

University of California, San Diego, 2015

Professor Michael Holst, Chair

The chief goal of this work is to explore a modern framework for the study and approximation of partial differential equations, recast common partial differential equations into this framework, and prove theorems about such equations and their approximations. A central motivation is to recognize and respect the essential geometric nature of such problems, and take it into consideration when approximating. The hope is that this process will lead to the discovery of more refined algorithms and processes and apply them to new problems.

In the first part, we introduce our quantities of interest and reformulate traditional boundary value problems in the modern framework. We see how Hilbert

complexes capture and abstract the most important properties of such boundary value problems, leading to generalizations of important classical results such as the Hodge decomposition theorem. They also provide the proper setting for numerical approximations. We also provide an abstract framework for evolution problems in these spaces: Bochner spaces. We next turn to approximation. We build layers of abstraction, progressing from functions, to differential forms, and finally, to Hilbert complexes. We explore finite element exterior calculus (FEEC), which allows us to approximate solutions involving differential forms, and analyze the approximation error.

In the second part, we prove our central results. We first prove an extension of current error estimates for the elliptic problem in Hilbert complexes. This extension handles solutions with nonzero harmonic part. Next, we consider evolution problems in Hilbert complexes and prove abstract error estimates. We apply these estimates to the problem for Riemannian hypersurfaces in  $\mathbb{R}^{n+1}$ , generalizing current results for open subsets of  $\mathbb{R}^n$ . Finally, we apply some of the concepts to a nonlinear problem, the Ricci flow on surfaces, and use tools from nonlinear analysis to help develop and analyze the equations. In the appendices, we detail some additional motivation and a source for further examples: canonical geometries that are realized as steady-state solutions to parabolic equations similar to that of Ricci flow. An eventual goal is to compute such solutions using the methods of the previous chapters.

# Introduction

Geometry is one of the oldest concepts known to human existence and often cited as the inauguration of the formal study of mathematics (in Euclid's *Elements*). How we perceive and consider the natural world has been of immense importance, and visualization is one of our most powerful tools. Another important ingredient for understanding the laws of nature has been the study of differential equations, first conceived by Isaac Newton. Since then, it has gradually been seen that many aspects of geometry enter into the structure of differential equations, and vice versa (Newton himself phrased all his work in the language of Euclid, despite having discovered calculus, in order to be able to communicate in the common language of his scientific peers). The interaction has been fruitful and elucidating. Our broad purpose is to explore that interaction—we examine how differential equations lead to interesting geometric structures, and reciprocally, how geometric problems can set up interesting differential equations. Because so many of the relevant equations lack closed-form analytical solutions, we must compute solutions numerically, which is why numerical approximation will also become a crucial part of this thesis—reflecting that it is difficult to truly understand a nontrivial differential equation without concrete, visualizable geometric representations. Computation is the only effective way to produce a realistic simulation of the solution of the differential equation. We may use such geometric information to formulate new conjectures and laws, clarify and elucidate old ones—

that is, do science, and try to answer, in however a small part, deep questions of the universe we live in.

## 0.1 The Main Problem: its Motivation and Antecedents

Having stated our general purpose, we present the specific problem we wish to explore in this work. One of the most fundamental model evolution problems in partial differential equations is the HEAT EQUATION: given a bounded open set  $U \subseteq \mathbb{R}^n$ , and some time interval  $T$ , and a function  $f : U \times [0, T] \rightarrow \mathbb{R}$  representing a time-varying heat source throughout  $U$ , and an initial temperature profile  $g : U \rightarrow \mathbb{R}$ , with zero boundary values, we seek the evolution of this temperature profile  $u : U \times [0, T] \rightarrow \mathbb{R}$ . We find that  $u$  must satisfy the partial differential equation [30, §2.3]

$$(0.1.1) \quad \begin{aligned} \frac{\partial u}{\partial t} - \Delta u &= f && \text{in } U \times (0, T) \\ u(x, t) &= 0 && \text{on } \partial U \times (0, T) \\ u(x, 0) &= g(x) && \text{in } U \times \{0\}, \end{aligned}$$

with  $\Delta = \sum \partial_i^2$  being the Euclidean Laplacian operator. Obviously, we may consider more general boundary conditions. This problem appears in many different guises throughout applied mathematics, so it is of great interest to find methods to approximate its solutions. It is generalizable in many different ways; the route we take is to examine a more geometric setting, in which we no longer require our domains to be open subsets of Euclidean space, but rather for  $U$  a Riemannian manifold-with-boundary (with Lipschitz smoothness). In addition, we want to be able to formulate similar equations for quantities more general than scalars, namely for differential forms. This allows us to think of these equations in a more invariant way. Specifically, for differential forms, we assume the existence of a Riemannian metric, and the exte-

rior codifferential  $\delta$  adjoint to the exterior derivative  $d$ , to instead arrive at the HODGE HEAT EQUATION, replacing occurrences of functions by possibly time-dependent differential  $k$ -forms [4]:

$$(0.1.2) \quad \begin{aligned} \frac{\partial u}{\partial t} + (\delta d + d\delta)u &= f(x, t) & \text{in } \Lambda^k(U) \times (0, T) \\ u(x, 0) &= g(x) & \text{in } \Lambda^k(U) \times \{0\}. \end{aligned}$$

$-(\delta d + d\delta)$  is the appropriate generalization of the Laplace operator, and so is also denoted by  $\Delta$ . The relevant boundary conditions are more complicated, and more interesting. We can consider the trace (tangential restriction) of  $u$  and its differential  $du$  to vanish on the boundary (corresponding to ESSENTIAL BOUNDARY CONDITIONS), or the trace of the *Hodge duals* of these quantities to vanish (corresponding to NATURAL BOUNDARY CONDITIONS). It generalizes the classical boundary conditions commonly encountered in electrostatics, namely tangential or normal continuity [42, 56, 85].

Central to both the solution and approximation of these problems is considering a WEAK FORMULATION via integration by parts. This enables us to use the modern methods of Sobolev spaces to describe the solutions and their approximations. The chief numerical method we are concerned with is the FINITE ELEMENT METHOD, which realizes an approximation by assuming the solutions lie in appropriately chosen finite-dimensional subspaces of our Sobolev spaces. Of course, if we wish to approximate solutions, we should also try to make an estimate of the error in our approximations, so that we know our numerical methods are sound. The error depends on the kind of finite element spaces we choose, as well as properties such as the regularity of the data and the domain. Principally, we usually seek estimates of the form  $C\|u\|h^\beta$  where  $h$  is an appropriate discretization parameter that accumulates to 0 and the corresponding finite element spaces “converge to the whole space” in some sense as  $h \rightarrow 0$ . The ORDER of the approximation is the power  $\beta$  and depends, again, on similar factors.



The relationship between the well-posedness of the problem and its discretization can be surprisingly subtle, and for the elliptic operators in our problems, was studied in-depth by Arnold, Falk, and Winther [5] for the Hodge Laplacian in Euclidean space.

Arnold, Falk, and Winther continue the theory established in [5] in a second work, [6], in which they place this problem in a more abstract framework, that of HILBERT COMPLEXES (introduced in [14])—sequences of Hilbert spaces  $W^k$ , with cochain operators  $d$  defined on domains  $V^k \subseteq W^k$ , that capture the main properties of the  $\mathcal{L}^2$  theory for differential forms. This approach is powerful, because we can understand precisely how concepts such as well-posedness of our equations comes about, and what abstract properties it depends on, which enables us to unify a multitude of problems. This approach also provides a framework for approximation, and in doing so, we can clarify the problem of well-posedness and stability of the numerical methods. The approach to the elliptic problem  $-\Delta u = (\delta d + d\delta)u = f$  considered is a MIXED FORMULATION, that is, rewriting it as a system and defining  $\sigma = \delta u$ :

$$(0.1.3) \quad \begin{aligned} \langle \sigma, \tau \rangle - \langle u, d\tau \rangle &= 0 & \forall \tau \in V^{k-1} \\ \langle d\sigma, v \rangle + \langle du, dv \rangle + \langle p, v \rangle &= \langle f, v \rangle & \forall v \in V^k \\ \langle u, q \rangle &= 0 & \forall q \in \mathfrak{H}^k, \end{aligned}$$

where  $\mathfrak{H}^k$  is the harmonic space, the abstraction of the concept in Hodge theory,  $p$  is the projection of the data on the harmonic space, which is necessary for the existence of a solution.<sup>1</sup> We also additionally need  $u$  to be perpendicular to the harmonic space for uniqueness. This mixed form turns out to also give the correct theory for discretization—because the theory is formulated abstractly, much of the theory carries

---

<sup>1</sup>The intuitive way of thinking of the requirement of the source being perpendicular to the harmonic space is that elliptic problems are often realized as steady-state solutions to parabolic problems, and a harmonic source term is like a constant source. A nonzero harmonic source term would therefore make the parabolic solution grow to infinity, and thus forbid the existence of a steady state.

over unchanged (restricting  $d$  to finite-dimensional subspaces  $V_h^k$ ).

The connection between the continuous and discrete spaces is established by certain bounded projection operators  $\pi_h^k : V^k \rightarrow V_h^k$ . Under reasonable hypotheses, the methods converge and are stable, expressing the error in terms of Hilbert space BEST APPROXIMATIONS, namely something of the form (for  $\omega$  to be approximated by  $\omega_h$ )

$$\|\omega - \omega_h\|_V \leq C \inf_{\eta \in V_h} \|\omega - \eta\|_V.$$

For the de Rham complex and various triangulations of the domain, we can translate this into estimates in terms of powers of  $h$ . For geometric problems, Holst and Stern [50] remove the restriction that the discrete spaces  $V_h^k$  be subspaces of the domain  $V^k$ , but rather are equipped with certain inclusion morphisms  $i_h^k$ . This contributes to the error, because the inner products on the approximating spaces need not coincide with the inner product on the image subspaces  $i_h V^k$ , i.e.,  $i_h$  may not be unitary, as they are in the special case that  $i_h$  is inclusion. Attempting to correct for this leads directly to additional error terms involving the norm  $\|I - i_h^* i_h\|$ , a precise measurement of the non-unitarity of the operator.

Turning back to time evolution, Gillette and Holst [40] approximate parabolic and hyperbolic evolution problems for the case of top-degree forms  $k = n$  by semidiscretization, generalizing the method of Thomée [106, Ch. 17] for domains in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ . Arnold and Chen [4] focus on parabolic problems but for any degree of differential form (specifically, the Hodge heat equation (0.1.2) above). They semidiscretize the solution in space, leading to evolution equations in certain finite-dimensional spaces. To compute the error, all of the above approaches compare the approximation to an ELLIPTIC PROJECTION of the solution—at each moment in time,  $u$  is already known, so  $u$  trivially solves an elliptic equation with data  $-\Delta u$ . Elliptic projection computes, using the methods developed in [6], another approximation  $\tilde{u}_h$  for  $u$  (i.e., it applies

the discrete solution operator to the known continuous data  $-\Delta u(t)$ ). The theory in [6] therefore gives the error in this approximation, namely, it compares the elliptic projection to the true solution. What remains to do is to compare the semidiscrete solution  $u_h$  to the elliptic projection  $\tilde{u}_h$ , so that we have the full error estimate we want, by the triangle inequality. These error estimates were shown by Thomée [106, Ch. 17] to have the form:

$$(0.1.4) \quad \|u_h(t) - u(t)\|_{L^2} \leq Ch^2 \left( \|u(t)\|_{H^2} + \int_0^t \|u_t\|_{H^2} ds \right),$$

$$(0.1.5) \quad \|\sigma_h(t) - \sigma(t)\|_{L^2} \leq Ch^2 \left( \|u(t)\|_{H^3} + \left( \int_0^t \|u_t\|_{H^2}^2 ds \right)^{1/2} \right).$$

The central equations that make these kinds of results possible are the error evolution equations of Thomée [106]: defining  $\rho = \|\tilde{u}_h(t) - u(t)\|$ ,  $\theta = \|u_h(t) - \tilde{u}_h(t)\|$ , and  $\varepsilon = \|\sigma_h(t) - \tilde{\sigma}_h(t)\|$ , he derives (in a slightly different notation)

$$\langle \theta_t, \phi_h \rangle + \langle d\varepsilon, \phi_h \rangle = -\langle \rho_t, \phi_h \rangle$$

$$\langle \varepsilon, \omega_h \rangle - \langle \theta, d\omega_h \rangle = 0.$$

From this, the error estimates are proved via Grönwall-type arguments.

It is the main project of this work to do for the parabolic problem (0.1.2) the same that has been done for the elliptic problems: find the analogue of (0.1.2) in a more abstract framework, examine the corresponding estimates in the general setting of Hilbert complexes, and clarify what is important in the error equations of Thomée.

It is of considerable interest to examine nonlinear problems, which are much more difficult. In this work we also investigate how some of the theory may apply to a

certain conformal factor equation, a nonlinear diffusion equation of the form

$$\frac{\partial u}{\partial t} = e^{2u}(\Delta u - K) + c$$

for  $K$  some function representing the Gaussian curvature of a background metric  $g_b$  on a compact surface,  $e^{2u}g_b$  representing an evolving metric, and  $c$  a constant that makes the equation have a steady state. This arises from considering the (normalized) Ricci flow equation on surfaces [18, Ch. 5], and indeed,  $e^{2u}g_b$  satisfies the Ricci flow equation. It is also the two-dimensional analogue of the Yamabe flow, and evolves a given initial metric to the constant curvature metric that is guaranteed to exist by the Uniformization Theorem. We describe how some of the previous theory still may apply, and a finite element method suited to it. This presents many challenges not present in the linear theory.

Finally, since the Ricci flow is an example of the intimate relation of parabolic problems to elliptic ones through steady-states (the limiting case as time goes to infinity), we give several examples of some of these steady-state solutions and develop some of their invariants. Our goal here is to provide some additional examples for which some of the numerical methods in the preceding chapters apply, solving for some of these geometries in a similar spirit to the Ricci flow example.

## 0.2 Part-by-Part Summary

We now present the general plan of this work, part-by-part. In the first part, we define our quantities of interest, differential forms, in order to be able to formulate our boundary value problems in an invariant fashion on manifolds, spaces more general than Euclidean space. Next, we introduce the relevant function spaces, in order to be able to use the modern methods of functional analysis to solve these boundary value

problems. We then recall some traditional boundary value problems and find their proper place in the modern framework, and speak of their solution and well-posedness using those functional-analytic methods. We build the theory up in varying layers of abstraction, bridging the classical and modern, in order to gain an understanding of the essential properties of such equations, which are made more obvious by the process of abstraction. The hope is that this process will lead to the discovery of more refined algorithms and processes that continue to respect the geometric nature of various problems. This culminates in the introduction of Hilbert complexes, which capture and abstract the most important properties of these boundary value problems, for their existence and wellposedness. This also leads to generalizations of major classical results such as the Hodge decomposition theorem for differential forms. We also describe an abstract framework for formulating evolution problems in these spaces: rigged Hilbert spaces and Bochner spaces.

We next turn to numerical methods and approximation theory, introducing the finite element method (FEM) to approximate elliptic problems, and the finite element exterior calculus (FEEC) to approximate the analogous problems for differential forms, as well as analyze discretization error. We again build up, as previously, progressing from functions, to forms, and finally, to Hilbert complexes. Indeed, Hilbert complexes provide the proper setting for numerical approximations: much of the same theory applies and gives well-posedness and stability of the approximations, provided we define the correct morphisms (representing the approximation properties of the spaces).

In the second part, we prove our main results; we use the setup developed in the first to place the problem we have described above in the setting of Hilbert complexes, and then apply the approximation theory developed. We also explore what happens with a nonlinear example, giving a sketch of how this theory may apply, indicating further research directions. We first prove an extension of the error estimates of Arnold,

Falk, and Winther [6] for the elliptic problem in Hilbert complexes, and Holst and Stern [50], for cases in which the approximating spaces need not be a subspace. This extension handles solutions with nonzero harmonic part. Next, we consider evolution problems in Hilbert complexes and prove abstract error estimates, and analyze the abstract analogue of the error equations of Thomée [106]. We apply these estimates to the problem for Riemannian hypersurfaces in  $\mathbb{R}^{n+1}$ , generalizing current results of Thomée [106], Gillette and Holst [40], and Arnold and Chen [4] for open subsets of  $\mathbb{R}^n$ . Finally, we apply some of the concepts to a nonlinear problem, the Ricci flow on surfaces [18, Ch. 5], and use tools from nonlinear analysis to help develop and analyze the equations.

Finally, in the appendices, we detail some additional motivation and a source for further examples from the work of Okikiolu [78, 77]: canonical geometries that are realized as steady-state solutions to parabolic equations similar to that of Ricci flow. The goal is to compute such solutions using the methods of the previous chapters.

## **Part I**

# **Background**

# Chapter 1

## Boundary Value Problems

The seasoned student of the theory of differential equations surely knows that solutions to such equations, directly posed, often involve one or more undetermined constants (in the theory of ordinary differential equations, ODE), or undetermined *functions* (in the theory of *partial* differential equations, PDEs). In other words, solutions to differential equations are usually not unique; we usually have a substantial number of degrees of freedom in the solution. To select a unique solution, we usually impose some form of BOUNDARY CONDITION: we constrain our solution to satisfy a certain condition on the boundary of the domain—for example, constraining its value to be equal to a prescribed function on the boundary (DIRICHLET CONDITIONS), or that the normal derivative of function in question is equal to a prescribed function (NEUMANN CONDITION). The problem of solving a differential equation, with one of these constraints, is called a BOUNDARY VALUE PROBLEM (BVP).

For an EVOLUTIONARY differential equation, i.e. one for which one of the independent variables is designated as “time,” we often consider an INITIAL VALUE PROBLEM CAUCHY PROBLEM or (IVP), namely, prescribing the values of the solution at the time  $t = 0$ . Although there are good reasons for making the distinction, it



actually is a special case of BVP: if the solution is defined on some Cartesian product  $\Omega \times [0, \infty)$ , then  $\Omega \times \{0\}$  is genuinely part of the boundary of the domain in question. The nature of solutions to initial value problems can be, in a real sense, very different from those which are traditionally called BVPS, so this distinction is not an artificial one in practice. In fact, the traditional division between initial value problems and BVPS has been claimed [86] to be an even more important distinction than the division of 2nd order PDES into elliptic, parabolic, and hyperbolic equations, especially for numerical considerations. One of our goals is to explore this and carefully discover this fact for ourselves in later chapters.

On the other hand, sometimes the space cannot be so nicely written as a Cartesian product of space and time variables—for example, the manifold of spacetime in the theory of relativity, the notions of space and time, and thus, “initial condition,” are not really so well-defined. Here the distinction is more or less replaced by using the Lorentzian nature of the spacetime metric and considering “initial” data on spacelike hypersurfaces (the use of Lorentz-geometrical methods is very useful even in flat space, such as the analysis of the wave equation in Euclidean space. It can be said that the distinction between hyperbolic and elliptic equations actually arises from the distinction between Riemannian and Lorentz metrics). But this just means that in theory, a time variable is not substantially different from a space one; it just augments the dimension of the problem by one. It is a metric that determines the timelike nature of a chosen coordinate. This shows that *geometric* considerations are essential in the formulation and solution of boundary and initial value problems.

Nevertheless, all of these problems require some kind of additional constraint to uniquely specify their solutions. Our project here is to investigate what are the essential properties of such problems, their higher-dimensional generalizations, and place these problems into a more abstract framework that captures those essential

properties. A general, interesting goal, which certainly will be the focus of further research, is to see how the “information” contained in the boundary or initial condition affects the nature of the unique solution it picks out. In this work, we focus mainly on the parabolic case and its approximation, and extend the existing theory. However, understanding hyperbolic equations is definitely one of the goals for future research.

We use the methods of modern real and functional analysis [34, 92, 112] to prove our results on boundary value problems. Indeed, using functional analysis in this manner (as in [30, 39, 97]) features some interesting uses of Sobolev spaces, and this develops a rich theory that makes PDE theory, as Evans [30] puts it, not just a branch of functional analysis. This method will also be the foundation upon which numerical methods will be built. There are several references for the fundamental boundary value problems in science and engineering texts (e.g. [43, 104]). Additionally, we recast these standard problems into more and more abstract frameworks to see exactly how the classical develops into the modern, following [5, 6]. Along the way, we shall see the essential geometric nature of these problems elucidated. Some of the same concepts involved also apply to numerical analysis (also detailed in [6]) and we also mention these connections where possible.

## 1.1 Differential Forms

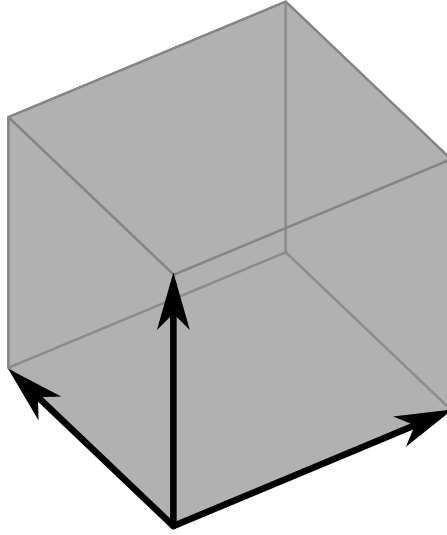
In this section, we define our quantities of interest, differential forms on manifolds, and consider their associated Sobolev spaces. These spaces will be essential, because this is where we use the Hilbert space methods of modern functional analysis that will give us well-posedness and good approximations to the boundary value problems we consider. Many of these spaces generalize classical spaces of vector fields, with curl and divergence being the appropriate derivative operators. Differential forms,

of course, were introduced as a method for generalizing the methods of vector calculus [67, 93] in a more invariant, geometric setting, generalizable to higher dimensions.

**1.1.1 Definition** (Basics of Differential Forms). Let  $M$  be a Lipschitz manifold with boundary (namely, a manifold whose transition charts in the usual sense of differentiable manifolds are LIPSCHITZ MAPPINGS, i.e., mappings  $\varphi$  satisfying  $|\varphi(x) - \varphi(y)| \leq L|x - y|$  in a chart domain for some  $L < \infty$ ). Our main example here would be a domain  $U \subseteq \mathbb{R}^n$  with Lipschitz boundary; frequently  $U$  is the union of some triangulation by  $n$ -simplices, leading to a boundary that would be smooth except where the faces of the triangulation up to dimension  $n - 2$  lie on the boundary. A section of the alternating tensor bundle  $\Lambda^k(M)$  is called a DIFFERENTIAL  $k$ -FORM, or just  $k$ -FORM. The vector space of smooth  $k$ -forms on  $M$  is denoted (following the notation of [62])  $\Omega^k(M)$ . Being alternating tensors, there is an operation, the WEDGE PRODUCT  $\wedge : \Lambda^k(M) \times \Lambda^\ell(M) \rightarrow \Lambda^{k+\ell}(M)$  which acts multilinearly in the vectors, as in the tensor product, but then further antisymmetrized: for tangent vectors  $X_1, \dots, X_{k+\ell} \in TM$ ,

$$(\omega \wedge \eta)(X_1, \dots, X_{k+\ell}) = \frac{1}{k!\ell!} \sum_{\sigma \in S_{k+\ell}} (\text{sgn } \sigma) \omega(X_{\sigma(1)}, \dots, X_{\sigma(k)}) \eta(X_{\sigma(k+1)}, \dots, X_{\sigma(k+\ell)}).$$

The convention of placing  $k!\ell!$  in the denominator (rather than the true average, which would be  $(k + \ell)!$ ) has the useful consequence (for our purposes) in terms of ELEMENTARY  $k$ -FORMS: given a local coordinate system  $(x^i)$ , we write  $\varepsilon^I := dx^{i_1} \wedge \dots \wedge dx^{i_k}$  where  $I$  is any ordered index set with  $1 \leq i_1 < \dots < i_k \leq n$ ; then we have  $\varepsilon^I \wedge \varepsilon^J = \varepsilon^{IJ}$  where  $IJ$  is simply the concatenation of the index sets. Geometrically speaking, it says that the coordinate volume of the  $(k + \ell)$ -dimensional parallelepiped determined with the coordinate vectors as sides is always 1, and is the product of the coordinate volumes of the corresponding  $k$ - and  $\ell$ -dimensional parallelepipeds (see Figure 1.1).



**Figure 1.1:** Vectors forming a parallelepiped.

**1.1.2 Definition** (Forms and Determinants). In general, an alternating  $k$ -tensor (or, for short,  $k$ -covector) at a point can be viewed as an assignment of volume to  $k$ -parallelepipeds (which are, in fact, the  $k$ th alternating product of the *tangent* space rather than the *cotangent* space) at that point; a differential  $k$ -form is then a field of these volume-assigning functions. This is useful in physics, because many field quantities such as forces, electric fields, magnetic fields, etc. can be expressed as functions of various elementary vectors (and parallelograms) such as velocity, displacement, and momentum, etc. [90, 91, 32]. This definition of the wedge product also can be expressed nicely in terms of determinants (unsurprisingly, since determinants are intimately related to the notion of volume): for vectors  $X_1, \dots, X_m$  and covectors  $\omega^1, \dots, \omega^m$ ,

$$(1.1.1) \quad \omega^1 \wedge \cdots \wedge \omega^m(X_1, \dots, X_m) = \det(\omega^i(X_j))_{i,j=1}^m.$$

There are two other ways that determinants interact with wedge products, one way involving linear transformations of the cotangent space, and another way involv-

ing Riemannian metrics. Given  $T$  a  $(1, 1)$ -tensor, i.e. a linear transformation of the cotangent space at every point (which can always be realized, also, as the transpose of a linear transformation of the tangent space at each point), we have, for 1-forms  $\omega^1, \dots, \omega^m$ ,

$$(1.1.2) \quad (T\omega^1) \wedge \cdots \wedge (T\omega^m) = (\det T)\omega^1 \wedge \cdots \wedge \omega^m.$$

Finally, if  $M$  is equipped with a Riemannian metric, we can induce a metric on all differential forms (as well as their dual space). First, we consider the induced metric on the cotangent space, whose components are given by the inverse of  $g_{ij}$  (often denoted with upper indices  $g^{ij}$  rather than the more clumsy  $(g^{-1})^{ij}$ , but since the context is clear, we just use the same notation  $g$  for that). We then define for covectors  $\omega^1, \dots, \omega^k$  and  $\eta^1, \dots, \eta^k$ :

$$(1.1.3) \quad \langle\langle \omega^1 \wedge \cdots \wedge \omega^k, \eta^1 \wedge \cdots \wedge \eta^k \rangle\rangle_g := \det(g(\omega^i, \eta^j)),$$

or, equivalently, we “lower the indices” of each  $\eta^i$  and use the previous relation (1.1.1). Then we extend multilinearly to all of  $\Lambda^k(M)$  and operate pointwise for fields.

**1.1.3 Definition** (Interior products). Rounding out the list of basic algebraic operations, we also consider contractions with tangent vectors. Given  $X \in TM$  a tangent vector, and  $\omega \in \Lambda^k(M)$ , we define the INTERIOR PRODUCT of  $X$  on  $\omega$ , written  $i_X\omega$  or  $X \lrcorner \omega$ , to be the  $(k-1)$ -form given by inserting  $X$  into the first slot namely, the multilinear, alternating map defined on the  $(k-1)$ -tuple  $(X_1, \dots, X_{k-1})$  by

$$X \lrcorner \omega(X_1, \dots, X_{k-1}) = \omega(X, X_1, \dots, X_{k-1}).$$

We extend this to act pointwise for vector fields and  $k$ -forms. One interesting property

it has, which we shall use often (and also is convenient for computation) is that it obeys a product rule: if  $\omega$  is a  $k$ -form and  $\eta$  is an  $\ell$ -form, then

$$X \lrcorner (\omega \wedge \eta) = (X \lrcorner \omega) \wedge \eta + (-1)^k \omega \wedge (X \lrcorner \eta).$$

This is unusual because the interior product is an *algebraic* operator (i.e. acting pointwise and independent of the behavior of sections in a *neighborhood* of a point), unlike most differential operators. This is also the same product rule obeyed by the exterior derivative, which we define soon.

One of the most useful properties of differential forms on manifolds is that they *pull back* under any smooth map, that is, given any  $F : M \rightarrow N$  and a  $k$ -form  $\omega$  on  $N$ , we can define  $F^* \omega$  on  $M$ , unlike the case for vector fields. The definition is very simple:

**1.1.4 Definition.** Let  $\omega$  be a  $k$ -form on  $N$  and  $F : M \rightarrow N$  a smooth map. We define a  $k$ -form on  $M$ , called the PULLBACK of  $\omega$  by  $F$ , and written  $F^* \omega$ , as follows:

$$(F^* \omega)_p(v_1, \dots, v_k) := \omega_{F(p)}(F_* v_1, \dots, F_* v_k),$$

i.e., we push forward all the vectors at  $p$  to vectors at  $F(p)$ , and evaluate the form  $\omega$  at  $F(p)$  on those pushed-forward vectors. Note that  $F^*$  can only pull back *sections* of  $\Lambda^k(M)$ , not individual  $k$ -covectors at each point of  $N$ , because in the latter case, we are faced with the task of defining for all covectors at every point of  $N$ , a corresponding covector at some points of  $M$ , whereas with a section on  $N$ , we only have to define for each point of  $M$  one particular covector from one single covector at the range point. This contrasts with the behavior of vectors and their fields; one can generally only push forward single vectors, but not their fields.

As mentioned on numerous occasions, differential forms are useful because they correctly generalize vector calculus. There is a differential operator,  $d$ , defined on smooth differential forms, which generalizes the classical gradient, curl, and divergence operators of classical vector calculus. These operators find application in many physical theories, e.g. electromagnetism and fluid mechanics [93, 42, 56].

**1.1.5 Definition.** Let  $\omega$  be a differential form. We define the EXTERIOR DERIVATIVE as follows. We first define it as the unique operator  $d$  satisfying:

1. (Linearity)  $d$  is linear.
2. (Cochain Property)  $d^2 = 0$ , that is,  $d(d\omega) = 0$  for any form  $\omega$ .
3. (Action on Functions) If  $f$  is a function,  $df$  is its differential, which is defined via the DIRECTIONAL or GÂTEAUX DERIVATIVE: for a tangent vector  $X$  at  $p$ , the differential of  $f$  at  $p$  is given by

$$df(X) = \left. \frac{d}{dt} \right|_{t=0} (f \circ \gamma)(t),$$

where  $\gamma$  is a curve such that  $\gamma(0) = p$  and  $\gamma'(0) = X$ . If we are in Euclidean space, we may, of course, take  $\gamma(t) = p + tX$ . In coordinates,  $df = \frac{\partial f}{\partial x^i} dx^i$ .

4. (Product Rule) If  $\omega$  and  $\eta$  are  $k$ - and  $\ell$ -forms, respectively,

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

That such an operator exists is proved in many texts, e.g. [62, 53, 17, 33]. It is worth noting that there is a more geometric interpretation of it, based on the definition of divergence given in [93] as the limit of the average flux (surface integral) per unit volume. This general geometric definition of the exterior derivative is given in [53]:

**1.1.6 Theorem.** *Let  $d$  be the exterior derivative for forms on  $U \subseteq \mathbb{R}^n$ . Then given a  $k$ -form  $\varphi$  on  $U$ , with at least  $C^1$  coefficients, and vectors  $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$  based at  $x$ ,*

$$(1.1.4) \quad d\varphi_x(\mathbf{v}_1, \dots, \mathbf{v}_{k+1}) = \lim_{h \rightarrow 0} \frac{1}{h^{k+1}} \int_{\partial P_x(h\mathbf{v}_1, \dots, h\mathbf{v}_{k+1})} \varphi$$

where  $P_x(h\mathbf{v}_1, \dots, h\mathbf{v}_{k+1})$  is the parallelepiped spanned by the vectors and based at  $x$ .

This requires us, of course, to define a notion of integration of differential forms, which we take up in the next section. We should mention a final important property of the exterior derivative which says how it relates to the pullback:

**1.1.7 Theorem** (Naturality of the exterior derivative). *Let  $\omega$  be a smooth  $k$ -form on  $N$ , and  $F: M \rightarrow N$  a smooth map. Then, of course,  $d\omega$  is a smooth  $(k+1)$ -form and may be pulled back by  $F$ , and we have*

$$(1.1.5) \quad d(F^* \omega) = F^*(d\omega),$$

as smooth  $(k+1)$ -forms on  $M$ .

## 1.2 Integration of Differential Forms and Hodge Duality

One of the most useful applications of differential forms is that they can be integrated over appropriately oriented submanifolds, which generalize the notion of vector line and surface integrals and integration over volumes (the top-dimensional case). What makes this work is that the behavior of forms under pullbacks can be used to reduce it to that top-dimensional case, which, along with partitions of unity, is reduced to (Lebesgue) integration over subsets of Euclidean space (we should make a note that all our integrals will be interpreted in the sense of Lebesgue, especially for



Sobolev space methods). This requires a notion of orientation, in order for combining results on different charts over a partition of unity to be well-defined (it is ultimately rooted in the fact that the general linear group  $GL_n(\mathbb{R})$  always has two disconnected components). The most efficient way to introduce orientation is via a certain line bundle, which keeps the unwieldy nature of multiple, consistently oriented charts in one place. Taking the tensor product (“twisting”) with this line bundle gives us differential pseudoforms ([36, §2.8 and §3.2], [9, §2.7],[15]), which are differential forms that take into account local orientation information, and actually are the most appropriate objects for volume and flux. It justifies the *streamline* or *field line* picture associated with flux.

**1.2.1 Definition** (Frames and Orientation). Let  $V$  be a finite-dimensional vector space (over  $\mathbb{R}$ ). Given two ordered bases or FRAMES  $\mathbf{e}_1, \dots, \mathbf{e}_n$  and  $\mathbf{f}_1, \dots, \mathbf{f}_n$ , there exists a unique  $A \in GL_n(\mathbb{R})$  such that for all  $j$ ,

$$\mathbf{f}_j = \sum_i \mathbf{e}_i A_j^i$$

which we abbreviate as  $\mathbf{f} = \mathbf{e}A$ . This is a convenient notational convention (we shall call it the FRAME POSTMULTIPLICATION CONVENTION)—if we consider frames as a “row vector” of basis vectors, to interpret that as matrix multiplication (see, for example, [36, Ch. 9 and §17.1b], and [95, pp. 261-262]); it acts most naturally on the right, and in fact, the *coordinates* in these frames transform correctly with  $A$  acting on the *left*. If  $\mathbf{e}$  acts on a column vector  $v \in \mathbb{R}^n$ , then  $\mathbf{e}v$  is a vector in  $V$  and  $v$  gives the coefficients of  $\mathbf{e}v$  in the basis; then to change the basis,  $\mathbf{f}w = \mathbf{e}Aw$ , which shows exactly how the dual behavior of  $A$  taking the  $\mathbf{e}$  frame to  $\mathbf{f}$  also takes coordinates in the  $\mathbf{f}$  frame to coordinates in the  $\mathbf{e}$  frame. We also often omit the summation over the dummy indices when dealing with tensor quantities, a standard technique in many texts in differential

geometry and relativity:

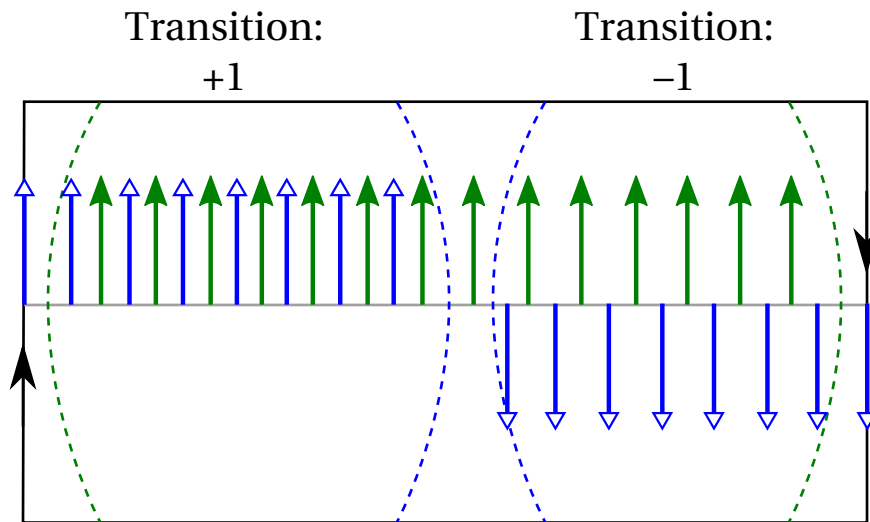
**1.2.2 The Einstein Summation Convention.** Given any tensor field quantity on a manifold  $M$  given in a frame, a formula with repeated indices, one in a lower position and one in an upper position, is regarded as a sum for values of the index up to the dimension of the manifold.

If  $A \in GL_n(\mathbb{R})$  has a positive determinant, we say  $\mathbf{e}$  and  $\mathbf{f}$  have the same ORIENTATION. Otherwise, we say that they are different. Since determinants preserve multiplication (i.e., the determinant is a group homomorphism from invertible matrices to nonzero scalars), this is an equivalence relation, and so frames for  $V$  define two equivalence classes. It is important to know that for a general (real) vector space  $V$ , there is *no canonical choice* of orientation; it must be specified in advance by external criteria. In  $\mathbb{R}^n$  itself, we take the orientation given by the standard basis written in the usual order (often called RIGHT-HANDED), but this cannot be transferred in an invariant way to an arbitrary vector space  $V$ , for the simple reason isomorphisms of  $V$  with  $\mathbb{R}^n$  are equivalent to choosing bases (and thus two bases of different orientations lead to equally good isomorphisms that differ in orientation).

By a similar argument as (1.1.2), an orientation is also a choice of half-line in the space of  $n$ -fold wedge products of frame vectors: given two frames  $\mathbf{e}$  and  $\mathbf{f}$  as above, related by  $A$ , we have

$$(1.2.1) \quad \mathbf{f}_1 \wedge \cdots \wedge \mathbf{f}_n = \det(A) \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n.$$

**1.2.3 Definition** (Orientation line bundles and orientation of manifolds). Now given a manifold-with-boundary  $M$ , we consider the line bundle  $L$  define by taking coordinate patches over  $M$  and taking the transition maps of the bundle to be the sign of the determinant [9, §2.7].  $M$  is ORIENTABLE if this bundle is trivial, i.e., we can find a



**Figure 1.2:** A nonorientable manifold: the Möbius strip and transition charts; the left and right edge are identified in opposite directions as indicated by the black arrows. The interior of the charts are indicated with the respective colored arrows and dashed curve boundaries.

covering of  $M$  by coordinate charts such that transition maps are all positive. A choice of charts, or a nonvanishing section of this bundle (which witnesses the triviality) is called an **ORIENTATION** of  $M$ . A **DIFFERENTIAL PSEUDOFORM** is a section of  $L \otimes \Lambda^k(M) := \Lambda_{\psi}^k(M)$  ( $\psi$  stands for *pseudo-*). Locally, a pseudoform looks like a form plus a choice of orientation over a coordinate patch. We define the exterior derivative to operate on the form portion (the fact that the transition functions are the constant functions  $\pm 1$  ensures this is well-defined), and can similarly extend the operations of wedge, interior, etc. products by doing the corresponding operation on the form parts and defining the product of orientations to be 1 if they agree and  $-1$  if they disagree. Given a form or pseudoform, whether or not it is pseudo- is referred to as its **PARITY** (this terminology originates from de Rham (who introduced the concept in [21]) referring to forms as “forms of the even kind” and pseudoforms as “forms of the odd kind”).

**1.2.4 Definition** (Integration of top forms). Given a smooth  $n$ -pseudoform or  $n$ -form supported in a single coordinate chart, we define its **INTEGRAL** to be the (Lebesgue)

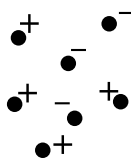
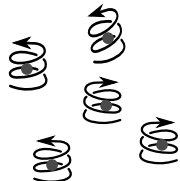
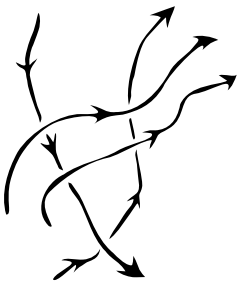
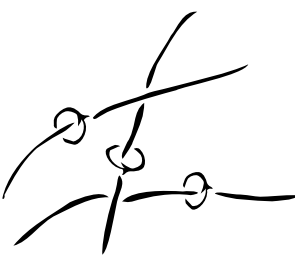

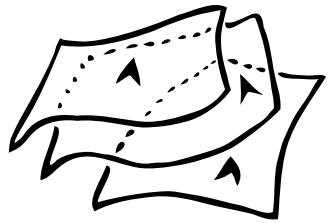
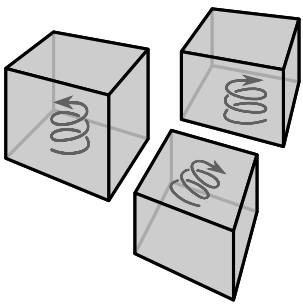
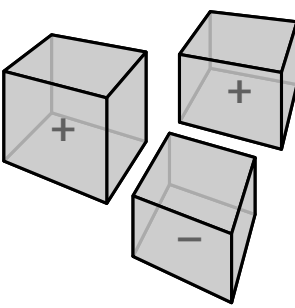
integral of its representation  $f dx^1 \wedge \cdots \wedge dx^n$  in the chart  $(U, \Phi = (x^i))$ :

$$(1.2.2) \quad \int_U \omega = \int_{\mathbb{R}^n} (\Phi^{-1})^* \omega = \int_{\mathbb{R}^n} f dx^1 \wedge \cdots \wedge dx^n.$$

Given an  $n$ -pseudoform defined over all of  $M$ , we use a partition of unity to patch together the integral of its restrictions to each chart (multiplying by a sign  $\pm 1$  according to whether the orientation part of the form agrees with the chart coordinates written in order). For an  $n$ -form, we require  $M$  to be orientable, and we either take the charts to all have the same orientation and integrate as before, ignoring orientation completely, or we consider the given orientation of  $M$  as being tensored with the form, making it into a pseudoform, so that the integral is defined as before. The formula (1.2.2) is invariant under diffeomorphism for pseudoforms, by the Change of Variables formula in Lebesgue integration [34, §2.5 and Ch. 11]; the pseudo-ness has the effect of putting a sign on the determinant for pullbacks (and indeed is the mathematical *raison d'être* for pseudoforms). Orientation-preserving diffeomorphisms for forms makes the sign unnecessary, if positively oriented charts are chosen, thus making integration of forms invariant under orientation-preserving transformations. Thus the use of pseudoforms is more fundamental for integration.

**1.2.5 Definition** (Integrals of  $k$ -forms over submanifolds). Integrals of  $k$ -forms or pseudoforms must proceed over  $k$ -submanifolds  $S$  rather than the whole space  $M$ . There is one catch, however;  $S$  must be appropriately oriented, and unlike the top-dimensional case, using pseudoforms does not completely eliminate the need for some form of orientability. Instead, being to integrate forms or pseudoforms depends on the *type* of orientation. Forms are integrated over *oriented* submanifolds, while pseudoforms are integrated over TRANSVERSELY ORIENTED submanifolds [36, §3.2] (write *trans*-oriented for short, and *cis*-oriented to distinguish it from the usual no-

**Table 1.1:** Examples of *cis*- and *trans*-oriented submanifolds  $S$  in  $\mathbb{R}^3$ . Notice the duality of “arrow”-like orientations (orientation via one vector) and “clock-face” orientations (orientation via two vectors), and signs vs. “corkscrews.” See also [35].

$S \subseteq M = \mathbb{R}^3$		
$k = \dim S$	<i>Cis</i> -oriented example	<i>Trans</i> -oriented example
$k = 0$	 <p>(a) Oriented by choice of signs.</p>	 <p>(b) Oriented by handedness of corkscrews or helices.</p>
$k = 1$	 <p>(c) Oriented by path traversal.</p>	 <p>(d) Oriented like a rotation axis.</p>
$k = 2$	 <p>(e) Oriented by clock sense.</p>	 <p>(f) Oriented by facing direction.</p>
$k = 3$	 <p>(g) Oriented by handedness of corkscrews or helices.</p>	 <p>(h) Oriented by choice of signs.</p>

tion). A *trans*-oriented submanifold is one that is oriented via normal vectors (i.e. orientations in the orthogonal complement of the tangent space), while a *cis*-oriented manifold is oriented (as before) by orientations in the tangent space. See Table 1.1 for the concept in  $\mathbb{R}^3$  and [35].

For *cis*-oriented submanifolds, integration of forms is short work: we define the integral to be the integral of the form pulled back by the inclusion map. This is then a top-dimensional form in the submanifold and can be integrated as previously. For *trans*-oriented manifolds, we also wish to pull back, which is not always possible for pseudoforms, because it is generally not possible to translate higher-dimensional orientations to lower-dimensional ones—a basic example being that the notion of congruence of shapes in the Euclidean plane by allowing (orientation-preserving) 3-dimensional rotations ends up realizing reflections in the plane, which is not orientation-preserving in 2 dimensions. However, transverse orientations fix this by specifying a consistent orientation that completes lower-dimensional bases to higher-dimensional ones. We then pull back a pseudoform by locally writing the pseudoform in the tensor product and pulling the form part back as usual. The orientation part is dealt with by considering top-dimensional orientations, aligning the first  $(n - k)$  vectors with the transverse orientation, and then throwing them out (i.e. taking repeated interior products) [36, §3.2], [21, §5]. Finally, it should be mentioned that for orientable ambient manifolds, *cis*- and *trans*-orientability are equivalent. Nevertheless, the *visualization* of these properties should still be kept separate, because sometimes the *trans* picture is much more natural (the most important example being flux).

Our final basic result is Stokes' Theorem:

**1.2.6 Theorem** (The Generalized Stokes' Theorem). *Let  $\omega$  be a smooth differential  $(n - 1)$ -(pseudo)form on an  $n$ -dimensional manifold-with-boundary  $M$ . We may*

orient the boundary transversely by taking the outward normal, or for cis-oriented  $M$ , this also leads to a corresponding cis-orientation of the boundary (see Table 1.2 for the concept in  $\mathbb{R}^3$ ). Then

$$\int_{\partial M} \omega = \int_{\partial M} i^* \omega = \int_M d\omega.$$

We note that we can extend this result (see the next section) for differential forms in Sobolev spaces.

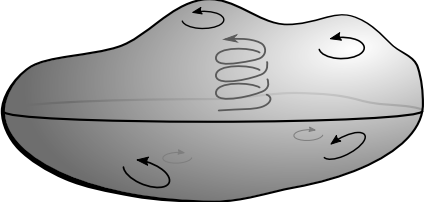
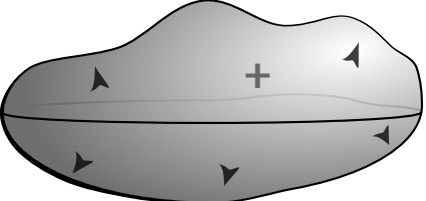
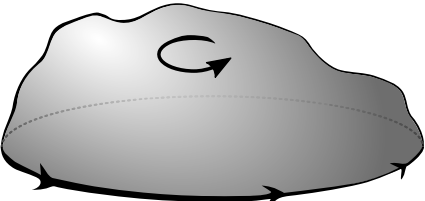
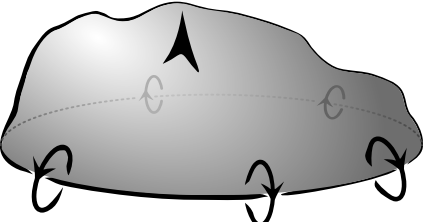
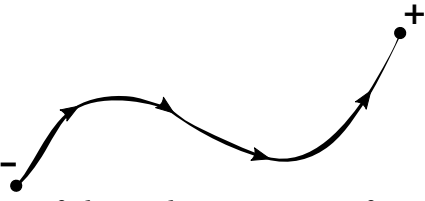
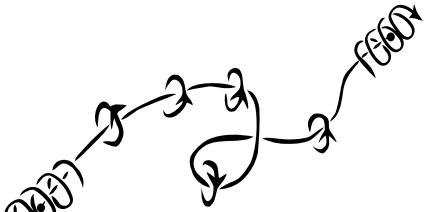
Given a Riemannian manifold-with-boundary, we can form a couple of other important operations that will be important for Sobolev spaces of forms. First, we can define a standard  $n$ -pseudoform, the RIEMANNIAN VOLUME FORM (written  $dV_g$ , though it is usually not  $d$  of anything), given in a coordinate chart by  $\sqrt{\det(g_{ij})} dx^1 \wedge \cdots \wedge dx^n$ , with the orientation given by writing the coordinates in order (i.e.  $dx^1 \wedge \cdots \wedge dx^n$  itself). For oriented Riemannian manifolds, we can say it is a differential  $n$ -form by assuming all our charts are positively oriented. This may be used to integrate *functions* by simply multiplying functions by the standard volume form. (Without a metric or other standard  $n$ -form, we cannot, in general, make invariant sense of the integral of a function.) Given any  $k$ -forms  $\eta, \omega$ , we define the  $\mathcal{L}^2$ -INNER PRODUCT

$$\langle \eta, \omega \rangle = \int_M \langle \langle \eta, \omega \rangle \rangle_g dV_g,$$

where the  $\langle \langle \cdot, \cdot \rangle \rangle_g$  is the pointwise inner product defined via determinants in (1.1.3) above. In the case we have complex-valued forms (which will be useful any time we deal with Fourier transforms), we place the complex conjugation on the *first* factor, the physics convention [64].

**1.2.7 Definition** (Hodge duals). Related to this is another independently useful opera-

**Table 1.2:** Orienting the boundary of *cis*- and *trans*-oriented manifolds.

$S \subseteq M = \mathbb{R}^3$		
$k = \dim S$	Transferring a <i>cis</i> -orientation	Transferring a <i>trans</i> -orientation
$k = 3$	 <p>(a) Push the helix through the surface so that the direction of traversal goes from the inside to the outside. The projection of the path onto the surface is a “clock sense” orientation.</p>	 <p>(b) For +, choose the outward direction, and for –, choose the inward direction.</p>
$k = 2$	 <p>(c) Bring the “clock-face” orientation to the edge. The part of the orientation closest to the boundary then unambiguously specifies a direction of traversal.</p>	 <p>(d) Bring the normal (“flagpole”) to the edge. Define an axial rotation sense on the boundary by making it pierce the surface in the same direction as the normal. Unlike the usual presentation (e.g., [67]), there is <i>no</i> arbitrary convention about the interior being on the left.</p>
$k = 1$	 <p>(e) If the path moves away from the endpoint, it gets –. If the path moves towards the endpoint, it gets + (consistent with the Fundamental Theorem of Calculus).</p>	 <p>(f) The outward direction completes the additional direction for corkscrew motion.</p>



tion, the HODGE DUAL (or STAR) OPERATOR

$$\star : \Lambda^k(M) \rightarrow \Lambda_{\psi}^{n-k}(M)$$

which maps to forms of opposite parity in such a manner that

$$\langle\langle \omega, \eta \rangle\rangle_g dV_g = \omega \wedge \star \eta.$$

Because a Riemannian metric determines a nonzero pseudoform in the line bundle  $L \otimes \Lambda^n(M)$ , it defines an isomorphism with  $\mathbb{R}$  by “dividing out the volume form”, an operation often conveniently written  $\omega \mapsto \omega / dV_g$ . Then we can equivalently realize  $\star \eta$  is constructed as the Riesz representative (relative to the pointwise inner product  $\langle\langle \cdot, \cdot \rangle\rangle_g$ ) of the mapping, for  $\xi \in \Lambda_{\psi}^{n-k}(M)$ ,

$$\xi \mapsto (\eta \wedge \xi) / dV_g,$$

which shows it exists and is unique. We note for convenience that  $dV_g = \star 1$ ,  $\star$  is defined on pseudoforms by pulling the orientation part out, and multiplying them according to the rule  $+1$  if they match up,  $-1$  if they otherwise, and with this,  $\star \star = (-1)^{k(n-k)}$  on  $k$ -forms. Finally,  $\star$  can very obviously be related to the notion of orthogonality by noting it sends orthonormal  $k$ -frames to orthonormal  $(n-k)$ -frames in such a manner that the wedge product of the orthonormal basis with its dual is the volume pseudoform (and thus the sign is chosen accordingly). This leads to the fundamental relations in  $\mathbb{R}^3$  (with the usual orientation):  $\star dx = dy \wedge dz$ ,  $\star dy = dz \wedge dx$  and  $\star dz = dx \wedge dy$ .

Having defined the operator  $\star$  as an algebraic operator, i.e. in each individual fiber, we extend it, as before, to act on sections (forms) by making it act pointwise. We

define the operator  $\delta$  on  $k$ -forms by  $(-1)^{n(k+1)+1} \star d \star$ . The reason for the sign is that we intend to make  $\delta$  the adjoint of  $d$  with respect to the inner product  $\langle \cdot, \cdot \rangle$ . Briefly, if  $\varphi \in \Omega_c^{k+1}(M)$ , i.e., a smooth form of compact support, and  $\eta$  is another smooth  $k$ -form (with any support), then  $\langle d\eta, \varphi \rangle_{\mathcal{L}^2} = \langle \eta, \delta\varphi \rangle_{\mathcal{L}^2}$ . This is useful for the analogue of distribution theory for forms (CURRENTS) and defining weak differentiation, as we will do in the next section on Sobolev spaces. The easiest way to remember the signs is with the following convenient commutation formula [6, §4]: for all  $\omega \in \Omega^k(M)$ ,

$$(1.2.3) \quad \star \delta \omega = (-1)^k d \star \omega$$

$$(1.2.4) \quad \star d \omega = (-1)^{k-1} \delta \star \omega.$$

**1.2.8 Hodge duals as constitutive relations.** Hodge duals can be viewed as *the* geometry-endowing structure, in the form of CONSTITUTIVE RELATIONS [97, Ch. 1], and as we have already seen, we can recover the metric from  $\star$  by first defining volume to be  $\star 1$ , then defining a metric structure by  $\langle\langle v, w \rangle\rangle_\star = (v \wedge \star w) / \star 1$ . That  $\star$  contains geometric information in the form of constitutive relations leads to an interpretation of general elliptic equations with different coefficients as being simply the Laplace equation in a different metric (we shall see this in our study of Hilbert complexes). For example, in electromagnetism, when rewriting Maxwell's equations in terms of differential forms, we see that we have relations between “flux-like” differential forms (2-forms) and “intensity”-like differential forms (1-forms). These are traditionally called “constitutive relations” with permeability and permittivity tensors [42, 56, 85], and the appropriate generalization here indeed is the use of Hodge operators; in fact, we can simply *define* new Hodge operators to be those tensors [36]. The ability to *have* different such operators is also essential for establishing certain compactness properties of our Sobolev spaces of differential forms, even over general Lipschitz

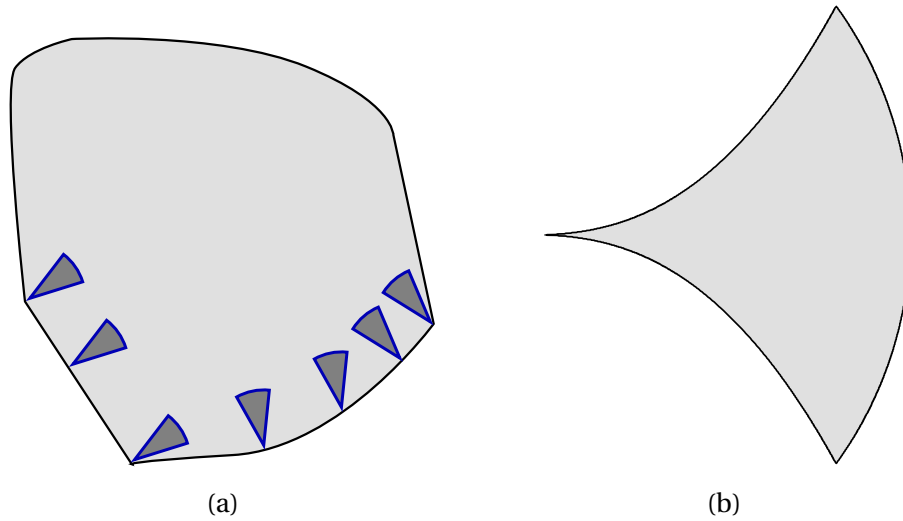
domains [84].

Finally, we remark that Hodge duals can be defined for *Lorentzian* metrics ([36, Ch.], [105, Ch], [69]), that is, the indefinite metrics of special and general relativity. This makes the formulation of Maxwell’s equations in spacetime even more clearly geometric—both the electric and magnetic fields are combined in one 2-form, the Faraday tensor, and its Hodge dual (incorporating both permeability and permittivity) gives another 2-form, the Maxwell tensor. Hodge duals of forms are used for source terms (a “handle to the source” [69, §15.1]), thus their common occurrence as mass or flux (or *quantity*), while forms are used to measure the amount of energy it takes to move test particles in the field (the field *intensity*, thus their common occurrence as *d* of potentials) [36, §3.5d].

### 1.3 Sobolev Spaces of Differential Forms

Here we assume all our manifolds are Lipschitz and compact, possibly with boundary; in particular, we will need them to satisfy some geometrical conditions such as cone conditions ([111, §I.2], [30, Ch. 5], [11, §§II.1-3]) for important theorems to work. In order to have a good theory for the existence and uniqueness of solutions to our boundary value problems on compact manifolds-with-boundary, we need, just as in the theory for functions on a bounded domain in Euclidean space [30, Chs. 5 and 6], Sobolev spaces of differential forms. One effective and useful way to define Sobolev spaces of forms is to work componentwise:

**1.3.1 Definition** (Sobolev spaces of differential forms). We define  $\mathcal{L}^p\Omega^k(M)$  to be all differential forms with  $\mathcal{L}^p$  coefficients in a coordinate basis (where the charts are assumed take values in bounded, open subsets of  $\mathbb{R}^n$ ). Since coordinate changes on compact manifolds with such charts can always be arranged in a manner so as to be



**Figure 1.3:** Demonstration of the cone condition and its violation: (1.3a): The cone condition. Note that the nontrivial cone fits in the corners (and of course, everywhere else) nicely, although it occasionally requires a rigid motion. (1.3b): This domain, with the cusp on its left end (here from the equation  $x^5 = y^2$  near the origin), does *not* satisfy the cone condition.

smooth (even  $C^1$  suffices), the Jacobians are all bounded, and the notions of  $\mathcal{L}^p$  are invariant under such mappings. It is known that this definition of  $\mathcal{L}^p$  works even if the mappings are Lipschitz (and thus the notion of Jacobian only makes sense almost everywhere) [111, §I.4]. Similarly, we define  $H^s\Omega^k(M)$  to be forms whose coefficients in every chart are  $H^s$  functions, for integer  $s$  in terms of number of  $\mathcal{L}^2$  weak partial derivatives [30, §5.2.1] of each component function, or for non-integer  $s$  in the sense of Fourier analysis or Slobodeckii spaces [111, §I.3]. These are called SOBOLEV SPACES of differential forms. They are Hilbert spaces by taking the componentwise Sobolev norms in each chart and gluing together with a partition of unity. Although a specific choice of norm depends on this partition of unity, the space  $H^s\Omega^k(M)$  itself does not depend on it.

Due to the componentwise nature of this definition, all the standard theorems for Sobolev spaces of functions [30, Ch. 5] extend to this case. In particular, we have

that smooth forms are dense in  $H^s\Omega^k$  (thus enabling a very standard technique of defining operators, namely defining the maps for smooth forms and showing they are bounded in the right norms, so a unique bounded extension can be made), forms on bounded domains in  $\mathbb{R}^n$  may be extended to all of  $\mathbb{R}^n$ , we have a trace operator which restricts the forms to the boundary, losing one degree of smoothness, and Sobolev embedding theorems hold (and are in fact *compact* embeddings, so long as our domains have smooth enough boundary, for example, satisfying the cone condition).

For differential forms, the trace operator carries an additional restriction (given by the pullback by the inclusion) in that only their operation on vectors tangent to the boundary needs to be considered. However, we also need a sharper form of the trace theorem which allows us to restrict  $H^s$  forms to  $H^{s-1/2}$  forms on the boundary (namely losing only half a degree of smoothness):

**1.3.2 Theorem** (Trace and Extension Theorems, [111], §I.8). *Let  $M$  be a Lipschitz manifold-with-boundary and  $s \geq \frac{1}{2}$ . Then there exists a bounded linear operator  $\text{Tr} : H^s\Omega^k(M) \rightarrow H^{s-1/2}\Omega^k(\partial M)$  such that for  $\omega \in \Omega^k(M)$ ,  $\text{Tr}\omega = i^*\omega$ , where  $i : \partial M \rightarrow M$  is the inclusion. Moreover, this operator is surjective, i.e. there exists a bounded linear inverse operator  $Z : H^{s-1/2}\Omega^k(\partial M) \rightarrow H^s\Omega^k(M)$  such that  $\text{Tr} Z\eta = \eta$ .*

However, we will need another space of  $k$ -forms,  $H\Omega^k(M)$  (with no superscript on the  $H$ ) which is in some sense more natural than the above definition. It is more natural for the simple reason that it takes into account the nature of the operator  $d$ , transcending its definition as some linear combination of partials (which is the viewpoint we have been stressing throughout this work). Indeed, we will see that  $H\Omega$  spaces contain forms that are generally less regular than those whose first weak partials all exist, namely the  $H^1\Omega$  spaces. This makes use of the Hodge duality and the codifferential operator  $\delta$ . We can use this to prove a version of Stokes' Theorem for

non-smooth forms as well (which will make use of extended trace theorems—see the next section—in order to define boundary restriction).

**1.3.3 Definition** (Weak derivatives and the Sobolev space  $H\Omega$ ).  $\omega \in \mathcal{L}^p\Omega^\ell(U)$  has a WEAK EXTERIOR DERIVATIVE  $\eta$  (which could be generally a current, i.e. linear functional on differential forms) if

$$\langle \omega, \delta\varphi \rangle = \langle \eta, \varphi \rangle$$

for all  $\varphi \in \Omega_c^{\ell+1}$  (note that according to our sign convention, making  $\delta$  the adjoint rather than the negative adjoint like for partial derivatives, we need no extra minus sign here). It necessarily is unique up to Lebesgue a.e. equivalence. If, additionally,  $\eta \in \mathcal{L}^p\Omega^{\ell+1}(U)$ , we say  $\omega \in W^p\Omega^\ell(U)$ . The space of greatest interest is actually when  $p = 2$ , for which we write  $H\Omega^\ell(U)$ —the space of all  $\mathcal{L}^2$  differential forms whose weak exterior differentials are also in  $\mathcal{L}^2$ . It is known [5, 6] that  $H\Omega^0(U)$  coincides with  $H^1(U)$  but in general, for  $\ell > 0$ ,  $H^1\Omega^\ell(U) \subsetneq H\Omega^\ell(U)$ . In fact, for forms of top degree,  $H\Omega^n(U) = \mathcal{L}^2\Omega^n(U)$ , since the exterior derivative of such forms is always zero (so, of course, it is trivial to generalize it to any degree of regularity we like). Similarly, we have  $\dot{H}\Omega^\ell(U)$  for the closure of  $\Omega_c^\ell(U)$  of forms vanishing on the boundary. We say such forms have vanishing TRACE; it will turn out that due to the progressively decreasing regularity with form degree, forms of vanishing trace become less and less restrictive class.

The spaces  $H\Omega^\ell$  are endowed with the GRAPH INNER PRODUCT

$$(1.3.1) \quad \langle \omega, \eta \rangle_{H\Omega} = \langle \omega, \eta \rangle_{\mathcal{L}^2} + \langle d\omega, d\eta \rangle_{\mathcal{L}^2}$$

(but recall that we still heavily rely on the  $\mathcal{L}^2$  inner product even when dealing with these spaces), and its corresponding graph norm  $\|\cdot\|_{H\Omega}$ . Of course, these spaces are complete:

**1.3.4 Theorem** (Completeness of  $H\Omega$  spaces [5, 6]).  $H\Omega^\ell(U)$  is complete in the norm defined by the graph inner product. This, in particular, makes  $d$  a closed operator (in the sense of functional analysis [34, 92, 97]). Moreover, smooth forms are dense in  $H\Omega^\ell(U)$ .

*Proof.* If  $\omega_n$  is Cauchy in the graph inner product, then both  $\omega_n$  and  $d\omega_n$  are Cauchy in  $\mathcal{L}^2$ . By the completeness of the respective  $\mathcal{L}^2$  spaces, they converge to  $\ell$ - and  $(\ell + 1)$ -forms  $\omega$  and  $\zeta$ , respectively. We only need to check  $\zeta$  is actually the weak exterior derivative. We simply recall that inner products are continuous with respect to the norms, so limits can be taken out of them:

$$\langle \zeta, \eta \rangle_{\mathcal{L}^2\Omega^{\ell+1}(U)} = \lim_{n \rightarrow \infty} \langle d\omega_n, \eta \rangle_{\mathcal{L}^2\Omega^{\ell+1}(U)} = \lim_{n \rightarrow \infty} \langle \omega_n, \delta\eta \rangle_{\mathcal{L}^2\Omega^\ell(U)} = \langle \omega, \delta\eta \rangle_{\mathcal{L}^2\Omega^\ell(U)}.$$

establishing that  $\zeta = d\omega$ . This, in particular, illustrates the power of the abstract Hilbert space approach: the raw materials of real analysis, with issues like integration and convergence, are neatly hidden under the umbrella in basic Hilbert space operations.  $\square$

Finally, we also define a Hodge dual version of the above spaces—these are not an entirely trivial definition, because, as we have observed (but not proved),  $H\Omega^\ell(U)$  gets progressively less regular as  $\ell$  increases: We define  $H^*\Omega^\ell(U) := \star H\Omega^{n-\ell}(U)$ , and  $\mathring{H}^*\Omega^\ell(U) := \star \mathring{H}\Omega^{n-\ell}(U)$  (in particular it does *not* mean their trace vanishes, but rather the trace of their *Hodge duals* vanish)<sup>1</sup>. We will have more to say about this in §1.5; but one can appreciate the difference between these two types of forms by looking at Figure 1.5 in that section. These spaces are important as the proper functional-analytic domain of the codifferential operator  $\delta$ : a function has a weak exterior coderivative precisely when its Hodge dual has a weak exterior derivative.

<sup>1</sup>For non-orientable manifolds, we should note the change in parity here: we notate the space  $H^*\Omega^\ell(U)_\psi$ .

For completeness (and that we need at least the definitions to state some important theorems on traces), we give these same definitions for distributions (these in turn allow us to define fractional-order Sobolev spaces via the Fourier transform [34, 64]).

**1.3.5 Definition.** We consider the partial and exterior derivatives of distributions and currents (the DISTRIBUTIONAL DERIVATIVE) to be defined by

$$(1.3.2) \quad (D^\alpha T)(\varphi) := \langle D^\alpha T, \varphi \rangle := \langle T, (-1)^{|\alpha|} D^\alpha \varphi \rangle = T((-1)^{|\alpha|} D^\alpha \varphi)$$

$$(1.3.3) \quad (dT)(\varphi) := \langle dT, \varphi \rangle := \langle T, \delta\varphi \rangle = T(\delta\varphi),$$

for functions and forms of compatible degree (namely,  $dT$  acts on forms of degree  $k+1$  if  $T$  acts on forms of degree  $k$ ). This makes the reason for choosing such notation pretty obvious. The difference between these kinds of derivatives and weak derivatives in Sobolev spaces is that the weak derivative of a function in  $\mathcal{L}_{\text{loc}}^1$  need not actually also be a function; it is when both a function *and* its distributional derivative lie in  $\mathcal{L}^p$  that we can say it is in the appropriate Sobolev space.

We use this to define the FOURIER TRANSFORM of distributions on  $\mathbb{R}^n$  (actually, this requires a slightly restricted class of distributions, called TEMPERED DISTRIBUTIONS, that extend to the Schwartz space of functions  $\mathcal{S}$ , functions which do not necessarily have compact support, but rather, vanish quickly at infinity along with all their derivatives [98, 99, 34]); we use the Fourier transform with the  $2\pi i$  in the exponent, following [98, 34, 64]:

$$(1.3.4) \quad \langle \hat{T}, \varphi \rangle := \langle T, \hat{\varphi} \rangle; \quad \hat{\varphi}(\xi) := \int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot x} \varphi(x) dx.$$



and multiplication by smooth functions with the appropriate growth conditions at infinity (in order to preserve the Schwartz space):

$$\langle \psi T, \varphi \rangle := \langle T, \psi \varphi \rangle.$$

This finally enables us to define the FRACTIONAL- and NEGATIVE-ORDER SOBOLEV SPACES: for  $s \in \mathbb{R}$ ,

$$H^s(\mathbb{R}^n) := \{T \text{ a tempered distribution} : (1 + 4\pi^2|\xi|^2)^{s/2} \hat{T} \in \mathcal{L}^2(\mathbb{R}^n)\}$$

(where we have used the variable  $\xi$  in the Fourier transform space). It should be noted that for  $\langle \cdot, \cdot \rangle$  denoting an extended kind of  $\mathcal{L}^2$  inner product discussed in more detail in Remark 1.9.5, we have that  $H^{-s}$  pairs with  $H^s$  in this way; this is easily verified by inserting the factors  $(1 + 4\pi^2|\xi|^2)$  raised to the appropriately oppositely signed powers. It, of course, also makes use of Plancherel's Theorem [34, 64] which says the Fourier transform (as we've defined it with the  $2\pi i$  in the exponent) preserves  $\mathcal{L}^2$  inner products. For domains satisfying nice properties, such as the uniform cone condition, they coincide with Slobodeckii spaces [111, §I.3], which are the  $\mathcal{L}^2$  analogue of the Hölder spaces.

## 1.4 The Extended Trace Theorem

If we are going to consider boundary value problems involving differential forms, we need some results on how to actually assign such boundary values. As noted before, since boundaries of compact Lipschitz manifolds-with-boundary (the domains of interest here) have measure zero, it does not make sense, from the standpoint of Lebesgue measure and integration, to restrict anything to such a boundary—any

function can be modified on a set of measure zero without affecting integrals. However, trace theorems (like Theorem 1.3.2 above) guarantee that it makes sense for functions in certain circumstances, namely for a high enough order Sobolev space. Roughly speaking, enough weak derivatives imply some of those derivatives become classical; the trace theorems are suitable generalizations of the Sobolev Embedding theorem [30, §§5.6-7]. By the corresponding theorems for functions, we immediately have the trace theorems for  $H^s\Omega$ . We now want a version of the trace theorem to work with  $H\Omega$ , which, recall, treats the exterior derivative as an organic whole, rather than a particular combination of partials. It turns out that we can use a dualization argument to apply the  $H^s\Omega$  theory to give us the theory for  $H\Omega$ .

**1.4.1 Theorem** (Extended Trace Theorem; Arnold, Falk, and Winther [5], p. 19). *Let  $U$  be a domain in  $\mathbb{R}^n$  with Lipschitz boundary. Then there exists a bounded linear map*

$$\text{Tr}: H\Omega^k(U) \rightarrow H^{-1/2}\Omega^k(\partial U)$$

*such that for all  $\omega \in \Omega^k(U)$ ,  $\text{Tr}\omega = i^*\omega$ , where  $i: \partial U \rightarrow U$  is the inclusion map.*

Recall that  $H^{-1/2}\Omega^k$  consists of  $k$ -currents that act on  $H^{1/2}\Omega^k$ , Sobolev forms of regularity  $1/2$ , and smooth forms act by the  $\mathcal{L}^2$  inner product in the inherited metric. Note, however, this extension is not surjective; this can be seen by realizing that for  $k=0$ ,  $H\Omega = H^1\Omega$ , so it is clear that not every such  $H^{-1/2}$  boundary function can be the trace of something. In order to show this proof, we need an extension of Stokes' Theorem.

**1.4.2 Theorem** (Stokes' Theorem for  $H^1\Omega$  forms [5], pp. 17-19). *Let  $\omega \in H^1\Omega^{n-1}(U)$ . Then*

$$(1.4.1) \quad \int_U d\omega = \int_{\partial U} \text{Tr}\omega.$$

*Proof.* We approximate the form  $\omega$  in the  $H^1$  norm by  $C^\infty$  differential forms  $\omega_m$ . Then  $d\omega_m \rightarrow d\omega$  in  $\mathcal{L}^2$  and  $\text{Tr}\omega_m \rightarrow \text{Tr}\omega$  in  $H^{1/2}$  (and in particular, in  $\mathcal{L}^2$ ). Therefore (recalling that the volume form is always  $\star 1$  in  $\mathcal{L}^2$  for a compact manifold-with-boundary, so integration on a manifold with any Riemannian metric can conveniently be represented as integrating against  $\star 1$ ),

$$\langle d\omega_m, \star 1 \rangle_U = \int_U d\omega_m = \int_{\partial U} i^* \omega_m = \int_{\partial U} \text{Tr}\omega_m = \langle \text{Tr}\omega_m, \star_{\partial U} 1 \rangle_{\partial U},$$

where we have written  $\star_{\partial U}$  for the Hodge star on the boundary with inherited metric. Therefore, taking the limit of both sides (as the  $\mathcal{L}^2$  inner products are continuous in the  $\mathcal{L}^2$  norms by definition), we get

$$\int_U d\omega = \langle d\omega, \star 1 \rangle_U = \langle \text{Tr}\omega, \star_{\partial U} 1 \rangle_{\partial U} = \int_{\partial U} \text{Tr}\omega.$$

□

Once we prove the extended trace theorem, the same proof above shows that Stokes' theorem holds for forms in  $H\Omega^{n-1}$  as well, except we replace convergence in  $\mathcal{L}^2$  for the  $\text{Tr}\omega_n$  with  $H^{-1/2}$ -convergence, and we can no longer necessarily interpret the latter as an integral (it will have to stay  $\langle \text{Tr}\omega, \star_{\partial U} 1 \rangle$ ).

Using this theorem and the product rule, we have two extensions of integration by parts, one for wedge products and one for inner products:

**1.4.3 Theorem** (Integration by Parts for forms). *Let  $\omega \in \Omega^k(U)$ ,  $\eta \in H\Omega^{n-k-1}(U)$ , and  $\xi \in H^1\Omega^{k+1}(U)$ . Then*

$$(1.4.2) \quad \langle d\omega, \xi \rangle = \langle \omega, \delta\xi \rangle + \int_{\partial U} \text{Tr}\omega \wedge \text{Tr}\star\xi$$

$$(1.4.3) \quad \int_U d\omega \wedge \eta = (-1)^{k+1} \int_U \omega \wedge d\eta + \int_{\partial U} \text{Tr}\omega \wedge \text{Tr}\eta.$$

(We of course extend this theorem later on for  $\omega \in H\Omega^k$  once we have the extended trace theorem, but we need this version to *prove* the extended trace theorem.)

*Proof.* Noting that  $H^1$ ,  $H^{1/2}$ , and  $\mathcal{L}^2$  functions are closed under multiplication by bounded, smooth functions, recalling the convenient commutation formula (1.2.3) for  $d$  and  $\delta$ , and using the Leibniz rule,

$$\begin{aligned} \langle dw, \xi \rangle &= \int_U d\omega \wedge \star \xi = \int_U d(\omega \wedge \star \xi) - (-1)^k \int_U \omega \wedge d(\star \xi) \\ &= \int_{\partial U} \text{Tr} \omega \wedge \text{Tr} \star \xi + \int_U \omega \wedge \star \delta \xi = \langle w, \delta \xi \rangle. \end{aligned}$$

□

*Proof of the extended trace theorem.* We need to show that given  $\omega \in H\Omega^k(U)$ , there exists a linear functional on  $H^{1/2}\Omega^k(\partial U)$  which reduces to the  $\mathcal{L}^2$  inner product by the trace of  $\omega$ , when  $\omega$  is smooth. We use a standard technique: prove that the relevant operators in the smooth case are bounded in the right norms, and use completeness to define an extension to the completion, which is all of  $H\Omega$  in this case. We follow the proof of [5], more directly using the inner product notation. Letting  $\omega$  be smooth up to the boundary, we consider the action of its trace on  $H^{1/2}\Omega^k(\partial U)$  by the  $\mathcal{L}^2$  inner product on  $\partial U$ . But if  $\xi$  is any form in  $H^{1/2}\Omega^k(\partial U)$ , then considering  $\rho = \star_{\partial U} \xi$ , for some  $\rho \in H^{1/2}\Omega^{n-k-1}(U)$ , then by the surjectivity of the trace operator (Theorem 1.3.2 above), there exists  $\eta = Z\rho \in H^1\Omega^{n-k-1}(U)$  such that  $\text{Tr} \eta = \rho$ , and moreover,  $\|\eta\|_{H^1\Omega^{n-k-1}(U)} \leq C' \|\rho\|_{H^{1/2}\Omega^{n-k-1}(\partial U)} \leq C \|\xi\|_{H^{1/2}\Omega^k(\partial U)}$ . This means

$$\begin{aligned} |\langle \text{Tr} \omega, \xi \rangle| &= \left| \int_{\partial U} \text{Tr} \omega \wedge \star_{\partial U} \rho \right| = \left| \int_U \text{Tr} \omega \wedge \text{Tr} \eta \right| \leq |\langle d\omega, \star \eta \rangle - \langle \omega, \delta(\star \eta) \rangle| \\ &\leq \|d\omega\| \|\star \eta\| + \|\omega\| \|\delta(\star \eta)\| \leq c \|\omega\|_{H\Omega^k(U)} \|\eta\|_{H^1\Omega^{n-k-1}(U)} \leq C \|\omega\|_{H\Omega^k(U)} \|\xi\|_{H^{1/2}\Omega^k(\partial U)}. \end{aligned}$$

(We used (1.4.2) to get the last term on the first line, and since the codifferential involves the componentwise weak partials along with the algebraic properties of the Hodge operator, the norm of the codifferential is bounded by that of the  $H^1$  norm.)

This shows that

$$\|\mathrm{Tr}\omega\|_{H^{-1/2}\Omega^k(\partial U)} \leq C\|\omega\|_{H\Omega^k(U)}$$

for all smooth forms  $\omega$ . If, now  $\omega$  is in  $H\Omega^k(U)$ , then it is the  $H\Omega$ -limit of some sequence of smooth forms  $\omega_n$ , and the boundedness in the right norms ensures that  $\mathrm{Tr}\omega_n$  is Cauchy in  $H^{-1/2}\Omega$ ; we let  $\mathrm{Tr}\omega$  be the limit, and the operator is bounded.  $\square$

By taking the limit of a sequence of smooth forms  $\omega_n$  and using that their traces converge in  $H^{-1/2}$  by the above theorem, we immediately have the following

**1.4.4 Corollary.** *The formulæ (1.4.2) and (1.4.3) continue to hold for  $\omega \in H\Omega^k(U)$ .*

We now can define the  $H\Omega$  forms of vanishing trace:

$$\mathring{H}\Omega^k(U) := \{\omega \in H\Omega^k(U) : \mathrm{Tr}\omega = 0\}.$$

We end with an application of this extended theorem to identify the adjoint  $d^*$  of the exterior differential  $d$  and its domain (in the full functional analytic sense [112, Ch. VII, §2]). It shows that the notion of duality and the Sobolev space equivalent of compact support—namely, having vanishing trace—are intertwined. We follow [6, §4.2].

**1.4.5 Theorem** (Arnold, Falk, Winther [6], Theorem 4.1). *Consider the space  $\mathcal{L}^2\Omega^k(U)$ , the space of all  $\mathcal{L}^2$  forms with the  $\mathcal{L}^2$  inner product. Then the weak exterior derivative operator  $d$  is an unbounded operator defined on a dense domain  $H\Omega^k(U)$  but in fact has closed graph (is a closed operator). Then there exists an adjoint operator  $d^*$*

defined on the domain  $\mathring{H}^* \Omega^k(U)$  (which, recall, is the space of Hodge duals to forms in  $\mathring{H} \Omega^{n-k}(U)$ ), and it is in fact the codifferential operator  $\delta$ .

*Proof.*  $d$  is by definition defined on all of  $H \Omega^k(U)$ , a space certainly dense in all of  $\mathcal{L}^2 \Omega^k(U)$ , since even smooth forms of compact support are ( $\mathcal{L}^2$ -)dense in  $\mathcal{L}^2 \Omega^k(U)$  (they are not, of course,  $H \Omega$ -dense in  $H \Omega^k(U)$ , however, but rather  $\mathring{H} \Omega^k(U)$ ). Thus, the adjoint operator  $d^*$  exists, and has a dense domain in  $\mathcal{L}^2 \Omega^k(U)$ .

Now, given  $\eta \in \mathring{H}^* \Omega^k(U)$ , we have for all  $\omega \in \Omega^{k-1}(U)$ , by (1.4.2),

$$\langle \omega, \delta \eta \rangle = \langle d \omega, \eta \rangle - \int_{\partial U} \text{Tr } \omega \wedge \text{Tr } \star \eta = \langle d \omega, \eta \rangle$$

(interpreting the integral as the action of  $\text{Tr } \star \eta$  as an operator on  $H^{1/2}$ , if necessary), since the trace of  $\star \eta$  is zero. Since smooth forms are dense, this establishes that  $\mathring{H}^* \Omega^k(U)$  is contained in the domain of  $d^*$ , and  $d^* = \delta$  there. On the other hand, if  $\eta$  is in the domain of the adjoint, then  $d^* \eta \in \mathcal{L}^2 \Omega^{k-1}(U)$  and by definition of the adjoint, for all  $\omega \in \Omega^{k-1}(U)$ ,

$$\langle \omega, d^* \eta \rangle = \langle d \omega, \eta \rangle.$$

This holds true, in particular, for forms  $\omega$  with compact support, so by the distribution definition of the weak exterior coderivative,  $d^* \eta = \delta \eta$ . This establishes that  $\eta \in H^* \Omega^k(U)$ . However,  $\delta \eta$  continues to follow that identity even for  $\omega$  not being of vanishing trace. Thus by (1.4.2), we have

$$\int_{\partial U} \text{Tr } \omega \wedge \text{Tr } \star \eta = \langle d \omega, \eta \rangle - \langle \omega, \delta \eta \rangle = 0,$$

which shows that  $\text{Tr } \star \eta$  vanishes as an operator on  $H^{1/2}$  (really, on a dense subspace). By the surjectivity of the trace operator, this means  $\text{Tr } \star \eta = 0$ , and thus,  $\eta \in \mathring{H}^* \Omega^k(U)$ .

□

## 1.5 Boundary Value Problems with the Hodge Laplacian

Having detailed differential forms, we now present a full recasting of some standard, classical BVPS in terms of them. The Hodge-theoretic formulation provides a complete story for many classical boundary value problems. We follow the development of Arnold, Falk, and Winther [6, §4.2 and §§6.1-2]. In addition, with the theory of weak solutions to come, we can pose a weak formulations of the problems, which sets things up for approximation via finite element methods (Chapter 2).

**1.5.1 Definition.** We recall, for  $\omega \in \Omega^k(M)$ ,

$$\Delta\omega := -(\delta d + d\delta)\omega.$$

A HARMONIC FORM is a form  $\omega$  such that  $\Delta\omega = 0$ ; this space is denoted  $\mathfrak{H}^k(M)$ . For greater precision, however, we should actually specify the domain of  $\Delta$ . Boundary conditions must be used to restrict the domain of  $\Delta$ , since, as observed above, the operator will no longer be an adjoint operator (due to the resulting extra boundary terms) without such a restriction. Since we must take  $d$  of  $\omega$ , we must have  $\omega$  at least be in  $H\Omega^k$ , and similarly since we must also take  $\delta$  of  $\omega$ , it at least must be in  $\mathring{H}^*\Omega^k$ . But it also must land in the domain of the other operator; in short, the proper domain is

$$D(-\Delta) = d^{-1}(\mathring{H}^*\Omega^{k+1}) \cap \delta^{-1}(H\Omega^{k-1}).$$

This allows us to formulate the following boundary value problem:

**1.5.2 The classic Hodge Laplacian boundary value problem for differential forms.**

The STANDARD HODGE LAPLACIAN BOUNDARY VALUE PROBLEM for  $\Delta$  is the problem

$$(1.5.1) \quad \begin{cases} -\Delta\omega = \eta \\ \text{Tr}_{\partial M}(\star\omega) = 0 \\ \text{Tr}_{\partial M}(\star d\omega) = 0 \end{cases}$$

for some given inhomogeneous (interior source) term  $\eta$ . We note that the boundary of a manifold-with-boundary is canonically transversely oriented by an outward normal, the normal  $\mathbf{n}$  to  $\partial M$  such that any curve approaching the boundary has a tangent vector making an acute angle with (having a positive dot product with)  $\mathbf{n}$ , so it makes sense to pull back the pseudoforms  $\star\omega$  and  $\star d\omega$ . As we saw previously in Theorem 1.4.5, the reason for the boundary conditions is because only in that case is  $\delta$  the adjoint of  $d$  relative to the inner products (recall that the adjoint of  $d$  on  $H\Omega^k$  is  $\delta$  restricted to  $\mathring{H}^*\Omega^k(M)$ ), so when we pass to the weak formulation, we have no boundary terms, and results from functional analysis are applicable.

We recall that  $\omega$  is a classical solution if it actually satisfies the above equations, using classical partial derivatives. If we interpret the derivatives as weak, we get what Gilbarg and Trudinger [39] call a STRONG SOLUTION, which is at first confusing because we are still using weak derivatives. What is called a WEAK SOLUTION is even weaker, because we use integration by parts (or adjoints in the inner product) to get expressions that may yield results that, *a priori*, could be outside the domain of  $\Delta$ . Again, this is no different from finding weak solutions for elliptic operators on functions in  $H^1$ , despite elliptic operators often needing the functions to be in  $H^2$  to literally be defined with weak derivatives ([39, 30]). So, we want to say  $\omega \in \mathring{H}\Omega^k(M)$  is a weak solution to the homogeneous problem if we have, for all  $v$  in the appropriate function space,

$$\langle d\omega, dv \rangle + \langle \delta\omega, \delta v \rangle = \langle \eta, v \rangle.$$



This is simply integrating it against a test form, and moving the  $d$ 's and  $\delta$ 's around. In fact, for convenience, it is common to *define* the operator  $-\Delta$  to map into the dual space  $(H\Omega^k)'$  by defining its action on test forms to be exactly the above, so that notationally things carry over identically. We must be careful, however, to not assume more of  $-\Delta$  and about what it is operating on, when we use the extended notation; consequently we try write things in explicit weak form as much as possible. In other words, we try to make things make sense even if  $\eta$  is a current. There actually are problems with this formulation (even in the case where everything is smooth): the harmonic forms are an obstruction to both existence and uniqueness. In addition, numerical methods based on this principle, for all but the easiest examples, are not stable [6, §2.3].

**1.5.3 How to allow for inhomogeneous boundary conditions.** In analogy to the theory for functions [30, Ch. 6], we can allow nonzero traces to the boundary of both  $\star\omega$  and  $\star d\omega$ , by simply using the (inverse) trace theorems (Theorem 1.3.2) above to extend the boundary forms to a form defined on all of the domain  $U$ , and modifying the *interior* inhomogeneous term ( $\eta$  in the above), to get a problem with homogeneous boundary conditions. We will say more about this in the next section on the theory of weak solutions.

**1.5.4 Problems with well-posedness.** As stated previously, the most directly stated boundary value problem for  $\Delta$  is not well-posed. To rectify this, we use another weak formulation (called the MIXED WEAK FORMULATION). This is motivated by recasting it as a *system* of first-order equations (mixed formulations are generally a useful technique and are covered in more generality in, e.g., [11, Ch. III], [54, Ch. 4], and [12, Ch. 12]). So suppose, for the moment, we define  $\sigma = \delta\omega$ . The weak formulation of this is

$\langle \sigma, \tau \rangle = \langle \omega, d\tau \rangle$  for all  $\tau \in H\Omega^{k-1}$ . Now we try to solve

$$d\sigma + \delta d\omega = \eta,$$

by moving things to the other side. Here we have  $\langle d\sigma, v \rangle + \langle d\omega, dv \rangle = \langle \eta, v \rangle$  for all  $v$ . But, a necessary condition for a solution to exist is that  $\langle \eta, h \rangle = 0$  for all harmonic forms  $h$ . This is because  $\langle \eta, h \rangle = \langle d\sigma, h \rangle + \langle d\omega, dh \rangle = \langle \sigma, \delta h \rangle = 0$  since both  $dh$  and  $\delta h$  vanish. To get around this, we orthogonally project  $\eta$  onto the harmonic forms, taking  $p$  to be that projection, and instead solve  $\langle d\sigma, v \rangle + \langle d\omega, dv \rangle + \langle p, v \rangle = \langle \eta, v \rangle$  so that  $\langle \eta - p, h \rangle = 0$  on all harmonic forms. Finally, because  $\Delta$  usually has a nontrivial kernel (the harmonic forms), we want to choose a unique solution. This can be done by constraining  $\omega$  to be orthogonal to the harmonic forms, namely  $\langle \omega, q \rangle = 0$  for all harmonic  $q$ .

Thus we arrive at the MIXED WEAK FORMULATION OF THE PROBLEM FOR THE HODGE LAPLACIAN (with vanishing traces) [6, §3.2], which is finding a solution

$$(\sigma, \omega, p) \in H\Omega^{k-1} \times H\Omega^k \times \mathfrak{H}^k$$

such that

$$(1.5.2) \quad \left\{ \begin{array}{ll} \langle \sigma, \tau \rangle - \langle \omega, d\tau \rangle = 0 & \forall \tau \in H\Omega^{k-1}(M) \\ \langle d\sigma, v \rangle + \langle d\omega, dv \rangle + \langle p, v \rangle = \langle \eta, v \rangle & \forall v \in H\Omega^k(M) \\ \langle \omega, q \rangle = 0 & \forall q \in \mathfrak{H}^k(M), \end{array} \right.$$

where all the inner products are taken relative to the  $\mathcal{L}^2\Omega$  inner products restricted to the  $H\Omega$ 's (and not the  $H\Omega$  inner products, which are more useful in estimates). The analogous problem for pseudoforms can also be posed; and indeed these versions

are extremely useful in higher degree forms such as those dealing with flux and mass. Note also that in this formulation, there are no  $\delta$ 's, and we do not directly deal with any spaces of the form  $\mathring{H}^* \Omega$  (we will see what this means when we try to fit the Dirichlet problem in to this framework). Nevertheless, the solutions *are* in fact in  $\mathring{H}^* \Omega$ , because both  $u$  and  $du$  satisfy the defining condition of having a weak coderivative (the first and second equations both have terms comparing it against  $d$  of something), and  $d^*$  has been established to have a domain  $\mathring{H}^* \Omega^k(U)$  (Theorem 1.4.5 above). The defining boundary conditions of the space  $\mathring{H}^* \Omega$  (namely  $\text{Tr} \star u = 0$  and  $\text{Tr} \star du = 0$ ) corresponds to the notion of *natural* boundary conditions, because they are enforced via Stokes' Theorem, and are not explicitly incorporated in the definition of the spaces directly used in the problem (1.5.2). It is often useful to think of  $\eta$  as a current, in which we do not yet know its Riesz representative, analogous to the spaces  $H^{-1}(M)$  in the theory for functions (we get to this in the weak solution theory; the details are in [30, Ch. 6]).

With these additional fixes, we have that the mixed weak formulation is well-posed [6] (the use of a bilinear form is also key in the weak solution theory):

**1.5.5 Theorem** (Arnold, Falk, Winther [6], Theorem 3.1). *Consider the mixed formulation above for  $(\sigma, \omega, p) \in H\Omega^{k-1}(M) \times H\Omega^k(M) \times \mathfrak{H}^k(M)$ . We consider the bilinear form (using  $\mathcal{L}^2 \Omega$  inner products)*

$$B(\sigma, \omega, p; \tau, v, q) := \langle \sigma, \tau \rangle - \langle \omega, d\tau \rangle + \langle d\sigma, v \rangle + \langle d\omega, dv \rangle + \langle p, v \rangle - \langle \omega, q \rangle.$$

*Then there exists a unique triplet  $(\sigma, \omega, p)$  such that  $B(\sigma, \omega, p; \tau, v, q) = (\eta, v)$  for all triplets  $(\tau, v, q) \in H\Omega^{k-1} \times H\Omega^k \times \mathfrak{H}^k$ . Moreover, we have the following a priori estimate:*

$$\|\sigma\|_{H\Omega^{k-1}} + \|\omega\|_{H\Omega^k} + \|p\|_{\mathcal{L}^2 \Omega^k} \leq C \|\eta\|_{\mathcal{L}^2 \Omega^k}$$

for some  $C$  depending only on the Poincaré constant  $c_P$  such that

$$\|\xi\|_{H\Omega^k} \leq c_P \|d\xi\|_{\mathcal{L}^2\Omega^{k+1}}$$

for all  $\xi$  orthogonal to cocycles; this holds true for functions vanishing on the boundary ([30, §5.6.1, Theorem 3]), which shows the solution depends continuously on the data.

The idea of the Poincaré inequality, as we have stressed in the introduction, is the key result that makes both the well-posedness and the numerical approximations work, and so we seek its generalization in §1.8. We now fit things into the existing framework (as detailed in [6, §4.2]).

**1.5.6 The Neumann Problem.** We consider the (strong) problem for  $k = 0$ , for a function  $u = \omega \in H\Omega^0(M) = H^1(M)$  and inhomogeneous term  $f = \eta \in \mathcal{L}^2$ . Note that it has vanishing weak coderivative, so all references to  $\sigma$  can be omitted. Now,  $\text{Tr}(\star u)$  is the trace of an  $n$ -pseudoform on the boundary, an  $n - 1$  dimensional manifold, so it vanishes. On the other hand,  $\text{Tr}(\star du)$  is interesting—we have that  $\star du = \nabla u \lrcorner dV_g$ , and orthogonally decomposing  $\nabla u = \nabla u^t + (\nabla u \cdot \mathbf{n})\mathbf{n}$  ( $\mathbf{n}$  the unit normal),

$$\text{Tr}(\star du) = \text{Tr}(\nabla u^t \lrcorner dV_g) + (\nabla u \cdot \mathbf{n}) \text{Tr}(\mathbf{n} \lrcorner dV_g) = \nabla u \cdot \mathbf{n} dS,$$

where  $dS$  the element of surface area on  $\partial M$ . Note that the tangential term vanishes, because its trace is a form that accepts  $n - 1$  vectors tangent to  $\partial M$ , and the interior product puts one more vector tangent to  $\partial M$ , thus it is a form that is evaluated on a linearly dependent set of vectors. This says that the normal derivative of  $u$ ,  $\frac{\partial u}{\partial \mathbf{n}}$ , vanishes. Finally, harmonic functions are constant. So we have the weak formulation

$$\langle du, dv \rangle = \langle f - p, v \rangle$$

for a function  $u$  of vanishing integral in  $H^1(M)$ , and where  $p$  is the orthogonal projection of  $f$  onto the constants. Thus the mixed formulation simply reduces to the standard theory for functions.

If we recast this as a minimization problem, namely, we try to find a form  $u$  minimizing

$$I(u) := \frac{1}{2} \|du\|^2 - fu$$

with the constraint  $\int u dV_g = 0$ , we actually find that the function  $p$  found above is the Lagrange multiplier.

**1.5.7 Pseudoinverse of the Gradient.** Given a 1-form  $\beta$ , can we find a function  $u$  such that  $du = \beta$ ? This is usually impossible, namely if the manifold is not simply connected and  $\beta$  represents a nontrivial cohomology class, but if we solve it in the LEAST SQUARES sense, we will get  $\delta du = \delta\beta$ , which is precisely the Neumann problem. Harmonic forms are isomorphic to the first cohomology, so the presence of simple connectivity in this problem is not a coincidence.

**1.5.8 The Dirichlet Problem.** We can formulate the de Rham complex *with boundary conditions* which is described in [6, §6.2] or Example 1.8.3 below; we can simply incorporate the boundary conditions directly, and much of the same arguments follow (we introduce the abstract Hilbert complex approach in §1.8 precisely to capture such properties that make the arguments work). However, it is, surprisingly, possible to include a discussion of the Dirichlet problem with *natural* boundary conditions: instead of seeking a function, let us seek a top degree,  $n$ -pseudoform  $\omega$ . Then the problem is  $\Delta\omega = \eta$ . Now  $\text{Tr}(\star d\omega) = 0$  automatically, because  $d\omega$  is an  $(n+1)$ -pseudoform, which always vanishes. But, writing  $u = \star\omega$  (a plain, not pseudo-) function,  $\text{Tr}(\star\omega) = u|_{\partial M} = 0$ . Because  $\Delta$  commutes with  $\star$  and  $\star$  is an isomorphism,  $\Delta\omega = \eta$  is equivalent to  $\Delta(\star\omega) = \Delta u = \star\eta$ . Writing  $f = \star\eta$ , we have then this is the

(strong) Dirichlet problem  $\Delta u = f$  and  $u|_{\partial M} = 0$ . As for the mixed weak formulation, though, we have that  $\sigma$  is no longer trivial. However, the harmonic  $n$ -forms are trivial, since we require compact support for the domain of  $\delta$ . Thus we have the problem

$$\langle \sigma, \tau \rangle - \langle \omega, d\tau \rangle = 0$$

for all  $\tau \in H\Omega_{\psi}^{n-1}(M)$  and

$$\langle d\sigma, v \rangle = \langle f, v \rangle$$

for all  $v \in H\Omega_{\psi}^n(M) = \mathcal{L}^2\Omega_{\psi}^n(M)$ . Taking duals, we find that we are actually solving for  $u \in \mathcal{L}^2$ , that is, the solution to the mixed weak formulation of the Dirichlet problem is possibly even less regular than the usual weak formulation of the Dirichlet problem, given, e.g., in [30, §6.1]. Since there are no explicit  $\delta$ 's or spaces  $\mathring{H}^*\Omega_{\psi}^n(M) = \star H_0^1(M)$ , this means that we need not restrict our test functions to those that vanish on the boundary. So although we work with the spaces  $\mathcal{L}^2$  and  $H\Omega^{n-1}$ , the boundary conditions are somehow incorporated in the structure of the inner products and weak form itself, i.e., they are *natural*. Of course,  $u$  may actually have much higher regularity (in fact it does, by standard elliptic regularity theory, at least if  $M$  is a smooth manifold and the boundary is smooth), but that fact is not, *a priori*, necessary.

We should note that seeking an  $n$ -pseudoform version is not artificial, because in the traditional formulation of the Dirichlet problem, the unknown function often represents the *concentration* of something. So to get the actual quantity of that something, one must integrate it over a volume, that is, we really seek an  $n$ -pseudoform (in the terminology of Frankel [36]).

**1.5.9 Example** (Fluid Flows). Consider the problem for  $n = 3$  and  $k = 2$ . Given a 2-pseudoform  $\omega$ , there exists a unique vector field  $\mathbf{u}$  such that  $\mathbf{u} \lrcorner dV = \omega$  (See [5, Table 2.1] for a reference on the different correspondences of vector fields to differential

forms in  $\mathbb{R}^3$ —such vector fields are called VECTOR PROXY FIELDS [5, p.26]). Thus 2-pseudofields correspond to velocity fields of fluids with uniform density. More generally, for a fluid of nonuniform density, we recall the momentum density field  $\rho \mathbf{v} \lrcorner dV$  is the interior product of the velocity field with a mass pseudoform  $\rho dV$ , or interior product of the momentum density vector field  $\rho \mathbf{u}$  with the volume form (the former description is the most natural one).

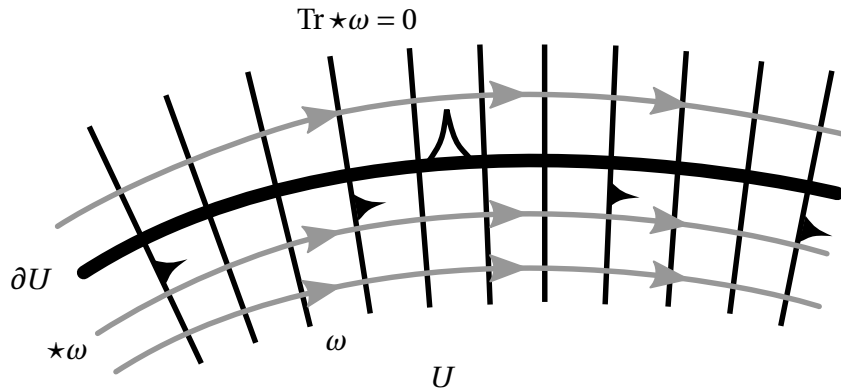
So the strong form of the problem is  $-\Delta\omega = \eta$ . In vector calculus notation,

$$\Delta\omega = \text{curl curl } \mathbf{u} - \text{grad div } \mathbf{u}.$$

In much of the literature, the vector calculus equivalents of  $H\Omega^1(M)$  and  $H\Omega^2(M)$  are, respectively, the classical Sobolev spaces  $H(\text{curl}; M)$  and  $H(\text{div}; M)$ . As for the boundary conditions, we have  $\text{Tr}(\star\omega) = 0$ , which says the corresponding 1-form vanishes on vectors tangent to the boundary. This says the corresponding velocity vector field is perpendicular to the boundary (usually written  $\mathbf{u} \times \mathbf{n} = 0$ ), or its tangential components vanish.

In terms of Weinreich's pictures [108], we form the Hodge dual by taking the sheets of a 1-form (the representation of a 1-form by level sets) so that the given 2-form (represented as field lines [108]) threads through it perpendicularly, and in the same direction, with magnitude made such that we once again have the volume pseudoform. To say that this 1-form vanishes at the boundary means any vectors tangent to the boundary vanish on it: the sheets of the 1-form are contained in the tangent space. Thus, again, we see the tangential component of the proxy vector field vanishes.  $\text{Tr}(\star d\omega) = 0$  means the divergence vanishes at the boundary in a very ordinary sense, namely, restriction of the function to the boundary is zero.

**1.5.10 Example** (The dual of a flow and equipotentials). Now, we examine the problem



**Figure 1.4:** A 1-form  $\omega$  (thin black level sets) whose hodge dual  $\star\omega$  (gray field lines) has vanishing trace on the boundary  $\partial U$ . This says the field lines of  $\star\omega$  are tangent to  $\partial U$ .

for  $n = 3$  and  $k = 1$ , this time choosing to solve for the momentum density as a 1-form, namely taking  $\omega = \mathbf{u}^\flat$  (i.e., the unique 1-form  $\omega$  such that the evaluation  $\omega(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$  for all vector fields  $\mathbf{v}$ , which is an isomorphism), rather than  $\mathbf{u} \lrcorner dV$ . Under this different identification, we find that the Laplacian still is  $\text{curl curl } \mathbf{u} - \text{grad div } \mathbf{u}$ , but the correspondence of operators switches  $d$  and  $\delta$  (namely,  $d$  on 2-forms and  $\delta$  on 1-forms correspond to  $\text{div}$ , and  $d$  on 1-forms and  $\delta$  on 2-forms correspond to  $\text{curl}$ , possibly with sign differences). Then  $\text{Tr}(\star\omega)$  is pulling the 2-pseudoform version of  $\omega$  back to the boundary, and its vanishing implies that  $\star\omega$  vanishes on pairs of vectors tangent to the boundary. This says that the fluxes of the material flow represented by  $\star\omega$  through all infinitesimal pieces of (transversely oriented) boundary are zero.

In more traditional vector calculus terms, now  $\star\omega$  is  $\mathbf{u} \lrcorner dV$ , so this means for any two vectors  $\mathbf{v}, \mathbf{w}$  tangent to  $\partial M$ ,  $0 = \star\omega(\mathbf{v}, \mathbf{w}) = dV(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ , that is, the parallelepiped they span is degenerate. In other words,  $\mathbf{u}$  is tangent to  $\partial M$  as well. So the vector calculus notation version of the boundary condition is  $\mathbf{u} \cdot \mathbf{n} = 0$ .

To see this in terms of Weinreich's visualizations, the procedure is to consider the threads of a 2-pseudoform to run perpendicularly through the sheets of the representative stack, in such a manner such that the density of the points of their



intersection represents the volume pseudoform (called a “swarm” by Weinreich [108]). Vanishing trace means they are tangent to the boundary, so therefore the original vector field was also tangent to the boundary, meaning, once again, its normal component vanishes. Finally, since  $d\omega$  is a 2-form,  $\text{Tr}(\star d\omega) = 0$  is simply  $(\star d\omega)^\sharp \times \mathbf{n} = 0$  (where  $\sharp$  is the inverse of the isomorphism  $\flat$  in Example 1.5.10), as in the previous example, or, traditionally,  $\text{curl } \mathbf{u} \times \mathbf{n} = 0$ .

The 1-form picture also is naturally encountered in electrostatics and other circumstances as force fields, and the surfaces defined by the 1-form are equipotentials.

**1.5.11 Example** (Flows in the complex plane, [74], Ch. 12). In the complex plane, the previous two examples are related via the notion of harmonic conjugate [74, Ch. 12]. The Cauchy-Riemann equations [3, 74] for holomorphic  $f = u + iv$  are

$$(1.5.3) \quad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$

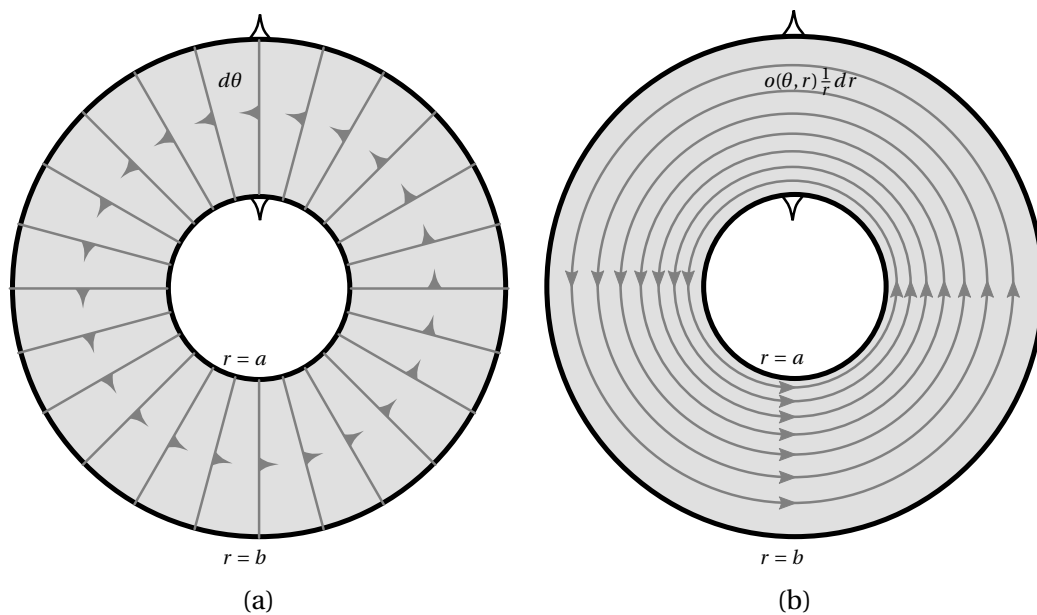
$$(1.5.4) \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x},$$

are invariantly stated as  $dv = \star du$  (where the orientation is specified by  $i$  being a rotation by  $\pi/2$  counterclockwise). It is, nevertheless, better to keep the pseudoform picture to keep things straight, i.e., we let one of the functions (say,  $du$ ) represent a collection of equipotentials, while  $dv$  should represent streamlines. This means that the real and imaginary parts of a holomorphic function contain the same information, but simply present themselves differently; in applications, usually one will be more natural than the other. See Figure 1.5 for an example on an annulus; here  $v = \log r$  and  $u = \theta$  (which is only an analytic function on the annulus minus a segment—but note that the 1-form is well-defined and smooth in the whole annulus). The two “functions” are HARMONIC CONJUGATES, and that they are both closed forms means there are no sources in the annulus.

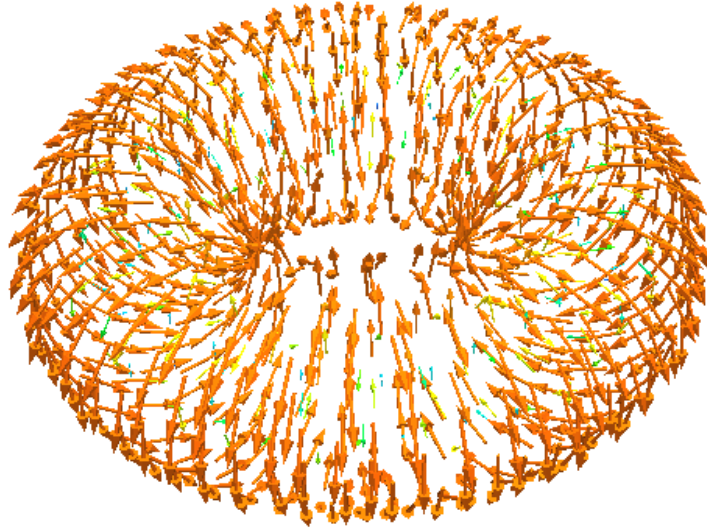
**1.5.12 Harmonic forms.** Harmonic forms are the kernel of the operator  $-\Delta$ , and by considering the equation  $\langle -\Delta\omega, \omega \rangle = 0$  in its weak formulation, this implies  $d\omega = 0$  and  $\delta\omega = 0$ . For manifolds with boundary, boundary conditions can profoundly influence what kind of solutions we can have (in similar analogy to the case for functions). The harmonic space  $\mathfrak{H}^k$  *only* includes forms satisfying the appropriate boundary conditions. Ultimately, this stems from the domain of the operator  $\delta$  having vanishing boundary integrals, in order to fulfill the conditions of an adjoint. This space is special, because it conveys topological information (the content of the Hodge decomposition theorem and de Rham cohomology theory—see §1.8 and [109, 107, 58]); in this case, we must either consider forms whose traces vanish, or forms for which the traces of their *Hodge duals* vanish (see Figure 1.5). It is a form of Poincaré duality in which we can formulate two different complexes, which in the smooth theory correspond to the theory for differential forms, and the theory for forms with compact support.

However, there are other harmonic forms (just as in the theory for functions) with other boundary conditions. The harmonic spaces are still relevant, because we *solve* for such forms by, recall, extending the prescribed boundary forms using the surjectivity of the trace theorem (Theorem 1.3.2) and then solving the homogeneous problem with a nonzero source term (and of course, this is what we do numerically). As we have seen, for functions, the mixed weak form is the Neumann problem, and the harmonic forms gotten by projecting the source term corresponds to the Lagrange multiplier for the solution with the constraint of vanishing integral. The interpretation for forms of degree different from zero is similar, the condition being now that the integral wedged with the Hodge dual of the harmonic forms (in the harmonic space) is zero—the Lagrange multiplier no longer needs to be a constant function [5].

For compact manifolds without boundary, of course, boundary conditions no longer need to be specified, and so the harmonic space does in fact represent



**Figure 1.5:** A form and pseudoform in  $\mathbb{R}^2$  dual to each other, with the two kinds of boundary conditions in the annulus  $A = \{a < r < b\}$ . (1.5a):  $d\theta$ , a harmonic form whose *Hodge dual* has vanishing trace on  $\partial A$ . (“ $d\theta$ ” actually is a form determined by overlaps,  $\theta \in (-\pi, \pi)$  and  $\theta \in (0, 2\pi)$ .) This represents a local equipotential; its level sets are oriented in the direction of (local) increase of  $\theta$ . (1.5b):  $o(\theta, r) \frac{1}{r} dr$ , a harmonic pseudoform with vanishing trace on  $\partial A$ . This models the flow of a circulating fluid. (See [36] for the notation  $o(\theta, r)$ .) The direction of flow was found by pulling it back to  $\theta = \text{const}$ , *trans*-oriented by direction of increase. Also see Figure 1.7.



**Figure 1.6:** Example of harmonic form on closed manifold (here, a torus).

all possible harmonic forms defined on the whole space. This conveys topological information, and the harmonic forms are isomorphic to the de Rham cohomology.

**1.5.13 Essential vs. natural boundary conditions.** The ESSENTIAL boundary conditions, in this formulation, are those on  $\omega$  and  $d\omega$ , while NATURAL boundary conditions are those on  $\star\omega$  and  $\star d\omega$ . Natural boundary conditions are handled by additional boundary integrals, using the Generalized Stokes' Theorem, essentially, the failure of  $\delta$  to be an adjoint of  $d$ , which occurs because of boundary terms. In general, the vanishing of the natural boundary conditions does not need to be explicitly included, because Theorem 1.4.5 above ensures (via Stokes' Theorem) that the boundary integrals must vanish for any test form.

In this framework we can also explicitly include boundary conditions, namely, impose the conditions  $\text{Tr}\omega = 0$  and  $\text{Tr}d\omega = 0$  rather than their Hodge duals, so that the theory is all formulated in terms of the spaces  $\mathring{H}\Omega^k(U)$  (here, the  $k = 0$  case is the Dirichlet problem, while the  $k = n$  case is the Neumann problem, i.e., the two classical BVPS have *switched places* in the framework). As previously remarked, the domain of

the adjoint is then  $H^*\Omega^k(U)$ , namely, we've switched where the vanishing is supposed to occur. As remarked before, the harmonic forms are also different (See Figure 1.5). This leaves the question, of course, of which spaces to choose; this generally does not have an immediately apparent answer, but geometry (e.g., in the form of constitutive relations) can provide it in some cases. In some sense, as in the Poincaré duality theory, what goes for  $k$ -forms has a corresponding, isomorphic problem for  $(n - k)$ -forms. On compact, oriented manifolds, this is especially nice, because then the two theories are exactly the same. The question becomes one of whether the most suitable boundary conditions are tangential or normal (there are also parity considerations). For example, if we want to represent streamlines for flows, then  $(n - 1)$ -pseudoforms are the most natural way to formulate the problem. Choosing the right way to represent things, even if they have other equivalent formulations, is important for models and problems, because there are fewer steps in translation, and the most natural operations (such as choosing between curl and divergence) suggest themselves.

**1.5.14 Our convention.** In this work, when using this framework, we mostly concern ourselves with the natural boundary conditions unless the problem is really more naturally formulated the other way. For example, for concentration problems of any kind,  $n$ -pseudoforms become the most appropriate objects to use, because they naturally live on the full dimensional cells and require integration to describe quantity. It is only the relative unfamiliarity of pseudoforms that induces one to almost reflexively use proxies of some kind.

## 1.6 The Hilbert Space Setting for Elliptic Problems

As stated before, our chief goal in developing the theory of Sobolev spaces is to try to solve PDEs by taking advantage of the notion of completeness. Our usual

spaces of smooth functions are not complete, at least under the norms we would like them to be complete in, so we have to make use of more sophisticated spaces to get completeness. Of course, we can always complete by taking equivalence classes of Cauchy sequences in our desired norm, but it is very useful to know that there are alternate characterizations to these completed spaces, because this helps us clear the clutter when trying to derive properties. For example, we now know that the completion of the space of all smooth  $\mathcal{L}^p$  functions, with one  $\mathcal{L}^p$  derivative, is the set of all  $\mathcal{L}^p$  functions with one *weak* derivative in  $\mathcal{L}^p$ . The chief thing is that we see that we can still have a notion of differentiation on this complete space, whereas using equivalence classes of Cauchy sequences gives us no additional insight into the nature of these spaces.

### 1.6.1 Recasting in terms of Sobolev Spaces

Our main goal in this chapter is to develop the theory of weak solutions to PDEs in the Hilbert spaces  $W^{k,2}(U) = H^k(U)$ . As before, we have the inner product

$$\langle u, v \rangle_{H^k(U)} := \sum_{|\alpha| \leq k} \langle D^\alpha u, D^\alpha v \rangle_{\mathcal{L}^2(U)}.$$

which induces the  $W^{k,2}$ -norm as before (which we know is complete, so that  $H^k(U)$  is indeed a Hilbert space).

The notion of *weak solution* to a PDE is defined by very similar means as the notion of weak derivatives: via integration by parts. The notion turns out to be even weaker (at least *a priori*) than that of solving differential equations with weak derivatives (replacing all occurrences of classical derivatives with weak ones). We'll explain this more thoroughly with an example in a moment.

**1.6.1 Example** (Weak Formulation of Poisson's Equation). Our notion of solution will

be made so that a “weak solution” to a second-order PDE need only have *one weak* derivative, and not even two weak derivatives (which are in turn weaker than two classical derivatives) as the problem would initially suggest. We motivate things here using stronger hypotheses. First, suppose  $U$  has a smooth boundary and we are indeed working with  $C^2$  functions continuous up to the boundary. If  $u \in C^2(U) \cap C(\bar{U})$  solves

$$\begin{aligned} -\Delta u &= f \\ u|_{\partial U} &= g \end{aligned}$$

then we have, for any  $v \in C_c^\infty(U)$ , by integration of the equation against  $v$ :

$$\int_U (-\Delta u)v \, dx = \int_U f v \, dx.$$

However, recall Green’s First Identity: since  $\nabla \cdot ((\nabla u)v) = \nabla u \cdot \nabla v + (\Delta u)v$  by a vector calculus version of the product rule, we have that  $-(\Delta u)v = \nabla u \cdot \nabla v - \nabla \cdot ((\nabla u)v)$ . Therefore,

$$\begin{aligned} \int_U f v \, dx &= \int_U (-\Delta u)v \, dx = \int_U \nabla u \cdot \nabla v \, dx \\ &\quad - \int_U \nabla \cdot (v \nabla u) \, dx = \int_U \nabla u \cdot \nabla v \, dx - \int_{\partial U} v \frac{\partial u}{\partial n} \, ds. \end{aligned}$$

by the Divergence Theorem (which requires some smoothness on the boundary to apply). However, since  $v$  vanishes on  $\partial U$ , we have that the boundary integral vanishes, and so

$$\int_U \nabla u \cdot \nabla v \, dx = \int_U f v \, dx$$

for all  $v \in C_c^\infty(U)$ . Thus, we have established:

**1.6.2 Theorem.** *Let  $U$  be an open subset of  $\mathbb{R}^n$ , and suppose  $f \in C(U)$ ,  $u \in C^2(U) \cap C(\bar{U})$*

$C(\bar{U})$ , and  $g \in C(\partial U)$  solve the Dirichlet problem for Poisson's equation: We have

$$(1.6.1) \quad -\Delta u = f$$

$$(1.6.2) \quad u|_{\partial U} = g.$$

Then for every  $v \in C_c^\infty(U)$ , the following integral formula holds:

$$\int_U \nabla u \cdot \nabla v \, dx = \int f v \, dx.$$

In fact, in rewriting in terms of  $\mathcal{L}^2$ -inner products (and shedding the  $\nabla$ s), we have

$$\langle du, dv \rangle_{\mathcal{L}^2(U)} = \langle f, v \rangle_{\mathcal{L}^2(U)}.$$

Now, this is what motivates our *definition* of weak solution. The crucial point in this is that in the integral formulæ, only *one derivative* of  $u$  is used (as well as one derivative of  $v$ ). Suppose the boundary is smooth enough to enable notions of traces described in the previous sections (so that we can define what boundary values even are). This, of course, does not need  $C^\infty$ -smoothness. We now *define* a weak solution as follows:

**1.6.3 Definition.** Given  $f \in H^{-1}(U) = H^1(U)'$  and  $g \in H^{1/2}(\partial U)$ ,  $u \in H^1(U)$  is called a WEAK SOLUTION to (1.6.1) with BOUNDARY CONDITION (1.6.2) if for all  $v \in H_0^1(U)$ , we have

$$(1.6.3) \quad \langle du, dv \rangle_U = \langle f, v \rangle_U$$

and  $\text{Tr } u = g$  (this is equivalent to requiring an extension of  $g$  to a  $H^1$  function on all of  $U$  using the surjectivity of the trace, Theorem 1.3.2 and saying  $u - g \in H_0^1(U)$ ). Of



course, all the gradients (exterior derivatives) in the preceding should be weak (vectors of weak derivatives). It is common to abbreviate the LHS of the preceding as  $B(u, v)$  and the RHS as  $F(v)$ . Note that  $B$  can of course be generally defined as a bilinear form on  $H^1(U)$ , and  $F$  a linear functional. We will see this notation is useful in more general examples. To summarize, the WEAK FORMULATION of the problem is to find  $u \in H^1(U)$  to solve

$$(1.6.4) \quad B(u, v) = F(v)$$

for all  $v \in H_0^1(U)$ , and such that  $\text{Tr } u = g$ . Note, in general, if  $F$  is a bounded linear functional (which is the case here by the Cauchy-Schwarz inequality), and  $B$  is also bounded (i.e. there exists  $M$  such that  $B(\varphi, \psi) \leq M \|\varphi\| \|\psi\|$  for all  $\varphi$  and  $\psi$  in  $H_0^1$ ), then it suffices to just consider  $v \in C_c^\infty(U)$  instead (which is dense in  $H_0^1(U)$  by definition—continuous maps are determined completely by what they do on dense subsets).

**1.6.4 Example** (Weak formulation for differential forms). The power of this approach is that we can immediately generalize it to spaces of differential forms, because they are also Hilbert spaces. We have, for the non-mixed problem  $-\Delta\omega = \eta$  for  $\omega \in D(-\Delta)$  and  $\eta \in \mathcal{L}^2\Omega^k(M)$ ,

$$B(\omega, \eta) = \langle d\omega, d\eta \rangle + \langle \delta\omega, \delta\eta \rangle.$$

However, we can also apply this abstract theory to the mixed form (1.5.2) and its variations treated in the previous section, by defining

$$B(\sigma, \omega, p; \tau, v, q) = \langle \sigma, \tau \rangle - \langle \omega, d\tau \rangle - \langle d\sigma, v \rangle - \langle d\omega, dv \rangle - \langle p, v \rangle + \langle \omega, q \rangle.$$

for all  $(\tau, v, p) \in H\Omega^{k-1}(M) \times H\Omega^k(M) \times \mathfrak{H}^k$ . We shall see more of this in §1.8.

**1.6.5 Treatment of inhomogeneous boundary value problems.** The standard pro-

cedure [30, Remark at end of §6.1.2] for dealing with boundary values is to use the surjectivity of the trace operator (Theorem 1.3.2) to transform the problem into homogeneous one (one with boundary values zero), i.e., extending  $g$  to a function defined on all of  $H^1$  and then considering the problem

$$-\Delta w = f + \Delta g$$

$$w|_{\partial U} = 0$$

and recovering the original equation as  $u = w + g$  (note we are using, again,  $\Delta$  as an operator into  $H^{-1}$ ). This means that we can instead solve for  $w$  in  $H_0^1(U)$ , so that we are seeking  $w \in H_0^1(U)$  such that

$$B(w, v) = F(v)$$

for all  $v \in H_0^1(U)$ . This is motivated of course by the classical problem; we should verify it works in the weak case: say  $w, u, f$ , and  $g$  are as above and  $u = w + g$ . Then:

$$\begin{aligned} \int_U \nabla u \cdot \nabla v \, dx &= \int_U \nabla(w + g) \cdot \nabla v \, dx = \int_U \nabla w \cdot \nabla v \, dx + \int_U \nabla g \cdot \nabla v \, dx \\ &= \int_U (f + \Delta g) v \, dx - \int_U (\Delta g) v \, dx = \int_U f v \, dx \end{aligned}$$

as desired.

Solutions as in the example directly above are, as mentioned, called WEAK SOLUTIONS. Solutions involving two weak derivatives in the example preceding the above are sometimes confusingly called STRONG SOLUTIONS. Solutions using classical derivatives are called CLASSICAL SOLUTIONS. So a classical solution is the strongest kind of solution we can demand. Regularity theory says that for  $f$  in a better function

space such as  $\mathcal{L}^2$ , or some Hölder space, the solution is also that smooth (we'll give a basic overview of this later).

**1.6.6 Example** (Sturm-Liouville Problem). Let  $p$  and  $q$  be smooth functions. Consider the problem

$$\begin{aligned} -\nabla \cdot (p \nabla u) + qu &= f \\ u|_{\partial U} &= g \end{aligned}$$

with, as the usual motivation,  $g \in C(\partial U)$  and  $f \in C(\bar{U})$ . This is usually called the STURM-LIOUVILLE PROBLEM, although that also often refers to the corresponding eigenvalue problem (which we'll give as another example). It reduces to Poisson's Equation when  $p \equiv 1$  and  $q \equiv 0$ . To rewrite this in its weak formulation, we once again appeal to Green's First Identity :

$$\nabla \cdot ((p \nabla u) v) = \nabla \cdot (p \nabla u) v + p \nabla u \cdot \nabla v.$$

So therefore, for  $v \in C_c^\infty(U)$ ,

$$\begin{aligned} \int_U -\nabla \cdot (p \nabla u) v \, dx &= \int_U p \nabla u \cdot \nabla v \, dx \\ - \int_U \nabla \cdot ((p \nabla u) v) \, dx &= \int_U p \nabla u \cdot \nabla v \, dx - \int_{\partial U} p \frac{\partial u}{\partial n} v \, ds. \end{aligned}$$

Because  $v \in C_c^\infty(U)$ , it vanishes at the boundary, so the second integral drops out. Therefore we have the following weak formulation: to find  $u \in H^1(U)$  such that for all  $v \in H_0^1(U)$ ,

$$\int_U p \nabla u \cdot \nabla v \, dx + \int_U quv \, dx = \int_U f v \, dx.$$

and such that  $u - g \in H_0^1(U)$ .

## 1.6.2 The General Elliptic Problem

The preceding examples were all special cases of very general elliptic partial differential equations. Here, we define them and give their weak formulation, and explore how all our usual examples are derived from this.

**1.6.7 Definition.** Let  $a^{ij}$ ,  $b^j$  and  $c$  be functions on  $U$ . In general the  $a^{ij}$  denote components of a symmetric contravariant 2-tensor  $A$ —often metric coefficients in Riemannian Geometry, and  $b^j$  are components of a vector field  $b$ . We require  $a^{ij}$  to be ELLIPTIC or COERCIVE, that is,

$$a^{ij}\xi_i\xi_j = A(\xi, \xi) > 0$$

at every point (again, using the Einstein Summation Convention, Remark 1.2.2 above), that is, the quadratic form  $A$  is positive-definite. We actually often require that  $A$  be UNIFORMLY ELLIPTIC: There exists a constant  $\theta > 0$  such that

$$a^{ij}(x)\xi_i\xi_j = A(\xi, \xi) \geq \theta|\xi|^2$$

at every point  $x$  and for all  $\xi \in \mathbb{R}^n$ . This says that not only the quadratic form  $A$  is positive definite at all points, but also that its smallest eigenvalue is always bounded below by the positive constant  $\theta$ . Plain ellipticity only requires that the smallest eigenvalue be positive at all points, which allows it to be arbitrarily close to 0, whereas uniform ellipticity forces it to be outside a whole fixed neighborhood of 0 (this condition is called BOUNDED AWAY FROM ZERO in most geometry and PDE literature).

Let us first write out the coordinate formulation in the DIVERGENCE FORM

which we shall see makes the weak formulation easier to write:

$$Lu := -D_i(a^{ij}D_j u) + b^j D_j u + cu = f.$$

$L$  is called an ELLIPTIC (DIFFERENTIAL) OPERATOR. There is a NONDIVERGENCE FORM which looks like, for functions  $\alpha^{ij}$ ,  $\beta^j$ , and  $\gamma$ :

$$-\alpha^{ij}D_i D_j u + \beta^j D_j u + \gamma u = f,$$

and provided that the coefficients are sufficiently smooth, the two formulations are equivalent; by expanding the divergence form using the product rule:

$$-a^{ij}D_i D_j u + (b^j + D_i a^{ij})D_j u + cu$$

so that  $\alpha^{ij} = a^{ij}$ ,  $\beta^j = b^j + D_i a^{ij}$ , and  $\gamma = c$ . Now the reason why we say that smoothness matters is that in the weak formulation, we can loosen the regularity assumptions on  $a^{ij}$ , because it will appear outside any derivative operator. The nondivergence form is useful for working with maximum principles [30, §6.4].

**1.6.8 Rewriting things more invariantly.** To rewrite the divergence form operator in a more invariant fashion, we define, for a (co)vector  $\xi$ ,  $A^\sharp(\xi) := a^{ij}\xi_j e_i$ , where  $e_i$  are the standard basis vectors—it is the vector whose  $i$ th component is  $a^{ij}\xi_j$ . The physical interpretation of  $-A^\sharp \xi$  is that it gives the direction of flow. For example, if it is just the Riemannian metric,  $-A^\sharp du$  points oppositely to  $\xi$ , saying that flow is from areas of higher concentration to lower concentration.

Also, if  $b$  is a vector field,

$$b^j D_j u = b \cdot \nabla u = b \lrcorner du = du(b) = \mathcal{L}_b u$$

the Lie (directional) derivative of  $u$  in the direction  $b$ . Thus we may rewrite the operator  $L$  as follows:

$$Lu = -\nabla \cdot (A^\sharp(\nabla u)) + \mathcal{L}_b u + cu = \delta(A^\sharp du) + \mathcal{L}_b u + cu.$$

This assists immensely in writing these equations on manifolds, which do not necessarily admit global coordinate charts, and also explains the name “divergence form” (the presence of the operator  $\nabla \cdot$ ). Note also that the geometric condition  $-A^\sharp(du) \cdot du \leq 0$  says that the flux from diffusion always travels opposite the gradient  $du$ , consistent with the usual constitutive laws of diffusive flux, and therefore, even when  $A$  is not given as a separate, prescribed Riemannian metric, and thus is anisotropic, it still flows from regions of higher concentration to lower concentration.

**1.6.9 The weak formulation.** We finally are ready to state the weak formulation. With the  $D_i$  conveniently placed outside everything, we can make the divergence theorem work for us, namely,  $D_i(a^{ij} D_j u)v = (D_i(a^{ij} D_j u))v + a^{ij} D_j u D_i v$ . So,

$$\int_U -D_i(a^{ij} D_j u) dx = \int_U a^{ij} D_i u D_j v dx - \int_U D_i(v a^{ij} D_j u) dx.$$

Invariantly, it is more transparent, and in fact nearly identical to the Sturm-Liouville situation (noting that  $A^\sharp(\xi) \cdot \eta = A(\xi, \eta)$ ):

$$\nabla \cdot (A^\sharp(\nabla u)v) = \nabla \cdot (A^\sharp(\nabla u))v + A^\sharp(\nabla u) \cdot \nabla v = \nabla \cdot (A^\sharp(\nabla u))v + A(\nabla u, \nabla v),$$

so that

$$\begin{aligned} & \int_U -\nabla \cdot (A^\sharp(\nabla u)) v \, dx \\ &= \int_U A(\nabla u, \nabla v) \, dx - \int_U \nabla \cdot (A^\sharp(\nabla u) v) \, dx = \int_U A(\nabla u, \nabla v) \, dx - \int_{\partial U} A(\nabla u, n) v \, ds. \end{aligned}$$

By the usual boundary conditions, the last boundary integral vanishes, so we have the full weak formulation of the problem  $Lu = f$ : To seek  $u \in H_0^1(U)$  such that for all  $v \in H_0^1(U)$

$$(1.6.5) \quad \int_U A(\nabla u, \nabla v) \, dx + \int_U (\mathcal{L}_b u) v \, dx + \int_U c u v \, dx = \int_U f v \, dx,$$

and finally,

$$\int A(du, dv) \, dx + \langle \mathcal{L}_b u, v \rangle + \langle cu, v \rangle = \langle f, v \rangle.$$

In coordinates,

$$\int_U a^{ij} D_i u D_j v \, dx + \int_U b^j D_j u v \, dx + \int_U c u v \, dx = \int_U f v \, dx.$$

Because no derivatives are involved on the coefficients  $a^{ij}$ ,  $b^j$  and  $c$ , we need only assume they are regular enough for the integrals to exist, which generally means they are in  $\mathcal{L}^2$  or  $\mathcal{L}^\infty$  or something of the sort. The use of the divergence theorem eliminates the minus sign.

As usual, we often abbreviate the LHS as  $B(u, v)$  and the RHS by  $F(v)$  and note that  $B$  is bilinear and  $F$  is a linear functional.

**1.6.10 Example** (All the preceding are special cases of the Elliptic Problem). If  $a^{ij} = \delta^{ij}$ ,  $b^j = 0$  and  $c = 0$ , then  $L$  is the Laplacian  $\Delta$ . If  $a^{ij} = p\delta^{ij}$  (a diagonal matrix with the scalar function  $p$  in its 3 entries),  $b^j = 0$ , and  $c = q$ , another function, then  $L$  is the

Sturm-Liouville operator.

**1.6.11 Example** (The Laplacian in Differential Geometry, [58, 19, 26]). Let  $(M, g)$  be a Riemannian manifold with boundary. Recall that on a Riemannian manifold, the Laplacian is defined in coordinates by

$$\Delta u := -\frac{1}{\sqrt{g}} \frac{\partial}{\partial x^j} \left( \sqrt{g} g^{ij} \frac{\partial u}{\partial x^i} \right)$$

where  $\sqrt{g}$  is the square root of the determinant  $\det(g_{ij})$ , and  $g^{ij}$  are the coefficients of the metric on the cotangent space (inverse metric). This is often called the LAPLACE-BELTRAMI OPERATOR. So, in coordinates (which is ultimately how we must compute), given  $f \in \mathcal{L}^2(M)$ , to solve  $\Delta u = f$ , in each coordinate chart, we must solve

$$-D_i(\sqrt{g} g^{ij} D_j u) = \sqrt{g} f$$

which says, in terms of our general elliptic problem, that  $a^{ij} = \sqrt{g} g^{ij}$  the “densitized metric,” and  $b^j = 0$ . If the patch we choose is precompact (has compact closure), then by the smoothness of the metric, it is uniformly elliptic (choose a constant coordinate vector field, say  $\frac{\partial}{\partial x^1}$ ; then  $g\left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^1}\right)$  has a positive minimum over the patch, that furnishes the positive lower bound required for uniform ellipticity. Thus we see that Laplacians on Riemannian manifolds become general elliptic problems in coordinates. This fact alone justifies study of general elliptic operators. In weak formulation, it looks like: Find all  $u \in H_0^1(U)$  such that for all  $v \in H_0^1(U)$ ,

$$\int_U \sqrt{g} g^{ij} D_i u D_j v \, dx = \int_U \sqrt{g} f v \, dx.$$

Finally, reinterpreting things back in coordinate-free terms, in geometry,  $\sqrt{g} dx = d\mu$



is, recall, the *Riemannian volume form* induced by the metric, and the integral is

$$\int_U g(\nabla u, \nabla v) d\mu = \int f v d\mu.$$

which formally looks exactly the same as it does in Euclidean space (after noting the usual Euclidean metric is just  $g_{ij} = \delta_{ij}$  and  $d\mu = dx$ ).

The physical interpretation of general elliptic operators is that the  $a^{ij}$  represent *diffusion* phenomena, which take into account (linear) anisotropic properties of the material (diffusion occurring more easily in some directions than others),  $b^j$  represent *convection* phenomena (say a fluid already flowing on the manifold), and  $c$  represents *source* phenomena (material being created or destroyed, e.g. through chemical reactions). The geometry of a manifold, of course, will alter the way diffusion operates, by its curved nature, which is why it is reasonable that metric coefficients can serve as the  $a^{ij}$ .

Actually, those pesky factors of  $\sqrt{g}$  tell us that, at least for concentration problems, we still have not gotten to the geometrically correct representation of the quantities at hand, as hinted in §1.5 where for such problems, the most appropriate thing is to consider  $u$  as an  $n$ -pseudoform. The  $a^{ij}$  similarly should modify  $\delta u$  appropriately, or simply just *become* the codifferential of  $u$  (see next remark), relative to a different metric, giving rise to a flux, an  $(n - 1)$ -pseudoform, which is most appropriately integrated over transversely oriented hypersurfaces. Then  $\langle -\delta u, \delta u \rangle \leq 0$  states that the flux takes material from areas of higher concentration to lower concentration.

**1.6.12 Redefining codifferentials and Hodge theory for coefficients.** The point of the preceding remarks about the metric in Riemannian manifolds is that we now can take the coefficients  $a^{ij}$  as a symmetric 2-tensor (ellipticity makes it positive-definite) and declare it a *new* inner product (as a sufficiently smooth, symmetric, positive-

definite 2-tensor), thus showing that the diffusion terms in a general elliptic problem always corresponds to *some* Hodge Laplacian problem (in fact, this fact is crucial for establishing some versions of the Sobolev embedding theorems for manifolds with Lipschitz boundary [84]). Uniform ellipticity shows that the inner product defined by the coefficients is equivalent to the  $\mathcal{L}^2$  inner product. This is also a way of formulating the Hodge operator as a kind of constitutive equation, and is also useful for formulating Maxwell's equations in terms of spacetime Hodge operators [36, §3.5 and Ch. 14]. To show that this works, we only need to demonstrate that the codifferential, hence the Hodge Laplacian,  $\delta$  is what we claim it to be (and then all the results of Hilbert complexes apply).

Provided, of course, that the coefficients are sufficiently smooth, we simply take  $g = a^{ij}$ ; structures such as  $\sqrt{g}$  apply automatically with the above. This is an example of how a problem *defines* a new geometry, perhaps because of some local structure inside the material, which is “anisotropic” only when seen from an ordinary Euclidean geometric point of view. We also must be careful, however, not to conflate it with other metrics, should they be given. In particular, we have to take care to note where and when we use such operators and other tools used in the existence theory, such as orthogonal projections and boundary conditions (both Dirichlet and Neumann, for the general case, with both  $\text{Tr} \star u$  and  $\text{Tr} \star du$  vanishing—see Example 1.5.2 above).

## 1.7 The Theory of Weak Solutions

As noted several times in the above, we rewrote all our example differential equations into the form  $B(u, v) = F(v)$ , where  $B : H \times H \rightarrow \mathbb{R}$  is some bilinear form defined on some Hilbert space of functions  $H$ , and  $F \in H'$  is some linear functional on

$H$ . The reason for this is that it expresses existence and uniqueness in terms of a very simple principle in the theory of Hilbert spaces: the Riesz representation theorem [34, §5.5]:

**1.7.1 Theorem.** *Let  $H$  be a (complex) Hilbert space. Then given any bounded linear functional  $F \in H'$ , there exists a unique  $u \in H$  such that*

$$F(v) = \langle u, v \rangle_H.$$

Moreover,  $\|u\|_H = \|F\|_{H'}$ .

Note that the appearance of  $u$  on the left factor of that inner product is actually why we prefer the conjugate on that factor when using complex Hilbert spaces; for the other convention, we have that  $F(v) = \langle v, u \rangle$ , that is, the  $u$  acts from the right (in [34], the theorem is stated and proved for this case). If  $B$  is a *symmetric* bilinear form (which will be the case, for example, if it arises from a general elliptic operator with *no* convection terms), it defines an inner product on  $H$  (called the ENERGY INNER PRODUCT), the Riesz representation theorem applies, and so given any  $F \in H'$ , there exists a unique  $B(u, v) = F(v)$  with  $\|u\|_B := B(u, u)^{1/2} = \|F\|_{H'}$ . If  $B$  is coercive, i.e. there exists  $\gamma > 0$  (the COERCIVITY CONSTANT) such that  $B(u, u) \geq \gamma \|u\|_H^2$ , then we have

$$\|u\|_H \leq \gamma^{-1/2} \|u\|_B \leq \gamma^{-1/2} \|F\|_{H'}.$$

so that the  $B$ -norm (the ENERGY NORM) is equivalent to the given Hilbert space norm. This constant  $\gamma$  is often referred as to the Poincaré constant, although we use it for a closely related quantity in the theory of Hilbert complexes below.

### 1.7.1 The Lax-Milgram Theorem

For our general elliptic problem, which includes convection terms (thus leading to a non-symmetric bilinear form  $B$ ), we need a theorem of greater generality.

**1.7.2 Theorem** (The Lax-Milgram Theorem, [30], §6.2.1, Theorem 1). *Let  $B : H \times H \rightarrow \mathbb{R}$  be a bounded, real, coercive bilinear form, and  $F \in H'$  be a linear functional. Then there exists a unique  $u \in H$  such that*

$$B(u, v) = F(v)$$

*for all  $v \in H$ , and moreover,  $\|u\|_H \leq \gamma^{-1} \|F\|_{H^{-1}}$  (the a priori estimate), where  $\gamma$  is the coercivity constant of  $B$ .*

Note that the constant here is  $\gamma^{-1}$  rather than the sharper  $\gamma^{-1/2}$  for symmetric coercive bilinear forms in the above. The proof, e.g., in Evans [30], is very illustrative of the important ideas and concepts that get built upon in the theory of Hilbert complexes. We use some of these ideas for dealing with some noncoercive bilinear forms, in §1.8. The key step is showing that the action of the bilinear form  $B$  is equivalent to a bounded linear operator acting on  $H$  in the first factor of the given inner product (the infinite-dimensional version of “index raising” for tensors). This allows us to reduce the question of existence to the Riesz representation theorem as before.

Actually, our most important operator  $-\Delta$  is *not* coercive; it does not become so until we deal with the fact it has a kernel  $\mathfrak{H}$ . We can use Fredholm theory to show that  $\mathfrak{H}$  is in fact finite-dimensional, due to certain compactness results. Thus, if we pose the problem on  $\mathfrak{H}^\perp$ , the bilinear form  $(u, v) \mapsto \langle du, dv \rangle$  on  $\mathfrak{H}^\perp$  is coercive. The constant here is given in the abstract Poincaré inequality in §1.8 below. On the other hand, we can also instead consider restricting  $-\Delta$  to  $H_0^1$ , where it is indeed coercive by

the Poincaré inequality.

## 1.7.2 Basic Existence Theorems

We catalogue some of the basic existence results, seeing what happens when we try to verify the Lax-Milgram theorem. We follow [30, §6.2] and also give a brief note on the existence of eigenfunctions (which are essential, of course, for Fourier series, and the basis of a technique for establishing the well-posedness of *parabolic* problems). This involves deriving estimates on the bilinear form, called ENERGY ESTIMATES, because the bilinear form usually corresponds to that concept for elastic energy (it is also why the corresponding norms, as remarked before, are called ENERGY NORMS).

**1.7.3 Theorem.** *Let  $B$  be the bilinear form corresponding to the general, uniformly elliptic operator on a domain  $U \subseteq \mathbb{R}^n$ . Then there exist  $\alpha, \beta > 0$  and  $\gamma \geq 0$  such that*

$$|B(u, v)| \leq \alpha \|u\|_{H^1(U)} \|v\|_{H^1(U)}$$

and GÅRDING'S INEQUALITY [106, Ch. 4]

$$B(u, u) \geq \beta \|u\|_{H^1(U)}^2 - \gamma \|u\|_{\mathcal{L}^2}^2$$

holds.

Note that if  $\gamma > 0$ , then the bilinear form actually does *not* satisfy the hypotheses of the Lax-Milgram theorem.

*Proof.* Bounding above is clear, from taking the  $L^\infty$ -norms (essential suprema) of the coefficients and using the Cauchy-Schwarz inequality. Bounding below (coercivity) is trickier: first we use Cauchy-Schwarz to bound below by the greatest negative

norm, and then use Cauchy's arithmetic-geometric mean inequality with  $\varepsilon$ , namely  $\alpha\beta = (\varepsilon\alpha)(\varepsilon^{-1}\beta) \leq \frac{1}{2}(\varepsilon^2\alpha^2 + \varepsilon^{-2}\beta^2)$  for any  $\varepsilon > 0$ , to split the term with  $\|u\|\|du\|$  into squares:

$$\begin{aligned}
B(u, u) &= \int_U (A(du, du) + (b \lrcorner du)u + cu^2) dx \\
&\geq \theta \|du\|^2 - \sum_i \|b^i\|_{\mathcal{L}^\infty} \|u\| \|du\| - \|c\|_{\mathcal{L}^\infty} \|u\|^2 \\
&\geq \theta \|du\|^2 - \frac{1}{2}\varepsilon^2 \|b\|_{\ell^1(\mathcal{L}^\infty)} \|du\|^2 - \frac{1}{2}\varepsilon^{-2} \|b\|_{\ell^1(\mathcal{L}^\infty)} \|u\|^2 - \|c\|_{\mathcal{L}^\infty} \|u\|^2 \\
&\geq \frac{1}{2}\theta \|du\|^2 - \left(\frac{1}{2}\|b\|_{\ell^1(\mathcal{L}^\infty)}\varepsilon^{-2} + \|c\|_{\mathcal{L}^\infty}\right) \|u\|^2,
\end{aligned}$$

where  $\varepsilon > 0$  has been chosen such that  $\|b\|_{\mathcal{L}^\infty}\varepsilon^2 \leq \theta$ . From this point, we can either simply add an additional  $\theta$  in the factor multiplying the second term (i.e., we take  $\gamma = \frac{1}{2}(\|b\|_{\ell^1(\mathcal{L}^\infty)}\varepsilon^{-2} + \theta) + \|c\|_{\mathcal{L}^\infty}$ ), or we consider subspaces on which some form of Poincaré inequality holds (which leads to better constants; we see this in a moment with some special cases). If on some  $V \subseteq H^1(U)$ , we have

$$\|u\|_{H^1} \leq c_P \|du\|, \quad \forall u \in V,$$

then

$$B(u, u) \geq \frac{1}{2}\theta c_P^{-2} \|u\|_{H^1}^2 - \left(\frac{1}{2}\|b\|_{\ell^1(\mathcal{L}^\infty)}\varepsilon^{-2} + \|c\|_{\mathcal{L}^\infty}\right) \|u\|^2.$$

□

The usual choices are either  $V = H_0^1(U)$  in which the Poincaré inequality follows from elliptic theory [30, Ch 5], or  $V = \mathfrak{H}^{0\perp} = \mathfrak{Z}^{0\perp}$ , the orthogonal complement of the constant functions (kernel of  $d$ ), for which the inequality holds by the theory of Hilbert complexes in the next section. This means, in general, that we need to add an extra term for the existence and uniqueness of weak solutions:

**1.7.4 Corollary.** *Suppose  $B$ , the bilinear form corresponding to some elliptic operator  $L$ , satisfies Gårding's inequality as above, and that a Poincaré inequality holds on a subspace  $V \subseteq H$ . Then for all  $\mu \geq \gamma$ , the equation  $Lu + \mu u = f$  has a unique weak solution  $u \in V$  for every  $f \in V'$ .*

**1.7.5 The advantage of a Poincaré inequality.** We look at a couple of special cases, which also illustrates the advantage of having a Poincaré inequality. First, suppose  $b = 0$  and  $c = 0$ . If we do not insist on a Poincaré inequality,  $\gamma$  in the above ends up being  $\frac{1}{2}\theta$ , which means we only have existence and uniqueness of weak solutions for the operator  $L + \mu I$  with  $\mu \geq \frac{1}{2}\theta$  (the advantage is that there is no constraint on the solution other than being in  $H^1(U)$ ). But if we restrict to  $V$  with a Poincaré inequality, there the  $\gamma$  is solely defined in terms of  $b$  and  $c$ , which are zero, so  $\gamma$  vanishes. Thus we can take  $\mu = 0$ , and we have existence and uniqueness of weak solutions for the operator  $L$  itself.

Another special case (considered in [11, §II.2]) is when  $b = 0$  and the function  $c$  is *bounded away from zero*, namely,  $c(x) \geq c' > 0$ . Then instead of bounding below in the inequality above with  $-\|c\|_{\mathcal{L}^\infty}\|u\|^2$ , we can instead bound below with  $c'\|u\|^2$ , that is,

$$B(u, u) \geq \theta \|du\|^2 + c' \|u\|^2.$$

Since there are no negative quantities, we have no need to finesse with Cauchy AM-GM inequality; instead, taking  $\theta' = \min\{\theta, c'\}$ , we directly have

$$B(u, u) \geq \theta' (\|du\|^2 + \|u\|^2) = \theta' \|u\|_{H^1}^2,$$

so it is coercive on all of  $H^1(U)$ , thus we have existence and uniqueness of weak solutions for  $L$  on all of  $H^1(U)$  and  $f \in H^1(U)'$ .

For the cases when the operator is not invertible (not restricting ourselves to

some  $V$ ), Fredholm theory allows us to deduce that the kernel of the operator  $L$  is finite-dimensional (really, it is from the compactness of the solution operators in the right norms). Of course, in that situation, some other criteria must be used to single out a solution; the problems that call for some function in the kernel for  $L$  are of a different nature; for example, harmonic functions often represent the average value of the solution to another problem, an often useful piece of data (and can participate in constraints).

**1.7.6 The solution operator, its compactness, and eigenfunctions.** Suppose, now that the Poincaré inequality holds on  $V \subseteq H^1(U)$  and  $B_\mu(u, v) = B(u, v) + \mu\langle u, v \rangle$  for  $\mu \geq \gamma$  in the above. Then the unique weak solution  $u$  such that  $B_\mu(u, v) = \langle f, v \rangle$  is a linear operator  $S$  on  $f$  mapping  $V'$  into  $V$ ; the *a priori* estimate of the Lax-Milgram theorem guarantees that  $S$  is bounded operator, and then further taking  $V \hookrightarrow H^1(U) \hookrightarrow \mathcal{L}^2(U)$ , we have by compactness of  $H^1(U) \hookrightarrow \mathcal{L}^2(U)$  that  $S: V' \rightarrow \mathcal{L}^2(U)$  is compact [30, §5.7]. Finally, restricting  $S$  to  $\mathcal{L}^2(U)$  gives  $S$  as a compact operator on  $\mathcal{L}^2(U)$ . Thus Fredholm theory applies to the operator  $S$ . Specifically,  $Lu = f$  is a weak solution if and only if  $Lu + \mu u = f + \mu u$ , if and only if  $u = S(f + \mu u)$ , and finally if and only if  $u - \mu Su = Sf$ ,  $\mu S$  is a compact operator, so the Fredholm alternative applies, and the existence of solutions to  $Lu = f$  is changed into questions about the existence of solutions for the operator  $I - \mu S$  [30, §6.2]. This gives the finite-dimensional kernel. If  $\mu = 0$  works, in particular,  $L$  itself is invertible with compact inverse  $S$ .

If  $B$  is symmetric, then we have spectral theory, because then,  $S$  is symmetric: For  $f, g \in \mathcal{L}^2$ ,

$$\langle Sf, g \rangle_{\mathcal{L}^2} = \langle g, Sf \rangle_{\mathcal{L}^2} = B(Sg, Sf) = B(Sf, Sg) = \langle f, Sg \rangle_{\mathcal{L}^2}.$$

By spectral theory,  $S$  has a complete, orthonormal set of eigenfunctions  $\{\phi_k\}$  with



corresponding positive real eigenvalues  $\mu_k$  with  $\mu_k \rightarrow 0$  as  $k \rightarrow \infty$ . Defining  $\lambda_k = \mu_k^{-1}$ , we have  $\lambda_k \rightarrow \infty$ .

## 1.8 Hilbert Complexes

Much of the theory of boundary value problems for differential forms can be very elegantly cast into the framework of HILBERT COMPLEXES, introduced by Brüning and Lesch [14]. This framework abstracts the key properties of differential forms that make them amenable to posing elliptic differential equations, and also very importantly for us, their approximation. It is useful, for example, to see exactly where concepts like the Poincaré inequality come from. Also, the framework unifies various disparate problems, explaining types of boundary conditions, realizing elliptic equations with general coefficients all as one kind of equation (but with different inner product), gives a very clear proof of the Hodge decomposition theorem, and sets up a framework for approximation. Questions of existence, uniqueness, and well-posedness are treated very cleanly here in general. Regularity theory remains separate, however (so, in particular, strong results on the Hodge decomposition theorem still need the standard regularity theory of general elliptic operators [109, Chapter IV]). Most of what we review here is as done by Arnold, Falk, and Winther [6], who also apply this theory to formulate stable numerical methods; indeed, it is our eventual goal in this work to explore those methods and build on them.

**1.8.1 Definition** (Hilbert complexes). We define a HILBERT COMPLEX  $(W, d)$  to be a sequence of Hilbert spaces  $W^k$  with possibly unbounded linear maps  $d^k : V^k \subseteq W^k \rightarrow V^{k+1} \subseteq W^{k+1}$ , such that each  $d^k$  has closed graph, densely defined, and satisfies the COCHAIN PROPERTY  $d^k \circ d^{k-1} = 0$  (this is often abbreviated  $d^2 = 0$ ; we often omit the superscripts when the context is clear). We call each  $V^k$  the DOMAIN of  $d^k$ . We will

often refer to elements of such Hilbert spaces as “forms,” being motivated by the canonical example of the de Rham complex. The Hilbert complex is called a **CLOSED COMPLEX** if each image space  $\mathfrak{B}^k = d^{k-1}V^{k-1}$  (called the  $k$ -**COBOUNDARIES**) is closed in  $W^k$ , and a **BOUNDED COMPLEX** if each  $d^k$  is in fact a bounded linear map. The most common arrangement in which one finds a bounded complex is by taking the sequence of domains  $V^k$ , the same maps  $d^k$ , but now with the **GRAPH INNER PRODUCT**

$$\langle v, w \rangle_V = \langle v, w \rangle + \langle d^k v, d^k w \rangle.$$

for all  $v, w \in V^k$ . This new complex is called the **DOMAIN COMPLEX**. Unsubscripted inner products and norms will always be assumed to be the ones associated to  $W^k$ . We will also omit superscripts on the  $d$  for clarity of notation when it is clear from the context.

**1.8.2 Example** (The de Rham Complex). Of course, this is motivated by the case of Sobolev spaces of differential forms,  $W^k = \mathcal{L}^2\Omega^k(M)$  and  $V^k = H\Omega^k(M)$  for a manifold-with-boundary  $M$ , with  $d$  the exterior derivative. By approximation with smooth forms, we see immediately  $H\Omega^k(M)$  is dense in  $\mathcal{L}^2\Omega^k(M)$ . To show that  $d$  has closed graph, we consider the sequence  $(\omega_m, d\omega_m)$  in the graph of  $d$  converging in the product norm  $\mathcal{L}^2$  to  $(\omega, \eta)$ . Then clearly  $\omega_m \rightarrow \omega$  and

$$\langle d\omega_m, \varphi \rangle = \langle \omega_m, \delta\varphi \rangle \rightarrow \langle \omega, \delta\varphi \rangle.$$

for any test form  $\varphi$  whose boundary trace vanishes. But  $\langle d\omega_n, \varphi \rangle \rightarrow \langle \eta, \varphi \rangle$  as well, so  $\langle \eta, \varphi \rangle = \langle \omega, \delta\varphi \rangle$  for all test forms  $\varphi$ . This shows that  $\eta = d\omega$ , by definition of distributional exterior derivative. It is a closed complex, although we show this later (it satisfies a compactness property [6, 84]).

**1.8.3 Example** (The de Rham Complex with Essential Boundary Conditions [6], §6.2). If  $M$  is a manifold with boundary, we can consider the complex with  $W^k = \mathcal{L}^2\Omega^k(M)$  as before, but now with domains  $V^k = \mathring{H}\Omega^k(M)$  and the exterior differential as before. Since  $d$  commutes with pullbacks, in particular,  $d$  commutes with the trace operator, so that  $d$  actually maps  $\mathring{H}\Omega^k(M)$  to  $\mathring{H}\Omega^{k+1}(M)$ . This actually shows this complex is a *subcomplex* of the de Rham complex above.

**1.8.4 Definition** (Cocycles, Coboundaries, and Cohomology). We have similar generalizations of differential form complexes for abstract Hilbert complexes. The kernel of the map  $d^k$  in  $V^k$  will be called  $\mathfrak{Z}^k$ , the  $k$ -COCYCLES and, as before, we have  $\mathfrak{B}^k = d^{k-1}V^{k-1}$ . Since  $d^k \circ d^{k-1} = 0$ , we have  $\mathfrak{B}^k \subseteq \mathfrak{Z}^k$ , so we have the  $k$ -COHOMOLOGY  $\mathfrak{Z}^k/\mathfrak{B}^k$ . The HARMONIC SPACE  $\mathfrak{H}^k$  is the orthogonal complement of  $\mathfrak{B}^k$  in  $\mathfrak{Z}^k$ . This means, in general, we have an orthogonal decomposition  $\mathfrak{Z}^k = \overline{\mathfrak{B}^k} \oplus \mathfrak{H}^k$ , and we have that  $\mathfrak{H}^k$  is isomorphic to  $\mathfrak{Z}^k/\overline{\mathfrak{B}^k}$ , the REDUCED COHOMOLOGY, which of course corresponds to the usual cohomology for closed complexes.

**1.8.5 Definition** (Dual complexes and adjoints). For a Hilbert complex  $(W, d)$ , we can form the DUAL COMPLEX  $(W^*, d^*)$  which consists of spaces  $W_k^* = W^k$ , maps  $d_k^* : V_k^* \subseteq W_k^* \rightarrow V_{k-1}^* \subseteq W_{k-1}^*$  such that  $d_{k+1}^* = (d^k)^*$ , the adjoint operator, that is:

$$\langle d_{k+1}^* v, w \rangle = \langle v, d^k w \rangle.$$

The operators  $d^*$  decrease degree, so this is a chain complex, rather than a cochain complex; the analogous concepts to cocycles and coboundaries extend to this case and we write  $\mathfrak{Z}_k^*$  and  $\mathfrak{B}_k^*$  for them.

**1.8.6 Example** (The de Rham complex). As noted before, the adjoint  $d^*$  of the operator  $d$  in the de Rham complex on a manifold-with-boundary is just the codifferential, but it must be noted that their *domains* are *not* all of  $H^*\Omega^k(M)$ , but rather the complex

$\mathring{H}^* \Omega^k(M)$ , forms whose Hodge duals have vanishing trace (Theorem 1.4.5 above; see also Figure 1.4), because we need the boundary terms to vanish in the integration by parts for the relevant operators to actually be adjoints. Of course, if  $M$  is a compact manifold without boundary, there is no boundary and it is indeed the whole space  $H^* \Omega^k(M)$ .

But, dually, the de Rham complex with boundary conditions has a dual complex *without* boundary conditions, showing that the vanishing at the boundary is something that gets carried along with information about duals (as well as their parity and degree). In short,

$$(H\Omega(M), d) \text{ has the dual complex } (\mathring{H}^* \Omega(M), \delta),$$

but

$$(\mathring{H}\Omega(M), d) \text{ has the dual complex } (H^* \Omega(M), \delta).$$

**1.8.7 Example** (de Rham Complex with Coefficients). If  $a^{ij}$  are smooth coefficients, or at least smooth enough to preserve the spaces  $H\Omega(M)$ , then we can define  $W^k$  to be  $\mathcal{L}^2 \Omega^k(M)$  with an equivalent inner product. Then  $d^*$  becomes a new codifferential operator, relative to the modified inner product. Thus, general elliptic problems (at least without convection terms) may be put into the same framework, provided that we use the equivalent inner product.

**1.8.8 Definition** (Morphisms of Hilbert complexes). Let  $(W, d)$  and  $(W', d')$  be two Hilbert complexes.  $f : W \rightarrow W'$  is called a MORPHISM OF HILBERT COMPLEXES if we have a sequence of bounded linear maps  $f^k : W^k \rightarrow W'^k$  such that  $d'^k \circ f^k = f^{k+1} \circ d^k$  (they commute with the differential).

With the above, we can show the following WEAK HODGE DECOMPOSITION:

**1.8.9 Theorem** (Weak Hodge Decomposition Theorem). *Let  $(W, d)$  be a Hilbert complex with domain complex  $(V, d)$ . Then we have the  $W$ - and  $V$ -orthogonal decompositions*

$$(1.8.1) \quad W^k = \overline{\mathfrak{B}^k} \oplus \mathfrak{H}^k \oplus \mathfrak{Z}^{k \perp W}$$

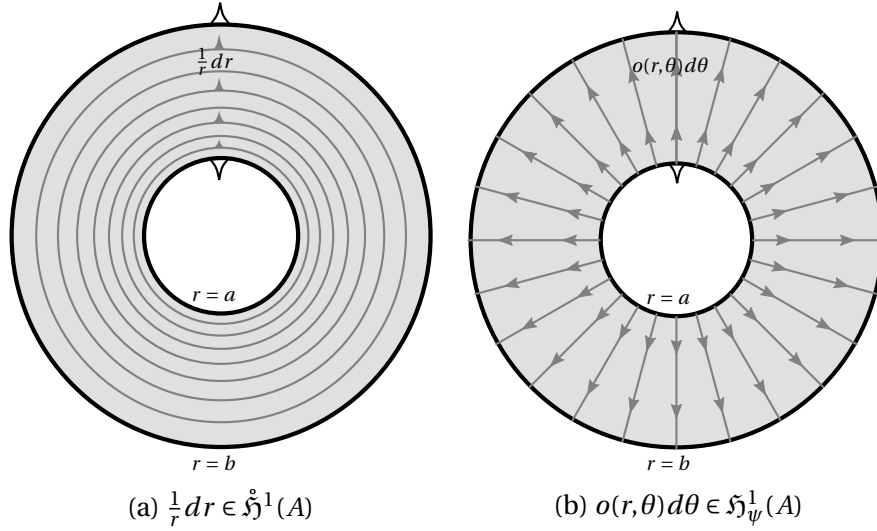
$$(1.8.2) \quad V^k = \overline{\mathfrak{B}^k} \oplus \mathfrak{H}^k \oplus \mathfrak{Z}^{k \perp V}.$$

where  $\mathfrak{Z}^{k \perp V} = \mathfrak{Z}^{\perp W} \cap V^k$ .

Of course, if  $\mathfrak{B}^k$  is closed, then the extra closure is unnecessary; it is referred to as the **STRONG HODGE DECOMPOSITION** or just **HODGE DECOMPOSITION**. We shall simply write  $\mathfrak{Z}^{k \perp}$  for  $\mathfrak{Z}^{k \perp V}$ , which will be the most useful orthogonal complement for our purposes. We note that by the abstract properties of adjoints [6, §3.1.2],  $\mathfrak{Z}^{k \perp W} = \overline{\mathfrak{B}_k^*}$ , and  $\mathfrak{B}^{k \perp W} = \mathfrak{Z}_k^*$ . This of course is also the generalization of the corresponding notions in the de Rham complex. We should note that the harmonic forms must incorporate the boundary conditions, so one must be careful, when computing them (and inferring topological results from them), to take note of those conditions, as noted in Example 1.5.12. See Figure 1.7.

**1.8.10 Example** (Harmonic forms for the de Rham Complex). If  $V = H\Omega(M)$ , then the harmonic forms  $\mathfrak{H}^k$  are  $\mathfrak{Z}^k \cap \mathfrak{Z}_k^*$ , that is  $\omega$  such that  $d\omega = 0$  and  $\delta\omega = 0$ , we must impose the additional requirement, since we have  $\mathfrak{Z}_k^* \subseteq \mathring{H}^* \Omega^k(M)$ , that  $\text{Tr} \star \omega = 0$ . For domains in  $\mathbb{R}^3$ , for example, taking  $k = 1$ , we get these 1-forms by “lowering the indices” of a proxy vector field (i.e. *work forms* in the terminology of [53]). This boundary condition says that the proxy vector field has vanishing normal component, as described in Example 1.5.10.

If  $k = 2$ , and the forms are of odd parity (2-pseudoforms), a harmonic 2-pseudoform is the contraction of the a proxy vector field with the volume pseudoform



**Figure 1.7:** Two generators for the harmonic forms for  $\mathfrak{H}^1(A)$  and  $\mathfrak{H}^1_\psi(A)$ , where  $A$  is the annulus  $\{a < r < b\} \subseteq \mathbb{R}^2$ , reflecting the different kinds of boundary conditions. Note how different they are, but at the same time, how they are dual in some sense, one having level sets that are the orthogonal trajectories of the other. Compare Figure 1.5.

(*flux form*). As noted in Example 1.5.9, this means the corresponding proxy vector field has vanishing tangential component.

**1.8.11 Example** (Harmonic forms for the de Rham Complex with boundary conditions). This time, we have the complex with  $V = \mathring{H}\Omega(M)$ , so the dual complex consists of the spaces  $V^* = H^*\Omega(M)$ , and thus the harmonic forms are  $\omega$  such that  $d\omega = 0$  and  $\delta\omega = 0$ , but now with  $\text{Tr}\omega = 0$  (and *not* its Hodge dual). It is easier to interpret the vanishing of the trace, since there is no dualization involved. If  $k = 1$ , then  $\text{Tr}\omega = 0$  means it vanishes on any vector tangent to the boundary. So the tangential component of the proxy field vanishes. If  $k = 2$ , then  $\text{Tr}\omega = 0$  means it vanishes on any pair of vectors tangent to the boundary, i.e. any parallelogram on the boundary. Since a parallelogram is perpendicular to the cross product of its sides, that means the normal component of the proxy field must vanish.

The following inequality is an important result crucial to the stability of our

solutions to the boundary value problems as well as the numerical approximations:

**1.8.12 Theorem** (Abstract Poincaré Inequality). *If  $(V, d)$  is a closed, bounded Hilbert complex, then there exists a constant  $c_P > 0$  such that for all  $v \in \mathfrak{Z}^{k\perp}$ ,*

$$\|v\|_V \leq c_P \|d^k v\|_V.$$

In the case that  $(V, d)$  is the domain complex associated to a closed Hilbert complex  $(W, d)$ ,  $(V, d)$  is again closed, and the additional graph inner product term vanishes:  $\|d^k v\|_V = \|d^k v\|$ . We now introduce the abstract version of the Hodge Laplacian and the associated problem.

**1.8.13 Definition** (Abstract Hodge Laplacian problems). We consider the operator  $L = dd^* + d^*d$  on a Hilbert complex  $(W, d)$ , called the **ABSTRACT HODGE LAPLACIAN**. Its domain is  $D_L = \{u \in V^k \cap V_k^* : du \in V_{k+1}^*, d^*u \in V^{k-1}\}$ , and the **HODGE LAPLACIAN PROBLEM** is to seek  $u \in V^k \cap V_k$ , given  $f \in W^k$ , such that

$$(1.8.3) \quad \langle du, dv \rangle + \langle d^*u, d^*v \rangle = \langle f, v \rangle$$

for all  $v \in V^k \cap V_k^*$ . This is simply the weak form of the Laplacian and any  $u \in V^k \cap V_k^*$  satisfying the above is called a **WEAK SOLUTION**. Owing to difficulties in the approximation theory for such a problem (it is difficult to construct finite elements for the space  $V^k \cap V_k^*$ ), Arnold, Falk, and Winther [6] formulated the **MIXED ABSTRACT HODGE LAPLACIAN PROBLEM** by defining auxiliary variables  $\sigma = d^*u$  and  $p = P_{\mathfrak{H}}f$ , the orthogonal projection of  $f$  into the harmonic space, and considering a *system* of equations,

to seek  $(\sigma, u, p) \in V^{k-1} \times V^k \times \mathfrak{H}^k$  such that

$$(1.8.4) \quad \begin{aligned} \langle \sigma, \tau \rangle - \langle u, d\tau \rangle &= 0 & \forall \tau \in V^{k-1} \\ \langle d\sigma, v \rangle + \langle du, dv \rangle + \langle p, v \rangle &= \langle f, v \rangle & \forall v \in V^k \\ \langle u, q \rangle &= 0 & \forall q \in \mathfrak{H}^k. \end{aligned}$$

The first equation is the weak form of  $\sigma = d^*u$ , the second is (1.8.3) modified to account for a harmonic term so that a solution exists, and the third enforces uniqueness by requiring perpendicularity to the harmonic space. With these modifications, the problem is well-posed by considering the bilinear form (writing  $\mathfrak{X}^k := V^{k-1} \times V^k \times \mathfrak{H}^k$ )  $B : \mathfrak{X}^k \times \mathfrak{X}^k \rightarrow \mathbb{R}$  defined by

$$(1.8.5) \quad B(\sigma, u, p; \tau, v, q) := \langle \sigma, \tau \rangle - \langle d\tau, u \rangle + \langle d\sigma, v \rangle + \langle du, dv \rangle + \langle p, v \rangle - \langle u, q \rangle.$$

and linear form  $F \in (\mathfrak{X}^k)'$  given by  $F(\tau, v, q) = \langle f, v \rangle$ . The form  $B$  is *not* coercive, but rather, for a closed Hilbert complex, satisfies the (LADYZHENSKAYA-BABUŠKA-BREZZI) INF-SUP CONDITION [6, 7]: there exists  $\gamma > 0$  (the STABILITY CONSTANT) such that

$$(1.8.6) \quad \inf_{(\sigma, u, p) \neq 0} \sup_{(\tau, v, q) \neq 0} \frac{B(\sigma, u, p; \tau, v, q)}{\|(\sigma, u, p)\|_{\mathfrak{X}} \|(\tau, v, q)\|_{\mathfrak{X}}} =: \gamma > 0.$$

where we have defined a standard norm on products:  $\|(\sigma, u, p)\|_{\mathfrak{X}} := \|\sigma\|_V + \|u\|_V + \|p\|$ .

This is sufficient to guarantee the well-posedness. To summarize, we have

**1.8.14 Theorem** ([6], Theorem 3.1). *The mixed variational problem (1.8.4) on a closed Hilbert complex  $(W, d)$  with domain  $(V, d)$  is well-posed: the bilinear form  $B$  satisfies the inf-sup condition, so for any  $F \in (X^k)'$ , there exists a unique solution  $(\sigma, u, p)$  to (4.2.4), i.e.,  $B(\sigma, u, p; \tau, v, q) = F(\tau, v, q)$  for all  $(\tau, v, q) \in \mathfrak{X}^k$ , and moreover*

$$\|(\sigma, u, p)\|_{\mathfrak{X}} \leq \gamma^{-1} \|F\|_{\mathfrak{X}'}$$



where  $\gamma$  is the stability constant; it depends only on the Poincaré constant.

Note that the bilinear form allows us to use any linear functional  $F \in (\mathfrak{X}^k)'$ , namely, there may be other nonzero quantities on the RHS of (4.2.4) besides  $\langle f, v \rangle$ . We shall need this result for parabolic problems.

One of the key ingredients in proving Theorem 1.8.14 is also something that we shall need, so we recall it here.

**1.8.15 Lemma.** *The inf-sup condition implies the existence and uniqueness of the solution as well as an a priori estimate: given  $B : H \times H \rightarrow \mathbb{R}$  satisfying the inf-sup condition*

$$\inf_{u \neq 0} \sup_{v \neq 0} \frac{|B(u, v)|}{\|u\|_H \|v\|_H} = \gamma > 0,$$

and  $F \in H'$ , there exists a unique  $u \in H$  such that

$$B(u, v) = F(v),$$

and moreover,  $\|u\|_H \leq \gamma^{-1} \|F\|_{H'}$ .

This is essentially an extension of the Lax-Milgram theorem for bilinear forms satisfying the inf-sup condition rather than coercivity.

*Proof of the lemma.* We base our proof on a modification of the argument in [30, §6.2.1] for the proof of the Lax-Milgram theorem (Theorem 1.7.2). Babuška [7] proves it in a bit more generality, in particular, when the two factors defining the bilinear form are not the same (i.e. if it is Petrov-Galérkin vs. just a Galérkin method—see §2.1.2). We assume  $B : H \times H \rightarrow \mathbb{R}$  is a bounded, symmetric, bilinear form satisfying the inf-sup condition:

$$\inf_{u \neq 0} \sup_{v \neq 0} \frac{|B(u, v)|}{\|u\|_H \|v\|_H} = \gamma > 0.$$

We use the same key tactic, namely to show that  $B$  is the inner product with a bounded linear operator  $A$  acting in one factor, then showing this operator  $A$  has closed range, which in fact must be the whole space, so that it is surjective; the existence and *a priori* estimate follows from the Riesz representation theorem.

Given  $w$ , the mapping  $v \mapsto B(w, v)$  is a bounded linear functional:

$$\sup_v B(w, v) \leq M \|w\|_H \|v\|_H$$

as before, so that the Riesz representation theorem says that there exists a unique  $Aw$  such that  $\langle Aw, v \rangle_H = B(w, v)$ , just as in the proof of the Lax-Milgram theorem. Moreover,  $\|Aw\|_H = \|B(w, \cdot)\|_{H'} \leq M \|w\|_H$ , so  $A$  is a bounded linear operator. To show that the range is closed, we first show that  $A$  is bounded away from zero. Since  $B(w, v) = \langle Aw, v \rangle_H$  for all  $v$ , the inf-sup condition implies that there exists  $v \neq 0$  such that  $|B(v', v)| \geq \gamma \|v'\|_H \|v\|_H$ , for all  $v' \in H$ . This means, in particular, it is true for  $v' = w$ :

$$\gamma \|w\|_H \|v\|_H \leq |B(w, v)| = |\langle Aw, v \rangle_H| \leq \|Aw\|_H \|v\|_H,$$

which, after canceling the  $\|v\|_H$ , gives  $\|Aw\|_H \geq \gamma \|w\|_H$ . Any sequence in the range of  $A$ , therefore, satisfies  $\|u_n - u_m\|_H \leq \gamma^{-1} \|Au_n - Au_m\|_H$ , so in particular, if the range sequence is Cauchy, so is the domain sequence, and converges to  $u^*$ ; the boundedness of  $A$  implies that the range sequence must converge to  $Au^*$ , just as in the proof of the Lax-Milgram theorem.

To show that the range is the entire space, we argue  $R(A)^\perp$  is the zero space. If  $w \in R(A)^\perp$ , then  $w \in R(A)^\perp$  means that given the same  $v$  witnessing the inf-sup condition as above,

$$\gamma \|w\|_H \|v\|_H \leq |B(w, v)| = |B(v, w)| = |\langle Av, w \rangle_H| = 0,$$

so that, since  $\gamma \neq 0$  and  $v \neq 0$ , we have  $w = 0$ . Here, the symmetry is vital, because it allows us to move  $A$  to the other factor in the inner product. It is unnecessary to consider this in the coercive case since there we used the same variable in both slots. So now, given  $F \in H'$ , we have that there exists a unique  $w'$  such that  $\langle w', v \rangle_H = F(v)$ , by the Riesz representation theorem as before, with  $\|w'\|_H = \|F\|_{H'}$ . Since  $A$  is surjective,  $w' = Au$ , and  $B(u, v) = \langle Au, v \rangle_H = F(v)$ . Thus

$$\|u\|_H \leq \gamma^{-1} \|Au\|_H = \gamma^{-1} \|w'\|_H = \gamma^{-1} \|F\|_{H'}.$$

□

**1.8.16 Compactness properties.** Finally, we make a note of some compact embedding properties of the spaces relevant to our purposes (following [6, §3.1.3]; see also [84]). The crucial property for our purposes is the compactness of the embedding  $V^k \cap V_k^* \hookrightarrow W^k$ ; complexes satisfying this are said to have the COMPACTNESS PROPERTY. This is the analogue for Hilbert complexes to the Rellich-Kondrachov theorem [30, §5.7] for elliptic equations for functions, and is how we establish that the Sobolev spaces relevant to problems on manifolds (namely,  $W^k = \mathcal{L}^2\Omega^k$  and  $V^k = H\Omega^k$ , etc.) are closed complexes.  $V^k \cap V_k^*$  has a natural norm combining the graph norms of both the  $V$  and  $V^*$  complexes, which reduces to the  $W^k$  norm (times a constant) on the harmonic space  $\mathfrak{H}^k = \mathfrak{Z}^k \cap \mathfrak{Z}_k^*$ . If the embedding is compact, then, restricted to  $\mathfrak{H}^k$ , it says the identity is compact—compactness of the identity is equivalent saying that  $\mathfrak{H}^k$  is finite-dimensional. Since  $\mathfrak{H}^k \cong \mathfrak{Z}^k / \overline{\mathfrak{B}^k}$  is finite-dimensional (a complex whose cohomology satisfies this property is referred to as being FREDHOLM), this implies  $\mathfrak{B}^k$  is closed in the  $\mathcal{L}^2$  norm. This says precisely that the complex is closed.

Compactness of the embeddings  $H\Omega \cap \mathring{H}^*\Omega$  follows from the usual Rellich-Kondrachov theorems *if* our manifolds are smooth, because in that case, the intersec-

tion actually lies in  $H^1\Omega$ , and so that theorem applies componentwise. For the case of Lipschitz boundaries (which is important for us, because our most common case is a shape-regular triangulation), this containment is false, but [84] establishes the result anyway. The essence of the argument in [84] is that the property is invariant under Lipschitz mappings locally (and in particular, it is independent of the possibly different metrics and thus different coefficients for the elliptic problem), so the property continues to hold on all Lipschitz manifolds (even if the intersection fails to be  $H^1$ ).

## 1.9 Evolutionary Partial Differential Equations

In many cases, it is informative to regard time-dependent partial differential equations (usually called EVOLUTIONARY PARTIAL DIFFERENTIAL EQUATIONS) as, actually, an *ordinary* differential equation in a *function space*: a solution  $u : M \times [0, T] \rightarrow \mathbb{R}$  can be thought of as a curve whose value at the time  $t$  is a function of space:

$$u(t)(x) = u(x, t),$$

i.e.  $u$  can be viewed as a function  $u : [0, T] \rightarrow \mathfrak{X}$ , where  $\mathfrak{X}$  is some (Banach) space of functions. We have often also used (and shall continue to use) the notation  $u(\cdot, t)$  for the value of  $u(t)$  as a function of the remaining slot where the dot is placed. This is in contrast to regarding the function as being defined on one single domain in spacetime (this is a very useful viewpoint as well, of course, and is one of the principal goals for future work). The theory of flows in standard ODE theory does in fact generalize to these cases of infinite-dimensional spaces, and in particular, we have Picard-like local existence theorems [61] in normed spaces. However, in many cases of interest, such as parabolic equations, the hypotheses are not satisfied, because the operators may not map back into the same space, at least with respect to the norms we want. Again,

as we did for elliptic problems in the previous chapter, we start out with the concrete motivations, and build our way to more abstract, clarifying theories, attempting to build bridges along the way.

### 1.9.1 Motivation: The Heat Equation

**1.9.1 Example and description of the issues.** We start with our standard model problem, the heat equation. Consider a bounded domain  $U \subseteq \mathbb{R}^n$ . We now consider the following equation for some  $u : U \times [0, T] \rightarrow \mathbb{R}$ :

$$\frac{\partial u}{\partial t} = \Delta u + f$$

where  $f : M \times [0, T] \rightarrow \mathbb{R}$  is some source term, for some boundary conditions in the space variable of  $u$ , and for some initial condition  $u(0) = g$ . We have deliberately not been precise about which function spaces we need our solution to lie in, because it is actually a subtle question. So, the objective of this example and indeed, this subsection is to establish exactly what kind of space we can formulate our problem, and see why we need a more comprehensive solution theory for our needs than the Picard-type theorems previously established. As we have seen, the Laplace operator  $\Delta$ , in general, maps the space  $H_0^1(U)$  into  $H^{-1}(U)$ . Generally, the Laplace operator maps a smaller space into a larger one because we lose regularity when applying the Laplacian. Even if, say, we define it from  $H^2$  into  $\mathcal{L}^2$ ,  $\Delta$  is not bounded if the same norm is used for both. This leads to a situation in which we cannot define the contraction operator that is instrumental in proving the Picard theorem.

**1.9.2 Using the weak form.** One way to proceed is to, of course, take advantage of the notion of weak formulations (actually, we already have done so, in saying  $\Delta$  maps into  $H^{-1}$ ): if we assume that at each time  $u(t) \in H_0^1(U)$ , or  $u : [0, T] \rightarrow H_0^1(U)$ , then, we

can infer from the heat equation what kinds of objects we should require of  $\frac{\partial}{\partial t}$  and  $f$ , given that we know where  $\Delta u$  lies. Since  $\Delta u(t) \in H^{-1}(U)$ , the equation tells us that  $\frac{\partial u}{\partial t}(t)$  and  $f(t) \in H^{-1}(U)$  also. As such, though  $u(t)$  itself may lie in  $H_0^1(U)$ , its time derivative must lie in the larger space  $H^{-1}(U)$ . This means, in particular, that though  $\Delta$  is ostensibly solely a spatial operator, its properties force an interaction between time and space derivatives, in the sense that  $\frac{\partial u}{\partial t}$  may, *a priori*, lie in a larger function space. Difference quotients are supposed to be definable for anything in the same function space, so it becomes a question of, in what sense, is the limit

$$\lim_{h \rightarrow 0} \frac{u(t+h) - u(t)}{h}$$

is to be taken (we never had to worry about this in the finite-dimensional case, since all norms on finite-dimensional spaces are equivalent!).

The technique we develop here, following [30, Ch. 7] and later, a more abstract generalization, [89, Ch. 11] relies on weakening the time derivative in some sense, as well (since weakening has been such a successful strategy for spatial equations, it is not surprising that it would enter into evolutionary equations as well). This requires some results on the integration theory of Banach-space valued functions [30, App. E]. The most basic definition is, of course,  $C(I, X)$ , continuous functions from  $I = [0, T]$  to  $X$ , which is definable since  $X$  has a metric and topology defined by the norm. Using similar notions of integration of simple functions, we define an integral, notions of measurability, and analogues of Lebesgue and Sobolev spaces (called BOCHNER SPACES. The spaces are also often said to be *time-parametrized Banach spaces*, although we reserve that term for a more literally evolving space (one of the goals of future work).

## 1.9.2 Bochner Spaces

In order to solve and approximate linear evolution problems, we introduce the framework of Bochner spaces, which realizes time-dependent functions as curves in Banach spaces (which will correspond to spaces of spatially-dependent functions in our problem). We continue the discussion in [30, Ch. 7], and introduce its more abstract counterpart as in [89, Ch. 11].

Let  $X$  be a Banach space and  $I := [0, T]$  an interval in  $\mathbb{R}$  with  $T > 0$ . We define

$$C(I, X) := \{u : I \rightarrow X \mid u \text{ bounded and continuous}\}.$$

In analogy to spaces of continuous, real-valued functions, we define a supremum norm on  $C(I, X)$ , making  $C(I, X)$  into a Banach space:

$$\|u\|_{C(I, X)} := \sup_{t \in I} \|u(t)\|_X.$$

We will of course need to deal with norms other than the supremum norm, which motivates us to define BOCHNER SPACES: to define  $\mathcal{L}^p(I, X)$ , we complete  $C(I, X)$  with the norm

$$\|u\|_{\mathcal{L}^p(I, X)} := \left( \int_I \|u(t)\|_X^p dt \right)^{1/p}.$$

Similarly, we have the space  $H^1(I, X)$ , the completion of  $C^1(I, X)$  with the norm

$$\|u\|_{H^1(I, X)} := \left( \int_I \|u(t)\|_X^2 + \left\| \frac{d}{dt} u(t) \right\|_X^2 dt \right)^{1/2}.$$

As mentioned before, there are more measure-theoretic notions which define the integral of Banach space-valued functions ([30, App. E]) and consider Lebesgue-measurable subsets of  $I$ . In particular, we make use of two key principles (which are

equivalent [38] for separable spaces, the case we are considering):

**1.9.3 Definition.** We say that  $u \in \mathcal{L}^2(I, X)$  has a WEAK DERIVATIVE  $v \in \mathcal{L}^2(I, X)$  (i.e.,  $H^1(I, X)$ ) if either of the two conditions hold:

1. (the Bochner integral method [30, App. E]) For all  $\phi \in C_c^\infty(I)$ ,

$$\int_I u(t)\phi'(t) dt = - \int_I v(t)\phi(t) dt.$$

2. (the distribution theory method [89, Ch. 11], [38]) Supposing  $X$  is a Hilbert space and defining  $\mathcal{D}(I, X)$  to be all classically differentiable functions of  $I$  into  $X$  with compact support, where the limit in the difference quotient is taken to be in the norm of  $X$  (i.e. the FRÉCHET DERIVATIVE), we have that for all  $w \in \mathcal{D}(I, X)$ ,

$$\int_I \langle u(t), w'(t) \rangle_X dt = - \int_I \langle v(t), w(t) \rangle_X dt.$$

The latter, of course, does not require any integration theory other than the usual Lebesgue theory on the line.

The usual setting, of course, is that  $X$  will be some space of functions depending on space, and the time-dependence is captured as being a curve in this function space (although this interpretation is only correct when we are considering  $C(I, X)$ —we must be careful about evaluating our functions at single points in time without an enclosing integral). Usually,  $X$  will be a space in some Hilbert complex, such as  $L^2\Omega^k(M)$  or  $H^s\Omega^k(M)$  where the forms are defined over a Riemannian manifold  $M$ .

**1.9.4 Definition (Rigged Hilbert Space).** We introduce this abstract framework in order to be able to formulate parabolic problems more generally. It turns out to be useful to consider the concept of RIGGED HILBERT SPACE or GELFAND TRIPLE, which consists of



a triple of separable Banach spaces

$$V \subseteq H \subseteq V^*$$

such that  $V$  is continuously and densely embedded in  $H$  and  $V^*$  is the dual space of  $V$  as a space of linear functionals. For example, if  $(V, d)$  is the domain complex of some Hilbert complex  $(W, d)$ , setting  $V = V^k$  and  $H = W^k$  works, as well as various combinations of their products (so that we can use mixed formulations).  $H$  is also continuously embedded in  $V^*$ . As another example, this is the proper setting of quantum mechanics, where  $H$  is  $\mathcal{L}^2$  as before, but now  $V$  is the Schwartz space and  $V^*$  is the space of tempered distributions. This legitimizes the use of many objects such as the Dirac delta, despite that they are not members of the Hilbert space  $\mathcal{L}^2$ .

**1.9.5 Warning about the use of the Riesz representation theorem.** The standard isomorphism (given by the Riesz representation theorem) between  $V$  and  $V^*$ , is not generally the composition of the inclusions, because the primary inner product of importance for weak formulations is the  $H$ -inner product. It coincides with the notion of distributions acting on test functions. Writing  $\langle \cdot, \cdot \rangle$  for the inner product on  $H$ , the setup is designed so that when it happens that some  $F \in V^*$  is actually in  $H$ , we have  $F(v) = \langle F, v \rangle$  (which is why we will often write  $\langle F, v \rangle$  to denote the action of  $F$  on  $v$  even if  $F$  is not in  $H$ ). In fact, in most cases of interest, the  $H$ -inner product is the restriction of a more general bilinear form between two spaces, in which elements of the left (acting) space are of less regularity than elements of  $H$ , while elements of the right space have more regularity. This motivation means  $H$  is identified with its own dual  $H^*$ , but we will *not* be using this identification for  $V$  and  $V^*$ .

**1.9.6 Explicit characterization of the maps and proof of density.** An explicit characterization of the map from  $H$  into  $V^*$  is the *adjoint* of the inclusion  $i : V \hookrightarrow H$ :

$\langle i^* v, w \rangle = \langle v, i(w) \rangle = \langle v, w \rangle$ , namely,  $i^*$  operates on linear functionals on  $H$  (identified with  $H$ ) as *restriction* to  $V$ . We should show that it actually extends boundedly, namely, that a restricted linear functional from  $H$  still gives bounded  $V$ -norm:

$$\|i^* F\|_{V^*} = \sup_{\|v\|_V \leq 1} |\langle F, i(v) \rangle| \leq \sup_{\|v\|_V \leq 1} \|F\|_H \|v\|_H \leq \sup_{\|v\|_V \leq 1} C_e \|F\|_H \|v\|_V \leq C_e \|F\|_H$$

where  $C_e$  is the embedding constant, i.e., bound for  $i$ : such that  $\|i(v)\|_H \leq C_e \|v\|_V$  that witnesses the continuity of the inclusion  $i$ . The density of  $V$  implies the injectivity of the mapping  $i^*$ , because  $i^* F = i^* G$  implies  $F = G$  on  $V$ , a dense subset, and thus by the continuity of the linear functionals  $F$  and  $G$ , they must agree on all of  $H$ .

That  $H$  is dense in  $V^*$  is a consequence of the fact that  $V$  is a reflexive Banach space (due to it having a Hilbert space structure), so that  $V^{**}$  is isomorphic to  $V$ . If  $v$  and  $w$  agree on  $H$  acting as the  $H$ -inner product, it follows that  $v - w$  is orthogonal to all of  $H$ , and  $v - w \in H$  also. This means, since  $H$  is complete, that  $v - w = 0$ . If, now, there is some  $w \in V^*$  that is of minimal, positive distance from  $\overline{H}$ , the Hahn-Banach theorem [30, Ch. 5] means there is  $v \in V^{**} = V$  such that  $\|v\|_V = 1$  and  $v$  vanishes on all of  $H$ , i.e. it agrees with the zero function on all of  $H$ , so must be zero, a contradiction.

Given  $A \in C(I, \mathcal{L}(V, V^*))$ , a time-dependent linear operator, we define the bilinear form

$$(1.9.1) \quad a(t, u, v) := \langle -A(t)u, v \rangle,$$

for  $(t, u, v) \in \mathbb{R} \times V \times V$ . As with the bilinear form theory described above in elliptic problems,  $a$  needs to satisfy some kind of coercivity condition for the theory to work. Elliptic problems, however, are concerned with inverting some operator, while parabolic problems do not have that same challenge—we'll see this very concretely

when we talk about numerical methods. So the condition we need on  $a$  is, not surprisingly, weaker than strict coercivity. It turns out that Gårding's Inequality, which played a role in the general existence theory in §1.7.2, is the right condition to use here:

$$(1.9.2) \quad a(t, u, u) \geq c_1 \|u\|_V^2 - c_2 \|u\|_H^2,$$

with  $c_1, c_2$  constants independent of  $t \in I$ . Then the following problem is the abstract version of linear, parabolic problems:

$$(1.9.3) \quad u_t = A(t)u + f(t)$$

$$(1.9.4) \quad u(0) = u_0.$$

This problem is well-posed:

**1.9.7 Theorem** (Existence of Unique Solution to the Abstract Parabolic Problem, [89], Theorem 11.3). *Let  $f \in \mathcal{L}^2(I, V^*)$  and  $u_0 \in H$ , and  $a$  the time-dependent quadratic form in (1.9.1). Suppose (1.9.2) holds. Then the abstract parabolic problem (1.9.3) has a unique solution*

$$u \in \mathcal{L}^2(I, V) \cap H^1(I, V^*).$$

*Moreover,  $u \in C(I, H)$  by the Sobolev embedding theorem, which allows us to unambiguously evaluate the solution at time zero, so the initial condition makes sense, and the solution indeed satisfies it:  $u(0) = u_0$ .*

**1.9.8 Key concepts in the proof.** The standard method ([89, p. 382] and [30, §7.2, for the the specific case of  $V = H_0^1(U)$ ]) is as follows: We take an orthonormal basis of  $H$  that is simultaneously orthogonal for  $V$  (a frequent situation occurring when, say, it is an orthonormal basis of eigenfunctions of the Laplace operator), formulate the problem in the finite-dimensional subspaces, and use *a priori* bounds on such

solutions to extract a weakly convergent subsequence via the Banach-Alaoglu theorem ([34, Ch. 4], [30, App. D]). That weak limit is then shown to actually satisfy the equation.

## Chapter 2

# Numerical Methods for Solving Partial Differential Equations

As one can see in the preceding theory, solving PDES analytically is often very tricky, if not impossible. There are several useful techniques that involve either explicit solutions or may be used to derive properties of solutions without knowing how to actually compute them (which of course may be sufficient for many purposes). However, being able to at least visualize some form of solution accurately is useful not only pedagogically, but also theoretically, as it can be then used to generate conjectures and seek new useful properties. Here we shall describe a kind of numerical method that is good for geometric analysis: the FINITE ELEMENT METHOD. There are other methods based on taking approximate difference quotients (FINITE DIFFERENCING), which are also important and useful, and in fact also have interesting visualizations, many of which are closer to the notions studied in algebraic topology. However, our goal in this work is to understand and apply the tools of modern analysis toward solving the PDES we encounter, so the finite element method is better suited for us. We mostly follow Braess [11] for the basics, tying them to the framework created for

differential forms, of Arnold, Falk, and Winther [5, 6], whose work we build upon in this work (and some of which has already been seen here for the proper formulation of many of these problems in terms of differential forms).

## 2.1 The Finite Element Method

The FINITE ELEMENT METHOD (FEM) is a method of numerically solving partial differential equations by reducing the (usually intractable) problem of *infinite-dimensional* linear algebra to (more tractable) finite-dimensional linear algebra by means of choosing appropriate subspaces of the relevant function spaces (usually Sobolev spaces). The method has several advantages over the more straightforward-seeming finite-difference methods, and it is especially suited to our needs because, first, it handles domains with complicated geometries quite well, and, it also works with the weak form of the PDES, enabling us to use modern methods of analysis [30, 34], to prove that our approximations are good. Also, weak formulations yield less stringent conditions on smoothness. The basic idea is very simple: we simply choose a finite-dimensional subspace of the relevant function space, and find the best approximation to the solution by using matrix equations set up by the weak form. We allow weak solutions not only because some equations (namely, hyperbolic ones) allow for non-smoothness in the initial data to be propagated over time, but more fundamentally, some of the most natural choices of approximating spaces, such as piecewise linear functions, may not consist of classically differentiable functions. The general method of approximating solutions this way is called the GALËRKIN METHOD. This, in turn, is motivated via minimization (over the finite-dimensional subspace) of the corresponding functionals (the RAYLEIGH-RITZ METHOD). The quality of the solution is, of course, dependent on the choice of appropriate basis functions—the

finite element method is a Galërkin method using bases (which are usually piecewise polynomial functions) constructed from geometrical properties of the domain.

### 2.1.1 The Rayleigh-Ritz Method

The Rayleigh-Ritz method [11, Ch. 2, §2] is a good way of motivating many of the constructions with the weak form of the differential equations. As noted before in Chapter 1, the idea is to realize the solution to a differential equation as a critical point of some functional on our spaces. The Rayleigh-Ritz method simply reduces this possibly intractable minimization (or at least critical point-seeking) problem (over an infinite-dimensional space) to a finite-dimensional one, where all the standard techniques of calculus can apply.

#### 2.1.1 Motivational example: variational form of the symmetric linear elliptic PDE.

Recall the standard variational calculus setup that we have explored in earlier chapters: we have a functional  $J$  for which the Euler-Lagrange equations give us the PDE on a domain  $U \subseteq \mathbb{R}^n$ , or a manifold-with-boundary. To recap, let's take the example corresponding to a linear elliptic PDE (using, as always, the Einstein summation convention):

$$\begin{aligned} J[u] &= \int_U \left( \frac{1}{2} a^{ij}(x) \partial_i u(x) \partial_j u(x) + \frac{1}{2} c(x) u(x)^2 - f(x) u(x) \right) dx \\ &= \int_U \left( \frac{1}{2} A(du, du) + \frac{1}{2} cu^2 - fu \right) dx. \end{aligned}$$

for symmetric, positive-definite matrix  $(a^{ij}(x))$ , and  $c(x) > 0$  (physically: diffusion with a proportional sink); hence we need not worry about boundary conditions for this example. We take the domain of  $J$  to be in the appropriate Sobolev space,  $H^1(U)$ . As noted before, this has a realization on abstract Hilbert complexes (§1.8 above) by taking  $a^{ij}$  to be a metric, considering  $W = \mathcal{L}^2(U)$ ,  $V = H^1(U)$ , and with a modified

inner product obtained by integrating  $a^{ij}\partial_i u \partial_j v + cuv$ . Then *all* the results of that section apply. Nevertheless, to help connect things up to the standard presentation of the theory, we note the computational aspects in components.

Of course, if we assume for the moment that we have enough differentiability, this gives the Euler-Lagrange equations in divergence form (the strong form of the equation):

$$-\partial_i(a^{ij}\partial_j u) + cu - f = 0.$$

or, in decreasing order of abstractness,

$$d^* du + cu = \delta(A(du)) + cu = -\nabla \cdot (A(\nabla u)) + cu = f.$$

$A$  is then a tensor that sends the differentials into the dual space, so producing a vector field for each  $du$ , which, recall, corresponds to constitutive relations. Computationally, it is a matrix-valued function defined by  $A(x)\xi = a^{ij}(x)\xi_i \mathbf{e}_j$ .

**2.1.2 Reduction to a finite-dimensional problem.** Now suppose we choose a basis of functions  $\{\varphi_i\}_{i=1}^N$ , which span a subspace  $V_h$  of  $H^1(U)$  (it is standard in FEM to use a subscript  $h$ , which denotes the mesh size). The goal now is to minimize the functional in this subspace: minimize

$$J \left[ \sum_k u^k \varphi_k \right]$$

with respect to the (finitely many!) variables  $(u^k)$ . The notation  $u^k$  has been chosen, in particular, to be reminiscent of vectors, simply because this is now a kind of “vector” in a finite-dimensional function space. The finite element method is vitally concerned about the corresponding dual spaces as well, so it is helpful to keep the distinction.

After this reduction to a finite-dimensional situation, minimization of this functional is subject to the usual requirements of multivariable calculus: take the



gradient with respect to the variables ( $u^k$ ) and set it to zero; additionally, one can check if the second derivative matrix (the Hessian) is positive-definite. In our standard case,  $a^{ij}(x)$  being positive-definite shows that it is indeed a minimum. We go through the details:

$$\begin{aligned} J \left[ \sum_k u^k \varphi_k \right] &= \sum_{k,\ell} \int_U \frac{1}{2} a^{ij}(x) u^k u^\ell \partial_i \varphi_k(x) \partial_j \varphi_\ell(x) + \frac{1}{2} u^k u^\ell c(x) \varphi_k(x) \varphi_\ell(x) - u^k f(x) \varphi_k(x) dx \\ &= \sum_{k,\ell} K_{k\ell} u^k u^\ell - u^k F_k = \frac{1}{2} \mathbf{u}^T \mathbf{K} \mathbf{u} - \mathbf{u} \cdot \mathbf{F}. \end{aligned}$$

where we have defined the matrix

$$K_{k\ell} = \int_U (a^{ij}(x) \partial_i \varphi_k \partial_j \varphi_\ell + c(x) \varphi_k \varphi_\ell) dx$$

and

$$F_k = \int_U f(x) \varphi_k(x) dx.$$

$K$  is clearly a symmetric matrix. It is positive-definite because  $\mathbf{u}^T \mathbf{K} \mathbf{u}$  is the integral of two always positive quantities (a consequence of the positive-definiteness of  $a^{ij}$  as well as of  $c$ ).

**2.1.3 Expressing the problem in terms of the bilinear weak form.** A very important point is to realize that  $K_{k\ell}$  is simply the matrix formed by considering the bilinear weak form of the differential equation evaluated on the basis:

$$B(u, v) := \int_U (a^{ij}(x) \partial_i u(x) \partial_j v(x) + c(x) u(x) v(x)) dx$$

and  $F(v) = \int_U f(x) v(x) dx$ . This is also just the inner product on our abstract Hilbert

complex  $W$ . This should be familiar from Chapter 1—the weak formulation is to seek  $u$  in  $H^1(U)$  such that for all  $v \in H^1(U)$ ,  $B(u, v) = F(v)$ . The Rayleigh-Ritz method, and more generally, the Galërkin method, reduces this to the problem of seeking a solution  $u_h = \sum_k u^k \varphi_k \in V_h$  such that

$$B(u_h, \varphi_k) = F(\varphi_k).$$

for all  $k$  (the function we seek is also only tested against functions in the subspace  $V_h$ , for otherwise the problem would still be infinite-dimensional!). It is nice how it corresponds exactly to the minimization condition (condition for a critical point) for the functional when such a functional exists. Anyway, we have not actually shown that the bilinear form equation is what we want: this is made plain by actually differentiating:

$$\begin{aligned} \frac{\partial}{\partial u^j} J \left[ \sum_i u^i \varphi_i \right] &= \sum_{k,\ell} \frac{\partial}{\partial u^j} \left( \frac{1}{2} K_{k\ell} u^k u^\ell - u^k F_k \right) \\ &= \frac{1}{2} \sum_{k,\ell} K_{k\ell} \delta_j^k u^\ell + \frac{1}{2} \sum_{k,\ell} K_{k\ell} \delta_j^\ell u^k - \sum_k \delta_j^k F_k = \sum_i K_{ij} u^i - F_j. \end{aligned}$$

(we have used the symmetry of  $K$  in that last equation). Writing it all as a matrix equation,

$$\nabla J \left[ \sum_\ell u^\ell \varphi_\ell \right] = K\mathbf{u} - \mathbf{F}.$$

For reasons soon to be described,  $K_{k\ell}$  is called the **STIFFNESS MATRIX**. The minimization condition is now a linear algebra problem: solve  $K\mathbf{u} - \mathbf{F} = 0$ , or  $K\mathbf{u} = \mathbf{F}$ . This has a solution, because, for  $c \geq 0$ ,  $K$  is positive-definite, therefore invertible. Writing  $K\mathbf{u} - \mathbf{F}$  in terms of its components, and using the definition of the matrix, we see that this is exactly solving the bilinear form equation  $B(\sum_k u^k \varphi_k, \varphi_j) = F(\varphi_j)$  where  $u = \sum_k u^k \varphi_k$ . Finally, that this really is a minimum comes from calculating the second derivative, which is just the matrix  $K$ .

### 2.1.2 The Galérkin Method

The Galérkin method picks up right at the observation above about the weak bilinear form associated to a differential equation. Namely, given some linear elliptic partial differential equation in weak (divergence) form,

$$\begin{aligned} B(u, v) &= \int_U a^{ij}(x) \partial_i u(x) \partial_j v(x) + b^j \partial_j u(x) v(x) + c(x) u(x) v(x) \\ &= \int_U A(du, dv) + (b \lrcorner du) v + cuv \, dx \end{aligned}$$

and

$$F(v) = \int_U f(x) v(x) \, dx,$$

or perhaps  $F \in H^{-1}(U)$ , we wish to solve for a function  $u \in H_0^1(U)$  such that

$$B(u, v) = F(v)$$

holds for all  $v \in H_0^1(U)$ . Note that the bilinear form is no longer necessarily symmetric (even if  $a^{ij}$  is), or positive-definite, so this does not necessarily have to come from a variational problem. The PETROV-GALÈRKIN METHOD is to find  $u_h \in V_h$  a subspace of  $H_0^1(U)$  such that for all  $w_h \in W_h$  (another finite-dimensional space),

$$B(u_h, w_h) = F(w_h)$$

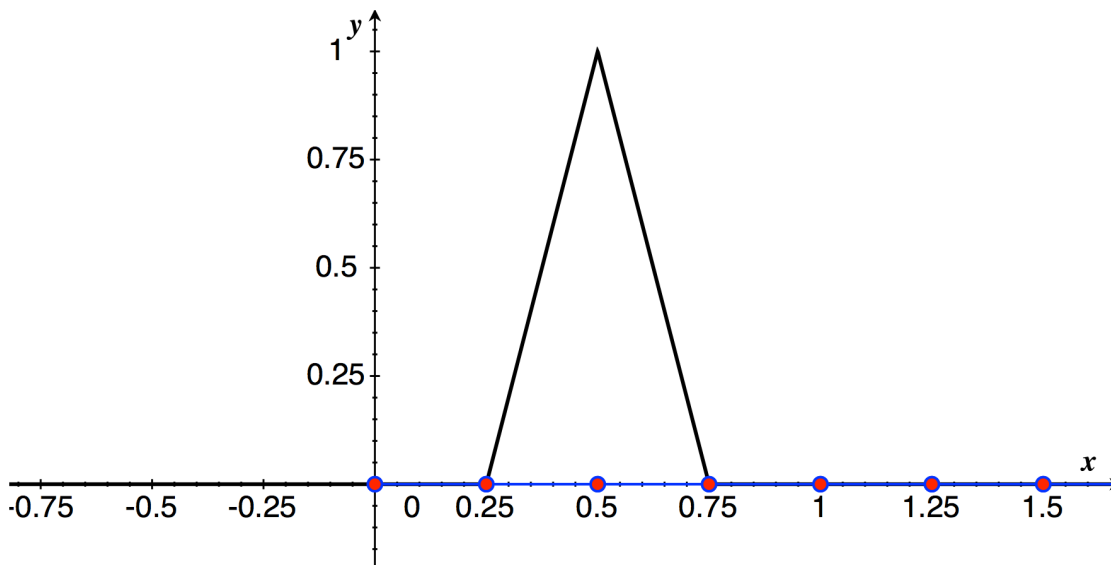
holds. Choosing bases  $\varphi_i$  and  $\psi_j$  for  $V_h$  and  $W_h$  respectively, we simply need to solve for  $u^k$  in

$$u^k B(\varphi_k, \psi_j) = F(\psi_j).$$

for all  $j$ . A GALËRKIN METHOD takes  $V_h = W_h$  and the same basis (and when it arises from a minimization problem, is simply the Rayleigh-Ritz method). We usually focus our attention on GalËrkin methods. As before, we can abbreviate  $B(\varphi_k, \varphi_j) = K_{kj}$ , the stiffness matrix, although it isn't necessarily symmetric or positive-definite anymore. The problem reduces to linear algebra as before: solving  $K\mathbf{u} = \mathbf{F}$ . It is not always straightforward to show that  $K$  is invertible, though, and it in fact may not be, until we do suitable restrictions of our Hilbert spaces (recall the process in the Hodge decomposition theorem).

## 2.2 Details of the Finite Element Method

We now get to some more specific details about the finite element method. We mostly follow [11, 54]. As noted before, theoretically, FEM is simply the (Petrov-)GalËrkin method with a specific choice of basis. Let  $U \subseteq \mathbb{R}^n$  be an open set with a smooth boundary (actually, we can get away with Lipschitz continuity, but we assume smoothness for now to motivate things). Suppose we have a TRIANGULATION of  $\bar{U}$ , that is a decomposition of  $U$  into  $n$ -simplices (often we just say  $U$  has been DISCRETIZED with a MESH). We will assume, also, that the mesh is CONFORMING: all the vertices only meet other simplices in other vertices, that is, no vertex of one simplex meets another along an edge, or face, and similarly edges only meet in other edges, and faces meet only in other faces, and so on. The diameter of the largest triangle in the triangulation is called the MESH SIZE or MESH PARAMETER and usually denoted with the letter  $h$  (quantities that depend on the triangulation, such as various approximations, are often subscripted with  $h$  to emphasize the dependence). As we shall see in a later section, there are a number of ways one can do triangulations, and there are various theorems in topology that guarantee that this can be done. Each simplex in the triangulation



**Figure 2.1:** Example tent function constructed for the node  $\frac{1}{2}$ ; where the nodes in the mesh are are  $\frac{k}{4}$ ,  $k = 0, \dots, 6$ .

is also called a (FINITE) ELEMENT. Here the word *finite* is used to distinguish it not from *infinite*, but rather *infinitesimal*, a use which is common among physicists and engineers. Mathematicians prefer to refer to things as being *discrete* or as having been *discretized* rather than as being *finite*.

### 2.2.1 The Basis

The triangulation of the domain enables us to choose a basis. First, suppose there are  $N$  vertices in the triangulation, and denote them by  $x_{(k)}$ . In the simplest FEM, we choose our basis to be piecewise linear and globally continuous. They are *uniquely* specified by the condition that  $\varphi_i(x_{(j)}) = \delta_{ij}$ . This means that the  $i$ th basis function  $\varphi_i$  is equal to 1 at precisely  $x_{(i)}$ , and it decreases to zero linearly along all the remaining faces, until it goes to 0 and stays there over the rest of the mesh (see Figure 2.1 for an example in one dimension). This simply means that basis functions are supported in a very limited subset of the mesh surrounding the vertex. In particular, they are compactly supported, and enjoy all the analytic advantages of such functions

(they are essentially the continuous piecewise linear analogues of characteristic functions, which are discontinuous). The fact they are piecewise linear and continuous means they are suitable to use as test functions in the weak formulation of second order equations (and in fact, their nondifferentiability makes the weak formulation essential).

Given such a basis on  $U$ , a function  $u : U \rightarrow \mathbb{R}$  has a PIECEWISE LINEAR AND CONTINUOUS APPROXIMATION or LINEAR INTERPOLATION relative to this basis, simply by evaluating at the points:

$$u_h(x) := \sum_{i=1}^N u(x_{(i)})\varphi_i(x).$$

It is customary to denote  $u(x_{(i)})$  by  $u_i$ . The collection of components  $(u_i)$  gives us a vector,  $\mathbf{u} = \sum_{i=1}^n u_i \mathbf{e}_i$ . We will actually depart slightly from the traditional notation and, being the geometers we are, write  $u^i$  with the  $i$  in the superscript position, so that  $u_h = u^i \varphi_i$  using the Einstein summation convention. Also, we often write  $\mathbf{u}\Phi = u^i \varphi_i$  (the notation commonly used in the theory of moving frames). The vector  $\mathbf{u}$  contains all the information of the piecewise linear discretization—recasting things in terms of their approximations using the basis is how we pass from the intractable infinite-dimensional things down to the finite-dimensional things that we can work with. As we saw in the above discussion about the Gal rkin methods, linear operators on function spaces such as the Laplacian also have their finite-dimensional, discretized versions—for example, linear, 2nd order operators are represented by the stiffness matrix.

Basis functions consisting of higher-order polynomials are also possible, and give more accurate results, although it takes considerably more work to deal with them, so we will leave the discussion of these elsewhere. Piecewise linear elements have

piecewise constant gradients and make the implementation considerably simpler to deal with, especially when dealing with numerical integration (quadratures)—we need to sample only one point per element—the barycenter (when each element is assumed to have uniform density—FEM can handle many problems including elasticity with variable densities, where the barycenter may be different from the usual geometric one).

### 2.2.2 Shape Functions

There are two ways of conceiving of the basis functions—first, as functions defined over the whole domain (extended by zero), with a value of 1 at its corresponding vertex (see Figure 2.1). On the other hand, if we look at a single element, there are  $n + 1$  vertices that are associated to it, so we often have to look at the part of each basis function that goes through the element. Each such restricted basis function is called a SHAPE FUNCTION. It is really the shape functions that are used to compute the stiffness matrices, as we have to integrate over the whole domain, and approximating the integral by a weighted sum over each individual element is a good start.

In practice, we really worry about the shape functions of only one true element, the MASTER ELEMENT, which is the unit simplex in  $\mathbb{R}^n$ . In the plane, it is the standard unit right triangle (diagonal half of a unit square) and similarly the unit tetrahedron in  $\mathbb{R}^3$  (long diagonal half of a cube). The rest may be derived by linear coordinate transformations (any simplex may be taken to any other by a linear transformation). The shape functions of the unit simplex in  $\mathbb{R}^n$  are totally determined by their values on each vertex. We number the vertices in the element by  $\mathbf{x}_{(i)}$ , in orientation-determining order, e.g. counterclockwise in  $\mathbb{R}^2$  and right-handed in  $\mathbb{R}^3$ , start our numbering at 0, and fix  $\mathbf{x}_{(0)}$  to be the origin. With this in mind, we can write down the shape functions

explicitly as:

$$\tilde{\varphi}_i(x^1, x^2, \dots, x^n) = x^i$$

for  $1 \leq i \leq n$ , and

$$\tilde{\varphi}_0(x^1, x^2, \dots, x^n) = 1 - \sum_i^n x^i.$$

For example, the shape functions for the triangular element in  $\mathbb{R}^2$  are

$$\tilde{\varphi}_0(x, y) = 1 - x - y$$

$$\tilde{\varphi}_1(x, y) = x$$

$$\tilde{\varphi}_2(x, y) = y$$

where the vertices are  $\mathbf{x}_{(0)} = (0, 0)$ ,  $\mathbf{x}_{(1)} = (1, 0)$ , and  $\mathbf{x}_{(2)} = (0, 1)$ , and similarly,

$$\tilde{\varphi}_0(x, y, z) = 1 - x - y - z$$

$$\tilde{\varphi}_1(x, y, z) = x$$

$$\tilde{\varphi}_2(x, y, z) = y$$

$$\tilde{\varphi}_3(x, y, z) = z$$

for the unit tetrahedron in  $\mathbb{R}^3$ .

How does one go from this to the general example? We use affine-linear transformations. Let  $\mathbf{x}_{(i)}$  now describe any simplex in  $\mathbb{R}^n$ , not necessarily the master element (now let  $\mathbf{y}_{(i)}$  denote the vertices of the master element instead). The numbering should still be in orientation-determining order, but otherwise arbitrary. As such, the functions we derive will of course be dependent on such a choice, but this is not a problem: it is equivalent to choice of parametrization, and hence it follows all the usual rules of dealing with coordinate transformations, and the usual expressions



in the right combinations are coordinate-invariant. The mapping of the simplex is simply determined by the difference vectors  $\mathbf{v}_{(i)} = \mathbf{x}_{(i)} - \mathbf{x}_{(0)}$  for  $1 \leq i \leq n$ ; we form a matrix  $A$  by placing them side-by-side as column vectors:  $A = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ . This works, because the unit vectors in  $\mathbb{R}^n$  are in fact the edges of the unit tetrahedron. Finally, of course, we have to add on  $\mathbf{x}_{(0)}$  to complete the transformation:

$$T(\mathbf{y}) = A\mathbf{y} + \mathbf{x}_{(0)}.$$

This transformation sends the unit tetrahedron to our simplex, with the origin mapping to  $\mathbf{x}_{(0)}$  and similarly,  $T(\mathbf{y}_{(i)}) = \mathbf{x}_{(i)}$ . Note that  $A$  is the derivative of  $T$ , and  $T^{-1}$  is also an affine transformation:

$$T^{-1}(\mathbf{x}) = B(\mathbf{x} - \mathbf{x}_{(0)}).$$

where  $B = A^{-1}$ .

What does this mean for computing shape functions? Given the shape function  $\varphi_i$  on the standard unit simplex, the shape function for the corresponding vertex is  $\varphi_i \circ T^{-1}$ , because what we need to do is take a point in the simplex, map it back to the unit triangle, and use the standard shape function defined there. What this also means is that their *gradients* transform inversely:  $\nabla(\varphi_i \circ T^{-1}) = \nabla\varphi_i A^{-1} = \nabla\varphi_i B$ . Really, we are using the 1-form, writing  $\nabla\varphi_i$  as the row matrix  $d\varphi_i$ , and thus we need to multiply by the derivative on the right—it matters, because  $A$  may not be an orthogonal transformation, and—if we do insist on working with gradient vectors, we would have to take into account the changed metric coefficients. It is easier to deal with 1-forms directly—we will see even more clearly that we need this viewpoint when we work on curved surfaces. Usually, the 1-form is more useful, and the vector is given only as an aid to those who have only had vector calculus.

### 2.2.3 Computation of the Stiffness Matrix

With all of this in mind, we now look at what this means for the computation of the stiffness matrix. We perform the integration by integrating over all the triangles and summing the results. Within each triangle, the integral is then easy to perform: one transforms the requisite gradients and includes the Jacobian of the transformation in the integral—writing  $U$  for the master element,  $T$  for the transformation defined in the above,  $T(U)$  is the triangle, and

$$\begin{aligned} U\text{'s contribution to } K_{k\ell} &= \int_{T(U)} a^{ij}(\mathbf{x}) \partial_i \varphi_k(\mathbf{x}) \partial_j \varphi_\ell(\mathbf{x}) + c(\mathbf{x}) \varphi_k(\mathbf{x}) \varphi_\ell(\mathbf{x}) d\mathbf{x} \\ &= \int_U \left( a^{ij}(T(\mathbf{y})) B_i^r \partial_r \tilde{\varphi}_k(\mathbf{y}) B_j^s \partial_s \tilde{\varphi}_\ell(\mathbf{y}) + c(T(\mathbf{y})) \tilde{\varphi}_k(\mathbf{y}) \tilde{\varphi}_\ell(\mathbf{y}) \right) |\det(A)| d\mathbf{y} \end{aligned}$$

where  $B$  is the matrix defined above (the inverse of  $A$ ,  $T$  without the extra  $+\mathbf{x}_{(0)}$ ). This looks messy, but in actual practice, it really is simple, especially in the piecewise linear case, since the computation of  $B\nabla\varphi_j$  is usually trivial, as  $\nabla\tilde{\varphi}_j$  is just a constant (1 or 0), and the quadrature only needs one integration point per simplex—the value at the barycenter. Finally, even more simplifying, since the  $\varphi_k$  are supported only within the directly neighboring simplices of the vertex, the integral is only nonzero for both  $k$  and  $\ell$  equal to the indices corresponding to the vertices of that single element. So, for example, if a triangular element is defined by  $\mathbf{x}_{(1)}$ ,  $\mathbf{x}_{(4)}$ , and  $\mathbf{x}_{(5)}$ , then the terms with  $\nabla\varphi_4 \cdot \nabla\varphi_1$  and  $\nabla\varphi_1 \cdot \nabla\varphi_5$  (corresponding to  $K_{41}$  and  $K_{15}$  in the stiffness matrix) might be nonzero, but terms containing  $\nabla\varphi_3 \cdot \nabla\varphi_2$  and  $\nabla\varphi_3 \cdot \nabla\varphi_1$  are definitely zero, since  $\varphi_3$  isn't supported in this element. This means the stiffness matrix is generally quite SPARSE, that is, has mostly zero elements, even in higher dimensions.

## 2.3 Adding Time Dependence

So far, we have considered only steady, i.e., purely time-independent problems. FEM is typically a method to discretize space. However, it is not difficult to allow handling of time as well. It is here that the viewpoint of evolutionary PDES as ODES in infinite-dimensional spaces shines. We consider a long thin elastic solid (a beam) with density  $\rho$  and elasticity  $E$  on an interval of length  $L$ . The PDE for longitudinal vibrations  $u$  in the beam is

$$\rho(x) \frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x} \left( E(x) \frac{\partial u}{\partial x} \right).$$

Now, if our domain does not vary with time, it is reasonable to assume that the the element (tent) functions  $\varphi_j$  do not vary in time. So if our discrete solution  $u^j \varphi_j$  is to approximate a continuous function of two variables,  $u(x, t)$ , it makes sense to have only the  $u^j$  vary in time. This is simply SEPARATION OF VARIABLES, and we connect it to the usual presentation of the technique in introductory books on PDES by offering another interpretation of what FEM is. So let us substitute  $u^j(t) \varphi_j(x)$  for  $u(x, t)$  in the equation:

$$\rho(x) \varphi_j(x) \frac{d^2 u^j(t)}{dt^2} = \frac{d}{dx} \left( E(x) u^j(t) \frac{d\varphi_j(x)}{dx} \right) = \frac{d}{dx} \left( E(x) \frac{d\varphi_j(x)}{dx} \right) u^j(t).$$

Now actually, this is only true in the distributional sense, because the  $E(x) \frac{d\varphi_j}{dx}$  are discontinuous. How do we deal with distributions? The usual way: integration. We integrate both sides against  $\varphi_k(x)$  and note that the integral is all in the space variables,

so we may pull out the coefficients  $u^j(t)$ , and do the usual integration by parts:

$$\begin{aligned} \left( \int_0^L \rho(x) \varphi_j(x) \varphi_k(x) dx \right) \frac{d^2 u^j(t)}{dt^2} &= \left( \int_0^L \frac{d}{dx} \left( E(x) \frac{d\varphi_j(x)}{dx} \right) \varphi_k(x) dx \right) u^j(t) \\ &= - \left( \int_0^L E(x) \frac{d\varphi_j(x)}{dx} \frac{d\varphi_k(x)}{dx} dx \right) u^j(t). \end{aligned}$$

The integral on the RHS is just the stiffness matrix  $K_{jk}$  as we argued previously, making the RHS  $K_{jk} u^j(t)$ . If we define  $M_{jk}$  to be the integral on the LHS, we have

$$M_{jk} \frac{d^2 u^j}{dt^2} = -K_{jk} u^j.$$

Finally, recalling the definition of matrix multiplication and writing  $\mathbf{u}$  for  $u^j$ , we have

$$M \frac{d^2 \mathbf{u}}{dt^2} = M \ddot{\mathbf{u}} = -K \mathbf{u}.$$

which is almost the equation of a mass-spring system. We say “almost,” because for  $\rho(x)$  with sufficiently large support, the matrix  $M$  is not diagonal (it is at most band tridiagonal in one dimension) because the functions  $\varphi_j(x)$  successively do have some overlap. We recover the spring-mass system by simply making  $\rho$  be the sum of point masses (in the sense of distributions), i.e.  $\rho(x) = \sum_i m_i \delta(x - x_i)$ , which is what a spring-mass system models anyway (the springs are usually regarded as “massless” in these simple models). In fact here we see exactly how we can take care of springs that have mass after all.

Now given all that, how do we actually solve it? We now have a 2nd-order (system of) *ordinary* differential equation(s),

$$\ddot{\mathbf{u}} = -M^{-1} K \mathbf{u}.$$

Actually it is not obvious that  $M$  is invertible, but it usually is, and is in fact symmetric and positive-definite. Also,  $K$  is usually positive-semidefinite. However,  $M^{-1}K$  is usually *not* symmetric, that is, self-adjoint with respect to the usual inner product. It is self-adjoint with respect to the inner product induced by  $M$ :  $((v, w)) = v^T M w$ , and as such, has a complete  $M$ -orthonormal basis of eigenvectors with corresponding eigenvalues that are real (that an operator is self-adjoint with respect to *any* metric at all guarantees real eigenvalues and that the matrix is non-defective; the dependence on metric only shows up in the orthonormality of the basis). It is instructive to note how the time-dependent problem is distinct from the elliptic problem, where the task is to invert  $K$  (in our prototypical elliptic problem, what would be the mass matrix is set to unity). Here,  $K$  does not have to be inverted, but rather, *exponentiated* in some manner, because that is how one solves linear differential equations. Thus this shows that the solution is unique (provided we give initial conditions for  $\mathbf{u}$  and  $\mathbf{u}_t$ ), and standard theory of dynamical systems [47] shows that the solution exists for all time, and the equilibrium point is a *center*.

Having said that, for the actual numerical method, it is better to keep the  $M$  on the LHS, for sparseness considerations. We can write it in block form as a system, defining  $\mathbf{v} = \dot{\mathbf{u}}$ :

$$(2.3.1) \quad \begin{pmatrix} I & 0 \\ 0 & M \end{pmatrix} \begin{pmatrix} \dot{\mathbf{u}} \\ \dot{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} 0 & I \\ -K & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}.$$

## 2.4 Numerical Methods for Evolutionary Equations

We have seen in principle how to compute the solution to our fundamental, evolutionary PDES. In fact, the only thing that has been discretized is space; our solutions above completely decoupled the time evolution from our spatial operators

and recast the problem into a (continuous time system of) ODE(s). We saw that for our canonical examples with linear differential operators, the solution is more or less explicitly known as the exponential or sines and cosines of matrices (generalizing rotation). However it is instructive to examine approximation in time (called TIMESTEP-PING) as well, since more complicated equations may not be solvable in terms of nice functions we know, and even in the linear case, computation of things like exponentials of very large matrices can be prohibitively expensive in both computing space and time. Fortunately, it turns out that it is very easy to see, at least conceptually, how to approximate time evolution. We follow [55] and [87, Ch. 11] for these fundamentals; this is obviously a much larger field, and we barely scratch the surface here.

### 2.4.1 Euler Methods

As a first stab, we try finite-differencing: pick a small  $\Delta t$  and approximate  $\frac{du_j}{dt}$  by a difference quotient:

$$\frac{du_j}{dt} = \frac{u_j(t + \Delta t) - u_j(t)}{\Delta t} = \frac{u_j^{k+1} - u_j^k}{\Delta t},$$

where it is traditional to write  $u_j^k$  for the value of  $u_j$  at the  $k$ th time step (thus we shall temporarily revert to using subscripts for vector components). Since this must hold for all  $j$ , we can actually use a vector difference quotient  $\frac{1}{\Delta t}(\mathbf{u}^{k+1} - \mathbf{u}^k)$ .

Setting this equal to the RHS of the space-discretized diffusion equation, we have

$$\frac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\Delta t} = (-M^{-1}K)\mathbf{u}^k$$

or, explicitly solving for  $\mathbf{u}^{k+1}$ ,

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \Delta t M^{-1} K \mathbf{u}^k = (I - \Delta t M^{-1} K) \mathbf{u}^k.$$

This is delightfully simple: to get our value at the next time step, simply apply the operator  $I - \Delta t M^{-1} K$  to our current time step. Since the equation is linear, this is just iterating the same map over and over again. It gives us a recipe for directly evolving our initial data forward in time. The intuition here is simple, too: imagining  $\mathbf{u} \mapsto -M^{-1} K \mathbf{u}$  as a vector field in some high-dimensional space, its value at  $\mathbf{u}^k$  determines a tangent vector (direction); one advances by  $\Delta t$  times this tangent vector to get to the next step along the integral curve. The error introduced here is due to moving along a (small) straight line segment instead of the (unknown) true curve that connects the points. Of course, if we keep things small, the approximation is not off by much. By analysis via Taylor series, it is easily shown that the error is proportional to the square of the timestep (it is a FIRST-ORDER METHOD). This simple method is called the EULER METHOD, and as one can guess by the name, dates back to the time of Euler.

**2.4.1 Instability and the Implicit Euler Method.** However, simplicity has its price: this method is very unstable if the timestep is too big. This is not simply a large approximation error that normally arises from discretization—but rather, catastrophic failures, such as the approximate solution going to infinity, when there is nothing of the sort in the true solution. In addition, it interacts badly with the spatial discretization: the size of the timestep required for stability is proportional to the *square* of the size of spatial discretization, so for even reasonably fine mesh sizes, say on the order of  $10^{-3}$ , we will need timesteps on the order of  $10^{-6}$ , which is prohibitively small for lengthy simulations, even on fast computers. Even if the available time and computational power is manageable, it is still better to figure out a way how to use computing resources more efficiently. A correct idea that fixes the stability problem, which is almost as simple, is to use the *future* timestep for evaluating discretized spatial operator:

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \Delta t M^{-1} K \mathbf{u}^{k+1}.$$

This simple modification presents its own conundrum however: how do we know the future timestep if that's what we're trying to compute in the first place? Here, the solution for linear equations is simple; we simply bring it to the other side:

$$\mathbf{u}^{k+1} + \Delta t M^{-1} K \mathbf{u}^{k+1} = (I + \Delta t M^{-1} K) \mathbf{u}^{k+1} = \mathbf{u}^k.$$

Thus, solving, we have

$$\mathbf{u}^{k+1} = (I + \Delta t M^{-1} K)^{-1} \mathbf{u}^k.$$

which we call the BACKWARD or IMPLICIT EULER METHOD. Comparing the two, we have that the usual Euler method iterates the map  $I - \Delta t M^{-1} K$  whereas the backward method iterates  $(I + \Delta t M^{-1} K)^{-1}$ , which, when using a small timestep, we can see is close to  $I - \Delta t M^{-1} K$  because of the geometric series.

In order to deal with nonhomogeneous terms, rederiving the equations with  $\Delta u + f$  instead of  $\Delta u$  gives an extra term  $\mathbf{f}$  on the RHS when integrating against  $\varphi_k$ .  $\mathbf{f}$  is the vector of coefficients  $\int f \varphi_k$ . We get

$$\dot{\mathbf{u}} = -M^{-1} K \mathbf{u} + M^{-1} \mathbf{f},$$

and discretizing in time, we have

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \Delta t M^{-1} K \mathbf{u}^{k+1} + \Delta t M^{-1} \mathbf{f},$$

and thus solving by Backward Euler,

$$\mathbf{u}^{k+1} = (I + \Delta t M^{-1} K)^{-1} (\mathbf{u}^k + \Delta t M^{-1} \mathbf{f}).$$

Thus it is almost as simple, in that now we iterate an affine map (inverting the operator



$I + \Delta t M^{-1} K$  as well as adding a term at each step) instead of a purely linear one. What happens is that at each step, we essentially “start off” with additional term  $\mathbf{f}$  as time evolves (this is a special case of a very general principle for evolutionary equations with inhomogeneous terms, known as DUHAMEL’S PRINCIPLE).

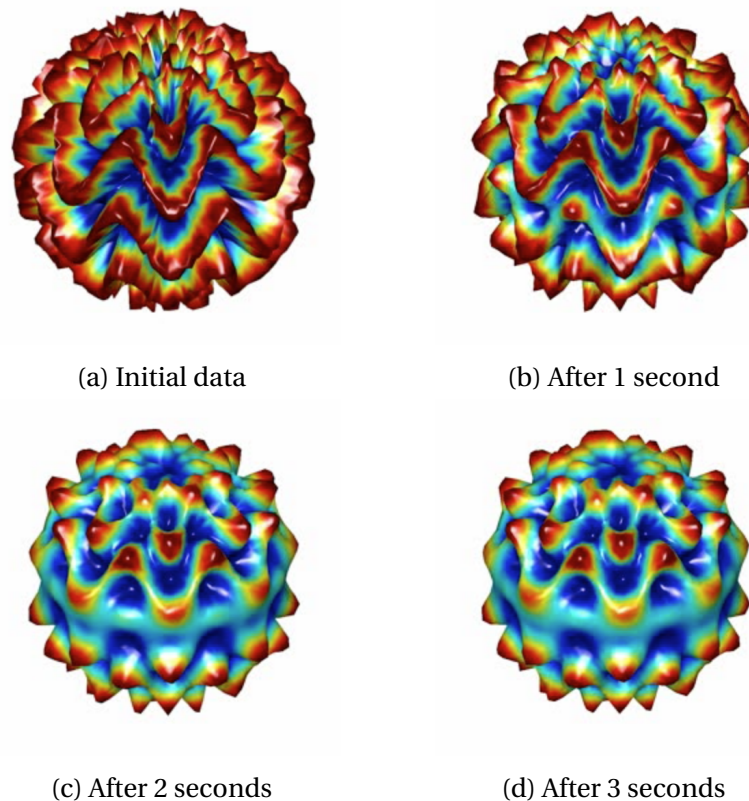
**2.4.2 Notes on actual implementation.** It should be noted that for actual implementation with linear-algebra solvers, it is better to write  $I + \Delta t M^{-1} K = M^{-1}(M + \Delta t K)$ , so that

$$\mathbf{u}^{k+1} = (M + \Delta t K)^{-1}(M\mathbf{u}^k + \Delta t\mathbf{f}),$$

or in the notation of Matlab,

$$\mathbf{u}^{k+1} = (M + \Delta t K) \setminus (M\mathbf{u}^k + \Delta t\mathbf{f}).$$

The reason why this is desirable is that the matrices  $M$ ,  $K$ , and  $M + \Delta t K$  are usually quite sparse, while  $I + \Delta t M^{-1} K$  may not be (the general rule is that the inverse of a sparse matrix need not be sparse, so anything that involves explicit evaluation of the inverse will lose its sparsity). Solving a system of equations with sparse matrices is much more efficient than with a full matrix, and the associative order can make a big difference. See Figure 2.2 and the supplemental files `heat-demo-basic.mov` and `heat-on-sphere.mpg` for examples for the heat equation in a square and on the sphere (the latter using surface finite element methods, as we shall detail in Chapter 4), which use exactly this timestepping scheme. For more general, nonlinear spatial operators, one may get a more complicated, nonlinear implicit equation for  $\mathbf{u}^{k+1}$  which is not nearly as easy to solve as just using the Euler method. One needs to use root-finding algorithms such as Newton’s method to solve for  $\mathbf{u}^{k+1}$ . However, such extra steps are usually an improvement over having to calculate a thousand times more timesteps just to get a solution worth visualizing.



**Figure 2.2:** The heat equation on a piecewise linear approximation of a sphere (3545 triangles). The solution is graphed in the normal direction of the sphere. The spatial discretization uses a surface finite element method detailed in Chapter 4 (based on [28]), and implemented using a modification of FETK [31], and the timestepping scheme is backward Euler. The supplemental file `heat-on-sphere.mpg` shows this as an animation at 60 frames per second.

## 2.4.2 Other Methods

The subject of approximation by ODEs is a vast subject in itself, so we do not treat them in great detail here. Our goal is to prove some general results on evolution equations, so we will not have a need to discuss specific choices of ODE methods in great detail. Nevertheless, we should mention some other methods to give an idea of how these concepts are used together.

**2.4.3 Runge-Kutta Methods.** For higher-order methods, Runge-Kutta methods are a popular choice ([55, Ch. 3], [87, §11.8]). The basic idea is to use some intermediate

stages in the computation of each timestep, which helps refine the estimate. It can also be viewed in terms of numerical integration, since the Fundamental Theorem of Calculus allows us to compute the next timestep exactly in terms of the current one by integrating the solution in between. This is exactly the basis of the error analysis, and what yields the higher order results. The tricky issue is that, unlike explicit integration of a function given in advance, the unknown function must be evaluated at some of the interior points, so we get, in general, implicit equations. For linear ODEs, of course, this does not pose such a problem—much like the backward Euler method, it is a matter of moving factors and their inverses around (although again, if we want to exploit sparse matrix structure, we have to be careful about how we write the equations). For nonlinear equations, this generally requires us to use root-finding methods (although once again, it is also something encountered in the backward Euler method). Finally, stability is of course an important issue (as it always is in numerical analysis).

**2.4.4 Symplectic Methods.** For differential equations with a certain special structure commonly encountered in mathematics and physics, namely Lagrangian and Hamiltonian equations of motion ([41, 1], [62, Ch. 22]), there are certain qualitative properties of solutions that we would like to see preserved (but usually are not under the previous approximation schemes). These equations arise naturally in the discretization of the wave equation (not surprisingly, because the derivation of the wave equation is based in Newtonian, and hence Lagrangian and Hamiltonian mechanics). The key property of these systems is that they conserve energy, and this has important physical implications which are not directly taken into consideration in the preceding algorithms. These methods are discussed at length in [63, 44]. We do give one simple example, namely, the symplectic Euler method. In some sense, it combines the approach of the two previous Euler methods for Hamilton's equations. One simply uses the forward method for the position variable and backward method for the momentum variable (or

vice versa). Heuristically, it is because the forward method tends to cause expansion in the phase space (which is related to its instability), while the backward method causes contraction. Thus the method is a sort of “goldilocks” compromise. Of course, that it seemingly so simply ends up combining the two is actually manifestation of something deeper.

**2.4.5 Example** (Symplectic Euler Method). To write it out in equations, in some Hamiltonian system we have some position variables  $q$ , momentum variables  $p$ , and a conserved energy, the Hamiltonian  $H$ . Thus Hamilton’s Equations are

$$\begin{aligned}\dot{q} &= \frac{\partial H}{\partial p} \\ \dot{p} &= -\frac{\partial H}{\partial q}.\end{aligned}$$

For a simple concrete example, for a mass on a spring (harmonic oscillator), with  $q$  being displacement from equilibrium, we have  $H(q, p) = \frac{1}{2}kq^2 + \frac{p^2}{2m}$ , which leads to the equations

$$\begin{aligned}\dot{q} &= \frac{p}{m} \\ \dot{p} &= -kq.\end{aligned}$$

Then using the same standard discretization procedures by rewriting  $\dot{q}$  and  $\dot{p}$  as a difference quotient, where the sequence of timesteps is denoted  $q^j$  and  $p^j$ . To evaluate the vector field side (RHS) of the equation, we use  $q^{j+1}$ , the future timestep for  $q$ , but  $p^j$ , the current timestep for  $p$  (forward Euler would insist on using  $j$  for both of them,

and backward Euler would always use  $j + 1$ ):

$$\begin{aligned}\frac{q^{j+1} - q^j}{\Delta t} &= \frac{1}{m} p^j \\ \frac{p^{j+1} - p^j}{\Delta t} &= -k q^{j+1},\end{aligned}$$

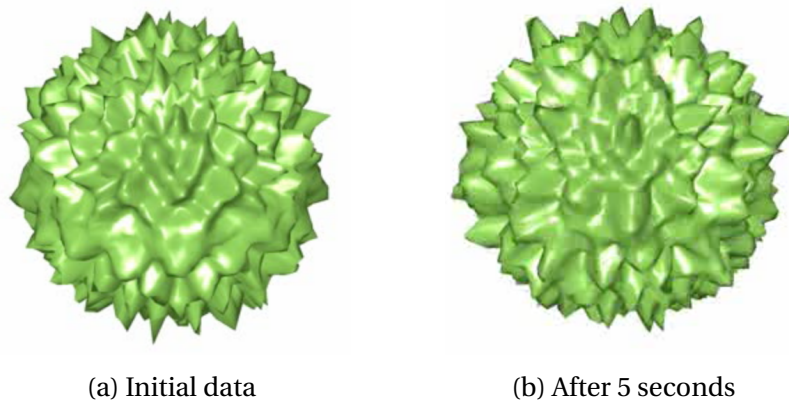
or, solving for the  $(j + 1)$ th timestep:

$$\begin{aligned}q^{j+1} &= q^j + \frac{\Delta t}{m} p^j \\ p^{j+1} &= p^j - k \Delta t q^{j+1}.\end{aligned}$$

This is an explicit algorithm, since the  $q^{j+1}$  already is expressed solely in terms of the variables at timestep  $j$ , so its calculation for  $p^{j+1}$  is already expressed in terms of known quantities. For the wave equation, we can write down the semidiscretized equation (2.3.1) (but here  $\mathbf{q} = \mathbf{u}$  and  $\mathbf{p} = M\mathbf{v}$ ). For  $H$ , instead of  $\frac{1}{2}kq^2$ , we have instead some quadratic form  $\frac{1}{2}\mathbf{q}^T K \mathbf{q}$ , and similarly,  $\frac{1}{2}\mathbf{p}^T M^{-1} \mathbf{p}$ , where  $K$  and  $M$  are resp. the stiffness and mass matrices. The customary warnings for exploitation of sparse matrix structure apply. See Figure 2.3 and the supplemental file `waves-on-sphere.mpg`.

## 2.5 Error Estimates for the Finite Element Method

We have mentioned that finite element methods give a very good framework for error analysis. Here, we list some main results and prove a couple of them to get a sense of how the analysis works. It will be important to establish these results so we can translate results about best approximation theorems (a natural consequence of Hilbert space theory) into concrete estimates based on mesh size. It leverages the use of modern Sobolev space methods [2, 30, 39]. Generally speaking, of course, we want



**Figure 2.3:** The wave equation on a piecewise linear approximation of a sphere (3545 triangles). The solution is graphed in the normal direction of the sphere. The spatial discretization uses a surface finite element method detailed in Chapter 4 (based on [28]), and implemented using a modification of FETK [31], and the timestepping scheme symplectic Euler. The supplemental file `waves-on-sphere.mpg` shows this as an animation at 60 frames per second.

our approximations  $u_h$  to converge to the true solution. The basic method, detailed in [11, §§II.6-7] and [13, Chs. 2-4], is, after choosing some finite element spaces  $V_h \subseteq V$  (with  $h$  a parameter accumulating to 0, which usually represents the size of elements in an approximating mesh), to define some type of linear operator  $I_h : V \rightarrow V_h$ , which represents some kind of approximation (called an INTERPOLATION operator). For example, if we choose  $V_h$  to be continuous piecewise polynomial functions of degree up to some  $r$ , then given enough interpolation points  $\{z_i\}$  in each simplex (the number of such points required is dependent on both the dimension and the degree of the polynomial), any continuous function  $u$  can be approximated by a unique polynomial  $I_h u$  whose values at  $z_i$  coincide with the value of  $u$ . Then the basic error of interpolation is

$$\|u - I_h u\|_\alpha$$

where  $\|\cdot\|_\alpha$  is some norm (usually one of the Sobolev norms). The kind of result we wish to establish is something of the form

$$(2.5.1) \quad \|u - I_h u\|_\alpha \leq C(\alpha, n, r) h^\beta \|u\|_\gamma,$$

that is, the interpolation error measured in some norm is dependent on some constant depending on geometric properties, the dimension, the degree, and so forth (but not on  $h$ ), then some power of the mesh size  $h$ , and then finally, the (possibly different) norm of the true solution  $u$ . In particular, we find that as  $h \rightarrow 0$ , the interpolations actually converge to the true function in this norm, at some rate  $\beta$ .

The key fact here is that, with an inner product  $\langle \cdot, \cdot \rangle_\alpha$ , the orthogonal projection  $P_{V_h}$  gives the BEST APPROXIMATION in the induced norm  $\|\cdot\|_\alpha$  ([48, §8.2, Theorem 4, Finite-dimensional case], [34, Theorem 5.24], [64, Lemma 2.8]), which is one of the reasons why we like Hilbert spaces and orthogonality:

$$(2.5.2) \quad \|u - P_{V_h} u\|_\alpha = \inf_{v_h \in V_h} \|u - v_h\|_\alpha \leq \|u - I_h u\|_\alpha.$$

**2.5.1 What this means for finite element methods: Céa's Lemma.** What does this mean for the error in finite element methods? Suppose we now have that  $u$  is the solution to some elliptic problem  $Lu = f$ , and using the Galërkin method (defining the bilinear form  $a(u, v) = \langle Lu, v \rangle$ ), we compute some approximation  $u_h \in V_h$  such that

$$a(u_h, v) = \langle f, v \rangle$$

for all  $v \in V_h$ . If we separately establish that the solution satisfies a QUASI-BEST AP-

PROXIMATION with respect to the  $\alpha$  norm, i.e.,

$$(2.5.3) \quad \|u - u_h\|_\alpha \leq C \inf_{v_h \in V_h} \|u - v_h\|_\alpha,$$

then, coupled with the estimates (2.5.2) and (2.5.1), we have

$$(2.5.4) \quad \|u - u_h\|_\alpha \leq C \|u - I_h u\|_\alpha \leq C'(\alpha, n, r) h^\beta \|u\|_\gamma.$$

In particular, if such an estimate as the above holds, this means the approximations  $u_h$  converge to the true solution  $u$  at rate  $\beta$ . In fact, such an estimate does hold:

**2.5.2 Theorem** (Céa's Lemma, [11], Theorem 4.2). *Let  $V$  be Hilbert space with an inner product  $\langle \cdot, \cdot \rangle_V$ , a bounded, coercive bilinear form, and  $\ell \in V'$  a bounded linear functional. Suppose that  $u$  is a solution to the weak form of the problem  $a(u, v) = \ell(v)$ , for all  $v \in V$ . Let  $V_h \subseteq V$  be some approximating spaces, and  $u_h$  be the Galërkin solution to the problem, namely,  $a(u_h, v_h) = \ell(v_h)$  for all  $v_h \in V_h$ . Then (taking  $\|\cdot\|_\alpha$  to be the norm  $\|\cdot\|_V$ ) we have*

$$\|u - u_h\|_V \leq M \gamma^{-1} \inf_{v_h \in V_h} \|u - v_h\|_V,$$

where  $M$  is the bound on  $a$  and  $\gamma$  is the coercivity constant, satisfying  $a(w, w) \geq \gamma \|w\|_V^2$ .

In applications,  $V$  is usually some Sobolev space, e.g.  $H^s(U)$  for bounded domains  $U \subseteq \mathbb{R}^n$ .

*Proof.* This is a good illustration of the use of orthogonality in different inner products. For all  $v \in V_h$ , we have that  $a(u, v_h) = \ell(v_h) = a(u_h, v_h)$ . The first equality is because  $u$  satisfies it for all  $v \in V$ , in particular,  $v_h \in V_h \subseteq V$ , and the second equality follows by



definition of the Galërkin solution and only holds for  $v_h \in V_h$ . So therefore

$$(2.5.5) \quad a(u - u_h, v_h) = 0$$

for all  $v_h \in V$ . Since  $u_h \in V_h$  also,  $a(u - u_h, u_h - v_h) = 0$ , also. By coercivity,

$$\begin{aligned} \gamma \|u - u_h\|_V^2 &\leq a(u - u_h, u - u_h) = a(u - u_h, u_h - v_h) + a(u - u_h, v_h - u) \\ &= a(u - u_h, v_h - u) \leq M \|u - u_h\|_V \|v_h - u\|_V. \end{aligned}$$

Canceling one factor of  $\|u - u_h\|_V$ , and noting that  $v_h \in V_h$  was arbitrary, gives the result.  $\square$

The trick of using coercivity or similar properties for bounds below, to cancel one factor in a bound above, is something we see over and over again. Also, note that the quantity  $a(u - u_h, u - u_h)$  is simply the (square of the) energy norm, and since coercivity makes the energy norm equivalent to the  $V$ -norm, we often refer to this as an ENERGY-NORM ESTIMATE. The above proof—specifically, (2.5.5)—also establishes that the solution  $u_h$  is in fact the best approximation to  $u$  in  $V$  *relative to the energy inner product*. Céa's lemma, therefore, relates this to the  $V$ -norm. and gives a specific bound on the constant.

To get good bounds, therefore, we need to formulate good interpolation operators  $I_h$ ; there are several different kinds for different purposes. Suppose that  $r$  is large enough such that the Sobolev space  $H^{r+1}(U)$  is in  $C^0$  [30, §5.6.3, Theorem 6], [11, §I.3]. If  $0 \leq s \leq r + 1$ , then [11, §§II.5-6], for continuous piecewise  $r$ th degree polynomials  $S_h$ , the polynomial interpolation operator  $I_h$  maps  $H^{r+1}$  boundedly to  $S_h$ , and

$$\|u - I_h u\|_{H^s(U)} \leq ch^{r+1-s} |u|_{r+1}$$

where  $|\cdot|_{r+1}$  is the seminorm, the  $\mathcal{L}^2$  norm of the vector of all  $(r+1)$ th derivatives of  $u$ . In other words, the power of  $h$  (namely,  $\beta$  in the generic estimate (2.5.1) above) is one more than the degree of the polynomials considered (we would expect that the higher the degree of the polynomial, the better the approximation rate), minus how refined a norm we choose (we would expect that if we demand an estimate that includes more derivatives, the worse the convergence rate). For some special cases, for example, if  $u$  is a polynomial of degree  $r$ , then the interpolation error is zero (in particular, our interpolations are idempotent), and if we choose the  $\mathcal{L}^2$  norm, then the convergence rate is indeed one more than the degree of the polynomials used.

**2.5.3 Why we need more general interpolation operators.** Other interpolation operators are possible, and in fact, necessary, because the above polynomial interpolation operators are limited to continuous functions, and we often want to prove estimates in Sobolev spaces which cannot be embedded into some Hölder space. The Clément interpolation is one common solution. We do not describe it here (although we give a few words about it for interpolating differential forms).

**2.5.4 *A priori* estimates: linking up to the general PDE theory.** In order to derive true *a priori* estimates for the error, that is, without knowing what the solution is, we need to be able to estimate that term  $\|u\|_\gamma$  in terms of known quantities, such as the data,  $f$ . This, of course, is done using the usual elliptic theory, described in Chapter 1. This depends on features of the domain, such as regularity and convexity. An important fact is that we cannot always approximate our domains via simplices, if we want good approximation results; which can be bad for domains with curved boundaries. However, we can make progress in this area via *variational crimes* [11, §III.1-2]: we approximate the domain itself with simplices (which give piecewise smooth, Lipschitz boundaries, satisfying the uniform cone condition), and see what the error is between the boundary and its approximation. There is much more to say about that later on.

## 2.6 Discretization of Differential Forms

Having stressed the importance of differential forms and exterior calculus, we should see how to compute with them numerically. Much of the existing methods of computation are still done simply using vector calculus methods. A first try at doing vector methods is simply solving for the component functions using the methods detailed earlier in this chapter. This sometimes works, especially for simple cases. Sometimes, however, these methods fail catastrophically: they become unstable, or they fail to converge, and it is difficult to pinpoint why. Aside from this obvious practical problem, there is a philosophical problem as well: recall, regarding vectors as mere lists of functions is not really capturing their geometric nature, and we have striven to avoid that kind of thinking throughout this whole work.

Of course, eventually some methods that do in fact, take that nature into account have been discovered, due to the importance of vector fields in fluid mechanics and electromagnetics [88, 72, 73, 8]. But these are really a part of a greater whole: the general theory of finite element methods for differential forms places many of these seemingly disparate concepts into a coherent framework and clarifies understanding of where certain conditions and restrictions come from (just as differential forms have similarly elucidated previous concepts studied in this work, such as Sobolev spaces, traces, and boundary value problems), and give us clues about how to analyze errors in approximations and improve our algorithms. This viewpoint was introduced by Arnold, Falk, and Winther [5, 6]. One of our goals in this theory is to show how to translate the vector calculus problems into this language, and ultimately derive greater insight into the problems at hand, or at least improve the underlying algorithms.

For differential forms, as for functions, there are several approaches; we describe here the basic analogues, for forms, of finite difference and finite element methods. Both of them rely on a discretization of the underlying space, as a simpli-

cial (and more generally, cell) complex. The first, discrete exterior calculus (DEC), views exterior calculus in algebraic topology terms: as linear functionals on chains (formal linear combinations of simplices of a certain dimension in the cell complex). This viewpoint has proved very fruitful, even in situations far removed from algebraic topology, for example, in movie ratings (or more generally, any multiple ranking type applications such as runoff elections). This is because the concept of cohomology, cycles versus boundaries, and path-dependence are familiar things in many applications where functions are involved. One advantage of this theory is that it works well with preservation of geometric invariants—so even if we do not have a coherent framework for error analysis like we do in FEEC (described below), we know certain features of the geometries will be preserved, and this is useful to ensure stability of algorithms. This is important, for example, in long-time simulation (which is, to a large extent, what we actually care about when we want to solve problems numerically) of, for example, the solar system (which informed some of our timestepping methods, namely the symplectic methods mentioned in Example 2.4.4 above).

The other approach is called finite element exterior calculus (FEEC), which, as its name implies, uses finite element methods, and is in fact the framework introduced in [5, 6]. Just as in the case of functions, the differential forms are approximated by considering forms with piecewise polynomial coefficients. Even in the case of piecewise linear forms, the discretization process is more subtle. The overall solution process is the same: write it in the appropriate weak form, form a matrix equation based on actually integrating against basis elements (the (FINITE ELEMENT) ASSEMBLY PROCESS). The subtlety (and often the challenge in real-world problems) is choosing the right kinds of basis for the problem. Finite element exterior calculus provides a large family of spaces for us to work with [11, 54, 13], [65, Ch. 3], which go well with the types of problems often encountered. The advantage of finite element exterior

calculus is that it provides a full framework for numerical analysis, including very precise error estimates (similar in spirit to those studied in the last section), which often are not available in DEC. It leverages the existing powerful theories of Hilbert complexes, and uses modern analysis concepts (which is essential as these forms are rarely smooth enough to allow classical exterior differentiation). For the theory in subsets of Euclidean space, we frequently consult the standard reference [6]. We also consider an extension of the theory to curved submanifolds of Euclidean space, using the analysis of [50], and present interesting examples.

### 2.6.1 Approximation in Hilbert Complexes

The weak formulations in §1.5 really pay off here, as most of the general work for approximating differential forms is done by considering the Hilbert space approach. Here, we describe a process for which we can approximate Hilbert complexes. It also explains a lot of the previous approximation theory. We consider a Hilbert complex  $(W, d)$  with domain  $(V, d)$ . For approximating this complex, Arnold, Falk, and Winther [6] introduce finite-dimensional subspaces  $V_h \subseteq V$  of the domain, such that the inclusion  $i_h : V_h \hookrightarrow V$  is a morphism, i.e.  $dV_h^k \subseteq V_h^{k+1}$ . With the weak form (1.8.4), we formulate the Galërkin method by restricting to the subspaces:

$$(2.6.1) \quad \begin{aligned} \langle \sigma_h, \tau \rangle - \langle u_h, d\tau \rangle &= 0 & \forall \tau \in V_h^{k-1} \\ \langle d\sigma_h, v \rangle + \langle du_h, dv \rangle + \langle p_h, v \rangle &= \langle f, v \rangle & \forall v \in V_h^k \\ \langle u_h, q \rangle &= 0 & \forall q \in \mathfrak{H}_h^k. \end{aligned}$$

We abbreviate by setting  $\mathfrak{X}_h^k := V_h^{k-1} \times V_h^k \times \mathfrak{H}_h^k$ . We must also assume the existence of bounded, surjective, and idempotent (projection) morphisms  $\pi_h : V \rightarrow V_h$ . It is generally not the orthogonal projection, as that fails to commute with the differentials.

We will see this corresponds to a kind of interpolation operator, like the previously considered polynomial interpolation operators. As a projection, it gives the following quasi-optimality result:

$$\|u - \pi_h u\|_V = \inf_{v \in V_h} \|(I - \pi_h)(u - v)\|_V \leq \|I - \pi_h\| \inf_{v \in V_h} \|u - v\|_V.$$

The problem (2.6.1) is then well-posed, with a Poincaré constant given by  $c_P \|\pi_h^k\|$ , where  $c_P$  is the Poincaré constant for the continuous problem, which we considered previously in our solution theory. This guarantees all the previous abstract results apply to this case. With this, we have the following error estimate, which is the Hilbert complex generalization of Céa's Lemma (Theorem 2.5.2):

**2.6.1 Theorem** (Arnold, Falk, and Winther [6], Theorem 3.9). *Let  $(V_h, d)$  be a family of subcomplexes of the domain  $(V, d)$  of a closed Hilbert complex, parametrized by  $h$  and admitting uniformly  $V$ -bounded cochain projections  $\pi_h$ , and let  $(\sigma, u, p) \in \mathfrak{X}^k$  be the solution of the continuous problem and  $(\sigma_h, u_h, p_h) \in \mathfrak{X}_h^k$  be the corresponding discrete solution. Then the following quasi-best approximation estimate holds:*

$$(2.6.2) \quad \begin{aligned} \|(\sigma - \sigma_h, u - u_h, p - p_h)\|_{\mathfrak{X}} &= \|\sigma - \sigma_h\|_V + \|u - u_h\|_V + \|p - p_h\| \\ &\leq C \left( \inf_{\tau \in V_h^{k-1}} \|\sigma - \tau\|_V + \inf_{v \in V_h^k} \|u - v\|_V + \inf_{q \in V_h^k} \|p - q\|_V + \mu \inf_{v \in V_h^k} \|P_{\mathfrak{B}} u - v\|_V \right) \end{aligned}$$

with  $\mu = \mu_h^k = \sup_{\substack{r \in \mathfrak{H}^k \\ \|r\|=1}} \|(I - \pi_h^k) r\|$ , the operator norm of  $I - \pi_h^k$  restricted to  $\mathfrak{H}^k$ .

**2.6.2 Corollary.** *If the  $V_h$  approximate  $V$ , that is, for all  $u \in V$ ,  $\inf_{v \in V_h} \|u - v\|_V \rightarrow 0$  as  $h \rightarrow 0$ , we have convergence of the approximations.*

In general, the harmonic spaces  $\mathfrak{H}^k$  and  $\mathfrak{H}_h^k$  do not coincide, but they are isomorphic under many circumstances we shall consider (namely, the spaces are isomorphic if for all harmonic forms  $q \in \mathfrak{H}^k$ , the error  $\|q - \pi_h q\|$  is at most the norm

$\|q\|$  itself [6, Theorem 3.4], and it *always* holds for the de Rham complex). For a quantitative estimate relating the two different kinds of harmonic forms, we have the following

**2.6.3 Theorem** ([6], Theorem 3.5). *Let  $(V, d)$  be a bounded, closed Hilbert complex,  $(V_h, d)$  a Hilbert subcomplex, and  $\pi_h$  a bounded cochain projection. Then*

$$(2.6.3) \quad \|(I - P_{\mathfrak{H}_h})q\|_V \leq \|(I - \pi_h^k)q\|_V, \forall q \in \mathfrak{H}^k$$

$$(2.6.4) \quad \|(I - P_{\mathfrak{H}})q\|_V \leq \|(I - \pi_h^k)P_{\mathfrak{H}}q\|_V, \forall q \in \mathfrak{H}_h^k.$$

## 2.6.2 Approximation with Variational Crimes

For geometric problems, it is essential to remove the requirement that the approximating complex  $V_h$  actually be subspaces of  $V$ . This is motivated by the example of approximating planar domains with curved boundaries by piecewise-linear approximations, resulting in finite element spaces that lie in a different function space [10]. Holst and Stern [50] extend the Arnold, Falk, Winther [6] framework by supposing that  $i_h : V_h \hookrightarrow V$  is an injective morphism which is not necessarily inclusion; they also require projection morphisms  $\pi_h : V \rightarrow V_h$  with the property  $\pi_h \circ i_h = \text{id}$ , which replaces the idempotency requirement of the preceding case. To summarize, our setup is that we are given  $(W, d)$  a Hilbert complex with domain  $(V, d)$ ,  $(W_h, d_h)$  another complex (whose inner product we denote  $\langle \cdot, \cdot \rangle_h$ ) with domain  $(V_h, d_h)$ , injective morphisms  $i_h : W_h \hookrightarrow W$ , and finally, projection morphisms  $\pi_h : V \rightarrow V_h$ . We then have the following generalized Galerkin problem:

$$\begin{aligned}
(2.6.5) \quad & \langle \sigma_h, \tau_h \rangle_h - \langle u_h, d_h \tau_h \rangle_h = 0 & \forall \tau_h \in V_h^{k-1} \\
& \langle d_h \sigma_h, v_h \rangle_h + \langle d_h u_h, d_h v_h \rangle_h + \langle p_h, v_h \rangle_h = \langle f_h, v_h \rangle_h & \forall v_h \in V_h^k \\
& \langle u_h, q_h \rangle_h = 0 & \forall q_h \in \mathfrak{H}_h^k,
\end{aligned}$$

where  $f_h$  is some interpolation of the given data  $f$  into the space  $W_h$  (we will discuss various choices of this operator later). This gives us a bilinear form

$$\begin{aligned}
(2.6.6) \quad B_h(\sigma_h, u_h, p_h; \tau_h, v_h, q_h) := & \langle \sigma_h, \tau_h \rangle_h - \langle u_h, d_h \tau_h \rangle_h \\
& + \langle d_h \sigma_h, v_h \rangle_h + \langle d_h u_h, d_h v_h \rangle_h + \langle p_h, v_h \rangle_h - \langle u_h, q_h \rangle_h.
\end{aligned}$$

This problem is well-posed, which again follows from the abstract theory as long as the complex is closed, and there is a corresponding Poincaré inequality:

**2.6.4 Theorem** (Holst and Stern [50], Theorem 3.5 and Corollary 3.6). *Let  $(V, d)$  and  $(V_h, d_h)$  be bounded closed Hilbert complexes, with morphisms  $i_h : V_h \hookrightarrow V$  and  $\pi_h : V \rightarrow V_h$  such that  $\pi_h \circ i_h = \text{id}$ . Then*

$$\|v_h\|_{V_h} \leq c_P \left\| \pi_h^k \right\| \left\| i_h^{k+1} \right\| \|d_h v_h\|_{V_h},$$

where  $c_P$  is the Poincaré constant corresponding to the continuous problem. If  $(V, d)$  and  $(V_h, d_h)$  are the domain complexes of closed complexes  $(W, d)$  and  $(W_h, d_h)$ , then  $\|d_h v_h\|_{V_h}$  is simply  $\|d_h v_h\|_h$  (since it is the graph norm and  $d^2 = 0$ ).

In other words, the norm of the injective morphisms  $i_h$  also contributes to the stability constant for this discrete problem. Analysis of this method results in two additional error terms (along with now having to explicitly reference the injective morphisms  $i_h$  which may no longer be inclusions), due to the inner product in the space



$V_h$  no longer necessarily being the restriction of that in  $V$ , the need to approximate the data  $f$ , and the failure of the morphisms  $i_h$  to be unitary:

**2.6.5 Theorem** (Holst and Stern [50], Corollary 3.11). *Let  $(V, d)$  be the domain complex of a closed Hilbert complex  $(W, d)$ , and  $(V_h, d_h)$  the domain complex of  $(W_h, d_h)$  with morphisms  $i_h : W_h \rightarrow W$  and  $\pi_h : V \rightarrow V_h$  as above. Then if we have a solutions  $(\sigma, u, p)$  and  $(\sigma_h, u_h, p_h)$  to (1.8.4) and (2.6.5) respectively, the following error estimate holds:*

$$(2.6.7) \quad \|\sigma - i_h \sigma_h\|_V + \|u - i_h u_h\|_V + \|p - i_h p_h\| \\ \leq C \left( \inf_{\tau \in i_h V_h^{k-1}} \|\sigma - \tau\|_V + \inf_{v \in i_h V_h^k} \|u - v\|_V + \inf_{q \in i_h V_h^k} \|p - q\|_V + \mu \inf_{v \in i_h V_h^k} \|P_{\mathfrak{B}} u - v\|_V \right. \\ \left. + \|f_h - i_h^* f\|_h + \|I - J_h\| \|f\| \right),$$

where  $J_h = i_h^* i_h$ , and  $\mu = \mu_h^k = \sup_{\substack{r \in \mathfrak{S}^k \\ \|r\|=1}} \|(I - i_h^k \pi_h^k) r\|$ .

The extra terms (the third line of the inequality above) are called VARIATIONAL CRIMES, which describe a situation in which the approximating weak (bilinear) forms are no longer necessarily the restriction of the weak form of the continuous problem. These terms are analogous to those described in the Strang lemmas ([11, §III.1], [13, Ch. 10]), which detail the analysis for functions on open subsets of  $\mathbb{R}^n$  and have diverse applications, such as approximating domains with curved boundaries and numerical quadrature. They are said to be *crimes*, a terminology of Strang [103], since they depart from the natural assumption of variational problems that the approximating spaces be subspaces. The main idea of the proof of Theorem 2.6.5 (which we will recall in more detail below, because we generalize it in proving our main results) is to form an intermediate complex by pulling the inner products in the complex  $(W, d)$  back to  $(W_h, d_h)$  by  $i_h$ , construct a solution to the problem there, and compare that solution with the solution we want. This modified inner product does not coincide with the

given one on  $W_h$  precisely when  $i_h$  is not unitary:

$$\langle v, w \rangle_{i_h^* W} = \langle i_h v, i_h w \rangle_h = \langle i_h^* i_h v, w \rangle_h = \langle J_h v, w \rangle_h.$$

Unitarity is then precisely the condition  $J_h = I$ . The complex  $W_h$  with the modified inner product now may be identified with a true subcomplex of  $W$ , for which the theory of [6] directly applies, yielding a solution  $(\sigma'_h, u'_h, p'_h) \in V_h^{k-1} \times V_h^k \times \mathfrak{H}_h'^k$ , where  $\mathfrak{H}_h'^k$  is the discrete harmonic space associated to the space with the modified inner product. This generally does not coincide with the discrete harmonic space  $\mathfrak{H}_h^k$ , since the discrete codifferential  $d_h^{*'}$  in that case is defined to be the adjoint with respect to the modified inner product, yielding a different Hodge decomposition. The estimate of  $\|i_h \sigma'_h - \sigma\|_V + \|i_h u'_h - u\|_V + \|i_h p'_h - p\|$  then proceeds directly from the preceding theory for subcomplexes (4.2.7). The variational crimes, on the other hand, arise from comparing the solution  $(\sigma_h, u_h, p_h)$  with  $(\sigma'_h, u'_h, p'_h)$ . Finally, the error estimate (2.6.7) proceeds by the triangle inequality (and the boundedness of the morphisms  $i_h$ ).

### 2.6.3 Polynomial Spaces and Error Estimates for Differential Forms

As in the theory of polynomial approximation of functions by polynomials, we can define polynomial spaces, and the relevant interpolation operators, for differential forms, and derive good estimates in terms of powers of the mesh size. Then, since we have analogue of Céa's Lemma using the abstract Hilbert complex theory above, we can, just as we did for functions, now express the approximation error in the concrete terms of the power of the mesh size, analogous to (2.5.1). A detailed description of how this is done, which we summarize and follow here, is given in the two standard papers of FEEC of Arnold, Falk, and Winther [5, §§4.5-5.3] and [6, §5]. Interesting is the construction of the bounded cochain operator  $\pi_h^k$ , which is central to making the

approximation of the Hilbert complex work.

**2.6.6 Polynomial spaces.** The first, most straightforward polynomial space is defined on  $\mathbb{R}^n$ :

$$(2.6.8) \quad \mathcal{P}_r \Lambda^k(\mathbb{R}^n) = \left\{ \omega \in \Omega^k(\mathbb{R}^n) : \omega = \sum_I a_I dx^I, a_I \text{ is a degree } r \text{ polynomial} \right\},$$

where  $I$  is, as usual, an increasing index set of length  $k$ , and  $dx^I$  is  $dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ . Despite its “list-of-functions” componentwise definition, this space is quite useful. Interpolation into this space is done more invariantly, and involves integration over the faces, rather than simple evaluation of each component at interior points (we’ll talk about that in a bit). However, we also will need another space of polynomials, which involves a geometric operator that is very much like a dual to the operator  $d$  (in fact, it is analogous to the cone operator in the Poincaré Lemma).

**2.6.7 Definition.** Let  $X$  be the radial vector field  $x^i \frac{\partial}{\partial x^i}$  in  $\mathbb{R}^n$ . We define

$$\kappa \omega := X \lrcorner \omega,$$

called the KOSZUL DIFFERENTIAL. It is called a “differential” because  $\kappa \circ \kappa = 0$  and it satisfies a product rule (this is clear from its definition as an interior product). Note that for a polynomial differential form, it replaces one of the  $dx^i$ ’s with  $x^i$ , so in particular, it *increases* the *polynomial* degree (multiplying everything by an extra  $x^i$ ), but *decreases* the *form* degree (there are fewer factors of  $dx^i$ ). This is the opposite effect of  $d$ .

Now we define  $\mathcal{H}_r \Lambda^k$  to be  $k$ -forms with homogeneous  $r$ th degree polynomial coefficients: only terms of degree  $r$  are permitted (and of course, zero). On these

HOMOGENEOUS FORMS,  $\kappa$  satisfies the property [6, Theorem 5.1]

$$(d\kappa + \kappa d)\omega = (r + k)\omega$$

which, in the terminology of algebraic topology, gives a CHAIN HOMOTOPY of  $(r + k)$  times the identity to 0 (again, similar to what is done for the Poincaré lemma), and shows that  $d$  is injective on the range of  $\kappa$  and  $\kappa$  is injective on the range of  $d$ . This also means that we can form chain complexes with  $\kappa$ . Also, if  $r, k \geq 0$  and  $r + k > 0$ ,  $\mathcal{H}_r \Lambda^k = \kappa \mathcal{H}_{r-1} \Lambda^{k+1} \oplus d \mathcal{H}_{r+1} \Lambda^{k-1}$ . We now define

$$\mathcal{P}_r^- \Lambda^k = \mathcal{P}_{r-1} \Lambda^k + \kappa \mathcal{H}_{r-1} \Lambda^{k+1}.$$

This sum is direct, i.e. any  $\omega \in \mathcal{P}_r^- \Lambda^k$ , it can be written in one and only one way as such a sum. It is also an AFFINE INVARIANT, namely, if we have an affine change of coordinates  $\mathbf{y} = \Phi(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ , with  $A$  a matrix and  $\mathbf{b}$  some fixed vector,

$$\Phi^* \left( \mathcal{P}_r \Lambda^k \right) = \mathcal{P}_r \Lambda^k$$

(this is obvious), and *also*

$$\Phi^* \left( \mathcal{P}_r^- \Lambda^k \right) = \mathcal{P}_r^- \Lambda^k,$$

despite that  $\kappa$  uses the coordinate-dependent radial field  $X$  (although the direct sum decomposition will not be the same). The spaces with  $\mathcal{P}_r^- \Lambda^k$  will be instrumental in defining the degrees of freedom (dual spaces) for the spaces  $\mathcal{P}_r \Lambda^k$ , and we shall see  $\mathcal{P}_r^- \Lambda^k$  is intimately related to the structure of the subfaces of the simplex.

We obviously have  $\mathcal{P}_{r-1} \Lambda^k \subseteq \mathcal{P}_r^- \Lambda^k \subseteq \mathcal{P}_r \Lambda^k$ . We use this to define various different complexes of polynomial spaces by considering different image spaces in the differential complex. Specifically, we can regard  $d$  as taking  $\mathcal{P}_r \Lambda^k$  into  $\mathcal{P}_r^- \Lambda^{k+1}$ , or

$\mathcal{P}_{r-1}\Lambda^{k+1}$ , and given the space  $\mathcal{P}_r^-\Lambda^k$ , we can also choose  $\mathcal{P}_r^-\Lambda^{k+1}$  or  $\mathcal{P}_{r-1}\Lambda^{k+1}$  (the general rule of thumb: one keeps the polynomial degree the same if one agrees to use the space with the  $-$ , and one decreases the polynomial degree when choosing the full space (and in all cases,  $d$  increases the form degree, as it always does). This forms  $2^{n-1}$  different possible complexes, with the complex consisting of  $\mathcal{P}_r^-$  for all spaces being the largest, and the complex decreasing  $\mathcal{P}_r \rightarrow \mathcal{P}_{r-1} \rightarrow \dots$  being the smallest.

**2.6.8 Geometric decomposition of the dual spaces of a simplex.** We now get to the reason why we care about the  $\mathcal{P}_r^-$  spaces in the first place. First, of course, given a simplex  $T$ , we can restrict polynomials to  $T$ , leading to the spaces  $\mathcal{P}_r\Lambda^k(T)$  and  $\mathcal{P}_r^-\Lambda^k(T)$ . Then for any face of the complex  $f$ , we define the various polynomial spaces  $\mathcal{P}_r\Lambda^k(f)$  and  $\mathcal{P}_r^-\Lambda^k(f)$  by pulling the forms back via the inclusion (i.e. using the trace). We then have

**2.6.9 Theorem** (Geometric decomposition of the dual, [6], Theorem 5.5). *Let  $r, k, n$  be integers with  $0 \leq k \leq n$  and  $r > 0$  and  $T$  be an  $n$ -simplex in  $\mathbb{R}^n$ . Then*

1. *To each  $f$  a face of  $T$ , we define the space  $W_r^k(T, f) \subseteq \mathcal{P}_r\Lambda^k(T)^*$ :*

$$W_r^k(T, f) := \left\{ \omega \mapsto \int_f \text{Tr}_{T,f} \omega \wedge \eta \mid \eta \in \mathcal{P}_{r+k-\dim f}^-\Lambda^{\dim f-k}(f) \right\}.$$

*Then  $W_r^k(T, f) \cong \mathcal{P}_{r+k-\dim f}^-\Lambda^{\dim f-k}(f)$  by identifying each  $\eta$  with its action via that integral, and*

$$\mathcal{P}_r\Lambda(T)^* \cong \bigoplus_{f \text{ a face of } T} W_r^k(T, f).$$

2. *To each face  $f$  of  $T$ , we define another space  $W_r^{k-}(T, f) \subseteq \mathcal{P}_r^-\Lambda^k(T)^*$ :*

$$W_r^{k-}(T, f) := \left\{ \omega \mapsto \int_f \text{Tr}_{T,f} \omega \wedge \eta \mid \eta \in \mathcal{P}_{r+k-\dim f-1}^-\Lambda^{\dim f-k}(f) \right\}.$$

Then  $W_r^{k-}(T, f) \cong \mathcal{P}_{r+k-\dim f-1} \Lambda^{\dim f-k}(f)$  by the same correspondence, and

$$\mathcal{P}_r^- \Lambda^k(T)^* \cong \bigoplus_{f \text{ a face of } T} W_r^{k-}(T, f).$$

The proof is given in [5, §§4.5-6]. Note how the  $\mathcal{P}_r^-$  spaces are involved in the dual to the  $\mathcal{P}_r$  space, and vice versa. These decompositions make a little more sense when doing interpolations, but in summary, they are the direct generalizations of the evaluation maps used for functions, called DEGREES OF FREEDOM. In the special case of  $k = \dim f$ , the basis function 1 is used, which corresponds simply integration of the trace over the face  $f$ , and the case  $\dim f = 0$ , i.e. a point, it is evaluation.

**2.6.10 Polynomial interpolation: using integration and Stokes's Theorem.** In order to interpolate into the polynomial spaces, we have interpolation operators similar to those for functions. Instead of evaluating the component functions of  $k$ -forms at points (which, again, would amount to simply treating differential forms as lists of functions, so therefore not what we want), we instead integrate over the  $k$ -faces of the simplex. This should not be so surprising, because generalizing integration to geometric problems is what brought about differential forms in the first place, so its use should be instrumental in taking the geometric nature of such objects into account for the interpolation process. We will also see that these operators commute with the differentials, by Stokes's Theorem. To interpolate, we consider the geometric decompositions in the above. Given  $\omega \in C^0 \Omega^k(T)$ , there exists a unique polynomial differential form  $I_h \omega$  such that for every face  $f$  of  $T$  and  $\eta \in P_{r+k-\dim f}^- \Lambda^{\dim f-k}(f)$ ,

$$\int_f \text{Tr}(\omega - I_h \omega) \wedge \eta = 0.$$

If  $k = 0$ , of course, this reduces to the usual polynomial interpolation. For an explicit

computation, we choose a basis for each  $P_{r+k-\dim f}^- \Lambda^{\dim f-k}(f)$ . Each one of these basis elements defines a degree of freedom, an element of  $\mathcal{P}_r \Lambda^k(T)^*$  as in the above, and together are a basis for this dual space, call it  $\{\varepsilon^\ell\}$ . Now we take the dual of this basis, call it  $\{\phi_j\}$ , a basis for  $\mathcal{P}_r \Lambda^k(T)$ ; the defining property of it is  $\varepsilon^\ell(\phi_j) = \delta_j^\ell$ . Then for any  $\omega \in \mathcal{P}_r \Lambda^k(T)$ , we define

$$I_h \omega := \sum_j \varepsilon^j(\omega) \phi_j.$$

Of course, computing  $\{\phi_j\}$  is not immediately obvious, but it is standard linear algebra: we start out with an “easier” basis, which for polynomial spaces is obvious: for every basis  $k$ -forms  $dx^I$  with  $I$  increasing, we consider  $\{1, x, x^2, \dots, x^r\}$ , the obvious polynomial basis. Then if we evaluate the degrees of freedom  $\varepsilon^j$  on this basis, we get coefficients of some matrix; by the usual results of linear algebra, the inverse of this matrix applied to each basis gives the dual basis.

To show that  $d$  commutes with  $I_h$ , this is Stokes’s Theorem and  $d$  commuting with pullbacks (and therefore, traces): for any  $\eta \in P_{r+(k+1)-\dim f}^- \Lambda^{\dim f-(k+1)}$ ,

$$\int_f (d\omega - d(I_h \omega)) \wedge \eta = \int_f (-1)^{k-1} (\omega - (I_h \omega)) \wedge d\eta + \int_{\partial f} (\omega - I_h \omega) \wedge \eta = 0$$

because  $d\eta \in P_{r+k-\dim f}^- \Lambda^{\dim f-k}$  so gives zero since  $I_h \omega$  is in fact the interpolation of  $\omega$ , and  $\partial f$  is the sum of faces of dimension  $\dim f - 1$ , so that simply regrouping the terms,

$$\eta \in P_{r+k-(\dim f-1)}^- \Lambda^{(\dim f-1)-k},$$

and again, since  $I_h \omega$  is the interpolation, this also vanishes. Since  $I_h(d\omega)$  is the unique form with this property, it follows that  $I_h(d\omega) = d(I_h \omega)$ .

**2.6.11 Finite element assembly.** Now for a general triangulation  $\mathcal{T}$  of a domain  $U$

with piecewise smooth, Lipschitz boundary, we ASSEMBLE the finite element spaces  $\mathcal{P}_r \Lambda^k(\mathcal{T})$  and  $\mathcal{P}_r^- \Lambda^k(\mathcal{T})$ . This is essentially assembling them piecewise, requiring certain interelement continuity conditions. These conditions are analogous to electrostatic and magnetostatic boundary conditions [42, 56, 85], namely, their traces to any common faces should be equal. This says for vectors tangent to the boundary, the forms on both sides must agree, but for vectors normal to the boundaries, they don't need to agree. Actually, the regularity of the  $H\Omega^k$  is the exact regularity needed. Namely, we have [6, Theorem 5.7]:

$$(2.6.9) \quad \mathcal{P}_r \Lambda^k(\mathcal{T}) = \{\omega \in H\Omega^k(U) : \omega|_T \in \mathcal{P}_r \Lambda^k(T) \forall T \in \mathcal{T}\}$$

$$(2.6.10) \quad \mathcal{P}_r^- \Lambda^k(\mathcal{T}) = \{\omega \in H\Omega^k(U) : \omega|_T \in \mathcal{P}_r^- \Lambda^k(T) \forall T \in \mathcal{T}\}$$

**2.6.12 Bounded Cochain Projections.** Our interpolation operators for differential forms are insufficient for precisely the same reasons they were for functions: they only work for continuous forms. As we have seen,  $H\Omega^k$  allows discontinuities in the normal or tangential components. Tracing onto lower dimensional simplexes usually cannot be done without higher regularity [6, 34]; tracing to various lower dimensional simplices require a degree of regularity between the usual Trace Theorem case and the Sobolev Embedding Theorem. We can use the analogue of the Clément interpolant for functions to interpolate forms. However, those operators fail to commute with the differentials, which is something we need for the bounded cochain operators required by the Hilbert complex theory.

The strategy is to take, for general  $\omega \in \mathcal{L}^2\Omega^k(U)$ , some form of smoothing which makes it continuous. Then the canonical interpolation operators above can be applied. For the smoothing, we use convolutions, using mollifiers [30, 6]. We average with the pullbacks of some translates to some distance  $\varepsilon$ ; this makes the operator



commute with the differentials. Along with the interpolations as above, we get some operator mapping  $\mathcal{L}^2\Omega^k$  into the desired polynomial space. The only problem with this is that the spaces the operators are not idempotent. This is fixed by establishing that the operators converge in the  $\mathcal{L}^2$  norm to the identity, uniformly in  $h$ . Composing with a fixed inverse of one of these interpolation operators with  $\varepsilon$  sufficiently small gives a smoothing operator that is idempotent. The details of this construction are presented in [5].

## Chapter 3

# Some Finite Element Methods for Nonlinear Equations

We would like to see if we can modify FEM for *nonlinear* differential operators, because most of the interesting problems in geometry are governed by such equations. In fact, it is even more critical to have good numerical methods at one's disposal, since such equations are difficult, if not impossible, to solve analytically. Here we follow Michael Holst's brief development in the documentation for his software, MCLite [52], and for more precision and detail regarding selection of the right function spaces, [37]. For the general nonlinear approximation theory and many results on Newton's method, we follow [97, §§10.2-4] and [53, §§2.8-9 and §A.5]. The theory of nonlinear equations is of course a very vast and difficult subject, so necessarily we only touch on a few techniques, and mention where the difficulties start.

### 3.1 Overview

The basic idea here is that, in solving nonlinear elliptic equations, we use Newton's Method to approximate solutions to the equations instead of directly solving the system using linear algebra. We assume for our purposes that the equation still may be written in some kind of weak form, so we will not consider fully general nonlinear equations. The theory is of course much harder, and finding the right function spaces in order even define the appropriate weak forms can be very subtle [37, Ch. 3]. For second-order equations, we will work with nonlinear equations of the form

$$F(u)(x) := -\nabla \cdot \mathbf{a}(x, u(x), \nabla u(x)) + b(x, u(x), \nabla u(x)) = f(x)$$

where  $u$  and  $f$  are in the appropriate function spaces,  $\mathbf{a} : \Omega \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a nonlinear vector field, depending also on  $u$  and  $\nabla u$  (really, a vector field on the 1-jet bundle), and  $b : \Omega \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$  is a scalar function on the 1-jet bundle. Of course, this includes the linear case

$$Lu := -\nabla \cdot (\mathbf{A}(x)(\nabla u)) + \mathbf{b}(x) \cdot \nabla u + cu$$

where  $\mathbf{a}(x, u, \nabla u) = \mathbf{A}(x)(\nabla u)$ , or,  $a^i(x, u, \nabla u) = a^{ij}(x) \partial_j u$ , and  $b(x, u, \nabla u) = \mathbf{b}(x) \cdot \nabla u + cu$ . But now  $\mathbf{a}$  can depend nonlinearly on  $\nabla u$  (as well as  $u$ ). On manifolds, we will actually prefer the 1-form  $du$  rather than the gradient vector  $\nabla u$ , as that gives the most natural formulation of the equations (and we treat it as such when dealing with the transformation rules we actually use in computing).

This is not the most general nonlinear second order equation we can come up with, due to the assumption that it is written in a kind of DIVERGENCE FORM (preceding the nonlinear vector field with a  $-\nabla \cdot$ ). We use divergence form for the same reason we used it in linear equations: we find a weak formulation of the problem, in

which integration by parts may be used to transfer that divergence onto something else, thus still allowing us to use a form of integration by parts (and also require less differentiability of the solution we are seeking).

**3.1.1 Nonlinear ellipticity.** A nonlinear, second order equation is called ELLIPTIC if its LINEARIZATION is elliptic at each point. To calculate the linearization, we employ the directional derivative (with the caveat that everything is infinite-dimensional; it is called the GÂTEAUX DERIVATIVE in this case) via the chain rule:

$$\begin{aligned} DF[u]w &= \left. \frac{d}{ds} \right|_{s=0} F(u + sw) = -\nabla \cdot \left( \sum_i \frac{\partial \mathbf{a}}{\partial u_{x_i}}(x, u, \nabla u) \partial_i w \right) - \nabla \cdot \left( \frac{\partial \mathbf{a}}{\partial u}(x, u, \nabla u) w \right) \\ &\quad + \frac{\partial b}{\partial u}(x, u, \nabla u) w + \sum_i \frac{\partial b}{\partial u_{x_i}}(x, u, \nabla u) \partial_i w \\ &= -\nabla \cdot \left( \sum_i \frac{\partial \mathbf{a}}{\partial u_{x_i}}(x, u, \nabla u) \partial_i w \right) + \sum_i \left( -\frac{\partial a^i}{\partial u}(x, u, \nabla u) + \frac{\partial b}{\partial u_{x_i}}(x, u, \nabla u) \right) \partial_i w \\ &\quad + \left( -\nabla \cdot \frac{\partial \mathbf{a}}{\partial u}(x, u, \nabla u) + \frac{\partial b}{\partial u}(x, u, \nabla u) \right) w, \end{aligned}$$

where at the end we made it look seemingly more complicated, in order to independently recognize the 2nd, 1st, and 0th order terms in a divergence-form linear operator we studied earlier. So, for a fixed  $u$ ,  $DF[u]$  acts on  $w$  as a linear operator, and it is (uniformly) elliptic precisely if there exists  $\theta > 0$ , depending on  $u$ , such that

$$\sum_{i,j} \frac{\partial a^j}{\partial u_{x_i}}(x, u, \nabla u) \xi_i \xi_j \geq \theta |\xi|^2.$$

So we say a nonlinear operator is ELLIPTIC precisely when the above holds for all  $u$ ,  $\xi \in \mathbb{R}^n$  and at all  $x \in \Omega$ . Elliptic nonlinear operators abound in differential geometry [19, 18].

It is worth mentioning some special cases of nonlinearity, because many problems also fall under these classes. The PDE is called SEMILINEAR if the only nonlinearity

in the equation is from the  $b$  term, that is,  $\mathbf{a}(x, u, \nabla u)$  in fact is like the highest order terms of a linear operator. There are also QUASILINEAR equations, which means when everything is expanded, the equation is linear in the second derivatives, with coefficients that may depend on lower order terms (i.e. like the semilinear case except the  $a$  may also depend on  $u$  and  $\nabla u$ ). But all divergence-form operators as described here are actually quasilinear, as one can check by using the Chain rule. The prototypical quasilinear equation which has been a large motivation in their study is the mean curvature equation [39].

**3.1.2 Weak formulation and discussion of function spaces.** As noted before, we wish to find a weak formulation, in order to be able to place things in a framework suitable for the finite element method. To find the weak formulation, we first operate formally and use integration by parts: the weak formulation is, for suitable  $v$  (assume, for now, that it is in  $C_c^\infty(\Omega)$ ),

$$(3.1.1) \quad \langle F(u), v \rangle := \int_{\Omega} \mathbf{a}(x, u(x), \nabla u(x)) \cdot \nabla v(x) + b(x, u, \nabla u) v(x) dx = \int_{\Omega} f(x) v(x) dx.$$

Because we only need one weak derivative of  $u$  to make this well-defined, a weak solution  $u$ , as in the linear case, does not need as many derivatives as the strong (classical) formulation would seem to indicate. However, difficulty arises from the nonlinearity, since, if we wish to realize the functional as being in some Sobolev space (so that it acts on  $v$  in another Sobolev space), the integral needs to always be well-defined (in order to be a bounded linear functional), thus imposing conditions on the nature of the coefficients  $\mathbf{a}$ . Analyzing the integral using Hölder's inequality, we can derive some conditions for polynomial growth of the coefficients in the  $(u, \nabla u)$  variables. For example, if they are continuous in  $u$  and  $\nabla u$ , and bounded by a  $(p-1)$ th order polynomials in  $u$  and  $\nabla u$ , this ensures it is well-defined for  $u, v \in W^{k,p}(\Omega)$ , for

suitable  $k$  [37, §§12-13]. However, in the following, we do consider a problem with exponential coefficients, which often also still works [37, §16]. Generally, one needs the theory of SOBOLEV-ORLICZ SPACES to find weak solutions with growth conditions like these.

However, for our purposes, we can establish the well-posedness of the continuous problem in a different manner—for example, if we can find, in fact, a classical solution to the equation on a compact manifold, then the solution, together with its derivative, is always bounded and in any  $W^{k,p}$  space we would like, and so the weak form of the equation is well-defined, by integration by parts. The weak form is still useful as a setup for the approximation theory.

**3.1.3 Example** (Ricci Flow on a Surface). Consider a Riemannian manifold  $(M, g_0)$  of dimension 2. Suppose we wish to solve the Ricci Flow equation [19, 18],

$$(3.1.2) \quad \frac{\partial g}{\partial t} = -2\text{Rc} = -2Kg$$

$$(3.1.3) \quad g(0) = g_0$$

where  $K$  is the Gaußian curvature of the surface (the simplification  $\text{Rc} = Kg$  is possible only in dimension 2). A further simplification can be made by initially supposing (making an ansatz) that the evolving metric is conformal to the initial metric, that is, there exists a “potential function”  $u(x, t)$  such that

$$g(x, t) = e^{2u(x,t)} g_0(x).$$

Substituting  $g(t) = e^{2u} g_0$  into the Ricci Flow equation, we have

$$2e^{2u} \frac{\partial u}{\partial t} g_0 = -2K[e^{2u} g_0] e^{2u} g_0.$$

Now we take advantage of the fact that  $K$  in the new metric is related to the original  $K$  by the following transformation formula:

$$K[e^{2u}g] = e^{-2u}(-\Delta u + K[g])$$

where  $\Delta$  is the Laplacian in the original metric. Thus the equation now reads

$$(3.1.4) \quad 2e^{2u} \frac{\partial u}{\partial t} g_0 = -2(-\Delta u + K) g_0.$$

Since  $g_0$  is nondegenerate, the scalars in the above must be equal, so that

$$(3.1.5) \quad \frac{\partial u}{\partial t} = e^{-2u}(\Delta u - K).$$

It is shown in [18, Ch. 5] that this equation is well-posed, exists for all time, and converges to the metric of constant curvature guaranteed by the Uniformization Theorem. This equation is a PDE in  $u$  and  $u$  alone, without reference to necessarily more complicated tensor quantities (only quantities derived from the *initial* metric such as  $\Delta$ ,  $\nabla$ , and  $K$ ). Finally, we can rewrite this in the nonlinear divergence form given above, by guessing the high-order term should look something like  $\nabla \cdot (e^{-2u} \nabla u)$ :

$$(3.1.6) \quad \nabla \cdot (e^{-2u} \nabla u) = \nabla(e^{-2u}) \cdot \nabla u + e^{-2u} \Delta u = -2e^{-2u} |\nabla u|^2 + e^{-2u} \Delta u.$$

So

$$e^{-2u} \Delta u = \nabla \cdot (e^{-2u} \nabla u) + 2e^{-2u} |\nabla u|^2$$

and we have

$$\frac{\partial u}{\partial t} = \nabla \cdot (e^{-2u} \nabla u) + 2e^{-2u} |\nabla u|^2 - e^{-2u} K$$

We define

$$F(u) = -\nabla \cdot (e^{-2u} \nabla u) - 2e^{-2u} |\nabla u|^2 + e^{-2u} K$$

to be the (negative) spatial part of the equation. Now  $F$  conforms to the divergence-form operator with  $\mathbf{a}(x, u, \nabla u) = e^{-2u} \nabla u$  and  $b(x, u, \nabla u) = -2e^{-2u} \|\nabla u\|^2 + e^{-2u} K$ . We then define

$$a^i(x, u, \nabla u) = e^{-2u} \partial_i u$$

so

$$\frac{\partial a^i}{\partial u_{x_j}}(x, u, \nabla u) = \frac{\partial}{\partial u_{x_j}}(e^{-2u} u_{x_i}) = e^{-2u} \delta_{ij}.$$

Simply choosing  $\theta(u) = \min_{x \in M} e^{-2u(x,t)} > 0$  (the minimum is guaranteed to be positive on a closed surface and compact interval of time), we see that  $F$  is a quasilinear elliptic operator. However, due to a coefficient being exponential in  $u$ , as mentioned above special considerations must be made to find the right spaces for a correct weak formulation.

**3.1.4 The correct function spaces for this problem.** If we have existence and uniqueness for this differential equation in  $u$  ([18, Ch. 5]), we have now actually shown, by multiplying  $g_0$  by  $e^{2u}$ , the calculation (3.1.4), and the uniqueness of solutions to Ricci flow, that any solution to Ricci Flow on the surface must indeed be given by a conformal change, with conformal factor satisfying the equation (3.1.6). We recall the spatial weak form for  $F(u) = f$ :

$$(3.1.7) \quad \langle F(u), v \rangle = \int_M e^{-2u} \nabla u \cdot \nabla v - 2e^{-2u} |\nabla u|^2 v + e^{-2u} K v \, d\mu = \int_M f v \, d\mu.$$

This is itself interesting to solve. The interpretation here is that  $F(u)$  gives the Gaussian curvature of the metric  $e^{2u} g$  and is studied in [59, 20]. If this problem is solvable for  $f$  given as a constant equal to the sign of the Euler characteristic of  $M$ , this gives



the UNIFORMIZATION THEOREM, which states that every compact Riemannian 2-manifold (surface) admits a metric of constant curvature, conformal to the given metric. The Ricci flow equation turns this into a parabolic question, and in fact attempts to realize equilibrium solution (solve elliptic problems) by taking the steady state of the corresponding parabolic problem (an interesting and useful technique in general). As we have seen, taking the parabolic view, the actual computation is quite different, because one is not attempting to invert the actual elliptic operator itself, and thus has less stringent requirements for existence and uniqueness. For example, for linear parabolic operators described in §1.9.2, the operator need only satisfy a Gårding inequality. However, more theory is still needed, because the question now is one of *long-time* existence and *convergence* of the solution.

Returning to our example, we linearize the operator  $F$ . It is what we will need in order to use finite elements to solve the problem. We can either substitute it into the formula we derived for the linearization of a general quasilinear operator above, or we can derive it directly:

(3.1.8)

$$\begin{aligned} (DF(u)w, v)_{\mathcal{L}^2} &= \left. \frac{d}{dt} \right|_{t=0} \int e^{-2(u+tw)} (\nabla(u+tw) \cdot \nabla v - 2|\nabla(u+tw)|^2 v + Kv) d\mu \\ &= \int -2e^{-2u} w (\nabla u \cdot \nabla v - 2|\nabla u|^2 v + Kv) + e^{-2u} (\nabla w \cdot \nabla v - 4\nabla u \cdot \nabla w v) d\mu \\ &= \int e^{-2u} (\nabla w \cdot \nabla v - 2\nabla u \cdot w \nabla v - 4\nabla u \cdot v \nabla w + (4|\nabla u|^2 - 2K)wv) d\mu. \end{aligned}$$

**3.1.5 Example** (Time-Dependent Integral Version). There actually is another way to formulate this equation, which is useful for analysis using maximum principles. As before, suppose  $u(x, t)$  is a solution to the equation we derived above (the Ricci flow

equation for the conformal factor):

$$\frac{\partial u}{\partial t} = e^{-2u}(\Delta u - K).$$

But for conformal changes of metric, the Laplacian transforms oppositely:  $\Delta_{g(t)} = \Delta_{e^{2u}g} = e^{-2u}\Delta_g$ . So therefore, we have

$$\frac{\partial u}{\partial t} = \Delta_{g(t)}u - e^{-2u}K.$$

This makes the weak form of the elliptic part easier to see:

$$\int_M \nabla_{g(t)}u \cdot \nabla_{g(t)}v - e^{-2u}Kv = \int_M f v.$$

This looks almost like the linear case, at least for the derivative term. However, the difficulty is that the metric changes in time. Thus, while the same setup for approximation applies here, it still, of course, leads to nonlinear equations. We will describe this more in detail in Chapter 5.

**3.1.6 Example** (Derivation of the Normalized Ricci Flow [18], Ch. 5). We take another detour into the general theory of Ricci flow. The ordinary Ricci flow equation often leads to singularities in finite time, because metrics degenerate or curvatures blow up. It is possible to examine what happens “in the limit,” that is, examine what the surface is approaching before the singularity time. This analysis is very important for using the Ricci flow to prove Thurston’s Geometrization Conjecture. However, we can often remove the problem of singularities forming in finite time by looking at the NORMALIZED RICCI FLOW (NRF), which essentially rescales the metric in time in such a manner that the surface area remains constant, and singularities in time are sent off to infinity. To obtain the normalized Ricci flow equation, we suppose there exists a

solution  $g(s)$  to the Ricci flow on some interval  $[0, T)$ , some (invertible) reparametrization of time  $\varphi : [0, T) \rightarrow [0, S)$  (write its inverse as  $\psi$ ) and a time-dependent conformal factor  $c(t) > 0$ . With this, we define

$$\tilde{g}(t) := c(\varphi(t))g(\varphi(t)).$$

To figure out what  $c$  and  $\varphi$  must be, we shall demand the metric  $\tilde{g}(t)$  have constant volume (i.e.  $\int_M d\tilde{\mu}(t)$  is actually time-independent). This seems a reasonable way to prevent a manifold from “collapsing” so that we can examine limiting behavior (similar to how difference quotient divides out the smallness, obtaining calculus without the use of infinitesimals). Differentiating with respect to  $t$ , we have

$$\begin{aligned} (3.1.9) \quad \frac{\partial \tilde{g}}{\partial t} &= c'(\varphi(t))\varphi'(t)g(\varphi(t)) + c(\varphi(t))\frac{\partial}{\partial t}g(\varphi(t)) \\ &= c'(\varphi(t))\varphi'(t)g(\varphi(t)) + c(\varphi(t))\frac{\partial g}{\partial s}(\varphi(t))\varphi'(t) \\ &= \varphi'(t) [c'(\varphi(t))g(\varphi(t)) - 2c(\varphi(t))\text{Rc}[g(\varphi(t))]] \\ &= \varphi'(t) \left( \frac{c'(\varphi(t))}{c(\varphi(t))} \tilde{g}(t) - 2c(\varphi(t))\text{Rc}[\tilde{g}(t)] \right). \end{aligned}$$

Here we have used the nontrivial fact [19, §1.5] that  $\text{Rc}[Ch] = \text{Rc}[h]$  for any  $C > 0$ , that is, Ricci curvature is invariant under constant conformal changes of metric. Now the demand of constant volume gives us

$$\begin{aligned} 0 &= \frac{d}{dt} \int_M d\tilde{\mu}(t) = \int_M \frac{\partial}{\partial t} \sqrt{\det(\tilde{g}_{ij}(t))} dx = \int_M \frac{\det(\tilde{g}_{ij}) \tilde{g}^{k\ell} \frac{\partial \tilde{g}_{k\ell}}{\partial t}}{2\sqrt{\det(\tilde{g}_{ij}(t))}} dx = \int_M \frac{1}{2} \tilde{g}^{ij} \frac{\partial \tilde{g}_{ij}}{\partial t} d\tilde{\mu} \\ &= \int_M \frac{1}{2} \tilde{g}^{ij} \varphi'(t) \left( \frac{c'(\varphi(t))}{c(\varphi(t))} \tilde{g}_{ij} - 2c(\varphi(t))\tilde{R}_{ij} \right) d\tilde{\mu} \\ &= \int_M \left[ \frac{n}{2} \left( \frac{c'(\varphi(t))}{c(\varphi(t))} \varphi'(t) \right) - c(\varphi(t))\tilde{R}\varphi'(t) \right] d\tilde{\mu} \end{aligned}$$

where we have differentiated under the integral sign and used the trace formula

$$[D \det(g_{ij})] \nu = \det(g_{ij}) g^{k\ell} \nu_{k\ell}.$$

Finally, we realize  $\tilde{R} = R[c(\varphi(t))g(\varphi(t))] = c(\varphi(t))^{-1}R[g(\varphi(t))]$  and that the parenthesized term is  $\frac{d}{dt} \log(c(\varphi(t)))$ , to finally get

$$\begin{aligned} 0 &= \int_M \left[ \frac{n}{2} \frac{d}{dt} \log(c(\varphi(t))) - R[g(\varphi(t))] \right] d\tilde{\mu} \\ &= \tilde{V} \frac{n}{2} \frac{d}{dt} \log(c(\varphi(t))) - \int R[g(\varphi(t))] \varphi'(t) d\tilde{\mu}. \end{aligned}$$

By hypothesis,  $\tilde{V}$  is constant. Rearranging, we have

$$\frac{d}{dt} \log(c(\varphi(t))) = \frac{2}{n} \frac{\int R[g(\varphi(t))] d\tilde{\mu}[g(\varphi(t))] \varphi'(t)}{\int d\tilde{\mu}} = \frac{2}{n} \frac{\int R[g(\varphi(t))] d\mu}{\int d\mu[g(\varphi(t))]} \varphi'(t)$$

where the last equality follows because the conformal factors are independent of space.

This is almost what we want, except we have too many  $\varphi$ 's entangled. Define  $r(s)$  to be the RHS of the above equation, without the  $\varphi$ 's:

$$r(s) = \frac{\int R[g(s)] d\mu[g(s)]}{\int d\mu[g(s)]},$$

the AVERAGE SCALAR CURVATURE. So

$$\frac{d}{dt} \log(c(\varphi(t))) = r(\varphi(t)) \varphi'(t).$$

But by the Chain Rule, this suggests the differential equation

$$\frac{d}{ds} \log(c(s)) = r(s),$$

and a reasonable initial condition being  $c(0) = 1$  and  $\varphi(0) = 0$ . This gives us

$$c(s) = \exp\left(\int_0^s r(\sigma) d\sigma\right).$$

Finally, to determine  $\varphi$ , we must make another restriction, related to how we want our differential equation to finally appear. In order to make it look like Ricci flow as much as possible (this is more than just for appearance—we want to make sure we have the same kind of “elliptic” part to ensure the same theories apply), we will demand that  $-2c(\varphi(t))\phi'(t)\widetilde{R}c = -2\widetilde{R}c$ , giving us  $c(\varphi(t))\phi'(t) = 1$  by consulting (3.1.9) above. This suggests defining  $C(s) = \int_0^s c(\sigma) d\sigma$ , which is an antiderivative of  $c$ . Therefore,  $C'(\varphi(t))\phi'(t) = 1$  by the Chain Rule. So  $C(\varphi(t)) = t + K$  for some constant  $K$ . Since we demand  $\varphi(0) = 0$ ,  $C(\varphi(0)) = C(0) = 0$ , so  $K = 0$ . This says that  $C = \psi$ , the inverse of  $\varphi$ .

Together, we have

$$\begin{aligned} c(s) &= \exp\left(\int_0^s r(\sigma) d\sigma\right) \\ \psi(s) &= \int_0^s c(\sigma) d\sigma. \end{aligned}$$

What equation does this give us for NRF? We have seen that  $c(\varphi(t))\phi'(t) = 1$  by definition. Now we just need to calculate the other factor in (3.1.9),  $\frac{d}{dt} \log(c(\varphi(t)))$ . This is

$$\frac{d}{dt} \int_0^{\varphi(t)} r(\sigma) d\sigma = r(\varphi(t))\phi'(t)$$

But now  $\phi'(t) = \frac{1}{\psi'(\varphi(t))} = \frac{1}{c(\varphi(t))}$ . So the factor is

$$\frac{r(\varphi(t))}{c(\varphi(t))} = \frac{\int c(\varphi(t))^{-1} R[g(\varphi(t))] d\tilde{\mu}}{\int d\tilde{\mu}} = \frac{\int R[c(\varphi(t))g(\varphi(t))] d\tilde{\mu}}{\int d\tilde{\mu}} = \tilde{r}(t).$$

Thus the full normalized Ricci flow is

$$\frac{\partial \tilde{g}}{\partial t} = -2\tilde{\text{Rc}} + \frac{2}{n}\tilde{r}\tilde{g}.$$

**3.1.7 Example** (Normalized Ricci flow in 2D). We return to scalar equations and see how the NRF looks, and compare it to what we derived before. We make the ansatz, as before, that the  $\tilde{g}$  remains in its conformal class as the derivative is taken:  $\tilde{g}(t) = e^{2u}\tilde{g}_0$ . Thus, using that  $2/n = 1$  and  $\tilde{r}$  is constant in time,

$$2\frac{\partial u}{\partial t}e^{2u}\tilde{g}_0 = -2\tilde{K}\tilde{g} + \tilde{r}e^{2u}\tilde{g}_0 = (-2(-\Delta u + \tilde{K}_0) + \tilde{r}e^{2u})\tilde{g}_0.$$

Thus, we derive the following equation for  $u$  (and  $u$  alone):

$$\frac{\partial u}{\partial t} = e^{-2u}(\Delta u - \tilde{K}_0) + \tilde{k}$$

where  $\tilde{k} = \tilde{r}/2$  is the average Gauß curvature. But by the Gauß-Bonnet Theorem,  $\tilde{k} = \chi(M)/\tilde{V}$ , which finally gives us

$$\frac{\partial u}{\partial t} = e^{-2u}(\Delta u - \tilde{K}_0) + \frac{2\pi\chi(M)}{\tilde{V}}.$$

Thus the normalized Ricci flow yields a conformal factor equation that contains an additional source term. We shall see that this is just enough to give us convergence in the limit. We should note that the properties of this flow is special to two dimensions; in three and higher dimensions, Ricci flow is not parabolic.

How do we apply FEM here? We set up the weak form of the problem as before: given  $F$  a *nonlinear* elliptic operator on  $u$ , we recall (3.1.1) above: to solve  $F(u) = f$  weakly, we integrate this equation against a function  $v$  in a suitable space of test

functions, and derive an analogous system of equations as in the linear case. However, the equations are now *nonlinear*, which are more difficult to solve.

### 3.2 Linearizing the Equation

We now explain the process of arriving at a system of equations in the quasilinear case. For more details about this and a more precise discussion of the function spaces involved, see [97, Ch. 10]. To solve  $F(u) = f$  weakly, integrating against a suitable  $v$ , we have (3.1.1):

$$\langle F(u), v \rangle = \int_M \mathbf{a}(x, u(x), \nabla u(x)) \cdot \nabla v(x) + b(x, u, \nabla u) v(x) d\mu = \int f v d\mu.$$

Proceeding as in the general development of FEM, we introduce a basis  $\{\varphi_i\}_{i=1}^N$ , and derive a system of equations for coefficients  $\mathbf{u} = (u^i)$  such that  $u = u^i \varphi_i$ :

$$\langle F(u^i \varphi_i), \varphi_j \rangle = \int_M \mathbf{a}(x, u^i \varphi_i, u^i \nabla \varphi_i) \cdot \nabla \varphi_j + b(x, u^i \varphi_i, u^i \nabla \varphi_i) \varphi_j d\mu = \int f \varphi_i d\mu.$$

Writing  $f_i = \int f \varphi_i d\mu$ ,  $\mathbf{f} = (f_i)_{i=1}^N$ ,  $F_j(\mathbf{u}) = \langle F(u^i \varphi_i), \varphi_j \rangle$ , and finally  $\mathbf{F}(\mathbf{u}) = (F_j(\mathbf{u}))_{j=1}^N$ , this gives us the *nonlinear* equation

$$\mathbf{F}(\mathbf{u}) = \mathbf{f}.$$

Here,  $\mathbf{F}$  is the nonlinear analogue of the stiffness matrix. In order to solve this equation, we can use any of the various methods from numerical analysis to solve nonlinear problems. This can be difficult, as there is no general theory that guarantees existence of solutions. However, in many cases, we can use Newton's method [53, §2.9], [97, §10.4], which often (but not always) gives good results (we discuss this in more depth

in §3.4). Various modifications this method have been devised to improve its reliability. Newton's method says that in order to approximate a solution to  $\mathbf{F}(\mathbf{u}) = \mathbf{f}$ , we chose an initial guess (starting point)  $\mathbf{u}_0$  and compute the sequence

$$\mathbf{u}_{n+1} = \mathbf{u}_n - \mathbf{DF}(\mathbf{u}_n)^{-1}(\mathbf{F}(\mathbf{u}_n) - \mathbf{f})$$

Standard techniques of linear algebra are used to compute the correction term

$$\mathbf{h}_n = -\mathbf{DF}(\mathbf{u}_n)^{-1}(\mathbf{F}(\mathbf{u}_n) - \mathbf{f}),$$

and in fact, each  $\mathbf{DF}(\mathbf{u}_n)$  is the LINEARIZED STIFFNESS MATRIX at  $\mathbf{u}_n$  (see Figure 3.1 for a graphical illustration in 1 dimension). In essence, this linearized problem for the correction  $h$  is the approximation to the solution, for *fixed*  $u$ , to the continuous linearized problem:

$$\langle DF(u)h, \varphi_j \rangle = \langle F(u) - f, \varphi_j \rangle.$$

But  $\langle DF(u)h, \varphi_j \rangle$  is precisely the linearization as before, which we use to check the ellipticity of the nonlinear operator  $F$ :

$$(3.2.1) \quad \langle DF(u)w, v \rangle = \int_M \left( \sum_i \frac{\partial \mathbf{a}}{\partial u_{x_i}}(x, u, \nabla u) \partial_i w \right) \cdot \nabla v \\ + \sum_i \left( -\frac{\partial a^i}{\partial u}(x, u, \nabla u) + \frac{\partial b}{\partial u_{x_i}}(x, u, \nabla u) \right) (\partial_i w) v \\ + \left( -\nabla \cdot \frac{\partial \mathbf{a}}{\partial u}(x, u, \nabla u) + \frac{\partial b}{\partial u}(x, u, \nabla u) \right) w v \, d\mu$$

(generally, it is easier to re-derive linearizations for specific nonlinear operators  $F$  than it is to remember this complicated general linearization formula).



### 3.3 Adding Time Dependence

Adding time dependence to a nonlinear equation also gives a similar situation.

The general setup is, for  $F$  an elliptic operator,

$$\frac{\partial u}{\partial t} = -F(u) + f.$$

for a source term  $f$  and a quasilinear elliptic operator  $F$  (note the use of the  $-$  is to be consistent with the fact that  $-\Delta$  is the positive elliptic operator, and the heat equation has a  $\Delta$ , not a  $-\Delta$  on the RHS). Choosing a time-independent basis  $\varphi_j$ , we use the method of separation of variables detailed before, in the linear case, to derive time-dependent coefficients,  $u^i$ : we assume we have a discretized solution  $u(x, t) = u^i(t)\varphi_i(x)$ , and integrate against another basis element as a test function:

$$\int_M \frac{du^i}{dt} \varphi_i \varphi_j dx = - \int_M \mathbf{a}(x, u^i \varphi_i, u^i \nabla \varphi_i) \cdot \nabla \varphi_j + b(x, u^i \varphi_i, u^i \nabla \varphi_i) \varphi_j d\mu + \int_M f \varphi_i d\mu,$$

which gives, using the abbreviations  $\mathbf{F}$ ,  $\mathbf{f}$ , etc., in the previous section, and the mass matrix  $M$  as before, we have

$$M\dot{\mathbf{u}} = -\mathbf{F}(\mathbf{u}) + \mathbf{f}.$$

**3.3.1 Example** (Discretization in time using backward Euler). We now discretize in time, using the backward Euler method. Writing  $\dot{\mathbf{u}} = \frac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\Delta t}$ , and expressing the spatial part using the future time  $\mathbf{u}^{k+1}$  we have the following equation for  $\mathbf{u}^{k+1}$ :

$$M(\mathbf{u}^{k+1} - \mathbf{u}^k) = \Delta t(\mathbf{f} - \mathbf{F}(\mathbf{u}^{k+1}))$$

which again is a nonlinear equation. We wish to solve for  $\mathbf{u}^{k+1}$  explicitly in terms of  $\mathbf{u}^k$ .

This again requires the assistance of Newton's method: we rewrite it as

$$M\mathbf{u}^{k+1} + \Delta t\mathbf{F}(\mathbf{u}^{k+1}) = M\mathbf{u}^k + \Delta t\mathbf{f}$$

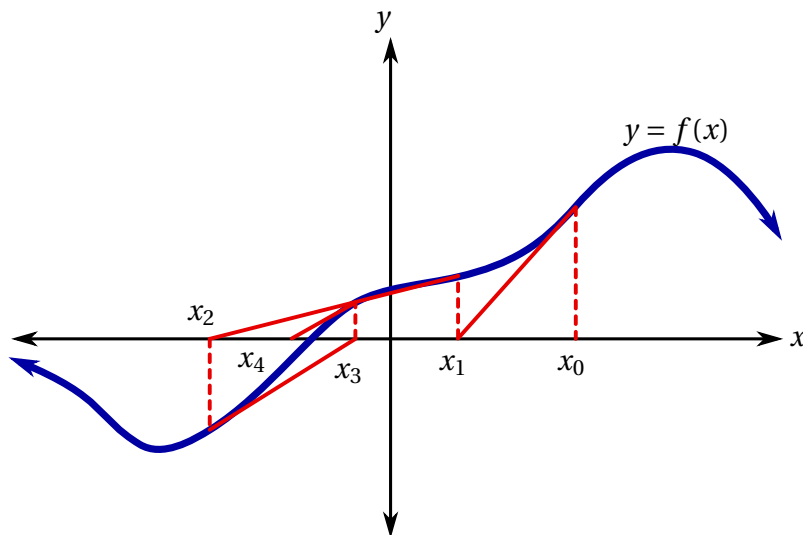
This is the setup for Newton's method. We start with an initial guess  $\mathbf{u}_0^{k+1}$ , which may reasonably be set to  $\mathbf{u}^k$ , and iterate:

$$\mathbf{u}_{n+1}^{k+1} = \mathbf{u}_n^{k+1} - (M + \Delta t\mathbf{DF}(\mathbf{u}_n^{k+1}))^{-1}(M(\mathbf{u}_n^{k+1} - \mathbf{u}^k) + \Delta t(\mathbf{F}(\mathbf{u}_n^{k+1}) - \mathbf{f})).$$

### 3.4 Newton's Method

The general solution of nonlinear problems via Newton's Method is so useful that we should devote a separate section to it, and prove some general theorems that will help us. Much of this material can be found in [53] and [97]. The general setup is as follows. Let  $F : U \subseteq \mathfrak{X} \rightarrow \mathfrak{X}$  be a mapping, where  $U$  is an open subset of a Banach space  $\mathfrak{X}$ . We would like to find  $u$  such that  $F(u) = 0$ . This incurs no loss of generality from before, where we solved  $F(u) = f$ , because we simply then define a new mapping  $G(u) = F(u) - f$  and solve  $G(u) = 0$  instead. The classical motivation is as follows. We start with a guess, that is, any point  $u_0 \in U$ , and, upon realizing that  $F(u_0)$  is not zero, we attempt to "correct"  $u_0$  by adding a term  $h$ : Find  $h$  such that linearize:  $F(u_0 + h) = 0$ . This, of course, is as hard as the original problem—all we've done was translate to a different point in space. However, linearizing about  $u_0$ , (and assuming  $F$  is Gâteaux differentiable in the sense of calculus in Banach spaces [17]):

$$F(u_0 + h) \approx F(u_0) + F'(u_0)h.$$



**Figure 3.1:** Graphical illustration of for Newton's Method on a function  $f$  (the graph  $y = f(x)$  is in blue). At each  $x_i$  on the  $x$ -axis, draw a vertical line (dashed red in the above) to the point  $(x_i, f(x_i))$ . From that point, draw a tangent line (in red). Then  $x_{i+1}$  is the intersection of the tangent line with the  $x$ -axis, which hopefully is closer to an actual intersection (i.e., root) of  $y = f(x)$  with the  $x$ -axis.

We set this linearization to 0, in order to solve for  $h$ :

$$F(u_0) + F'(u_0)h = 0,$$

which gives  $h = -F'(u_0)^{-1}F(u_0)$ . Thus defining  $u_1 = u_0 + h = u_0 - F'(u_0)^{-1}F(u_0)$ , this yields a result that hopefully makes  $F(u_1)$  closer to zero. For  $\mathfrak{X} = \mathbb{R}$ , this is drawing a tangent line to the graph of  $F$  at  $u_0$ , and finding out where it meets the  $x$ -axis—if  $F$  is sufficiently well-behaved, then  $F$  behaves much like its linearization, so the tangent line is not too far off when hitting the  $x$ -axis.

Of course,  $F'(u_0)$  may fail to be invertible (for  $\mathfrak{X} = \mathbb{R}$ , the tangent can be horizontal), which forces us to have to choose a new guess.

If  $F(u_1)$  is in fact zero, we are done. Otherwise,  $u_1$  can serve as a new guess; we try again: find  $h$  such that  $F(u_1 + h)$  is zero, or at least its approximation: solve  $F(u_1) + F'(u_1)h = 0$  for  $h$ , and define  $u_2 = u_1 + h$ . Continuing, we construct the sequence of

approximations

$$u_{n+1} = u_n - F'(u_n)^{-1}F(u_n),$$

with an arbitrary choice of  $u_0 \in U$ . It need not be completely random—for example, we may have some rough idea or intuitive sense of where a root should be, thus allowing us to make an informed guess. For standard ODE solvers such as Runge-Kutta methods, for example, the natural start point is the result at the current timestep (or the initial condition). What we desire, of course, is that this sequence actually converge to a solution. Intuitively, since  $h$  “corrects” the guess  $u_n$  by linearizing and solving,  $u_n + h$  is closer to the true root. Then, linearizing at  $u_n + h$  is likely to give an even better linear approximation to  $F$  close to the root. This “virtuous cycle” should allow us to hone in on the solution very quickly.

Many things can go wrong, however; for example, some  $F'(u_n)$  fails to be invertible, the sequence never converges, the sequence converges to something completely different, etc. The trouble, at least in 1 dimension, occurs when there are oscillations ( $x_1$  and  $x_2$  in Figure 3.1 have a larger gap than  $x_0$  to  $x_1$ , with oscillations in  $f$  there), because the slope can change sign or reduce drastically in magnitude. This bad local behavior has global significance, because we are extending the tangent line *as far as necessary* for an intersection. It would be useful to have a few theorems for guidance. We are interested in some theorems that give a guarantee that the sequence converges, and not only that, converges nearby, and quickly. Very little is known about the global behavior of Newton’s method, and in fact, partitioning the domain into different regions, according to which root a point starting in the region converges, yields complicated, fractal sets [55, §6.1], [79], thus showing that there is no neat, clear-cut test to find where a given starting point will converge.

### 3.4.1 Kantorovitch's Theorem

One of the reason Newton's method is very well-liked is that we can get it to SUPERCONVERGE, namely, have it converge so quickly that, roughly, the number of accurate digits doubles with each iteration. This overwhelms the precision of computers very quickly. In this section, we describe a sufficient criterion for superconvergence. This is especially good for numerical approximations to ODES because the operators approach the identity as the timesteps get smaller, leading to a very well-conditioned problem for Newton's method—the error that is the result of stopping the Newton iteration at finitely many steps becomes an insignificant contributor to the total error in the problem. This theorem can be found in [97, Theorem 10.7.1] and (in the finite-dimensional case, along with its proof) [53, §2.9 and §A.5].

**3.4.1 Theorem** (Kantorovitch's Theorem). *Let  $F : U \rightarrow \mathfrak{X}$  be a  $C^1$  mapping of Banach spaces. Suppose that there exists  $u_0 \in U$  such that  $F'(u_0)$  is invertible. Define  $h_0 = -F'(u_0)^{-1}F(u_0)$  and  $u_1 = u_0 + h_0$ . Suppose that in  $U_0 = B_{\|h_0\|}(u_1)$ ,  $F'$  satisfies a Lipschitz condition*

$$\|F'(x) - F'(y)\| \leq M\|x - y\|.$$

*for  $x$  and  $y$  in  $U_0$ . Finally, suppose that the following holds at  $u_0$ :*

$$\|F'(u_0)^{-1}\|^2 \|F(u_0)\| M = k \leq \frac{1}{2}.$$

*Then Newton's Method, starting at  $u_0$ , converges to a solution  $u$ , i.e.  $F(u) = 0$ . Moreover,  $u$  is the unique solution in  $U_0$ . If, moreover, strict inequality holds, that is,  $k < 1/2$ , then defining*

$$c = \frac{1 - k}{1 - 2k} \frac{M}{2} \|F'(u_0)^{-1}\|,$$

if at some point,  $\|u_{n+1} - u_n\| \leq \frac{1}{2c}$ ,

$$\|u_{n+m+1} - u_{n+m}\| \leq \frac{1}{c} \left(\frac{1}{2}\right)^{2^m},$$

that is to say, the distance between successive iterates shrinks hyper-exponentially.

Practically, this means that once we're near the solution, the convergence is extremely fast. In terms of decimal or binary expansions, this says that the number of correct digits roughly *doubles* with each iteration.

### 3.4.2 Globalizing Newton's Method

As mentioned before, little is known about the global behavior of Newton's Method. However, we should say what little we *do* know. Much of this follows the discussion in [97, §10.7]. One method we can use is that of *damping*. Many problems of the form  $F(u) = 0$  for  $F : U \subseteq \mathfrak{X} \rightarrow X$  can be recast as a minimization of some functional,  $J : U \rightarrow \mathbb{R}$ . For differential equations, for example, we have Euler-Lagrange equations. If  $\mathfrak{X}$  is a Hilbert space, we can always construct a functional  $J(u) = \frac{1}{2} \|F(u)\|_{\mathfrak{X}}^2$ . The key concept is that  $J$  hits its minimum, 0, if and only if  $F$  vanishes. If  $F$  has a unique solution  $u$  (or at least it has a unique solution in some neighborhood  $U_0 \subseteq U$ ), then  $J$  has a unique global minimum at  $u$  (or minimum in  $U_0$ ). Of course, minimization of functions is its own highly nontrivial problem, so it is not clear we gain anything at all by switching our viewpoint to minimizing  $J$  instead of finding a root of  $F$ . However,  $J$  can be used to improve the robustness of Newton's method. The concept is very simple: if the next iterate of Newton's method is a better approximation of a root of  $F$ , then  $J$  should decrease. What could possibly interfere with  $J$  decreasing? For example, if  $J$  is sufficiently differentiable, and if the increment  $h_n = -F'(u_n)^{-1}F(u_n)$  is too large, then quadratic terms in a Taylor expansion of  $J$  at  $u_n$  can dominate the local behavior,

swamping any decrease.

However, all is not lost in such cases. What we can show is that there exists  $\lambda \in [0, 1]$  such that for all  $\alpha \in (0, \lambda)$ ,  $J(u + \alpha h) < J(u)$  whenever  $h = -F'(u)^{-1}F(u)$ : we can guarantee that  $J$  descends as we move in the direction of the increment, but only in a sufficiently small neighborhood of  $u$ . This is why this is called *damping*: we still move in the direction dictated by Newton's Method, but possibly not as much. How do we convert this into an algorithm? We simply set  $h_n = -F'(u_n)^{-1}F(u_n)$  as before, and see if  $J(u_n + h_n) < J(u_n)$ . If this holds, then the regular Newton iteration does indeed work, and we set  $u_{n+1} = u_n + h_n$ . Otherwise, we run another loop: we test  $J(u_n + \lambda_k h_n) < J(u_n)$  for some sequence  $\lambda_k$ , where  $\lambda_k$  decreases to 0 (typically  $2^{-k}$ ). The first time  $\ell$  such that the inequality holds, that is,  $J(u_n + \lambda_\ell h_n) < J(u_n)$  but  $J(u_n + \lambda_k h_n) \geq J(u_n)$  for all  $k < \ell$ , we say the Newton iteration is finished, setting  $u_{n+1} = u_n + \lambda_\ell h_n$ . By the descent guarantee, each loop is guaranteed to terminate, since a sequence decreasing to 0 must eventually make it through the neighborhood  $(0, \lambda)$ .

How do we prove the descent guarantee? We simply show that the directional derivative

$$J'(u)h = \left. \frac{d}{d\alpha} \right|_{\alpha=0} J(u + \alpha h) < 0$$

for  $h = -F'(u)^{-1}F(u)$ . Since  $J$  is  $C^1$ , so is the one-variable function  $f(\alpha) = J(u + \alpha h)$ , and since  $f'(0) = J'(u)h < 0$ , it is  $< 0$  in a whole neighborhood of 0. Thus  $f$  must actually be decreasing in this neighborhood, that is  $f(\alpha) = J(u + \alpha h)$  is decreasing for  $\alpha$  close enough to 0. Alternatively, one could speak of this in terms of Taylor series:

$$f(\alpha) = f(0) + f'(0)\alpha + O(\alpha^2) = J(u) + J'(u)h\alpha + O(\alpha^2).$$

Thus, close to 0, the linear term dominates (and is decreasing). For  $J(u) = \frac{1}{2}\|F(u)\|_{\mathbb{X}}^2$ ,

we have

$$\frac{d}{d\alpha} J(u + \alpha h) = J'(u)h = \frac{1}{2} 2(F(u), F'(u)h)_{\mathfrak{X}} = (F(u), F'(u)h)_{\mathfrak{X}}.$$

Now if  $h = -F'(u)^{-1}F(u)$ , then

$$J'(u)h = (F(u), -F'(u)F'(u)^{-1}F(u))_{\mathfrak{X}} = -\|F(u)\|_{\mathfrak{X}}^2 \leq 0.$$

If  $-\|F(u)\|^2 = 0$ , then we are actually done, for this means that  $F(u) = 0$ . On the other hand, if it is strictly less than 0, this proves the descent guarantee.

How good is this method? It guarantees descent in  $\|F(u)\|$ , and if  $J$  has some nice properties such as convexity and properness, it is easy to show that  $u_n$  converges. Since  $J$  is bounded below by 0, and the sequence  $J(u_n)$  is decreasing by construction, this sequence must converge. If  $u_n \rightarrow u$ , and  $F'(u)$  is invertible, then the sequence  $F'(u_n)^{-1}$  is invertible and hence  $-F'(u_n)^{-1}F(u_n)$  converges.



## **Part II**

# **Applications to Evolution Problems**

## Chapter 4

# Approximation of Parabolic Equations in Hilbert Complexes

This chapter is in preparation as a separate published article (joint work with Michael Holst), and therefore may depart from some conventions established earlier, and some material may be duplicated. We prove our main results in this chapter.

### 4.0 Abstract

Arnold, Falk, and Winther [5, 6] introduced the Finite Element Exterior Calculus (FEEC) as a general framework for linear mixed variational problems, their numerical approximation by mixed methods, and their error analysis. They recast these problems using the ideas and tools of *Hilbert complexes*, leading to a more complete understanding. Subsequently, Holst and Stern [50] extended the Arnold–Falk–Winther framework to include *variational crimes*, allowing for the analysis and numerical approximation of linear and geometric elliptic partial differential equations on Riemannian manifolds of arbitrary spatial dimension, generalizing the existing sur-

face finite element approximation theory in several directions. Gillette and Holst [40] extended the FEEC in another direction, namely to parabolic and hyperbolic evolution systems by combining recent work on the FEEC for elliptic problems with a classical approach of Thomée [106] to solving evolution problems using semi-discrete finite element methods, by viewing solutions to the evolution problem as lying in Bochner spaces (spaces of Banach-space valued parametrized curves). Arnold and Chen [4] independently developed related work, for generalized Hodge Laplacian parabolic problems for differential forms of arbitrary degree. In this article, we aim to combine the approaches of the above articles, extending the work of Gillette and Holst [40] and Arnold and Chen [4] to parabolic evolution problems on Riemannian manifolds by using the framework of Holst and Stern [50].

## 4.1 Introduction

Before introducing the abstract framework, we motivate the continuous problem concretely by considering an evolution equation for differential forms on a manifold; then we rephrase it as a mixed problem as an intermediate step toward semidiscretization using the finite element method. We then see how this allows us to leverage existing *a priori* error estimates for parabolic problems, and see how it fits in the framework of Hilbert complexes.

**4.1.1 The Hodge heat equation and its mixed form.** Let  $M$  be a compact oriented Riemannian  $n$ -manifold embedded in  $\mathbb{R}^{n+1}$ . The HODGE HEAT EQUATION is to find time-dependent  $k$ -form  $u: M \times [0, T] \rightarrow \Lambda^k(M)$  such that

$$(4.1.1) \quad \begin{aligned} u_t - \Delta u &= u_t + (\delta d + d\delta)u = f && \text{in } M, \quad \text{for } t > 0 \\ u(\cdot, 0) &= g && \text{in } M. \end{aligned}$$

where  $g$  is an initial  $k$ -form, and  $f$ , a possibly time-dependent  $k$ -form, is a source term. Note that no boundary conditions are needed for manifolds without boundary. This is the problem studied by Arnold and Chen [4], and in the case  $k = n$ , one of the problems studied by Gillette and Holst [40], building upon work in special cases for domains in  $\mathbb{R}^2$  and  $\mathbb{R}^3$  by Johnson and Thomée [57, 106].

For the stability of the numerical approximations with the methods of [51] and [6], we recast the problem in mixed form, converting the problem into a system of differential equations. Motivating the problem by setting  $\sigma = \delta u$  (recall that for the Dirichlet problem and  $k = n$ ,  $\delta$  here corresponds to the gradient in Euclidean space, and is the adjoint  $d$ , corresponding to the *negative* divergence), and taking the adjoint, we have

$$\begin{aligned}
 \langle \sigma, \omega \rangle - \langle u, d\omega \rangle &= 0, & \forall \omega \in H\Omega^{k-1}(M), \quad t > 0, \\
 \langle u_t, \varphi \rangle + \langle d\sigma, \varphi \rangle + \langle du, d\varphi \rangle &= \langle f, \varphi \rangle, & \forall \varphi \in H\Omega^k(M) \quad t > 0. \\
 u(0) &= g.
 \end{aligned}
 \tag{4.1.2}$$

Unlike the elliptic case, we do not have to explicitly account for harmonic forms in the formulation of the equations themselves, but they will definitely play a critical role in our analysis and bring new results not apparent in the  $k = n$  case.

**4.1.2 Semidiscretization of the equation.** In order to analyze the numerical approximation, we semidiscretize our problem in space. In our case, we shall assume, following [50], that we have a family of approximating surfaces  $M_h$  to the hypersurface  $M$ , given as the zero level set of some signed distance function, all contained in a tubular neighborhood  $U$  of  $M$ , and a projection  $a : M_h \rightarrow M$  along the surface normal (of  $M$ ). The surfaces may be a triangulations, i.e., piecewise linear (studied by Dziuk and Demlow in [27, 25]), or piecewise polynomial (obtained by Lagrange interpolation over a triangulation of the projection  $a$ , as later studied by Demlow in [24]). We pull forms

on  $M_h$  to  $M$  back via the inverse of the normal projection, which furnishes injective morphisms  $i_h^k : \Lambda_h^k \hookrightarrow H\Omega^k(M)$  as required by the theory in [50], which we shall review in Section 4.2 below. Finally, we need a family of linear projections  $\Pi_h^k : H\Omega^k(M) \rightarrow \Lambda_h^k$  such that  $\Pi_h \circ i_h = \text{id}$  which allow us to interpolate given data into the chosen finite element spaces—this is necessary because some of the more obvious, natural seeming choices of operators, such as  $i_h^*$ , can be difficult to compute (nevertheless,  $i_h^*$  will still be useful theoretically).

We now can formulate the semidiscrete problem: we seek a solution  $(\sigma_h, u_h) \in H_h \times S_h \subseteq H\Omega^{k-1} \times H\Omega^k$  such that

(4.1.3)

$$\begin{aligned} \langle \sigma_h, \omega_h \rangle_h - \langle u_h, d\omega_h \rangle_h &= 0, & \forall \omega_h \in H_h, \quad t > 0 \\ \langle u_{h,t}, \varphi_h \rangle_h + \langle d\sigma_h, \varphi_h \rangle_h + \langle du_h, d\varphi_h \rangle_h &= \langle \Pi_h f, \varphi_h \rangle_h, & \forall \varphi_h \in S_h \quad t > 0 \\ u_h(0) &= g_h. \end{aligned}$$

We shall describe how to define  $g_h \in S_h$  shortly; it is to be some suitable interpolation of  $g$ . As  $S_h$  and  $H_h$  are finite-dimensional spaces, we can reduce this to a system of ODEs in Euclidean space by choosing bases  $(\psi_i)$  for  $S_h$  and  $(\phi_k)$  for  $H_h$ ; expanding the unknowns  $\sigma_h = \sum_i \Sigma^i(t) \psi_i$  and  $u_h = \sum_k U^k(t) \phi_k$ ; substituting these basis functions as test functions to form matrices  $A_{k\ell} = \langle \phi_k, \phi_\ell \rangle$ ,  $B_{ik} = \langle d\psi_i, \phi_k \rangle$ , and  $D_{ij} = \langle \psi_i, \psi_j \rangle$ ; and finally forming the vectors for the load data  $F$  defined by  $F_k = \langle F, \phi_k \rangle$ , and initial condition  $G$  defined by  $g_h = \sum G^k \phi_k$ . We thus arrive at the matrix equations for the unknown, time-dependent coefficient vectors  $\Sigma$  and  $U$ :

$$\begin{aligned} D\Sigma - B^T U &= 0, \\ AU_t + B\Sigma + KU &= F, \text{ for } t > 0 \\ U(0) &= G. \end{aligned}$$

The matrices  $A$  and  $D$  are positive definite, hence invertible. Substituting  $\Sigma = D^{-1}B^T U$ , we have the system of ODEs

$$AU_t + (BD^{-1}B^T + K)U = F, \text{ for } t > 0, \quad U(0) = G,$$

which has a unique solution by the usual ODE theory. For purposes of actually numerically integrating the ODE, namely, discretizing fully in space and time, it is better not to use the above formulation, because it can lead to dense matrices. Computationally, this is due to the explicit presence of an inverse,  $D^{-1}$ , not directly multiplying the variable; conceptually, this is actually a statement about the discrete adjoint to the codifferential  $d_h^*$  generally having global support even if the finite element functions are only locally supported [4]. Instead, we differentiate the first equation with respect to time, getting  $D\Sigma_t - B^T U_t = 0$ , which leads to the block system

$$(4.1.4) \quad \frac{d}{dt} \begin{pmatrix} D & -B^T \\ 0 & A \end{pmatrix} \begin{pmatrix} \Sigma \\ U \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -B & -K \end{pmatrix} \begin{pmatrix} \Sigma \\ U \end{pmatrix} + \begin{pmatrix} 0 \\ F \end{pmatrix}$$

which is still well-defined ODE for  $\Sigma$  and  $U$ , as the invertible matrices  $A$  and  $D$  appear on the diagonal. This differentiated equation also plays a role in the showing that the continuous problem is well-posed.

These equations only differ from those studied by Gillette and Holst [40], Arnold and Chen [4], and Thomée [106] by the choice of finite element spaces—here we are assuming them to be in some Sobolev space of differential forms on manifolds (or in a triangulated mesh in a tubular neighborhood) rather than subsets of Euclidean space. This suggests that we should try to gather these commonalities, examine what happens in abstract Hilbert complexes, and see how general a form of error estimate we can get this way.

**4.1.3 Error analysis.** The general idea of the method of Thomée [106] is to compare the semidiscrete solution to an ELLIPTIC PROJECTION of the data, a method first explored by Wheeler [110]. If we assume that we already have a solution  $u$  to the continuous problem, then for each fixed time  $t$ ,  $u(t)$  can be considered as trivially solving an elliptic equation with data  $-\Delta u(t)$ . Thus, using the methods developed in [6], we consider the discrete solution  $\tilde{u}_h$  for  $u$  in this elliptic problem (namely, applying the discrete solution operator  $T_h$  to  $-\Delta u(t)$ ). This may be compared to the true solution (at each fixed time) using the error estimates in [6]. What remains is to compare the semidiscrete solution  $u_h$  (as defined by the ODEs (4.1.3) above) to the elliptic projection, so that we have the full error estimate by the triangle inequality. Thomée derives the following estimates, for finite elements in the plane ( $n = 2$ ) of top-degree forms ( $k = 2$ , there represented by a scalar proxy), for  $g_h$  the elliptic projection of the initial condition  $g$  and  $t \geq 0$ :

$$(4.1.5) \quad \|u_h(t) - u(t)\|_{L^2} \leq ch^2 \left( \|u(t)\|_{H^2} + \int_0^t \|u_t(s)\|_{H^2} ds \right),$$

$$(4.1.6) \quad \|\sigma_h(t) - \sigma(t)\|_{L^2} \leq ch^2 \left( \|u(t)\|_{H^3} + \left( \int_0^t \|u_t(s)\|_{H^2}^2 ds \right)^{1/2} \right).$$

Gillette and Holst [40], and Arnold and Chen [4] generalize these estimates and represent them in terms of Bochner norms. These estimates describe the accumulation of error up to fixed time value  $t$ , assuming, of course, that the spaces finite elements are sufficiently regular to allow those estimates. The key equation that makes these estimates possible are Thomée's error evolution equations: defining  $\rho = \|\tilde{u}_h(t) - u(t)\|$ ,  $\theta = \|u_h(t) - \tilde{u}_h(t)\|$ , and  $\varepsilon = \|\sigma_h(t) - \tilde{\sigma}_h(t)\|$ , we have

$$\langle \theta_t, \phi_h \rangle - \langle \operatorname{div} \varepsilon(t), \phi_h \rangle = -\langle \rho_t, \phi_h \rangle$$

$$\langle \varepsilon, \omega_h \rangle + \langle \theta, \operatorname{div} \omega_h \rangle = 0.$$

These are used to derive certain differential inequalities and make Grönwall-type estimates. In this chapter, we examine the above error equations and place them in a more abstract framework. We use Bochner spaces (also used by [40]) to describe time evolution in Hilbert complexes, building on their successful use in elliptic problems. We investigate Thomée’s method in this framework to gain further insight into how finite element error estimates evolve in time.

**4.1.4 Summary of the chapter.** The remainder of this chapter is structured as follows. In Section 4.2, we review the finite element exterior calculus (FEEC) and the variational crimes framework of Holst and Stern [50]. We prove some extensions in order to account for problems with prescribed harmonic forms; this is what allows the elliptic projection to work in the case where harmonic forms are present. In Section 4.3, we formulate abstract parabolic problems in Bochner spaces and extend some standard results on the existence and uniqueness of strong solutions, and describe how this problem fits into that framework. In Section 4.4, we extend the *a priori* error estimates for Galerkin mixed finite element methods to parabolic problems on Hilbert complexes. Then, we relate the results to the problem on manifolds. The main abstract result is Theorem 4.4.4, which uses the previous results from the FEEC framework with variational crimes, in order to understand how those error terms evolve with time. We then specialize, in Section 4.5 to parabolic equations on Riemannian manifolds, our original motivating example, and see how this generalizes the error estimates of Thomée [106], Gillette and Holst [40], and Holst and Stern [50]. In Section 4.6, we present a numerical experiment comparing the methods based on this mixed form to more straightforward implementations in the scalar heat equation case.



## 4.2 The Finite Element Exterior Calculus

We review here the relevant results from the finite element exterior calculus (FEEC) that we will need for this paper. FEEC was introduced in Arnold, Falk and Winther [5, 6] as a framework for deriving error estimates and formulating stable numerical methods for a large class of elliptic PDE. One of the central ideas which helped unify many of these distinct methods into a structured framework has been the idea of HILBERT COMPLEXES [14], which abstracts the essential features of the cochain complexes commonly found in exterior calculus and places them in a context where modern methods of functional analysis may be applied. This assists in formulating and solving boundary value problems, in direct analogy to how Sobolev spaces have helped provide a framework for solving such problems for functions. Arnold, Falk, and Winther [6] place numerical methods into this framework by choosing certain finite-dimensional subspaces satisfying certain compatibility and approximation properties. Holst and Stern [50] extended this framework by considering the case in which there is an injective morphism from a finite-dimensional complex to the complex of interest, without it necessarily being inclusion. This allows the treatment of geometric VARIATIONAL CRIMES [10, 13], where an approximating manifold (on which it may be far easier to choose finite element spaces) no longer coincides with the actual manifold on which we seek our solution. We review the theory as detailed in [50] and refer the reader there for details.

### 4.2.1 Hilbert Complexes

As stated before, the essential details of differential complexes, such as the de Rham complex, are nicely captured in the notion of Hilbert complexes. This enables us to see clearly where many elements of boundary value problems come from, in

particular, the Laplacian, Hodge decomposition theorem, and Poincaré inequality. In addition, it allows us to see how to carry these notions over to numerical approximations.

**4.2.1 Definition** (Hilbert complexes). We define a HILBERT COMPLEX  $(W, d)$  to be sequence of Hilbert spaces  $W^k$  with possibly unbounded linear maps  $d^k : V^k \subseteq W^k \rightarrow V^{k+1} \subseteq W^{k+1}$ , such that each  $d^k$  has closed graph, densely defined, and satisfies the COCHAIN PROPERTY  $d^k \circ d^{k-1} = 0$  (this is often abbreviated  $d^2 = 0$ ; we often omit the superscripts when the context is clear). We call each  $V^k$  the DOMAIN of  $d^k$ . We will often refer to elements of such Hilbert spaces as “forms,” being motivated by the canonical example of the de Rham complex. The Hilbert complex is called a CLOSED COMPLEX if each image space  $\mathfrak{B}^k = d^{k-1}V^{k-1}$  (called the  $k$ -COBOUNDARIES is closed in  $W^k$ , and a BOUNDED COMPLEX if each  $d^k$  is in fact a bounded linear map. The most common arrangement in which one finds a bounded complex is by taking the sequence of domains  $V^k$ , the same maps  $d^k$ , but now with the GRAPH INNER PRODUCT

$$\langle v, w \rangle_V = \langle v, w \rangle + \langle d^k v, d^k w \rangle.$$

for all  $v, w \in V^k$ . Unsubscripted inner products and norms will always be assumed to be the ones associated to  $W^k$ .

**4.2.2 Definition** (Cocycles, Coboundaries, and Cohomology). The kernel of the map  $d^k$  in  $V^k$  will be called  $\mathfrak{Z}^k$ , the  $k$ -COCYCLES and, as before, we have  $\mathfrak{B}^k = d^{k-1}V^{k-1}$ . Since  $d^k \circ d^{k-1} = 0$ , we have  $\mathfrak{B}^k \subseteq \mathfrak{Z}^k$ , so we have the  $k$ -COHOMOLOGY  $\mathfrak{Z}^k/\mathfrak{B}^k$ . The HARMONIC SPACE  $\mathfrak{H}^k$  is the orthogonal complement of  $\mathfrak{B}^k$  in  $\mathfrak{Z}^k$ . This means, in general, we have an orthogonal decomposition  $\mathfrak{Z}^k = \overline{\mathfrak{B}^k} \oplus \mathfrak{H}^k$ , and we have that  $\mathfrak{H}^k$  is isomorphic to  $\mathfrak{Z}^k/\overline{\mathfrak{B}^k}$ , the REDUCED COHOMOLOGY, which of course corresponds to the usual cohomology for closed complexes.

**4.2.3 Definition** (Dual complexes and adjoints). For a Hilbert complex  $(W, d)$ , we can form the DUAL COMPLEX  $(W^*, d^*)$  which consists of spaces  $W_k^* = W^k$ , maps  $d_k^* : V_k^* \subseteq W_k^* \rightarrow V_{k-1}^* \subseteq W_{k-1}^*$  such that  $d_{k+1}^* = (d^k)^*$ , the adjoint operator, that is:

$$\langle d_{k+1}^* v, w \rangle = \langle v, d^k w \rangle.$$

The operators  $d^*$  decrease degree, so this is a chain complex, rather than a cochain complex; the analogous concepts to cocycles and coboundaries extend to this case and we write  $\mathfrak{Z}_k^*$  and  $\mathfrak{B}_k^*$  for them.

**4.2.4 Definition** (Morphisms of Hilbert complexes). Let  $(W, d)$  and  $(W', d')$  be two Hilbert complexes.  $f : W \rightarrow W'$  is called a MORPHISM OF HILBERT COMPLEXES if we have a sequence of bounded linear maps  $f^k : W^k \rightarrow W'^k$  such that  $d'^k \circ f^k = f^{k+1} \circ d^k$  (they commute with the differentials).

With the above, we can show the following WEAK HODGE DECOMPOSITION:

**4.2.5 Theorem** (Hodge Decomposition Theorem). *Let  $(W, d)$  be a Hilbert complex with domain complex  $(V, d)$ . Then we have the  $W$ - and  $V$ -orthogonal decompositions*

$$(4.2.1) \quad W^k = \overline{\mathfrak{B}^k} \oplus \mathfrak{H}^k \oplus \mathfrak{Z}^{k \perp W}$$

$$(4.2.2) \quad V^k = \overline{\mathfrak{B}^k} \oplus \mathfrak{H}^k \oplus \mathfrak{Z}^{k \perp V}.$$

where  $\mathfrak{Z}^{k \perp V} = \mathfrak{Z}^{k \perp W} \cap V^k$ .

Of course, if  $\mathfrak{B}^k$  is closed, then the extra closure is unnecessary, and we omit the term “weak”. We shall simply write  $\mathfrak{Z}^{k \perp}$  for  $\mathfrak{Z}^{k \perp V}$ , which is will be the most useful orthogonal complement for our purposes. The orthogonal projections  $P_U$  for a subspace  $U$  will be in the  $W$ -inner product unless otherwise stated (although again,

due to the two inner products coinciding on  $\mathfrak{Z}^k$  and its subspaces, they may be the same). We note that by the abstract properties of adjoints ([6, §3.1.2]),  $\mathfrak{Z}^{k\perp W} = \overline{\mathfrak{B}_k^*}$ , and  $\mathfrak{B}^{k\perp W} = \mathfrak{Z}_k^*$ . Also very useful is that the  $V$ - and  $W$ -norms agree on  $\mathfrak{Z}$  and hence on  $\mathfrak{B}$  and  $\mathfrak{H}$ .

The following inequality is an important result crucial to the stability of our solutions to the boundary value problems as well as the numerical approximations:

**4.2.6 Theorem** (Abstract Poincaré Inequality). *If  $(V, d)$  is a closed, bounded Hilbert complex, then there exists a constant  $c_P > 0$  such that for all  $v \in \mathfrak{Z}^{k\perp}$ ,*

$$\|v\|_V \leq c_P \|d^k v\|_V.$$

In the case that  $(V, d)$  is the domain complex associated to a closed Hilbert complex  $(W, d)$ ,  $(V, d)$  is again closed, and the additional graph inner product term vanishes:  $\|d^k v\|_V = \|d^k v\|$ . We now introduce the abstract version of the Hodge Laplacian and the associated problem.

**4.2.7 Definition** (Abstract Hodge Laplacian problems). We consider the operator  $L = dd^* + d^*d$  on a Hilbert complex  $(W, d)$ , called the **ABSTRACT HODGE LAPLACIAN**. Its domain is  $D_L = \{u \in V^k \cap V_k^* : du \in V_{k+1}^*, d^*u \in V^{k-1}\}$ , and the **HODGE LAPLACIAN PROBLEM** is to seek  $u \in V^k \cap V_k^*$ , given  $f \in W^k$ , such that

$$(4.2.3) \quad \langle du, dv \rangle + \langle d^*u, d^*v \rangle = \langle f, v \rangle$$

for all  $v \in V^k \cap V_k^*$ . This is simply the weak form of the Laplacian and any  $u \in V^k \cap V_k^*$  satisfying the above is called a **WEAK SOLUTION**. Owing to difficulties in the approximation theory for such a problem (it is difficult to construct finite elements for the space  $V^k \cap V_k^*$ ), Arnold, Falk, and Winther [6] formulated the **MIXED ABSTRACT HODGE**

LAPLACIAN PROBLEM by defining auxiliary variables  $\sigma = d^* u$  and  $p = P_{\mathfrak{H}} f$ , the orthogonal projection of  $f$  into the harmonic space, and considering a *system* of equations, to seek  $(\sigma, u, p) \in V^{k-1} \times V^k \times \mathfrak{H}^k$  such that

$$(4.2.4) \quad \begin{aligned} \langle \sigma, \tau \rangle - \langle u, d\tau \rangle &= 0 & \forall \tau \in V^{k-1} \\ \langle d\sigma, v \rangle + \langle du, dv \rangle + \langle p, v \rangle &= \langle f, v \rangle & \forall v \in V^k \\ \langle u, q \rangle &= 0 & \forall q \in \mathfrak{H}^k. \end{aligned}$$

The first equation is the weak form of  $\sigma = d^* u$ , the second is the main equation (4.2.3) modified to account for a harmonic term so that a solution exists, and the third enforces uniqueness by requiring perpendicularity to the harmonic space. With these modifications, the problem is well-posed by considering the bilinear form (writing  $\mathfrak{X}^k := V^{k-1} \times V^k \times \mathfrak{H}^k$ )  $B : \mathfrak{X}^k \times \mathfrak{X}^k \rightarrow \mathbb{R}$  defined by

$$(4.2.5) \quad B(\sigma, u, p; \tau, v, q) := \langle \sigma, \tau \rangle - \langle d\tau, u \rangle + \langle d\sigma, v \rangle + \langle du, dv \rangle + \langle p, v \rangle - \langle u, q \rangle.$$

and linear functional  $F \in (\mathfrak{X}^k)^*$  given by  $F(\tau, v, q) = \langle f, v \rangle$ . The form  $B$  is *not* coercive, but rather, for a closed Hilbert complex, satisfies an INF-SUP CONDITION [6, 7]: there exists  $\gamma > 0$  (the STABILITY CONSTANT) such that

$$\inf_{(\sigma, u, p) \neq 0} \sup_{(\tau, v, q) \neq 0} \frac{B(\sigma, u, p; \tau, v, q)}{\|(\sigma, u, p)\|_{\mathfrak{X}} \|(\tau, v, q)\|_{\mathfrak{X}}} =: \gamma > 0.$$

where we have defined a standard norm on products:  $\|(\sigma, u, p)\|_{\mathfrak{X}} := \|\sigma\|_V + \|u\|_V + \|p\|$ .

This is sufficient to guarantee the well-posedness [7]. To summarize:

**4.2.8 Theorem** (Arnold, Falk, and Winther [6], Theorem 3.1). *The mixed variational problem (4.2.4) on a closed Hilbert complex  $(W, d)$  with domain  $(V, d)$  is well-posed: the bilinear form  $B$  satisfies the inf-sup condition with constant  $\gamma$ , so for any  $F \in (\mathfrak{X}^k)^*$ ,*

there exists a unique solution  $(\sigma, u, p)$  to (4.2.4), i.e.,  $B(\sigma, u, p; \tau, v, q) = F(\tau, v, q)$  for all  $(\tau, v, q) \in \mathfrak{X}^k$ , and moreover,

$$\|(\sigma, u, p)\|_{\mathfrak{X}} \leq \gamma^{-1} \|F\|_{\mathfrak{X}^*}.$$

The STABILITY CONSTANT  $\gamma^{-1}$  depends only on the Poincaré constant.

Note that the general theory ([7] and §1.6 above) guarantees a unique solution exists for *any* bounded linear functional  $F \in (\mathfrak{X}^k)^*$ , which in this case with product spaces, means that the problem is still well-posed when there are other nonzero linear functionals on the RHS of (4.2.4) besides  $\langle f, v \rangle$ . We shall need this result for parabolic problems, where we assume  $u$  has a harmonic part ( $P_{\mathfrak{H}} u \neq 0$ ).

## 4.2.2 Approximation of Hilbert Complexes

We now approximate solutions to the abstract mixed Hodge Laplacian problem. To do so, Arnold, Falk, and Winther [6] introduce finite-dimensional subspaces  $V_h \subseteq V$  of the domain complex, such that the inclusion  $i_h : V_h \hookrightarrow V$  is a morphism, i.e.  $dV_h^k \subseteq V_h^{k+1}$ . With the weak form (4.2.4), we formulate the Galerkin method by restricting to the subspaces:

$$(4.2.6) \quad \begin{aligned} \langle \sigma_h, \tau \rangle - \langle u_h, d\tau \rangle &= 0 & \forall \tau \in V_h^{k-1} \\ \langle d\sigma_h, v \rangle + \langle du_h, dv \rangle + \langle p_h, v \rangle &= \langle f, v \rangle & \forall v \in V_h^k \\ \langle u_h, q \rangle &= 0 & \forall q \in \mathfrak{H}_h^k. \end{aligned}$$

We abbreviate by setting  $\mathfrak{X}_h^k := V_h^{k-1} \times V_h^k \times \mathfrak{H}_h^k$ . We must also assume the existence of bounded, surjective, and idempotent (projection) morphisms  $\pi_h : V \rightarrow V_h$ . It is generally not the orthogonal projection, as that fails to commute with the differentials.

As a projection, it gives the following QUASI-OPTIMALITY result:

$$\|u - \pi_h u\|_V = \inf_{v \in V_h} \|(I - \pi_h)(u - v)\|_V \leq \|I - \pi_h\| \inf_{v \in V_h} \|u - v\|_V.$$

The problem (4.2.6) is then well-posed, with a Poincaré constant given by  $c_P \|\pi_h^k\|$ , where  $c_P$  is the Poincaré constant for the continuous problem. This guarantees all the previous abstract results apply to this case. With this, we have the following error estimate:

**4.2.9 Theorem** (Arnold, Falk, and Winther [6], Theorem 3.9). *Let  $(V_h, d)$  be a family of subcomplexes of the domain  $(V, d)$  of a closed Hilbert complex, parametrized by  $h$  and admitting uniformly  $V$ -bounded cochain projections  $\pi_h$ , and let  $(\sigma, u, p) \in \mathfrak{X}^k$  be the solution of the continuous problem and  $(\sigma_h, u_h, p_h) \in \mathfrak{X}_h^k$  be the corresponding discrete solution. Then the following error estimate holds:*

$$(4.2.7) \quad \begin{aligned} \|(\sigma - \sigma_h, u - u_h, p - p_h)\|_{\mathfrak{X}} &= \|\sigma - \sigma_h\|_V + \|u - u_h\|_V + \|p - p_h\| \\ &\leq C \left( \inf_{\tau \in V_h^{k-1}} \|\sigma - \tau\|_V + \inf_{v \in V_h^k} \|u - v\|_V + \inf_{q \in V_h^k} \|p - q\|_V + \mu \inf_{v \in V_h^k} \|P_{\mathfrak{B}} u - v\|_V \right) \end{aligned}$$

with  $\mu = \mu_h^k = \sup_{\substack{r \in \mathfrak{H}^k \\ \|r\|=1}} \|(I - \pi_h^k) r\|$ , the operator norm of  $I - \pi_h^k$  restricted to  $\mathfrak{H}^k$ .

**4.2.10 Corollary.** *If the  $V_h$  approximate  $V$ , that is, for all  $u \in V$ ,  $\inf_{v \in V_h} \|u - v\|_V \rightarrow 0$  as  $h \rightarrow 0$ , we have convergence of the approximations.*

In general, the harmonic spaces  $\mathfrak{H}^k$  and  $\mathfrak{H}_h^k$  do not coincide, but they are isomorphic under many circumstances we shall consider (namely, the spaces are isomorphic if for all harmonic forms  $q \in \mathfrak{H}^k$ , the error  $\|q - \pi_h q\|$  is at most the norm  $\|q\|$  itself [6, Theorem 3.4], and it *always* holds for the de Rham complex). For a quantitative estimate relating the two different kinds of harmonic forms, we have the following

**4.2.11 Theorem** ([6], Theorem 3.5). *Let  $(V, d)$  be a bounded, closed Hilbert complex,  $(V_h, d)$  a Hilbert subcomplex, and  $\pi_h$  a bounded cochain projection. Then*

$$(4.2.8) \quad \|(I - P_{\mathfrak{S}_h})q\|_V \leq \|(I - \pi_h^k)q\|_V, \forall q \in \mathfrak{S}^k$$

$$(4.2.9) \quad \|(I - P_{\mathfrak{S}})q\|_V \leq \|(I - \pi_h^k)P_{\mathfrak{S}}q\|_V, \forall q \in \mathfrak{S}_h^k.$$

For geometric problems, it is essential to remove the requirement that the approximating complex  $V_h$  actually be subspaces of  $V$ . This is motivated by the example of approximating planar domains with curved boundaries by piecewise-linear approximations, resulting in finite element spaces that lie in a different function space [10]. Holst and Stern [50] extend the Arnold, Falk, Winther [6] framework by supposing that  $i_h : V_h \hookrightarrow V$  is an injective morphism which is not necessarily inclusion; they also require projection morphisms  $\pi_h : V \rightarrow V_h$  with the property  $\pi_h \circ i_h = \text{id}$ , which replaces the idempotency requirement of the preceding case. To summarize, given  $(W, d)$  a Hilbert complex with domain  $(V, d)$ ,  $(W_h, d_h)$  another complex (whose inner product we denote  $\langle \cdot, \cdot \rangle_h$ ) with domain  $(V_h, d_h)$ , injective morphisms  $i_h : W_h \hookrightarrow W$ , and finally, projection morphisms  $\pi_h : V \rightarrow V_h$ . We then have the following generalized Galerkin problem:

$$(4.2.10) \quad \begin{aligned} \langle \sigma_h, \tau_h \rangle_h - \langle u_h, d_h \tau_h \rangle_h &= 0 & \forall \tau_h \in V_h^{k-1} \\ \langle d_h \sigma_h, v_h \rangle_h + \langle d_h u_h, d_h v_h \rangle_h + \langle p_h, v_h \rangle_h &= \langle f_h, v_h \rangle_h & \forall v_h \in V_h^k \\ \langle u_h, q_h \rangle_h &= 0 & \forall q_h \in \mathfrak{S}_h^k, \end{aligned}$$

where  $f_h$  is some interpolation of the given data  $f$  into the space  $W_h$  (we will discuss



various choices of this operator later). This gives us a bilinear form

$$(4.2.11) \quad B_h(\sigma_h, u_h, p_h; \tau_h, v_h, q_h) := \langle \sigma_h, \tau_h \rangle_h - \langle u_h, d_h \tau_h \rangle_h \\ + \langle d_h \sigma_h, v_h \rangle_h + \langle d_h u_h, d_h v_h \rangle_h + \langle p_h, v_h \rangle_h - \langle u_h, q_h \rangle_h.$$

This problem is well-posed, which again follows from the abstract theory as long as the complex is closed, and there is a corresponding Poincaré inequality:

**4.2.12 Theorem** (Holst and Stern [50], Theorem 3.5 and Corollary 3.6). *Let  $(V, d)$  and  $(V_h, d_h)$  be bounded closed Hilbert complexes, with morphisms  $i_h : V_h \hookrightarrow V$  and  $\pi_h : V \rightarrow V_h$  such that  $\pi_h \circ i_h = \text{id}$ . Then for all  $v_h \in \mathfrak{Z}_h^{k\perp}$ , we have*

$$\|v_h\|_{V_h} \leq c_P \left\| \pi_h^k \right\| \left\| i_h^{k+1} \right\| \|d_h v_h\|_{V_h},$$

where  $c_P$  is the Poincaré constant corresponding to the continuous problem. If  $(V, d)$  and  $(V_h, d_h)$  are the domain complexes of closed complexes  $(W, d)$  and  $(W_h, d_h)$ , then  $\|d_h v_h\|_{V_h}$  is simply  $\|d_h v_h\|_h$  (since it is the graph norm and  $d^2 = 0$ ).

In other words, the norm of the injective morphisms  $i_h$  also contributes to the stability constant for this discrete problem. Analysis of this method results in two additional error terms (along with now having to explicitly reference the injective morphisms  $i_h$  which may no longer be inclusions), due to the inner products in the space  $V_h$  no longer necessarily being the restriction of that in  $V$ : the need to approximate the data  $f$ , and the failure of the morphisms  $i_h$  to be unitary:

**4.2.13 Theorem** (Holst and Stern [50], Corollary 3.11). *Let  $(V, d)$  be the domain complex of a closed Hilbert complex  $(W, d)$ , and  $(V_h, d_h)$  the domain complex of  $(W_h, d_h)$  with morphisms  $i_h : W_h \rightarrow W$  and  $\pi_h : V \rightarrow V_h$  as above. Then if we have a solutions  $(\sigma, u, p)$  and  $(\sigma_h, u_h, p_h)$  to (4.2.4) and (4.2.10) respectively, the following error*

estimate holds:

$$\begin{aligned}
(4.2.12) \quad & \|\sigma - i_h \sigma_h\|_V + \|u - i_h u_h\|_V + \|p - i_h p_h\| \\
& \leq C \left( \inf_{\tau \in i_h V_h^{k-1}} \|\sigma - \tau\|_V + \inf_{v \in i_h V_h^k} \|u - v\|_V + \inf_{q \in i_h V_h^k} \|p - q\|_V + \mu \inf_{v \in i_h V_h^k} \|P_{\mathfrak{B}} u - v\|_V \right. \\
& \qquad \qquad \qquad \left. + \|f_h - i_h^* f\|_h + \|I - J_h\| \|f\| \right),
\end{aligned}$$

where  $J_h = i_h^* i_h$ , and  $\mu = \mu_h^k = \sup_{\substack{r \in \mathfrak{H}_h^k \\ \|r\|=1}} \|(I - i_h^k \pi_h^k) r\|$ .

The extra terms (in the third line of the inequality) are analogous the terms described in the Strang lemmas [11, §III.1]. The main idea of the proof of Theorem 4.2.13 (which we will recall in more detail below, because we will need to prove a generalization of it as part of our main results) is to form an intermediate complex by pulling the inner products in the complex  $(W, d)$  back to  $(W_h, d_h)$  back by  $i_h$ , construct a solution to the problem there, and compare that solution with the solution we want. This modified inner product does not coincide with the given one on  $W_h$  precisely when  $i_h$  is not unitary:

$$\langle v, w \rangle_{i_h^* W} = \langle i_h v, i_h w \rangle_h = \langle i_h^* i_h v, w \rangle_h = \langle J_h v, w \rangle_h.$$

Unitarity is then precisely the condition  $J_h = I$ . The complex  $W_h$  with the modified inner product now may be identified with a true subcomplex of  $W$ , for which the theory of [6] directly applies, yielding a solution  $(\sigma'_h, u'_h, p'_h) \in V_h^{k-1} \times V_h^k \times \mathfrak{H}_h^k$ , where  $\mathfrak{H}_h^k$  is the discrete harmonic space associated to the space with the modified inner product. This generally does not coincide with the discrete harmonic space  $\mathfrak{H}_h^k$ , since the discrete codifferential  $d_h^{*'}$  in that case is defined to be the adjoint with respect to the modified inner product, yielding a different Hodge decomposition. The estimate

of  $\|i_h\sigma'_h - \sigma\|_V + \|i_h u'_h - u\|_V + \|i_h p'_h - p\|$  then proceeds directly from the preceding theory for subcomplexes (4.2.7). The variational crimes, on the other hand, arise from comparing the solution  $(\sigma_h, u_h, p_h)$  with  $(\sigma'_h, u'_h, p'_h)$ . Finally, the error estimate (4.2.12) proceeds by the triangle inequality (and the boundedness of the morphisms  $i_h$ ).

### 4.2.3 Extension of Elliptic Error Estimates for a Nonzero Harmonic Part

Our objective in the remainder of this section is to prove one of our main results, a generalization of Theorem 4.2.13 which allows the possibility of the solution  $u$  having a nonzero harmonic part  $w$ . We first need a couple of lemmas.

**4.2.14 Lemma.** *Theorem 4.2.9 continues to apply when we have  $\langle u, p \rangle = \langle w, p \rangle$  where  $w \in \mathfrak{H}^k$  is prescribed (i.e.,  $P_{\mathfrak{H}} u = w$ , which may generally not be zero).*

*Proof.* We closely follow the proof, in [6], of Theorem 4.2.9 above, noting where the modifications must occur. Let  $B$  be the bounded bilinear form (4.2.5); then  $(\sigma, u, p)$  satisfies, for all  $(\tau_h, v_h, q_h) \in \mathfrak{X}_h^k$ ,

$$B(\sigma, u, p; \tau_h, v_h, q_h) = \langle f, v_h \rangle - \langle u, q_h \rangle.$$

We  $V$ -orthogonally project  $(\sigma, u, p)$  in each factor to  $(\tau, v, q) \in \mathfrak{X}_h^k$ . Then for any

$$(\tau_h, v_h, q_h) \in \mathfrak{X}_h^k,$$

$$\begin{aligned}
(4.2.13) \quad & B(\sigma_h - \tau, u_h - v, p_h - q; \tau_h, v_h, q_h) \\
&= B(\sigma - \tau, u - v, p - q; \tau_h, v_h, q_h) + \langle u, q_h \rangle - \langle w, q_h \rangle \\
&= B(\sigma - \tau, u - v, p - q; \tau_h, v_h, q_h) + \langle P_{\mathfrak{S}_h}(u - w), q_h \rangle \\
&\leq C (\|\sigma - \tau\|_V + \|u - v\|_V + \|p - q\| + \|P_{\mathfrak{S}_h}(u - w)\|) (\|\tau_h\|_V + \|v_h\|_V + \|q_h\|).
\end{aligned}$$

Noticing that the factor  $p_h - q$  in the bilinear form above is in the original domain  $\mathfrak{S}_h^k$ , we can now choose the appropriate  $(\tau_h, v_h, q_h)$  that verifies inf-sup condition of  $B$ :

$$\begin{aligned}
& B(\sigma_h - \tau, u_h - v, p_h - q; \tau_h, v_h, q_h) \\
&\geq \gamma (\|\sigma_h - \tau\|_V + \|u_h - v\|_V + \|p_h - q\|) (\|\tau_h\|_V + \|v_h\|_V + \|q_h\|).
\end{aligned}$$

Comparing this to (4.2.13) above, we may cancel the common factor, and divide by  $\gamma$  to arrive at

$$\begin{aligned}
(4.2.14) \quad & \|\sigma_h - \tau\|_V + \|u_h - v\|_V + \|p_h - q\| \\
&\leq C\gamma^{-1} (\|\sigma - \tau\|_V + \|u - v\|_V + \|p - q\| + \|P_{\mathfrak{S}_h}(u - w)\|).
\end{aligned}$$

This differs (aside from the notation) from [6] in that we now have, rather than  $P_{\mathfrak{S}_h} u$ , instead  $P_{\mathfrak{S}_h}(u - w)$ , with the harmonic part subtracted off. Removing the harmonic part allows us to continue as in [6]: the Hodge decomposition  $u - w = u - P_{\mathfrak{S}} u$  consists only of coboundary and perpendicular terms  $u_{\mathfrak{B}} + u_{\perp} \in \mathfrak{B}^k \oplus \mathfrak{Z}^{k\perp}$ . With  $\mathfrak{S}_h^k$  contained in  $\mathfrak{Z}^k$ , it follows  $P_{\mathfrak{S}_h} u_{\perp} = 0$ , and  $P_{\mathfrak{S}_h} \pi_h u_{\mathfrak{B}} = 0$ . Also,  $(I - \pi_h) u_{\mathfrak{B}}$  is perpendicular to  $\mathfrak{S}_h^k$ .

Therefore, for all  $q \in \mathfrak{S}_h^k$ ,

$$\begin{aligned} \langle P_{\mathfrak{S}_h}(u - P_{\mathfrak{S}}u), q \rangle &= \langle P_{\mathfrak{S}_h}u_{\mathfrak{B}}, q \rangle = \langle P_{\mathfrak{S}_h}(u_{\mathfrak{B}} - \pi_h u_{\mathfrak{B}}), q \rangle \\ &= \langle u_{\mathfrak{B}} - \pi_h u_{\mathfrak{B}}, q \rangle = \langle u_{\mathfrak{B}} - \pi_h u_{\mathfrak{B}}, (I - P_{\mathfrak{S}})q \rangle. \end{aligned}$$

Now, setting

$$q = \frac{P_{\mathfrak{S}_h}(u - P_{\mathfrak{S}}u)}{\|P_{\mathfrak{S}_h}(u - P_{\mathfrak{S}}u)\|} \in \mathfrak{S}_h^k,$$

we have

$$\begin{aligned} \|P_{\mathfrak{S}_h}(u - P_{\mathfrak{S}}u)\| &= \langle P_{\mathfrak{S}_h}(u - P_{\mathfrak{S}}u), q \rangle = \langle u_{\mathfrak{B}} - \pi_h u_{\mathfrak{B}}, (I - P_{\mathfrak{S}})q \rangle \\ &\leq \|u_{\mathfrak{B}} - \pi_h u_{\mathfrak{B}}\| \|(I - P_{\mathfrak{S}})q\| \leq C \|(I - P_{\mathfrak{S}})q\| \inf_{v \in V_h^k} \|u_{\mathfrak{B}} - v\|_V. \end{aligned}$$

Finally, by the second estimate of Theorem 4.2.11 above, we can bound  $\|(I - P_{\mathfrak{S}})q\|$  by  $\|(I - \pi_h)P_{\mathfrak{S}}q\|$ , giving us

$$\|(I - P_{\mathfrak{S}})q\| \leq \|(I - \pi_h)P_{\mathfrak{S}}q\| \leq \sup_{\substack{\|r\|=1 \\ r \in \mathfrak{S}_h^k}} \|(I - \pi_h)r\| \|P_{\mathfrak{S}}q\| \leq \mu.$$

From the triangle inequality, we derive the estimate

$$\begin{aligned} &\|\sigma - \sigma_h\|_V + \|u - u_h\|_V + \|p - p_h\| \\ &\leq \|\sigma - \tau\|_V + \|u - v\|_V + \|p - q\| + \|\tau - \sigma_h\|_V + \|u_h - v\|_V + \|q - p_h\| \\ &\leq (1 + C\gamma^{-1}) \left( \|\sigma - \tau\|_V + \|u - v\|_V + \|p - q\| + \mu \inf_{v \in i_h V_k^h} \|P_{\mathfrak{B}}u - v\|_V \right). \end{aligned}$$

Using best approximation property of orthogonal projections, we can express the remaining terms with the infima, and this gives the result.  $\square$

We also need a technical lemma which enables us to identify the orthogonal projection onto the identified subcomplex  $i_h \mathfrak{X}_h^k$  in order to be able to make additional estimates of the variational crimes in terms of the operator norms  $\|I - J_h\|$ . It is the infinite-dimensional analogue of taking a Moore-Penrose pseudoinverse [102, §3.3] for infinite-dimensional spaces:

**4.2.15 Lemma.** *Let  $i_h : W_h \rightarrow W$  be an injective map of Hilbert spaces, and  $J = i_h^* i_h$ . Then  $J_h$  is invertible, and  $J_h^{-1} i_h^*$  is the Moore-Penrose pseudoinverse of  $i_h$ , i.e. it maps  $i_h W_h$  isometrically back to  $W_h$  with the modified inner product.*

We write  $i_h^+$  for  $J_h^{-1} i_h^*$ .

*Proof.* The invertibility of  $J_h$  follows directly from the injectivity of  $i_h$ , which makes  $\langle J_h \cdot, \cdot \rangle_h$  a positive-definite form. Now,  $(J_h^{-1} i_h^*) i_h = J_h^{-1} J_h = \text{id}_{W_h}$ , which shows that it is in fact a left inverse, as required for pseudoinverses. To show the orthogonality, minimizing  $\frac{1}{2} \|i_h w - b\|^2$  for any  $b \in W$  yields, by the completeness of  $W_h$ , the solution  $w = J_h^{-1} i_h^* b$ , showing that it is a least squares solution, therefore the Moore-Penrose pseudoinverse.  $\square$

We are now ready to prove our main elliptic error estimate, an extension of Theorem 4.2.13.

**4.2.16 Theorem** (Extension of elliptic error estimates to allow for a harmonic part). *Consider the problems (4.2.4) and (4.2.10) but instead with now with prescribed, possibly nonzero harmonic part  $w$ : Given  $f \in W^k$  and  $w \in \mathfrak{H}^k$ , we seek  $(\sigma, u, p) \in \mathfrak{X}^k$  such that*

$$(4.2.15) \quad \begin{aligned} \langle \sigma, \tau \rangle - \langle u, d\tau \rangle &= 0 & \forall \tau \in V^{k-1} \\ \langle d\sigma, v \rangle + \langle du, dv \rangle + \langle p, v \rangle &= \langle f, v \rangle & \forall v \in V^k \\ \langle u, q \rangle &= \langle w, q \rangle & \forall q \in \mathfrak{H}^k. \end{aligned}$$

The solution to this problem exists and is unique, with  $w$  indeed equal to  $P_{\mathfrak{S}}u$ , and is bounded by  $c(\|f\| + \|w\|)$ , with  $c$  depending only on the Poincaré constant. Now, consider the discrete problem, with  $f_h, w_h \in V_h^k$ :

$$(4.2.16) \quad \begin{aligned} \langle \sigma_h, \tau_h \rangle_h - \langle u_h, d_h \tau_h \rangle_h &= 0 & \forall \tau_h \in V_h^{k-1} \\ \langle d_h \sigma_h, v_h \rangle_h + \langle d_h u_h, d_h v_h \rangle_h + \langle p_h, v_h \rangle_h &= \langle f_h, v_h \rangle_h & \forall v_h \in V_h^k \\ \langle u_h, q_h \rangle_h &= \langle w_h, q_h \rangle_h & \forall q_h \in \mathfrak{S}_h^k. \end{aligned}$$

This problem is also well-posed, with the modified Poincaré constant in Theorem 4.2.12. Then we have the following generalization of the error estimate (4.2.12) above:

$$(4.2.17) \quad \begin{aligned} &\|\sigma - i_h \sigma_h\|_V + \|u - i_h u_h\|_V + \|p - i_h p_h\| \\ &\leq C \left( \inf_{\tau \in i_h V_h^{k-1}} \|\sigma - \tau\|_V + \inf_{v \in i_h V_h^k} \|u - v\|_V + \inf_{q \in i_h V_h^k} \|p - q\|_V + \mu \inf_{v \in i_h V_h^k} \|P_{\mathfrak{B}} u - v\|_V \right. \\ &\quad \left. + \inf_{\xi \in i_h V_h^k} \|w - \xi\|_V + \|f_h - i_h^* f\|_h + \|w_h - i_h^* w\|_h + \|I - J_h\| (\|f\| + \|w\|) \right), \end{aligned}$$

where, as before,  $J_h = i_h^* i_h$ , and  $\mu = \mu_h^k = \sup_{\substack{r \in \mathfrak{S}_h^k \\ \|r\|=1}} \|(I - i_h^k \pi_h^k) r\|$ .

We see that three new error terms arise from the approximation of the harmonic part, one being the data interpolation error (but measured in the  $V_h$ -norm, which partially captures how  $d$  fails to commute with  $i_h^*$  and how  $w_h$  may not necessarily be a discrete harmonic form), another best approximation term, and finally another term from the non-unitarity. The relation of  $f_h$  to  $f$  and  $w_h$  to  $w$  need not be further specified, because the theorem directly expresses such a dependence in terms of their relation to  $i_h^* f$  and  $i_h^* w$ ; it has been isolated as a separate issue. However as mentioned in the introduction, and following [50], we often take  $f_h = \Pi_h f$ , where  $\Pi_h$  is some family of linear interpolation operators with  $\Pi_h \circ i_h = \text{id}$ . Another seemingly

obvious choice is  $i_h^*$  itself (thus making those corresponding error terms zero), but as mentioned in [50], this can be difficult to compute, so we do not restrict ourselves to this case. Various choices of interpolation will be crucial in deciding which estimates to make in the parabolic problem. We split the proof of this theorem into two parts, the first of which derives the quantities on the second line of (4.2.17), and the second part, we derive the quantities on the third line of (4.2.17). Generally, we follow the pattern of proof in [6, Theorem 3.9] and [50, Theorem 3.10], noting the necessary modifications, as well as a similar technique given for the improved error estimates by Arnold and Chen [4].

*First part of the proof of Theorem 4.2.16.* First, following Holst and Stern [50] as above, we construct the complex  $W_h$  but with the modified inner product  $\langle J_h \cdot, \cdot \rangle$  (the associated domain complex  $V_h$  remains the same). This gives us a discrete Hodge decomposition with another type of orthogonality and corresponding discrete harmonic forms and orthogonal complement (due to a different adjoint  $d_h^{*'}):$

$$V_h^k = \mathfrak{B}_h^k \oplus \mathfrak{H}_h^{\prime k} \oplus \mathfrak{Z}_h^{k\perp'}$$

(generally, primed objects will represent the corresponding objects defined with the modified inner product; the discrete coboundaries are in fact the same as before, because  $d$  and  $d_h$  do not depend on the choice of inner product). The main complications arise in having to keeping careful track of the different harmonic forms involved, because their nonequivalence and possible non-preservation by the operators contribute directly to the error. We then define, as in [50], the intermediate



solution  $(\sigma'_h, u'_h, p'_h) \in V_h^{k-1} \times V_h^k \times \mathfrak{S}_h^{k'}$  (which we abbreviate as  $\mathfrak{X}'_h$ ):

(4.2.18)

$$\begin{aligned} \langle J_h \sigma'_h, \tau_h \rangle_h - \langle J_h u'_h, d_h \tau_h \rangle_h &= 0 & \forall \tau_h \in V_h^{k-1} \\ \langle J_h d_h \sigma'_h, v_h \rangle_h + \langle J_h d_h u'_h, d_h v_h \rangle_h + \langle J_h p'_h, v_h \rangle_h &= \langle i_h^* f, v_h \rangle_h & \forall v_h \in V_h^k \\ \langle J_h u'_h, q'_h \rangle_h &= \langle i_h^* w, q'_h \rangle_h & \forall q'_h \in \mathfrak{S}_h^{k'}, \end{aligned}$$

and the corresponding bilinear form  $B'_h : \mathfrak{X}'_h \times \mathfrak{X}'_h \rightarrow \mathbb{R}$  given by

$$\begin{aligned} (4.2.19) \quad B'_h(\sigma'_h, u'_h, p'_h; \tau_h, v_h, q'_h) &:= \langle J_h \sigma'_h, \tau_h \rangle_h - \langle J_h u'_h, d_h \tau_h \rangle_h \\ &+ \langle J_h d_h \sigma'_h, v_h \rangle_h + \langle J_h d_h u'_h, d_h v_h \rangle_h + \langle J_h p'_h, v_h \rangle_h - \langle J_h u'_h, q'_h \rangle_h. \end{aligned}$$

This satisfies the inf-sup condition with Poincaré constant  $c_P \|\pi_h\|$ . Note that we will need to extend all the bilinear forms  $B_h$ , and  $B'_h$  in the last factor to all of  $V_h^k$  in order to compare the two, since they are initially only defined on the respective, *differing* harmonic form spaces. This is not a problem so long as we remember to invoke the inf-sup condition only when using the non-extended versions. The idea is, again, to use the triangle inequality:

$$(4.2.20) \quad \|\sigma - i_h \sigma_h\|_V + \|\tau - i_h \tau_h\|_V + \|p - i_h p_h\| \leq$$

$$(4.2.21) \quad \|\sigma - i_h \sigma'_h\|_V + \|\tau - i_h \tau'_h\|_V + \|p - i_h p'_h\|$$

$$(4.2.22) \quad + \|i_h(\sigma'_h - \sigma_h)\|_V + \|i_h(\tau'_h - \tau_h)\|_V + \|i_h(p'_h - p_h)\|.$$

These quantities can be estimated using only geometric properties of the domain; we have no need to actually explicitly compute  $(\sigma'_h, u'_h, p'_h)$ . To estimate the term (4.2.21) (which we shall refer to as the PDE approximation term, whereas (4.2.22) will be called variational crimes), we recall that  $i_h$  is an isometry of  $W_h$  with the modified inner

product to its image, which is a subcomplex.

Thus, Lemma 4.2.14 above applies, with the approximation  $(i_h \sigma'_h, i_h u'_h, i_h p'_h)$  on identified subcomplex  $i_h \mathfrak{X}'^k_h$ . This gives us the terms on the second line of (4.2.17).  $\square$

To finish our main proof, we need to consider the variational crimes (4.2.22). Since the maps  $i_h$  are bounded, and we eventually absorb their norms into the constant  $C$  above, it suffices to consider  $\|\sigma_h - \sigma'_h\|_{V_h} + \|u_h - u'_h\|_{V_h} + \|p_h - p'_h\|_h$ , which we shall state as a separate theorem.

**4.2.17 Theorem.** *Let  $(\sigma_h, u_h, p_h) \in \mathfrak{X}^k_h$  be a solution to (4.2.16),  $(\sigma'_h, u'_h, p'_h) \in \mathfrak{X}'^k_h$  a solution to (4.2.18), and  $w = P_{\mathfrak{S}} u$ , the prescribed harmonic part of the continuous problem. Then*

$$(4.2.23) \quad \|\sigma_h - \sigma'_h\|_{V_h} + \|u_h - u'_h\|_{V_h} + \|p_h - p'_h\|_h \\ \leq C(\|f_h - i_h^* f\|_h + \|w_h - i_h^* w\|_{V_h} + \|I - J_h\|(\|f\| + \|w\|) + \inf_{\xi \in i_h V_h^k} \|w - \xi\|_V),$$

*i.e., they are bounded by the terms on the third line in (4.2.17).*

*Proof of Theorem 4.2.17 and second part of the proof of Theorem 4.2.16.* We follow the proof of Holst and Stern [50, Theorem 3.10] and note the modifications. Let  $(\tau, v, q)$  and  $(\tau_h, v_h, q_h) \in \mathfrak{X}^k_h$ . Consider the bilinear form  $B_h$ , (4.2.11) above, and write

$$B_h(\sigma_h - \tau, u_h - v, p_h - q; \tau_h, v_h, w_h) = B_h(\sigma_h - \sigma'_h, u_h - u'_h, p_h - p'_h; \tau_h, v_h, q_h) \\ + B_h(\sigma'_h - \tau, u'_h - v, p'_h - q; \tau_h, v_h, q_h).$$

We then have, recalling the modified bilinear form  $B'_h$ , (4.2.19) above, and extending it

in the last factors to all of  $V_h^k$ ,

$$\begin{aligned}
& B_h(\sigma'_h, u'_h, p'_h; \tau_h, v_h, q_h) \\
&= B'_h(\sigma'_h, u'_h, p'_h; \tau_h, v_h, q_h) + \langle (I - J_h)\sigma'_h, \tau_h \rangle_h - \langle (I - J_h)u'_h, d_h \tau_h \rangle_h \\
&\quad + \langle (I - J_h)d_h \sigma'_h, v_h \rangle_h + \langle (I - J_h)d_h u'_h, d_h v_h \rangle_h + \langle (I - J_h)p'_h, v_h \rangle_h \\
&\qquad\qquad\qquad - \langle (I - J_h)u'_h, q_h \rangle_h.
\end{aligned}$$

Substituting the respective solutions (4.2.16) and (4.2.18) (and noting the slight discrepancy in the use of different harmonic forms), we have

$$\begin{aligned}
B'_h(\sigma'_h, u'_h, p'_h; \tau_h, v_h, q_h) &= \langle i_h^* f, v_h \rangle_h - \langle J_h u'_h, q_h \rangle_h \\
B_h(\sigma_h, u_h, p_h; \tau_h, v_h, q_h) &= \langle f_h, v_h \rangle_h - \langle w_h, q_h \rangle_h,
\end{aligned}$$

so

$$\begin{aligned}
& B_h(\sigma_h - \sigma'_h, u_h - u'_h, p_h - p'_h; \tau_h, v_h, q_h) \\
&\quad = \langle f_h - i_h^* f, v_h \rangle_h + \langle u'_h, q_h \rangle_h - \langle w_h, q_h \rangle_h \\
&\quad\quad - \langle (I - J_h)\sigma'_h, \tau_h \rangle_h + \langle (I - J_h)u'_h, d_h \tau_h \rangle_h \\
&\quad\quad - \langle (I - J_h)d_h \sigma'_h, v_h \rangle_h - \langle (I - J_h)d_h u'_h, d_h v_h \rangle_h - \langle (I - J_h)p'_h, v_h \rangle_h.
\end{aligned}$$

As before, we bound the form above and below. For the upper bound, using Cauchy-

Schwarz to estimate the extra inner product terms, we arrive at

$$\begin{aligned}
& B_h(\sigma_h - \tau, u_h - v, p_h - q; \tau_h, v_h, q_h) \\
& \leq C (\|f_h - i_h^* f\|_h + \|P_{\mathcal{S}_h}(u'_h - w_h)\|_h + \|I - J_h\| (\|\sigma'_h\|_{V_h} + \|u'_h\|_{V_h} + \|p'_h\|_h) \\
& \quad + \|\sigma'_h - \tau\|_{V_h} + \|u'_h - v\|_{V_h} + \|p'_h - q\|_h) (\|\tau_h\|_{V_h} + \|v_h\|_{V_h} + \|q_h\|_h).
\end{aligned}$$

For the lower bound, we again choose  $(\tau_h, \sigma_h, q_h) \in \mathfrak{X}_h^k$  to verify the inf-sup condition this time for  $B_h$ :

$$\begin{aligned}
& B_h(\sigma_h - \tau, u_h - v, p_h - q; \tau_h, v_h, q_h) \\
& \geq \gamma_h (\|\sigma_h - \tau\|_{V_h} + \|u_h - v\|_{V_h} + \|p_h - q\|_h) (\|\tau_h\|_{V_h} + \|v_h\|_{V_h} + \|q_h\|_h)
\end{aligned}$$

and  $\gamma_h$  depends only on the Poincaré constant  $c_P \|i_h\| \|\pi_h\|$ , uniformly bounded in  $h$ .

Comparing with the upper bound and dividing out the common factor as before, this leads to:

$$\begin{aligned}
& \|\sigma_h - \tau\|_{V_h} + \|u_h - v\|_{V_h} + \|p_h - q\|_h \\
& \leq C \gamma_h^{-1} (\|f_h - i_h^* f\|_h + \|P_{\mathcal{S}_h}(u'_h - w_h)\|_h + \|I - J_h\| (\|\sigma'_h\|_{V_h} + \|u'_h\|_{V_h} + \|p'_h\|_h) \\
& \quad + \|\sigma'_h - \tau\|_{V_h} + \|u'_h - v\|_{V_h} + \|p'_h - q\|_h).
\end{aligned}$$

Choosing  $(\tau, v, q) = (\sigma'_h, u'_h, P_{\mathcal{S}_h} p'_h)$ , applying the triangle inequality with  $p'_h$  to account for the mismatch in the harmonic spaces, and using the well-posedness of the continuous problem (4.2.18),

$$\begin{aligned}
& \|\sigma_h - \sigma'_h\|_{V_h} + \|u_h - u'_h\|_{V_h} + \|p_h - p'_h\|_h \\
& \leq C (\|f_h - i_h^* f\|_h + \|P_{\mathcal{S}_h}(u'_h - w_h)\|_h + \|I - J_h\| (\|f\| + \|w\|) + \|p'_h - q\|_h).
\end{aligned}$$

This differs from [50] in that we have the bound in terms of  $\|f\| + \|w\|$ , and that we must estimate  $\|P_{\mathfrak{S}_h}(u'_h - w_h)\|_h$  rather than  $\|P_{\mathfrak{S}_h} u'_h\|_h$  alone. First, we use the modified Hodge decomposition to uniquely write  $u'_h$  as  $u'_{\mathfrak{B}} + P_{\mathfrak{S}'_h} u'_h + u'_\perp$  with  $u'_{\mathfrak{B}} \in \mathfrak{B}_h^k$  and  $u'_\perp \in \mathfrak{Z}_h^{k\perp'}$ , and

$$\|P_{\mathfrak{S}_h}(u'_h - w_h)\|_h \leq \|P_{\mathfrak{S}_h}(u'_{\mathfrak{B}} + u'_\perp)\|_h + \|P_{\mathfrak{S}_h}(P_{\mathfrak{S}'_h} u'_h - w_h)\|_h.$$

(The projection  $P_{\mathfrak{S}'_h}$  is respect to the *modified* inner product). For the first term, we proceed exactly as in [50]: we have  $P_{\mathfrak{S}_h} u'_{\mathfrak{B}} = 0$  since the coboundary space is still the same, and thus only the term  $u'_\perp$  contributes. Now  $u'_\perp \in \mathfrak{Z}_h^{k\perp'}$  so, using  $J_h$  to express it in terms of  $V$ -orthogonality, we have  $J_h u'_\perp \perp \mathfrak{Z}_h^k$ , and thus  $P_{\mathfrak{S}_h} J_h u'_\perp = 0$ . Therefore, we have

$$\|P_{\mathfrak{S}_h}(u'_{\mathfrak{B}} + u'_\perp)\|_h = \|P_{\mathfrak{S}_h} u'_\perp\|_h = \|P_{\mathfrak{S}_h}(I - J_h)u'_\perp\|_h \leq C\|I - J_h\|(\|f\| + \|w\|).$$

For the  $p'_h$  term, this also proceeds as in [50] unchanged (except for, of course, the extra  $\|w\|$  term): using the (unmodified) discrete Hodge decomposition, we have  $p'_h = P_{\mathfrak{B}_h} p'_h + P_{\mathfrak{S}_h} p'_h = P_{\mathfrak{B}_h} p'_h + q$ . Since  $p'_h \in \mathfrak{S}_h^k$ , a similar argument gives  $J_h p'_h \perp \mathfrak{B}_h^k$ , so  $P_{\mathfrak{B}_h} J_h p'_h = 0$  and

$$\|p'_h - q\|_h = \|P_{\mathfrak{B}_h} p'_h\|_h = \|P_{\mathfrak{B}_h}(I - J_h)p'_h\|_h \leq C\|I - J_h\|(\|f\| + \|w\|).$$

Finally, we must consider the term  $\|P_{\mathfrak{S}_h}(P_{\mathfrak{S}'_h} u'_h - w_h)\|_h$ . Expressing  $u'_h$  in terms of  $w$ , the terms do not combine as easily as the analogous terms involving  $f_h$  and  $i_h^* f$ , because their action as linear functionals operate on different harmonic spaces.

Continuing with the proof of the theorem, we recall the third equation of

(4.2.18):

$$\langle J_h u'_h, q' \rangle_h = \langle i_h^* w, q' \rangle_h = \langle J_h (J_h^{-1} i_h^* w), q' \rangle_h$$

which therefore says  $P_{\mathcal{S}'_h} u'_h = P_{\mathcal{S}'_h} i_h^+ w$ . This enables us to properly work with the modified orthogonal projection  $P_{\mathcal{S}'_h}$ . Because  $i_h^+$  is an isometry of the subspace  $i_h W_h$  to  $W_h$ , we have

$$P_{\mathcal{S}'_h} i_h^+ w = i_h^+ P_{i_h \mathcal{S}'_h} w.$$

where now  $P_{i_h \mathcal{S}'_h}$  is the orthogonal projection onto the identified image harmonic space. Then, using the triangle inequality again,

$$\begin{aligned} & \|P_{\mathcal{S}'_h} (P_{\mathcal{S}'_h} u'_h - w_h)\|_h \\ & \leq \left\| P_{\mathcal{S}'_h} \left( P_{\mathcal{S}'_h} i_h^+ w - i_h^+ w \right) \right\|_h + \|P_{\mathcal{S}'_h} (J_h^{-1} i_h^* w - i_h^* w)\|_h + \|P_{\mathcal{S}'_h} (i_h^* w - w_h)\|_h \\ & \leq \|P_{\mathcal{S}'_h}\| \left( \|i_h^+\| \left\| \left( I - P_{i_h \mathcal{S}'_h} \right) w \right\| + \|J_h^{-1}\| \|I - J_h\| \|i_h^* w\|_h + \|i_h^* w - w_h\|_h \right) \\ & \leq C \left( \left\| \left( I - P_{i_h \mathcal{S}'_h} \right) w \right\| + \|I - J_h\| \|w\| + \|i_h^* w - w_h\|_h \right). \end{aligned}$$

The last term is the data approximation error for  $w$ , and the second term combines with the previous errors that reflect the non-unitarity of the operator. So, all that remains is to estimate the first term. Since it is in the subcomplex  $i_h W_h$ , the first estimate of Theorem 4.2.11 applies:

$$(4.2.24) \quad \left\| \left( I - P_{i_h \mathcal{S}'_h} \right) w \right\| \leq \|(I - \pi'_h) w\| \leq C \inf_{\xi \in i_h V_h^k} \|w - \xi\|_V,$$

by quasi-optimality. □

*Concluding remarks of the proof of Theorem 4.2.16.* To summarize, we have proved Theorem 4.2.16 by defining an intermediate solution on a modified complex that we identify with a subcomplex, and analyzing the result via the Arnold, Falk, and

Winther [6] framework. That theorem holds, with the estimate unchanged, though now  $u$  and  $u_h$  no longer are perpendicular to their respective harmonic spaces. The place where the extra terms all show up is in the variational crimes. In the process of arriving at a term that looks like  $i_h^* w - w_h$ , working with the different harmonic forms produces two more non-unitarity terms  $\|I - J_h\|(\|f\| + \|w\|)$ , and finally, using Theorem 4.2.11 yields a direct estimate of how  $w$  fails to be a modified discrete harmonic form, giving the last best approximation term  $\inf_{\xi \in i_h V_h^k} \|w - \xi\|_V$ .  $\square$

We also note for future reference that in spaces where we have improved error estimates (which means  $\pi_h$  are  $W$ -bounded maps) that we can replace that last  $V$ -norm in (4.2.24) to be the  $W$ -inner product. Finally, we remark that, for a certain types of data interpolation, the errors  $\|f_h - i_h^* f\|$  and  $\|w_h - i_h^* w\|$  can be rewritten in terms of the other errors and another best approximation term. This will be useful for us in our examples.

**4.2.18 Theorem** (Holst and Stern [50], Theorem 3.12). *If  $\Pi_h : W^k \rightarrow W_h^k$  is a family of linear projections uniformly bounded with respect to  $h$ , then for all  $f \in W^k$ ,*

$$(4.2.25) \quad \|\Pi_h f - i_h^* f\| \leq C \left( \|I - J_h\| \|f\| + \inf_{\phi \in i_h W_h^k} \|f - \phi\| \right).$$

### 4.3 Abstract Evolution Problems

In order to solve and approximate linear evolution problems, we introduce the framework of Bochner spaces (also following Gillette and Holst [40]), which realizes time-dependent functions as curves in Banach spaces (which will correspond to spaces of spatially-dependent functions in our problem). We follow mostly [89] and [30] for this material.

### 4.3.1 Overview of Bochner Spaces and Abstract Evolution Problems

Let  $X$  be a Banach space and  $I := [0, T]$  an interval in  $\mathbb{R}$  with  $T > 0$ . We define

$$C(I, X) := \{u : I \rightarrow X \mid u \text{ bounded and continuous}\}.$$

In analogy to spaces of continuous, real-valued functions, we define a supremum norm on  $C(I, X)$ , making  $C(I, X)$  into a Banach space:

$$\|u\|_{C(I, X)} := \sup_{t \in I} \|u(t)\|_X.$$

We will of course need to deal with norms other than the supremum norm, which motivates us to define BOCHNER SPACES: to define  $\mathcal{L}^p(I, X)$ , we complete  $C(I, X)$  with the norm

$$\|u\|_{L^p(I, X)} := \left( \int_I \|u(t)\|_X^p dt \right)^{1/p}.$$

Similarly, we have the space  $H^1(I, X)$ , the completion of  $C^1(I, X)$  with the norm

$$\|u\|_{H^1(I, X)} := \left( \int_I \|u(t)\|_X^2 + \left\| \frac{d}{dt} u(t) \right\|_X^2 dt \right)^{1/2}.$$

There are methods of formulating this in a more measure-theoretic way ([30, Appendix E]), considering Lebesgue-measurable subsets of  $I$ .

As mentioned before, for our purposes,  $X$  will be some space of spatially-dependent functions, and the time-dependence is captured as being a curve in this function space (although this interpretation is only correct when we are considering  $C(I, X)$ —we must be careful about evaluating our functions at single points in time without an enclosing integral). Usually,  $X$  will be a space in some Hilbert complex, such as  $L^2\Omega^k(M)$  or  $H^s\Omega^k(M)$  where the forms are defined over a Riemannian manifold  $M$ .



We introduce this framework in order to be able to formulate parabolic problems more generally. It turns out to be useful to consider the concept of *rigged Hilbert space* or *Gelfand triple*, which consists of a triple of separable Banach spaces

$$V \subseteq H \subseteq V^*$$

such that  $V$  is continuously and densely embedded in  $H$ . For example, if  $(V, d)$  is the domain complex of some Hilbert complex  $(W, d)$ , setting  $V = V^k$  and  $H = W^k$  works, as well as various combinations of their products (so that we can use mixed formulations).  $H$  is also continuously embedded in  $V^*$ . The standard isomorphism (given by the Riesz representation theorem) between  $V$  and  $V^*$ , is not generally the composition of the inclusions, because the primary inner product of importance for weak formulations is the  $H$ -inner product. It coincides with the notion of distributions acting on test functions. Writing  $\langle \cdot, \cdot \rangle$  for the inner product on  $H$ , the setup is designed so that when it happens that some  $F \in V^*$  is actually in  $H$ , we have  $F(v) = \langle F, v \rangle$  (which is why we will often write  $\langle F, v \rangle$  to denote the action of  $F$  on  $v$  even if  $F$  is not in  $H$ ). In fact, in most cases of interest, the  $H$ -inner product is the restriction of a more general bilinear form between two spaces, in which elements of the left (acting) space are of less regularity than elements of  $H$ , while elements of the right space have more regularity.

Given  $A \in C(I, \mathcal{L}(V, V^*))$ , a time-dependent linear operator, we define the bilinear form

$$(4.3.1) \quad a(t, u, v) := \langle -A(t)u, v \rangle,$$

for  $(t, u, v) \in \mathbb{R} \times V \times V$ . To proceed, as in elliptic problems, we need  $a$  to satisfy some kind of coercivity condition, although it need not be as strong. It turns out that

Gårding's Inequality is the right condition to use here:

$$(4.3.2) \quad a(t, u, u) \geq c_1 \|u\|_V^2 - c_2 \|u\|_H^2,$$

with  $c_1, c_2$  constants independent of  $t \in I$ . Then the following problem is the abstract version of linear, parabolic problems:

$$(4.3.3) \quad u_t = A(t)u + f(t)$$

$$(4.3.4) \quad u(0) = u_0.$$

This problem is well-posed:

**4.3.1 Theorem** (Existence of Unique Solution to the Abstract Parabolic Problem, [89], Theorem 11.3). *Let  $f \in L^2(I, V^*)$  and  $u_0 \in H$ , and  $a$  the time-dependent quadratic form in (4.3.1). Suppose (4.3.2) holds. Then the abstract parabolic problem (4.3.3) has a unique solution*

$$u \in L^2(I, V) \cap H^1(I, V^*).$$

*Moreover, the Sobolev embedding theorem implies  $u \in C(I, H)$ , which allows us to unambiguously evaluate the solution at time zero, so the initial condition makes sense, and the solution indeed satisfies it:  $u(0) = u_0$ .*

This theorem is proved via standard methods ([89, p. 382]); we take an orthonormal basis of  $H$  that is simultaneously orthogonal for  $V$  (a frequent situation occurring when, say, it is an orthonormal basis of eigenfunctions of the Laplace operator), formulate the problem in the finite-dimensional subspaces, and use *a priori* bounds on such solutions to extract a weakly convergent subsequence. With this framework, we can show that a wide class of PDE problems, particularly ones that are suited to finite element approximations, are well-posed.

### 4.3.2 Recasting the Problem as an Abstract Evolution Equation

Let us now see how these results apply in the case of the Hodge heat equation (4.1.1) on manifolds. We take a slightly different approach from what is done in [40] and [4], solving an equivalent problem. This sets things up for our modified numerical method detailed in later sections.

Let  $(W, d)$  be a closed Hilbert complex, with domain complex  $(V, d)$ , the standard setup in the above—in particular, we have the Poincaré inequality and the well-posedness of the continuous Hodge Laplacian problem. We consider the space  $\mathfrak{Y}^k := V^{k-1} \times V^k$  and its dual  $\mathfrak{Y}' = (V^{k-1})' \times (V^k)'$  with the obvious product norms (we use primes to denote dual spaces so as not to conflict with the dual complex with respect to the Hodge star defined earlier, though these uses are related). This, along with  $H = W^{k-1} \times W^k$ , gives rigged Hilbert space structure

$$\mathfrak{Y} \subseteq H \subseteq \mathfrak{Y}'.$$

The embeddings are dense and continuous by definition of the graph inner product and that the operators  $d$  have dense domain. We consider the BOCHNER MIXED WEAK PARABOLIC PROBLEM: to seek a weak solution  $(u, \sigma) \in L^2(I, \mathfrak{Y}) \cap H^1(I, \mathfrak{Y}')$  to the mixed problem

$$(4.3.5) \quad \begin{aligned} \langle \sigma, \omega \rangle - \langle u, d\omega \rangle &= 0, & \forall \omega \in V^{k-1}, \quad t \in I, \\ \langle u_t, \varphi \rangle + \langle du, d\varphi \rangle + \langle d\sigma, \varphi \rangle &= \langle f, \varphi \rangle, & \forall \varphi \in V^k, \quad t \in I, \\ u(0) &= g, \end{aligned}$$

this makes it suitable for approximation using finite-dimensional subspaces of  $\mathfrak{Y}'$  (e.g. degrees of freedom for finite element spaces). We see that (4.3.5) is the mixed form

of (4.1.1), which amounts to defining a *system* of differential equations, introducing the variable  $\sigma$  defined by  $\sigma = d^* u$ , where  $d^*$  is the adjoint of the operator  $d$ . We write the equation weakly (namely, moving  $d^*$  back to the other side), which makes no difference at the continuous level, but will make a significant difference when discretizing.

In order to use the abstract machinery above, we need a term with  $\sigma_t$ . Formally differentiating the first equation of (4.1.2), and substituting  $\varphi = d\omega$  in the second equation, we obtain

$$0 = \langle \sigma_t, \omega \rangle - \langle u_t, d\omega \rangle = \langle \sigma_t, \omega \rangle - \langle f, d\omega \rangle + \langle d\sigma, d\omega \rangle + \langle du, dd\omega \rangle.$$

Since  $d^2 = 0$ , that last term vanishes, and so, together with the equation for  $u_t$ , we have the following system:

$$\begin{aligned} \langle \sigma_t, \omega \rangle + \langle d\sigma, d\omega \rangle &= \langle f, d\omega \rangle, \quad \forall \omega \in V^{k-1}, \quad t \in I, \\ (4.3.6) \quad \langle u_t, \varphi \rangle + \langle d\sigma, \varphi \rangle + \langle du, d\varphi \rangle &= \langle f, \varphi \rangle, \quad \forall \varphi \in V^k, \quad t \in I, \\ u(0) &= g. \end{aligned}$$

**4.3.2 Theorem.** *Suppose the initial condition  $g$  is in the domain of the adjoint  $V^*$  and  $f \in L^2(I, (V^k)')$ . Then the problem (4.3.6) is well-posed: there exists a unique solution  $(\sigma, u) \in L^2(I, \mathfrak{Y}) \cap H^1(I, \mathfrak{Y}') \cap C(I, H)$  with  $(\sigma(0), u(0)) = (d^* g, g)$ .*

*Proof.* We see that given  $f \in L^2(I, (V^k)')$ , we have that the functional  $F : (\tau, v) \mapsto \langle f, d\tau \rangle + \langle f, v \rangle$  is in  $L^2(I, \mathfrak{Y}')$ , since  $d$  maps  $V^{k-1}$  to  $V^k$ . For an initial condition on  $\sigma$ , we can demand that  $\sigma(0)$  be the unique  $\sigma_0$  satisfying  $\langle \sigma_0, \tau \rangle - \langle g, d\tau \rangle = 0$ . For this to reasonably hold, we must actually have at least  $u_0 \in V_k^*$ , the domain of the adjoint operator  $d^*$ , that is,  $\sigma_0 = d^* g$ . We equip the spaces with the standard inner products

for product spaces:

$$(4.3.7) \quad \langle (\sigma, u), (\tau, v) \rangle_H := \langle \sigma, \tau \rangle + \langle u, v \rangle$$

$$(4.3.8) \quad \langle (\sigma, u), (\tau, v) \rangle_{\mathfrak{Y}} := \langle \sigma, \tau \rangle_V + \langle u, v \rangle_V.$$

Consider the operator  $A : \mathfrak{Y} \rightarrow \mathfrak{Y}'$  defined by

$$a(\sigma, u; \omega, \varphi) = \langle -A(\sigma, u), (\omega, \varphi) \rangle = \langle d\sigma, d\omega \rangle + \langle d\sigma, \varphi \rangle + \langle du, d\varphi \rangle.$$

With the functional  $F$  defined as above, we have  $F \in L^2(I, \mathfrak{Y}')$ , and so (4.3.6) is equivalent to the problem

$$(4.3.9) \quad (\sigma, u)_t = A(\sigma, u) + F.$$

We now need to verify that the bilinear form  $a$  satisfies Gårding's Inequality:

$$\begin{aligned} a(\sigma, u; \sigma, u) &= \|d\sigma\|^2 + \langle d\sigma, u \rangle + \|du\|^2 \\ &= \|\sigma\|_V^2 - \|\sigma\|^2 + \langle d\sigma, u \rangle + \|u\|_V^2 - \|u\|^2 \\ &\geq \|\sigma\|_V^2 - \|\sigma\|^2 - \|d\sigma\| \|u\| + \|u\|_V^2 - \|u\|^2 \\ &\geq \|\sigma\|_V^2 - \|\sigma\|^2 - \frac{1}{2}\|\sigma\|_V^2 - \frac{1}{2}\|u\|_V^2 + \|u\|_V^2 - \|u\|^2 \\ &= \frac{1}{2}\|(\sigma, u)\|_{\mathfrak{Y}}^2 - \|(\sigma, u)\|_H^2. \end{aligned}$$

Thus, the abstract theory applies, and noting that the initial conditions  $(d^*g, g) \in H$ , we have that

$$(\sigma, u) \in L^2(I, \mathfrak{Y}) \cap H^1(I, \mathfrak{Y}') \cap C(I, H)$$

is the unique solution to (4.3.6) with initial conditions given by  $u(0) = g \in V_k^*$  and

$$\sigma(0) = d^* g. \quad \square$$

Given this, however, we must still establish that we also have a solution to the original mixed problem (which will be crucial in our error estimates):

**4.3.3 Theorem.** *Let  $(\sigma, u) \in L^2(I, \mathfrak{U}) \cap H^1(I, \mathfrak{U}') \cap C(I, H)$  solve (4.3.6) above with the initial conditions. Then, in fact,  $(\sigma, u)$  also solves (4.3.5).*

*Proof.* The second equation already holds, as it is incorporated unchanged into the equations (4.3.6). To show the first equation, we show

$$\langle \sigma_t, \omega \rangle - \langle u_t, d\omega \rangle = 0$$

for all time  $t$ . Then, since the original mixed equation holds at the initial time, standard uniqueness ensures it holds for all  $t \in I$ . We simply realize it is setting  $\varphi = -d\omega$ :

$$\begin{aligned} \langle \sigma_t, \omega \rangle - \langle u_t, d\omega \rangle &= \langle (\sigma, u)_t, (\omega, -d\omega) \rangle_H = a(\sigma_t, u_t; \omega, -d\omega) + \langle f, d\omega \rangle + \langle f, -d\omega \rangle \\ &= \langle d\sigma, d\omega \rangle + \langle d\sigma, -d\omega \rangle + \langle du, dd\omega \rangle = 0. \end{aligned}$$

□

## 4.4 *A Priori* Error Estimates for the Abstract Parabolic Problem

We now combine all the preceding abstract theory (the Holst-Stern [50] framework recalled in §4.2.2, and the abstract evolution problems framework recalled in §4.3) to extend the error estimates of Gillette and Holst [40] and in particular, recover

the case of approximating parabolic equations on compact, oriented<sup>1</sup> Riemannian hypersurfaces in  $\mathbb{R}^{n+1}$  with triangulations in a tubular neighborhood. The key equation in the derivation of the estimates are the generalizations of Thomée's evolution equations for the error terms. We shall see that these equations lead most naturally to the use of certain Bochner norms for the error estimates that are different for each component in the equation.

Let  $(W, d)$  be a closed Hilbert complex with domain  $(V, d)$ , and the Gelfand triple  $\mathfrak{Y} \subseteq H \subseteq \mathfrak{Y}'$  on this complex as above. Now consider our previous standard setup of finite-dimensional approximating complexes  $(W_h, d)$  with domain  $(V_h, d)$ , with corresponding spaces  $\mathfrak{Y}_h^k = V_h^{k-1} \times V_h^k$  (it is  $\mathfrak{X}_h^k$  missing the harmonic factor),  $i_h : V_h \hookrightarrow V$  injective morphisms (that are  $W$ -bounded),  $\pi_h : V_h \rightarrow V$  projection morphisms (which may be merely  $V$ -bounded), and  $\pi_h \circ i_h = \text{id}$ . Finally, we consider data interpolation operators  $\Pi_h : W \rightarrow W_h$ , such that  $\Pi_h \circ i_h = \text{id}$  that realize which projections for the inhomogeneous and prescribed harmonic terms ( $f_h$  and  $w_h$  in the abstract theory above) that we use.

**4.4.1 Discretization of the weak form.** Suppose we have  $f \in L^2(I, (V^k)')$  and  $g \in V_k^*$ . Let  $(\sigma, u) \in L^2(I, \mathfrak{Y}) \cap H^1(I, \mathfrak{Y}') \cap C(I, H)$  be the unique (continuous) solution to (4.3.5), as covered in §4.3. As in [40], we can consider approximations to this solution as functionals on finite-dimensional spaces  $\mathfrak{Y}_h$ , e.g. finite element spaces. With the above considerations, we formulate the SEMI-DISCRETE BOCHNER PARABOLIC PROBLEM: Find

---

<sup>1</sup>Using differential pseudoforms ([36, §2.8], [108], and §1.2 above), we can eliminate this restriction. However, more theory needs to be developed for that case; the normal projection, in particular. We consider this in future work.

$(\sigma_h, u_h) : I \rightarrow \mathfrak{V}_h$  such that

(4.4.1)

$$\begin{aligned} \langle \sigma_h, \omega_h \rangle_h - \langle u_h, d\omega_h \rangle_h &= 0, & \forall \omega_h \in V_h^{k-1}, \quad t \in I, \\ \langle u_{h,t}, \varphi_h \rangle_h + \langle d\sigma_h, \varphi_h \rangle_h + \langle du_h, d\varphi_h \rangle_h &= \langle \Pi_h f, \varphi_h \rangle_h, & \forall \varphi_h \in V_h^k, \quad t \in I, \\ u_h(0) &= g_h. \end{aligned}$$

(We use the notation of Thomée for the test forms.) We define  $g_h$ , the projected initial data, shortly. A similar argument as in §4.3 above, differentiating the first equation with respect to time, considering the Gelfand triple  $\mathfrak{V}_h^k \subseteq W_h^{k-1} \times W_h^k \subseteq (\mathfrak{V}_h^k)'$  gives that this problem is well-posed (or more simply, we choose bases and reduce to standard ODE theory as in (4.1.3) above). Following Gillette and Holst [40], we define the TIME-IGNORANT DISCRETE PROBLEM, using the idea of elliptic projection [110] which we use to define a discrete solution via elliptic projection of the continuous solution at each time  $t_0 \in I$ : We seek  $(\bar{\sigma}_h, \bar{u}_h, \bar{p}_h) \in \mathfrak{X}_h^k$  such that

(4.4.2)

$$\begin{aligned} \langle \bar{\sigma}_h, \omega_h \rangle_h - \langle \bar{u}_h, d\omega_h \rangle_h &= 0, & \forall \omega_h \in V_h^{k-1}, \\ \langle d\bar{\sigma}_h, \varphi_h \rangle_h + \langle d\bar{u}_h, d\varphi_h \rangle_h + \langle \bar{p}_h, \varphi_h \rangle_h &= \langle \Pi_h(-\Delta u(t_0)), \varphi_h \rangle_h, & \forall \varphi_h \in V_h^k, \\ \langle \bar{u}_h, q_h \rangle_h &= \langle \Pi_h(P_{\mathfrak{H}} u(t_0)), q_h \rangle_h & \forall q_h \in \mathfrak{H}_h^k \end{aligned}$$

Note that we have included a prescribed harmonic form given by the harmonic part of  $u$  (following [4]). We then take the initial data  $g_h$  to be  $\bar{u}_h(0)$ ; it is just the solution to the elliptic problem with load data  $\Pi_h(-\Delta g)$ , since  $u(0) = g$ . Note we do not directly interpolate  $g$  itself via  $\Pi_h$  for the data; the reason for this will be seen shortly. This discrete problem is well-posed, i.e., a unique solution  $u_h(t_0)$  always exists for every time  $t_0 \in I$ , by the first part of Theorem 4.2.16 above. The presence of an additional term  $\bar{p}_h$  and equation involving harmonic forms departs from Gillette and Holst [40],



because the theory there is facilitated by the fact that there are no harmonic  $n$ -forms on open domains in  $\mathbb{R}^n$  (the *natural* boundary conditions for such spaces are Dirichlet boundary conditions, in contrast to the more classical example of 0-forms, i.e. functions). Here, however, we must consider harmonic forms, since we may not be working at the end of an abstract Hilbert complex. For our model problem, namely differential forms on compact orientable manifolds (without boundary), even in the case of  $n$ -forms, the theory is completely symmetric (by Poincaré duality [9, 58, 83]).<sup>2</sup> In addition, the linear projections  $\Pi_h$  may not preserve the harmonic space, which gives the possibility of a nonzero  $\tilde{p}_h$ , despite  $-\Delta u$  having zero harmonic part (so it is its own error term).

**4.4.2 Determining the error terms and their evolution.** Continuing the method of Thomée [106], we use the time-ignorant discrete solution as an intermediate reference, and estimate the total errors by comparing to this reference and using the triangle inequality. Roughly speaking, we try to estimate as follows:

$$(4.4.3) \quad \|i_h \sigma_h(t) - \sigma(t)\|_V \leq \|i_h \sigma_h(t) - i_h \tilde{\sigma}_h(t)\|_V + \|i_h \tilde{\sigma}_h(t) - \sigma(t)\|_V$$

$$(4.4.4) \quad \|i_h u_h(t) - u(t)\|_V \leq \|i_h u_h(t) - i_h \tilde{u}_h(t)\|_V + \|i_h \tilde{u}_h(t) - u(t)\|_V.$$

It turns out that this grouping of the terms is not the most natural for our purposes.

We shall see it is the structure of the error evolution equations that groups the terms

---

<sup>2</sup>Despite this, there are a number of reasons why one should still prefer to continue to phrase problems in terms of  $n$ -forms if the problem calls for it ([36] describes how it affects the interpretation of certain quantities); and we shall see that it does in fact still make a difference at the discrete level.

more naturally as:

$$(4.4.5) \quad \|i_h u_h(t) - u(t)\|$$

$$(4.4.6) \quad \|i_h \sigma_h(t) - \sigma(t)\| + \|d(i_h u_h(t) - u(t))\|$$

$$(4.4.7) \quad \|d(i_h \sigma_h(t) - \sigma(t))\|.$$

The sum of these three terms is the sum of the two  $V$ -norms above. In addition, we shall see in our application to hypersurfaces that this particular grouping of the error terms also corresponds more precisely to the order of approximations in the improved estimates for the elliptic projection (namely, they are of orders  $h^{r+1}$ ,  $h^r$ , and  $h^{r-1}$ , respectively, for degree- $r$  polynomial differential forms).

The plan is to use the theory of Holst and Stern [50] reviewed in §4.2.2 above to estimate the sum of the two second terms in (4.4.3) and (4.4.4); the elliptic projection simply is an approximation, at each fixed time, of the trivial case of  $u$  being the solution of the continuous problem with data given by its own Laplacian,  $-\Delta u$ . The harmonic form portion will come up naturally as part of the calculation. Using the notation of Thomée [106], we define the error functions

$$(4.4.8) \quad \rho(t) := \tilde{u}_h(t) - i_h^* u(t)$$

$$(4.4.9) \quad \theta(t) := u_h(t) - \tilde{u}_h(t)$$

$$(4.4.10) \quad \psi(t) = \sigma_h(t) - i_h^* \sigma(t).$$

$$(4.4.11) \quad \varepsilon(t) := \sigma_h(t) - \tilde{\sigma}_h(t)$$

(Thomée does not define the third term  $\psi$ ; we have added it for convenience.) In the case that there are no variational crimes (i.e.,  $J_h$  is unitary), the error terms  $\rho$  and  $\psi$  are bounded above by the elliptic projection errors (because there,  $i_h^*$  is the orthogonal

projection, and  $\|i_h^*\| = \|i_h\| = 1$ ), so that we have, for example, that  $\|i_h u_h - u\| \leq \|\theta\| + \|\rho\|$ , corresponding to the use of  $\rho$  in [106, 40]. For our purposes, however, the choice of  $\rho$  here does not correspond as neatly, now being an intermediate quantity that helps us estimate  $\theta$  in terms the elliptic projection error (the second term in (4.4.4)). We find that it contributes more terms with  $\|I - J_h\|$ . Similar remarks apply for  $\sigma$  and  $\psi$ . We use the method of Thomée to estimate the terms  $\theta$  and  $\varepsilon$  in terms of (the time derivatives of)  $\rho$  and  $\psi$ , and the elliptic projection error; In order to do this, we need an analogue of Thomée's error equations.

**4.4.3 Lemma** (Generalized Thomée error equations). *Let  $\theta$ ,  $\rho$ , and  $\varepsilon$  be defined as above. Then for all  $t \in I$ ,*

(4.4.12)

$$\begin{aligned} \langle \varepsilon, \omega_h \rangle_h - \langle \theta, d\omega_h \rangle_h &= 0 & \forall \omega_h \in V_h^{k-1}, \\ \langle \theta_t, \varphi_h \rangle_h + \langle d\varepsilon, \varphi_h \rangle_h + \langle d\theta, d\varphi_h \rangle_h &= \langle -\rho_t + \tilde{p}_h + (\Pi_h - i_h^*)u_t, \varphi_h \rangle_h & \forall \omega_h \in V_h^k. \end{aligned}$$

This differs from Thomée [106] and Gillette and Holst [40] with the harmonic term  $\tilde{p}_h$ , which accounts for the projections  $\Pi_h$  possibly not sending the harmonic forms to the discrete harmonic forms, an extra  $d\theta$  term which accounts for possibly working away from the end of the complex (for differential forms on an  $n$ -manifold, forms of degree  $k < n$ ), and another data interpolation error term for  $u_t$  (which also distinguishes it from Arnold and Chen [4]).

*Proof.* The first equation is simply weakly expressing  $\varepsilon$  as  $d_h^* \theta$ . This follows immediately from the corresponding equations in the semidiscrete problem and the time-ignorant discrete problem. For the second term, consider the expression

$$(4.4.13) \quad B := \langle \theta_t, \varphi_h \rangle_h + \langle d\varepsilon, \varphi_h \rangle_h + \langle d\theta, d\varphi_h \rangle_h + \langle \rho_t, \varphi_h \rangle_h,$$

and expand it using the definitions to obtain

$$B = \langle u_{h,t}, \varphi_h \rangle_h - \langle \tilde{u}_{h,t}, \varphi_h \rangle_h \\ + \langle d\sigma_h - d\tilde{\sigma}_h, \varphi_h \rangle_h + \langle du_h - d\tilde{u}_h, d\varphi \rangle_h + \langle \tilde{u}_{h,t}, \varphi_h \rangle_h - \langle i_h^* u_t, \varphi_h \rangle_h.$$

We cancel the  $\tilde{u}_{h,t}$  terms, and apply the semidiscrete equation (4.4.1) to cancel the  $d\sigma_h$  and  $du_h$  terms, which gives us

$$B = \langle \Pi_h f, \varphi_h \rangle_h - \langle d\tilde{\sigma}_h, \varphi_h \rangle_h - \langle d\tilde{u}_h, d\varphi_h \rangle_h - \langle i_h^* u_t, \varphi_h \rangle_h,$$

and finally, using the second equation of (4.4.2) to account for the middle terms, we have

$$B = \langle \Pi_h f, \varphi_h \rangle_h + \langle \Pi_h(\Delta u), \varphi \rangle_h + \langle \tilde{p}_h, \varphi_h \rangle_h - \langle i_h^* u_t, \varphi_h \rangle_h \\ = \langle \Pi_h(\Delta u + f - u_t), \varphi_h \rangle_h + \langle \tilde{p}_h, \varphi_h \rangle_h + \langle (\Pi_h - i_h^*) u_t, \varphi_h \rangle_h.$$

But since  $u_t = \Delta u + f$  is the strong form of the equation, which we know is satisfied by the uniqueness, it follows that  $B = \langle \tilde{p}_h + (\Pi_h - i_h^*) u_t, \varphi_h \rangle_h$ . Subtracting the  $\rho_t$  from both sides gives the result.  $\square$

Now we present our main theorem.

**4.4.4 Theorem** (Main parabolic error estimates). *Let  $(\sigma, u)$  be the solution to the continuous problem (4.3.5),  $(\sigma_h, u_h)$  be the semidiscrete solution (4.4.1),  $(\tilde{\sigma}_h, \tilde{u}_h)$  the elliptic projection (4.4.2), and the error quantities (4.4.8)-(4.4.11) be defined as above. Then we have the following error estimates:*

$$(4.4.14) \quad \|\theta(t)\|_h \leq \|\rho_t\|_{L^1(I, W_h)} + \|\tilde{p}_h\|_{L^1(I, W_h)} + \|(\Pi_h - i_h^*)u_t\|_{L^1(I, W_h)}$$

$$(4.4.15) \quad \|d\theta(t)\|_h + \|\varepsilon(t)\|_h \leq C \left( \|\rho_t\|_{L^2(I, W_h)} + \|\tilde{p}_h\|_{L^2(I, W_h)} + \|(\Pi_h - i_h^*)u_t\|_{L^2(I, W_h)} \right)$$

$$(4.4.16) \quad \|d\varepsilon(t)\|_h \leq C \left( \|\psi_t\|_{L^2(I, W_h)} + \|d_h^*(\Pi_h - i_h^*)u_t\|_{L^2(I, W_h)} \right),$$

with

$$(4.4.17) \quad \|\rho_t\|_{L^2(I, W_h)} \leq C \left( \|i_h \tilde{u}_{h,t} - u_t\|_{L^2(I, W)} + \|I - J_h\|_{\mathcal{L}(W_h)} \|u_t\|_{L^2(I, W)} \right)$$

$$(4.4.18) \quad \|\psi_t\|_{L^2(I, W_h)} \leq C \left( \|i_h \tilde{\sigma}_{h,t} - \sigma_t\|_{L^2(I, W)} + \|I - J_h\|_{\mathcal{L}(W_h)} \|\sigma_t\|_{L^2(I, W)} \right).$$

We may further combine these terms, which we shall do in a separate corollary, but it is useful to keep things separate, which allows terms to be analyzed individually when considering specific choices of  $V$  and  $V_h$ . The error terms  $i_h \tilde{\sigma}_h - \sigma$  and  $i_h \tilde{u}_h - u$  and their time derivatives are furthermore estimated in terms of best approximation norms and variational crimes via the theory of Holst and Stern [50]. The different Bochner norms involved arise from the structure of the error evolution equations.

*Proof.* We adapt the proof technique in [106, 40] to our situation, and for ease of notation, unsubscripted norms will denote the  $W$ -norms and norms subscripted with just  $h$  will denote norms on the approximating complex. We now assemble the estimates above separately by computing the  $W$ -norms of the errors and their differentials. We begin by estimating  $\|\theta(t)\|_h$ . We use the standard technique of using the solutions as their own test functions: Set  $\varphi_h = \theta$  and  $\omega_h = \varepsilon$  in (4.4.12). Adding the two equations together yields

$$(4.4.19) \quad \frac{1}{2} \frac{d}{dt} \|\theta\|_h^2 + \|\varepsilon\|_h^2 + \|d\theta\|_h^2 = \langle -\rho_t + \tilde{p}_h + (\Pi_h - i_h^*)u_t, \theta \rangle_h, \quad t \in I$$

Following Thomée [106], we introduce  $\delta > 0$  to account for non-differentiability at  $\theta = 0$ , and observe that

$$\begin{aligned} (\|\theta\|_h^2 + \delta^2)^{1/2} \frac{d}{dt} (\|\theta\|_h^2 + \delta^2)^{1/2} &= \frac{1}{2} \frac{d}{dt} (\|\theta\|_h^2 + \delta^2) \\ &= \frac{1}{2} \frac{d}{dt} \|\theta\|_h^2 \leq (\|\rho_t\|_h + \|\tilde{p}_h\|_h + \|(\Pi_h - i_h^*)u_t\|_h) \|\theta\|_h, \end{aligned}$$

using (4.4.19), the Cauchy-Schwarz inequality, and the definition of operator norms (our goal is to get all of those quantities on the right side of the equation close to zero, so we need not care too much about their sign). Thus, since  $\|\theta\|_h \leq (\|\theta\|_h^2 + \delta^2)^{1/2}$ , we have, canceling  $\|\theta\|_h$ ,

$$\frac{d}{dt} (\|\theta\|_h^2 + \delta^2)^{1/2} \leq \|\rho_t\|_h + \|\tilde{p}_h\|_h + \|(\Pi_h - i_h^*)u_t\|_h.$$

Now, using the Fundamental Theorem of Calculus, we integrate from 0 to  $t$  to get

(4.4.20)

$$\|\theta(t)\|_h = \|\theta(0)\|_h + \lim_{\delta \rightarrow 0} \int_0^t \frac{d}{dt} (\|\theta\|_h^2 + \delta^2)^{1/2} \leq \int_0^t (\|\rho_t\|_h + \|\tilde{p}_h\|_h + \|(\Pi_h - i_h^*)u_t\|_h).$$

$\theta(0)$  vanishes by our choice of initial condition as the elliptic projection.

Next, continuing to follow [40], we consider  $\|\varepsilon(t)\|_h$ . We differentiate the first error equation and substitute  $\varphi_h = 2\theta_t$  and  $\omega_h = 2\varepsilon$ , so that

$$(4.4.21) \quad \langle \varepsilon_t, 2\varepsilon \rangle_h - \langle \theta_t, 2d\varepsilon \rangle_h = 0$$

$$(4.4.22) \quad \langle \theta_t, 2\theta_t \rangle_h + \langle d\varepsilon, 2\theta_t \rangle_h + \langle d\theta, 2d\theta_t \rangle_h = \langle -\rho_t + \tilde{p}_h + (\Pi_h - i_h^*)u_t, 2\theta_t \rangle_h.$$

Adding the two equations as before, we have, by Cauchy-Schwarz and the AM-GM

inequality,

$$\begin{aligned} \frac{d}{dt} \|\varepsilon\|_h^2 + 2\|\theta_t\|_h^2 + \frac{d}{dt} \|d\theta\|_h^2 &\leq 2\|\rho_t\|_h \|\theta_t\|_h + 2\|\tilde{p}_h\|_h \|\theta_t\|_h + 2\|(\Pi_h - i_h^*)u_t\|_h \|\theta_t\|_h \\ &\leq 2(\|\rho_t\|_h^2 + \|\tilde{p}_h\|_h^2 + \|(\Pi_h - i_h^*)u_t\|_h^2) + \frac{3}{2}\|\theta_t\|_h^2. \end{aligned}$$

Again, dropping some positive terms (this time  $\|\theta_t\|_h^2$ ), using the Fundamental Theorem of Calculus and noting the initial conditions vanish by the choice of elliptic projection, we have

$$(4.4.23) \quad \|\varepsilon\|_h^2 + \|d\theta\|_h^2 \leq 2 \int_0^t (\|\rho_t\|_h^2 + \|\tilde{p}_h\|_h^2 + \|(\Pi_h - i_h^*)u_t\|_h^2).$$

Finally, we estimate  $\|d\varepsilon\|_h$ . As in the estimate above, we differentiate the first equation with respect to time, and substitute  $\omega = 2\varepsilon_t$ ,  $\varphi = 2d\varepsilon_t$ ,

$$(4.4.24) \quad \langle \varepsilon_t, 2\varepsilon_t \rangle_h - \langle \theta_t, 2d\varepsilon_t \rangle_h = 0$$

$$(4.4.25) \quad \langle \theta_t, 2d\varepsilon_t \rangle_h + \langle d\varepsilon, 2d\varepsilon_t \rangle_h + \langle d\theta, 2dd\varepsilon_t \rangle_h = \langle -\rho_t + \tilde{p}_h + (\Pi_h - i_h^*)u_t, 2d\varepsilon_t \rangle_h.$$

Noting that  $d^2 = 0$ ,  $\tilde{p}_h$  is perpendicular to the coboundaries, and  $\psi = d_h^* \rho$ , we add the equations to get

$$\begin{aligned} 2\|\varepsilon\|_h^2 + \frac{d}{dt} \|d\varepsilon\|_h^2 &= 2\langle -\rho_t + (\Pi_h - i_h^*)u_t, d\varepsilon_t \rangle_h = 2\langle -\psi_t + d_h^*(\Pi_h - i_h^*)u_t, \varepsilon_t \rangle_h \\ &\leq \|\psi_t\|_h^2 + \|d_h^*(\Pi_h - i_h^*)u_t\|_h^2 + 2\|\varepsilon\|_h^2. \end{aligned}$$

By the Fundamental Theorem of Calculus, and noting vanishing initial conditions

(and an exact cancellation of positive terms), we have

$$(4.4.26) \quad \|d\varepsilon\|_h^2 \leq \int_0^t (\|\psi_t\|_h^2 + \|d_h^*(\Pi_h - i_h^*)u_t\|_h^2).$$

We now estimate  $\rho$  and  $\psi$ . We note that the time derivative of the solutions are also solutions to the mixed formulation, at least provided that  $u_t$  and other associated quantities are sufficiently regular (in the domain of the Laplace operator) for the norms and derivatives to make sense. Then (recalling  $i_h^+ = J_h^{-1}i_h^*$ ), we have

$$(4.4.27) \quad \begin{aligned} \|\rho(t)\|_h = \|\tilde{u}_h - i_h^*u\| &\leq \|\tilde{u}_h - i_h^+u\| + \|i_h^+u - i_h^*u\| \\ &\leq \|i_h^+\| (\|i_h\tilde{u}_h - u\| + \|I - J_h\| \|u\|), \end{aligned}$$

and

$$(4.4.28) \quad \begin{aligned} \|\psi(t)\|_h = \|\tilde{\sigma}_h - i_h^*\sigma\| &\leq \|\tilde{\sigma}_h - i_h^+\sigma\| + \|i_h^+\sigma - i_h^*\sigma\| \\ &\leq \|i_h^+\| (\|i_h\tilde{\sigma}_h - \sigma\| + \|I - J_h\| \|\sigma\|). \end{aligned}$$

The same estimates hold for the time derivatives. The first terms are the estimates that allow us to use the theory of §4.2.2. We note that the theory actually uses  $V$ -norms, but it will work. We cannot improve this in the abstract theory; instead, we use theory for specific choices of  $V$ ,  $W$ , and  $V_h$ , such as appropriately chosen de Rham complexes and approximations to improve the estimates ([6, §3.5], [4, Theorem 3.1]). For these cases, it is helpful to keep the individual estimates on  $\|\varepsilon\|^2$ ,  $\|\theta\|^2$ , etc. separated. We have combined terms because the abstract theory gives us all the variational crimes together, as it makes heavy use of the bilinear forms above. Additional improvement of estimates based on regularity as done in [6] cannot be made for the variational crimes, as discussed in [50, §3.4]. We give the relevant example and result in the next section.  $\square$



**4.4.5 Corollary** (Combined  $L^1$  estimate). *Let  $\theta$ ,  $\rho$ ,  $\psi$ , and  $\varepsilon$  be as above. Then we have*

$$(4.4.29) \quad \begin{aligned} & \|i_h \sigma_h - \sigma\|_{L^1(I,V)} + \|i_h u_h - u\|_{L^1(I,V)} \leq \\ & C \left( \|\rho_t\|_{L^2(I,W_h)} + \|(\Pi_h - i_h^*) u_t\|_{L^2(I,W_h)} + \|\psi_t\|_{L^2(I,W_h)} + \|d_h^* (\Pi_h - i_h^*) u_t\|_{L^2(I,W_h)} \right. \\ & \quad \left. + \|\tilde{\rho}_h\|_{L^2(I,W_h)} + \|i_h \tilde{\sigma}_h - \sigma\|_{L^2(I,V)} + \|i_h \tilde{u}_h - u\|_{L^2(I,V)} \right). \end{aligned}$$

*Further expanding the time derivative terms, we have*

$$\begin{aligned} & \|i_h \sigma_h - \sigma\|_{L^1(I,V)} + \|i_h u_h - u\|_{L^1(I,V)} \leq \\ & C \left( \|i_h \tilde{u}_{h,t} - u_t\|_{L^2(I,W)} + \|i_h \tilde{\sigma}_{h,t} - \sigma_t\|_{L^2(I,W)} \right. \\ & \quad + \|I - J_h\| \|u_t\|_{L^2(I,W)} + \|I - J_h\| \|\sigma_t\|_{L^2(I,W)} \\ & \quad + \|(\Pi_h - i_h^*) u_t\|_{L^2(I,W_h)} + \|d_h^* (\Pi_h - i_h^*) u_t\|_{L^2(I,W_h)} \\ & \quad \left. + \|i_h \tilde{\rho}_h\|_{L^2(I,W)} + \|i_h \tilde{\sigma}_h - \sigma\|_{L^2(I,V)} + \|i_h \tilde{u}_h - u\|_{L^2(I,V)} \right). \end{aligned}$$

These terms are organized as follows: the  $W$ -error in the approximations of the time derivatives, the variational crimes with  $\|I - J_h\|$ , the data approximation error for the time derivatives, and finally the  $V$ -approximation errors for the elliptic projection. These can be further expanded in terms of best approximation errors, but we will not have use for that outside of specific examples where the computation is easier done with the previous theorems. This corollary is simply stated for conceptual clarity and a qualitative sense of all the different individual contributions to the error.

*Proof.* First, we note that by the Cauchy-Schwarz inequality, the estimate for  $\|d\theta\|$  (4.4.14) can be rewritten as using  $L^2(I, W)$  norms to match the squared terms (4.4.23)

and (4.4.26). Combining and absorbing constants, we arrive at

$$\begin{aligned} \|i_h \sigma_h(t) - \sigma(t)\|_V + \|i_h u_h(t) - u(t)\|_V &\leq C \left( \|\rho_t\|_{L^2(I, W_h)} + \|(\Pi_h - i_h^*) u_t\|_{L^2(I, W_h)} \right. \\ &\quad \left. + \|\psi_t\|_{L^2(I, W_h)} + \|d_h^*(\Pi_h - i_h^*) u_t\|_{L^2(I, W_h)} + \|\tilde{p}_h\|_{L^2(I, W_h)} \right) \\ &\quad + \|i_h \tilde{\sigma}_h(t) - \sigma(t)\|_V + \|i_h \tilde{u}_h(t) - u(t)\|_V. \end{aligned}$$

Integrating from 0 to  $T$ , the latter two  $V$ -norm terms become  $L^1(I, V)$  norms (and absorb the factor of  $T$  from integrating the first into the constant). Finally, using Cauchy-Schwarz to change the  $L^1(I, V)$  norm into an  $L^2(I, V)$  norm, and substituting for  $\rho_t$  and  $\psi_t$  gives the result.  $\square$

## 4.5 Parabolic Equations on Compact Riemannian Manifolds

As an application of the preceding results, we return to our original motivating example of de Rham complex to explore an example with the Hodge heat equation on hypersurfaces of Euclidean space, generalizing the discussion in [50, 40]. Let  $M$  be compact hypersurface embedded in  $\mathbb{R}^{n+1}$ .  $M$  inherits a Riemannian metric from the Euclidean metric of  $\mathbb{R}^{n+1}$ .

**4.5.1 The de Rham Complex on a Manifold.** We define the  $L^2$  differential  $k$ -forms on  $M$  given by

$$L^2 \Omega^k(M) := \left\{ \sum_{1 \leq i_1 < \dots < i_k \leq n} a_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k} \in \Omega^k(M) : a_{i_1 \dots i_k} \in L^2(M) \right\},$$

the standard indexing of differential form basis elements, namely strictly increasing sequences from  $\{1, \dots, n\}$ . The inner product is given by  $\langle \omega, \eta \rangle = \int \omega \wedge \star \eta$ , where  $\star$  is

the Hodge operator corresponding to the metric.

The weak exterior derivative  $d^k$  is defined on the domains  $H\Omega^k(M)$ , and we have a Hilbert complex  $(L^2\Omega, d)$  with domain complex  $(H\Omega(M), d)$ , with  $d^{k+1} \circ d^k = 0$ :

$$0 \longrightarrow H\Omega^0 \xrightarrow{d^0} H\Omega^1 \xrightarrow{d^1} \dots \xrightarrow{d^{n-1}} H\Omega^n \longrightarrow 0.$$

As required in the abstract Hilbert complex theory, each domain space carries the graph inner product:

$$\langle u, v \rangle_{H\Omega^k(M)} := \langle u, v \rangle_{L^2\Omega^k(M)} + \langle d^k u, d^k v \rangle_{L^2\Omega^{k+1}(M)}.$$

For open subsets  $U \subseteq \mathbb{R}^n$ , the ends ( $k = 0$  and  $k = n$ ) of this complex are familiar Sobolev spaces of vector fields with the traditional gradient, curl, and divergence operators of vector analysis:

$$0 \longrightarrow H^1(U) \xrightarrow{\text{grad}} H(\text{curl}) \xrightarrow{\text{curl}} \dots \longrightarrow H(\text{div}) \xrightarrow{\text{div}} L^2(U) \longrightarrow 0.$$

Similarly, the dual complex is  $H^*\Omega(M)$  defined by  $H^*\Omega^k(M) := \star H\Omega^{n-k}(M)$ , consisting of Hodge duals of  $(n - k)$ -forms. We have that the embedding  $H\Omega^k(M) \cap H^*\Omega^k(M) \hookrightarrow \mathcal{L}^2\Omega^k(M)$  is compact, which enables a Poincaré Inequality to hold and the resulting Hilbert complex  $(L^2\Omega^k(M), d)$  to be a closed complex [84, 6]. To summarize, we have the following:

**4.5.2 Theorem.** *Let  $M$  be a compact Riemannian hypersurface in  $\mathbb{R}^{n+1}$ . Then taking  $W^k = L^2\Omega^k(M)$ , with maps  $d^k$  the exterior derivative defined on the domains  $V^k = H\Omega^k(M)$ ,  $(W, d)$  is a closed Hilbert complex with domain  $(V, d)$ .*

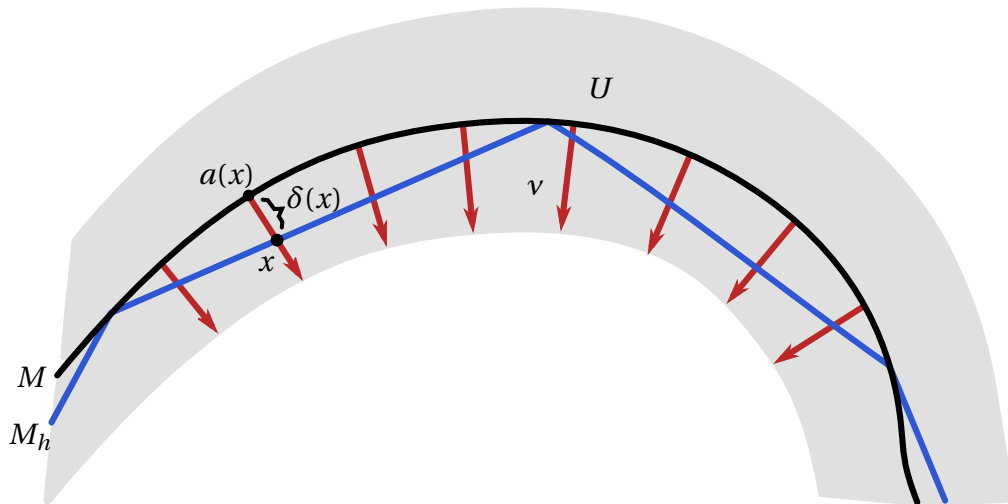
We thus are able to define Hodge Laplacians, and see all the abstract theory for the continuous problems (4.2.15) and (4.3.6) applies with these choices of spaces.

**4.5.3 Approximation of a hypersurface in a tubular neighborhood.** In order to approximate the problems (4.2.15) and (4.3.6), we consider, following [50], a family of approximating hypersurfaces  $M_h$  to an oriented hypersurface  $M$  all contained in a tubular neighborhood  $U$  of  $M$ . The surfaces  $M_h$  generally will be piecewise polynomial (say, of degree  $s$ ); the case  $s = 1$  corresponds to (piecewise linear) triangulations, studied in [27, 25], and generalized for  $s > 1$  in [24]. However, the piecewise linear case still is instrumental in the analysis and indeed, the definition of the spaces (via Lagrange interpolation), and so we shall denote it by  $T_h$  (the triangulation, i.e., set of simplices, will be correspondingly denoted by  $\mathcal{T}_h$ , and their images under the interpolation will be denoted  $\hat{\mathcal{T}}_h$ ). It is convenient, also, to assume that the vertices of the both the triangulation and the higher-degree interpolated surfaces actually lie on the true hypersurface.

The normal vector  $\nu$  to the  $M$  allows us to define a signed distance function  $\delta : U \rightarrow \mathbb{R}$  given by

$$\delta(x) = \pm \operatorname{dist}(x, M) = \pm \inf_{y \in M} |x - y|$$

where the sign is chosen in accordance to which side of the normal  $x$  lies on. By elementary theorems in Riemannian geometry [26, Ch. 6],  $\delta$  is smooth, provided  $U$  is small enough; the maximum distance for which it exists is controlled by the sectional curvature of  $M$ . The normal  $\nu$  can be extended to the whole neighborhood; in fact it is the gradient  $\nabla\delta$ . It is also convenient to define the normals  $\nu_h$  to the approximating surfaces  $M_h$ . In most of the examples we consider, we assume the vertices of  $M_h$  (and  $T_h$ ) lie on  $M$ , but this is not a strict requirement. Instead, we need a condition to ensure that the hypersurfaces  $M_h$  are diffeomorphic to  $M$ , eliminating the possibility of a double covering (e.g., as pictured in [28, Fig. 1, p. 12]). In particular, we want  $M_h$  to have the same topology as  $M$ . This is again restriction on the size of the tubular neighborhood. In such a neighborhood  $U$ , every  $x \in U$  decomposes *uniquely* as



**Figure 4.1:** A curve  $M$  with a triangulation (blue polygonal curve  $M_h$ ) within a tubular neighborhood  $U$  of  $M$ . Some normal vectors  $v$  are drawn, in red; the distance function  $\delta$  is measured along this normal. The intersection  $x$  of the normal with  $M_h$  defines a mapping  $a$  from  $x$  to its base point  $a(x) \in M$ .

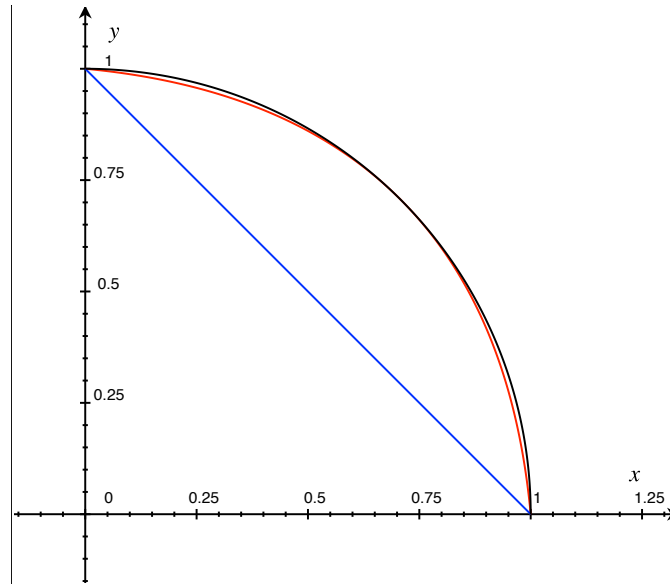
$$(4.5.1) \quad x = a(x) + \delta(x)v(x),$$

where  $a(x) \in M$ , and  $a : U \rightarrow M$  is in fact a smooth function, called the **NORMAL PROJECTION**.  $a$  can then be used to define the degree- $s$  Lagrange interpolated hypersurfaces by considering the image of  $T_h$  under the degree- $s$  Lagrange interpolation of  $a$  over each simplex in  $\mathcal{T}_h$  (we write  $a_k : T_h \rightarrow M_h$  for this) [24, §2.3]. Now, Holst and Stern [50] show, for hypersurfaces, the following result for the variational crime  $\|I - J_h\|$ :

**4.5.4 Theorem** (Holst and Stern [50], Theorem 4.4). *Let  $M$  be an oriented, compact  $m$ -dimensional hypersurface in  $\mathbb{R}^{m+1}$ , and  $M_h$  be a family of hypersurfaces lying in a tubular neighborhood  $U$  of  $M$  transverse to its fibers, such that  $\|\delta\|_\infty \rightarrow 0$  and  $\|v - v_h\|_\infty \rightarrow 0$  as  $h \rightarrow 0$ . Then for sufficiently small  $h$ ,*

$$(4.5.2) \quad \|I - J_h\| \leq C(\|\delta\|_\infty + \|v - v_h\|_\infty^2).$$

A result of Demlow [24, Proposition 2.3] states that, in the case that  $M_h$  is



**Figure 4.2:** Approximation of a quarter unit circle (black) with a segment (blue) and quadratic Lagrange interpolation for the normal projection (red). Even though the underlying triangulation is the same (and thus also the mesh size), notice how much better the quadratic approximation is.

obtained by degree- $s$  Lagrange interpolation, that  $\|\delta\|_\infty < Ch^{s+1}$  and  $\|v - v_h\|_\infty < Ch^s$ .

Thus, putting these results together, we have that

$$(4.5.3) \quad \|I - J_h\| \leq Ch^{s+1}.$$

Now, the three best approximation error terms (4.2.17) for finite element approximation by polynomials of degree  $r$  are bounded by  $Ch^r$ ,  $Ch^{r+1}$ , or  $Ch^{r-1}$ , depending on the component chosen, so it is crucial to allow for this case, and the convergence rate is optimal when  $r = s$ . Figure 4.2 also dramatically demonstrates how much better a higher-order approximation can be with a given mesh size.

Restricting  $a$  to the surfaces  $M_h$  gives diffeomorphisms

$$a|_{M_h} : M_h \rightarrow M.$$

$a: M_h \rightarrow M$  is therefore a diffeomorphism when restricted to each polyhedron (and is at least globally Lipschitz continuous, the maximum degree of regularity in the piecewise linear case. This is not a problem for Hodge theory, because the form spaces are at most  $H^1$  where regularity is concerned; see [111]). See Figure 4.1.

**4.5.5 Finite element spaces.** We thus choose finite-dimensional subspaces  $\Lambda_h^k$  of  $H\Omega^k(M_h)$  for each  $k$ , satisfying the subcomplex property  $d_h\Lambda_h^k \subseteq \Lambda_h^{k+1}$ . We can then pull forms on  $M_h$  back to forms on  $M$  via the inverse of the normal projection, which furnishes the injective morphisms  $i_h^k: \Lambda_h^k \hookrightarrow H\Omega^k(M)$  (since pullbacks commute with  $d$ ) required by the theory above in Section 4.2.

The main finite element spaces relevant for our purposes are two families of piecewise polynomials, discussed in detail in [5, 6]. We must choose these spaces for our equations in a specific relationship in order for the numerical methods and theory detailed above to apply, and for the approximations to work. This is why we prefer a piecewise polynomial approximation of  $M$  as opposed to a curved triangulation of  $M$  itself; these are shown to have these necessary properties.

**4.5.6 Definition** (Polynomial differential forms). Let  $\mathcal{P}_r$  denote polynomials of degree at most  $r$ , in  $n$  variables, and  $\mathcal{H}_r$  be the subspace of homogeneous polynomials. We define the first family, denoted  $\mathcal{P}_r\Lambda^k(\mathcal{T})$ , to consist of all  $k$ -forms with coefficients belonging to  $\mathcal{P}_r$  when restricted to each  $n$ -simplex of  $\mathcal{T}$ . The continuity condition is that the polynomials on two simplices having a common face must have the same trace to that face. The second family, denoted  $\mathcal{P}_r^-\Lambda^k(\mathcal{T})$ , are intermediate spaces, between the spaces of the first class:

$$\mathcal{P}_{r-1}\Lambda^k(\mathcal{T}) \subsetneq \mathcal{P}_r^-\Lambda^k(\mathcal{T}) \subsetneq \mathcal{P}_r\Lambda^k(\mathcal{T}).$$

These are defined as follows: first, consider the radial vector field  $X = x^i \frac{\partial}{\partial x^i}$ , that is,

at each  $x$ , it is a radially pointing vector of length  $|x|$ , and then define the KOSZUL OPERATOR  $\kappa\omega := X \lrcorner \omega$ , the interior product with  $X$ . Then

$$\mathcal{P}_r^- \Lambda^k(\mathcal{T}) := \mathcal{P}_{r-1} \Lambda^k \oplus \kappa \mathcal{H}_{r-1} \Lambda^{k+1}.$$

This is a direct sum, since  $\kappa$  always raises polynomial degree and decreases form degree, so yields homogeneous polynomials of degree  $r$ .  $\kappa$  is in some ways dual to the operator  $d$  (which, in particular, increases form degree and decreases polynomial degree), and by the properties of interior products,  $\kappa^2 = 0$ .

These polynomial spaces generalize existing finite element spaces, such as Whitney forms, Nédélec elements, and Raviart-Thomas elements (see [40, 6] for these examples and more), realizing the collection and clarification of previous results respecting vector methods, as we have mentioned numerous times throughout this work. The important property of these spaces is that they admit the cochain projections whose role we have seen is so important in the theory. First, we describe the case where  $M = U$  is a domain in  $\mathbb{R}^n$  with smooth or Lipschitz boundary.

$$(4.5.4) \quad \pi_h^k : L^2 \Omega^k \rightarrow \Lambda_h^k \quad \text{where } \Lambda_h^k \in \{\mathcal{P}_r \Lambda^k(\mathcal{T}), \mathcal{P}_r^- \Lambda^k(\mathcal{T})\}.$$

These operators, by virtue of their construction, are uniformly bounded (in  $L^2 \Omega^k$ , not just  $H\Omega^k$ ) with respect to  $h$ . Finally, the following theorem explicitly expresses the projection error (and hence, best approximation error) in terms of powers of the mesh size  $h$  and the norms of the solution.

**4.5.7 Theorem** (Arnold, Falk, and Winther [6], Theorem 5.9).

(i.) Let  $\Lambda_h^k$  be one of the spaces  $\mathcal{P}_{r+1}^- \Lambda^k(\mathcal{T})$  or, if  $r \geq 1$ ,  $\mathcal{P}_r \Lambda^k(\mathcal{T})$ . Then  $\pi_h^k$  is a



cochain projection onto  $\Lambda_h^k$  and satisfies

$$\|\omega - \pi_h^k \omega\|_{L^2 \Omega^k(U)} \leq ch^s \|\omega\|_{H^s \Omega^k(U)}, \quad \omega \in H^s \Omega^k(U),$$

for  $0 \leq s \leq r+1$ . Moreover, for all  $\omega \in L^2 \Omega^k(U)$ ,  $\pi_h^k \omega \rightarrow \omega$  in  $L^2$  as  $h \rightarrow 0$ .

(ii.) Let  $\Lambda_h^k$  be one of the spaces  $\mathcal{P}_r \Lambda^k(\mathcal{T})$  or  $\mathcal{P}_r^- \Lambda^k(\mathcal{T})$  with  $r \geq 1$ . Then

$$\|d(\omega - \pi_h^k \omega)\|_{L^2 \Omega^k(U)} \leq ch^s \|d\omega\|_{H^s \Omega^k(U)}, \quad \omega \in H^s \Omega^k(U),$$

for  $0 \leq s \leq r$ .

These bounded cochain operators are explicitly constructed in [5, 6]; they are the natural interpolation operators defined for continuous differential forms and analogous to polynomial interpolation operators on functions, but combined with smoothings to allow extension to  $H^s$  differential forms which may not necessarily be continuous.

**4.5.8 Example** (The Mixed Hodge Laplacian problem on an open subset of  $\mathbb{R}^n$ ). For the mixed Hodge Laplacian problem we considered above, we must choose  $\Lambda_h^{k-1}$  and  $\Lambda_h^k$  in such a manner such that  $d\Lambda_h^{k-1} \subseteq \Lambda_h^k$ ; one cannot make the choices of spaces completely independent of one another for our mixed problem [6, §5.2]. For example, if we choose  $\Lambda_h^{k-1} = \mathcal{P}_r \Lambda^{k-1}(\mathcal{T}_h)$ , we necessarily must choose

$$\Lambda_h^k \in \left\{ \mathcal{P}_{r-1} \Lambda^k(\mathcal{T}_h), \mathcal{P}_r^- \Lambda^k(\mathcal{T}_h) \right\}.$$

Similarly, for  $\Lambda_h^{k-1} = \mathcal{P}_r^- \Lambda^{k-1}(\mathcal{T}_h)$ , we choose

$$\Lambda_h^k \in \left\{ \mathcal{P}_r^- \Lambda^k(\mathcal{T}_h), \mathcal{P}_{r-1} \Lambda^k(\mathcal{T}_h) \right\}.$$

Continuing in this manner down the complex, there are  $2^n$  possible full cochain subcomplexes one can form with these choices of spaces. Of course, for one single Hodge Laplacian problem, we only need to work with three spaces in the chain, since the equations only involve  $(k-1)$ - and  $k$ -forms and their differentials.

**4.5.9 Example** (Finite Element Spaces on Riemannian manifolds). Now, suppose we are back in the situation with a Riemannian hypersurface  $M \subseteq \mathbb{R}^{n+1}$ , with a family of degree- $s$  Lagrange-interpolated surfaces  $M_h$ , over a triangulation  $T_h$ . We can still consider the polynomial finite element spaces on the triangulation  $\mathcal{T}_h$  as before; the only difference here is that the simplices may not join up smoothly (i.e., as a manifold, it may have corners). This is not a problem, because the continuity conditions enforced by the finite element spaces also allow for discontinuities or non-classical-differentiability on the simplicial boundary faces. To define the analogous polynomial spaces on the possibly curved triangulations  $M_h$ , we simply say a form is in the analogous polynomial spaces  $\mathcal{P}_r \Lambda^k(\hat{\mathcal{T}}_h)$  if its pullback by the inverse of the interpolated normal projections  $a_k : T_h \rightarrow M_h$  to  $T_h$  is in  $\mathcal{P}_r \Lambda^k(\mathcal{T}_h)$  [24, §2.5]. Now, from  $\mathcal{P}_r \Lambda^k(\hat{\mathcal{T}}_h)$ , we pull these forms back to the surface  $M$  via the normal projections  $(a|_{M_h})^{-1}$ . This gives the injective morphisms  $i_h^k : \Lambda_h^k \rightarrow H\Omega^k(M)$ ; it commutes with the differentials, since the pullbacks do.

For the bounded cochain operators, the situation is similar. We have  $\pi_h'^k : H\Omega^k(M_h) \rightarrow \Lambda_h^k$  a cochain projection defined by pulling forms defined in neighborhoods back to the triangulations (using the trace theorem if necessary), as constructed in [5, 6]. Then we compose with the pullbacks  $(a|_{M_h})^*$ . This gives us the cochain projections  $\pi_h^k : H\Omega^k(M) \rightarrow \Lambda_h^k$  (by [50, Theorem 3.7]).

**4.5.10 Estimates for the Mixed Hodge Laplacian problem on manifolds.** With this, we can then integrate the terms from [50, Example 4.6] to get the results for the parabolic equations (or, equivalently, add the variational crimes to [40, 4]). Let us

consider now the mixed Hodge Laplacian problem on Riemannian hypersurfaces, considering the setup in the previous example. Namely, we consider  $W^k = \mathcal{L}^2\Omega^k(M)$ ,  $V^k = H\Omega^k(M)$  as above, the approximating spaces  $V_h^{k-1} = \mathcal{P}_{r+1}\Lambda^{k-1}(\hat{\mathcal{T}}_h)$  and  $V_h^k = \mathcal{P}_r\Lambda^k(\hat{\mathcal{T}}_h)$ , and finally the inclusion and projection morphisms as above (possibly with additional pullbacks for interpolation degree  $s > 1$ ). Of course, as mentioned before, these are not the only ways of choosing the spaces, but we stay with, and make estimates based on, this choice for the remainder of this example (the same choice made in [50, Example 4.6]). For a function  $\tilde{f} \in L^2\Omega^k(M)$ , we have an approximate solution  $(\sigma'_h, u'_h, p'_h) \in i_h\mathfrak{X}'_h$  to the elliptic problem, on the true subcomplex  $i_hW_h$  (with modified inner product, as in the theory of §4.2.3). For  $\tilde{f}$  sufficiently regular, and  $(\sigma, u, p)$  satisfying the regularity estimate [6, 40]

$$(4.5.5) \quad \|u\|_{H^{s+2}} + \|p\|_{H^{s+2}} + \|\sigma\|_{H^{s+1}} \leq C\|\tilde{f}\|_{H^s},$$

for  $0 \leq s \leq s_{\max}$ , then, since we are in the de Rham complex, where the cochain projections are  $W$ -bounded, we have the improved error estimates of Arnold, Falk, and Winther [6, §3.5 and p. 342] for the elliptic problem:

$$(4.5.6) \quad \|u - i_h u'_h\| + \|p - i_h p'_h\| \leq Ch^{r+1}\|\tilde{f}\|_{H^{r-1}}$$

$$(4.5.7) \quad \|d(u - i_h u'_h)\| + \|\sigma - i_h \sigma'_h\| \leq Ch^r\|\tilde{f}\|_{H^{r-1}}$$

$$(4.5.8) \quad \|d(\sigma - i_h \sigma'_h)\| \leq Ch^{r-1}\|\tilde{f}\|_{H^{r-1}}.$$

We should also note that Arnold and Chen [4] prove that this also works for a nonzero harmonic part [4, Theorem 3.1]. Holst and Stern [50] augment these estimates to include the variational crimes, so that (changing the notation to suit our problem) for  $(\tilde{\sigma}_h, \tilde{u}_h, \tilde{p}_h) \in \mathfrak{X}_h$ , the discrete solution to the elliptic problem now on the approximat-

ing complexes we have chosen, we have the estimates

$$(4.5.9) \quad \|u - i_h \tilde{u}_h\| + \|p - i_h \tilde{p}_h\| + h(\|d(u - i_h \tilde{u}_h)\| + \|\sigma - i_h \tilde{\sigma}_h\|) \\ + h^2 \|d(\sigma - i_h \tilde{\sigma}_h)\| \leq C(h^{r+1} \|\tilde{f}\|_{H^{r-1}} + h^{s+1} \|\tilde{f}\|).$$

We note the terms associated to the different powers of  $h$  above correspond exactly to the breakdown (4.4.14)-(4.4.16) above. For the elliptic projection in our problem, we also need to account for the nonzero harmonic part of the solution. Setting  $\tilde{w} = P_{\mathcal{H}} \tilde{u}$  and  $\tilde{w}_h = \Pi_h \tilde{w}$ , we have that our three additional terms (given by Theorem 4.2.16 above) are the corresponding best approximation error  $\inf_{v \in V_h^k} \|\tilde{w} - v\|_V$ , the  $\|I - J_h\|$  term, and the data approximation  $\|\tilde{w}_h - i_h^* \tilde{w}\|_h$ . For the best approximation, we make use of our observation about the inequality (4.2.24), in which we may instead use the  $W$ -norm instead of the  $V$ -norm in the case that the projections are  $W$ -bounded, as they are here in the de Rham complex. Because  $\tilde{w}$  is harmonic, it is smooth (and in particular, in  $H^{r+1}$ ), so we may apply Theorem 4.5.7 to find that it is of order  $Ch^{r+1} \|\tilde{w}\|_{H^{r+1}}$ . The  $\|I - J_h\|$  term has already been shown to be of order  $Ch^{s+1}$  above in Theorem 4.5.4. Finally, by Theorem 4.2.18 above, we have that data approximation splits into the other two terms. Therefore, to summarize, we have

**4.5.11 Theorem** (Estimates for the elliptic projection). *Consider  $(\sigma(t), u(t))$ , the solution to the parabolic problem (4.3.6) and  $(\sigma_h(t), u_h(t))$  the semidiscrete solution in (4.4.1) above. Then we have the following estimates for the elliptic projection  $(\tilde{\sigma}_h, \tilde{u}_h, \tilde{p}_h)$ :*

$$(4.5.10) \quad \|u - i_h \tilde{u}_h\| + \|i_h \tilde{p}_h\| + h(\|d(u - i_h \tilde{u}_h)\| + \|\sigma - i_h \tilde{\sigma}_h\|) \\ + h^2 \|d(\sigma - i_h \tilde{\sigma}_h)\| \leq C(h^{r+1} (\|\Delta u\|_{H^{r-1}} + \|\tilde{w}\|_{H^{r+1}}) + h^{s+1} (\|\Delta u\| + \|\tilde{w}\|)).$$

(We note  $p = P_{\mathcal{H}}(-\Delta u) = 0$ .) We now would like use the our main parabolic estimates to analyze the analogous quantity

(4.5.11)

$$\|u(t) - i_h u_h(t)\| + h(\|d(u(t) - i_h u_h(t))\| + \|\sigma(t) - i_h \sigma_h(t)\|) + h^2 \|d(\sigma(t) - i_h \sigma_h(t))\|,$$

and its integral, i.e. Bochner  $L^1$  norm.

**4.5.12 Theorem** (Main combined error estimates for Riemannian hypersurfaces). *Let  $(\sigma(t), u(t))$ ,  $(\sigma_h(t), u_h(t))$ , and all terms involving the elliptic projection are defined as above, and the regularity estimate (4.5.5) is satisfied. Then*

$$\begin{aligned} & \|u - i_h u_h\|_{L^1(W)} + h(\|d(u - i_h u_h)\|_{L^1(W)} + \|\sigma - i_h \sigma_h\|_{L^1(W)}) + h^2 \|d(\sigma - i_h \sigma_h)\|_{L^1(W)} \\ & \leq C [h^{r+1} ((T+1)(\|\Delta u\|_{L^1(H^{r-1})} + \|\tilde{w}\|_{L^1(H^{r+1})}) + T(\|\Delta u_t\|_{L^1(H^{r-1})} + \|\tilde{w}_t\|_{L^1(H^{r+1})})) \\ & \quad + h^{s+1} ((T+1)(\|\Delta u\|_{L^1(W)} + \|\tilde{w}\|_{L^1(W)}) + T(\|\Delta u_t\|_{L^1(W)} + \|\tilde{w}_t\|_{L^1(W})))] . \end{aligned}$$

(We abbreviate  $L^p(I, X)$  as  $L^p(X)$ .) The constants  $T$ , of course, can be further rolled into the constant  $C$ . We remark that in previous results, factors of  $T$  show up on the  $\|\Delta u_t\|$  terms, and, heuristically speaking, this is due to the  $u_t$  being a physically different quantity, namely, a rate of change. However, the appearance of the factor of  $T$  on the  $\|\Delta u\|$  comes from the harmonic approximation error  $\tilde{p}_h$ , which is, physically speaking, a harmonic source term. The details depend on the nature of the approximation operators  $\Pi_h$ .

*Proof.* By the triangle inequality, we have that (4.5.11) breaks up into something of the form (4.5.9) (taking  $(\tilde{\sigma}_h, \tilde{u}_h, \tilde{p}_h)$  to be elliptic projection with  $\tilde{f} = -\Delta u(t)$  and  $\tilde{p} = 0$ ; here  $\tilde{f}$  is not to be confused with the *parabolic* source term  $f(t)$ ) and

$$(4.5.12) \quad \|i_h\| (\|\theta(t)\|_h + h(\|\varepsilon(t)\|_h + \|d u(t)\|_h) + h^2 \|d \varepsilon(t)\|_h),$$

recalling the error quantities defined in (4.4.8)-(4.4.11). Now, substituting our estimates (4.4.14)-(4.4.16), we then have

$$\begin{aligned}
(4.5.13) \quad \|\theta(t)\|_h &\leq \|\rho_t\|_{L^1(W_h)} + \|\tilde{p}_h\|_{L^1(W_h)} + \|(\Pi_h - i_h^*)u_t\|_{L^1(W_h)} \\
&\leq C \left( \|i_h \tilde{u}_{h,t} - u_t\|_{L^1(W)} + \|\tilde{p}_h\|_{L^1(W_h)} + \|I - J_h\| \|u_t\|_{L^1(W)} + \|(\Pi_h - i_h^*)u_t\|_{L^1(W_h)} \right) \\
&\leq C_1 h^{r+1} \left( \|\Delta u\|_{L^1(H^{r-1})} + \|\Delta u_t\|_{L^1(H^{r-1})} + \|\tilde{w}\|_{L^1(H^{r+1})} + \|\tilde{w}_t\|_{L^1(H^{r+1})} \right) \\
&\quad + C_2 h^{s+1} \left( \|\Delta u\|_{L^1(W)} + \|\Delta u_t\|_{L^1(W)} + \|\tilde{w}\|_{L^1(W)} + \|\tilde{w}_t\|_{L^1(W)} \right).
\end{aligned}$$

For  $\|d\theta\|_h + \|\varepsilon\|_h$ , the computation is almost exactly the same, except with possibly different constants, to account for using  $L^2$  Bochner norms, and that :

$$\begin{aligned}
&\|d\theta(t)\|_h + \|\varepsilon(t)\|_h \\
&\leq C \left( \|i_h \tilde{u}_{h,t} - u_t\|_{L^2(W)} + \|\tilde{p}_h\|_{L^2(W_h)} + \|I - J_h\| \|u_t\|_{L^2(W)} + \|(\Pi_h - i_h^*)u_t\|_{L^2(W_h)} \right) \\
&\quad C_3 h^{r+1} \left( \|\Delta u\|_{L^2(H^{r-1})} + \|\Delta u_t\|_{L^2(H^{r-1})} + \|\tilde{w}\|_{L^2(H^{r+1})} + \|\tilde{w}_t\|_{L^2(H^{r+1})} \right) \\
&\quad + C_4 h^{s+1} \left( \|\Delta u\|_{L^2(W)} + \|\Delta u_t\|_{L^2(W)} + \|\tilde{w}\|_{L^2(W)} + \|\tilde{w}_t\|_{L^2(W)} \right).
\end{aligned}$$

These terms are actually absorbed into the lower order terms by the extra factor of  $h$ , due to consisting entirely of the same order terms except using a different norm. However, the situation is slightly different for  $\|d\varepsilon\|_h$ ; namely we use (4.5.7) to get a term of order  $h^r$ , and the  $d_h^*$  on the variational crime part also removing a factor of  $h$ :

$$\begin{aligned}
\|d\varepsilon(t)\|_h &\leq C \left( \|\psi_t\|_{L^2(W_h)} + \|d_h^*(\Pi_h - i_h^*)u_t\|_{L^2(W_h)} \right) \\
&\leq C \left( \|i_h \tilde{\sigma}_{h,t} - \sigma_t\|_{L^2(W)} + \|I - J_h\| \|\sigma_t\|_{L^2(W)} + \|d_h^*(\Pi_h - i_h^*)u_t\|_{L^2(W_h)} \right) \\
&\leq C_5 h^r \left( \|\Delta u_t\|_{L^2(H^{r-1})} + \|\tilde{w}\|_{L^2(H^{r+1})} \right) \\
&\quad + C_6 h^s \left( \|\Delta u\|_{L^2(W)} + \|\Delta u_t\|_{L^2(W)} + \|\tilde{w}\|_{L^2(W)} + \|\tilde{w}_t\|_{L^2(W)} \right)
\end{aligned}$$

However, we see that multiplying by  $h^2$ , this term also gets absorbed; thus we need only consider the error from  $\|d\theta\|_h$  in further calculation of the combined estimate.

We have, thus far:

(4.5.14)

$$\begin{aligned} & \|u(t) - i_h u_h(t)\| + h(\|d(u(t) - i_h u_h(t))\| + \|\sigma(t) - i_h \sigma_h(t)\|) + h^2 \|d(\sigma(t) - i_h \sigma_h(t))\| \\ & \leq C_1 h^{r+1} (\|\Delta u\|_{L^1(H^{r-1})} + \|\Delta u_t\|_{L^1(H^{r-1})} + \|\tilde{w}\|_{L^1(H^{r+1})} + \|\tilde{w}_t\|_{L^1(H^{r+1})}) \\ & \quad + C_2 h^{s+1} (\|\Delta u\|_{L^1(W)} + \|\Delta u_t\|_{L^1(W)} + \|\tilde{w}\|_{L^1(W)} + \|\tilde{w}_t\|_{L^1(W)}) \\ & \quad + C (h^{r+1} (\|\Delta u(t)\|_{H^{r-1}} + \|\tilde{w}(t)\|_{H^{r+1}}) + h^{s+1} (\|\Delta u(t)\| + \|\tilde{w}(t)\|)). \end{aligned}$$

Integrating with respect to  $t$  from 0 to  $T$ , we find that the already-present Bochner norms are constant and thus introduce an extra factor of  $T$ . Absorbing the constants except  $T$  gives the result.  $\square$

This shows, in particular, that the optimal rate of convergence occurs when  $r = s$ , i.e., the polynomial degree of the finite element functions matches the degree of polynomials used to approximate the hypersurface. This tells us, for example, it is not beneficial to use higher-order finite elements on, say, a piecewise linear triangulation. Finally, to put these estimates into some perspective and help develop some intuition for their meaning, we present the generalization of the estimates of Thomée from the introduction.

**4.5.13 Corollary** (Generalization of [106, 40, 4]). *Focusing on just the components  $u$  and  $\sigma$  separately, we have the following estimates (assuming the regularity estimates (4.5.5) are satisfied), and supposing  $r = s$ , i.e., the finite element spaces considered consist of polynomials of the same degree as the interpolation on the surface:*

$$\|u(t) - i_h u_h(t)\| \leq Ch^{r+1} \left( \|u(t)\|_{H^{r+1}} + \int_0^t (\|u(s)\|_{H^{r+1}} + \|u_t(s)\|_{H^{r+1}}) ds \right)$$

$$\|\sigma(t) - i_h \sigma_h(t)\| \leq Ch^{r+1} \left( \|u(t)\|_{H^{r+2}} + \left( \int_0^t (\|u(s)\|_{H^{r+1}}^2 + \|u_t(s)\|_{H^{r+1}}^2) ds \right)^{1/2} \right)$$

This easily leads to an estimate in a Bochner  $L^\infty$  norm (simply take the sup in the non-Bochner norm terms and  $t = T$  in the integrals); this shows that the error in time is small at every  $t \in I$ . Similar estimates hold for  $L^2(I, W)$  norms.

*Proof.* We consider the improved error estimate and variational crimes in  $u$  and  $\sigma$  separately. We first have, by expanding the terms in (4.4.20) as in the derivation of (4.5.13),

$$\begin{aligned} \|u(t) - i_h u_h(t)\| &\leq C (\|u(t) - i_h \tilde{u}_h(t)\| + \|\theta(t)\|) \\ &\leq Ch^{r+1} \left( \|\Delta u(t)\|_{H^{r-1}} + \|\tilde{w}(t)\|_{H^{r+1}} \right. \\ &\quad \left. + \int_0^t (\|\Delta u(s)\|_{H^{r-1}} + \|\Delta u_t(s)\|_{H^{r-1}} + \|\tilde{w}(s)\|_{H^{r-1}} + \|\tilde{w}_t(s)\|_{H^{r-1}}) ds \right). \end{aligned}$$

The result follows by noting that  $\|u\|_{H^{r+1}}$  includes estimates on all the second derivative terms in  $u$ , and  $\tilde{w} = P_{\mathcal{S}} u$ , so those two norms can all be combined (with possibly different constants). Next, we consider  $\sigma$ . The improved error estimates [6, p. 342] imply that if we do not combine estimates involving  $du$  with those of  $\sigma$  for the modified solution, and  $\tilde{f}$  is regular enough to use the  $H^r$ - rather than  $H^{r-1}$ -norm, then we can gain back one factor of  $h$ , so that it is of order  $h^{r+1}$  (rather than  $h^r$  as in (4.5.7)). On the other hand, the elliptic projection error  $\|\varepsilon(t)\|$  still can be taken along with  $\|d\sigma(t)\|$



and was of order  $h^{r+1}$  to begin with. Thus, applying (4.4.23), we have

$$\begin{aligned} \|\sigma(t) - i_h \sigma_h(t)\| &\leq C (\|\sigma(t) - i_h \tilde{\sigma}_h(t)\| + \|\varepsilon(t)\| + \|du(t)\|) \\ &\leq Ch^{r+1} \left( \|\Delta u(t)\|_{H^r} + \|\tilde{w}(t)\|_{H^{r+1}} \right. \\ &\quad \left. + \left[ \int_0^t (\|\Delta u(s)\|_{H^{r-1}}^2 + \|\Delta u_t(s)\|_{H^{r-1}}^2 + \|\tilde{w}(s)\|_{H^{r-1}}^2 + \|\tilde{w}_t(s)\|_{H^{r-1}}^2) ds \right]^{1/2} \right) \\ &\leq Ch^{r+1} \left( \|u(t)\|_{H^{r+2}} + \left[ \int_0^t (\|u(s)\|_{H^{r+1}}^2 + \|u_t(s)\|_{H^{r+1}}^2) ds \right]^{1/2} \right), \end{aligned}$$

where we have used the same consolidation techniques for the norms on  $\Delta u$  and  $\tilde{w}$  into norms on  $u$  as before.  $\square$

We see the variational crimes (arising from the extra  $\tilde{p}_h$ ) account for the sole additional term in the integrals. This cannot be improved without further information on the projections  $\Pi_h$ . Otherwise, for  $r = 1$ , which correspond to piecewise linear discontinuous elements for 2-forms ( $u$ ), and piecewise quadratic elements for 1-forms ( $\sigma$ ) with normal continuity (Raviart-Thomas elements), as studied by Thomée, we obtain the estimates he derived (and since the  $\tilde{p}_h$  is not there in his case, we have that the extra terms with  $u$  do not appear under the integral sign).

## 4.6 Numerical Experiments and Implementation Notes

In order to actually simulate a solution to the Hodge heat equation, we consider the scalar heat equation on a domain in  $M \subseteq \mathbb{R}^2$ , but now using a mixed method with 2-forms rather than the functions. We return to the evolution equation for both  $\sigma$  and

$u$ , (4.3.6) above, which we recall here:

$$(4.6.1) \quad \begin{aligned} \langle \sigma_t, \omega \rangle + \langle d\sigma, d\omega \rangle &= \langle f, d\omega \rangle, \quad \forall \omega \in V^{k-1}, \quad t \in I, \\ \langle u_t, \varphi \rangle + \langle d\sigma, \varphi \rangle + \langle du, d\varphi \rangle &= \langle f, \varphi \rangle, \quad \forall \varphi \in V^k, \quad t \in I, \\ u(0) &= g. \end{aligned}$$

Given  $S_h \subseteq V^k = H\Omega^2(M)$  and  $H_h \subseteq V^{k-1} = H\Omega^1(M)$ , we choose bases, and use the semidiscrete equations (4.1.4), which we recall here (setting  $U$  to be the coefficients of  $u_h$  in the basis for  $S_h$ , and  $\Sigma$  to be the coefficients of  $\sigma_h$  in the basis for  $H_h$ )

$$(4.6.2) \quad \frac{d}{dt} \begin{pmatrix} D & -B^T \\ 0 & A \end{pmatrix} \begin{pmatrix} \Sigma \\ U \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -B & -K \end{pmatrix} \begin{pmatrix} \Sigma \\ U \end{pmatrix} + \begin{pmatrix} 0 \\ F \end{pmatrix}$$

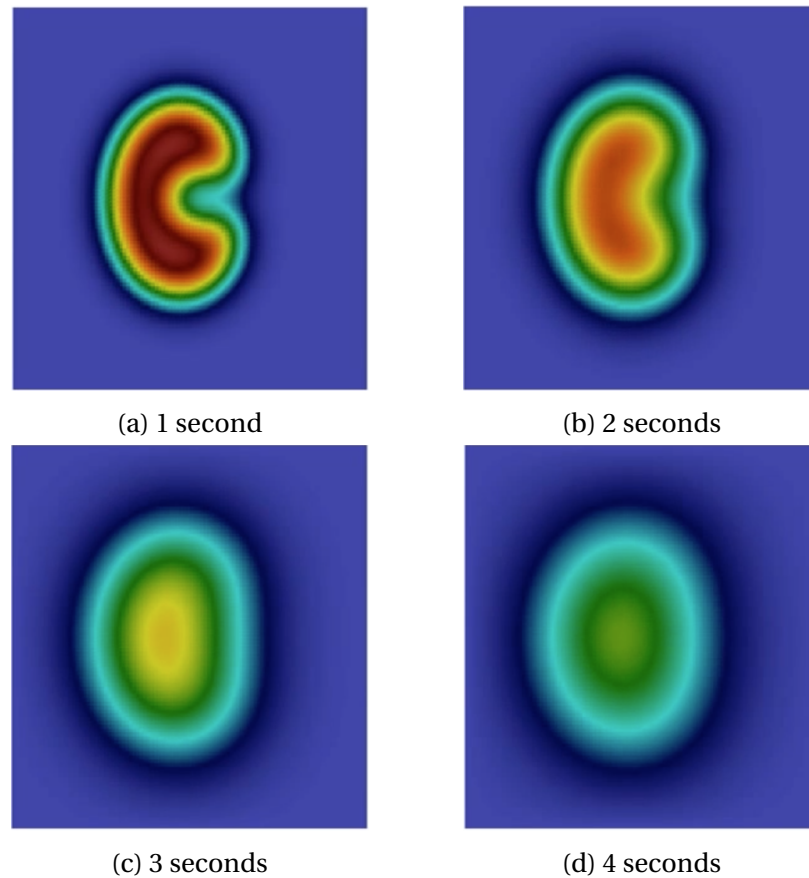
This may be discretized via standard methods for ODEs. For our implementation, we use the backward Euler method. This means we consider sequences  $(\Sigma^n, U^n)$  in time, and then rewrite the derivative instead as a finite difference, evaluating the vector field portion on the right side at timestep  $n+1$ , taking  $M = \begin{pmatrix} D & -B^T \\ 0 & A \end{pmatrix}$ :

$$\frac{1}{\Delta t} M \left( \begin{pmatrix} \Sigma^{n+1} \\ U^{n+1} \end{pmatrix} - \begin{pmatrix} \Sigma^n \\ U^n \end{pmatrix} \right) = \begin{pmatrix} 0 & 0 \\ -B & -K \end{pmatrix} \begin{pmatrix} \Sigma^{n+1} \\ U^{n+1} \end{pmatrix} + \begin{pmatrix} 0 \\ F^{n+1} \end{pmatrix}$$

or

$$\left( M + \Delta t \begin{pmatrix} 0 & 0 \\ B & K \end{pmatrix} \right) \begin{pmatrix} \Sigma^{n+1} \\ U^{n+1} \end{pmatrix} = M \begin{pmatrix} \Sigma^n \\ U^n \end{pmatrix} + \Delta t \begin{pmatrix} 0 \\ F^{n+1} \end{pmatrix}.$$

We now have written the system as a sparse matrix times the unknown,  $(\Sigma^{n+1}, U^{n+1})$ . This allows us to solve the system directly using sparse matrix algorithms without explicitly inverting any matrices, making the iterations efficient. To analyze the error



**Figure 4.3:** Hodge heat equation for  $k = 2$  in a square modeled as a  $100 \times 100$  mesh, using the mixed finite element method given above. Initial data is given as the (discontinuous) characteristic function of a C-shaped set in the square. The timestepping method is given by the backward Euler discretization, with timestep  $\Delta t = 5 \times 10^{-5}$ . The frames are from the supplemental file `heat-demo-hodge.mov` which runs at 60 frames per second.

of the approximations, we can combine the above error estimates with the standard error analysis of Euler methods. See Figure 4.3.

## 4.7 Conclusion and Future Directions

We have seen that the abstract theory of Hilbert complexes, as detailed by Arnold, Falk, and Winther [6], and Bochner spaces, as detailed in Gillette and Holst [40] and Arnold and Chen [4], has been very useful in clarifying the important aspects

of elliptic and parabolic equations. The mixed formulation gives great insight into questions of existence, uniqueness, and stability of the numerical methods (linked by the cochain projections  $\pi_h$ ). The method of Thomée [106] allows us to leverage the existing theory for elliptic problems to apply to parabolic problems, taking care of the remaining error terms by the use of differential inequalities and Grönwall estimates (in the important error evolution equations (4.4.12) above). Incorporating the analysis of variational crimes allow us to carry this theory over to the case of surfaces and their approximations.

We remark on some possible future directions for this work. Some existing surface finite elements for parabolic equations have been studied by Dziuk and Elliott [28] (see supplemental files `heat-demo-basic.mov` and `heat-on-sphere.mpg` for demonstrations on a piecewise linear approximation to the sphere), and much other work by Dziuk, Elliott, Deckelnick [23, 22], which actually treat the case of an evolving surface, and treat a nonlinear equation, the mean curvature flow. Generally speaking, this translates to an additional time dependence for evolving metric coefficients, and a logical place to start is in the Thomée error evolution equations (4.4.12). Nonlinear evolution equations for evolving metrics also suggests the Ricci flow [82, 18, 19], instrumental in showing the Poincaré conjecture. The challenge there, besides nonlinearity, is that tensor equations do not necessarily fit in the framework for FEED. On the other hand, the Yamabe flow [94], which solves for a conformal factor for the metric (and is equivalent to the Ricci flow in dimension 2) suggests an interesting nonlinear scalar evolution equation for which this analysis may be useful.

Gillette and Holst [40] also analyzed hyperbolic equations in this framework, and it would be interesting and useful to analyze methods on surfaces (including the evolving case), as well as taking a more integrated approach in spacetime. This is usually taken care of using the discrete exterior calculus (DEC), the finite-difference coun-

terpart to FEEC to analyze hyperbolic equations [66]. A basic piecewise-linear implementation of this method on the sphere is demonstrated in `waves-on-sphere.mpg`.

## 4.8 Acknowledgements

This chapter, in full, is currently being prepared for submission for publication. The material may appear as M. Holst and C. Tsee, *Approximation of Parabolic Equations in Hilbert Complexes*. The dissertation author was the primary investigator and author of this paper.

# Chapter 5

## Finite Element Methods for Ricci Flow on Surfaces

This chapter is in preparation as a separate published article (joint work with Michael Holst), and therefore may depart from some conventions established earlier, and some material may be duplicated. It is a sketch of how to apply the main results proved in the previous chapter.

### 5.1 Introduction

In this paper, we simulate Ricci flow on surfaces and visualize several examples, exploring interesting geometrical and numerical questions along the way. The Ricci flow is a weakly parabolic evolution equation, for a metric on a manifold. Heuristically, its effect is to smooth out inhomogeneities in an arbitrary initial geometry of some manifold, to eventually yield some kind of canonical geometry, much like how the classical, scalar diffusion equation smoothes out rough initial data, evolving it towards a constant function. Ricci flow was first introduced by Richard Hamilton [45], in which

he proved that given an initial metric of positive Ricci curvature, one can use Ricci flow to evolve the metric to one of constant positive sectional curvature. This technique has gained a lot of prominence in recent years because of the work of Grigori Perelman ([80, 82, 81]), in which he generalizes Hamilton's method, proposing to use Ricci flow to solve the Geometrization Conjecture of Thurston. The nonlinearity of the equation presents numerous challenges (requiring surgeries to continue past singularities that may occur).

In this work, we consider surface finite element method for diffusion equations on evolving surfaces. Surface finite element methods were first considered by Dziuk [27] for numerically solving elliptic PDEs on a piecewise linear approximation to a surface. Subsequently, Demlow [24] treated the case for elliptic equations for higher-order piecewise-polynomial approximations, on higher-dimensional hypersurfaces in  $\mathbb{R}^{n+1}$ . For evolution equations, previous work on surface finite elements include, for example, methods for linear equations on evolving surfaces [28], mean curvature flow [23, 29], diffusion-induced motion of grain boundaries [16, 68], and other applications. We recast and refine the error analysis into the general framework detailed by Holst and Stern [50]. We then conduct actual numerical experiments. Here we use the software MCLite, part of the Finite Element Toolkit (FETK) written by one of the authors. We discuss issues of embedding surfaces into Euclidean space and its interaction with Ricci flow.

## 5.2 Notation and Conventions

We summarize some standard definitions and results from differential geometry and functional analysis, standard material that can be found in, e.g., [62, 26, 58, 30]. We will be working on compact, smooth, orientable manifolds  $M$ , of dimension  $n = 2$ ,

without boundary. A Riemannian metric  $g$  is a symmetric and positive-definite section of the tensor bundle  $T^*M \otimes T^*M$ . We use the notation  $g^{ij}$  for the inverse metric components and use the Einstein Summation convention. For sufficiently differentiable  $g$ , we can define its Levi-Civita connection  $\nabla$  and curvature tensor  $Rm$ . On surfaces, the full Riemann curvature tensor and Ricci tensor are completely determined by the scalar curvature  $R$ , which is twice the Gaussian curvature  $K$ . The Ricci tensor is given by  $Rc = \frac{1}{2}Rg = Kg$ . Since we will be working with different metrics on the same surface, we shall write  $K[g]$ ,  $R[g]$ ,  $Rc[g]$ , etc. to emphasize the dependence of the tensors on the specific metric  $g$ . In the next section, we shall see in detail that the mappings  $g \mapsto K[g]$ ,  $g \mapsto R[g]$ , etc. are nonlinear, second order differential operators on the metric. The metric induces an area 2-form  $dA[g] = \sqrt{\det(g_{ij})} du \wedge dv$  in some smooth coordinates  $u$  and  $v$ . As  $M$  is compact, we may cover it with finitely many charts and define integration via a partition of unity, by taking the Lebesgue integral over each patch, and summing. In particular, by integrating the constant function 1 over a patch, this gives us a measure. So long as the metric coefficients  $g_{ij}$  are  $L^\infty$  over each coordinate chart, this is well-defined. The resulting construction is independent of partition of unity, and in fact, of choice of  $C^1$  metric (the induced norms are equivalent, and at least one such metric always exists).

We now recall the main ideas from Sobolev space theory that we shall need. This requires a metric smooth enough to not interfere with the interpretations of any of the operators in the standard theory, which will not be a problem, since we shall only need to use norms and inner products relative to a smooth background metric (which will be denoted  $g_b$  in the following sections). So given such a metric  $g$ , we have a norm on smooth functions  $f \in C^\infty(M)$  given by

$$(5.2.1) \quad \|f\|_{L^2(M,g)} = \left( \int f^2 dA[g] \right)^{1/2}.$$



This also defines an  $L^2$  inner product. Additionally, given  $f$  smooth, the pointwise norm of its differential  $|df|_g = (g^{ij}\partial_i f \partial_j f)^{1/2}$  is continuous, and thus also integrable, and we define

$$\|f\|_{H^1(M,g)} = \left( \int_M f^2 dA[g] + \int |df|_g^2 dA[g] \right)^{1/2} = (\|f\|_{L^2(M,g)}^2 + \|df\|_{L^2(M,g)}^2)^{1/2}.$$

We then define

$$(5.2.2) \quad L^2(M) = \text{completion of } C^\infty(M) \text{ in the } L^2 \text{ norm}$$

$$(5.2.3) \quad H^1(M) = \text{completion of } C^\infty(M) \text{ in the } H^1 \text{ norm.}$$

Again, these function spaces are independent of metric, but any actual norm on it must use a metric. It can also be shown that this is equivalent to coordinate representations of  $f$  being  $H^1$  over any coordinate patch, so that we may now define tensors and forms to be  $H^1$  if and only if all their coordinate component functions are. It is also important to consider Sobolev spaces of differential  $p$ -forms that are more general than requiring that their components be in  $H^1$ , namely, forms with a notion of weak *exterior* differentiability. This notion treats  $d$  as an organic whole, rather than a linear combination of partial derivatives, and indeed, they may be less regular than  $H^1$  forms [5, 6].

In order to do this, we first define an  $L^2$  inner product on forms, by integrating the pointwise inner product induced from a given metric (inner product on 1-forms), and consider its associated norm. We write  $L^2 \Lambda^k(M)$  for the completion of smooth  $k$ -forms in this norm. We then define the Hodge dual and codifferential. The Hodge dual is simply defined pointwise to take wedge products of  $(g)$ -orthonormal basis forms to wedge products of the basis forms in the complementary set of indices (keeping in mind the orientation): given  $\omega^1$  and  $\omega^2$  in the cotangent space  $T^*M$ ,

and the orientation (and volume form) specified by  $\omega^1 \wedge \omega^2$ , we have  $\star 1 = \omega^1 \wedge \omega^2$ ,  $\star \omega^1 \wedge \omega^2 = 1$ ,  $\star \omega^1 = \omega^2$ , and  $\star \omega^2 = -\omega^1$ . The CODIFFERENTIAL is  $\delta = -\star d\star$  on all forms, and using the covariant derivative, we have

$$(5.2.4) \quad \delta \alpha = -g^{ij} \nabla_i \alpha_j$$

on 1-forms  $\alpha$ , a form of DIVERGENCE. We find that the  $L^2$  inner product associated to  $g$  for forms is now succinctly expressed

$$(5.2.5) \quad (\omega, \eta)_{L^2 \Lambda^k(M)} = \int \omega \wedge \star \eta.$$

We can now define the WEAK EXTERIOR DERIVATIVE: An  $L^2$  differential form  $\omega$  has a weak exterior derivative  $\zeta$  if for all smooth  $(k+1)$ -forms  $\eta$  of compact support,

$$(\zeta, \eta)_{L^2 \Lambda^{k+1}(M)} = (\omega, \delta \eta)_{L^2 \Lambda^k(M)}.$$

If  $\zeta$  exists, it is unique (up to Lebesgue a.e. equivalence), and we write  $\zeta = d\omega$ . We then consider the SOBOLEV SPACE OF DIFFERENTIAL FORMS  $H\Lambda^k(M)$  for all  $L^2$   $k$ -forms  $\omega$  on  $M$  such that  $d\omega$  exists and is also in  $L^2$ .

We finally remark on the Laplacian operator. On forms, we have that  $\Delta_g = -(d\delta + \delta d)$  (it is chosen to have negative eigenvalues). On functions, and in a coordinate chart, this is equivalent to

$$\Delta_g u = \frac{1}{\sqrt{g}} \partial_i (g^{ij} \sqrt{g} \partial_j u)$$

where  $\sqrt{g} = \sqrt{\det g_{ij}}$ . We can recast this as a bilinear weak form:

$$\int (-\Delta_g u)v = \int -\partial_i(g^{ij}\sqrt{g}\partial_j u)v \, dx = \int g^{ij}\sqrt{g}\partial_j u\partial_i v \, dx = (du, dv)_{L^2\Lambda^1(M,g)},$$

so therefore the bilinear weak form corresponding to the Laplacian is exactly the  $L^2$  norm (with the same metric), as it is in the case of (subsets) of Euclidean space. Because, in a chart, the coefficients  $g^{ij}\sqrt{g}$  are  $C^1$ , that means (see [30]) all the standard elliptic weak solution theory carries over locally—a solution  $u$  to  $-\Delta_g u = f$  exists for  $f \in L^2$  satisfying  $\int f \, dA[g] = 0$ , and by theorems on interior regularity,  $u \in H^2(M)$ . By the Sobolev Embedding Theorem (since the dimension is 2), this implies that  $u$  is Hölder continuous for any exponent less than 1.

### 5.3 The Ricci Flow on Surfaces

In this section, we present the Ricci flow equation on surfaces, and show how it can be used to derive an equivalent, scalar equation for a conformal factor. We then further recast it in a normalized form (involving a reparametrization of time and conformal scaling of space, which preserves area).

Let  $(M, g_0)$  be a closed Riemannian surface without boundary. The RICCI FLOW equation with initial metric  $g_0$  is the initial-value problem

$$(5.3.1) \quad \frac{\partial g}{\partial t} = -2\text{Rc} = -2Kg$$

$$(5.3.2) \quad g(0) = g_0$$

for the metric, where  $\text{Rc} = \text{Rc}[g(t)]$  is the Ricci curvature of the evolving metric  $g(t)$  and  $K = K[g(t)]$  is the Gaussian curvature of  $g(t)$  (the simplification  $\text{Rc} = Kg$  is possible only in dimension 2). A further simplification can be made by observing that the

evolution preserves the conformal class of the metric (since the time derivative  $-2Kg$  is a conformal to  $g$ ). If we suppose the evolving metric is conformal to some *background* metric, that is, there exists a (time-independent) metric  $g_b$  and a function  $u(x, t)$  such that

$$g(t) = e^{2u(\cdot, t)} g_b.$$

Substituting  $g = e^{2u} g_b$  into the Ricci Flow equation, we have

$$2e^{2u} \frac{\partial u}{\partial t} g_b = -2K[e^{2u} g_b] e^{2u} g_b.$$

We now use (see [18]) that the Gaussian curvature satisfies the following, under conformal change of metric:

$$K[e^{2u} g] = e^{-2u} (-\Delta_g u + K[g]),$$

Thus the equation now reads

$$(5.3.3) \quad 2e^{2u} \frac{\partial u}{\partial t} g_b = -2(-\Delta_{g_b} u + K[g_b]) g_b,$$

and so equating the factors, we finally have

$$(5.3.4) \quad \frac{\partial u}{\partial t} = e^{-2u} (\Delta_{g_b} u - K[g_b]) = e^{-2u} (\Delta u - K).$$

(We make the convention that unsubscripted geometrical quantities are associated to the background metric  $g_b$ .) This is a PDE in  $u$ , and  $u$  alone—thus we decouple ourselves, in this case, from concerns about tensor equations. We shall, for the purposes of this work, call (5.3.4) the CONFORMAL FACTOR EQUATION. There are other equivalent ways of formulating the equation that may be useful, for example, taking  $F = e^{2u}$  (so

that the conformal factor is  $g = Fg_b$ , we instead get [19, App. B]

$$\frac{\partial F}{\partial t} = \Delta_{g_b}(\log F) - 2K[g_b].$$

We shall find this form useful from time to time. In particular, this can be viewed as a formal limit as  $m \rightarrow 0$  of the porous medium equation,

$$\frac{\partial F}{\partial t} = \Delta_{g_b}(u^m).$$

Also useful is that  $\Delta_{g(t)} = \Delta_{e^{2u}g_b} = e^{-2u}\Delta_{g_b}$  so that we have

$$(5.3.5) \quad \frac{\partial u}{\partial t} = \Delta_{g(t)}u - 2e^{-2u}K.$$

This says that  $u$  satisfies a kind of heat equation, although it is still nonlinear, since  $\Delta_{g(t)}$  depends on the evolving metric (and, of course, that  $e^{2u}$  is still present multiplying  $K$ ).

There is another variant of the Ricci flow equation, which rescales time and space to give a better-behaved equation (it turns out, for example, it exists for all time). The rescaling allows for the existence of a steady state, while the original equation may yield curvature that blows up in finite time. The reparametrization simply sends the blow-up to temporal infinity, while the rescaling allows us to see how the geometry evolves without it actually shrinking to a singularity. Here we assume the metric  $g(s)$  satisfies the Ricci flow equation, and we define  $\tilde{g}(t) = c(\varphi(t))g(\varphi(t))$ , where we seek  $\varphi(t)$  a reparametrization, and  $c(s) > 0$  is a spatially constant rescaling of the metric. We then impose the condition that the surface area should remain constant, and finally see what kind of evolution equation we get. It turns out that  $c(s) := \exp(\int_0^s r(\sigma)d\sigma)$ , where  $r$  is the average scalar curvature of the metric  $g$ , and  $\psi(s) := \int_0^s c(\sigma)d\sigma$  gives the inverse of  $\varphi$ . Details are given in [19]. This leads to the following equation for  $\tilde{g}$ , which

can be thought of as Ricci flow with a “cosmological constant”:

$$(5.3.6) \quad \frac{\partial \tilde{g}}{\partial t} = -2\tilde{\text{Rc}} + \tilde{r}\tilde{g}$$

$$(5.3.7) \quad \tilde{g}(0) = g_0,$$

where  $\tilde{r}$  is the AVERAGE SCALAR CURVATURE

$$\tilde{r} = \frac{1}{\tilde{A}} \int_M \tilde{R} = \frac{2}{\tilde{A}} \int_M \tilde{K} = \frac{4\pi\chi(M)}{A}.$$

Because we demand that the area be constant in time, by the Gauss-Bonnet theorem,

$\tilde{r}$  is constant in time, equal to

$$\frac{1}{A_0} \int_M R_0 = \frac{4\pi\chi(M)}{A_0}.$$

Thus  $\tilde{r} = r_0$ , and a similar calculation as above gives the NORMALIZED CONFORMAL FACTOR EQUATION

$$(5.3.8) \quad \frac{\partial u}{\partial t} = e^{-2u}(\Delta u - K) + \frac{2\pi\chi(M)}{A_0},$$

which is like adding an additional source term to the original conformal factor equation. We have the following theorem that this problem is well-posed (which also, in particular, shows this source term is in fact exactly enough to give a steady state):

**5.3.1 Theorem.** *The conformal factor equation is well-posed, in fact, for all time: given a smooth initial metric  $g_0$ , which we take to be the background metric, there exists a unique, smooth solution  $u : M \times [0, \infty) \rightarrow \mathbb{R}$  of the conformal factor equation such that  $e^{2u(x,t)} g_0$  solves the Ricci flow equation, and moreover, the solution converges, as  $t \rightarrow \infty$ , to a smooth function  $u_\infty$  such that  $e^{2u_\infty} g_0$  is a metric in the same conformal*

class as  $g_0$ , with the same area, and constant curvature equal to the average scalar curvature of  $g_0$ , such that the convergence of  $g$  to its uniformization exponentially fast in any  $C^k$  norm.

Chow and Knopf [18, Theorem 5.1] establish this result in directly for the evolution equation of the metric. Since, as we have observed, the conformal class of the metric does not change, we also have that a solution to the conformal factor equation exists. If we show that it is unique, then it follows that any solution to the conformal factor equation must arise from the corresponding Ricci flow solution of the metric.

## 5.4 Weak Form of the Equation

Here we find a weak formulation for the conformal factor equation, which will be essential for the finite element methods and their analysis. It is convenient to attempt to recast this into quasilinear divergence form [49]:

$$(5.4.1) \quad F(u) = -\nabla \cdot \mathbf{a}(x, u, \nabla u) + b(x, u, \nabla u),$$

where  $\mathbf{a} : T^*M \times \mathbb{R} \rightarrow TM$  is a vector field on  $M$ , and  $b : T^*M \times \mathbb{R} \rightarrow \mathbb{R}$  is a scalar function. Such an operator defined to be ELLIPTIC if its linearization is elliptic, that is, the matrix  $\frac{\partial a^i}{\partial u_{x_j}}$  is positive-definite in coordinates.

We begin with spatial part of the conformal factor equation,  $e^{-2u}(\Delta u - K)$ , and attempt to rewrite it into divergence form. If we consider  $\nabla \cdot (e^{-2u}\nabla u)$ , we have

$$(5.4.2) \quad \nabla \cdot (e^{-2u}\nabla u) = \nabla(e^{-2u}) \cdot \nabla u + e^{-2u}\Delta u = -2e^{-2u}|\nabla u|^2 + e^{-2u}\Delta u.$$

So, rearranging,

$$e^{-2u} \Delta u = \nabla \cdot (e^{-2u} \nabla u) + 2e^{-2u} |\nabla u|^2.$$

So we can rewrite the original equation as

$$\frac{\partial u}{\partial t} = \nabla \cdot (e^{-2u} \nabla u) + 2e^{-2u} |\nabla u|^2 - e^{-2u} K.$$

We define

$$F(u) := -\nabla \cdot (e^{-2u} \nabla u) - 2e^{-2u} |\nabla u|^2 + e^{-2u} K$$

to be the (negative) spatial part of the equation. Now  $F$  is a quasilinear divergence-form operator, as above, with  $\mathbf{a}(x, u, \nabla u) = e^{-2u} \nabla u$  and  $b(x, u, \nabla u) = -2e^{-2u} |\nabla u|^2 + e^{-2u} K$ . Choosing coordinates, we see that

$$a^i(x, u, \nabla u) = e^{-2u} \partial_i u = e^{-2u} u_{x_i},$$

so it follows that

$$\frac{\partial a^i}{\partial u_{x_j}}(x, u, \nabla u) = \frac{\partial}{\partial u_{x_j}}(e^{-2u} u_{x_i}) = e^{-2u} \delta_{ij},$$

which is clearly positive-definite at any  $u$ . This shows  $F$  is, in fact, a quasilinear *elliptic* operator. Integrating against a test function  $v$ , we have the spatial weak form for  $F(u) = f$ :

$$(5.4.3) \quad (F(u), v)_{L^2} = \int_M e^{-2u} \nabla u \cdot \nabla v - 2e^{-2u} |\nabla u|^2 v + e^{-2u} K v = \int_M f v.$$

Because we already know the strong form of the problem is well-posed, the solution  $u$  exists and is bounded, so this is a well-defined form on our function spaces of interest (since we do not have to *solve* for  $u$  in weak form, we need not, for our purposes, consider the more general Sobolev spaces that often occur in nonlinear theory). The



interpretation of the nonlinear operator here [59, 20] is that  $F(u)$  gives the Gaussian curvature of the metric  $e^{2u}g$ . If this problem is solvable for  $f$  given as a constant equal to the sign of the Euler characteristic of  $M$ , this gives the Uniformization Theorem, which states that every compact Riemannian 2-manifold (surface) admits a metric of constant curvature, conformal to the given metric. The Ricci flow equation turns this into a parabolic question, and in fact attempts to realize equilibrium solution (solve elliptic problems) by taking the steady state of the corresponding parabolic problem. As we have seen, taking the parabolic view, the actual computation is quite different, because one is not attempting to invert the actual elliptic operator itself (which can be noninvertible for Neumann and closed manifolds).

There actually is another way to formulate this equation, which is useful for analysis using maximum principles. Recall (5.3.5):

$$\frac{\partial u}{\partial t} = \Delta_{g(t)} u - e^{-2u} K.$$

This makes the weak form of the elliptic part easier to see:

$$\int_M \nabla_{g(t)} u \cdot \nabla_{g(t)} v + e^{-2u} K v = \int_M f v.$$

However, the difficulty is that the metric changes in time. Thus, while the same setup for approximation applies here, it still, of course, leads to nonlinear equations. Here, the evolution of the surface also is dependent on the solution we are trying to find. In the original form, we decouple the surface evolution from the evolution of  $u$ , which ends up being a special case of the surface finite element method of Dziuk and Elliot [28], because the surface itself, for the analysis, is not considered to be evolving (for actual visualization purposes, there is the separate issue of embedding; in our examples, numerical integration suffices). We explain this in detail next.

## 5.5 Numerical Method

As previously mentioned, we shall use a modification of the surface finite element methods for evolution equations, principally, a modification of the method for linear equations of Dziuk and Elliott [28]. Other treatments of nonlinear equations, which will also inform our methods, are the treatments of mean curvature flow given in [23, 29]. The general procedure for solving linear problems via FEM is to reformulate the problem weakly, so that we may set up a system of linear equations by choosing bases in the appropriate Sobolev spaces. The weak formulation, called the GALÈRKIN METHOD also enables us to prove error estimates using modern techniques. The general setup is as follows: Given some linear elliptic differential operator  $L$ , in order to solve the elliptic problem

$$(5.5.1) \quad Lu = f$$

with  $f$  in a function space, say,  $L^2$ , for  $u$  in a nicer function space (say  $H^1$ ), we integrate against test functions  $v$ , and recast the problem as seeking  $u \in H^1$  such that the following equation holds for all test functions  $v$ :

$$(5.5.2) \quad B(u, v) := (Lu, v)_{L^2} = (f, v)_{L^2}.$$

where  $B$  is the weak BILINEAR FORM. In order to approximate the solution  $u$ , we discretize the solution by choosing a finite-dimensional subspace  $X_h$ , and seek an approximate  $u_h \in X_h$  such that

$$B(u_h, v_h) = (f, v_h)_{L^2}$$

for all  $v_h \in X_h$ . By choosing a basis for  $X_h$ , this gives us a set of linear equations. For piecewise linear finite element methods, we choose  $X_h$  by triangulating the domain and defining the basis functions to be the unique piecewise linear functions  $\varphi_i$  such that their value on the nodes of the triangulation  $(x_i)$  are given by  $\varphi_i(x_j) = \delta_{ij}$ . It is of course possible to approximate using piecewise polynomials of higher degree, but here we shall only consider piecewise linear approximation. The innovation introduced by Dziuk in [27] was to formulate the method for general embedded surfaces in  $\mathbb{R}^3$  (much of which depends merely on being hypersurfaces of codimension 1). This introduces some complexity, because the approximating triangulation is not necessarily a subset of the surface itself (whereas this is always the case when triangulating flat domains in Euclidean space, that is, domains of codimension zero), and thus, the approximating function space  $X_h$  is, similarly, not an actual subspace of  $H^1(M)$ , the Sobolev space on the surface.

To deal with nonlinearity, we attempt to do the same thing as before: integrate against test functions to obtain a weak formulation. If  $F$  is a quasilinear elliptic differential operator, such as that defined in (5.4.1), integrating against a test function  $v$  gives us

$$(5.5.3) \quad (F(u), v)_{L^2} = \int \mathbf{a}(x, u, \nabla u) \cdot \nabla v + b(x, u, \nabla u) v,$$

where we have integrated by parts as before, to move the divergence to the other side. Since we work on closed surfaces in this paper, we need not worry about boundary terms. The ability to use integration by parts is why we choose to work with nonlinear operators that still have some sort of divergence term. This is indeed still useful for a very wide class of problems, especially those occurring in differential geometry. Note

now that the operator

$$B(u, v) := (F(u), v)_{L^2}$$

is now linear in its second variable, but *not necessarily* the first. Indeed, approximating a weak solution  $u$  to  $F(u) = f$  by discretizing (using the same kinds of finite-dimensional subspaces  $X_h$ ) requires us to consider solving the *nonlinear* system of equations

$$(F(u_h), v_h)_{L^2} = (f, v_h)_{L^2}.$$

More precisely, given a basis  $(\varphi_i)$  for  $X_h$ , we wish to solve for the components  $\mathbf{u} = (u^i)$  such that for each  $j$ ,

$$\left( F\left( \sum u^i \varphi_i \right), \varphi_j \right)_{L^2} = (f, \varphi_j)_{L^2}.$$

Writing  $\mathbf{F}(\mathbf{u})$  to be the LHS of the preceding (taking the index  $j$  as denoting vector components), and  $\mathbf{f}$  for the RHS, we solve  $\mathbf{F}(\mathbf{u}) = \mathbf{f}$ . To do this, we shall use Newton's method: iterating

$$\mathbf{u}_{n+1} = \mathbf{u}_n - \mathbf{DF}(\mathbf{u}_n)^{-1}(\mathbf{F}(\mathbf{u}_n) - \mathbf{f}).$$

with the appropriate choice of start point. In our parabolic problems (adding a time dependence), the choice will be obvious. We derive  $\mathbf{DF}(\mathbf{u})$  by using the LINEARIZED WEAK FORM

$$(DF(u)w, v)_{L^2} := \left. \frac{d}{dt} \right|_{t=0} (F(u + tw), v)_{L^2}.$$

$\mathbf{DF}(\mathbf{u})$  is the LINEARIZED STIFFNESS MATRIX.

As for adding the time dependence, we also use Newton's method, although in a slightly different context. The general setup is, for  $F$  an elliptic operator,

$$\frac{\partial u}{\partial t} = -F(u) + f.$$

for a source term  $f$  and a quasilinear elliptic operator  $F$  (note the use of the  $-$  is to be consistent with the fact that  $-\Delta$  is the positive elliptic operator, and the heat equation has a  $\Delta$ , not a  $-\Delta$  on the RHS). Choosing a time-independent basis  $\varphi_j$ , we use the method of separation of variables detailed before, in the linear case, to derive time-dependent coefficients,  $u^i$ : a discretized solution  $u(x, t) = u^i(t)\varphi_i(x)$ , and integrate against the test function:

$$\int \frac{du^i}{dt} \varphi_i \varphi_j = - \int_M \mathbf{a}(x, u^i \varphi_i, u^i \nabla \varphi_i) \cdot \nabla \varphi_j + b(x, u^i \varphi_i, u^i \nabla \varphi_i) \varphi_j d\mu + \int_M f \varphi_i,$$

which gives, using the abbreviations  $\mathbf{F}$ ,  $\mathbf{f}$ , etc., as above, and the MASS MATRIX  $M$  defined by

$$M_{ij} = \int \varphi_i \varphi_j,$$

we have

$$M\dot{\mathbf{u}} = -\mathbf{F}(\mathbf{u}) + \mathbf{f}.$$

We shall discretize in time using the backward Euler method, which is a stable, first-order method. Writing  $\dot{\mathbf{u}} = \frac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\Delta t}$ , and expressing the spatial part using the *future time*  $\mathbf{u}^{k+1}$  we have the following equation for  $\mathbf{u}^{k+1}$ :

$$M(\mathbf{u}^{k+1} - \mathbf{u}^k) = \Delta t(\mathbf{f} - \mathbf{F}(\mathbf{u}^{k+1}))$$

which again is a nonlinear equation. We wish to solve for  $\mathbf{u}^{k+1}$  explicitly in terms of  $\mathbf{u}^k$ . This again requires the assistance of Newton's method: we rewrite it as

$$M\mathbf{u}^{k+1} + \Delta t\mathbf{F}(\mathbf{u}^{k+1}) = M\mathbf{u}^k + \Delta t\mathbf{f}.$$

This is the setup for Newton's method. We start with an initial guess  $\mathbf{u}_0^{k+1}$ , which may

reasonably be set to  $\mathbf{u}^k$ , and iterate:

$$\mathbf{u}_{n+1}^{k+1} = \mathbf{u}_n^{k+1} - (M + \Delta t \mathbf{DF}(\mathbf{u}_n^{k+1}))^{-1} (M(\mathbf{u}_n^{k+1} - \mathbf{u}^k) + \Delta t (\mathbf{F}(\mathbf{u}_n^{k+1}) - \mathbf{f})).$$

## 5.6 A Numerical Experiment

The following numerical experiment takes place on the unit sphere. Selecting the background metric  $g_b$  to be the standard round (Euclidean) metric, with constant curvature of 1, we have  $2\pi\chi(S^2) = 4\pi$ , and we derive the normalized conformal factor equation

$$\frac{\partial u}{\partial t} = e^{-2u}(\Delta u - 1) + \frac{4\pi}{A_0} = \nabla \cdot (e^{-2u} \nabla u) + 2e^{-2u} |\nabla u|^2 - e^{-2u} + \frac{4\pi}{A_0}.$$

In this experiment, we choose the initial data

$$u_0(\varphi, \theta) = \frac{1}{2} \log(1 + 0.09 \sin(12 \cos \varphi))$$

This gives initial metric

$$(1 + 0.09 \sin(12 \cos \varphi)) g_e$$

and an area of  $4\pi$ , since the area in 2D is simply the integral of the conformal factor, and the trigonometric terms have vanishing integral by symmetry. We note, in particular, that this metric is rotationally symmetric (and in fact, arises from a surface of revolution—we shall see this shortly). See Figure 5.1a for an embedding realizing this geometry (we describe how this was computed in detail shortly).

This initial data is smooth, including at the poles—a general sufficient condition for smoothness at the poles is simply that all odd-order derivatives vanish [83]. Since  $u_0$  is an even function and is real-analytic on all of  $\mathbb{R}$ , this condition is met

at 0; at  $\pi$  we simply use the fact  $\cos(\pi - u) = -\cos(u)$  and argue by symmetry. This is important to note, since we want a solution that truly is smooth on a sphere, as opposed to one that is more naturally a smooth solution on a cylinder with ends. Even when a rotationally symmetric solution exists, it may not be realizable via embedding, because the embedding equations involve square roots of quantities that become negative for solutions with sufficiently large derivative—embeddability imposes more stringent requirements on the solution than mere existence and uniqueness. Being able to have a true picture of what is happening is very valuable, so despite being a more restricted class of metrics, it is still a worthwhile endeavor to study them.

In order to derive (and solve) the embedding equations, we seek a smooth embedding  $\Phi(\varphi, \theta) = (R(\varphi), \theta, Z(\varphi))$  where the triplet  $(R, \theta, Z)$  in the destination  $\mathbb{R}^3$  denotes *cylindrical* coordinates. The dependence only on  $\varphi$  and not  $\theta$  is how we enforce the condition of rotational symmetry.

The Euclidean metric is then  $dR^2 + R^2 d\theta^2 + dZ^2$ , which, when pulled back via  $\Phi$ , gives us

(5.6.1)

$$\Phi^*(dR^2 + R^2 d\theta^2 + dZ^2) = (R' d\varphi)^2 + R^2 d\theta^2 + (Z' d\varphi)^2 = ((R')^2 + (Z')^2) d\varphi^2 + R^2 d\theta^2.$$

To realize the embedding of our solution, we simply demand that this pullback be equal to  $e^{2u}(d\varphi^2 + \sin^2 \varphi d\theta^2)$ , which gives us, equating coefficients as before,

$$(5.6.2) \quad (R')^2 + (Z')^2 = e^{2u}$$

$$(5.6.3) \quad R^2 = e^{2u} \sin^2 \varphi.$$

We directly see that  $R = e^u \sin(\varphi)$  works. To derive an equation for  $Z$ , we first note that

$$R'(\varphi) = e^u \frac{\partial u}{\partial \varphi} \sin \varphi + e^u \cos \varphi,$$

and substituting back in to the first equation,

$$e^{2u} \left( \frac{\partial u}{\partial \varphi} \sin \varphi + \cos \varphi \right)^2 + (Z')^2 = e^{2u}.$$

Solving for  $Z'$ , we then have

$$Z'(\varphi) = -e^{u(\varphi)} \sqrt{1 - \left( \frac{\partial u}{\partial \varphi} \sin \varphi + \cos \varphi \right)^2}$$

(We choose the negative square root because for  $u \equiv 0$ ,  $Z$  decreases as  $\varphi$  increases, so its derivative should be everywhere nonpositive). Note that  $Z$  itself does not appear in this equation, so the solution is given by integration:

$$(5.6.4) \quad Z(\varphi, t) = Z(0, t) - \int_0^\varphi e^{u(\sigma, t)} \sqrt{1 - \left( \frac{\partial u}{\partial \varphi}(\sigma, t) \sin \sigma + \cos \sigma \right)^2} d\sigma$$

$$(5.6.5) \quad R(\varphi, t) = e^{u(\varphi, t)} \sin(\varphi)$$

The freedom of the value  $Z(0, t)$  reflects the fact that post-composing the embedding with an isometry of Euclidean space (here a translation) should not affect the Euclidean metric. In our example, we choose  $Z(0, t) \equiv 1$ , so as to fix the north pole for all time.

That a square root is taken and we are subtracting the term  $(\partial u / \partial \varphi \sin \varphi + \cos \varphi)^2$  underneath it means that it is certainly possible for the integrand to become complex, and thus derive an inadmissible embedding. To guarantee that a solution exists, we must have

$$\left| \frac{\partial u}{\partial \varphi} \sin \varphi + \cos \varphi \right| \leq 1$$



or

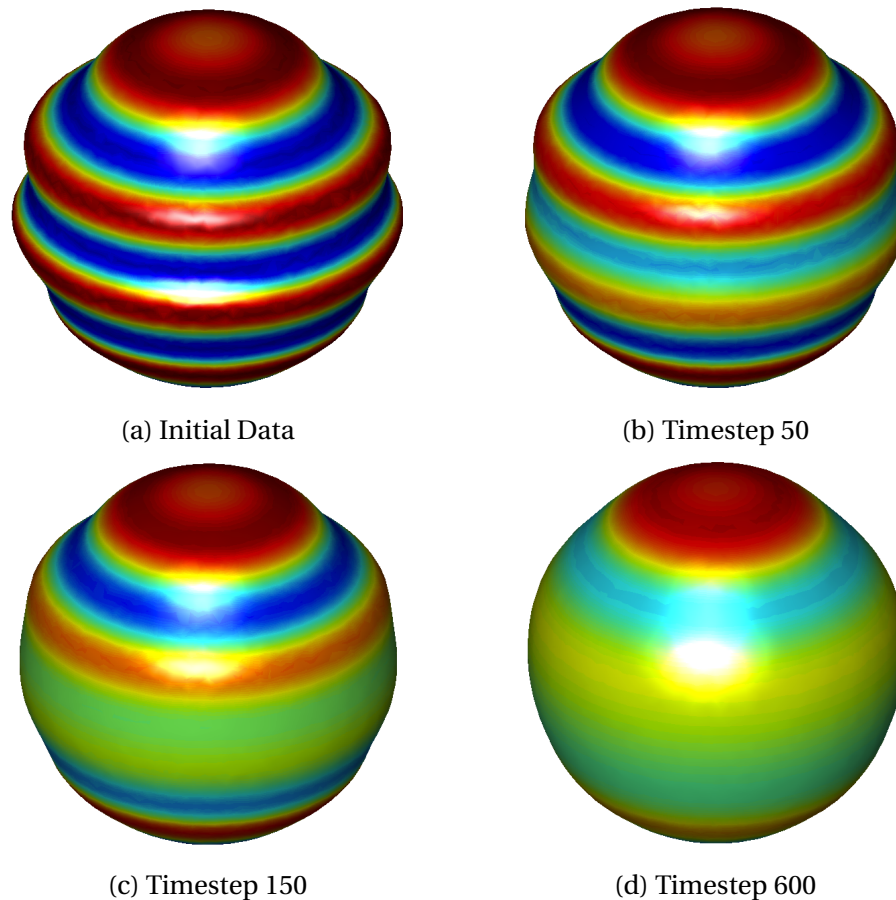
$$-\cot(\varphi/2) \leq \frac{\partial u}{\partial \varphi} \leq \tan(\varphi/2)$$

This condition is easily satisfied by many functions. See Figure 5.1. Note that in non-normalized Ricci flow, the sphere shrinks and becomes smooth in finite time (although becomes round in the limit).

In order to actually evaluate the integral, we first compute the derivative  $\frac{\partial u}{\partial \varphi}$  via the chain rule (the extrinsic spatial derivatives are computed numerically using the midpoints in the finite element basis), and determine the corresponding spherical coordinate  $\varphi$  for each point. Next, we choose a fine mesh for quadrature, evaluate the integrand at the midpoints of each interval, and take the cumulative sum (the trapezoidal rule). Finally, we translate back to the actual points in question again by linear interpolation, and get a collection of new vertices  $(R, \theta, Z)$ . The mesh was provided by CGAL's implicit meshing function, consisting of 3545 vertices on the sphere, with vertex angles being no less than  $30^\circ$ .

## 5.7 Conclusion and Future Work

In this chapter, we have used techniques from nonlinear analysis to apply a finite element method to solving a geometrical evolution equation. We derived this evolution equation by considering Ricci flow on surfaces, which can be reduced to an evolution equation for a conformal factor, since Ricci flow preserves the conformal class of a metric. The nonlinear operator is closely related to that which was analyzed by Kazdan and Warner [59]. Next, we recast our problem into weak form, in preparation for the finite element method, which requires this form, and described the algorithm using Newton's method. Finally, we presented a numerical example, which included an additional step of choosing an embedding and deriving more equations



**Figure 5.1:** Embedded spheres for the metrics  $e^{2u}g$  at time steps 1, 50, 150, and 300 (the timestep  $\Delta t$  is  $1/72000$ ). This is a picture of the *true* geometry, using the embedding equations (5.6.4)-(5.6.5). As one can see, the geometry near the equator dissipates faster than that near the poles, because the value of  $u$  is concentrated over a smaller area, and the factor  $e^{-2u}$  slows the rate of diffusion. Also see the supplementary file `ricci-flow-on-sphere.mov`.

based on that.

An interesting future direction is to provide is to take advantage of the finite element theory to do error analysis. Methods such as the finite element exterior calculus (FEEC) [5, 6, 50] allow discretization of more general differential forms. In addition, [51] provides some results for semilinear operators. With this we can sketch a plan for the error analysis. We continue to work with our semi-discretization in time, and then in space, except now recasting it in the mixed form. This is because

the Newton iterations involved in time evolution are far more stable than that in elliptic problems, for the simple reason that it is adding a small multiple to the identity (actually, at the computational level, we have an extra mass matrix term). The error in our solution therefore breaks up into 5 errors that form a recurrence relation. We suppose that, at timestep  $n$ , we have the true solution  $u^n = u(t^n)$ , and a discrete solution  $u^{n,h}$ . Then the error  $\|u^n - u^{n,h}\|$  breaks up (via the Triangle Inequality) as follows:

1. The error due to the continuous flow acting on two different points  $u^n$  and  $u^{n,h}$ . This is the usual term involved in Grönwall-type inequalities.
2. The error due to approximating the continuous flow with a discrete mapping, starting at the same point  $u^{n,h}$ . This is the usual error introduced by moving from the ODE to methods like Euler, Runge-Kutta, etc.
3. The error due to spatial discretization of the nonlinear operator—the discrete operator is considering the restriction of  $F$  to a finite-dimensional affine subset (initial point plus a finite element space), and orthogonally project the range onto another affine subspace of the same dimension. Then the errors accumulate in the Newton iterations. This splits into two further errors:
  - a. The error resulting from doing Newton iterations (of the continuous operator) on two neighboring start points. This involves various Lipschitz conditions, the inverse of the derivative squared, and the value (i.e., exactly what is necessary for the Kantorovich condition [96, Chapter 10]).
  - b. The error resulting from doing Newton iterations with the discretized operator instead of the continuous operator—it is here where the linearized finite element theory comes in, because the linear operator is  $F'(u_m^{n+1,h})$  and the data is  $F(u_m^{n+1,h})$ , and so is directly estimable using Céa and best

approximation type lemmas. The caveat here is that the “constants” do depend on each iterate  $u$ , but they can be controlled by taking  $L^\infty$  norms and various Lipschitz constants (actually these two errors are in exactly the same spirit as the first two in the above)

4. The error due to cutting off the Newton iterations after only finitely many steps. For sufficiently small timesteps, we can *always* arrange things so that the Kantorovich condition [96, Chapter 10], [53, Section 2.9, which only applies to the finite-dimensional case] holds, so this error will be by far the smallest.

We control the error (3b) using appropriate FEEC estimates, provided, of course, we choose our finite element spaces consistent with what FEEC requires. From those five errors, we form a recurrence relation, and we can estimate the total error via a discrete Grönwall estimate [87].

Another remaining challenge is the question of embedding for numerical simulation and visualization of Ricci flow (and Yamabe flow) in general. This is important because one of our aims is to use visualizations as a method of exploring properties of differential equations and the essential features of geometric flows, in order to generate new conjectures. We want to clearly understand already known solutions as well, since that can only improve our ability to understand how to prove such new conjectures. The version for surfaces is quite unrepresentative, because the flow is actually smooth for all time, and no singularities develop; this is not true in higher dimensions. One way in which singularities are forced to form in higher dimensions is the slowing down of the diffusion—the diffusion is slowed down sufficiently that the concentration terms dominate. In two dimensions, this is represented by the factor  $e^{-2u}$  multiplying the Laplacian, but this is insufficient to cause singularity formation. In higher dimensions, however, the more complicated conformal transformation of the Laplacian yields more concentration terms as well. It should be interesting to

visualize this singularity formation in some manner.

## 5.8 Acknowledgements

This chapter, in full, is currently being prepared for submission for publication. The material may appear as M. Holst and C. Tve, *Finite Elements for the Ricci Flow on Surfaces*. The dissertation author was the primary investigator and author of this paper.

## **Part III**

# **Appendices**

# Appendix A

## Elliptic Equations, Canonical Geometries, and the Robin Mass

As previously mentioned, one of the major motivations of Ricci Flow is the exploration of canonical geometries—a special case of a venerable method of studying evolution equations by their equilibrium solutions. In these appendices, we explore the work of Okikiolu [78, 77] and present some conjectures about some of these equilibrium geometries.

### A.1 Introduction to Spectral Geometry

Spectral geometry is the study of invariants of the Laplace operator. Specifically, those that concern the eigenvalues of the Laplacian, studied in the context of Riemannian geometry. The goal is to develop this theory to gain greater insight into the geometrical meaning of these invariants, which should be useful as much of this subject stands at the crossroads of many different mathematical disciplines such as differential equations, analysis, number theory, differential geometry, algebraic

geometry, etc. The slogan for spectral geometry is “Can you hear the shape of a drum?”, or more formally, *Do the natural frequencies of an object completely specify its shape?* As seen in Chapter 1, the reason why these are “natural frequencies” comes out of solving the wave equation via the method of separation of variables.

The specific problem we have chosen to look at so far is what happens to our invariants when we make a conformal change of metric, and how much of it is a local (geometric) question, and how much of it depends on the global (topological) aspects. Differential operators like the Laplacian, and tensors like the metric, are inherently local objects: its effects only depend on what happens in a vanishingly small neighborhood of a point. However, in solving differential equations, we get integral formulæ, which are inherently global (integration always involves summing over the *whole* manifold). In other words, in determining the inverse of our operators, we somehow involve the global nature. From the standpoint of, for example, the shape of a drum, of course the global aspect has everything to do with how it sounds. The natural frequencies (corresponding to the eigenvalues of the Laplacian) are global quantities, not local ones.

One important link between the spectrum of the Laplacian, and the geometry of our surface, is an invariant called the Robin mass. It is a function on our manifold corresponding to what happens when the appropriate singularity in the Green's function at the diagonal is subtracted off. The integral of the Robin mass is equal to the regularized trace of the inverse of the Laplacian (this is given by summing up the inverses of the eigenvalues, using analytic continuation if necessary—this is why it is called the spectral  $\zeta$  function; in the special case of the circle, it is the Riemann  $\zeta$  function).

The very interesting thing that has been discovered so far is that there are certain canonical metrics which satisfy extremal properties of the Robin mass. For



example, the standard round  $n$ -sphere is a minimum for the Robin mass in its area-preserving conformal class (the area-preserving conformal class of a metric is just the set of all metrics given by multiplying the original metric by a function, and having the same area as the original metric), and moreover, this mass is always positive (establishing that a sphere is optimal in yet another sense).

On the other hand, it has recently been shown by Okikiolu [77] that one can construct a metric with negative mass on a 2-torus, so that in particular, the behavior of the mass is influenced by the genus of the surface in ways that are not straightforward to understand. Okikiolu's proof, unfortunately, does not generalize to (compact Riemann) surfaces of higher genus; we would like to see what happens here and give some conjectures. In particular, this requires examining what happens if we cut out a disc on a Riemann surface and sew in a handle (the standard genus-increasing operation). This in turn requires us to study the Robin mass and its transformation properties on manifolds *with boundary*, which is also a previously unexplored area. The Green's functions for Laplacians on manifolds with boundary, of course, are slightly different, because we have to take into account either Dirichlet or Neumann conditions (the Neumann case is very similar to the case on closed manifolds), so the Robin mass will also satisfy a different behavior with respect to conformal changes. We also would like to examine the Robin mass on the flat and hyperbolic discs and have some form of comparison, which gives us at least two major directions to proceed in: first, to see whether the disk satisfies a similar optimality property, since it is in fact the negative-curvature model space just as the sphere is the positive-curvature model space, and second, to examine the implications for compact Riemann surfaces, if any (since we know the disk is the universal cover of all the compact Riemann surfaces of genus greater than 1).

## A.2 Solving Poisson's Equation

The Laplace operator is ubiquitous in mathematics and the physical sciences [32]. So, of course, mathematicians like to analyze its properties, give some reasonable generalizations, and above all, study its invariants. This gives enormous insight into the nature of the operator. The most basic occurrence, as we've seen in the preceding chapters is, of course, POISSON'S EQUATION: given  $f$ , we would like to solve the equation

$$-\Delta u = f$$

for  $u$ . As a warning, spectral geometers tend to use  $\Delta$  to mean the negative of what we have here; we notate this in order to be consistent with the previous chapters. Of course, we have to specify what domain we're working in and what boundary conditions, in order properly pose the problem. For now, assume we're in a bounded domain  $\Omega \subseteq \mathbb{R}^n$ . Recall that the DIRICHLET PROBLEM, i.e., the task of solving Poisson's equation, subject to DIRICHLET CONDITIONS is: Given  $f : \Omega \rightarrow \mathbb{R}$  and  $\varphi : \partial\Omega \rightarrow \mathbb{R}$  sufficiently nice (say, continuous), we want to find some  $u : \Omega \rightarrow \mathbb{R}$  solving

$$(A.2.1) \quad \begin{cases} \Delta u = f \text{ in } \Omega \\ u|_{\partial\Omega} = \varphi. \end{cases}$$

As we saw, using Sobolev space methods in the previous chapters (or in [30, Ch. 5]), for sufficiently well-behaved  $f$ ,  $\varphi$ , and boundary  $\partial\Omega$ , the solution in fact exists and is unique.

Here we describe a different, more classical approach to the problem. We now consider solving the problem via a (DIRICHLET) GREEN'S FUNCTION, namely an

integration kernel  $G_{\mathcal{D}} : \bar{\Omega} \times \bar{\Omega} \setminus D \rightarrow \mathbb{R}$  where  $D$  is the diagonal  $\{(x, x) : x \in \Omega\}$  such that:

$$(A.2.2) \quad u(x) = \int_{\Omega} G_{\mathcal{D}}(x, y) f(y) dV(y) + \int_{\partial\Omega} \frac{\partial G_{\mathcal{D}}}{\partial n_y}(x, y) \varphi(y) dS(y)$$

where  $\frac{\partial G_{\mathcal{D}}}{\partial n_y}(x, y) = \nabla_y G_{\mathcal{D}}(x, y) \cdot n(y)$  denotes the normal derivative of  $G_{\mathcal{D}}$  with respect to the  $y$  variable,  $dV$  represents the volume element for  $\Omega$ , and  $dS$  the surface element for  $\partial\Omega$ . (We use the subscript  $\mathcal{D}$  to signify that it is the Green's function for *Dirichlet* conditions; but if it is clear we are talking about Dirichlet conditions, we'll drop the subscript  $\mathcal{D}$ ).  $\frac{\partial G}{\partial n_y}$  is called the POISSON KERNEL. The first term solves Poisson's equation  $\Delta u = f$  with homogeneous boundary values, and the second solves Laplace's equation  $\Delta u = 0$  with boundary values  $\varphi$ . For "sufficiently nice"  $\partial\Omega$ , the solution attains the boundary values  $\varphi$  at every point of continuity. Moreover, this solution is *unique*.

It is shown in standard texts on PDES, e.g. [30, 39, 97, 43], that the Green's function itself is a solution to Poisson's equation with Dirichlet conditions, in the sense of distributions [105, 97, 89]:

$$(A.2.3) \quad \begin{cases} \Delta_x G_{\mathcal{D}}(x, y) = \delta_0(x - y) = \delta_y(x) & \text{for all } x, y \in \Omega \\ G_{\mathcal{D}}(x, y) = 0 & \text{for all } x \in \partial\Omega \end{cases}$$

where  $\delta_y$  is the point-mass (the Dirac  $\delta$  "function") at  $y$ . Heuristically, this says that a solution to Poisson's equation with  $\varphi = 0$  is given by resolving into a continuum of impulse solutions for the point masses, each weighted according to  $f$ , and summing.

In summary, for  $\varphi = 0$ , there exists, for every  $f : \Omega \rightarrow \mathbb{R}$  sufficiently regular, a unique  $u : \Omega \rightarrow \mathbb{R}$ , such that  $\Delta u = f$ ,  $u$  vanishes on  $\partial\Omega$ , and is given by

$$u(x) = \int_{\Omega} G(x, y) f(y) dV(y).$$

that is  $G$  is the integration kernel for the *inverse* of the Laplacian, which exists when we restrict to the appropriate (Sobolev) space of functions vanishing on  $\partial\Omega$ .

### A.3 Finding Dirichlet Green's Functions

Green's functions certainly show themselves to be a powerful construct: once we have them, we have solved, in principle, any reasonable Poisson's equation we please. But finding explicit Green's functions can itself be very difficult. The chief thing that makes Green's functions work is that precisely their "singular" behavior on the diagonal: in the neighborhood of  $(p, p)$  for all  $p \in \Omega$ , the Green's function is unbounded. The precise nature of the behavior of  $G(p, q)$  for fixed  $q$  and  $p$  in a neighborhood of  $q$  is a dimension-dependent blow-up: it looks roughly like  $C_n|p - q|^{2-n}$  for  $n \neq 2$  where  $C_n$  is a dimension-dependent constant (involving the volume of the  $n$ -dimensional unit ball and such), and  $-\frac{1}{2\pi} \log(|p - q|)$  for the special case of dimension 2 (a logarithmic singularity). In dimension 1 there is no blow-up; it's just absolute value; i.e. the badness only happens in the derivative. This case is often ignored in books but we'll compute with it because it helps to give the feel of the mass. If the singular behavior were not present, then the Green's identities would instead imply that  $\int G(x, y) f(y) dV(y) = 0$ .

The reason for this is that, what figures deriving the Green's function is the use of the FUNDAMENTAL SOLUTION [30, 89] to the Laplace equation,

$$\Delta\Phi(x) = \delta(x) \text{ in } \mathbb{R}^n$$

which yields the radial solutions

$$\Phi(x) = \begin{cases} C_n |x|^{2-n} & n \neq 2 \\ -\frac{1}{2\pi} \log|x| & n = 2 \end{cases}$$

Again, this is derived in texts on PDES. This is related to the Green's functions via GREEN'S REPRESENTATION FORMULA [30, 46, 39]:

$$u(x) = \int_{\Omega} \Phi(x-y) \Delta u(y) dV(y) + \int_{\Omega} \left( \frac{\partial}{\partial n_y} \Phi(x-y) u(y) - \frac{\partial u}{\partial n}(y) \Phi(x-y) \right) dS(y)$$

So if we substitute  $f$  for  $\Delta u$  in the volume integral and  $\varphi$  for  $u$  on the boundary integral in the above, then this is almost what we want; it isn't quite because if we are only given  $f$  and  $\varphi$ , we still don't know  $\partial u / \partial n$ . The idea is to introduce a CORRECTOR FUNCTION  $h$  as follows:

$$(A.3.1) \quad \begin{cases} \Delta_x h(x, y) = 0 & \text{for all } x \in \Omega \\ h(x, y) = -\Phi(x-y) & \text{for all } x \in \partial\Omega \end{cases}$$

Note that  $h$  in the above will satisfy Laplace's equation in  $x$  throughout the whole interior of  $\Omega$ , not just on  $\Omega \setminus \{y\}$ . Then, assuming that the  $h$  actually exists, we have that the Green's function satisfies  $G(x, y) = \Phi(x-y) + h(x, y)$ . Using  $G(x, y)$  in place of  $\Phi(x-y)$  in the formulæ above conveniently eliminates the term with  $\frac{\partial u}{\partial n}$  and yields the formula (A.2.2), so that it in fact works as advertised. Since the  $h$  is perfectly well-defined on the diagonal, it follows that behavior  $G(x, y)$  at a singularity is exactly the same for  $\Phi(x-y)$ . We should note another important property of  $G$ , namely that it is symmetric in the variables  $x$  and  $y$ :  $G(x, y) = G(y, x)$ . By the similar symmetry for  $\Phi(x-y)$  this also carries over to  $h$ .

We can now define the Dirichlet Robin mass.

**A.3.1 Definition.** The DIRICHLET ROBIN MASS for  $\Delta$  on a domain  $\Omega \subseteq \mathbb{R}^n$  is just the corrector function for the Dirichlet Green's function, at the diagonal:

$$m_{\mathcal{D}}(x) := m(x) := h(x, x) = \lim_{y \rightarrow x} (G(x, y) - \Phi(x - y)).$$

Again, recall that  $h$  does not exhibit any bad behavior at the diagonal. In other words, the Robin mass is the leftover when the singular part of the Green's function is subtracted off, at the diagonal. We'll give some examples in the next section.

## A.4 The Dirichlet Problem

Let  $(M, g)$  be a Riemannian manifold with boundary. We can in fact define a Laplace operator on  $M$  which is the appropriate analogue of the version on  $\mathbb{R}^n$ . There is a version with the Christoffel symbols  $\Gamma_{ij}^k$ , but there is also a more elegant formula. Note that  $g$  defines a volume element, which looks like, in coordinates  $(x^i)$ ,  $dV = \sqrt{|\det(g_{ij})|} |dx^1 \wedge \cdots \wedge dx^n|$ . We write  $\sqrt{g} = \sqrt{|\det(g_{ij})|}$ , which is unambiguous because you can't take the square root of a tensor anyway. The Laplacian is defined as

$$\Delta u = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left( \sqrt{g} g^{ij} \frac{\partial u}{\partial x^j} \right)$$

If  $M$  has a boundary, we can also adapt the forgoing theory accordingly: we can solve Poisson's equation, subject to Dirichlet conditions. The issue with Green's functions is a little stickier, because the fundamental solutions, if they exist, are not as clean to write as the one on  $\mathbb{R}^n$ . We still may speak of point masses, though they are slightly tricky because the concept of the measure which assigns a set containing the point in question, the value 1, and 0 otherwise, has nothing to do with the metric,

but defining a delta “function” which can appear under the integral sign does involve the metric as the volume element. However we shall just suppose we have the right distribution-theoretic approach (the essential point is that  $\delta$  transforms with factors of  $\sqrt{g}$  under coordinate changes in an opposite manner as the volume element. This is not surprising as “integration of a delta function against the volume element” should yield 1). In summary, the Green’s function is a function defined on  $M \times M$  which solves

$$(A.4.1) \quad \begin{cases} \Delta_x G(x, y) = \delta_y(x) & \text{for all } x \in \text{Int } M \\ G(x, y) = 0 & \text{for all } x \in \partial M, \end{cases}$$

for all  $y \in M$ . At the diagonal,  $G(x, y)$  has a singularity which can be expressed in some polynomial of the reciprocal of the Riemannian geodesic distance  $1/d(x, y)$ , of degree up to  $n - 2$ , and where the “constant term” is really logarithmic on the distance. If we are on a flat Riemannian manifold, the Green’s function is similar to the case of  $\mathbb{R}^n$ , namely, it blows up like  $d(x, y)^{2-n}$  for  $n \neq 2$  and like logarithm for  $n = 2$ . The Dirichlet Robin mass is then again defined to be the leftover part after all the singularities is subtracted off. We shall assume these results for now. Examples are also in the next section. We should note that just as in the case of domains in  $\mathbb{R}^n$ , solutions to the Dirichlet problem are unique: once we prescribe boundary values,  $\Delta$  is invertible. What this means is, if we restrict to all functions which vanish at the boundary, we have that the inverse satisfies

$$(\Delta_g^{-1} f)(p) = \int_M G_g(p, q) f(q) dV_g(q).$$

If we examine what happens to the Green’s function at the diagonal, we have the

expansion

$$G(p, q) = C_{n,0}d_g(p, q)^{2-n} + C_{n,1}d_g(p, q)^{2-n+1} + \dots \\ + C_{n,n-2} \log(d_g(p, q)) + m_g(p) + o(d_g(p, q))$$

where if  $n = 2$  we use the logarithm, for  $p$  sufficiently close to  $q$ ,  $m_g$  is the Robin mass, and the  $C_{n,j}$  are dimension- and metric-dependent coefficients. This is how we define the (Dirichlet) Robin mass on a manifold. More explicitly,

$$m_{\mathcal{D},g}(p) := m_g(p) := \lim_{q \rightarrow p} \left( G(p, q) - \sum_{i=0}^{n-3} C_{n,i} d_g(p, q)^{2-n+i} - C_{n,n-2} \log(d_g(p, q)) \right)$$

For the case of surfaces, our prime area of interest, of course, we only have a log term and  $C_{2,0} = -\frac{1}{2\pi}$ , that is,

$$m_{\mathcal{D},g}(p) = \lim_{q \rightarrow p} \left( G(p, q) + \frac{1}{2\pi} \log(d_g(p, q)) \right).$$

That this is actually the expansion of the Green's function can be seen directly by using polar normal coordinates: after subtracting off the logarithmic singularity, we get a harmonic function, so therefore it has a Taylor expansion in the radial coordinate which is precisely geodesic distance.

Actually, there is more well-known kind of Robin mass, defined on manifolds of even dimension, which arises from a "Green's function" that always has a logarithmic singularity and nothing else; it generalizes the Laplace operator in a different direction. The difference is now that the so-called Green's function is the integral kernel for (the inverse of) a different differential operator, one of order  $n$ , called the PANEITZ OPERATOR. It is a differential operator  $\square_g$  which transforms under conformal changes of metric as  $\square_{Fg} = F^{-1}\square_g$ . This is in contrast to  $\Delta$  which transforms as  $\Delta_{Fg} = F^{-n/2}\Delta +$



many other unpleasant terms (and is of course the same as  $\square_{Fg}$  if  $n = 2$ ).

## A.5 The Neumann Problem

We haven't mentioned what happens on manifolds with boundary where we prescribe the normal derivative, namely the term  $\frac{\partial u}{\partial n}$ , rather than the boundary values themselves. This is called the Neumann problem. The interpretation of this, for a zero normal derivative, is the quantity that  $u$  represents does not "flow" across the boundary. It turns out that this determines a solution to the Laplacian only up to a constant; i.e. the Laplace operator has a nontrivial kernel when restricted to the space of functions with vanishing normal derivative.

Now if our manifold is compact without boundary, there are no boundary conditions to satisfy at all. It is easy to show that harmonic functions on a closed manifold are just constants (this is the analogue of the Liouville theorem in complex analysis). In other words, the Laplacian has the same kernel as it would have if we were considering a manifold with boundary and Neumann boundary condition. By Stokes' theorem, we have  $\int_M \Delta u \, dV = \int_{\partial M} \frac{\partial u}{\partial n} \, dS = 0$  (for either Neumann boundary conditions, where  $\frac{\partial u}{\partial n} = 0$ , or if  $M$  is closed so  $\partial M = \emptyset$ ), for all  $C^2$  functions  $u$ , so that we should restrict the range to only those functions whose total integral is 0. This is called NORMALIZING. There is an analogous notion of Green's functions here, given by

$$(A.5.1) \quad \Delta_x G_{\mathcal{N},g}(x, y) = \delta_y(x) - \frac{1}{V_g} \text{ for all } x, y \in M$$

where  $V_g = \int_M dV_g$ , the volume of  $M$  with respect to the volume element  $dV_g$ . What this does to a function is that it inverts the Laplacian and subtracts off the average value of the solution; in other words, all solutions are normalized to have zero average. The subscript  $\mathcal{N}$  means Neumann conditions, and this includes the case if  $M$  is closed

(we drop the subscript when it is clear what kind of Green's functions we are working with). We also will drop the subscript  $g$  from time to time if the metric is clear.

Despite the extra subtraction of the volume, the astute reader may note this still does not uniquely specify  $G_{\mathcal{N}}$ . The solution is to make  $G_{\mathcal{N}}$  itself have total integral 0 over the whole manifold, in one of the variables. We also must separately enforce the symmetry of  $G_{\mathcal{N}}$  in its two variables (it was automatic for the Dirichlet case). Giving it total integral zero amounts to specifying that the kernel of the inverse operator is also the constants. This is a natural consequence of the weak solution theory using Hilbert space methods studied in §1.7.

Abusing notation, we shall still write (or in fact *define*) for  $f$  with vanishing normal derivative (if there is a boundary at all—otherwise, for arbitrary  $f$  in a suitable Sobolev space)

$$(\Delta_{\mathcal{N},g}^{-1}f)(p) := \int_M G_{\mathcal{N},g}(p, q) f(q) dV_g(q).$$

It abuses notation because  $\Delta_g$  is not one-to-one, so has no true inverse (and we'll quickly start losing subscripts at this point). For  $u$  with vanishing normal derivative, we have

$$\Delta^{-1}\Delta u = u - \frac{1}{V_g} \int_M u dV_g.$$

namely,  $\Delta^{-1} \circ \Delta$  is the operator which subtracts off the average value of  $u$ . Similarly,

$$\Delta\Delta^{-1}f = f - \frac{1}{V} \int f dV_g.$$

That is to say,  $\Delta$  and  $\Delta^{-1}$  are inverses whenever all functions in question have vanishing total integral and normal derivative.

This can be heuristically seen by “integration by parts” with a  $\delta$  “function” (and

also the additional  $-1/V$  term):

$$\begin{aligned} (\Delta^{-1}\Delta u)(p) &= \int_M G(p, q)\Delta u(q) dV(q) = \int_M \Delta_q G(p, q)u(q) dV(q) \\ &= \int_M \left(-\frac{1}{V}\right)u(q) dV(q) + \int_M \delta_p(q)u(q) dV(q) = u(p) - \frac{1}{V} \int_M u dV \end{aligned}$$

(there are no boundary terms in switching the Laplacian over, because either the normal derivative of  $u$  vanishes there, or the boundary doesn't exist).

Note that for the (true) Neumann problem (i.e. when  $\partial M \neq \emptyset$ ), there is an additional compatibility condition we must have, that is not present in either the Dirichlet problem or the problem on closed manifolds: the volume integral of  $f$  must equal the surface integral of the prescribed normal derivative. Of course if we restrict to functions with vanishing integral, and consider a zero normal derivative, this condition is satisfied (besides, without those vanishing boundary conditions, integration against  $G$  no longer inverts the operator even in the Dirichlet case; remember the true representation formula with arbitrary boundary conditions also involves an additional surface integral term): If  $f : \Omega \rightarrow \mathbb{R}$  and  $\psi : \partial\Omega \rightarrow \mathbb{R}$ , such that  $\int_M f dV = \int_{\partial M} \psi dS$ , then

$$u(x) = \int_M G(x, y)f(y) dV(y) - \int_{\partial M} G(x, y)\psi(y) dS(y)$$

solves our problem. Again arguing heuristically with  $\delta$ , we have

$$\begin{aligned} \Delta u(x) &= \int_\Omega \Delta_x G(x, y)f(y) dV(y) - \Delta_x \int_{\partial\Omega} G(x, y)\psi(y) dS(y) \\ &= f(x) - \frac{1}{V_g} \int_M f(y) dV(y) - \Delta_x \int_{\partial\Omega} G(x, y)\psi(y) dS(y) \\ &= f(x) - \int_{\partial\Omega} (\Delta_x G(x, y) + 1/V_g)\psi(y) dS(y) = f(x), \end{aligned}$$

the last interchange giving 0 because  $x$  is an interior point and so the singularity is

never encountered on the integral over the surface. The Green's function exhibits exactly the same kind of singularity as it does in the Dirichlet case, so we can define a Robin mass for it:

$$m_{\mathcal{N},g}(p) := m_g(p) := \lim_{q \rightarrow p} \left( G(p, q) - \sum_{i=0}^{n-3} C_{n,i} d_g(p, q)^{2-n+i} - C_{n,n-2} \log(d_g(p, q)) \right)$$

where of course if  $n = 2$  it only has a logarithm (the case we shall be most interested in).

# Appendix B

## Examples of Green's Functions and Robin Masses

Here, we do some calculations and to get an idea of what this Robin mass is. We start off with the simplest case: one dimension. The “singularity” turns out to actually be a corner (so the function is equal to its limiting value there but not a continuous first derivative); this means the Robin mass can be obtained by directly setting  $x = y$ .

### B.1 In One Dimension

**B.1.1 Example** (The Interval). In one dimension, Green's functions are relatively easy to solve for, because the Laplace equation is an ODE. Let  $I = [-\pi, \pi]$  (this will be convenient because we will re-use many of our calculations for the circle). So for the

Dirichlet problem on the interval, we are looking for  $u$  satisfying

$$(B.1.1) \quad \begin{cases} -u''(x) = f(x) & x \in I \\ u(-\pi) = a \\ u(\pi) = b. \end{cases}$$

To find the fundamental solution  $\Phi$  on  $\mathbb{R}$ , we solve

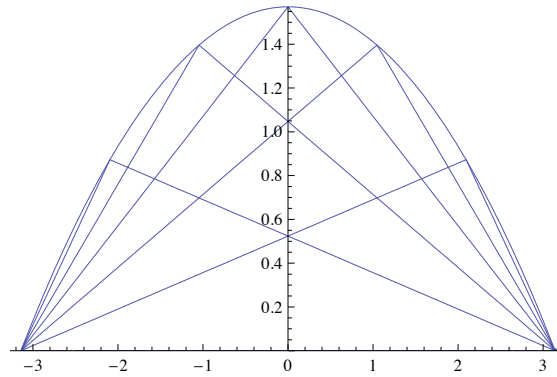
$$-\Phi''(x) = \delta(x).$$

But  $\delta$  as we should recall is the distributional derivative of the unit step function.

$$U(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0. \end{cases}$$

Integrating once,  $\Phi'(x) = -U(x) + C$ , and twice,  $\Phi(x) = -xU(x) + Cx + D$ . If we relate this to what happens on the unit “ball” in  $\mathbb{R}$  namely  $[-1, 1]$ , we should remark that the “area” of the unit “sphere” in  $\mathbb{R}$  is 2 (sum of the two points  $-1$  and  $1$  each having counting measure 1). So we should set  $\Phi(-1) = -\frac{1}{2}$ . So  $-C + D = -\frac{1}{2}$  since  $U(-1) = 0$ , so that  $D = C - \frac{1}{2}$ . We should also have  $\Phi(1) = -\frac{1}{2}$ , so that  $-1 + C + C - \frac{1}{2} = -\frac{1}{2}$ . This says  $2C - 1 = 0$  or  $C = \frac{1}{2}$  and  $D = 0$ . So therefore

$$\Phi(x) = -xU(x) + \frac{1}{2}x = \begin{cases} \frac{1}{2}x & \text{if } x < 0 \\ \frac{1}{2}x - x = -\frac{1}{2}x & \text{if } x > 0 \end{cases}$$



**Figure B.1:** Graph of the Green's function for a few values of  $y$ , along with the Robin mass.

i.e.  $\Phi(x) = -\frac{1}{2}|x|$ . The corrector function  $h$  therefore solves

$$-\frac{\partial^2}{\partial x^2} h(x, y) = 0$$

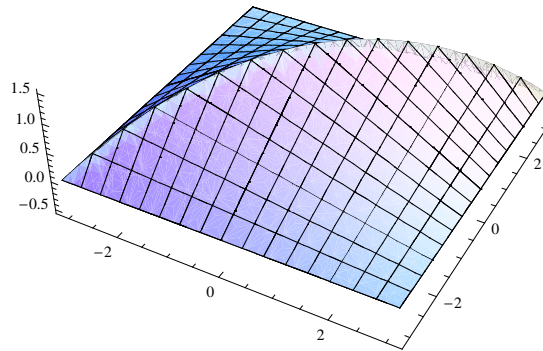
with  $h(-\pi, y) = \frac{1}{2}|\pi + y|$  and  $h(\pi, y) = \frac{1}{2}|\pi - y|$ . Integrating twice,  $h(x, y) = Ax + B$ . Therefore  $-\pi A + B = \frac{1}{2}|\pi + y|$  and  $\pi A + B = \frac{1}{2}|\pi - y|$ . Adding the equations, this says  $B = \frac{1}{4}(|\pi + y| + |\pi - y|)$  and  $A = \frac{1}{4\pi}(|\pi - y| - |\pi + y|)$ . Since  $y \in [-\pi, \pi]$ , these simplify considerably, for we can remove the absolute value signs:  $A = -\frac{y}{2\pi}$  and  $B = \frac{\pi}{2}$  (we would expect that the  $B$  not depend on  $y$  because the Green's function is symmetric in  $x$  and  $y$ ). So we have, therefore,

$$h(x, y) = -\frac{xy}{2\pi} + \frac{\pi}{2}$$

and as a bonus, the Dirichlet Robin mass of the interval is evaluating at  $(x, x)$ :

$$m_{\mathcal{D}}(x) = -\frac{x^2}{2\pi} + \frac{\pi}{2}.$$

This makes the total Green's function



**Figure B.2:** Full graph of the Green's function in two variables.

$$(B.1.2) \quad G(x, y) = -\frac{1}{2}|x - y| - \frac{xy}{2\pi} + \frac{\pi}{2}.$$

The graph of the Green's function for a fixed  $y$  is a triangle with base vertices at  $(\pm\pi, 0)$  and peak at  $(y, y^2/2)$ , i.e. the peak as  $y$  varies is precisely the Robin mass evaluated at that point (see Figures B.1 and B.2). Integrating the Robin mass, we have

$$\int_{-\pi}^{\pi} m(x) dx = -\int_{-\pi}^{\pi} \frac{x^2}{2\pi} dx + \pi^2 = -\frac{\pi^2}{3} + \pi^2 = \frac{2\pi^2}{3}.$$

This is equal to  $4\zeta(2)$ ; its relation to the Riemann  $\zeta$  function is not coincidental. We do not explain it here; instead, we refer the reader to the research literature in spectral geometry [76, 100, 101, 70, 71].

We now examine how differing 1-dimensional topologies can change things.

**B.1.2 Example (The Circle).** Now we give an example on a closed 1-manifold, the only connected example of which is a circle. The computation is remarkably similar to that of the interval, precisely because it is equivalent to enforcing periodic boundary conditions ( $f(-\pi) = f(\pi)$  instead of requiring the value at the endpoints to actually be zero). However, we do have that extra volume term to take into account for normalization.



That is, we solve

$$\frac{\partial^2}{\partial x^2} G(x, y) = \frac{1}{2\pi} - \delta(x - y).$$

(we remind the reader that  $\frac{\partial^2}{\partial x^2}$  is the negative of the Laplacian in our sign convention).

Integrating twice, we have

$$G(x, y) = \frac{x^2}{4\pi} - (x - y)U(x - y) + B(y)x + C(y)$$

Plugging in  $G(-\pi, y) = G(\pi, y)$  we have

$$\frac{\pi^2}{4} - B(y)\pi + C(y) = \frac{\pi^2}{4} - (\pi - y) + B(y)\pi + C(y)$$

or  $-B(y)\pi = y + B(y)\pi - \pi$ . This says  $2B(y)\pi = \pi - y$  or

$$B(y) = \frac{\pi - y}{2\pi}.$$

Periodic boundary conditions cannot determine  $C$  since a function that does not depend on  $x$  is, rather trivially, periodic in  $x$ . We'll set  $C(y) = \frac{y^2}{4\pi} - \frac{y}{2} + D$ , because this will make  $G$  symmetric in  $x$  and  $y$ ; the constant  $D$  will be determined as the constant that makes the average value zero.

This means

$$\begin{aligned}
 G(x, y) &= \frac{x^2 + y^2}{4\pi} - (x - y)U(x - y) + \frac{\pi - y}{2\pi}x - \frac{y}{2} + D \\
 &= \begin{cases} \frac{x^2 + y^2}{4\pi} + \frac{\pi - y}{2\pi}x - \frac{y}{2} + D & \text{if } x < y \\ \frac{x^2 + y^2}{4\pi} - (x - y) + \frac{\pi - y}{2\pi}x - \frac{y}{2} + D & \text{if } x > y \end{cases} \\
 &= \begin{cases} \frac{x^2 + y^2}{4\pi} - \frac{xy}{2\pi} - \frac{y}{2} + \frac{x}{2} + D & \text{if } x < y \\ \frac{x^2 + y^2}{4\pi} + y - x - \frac{xy}{2\pi} - \frac{y}{2} + \frac{x}{2} + D & \text{if } x > y \end{cases} \\
 &= \begin{cases} \frac{x^2 + y^2}{4\pi} - \frac{xy}{2\pi} + \frac{1}{2}(x - y) + D & \text{if } x < y \\ \frac{x^2}{4\pi} - \frac{xy}{2\pi} + \frac{1}{2}(y - x) + D & \text{if } x > y \end{cases} \\
 &= \frac{x^2 + y^2}{4\pi} - \frac{xy}{2\pi} - \frac{1}{2}|x - y| + D.
 \end{aligned}$$

Note that this differs from the case for the interval only in the fact that the term  $\frac{x^2 + y^2}{4\pi}$  is replaced by the constant  $\frac{\pi}{2}$ . The Robin mass then satisfies  $m(x) \equiv D$ : it is constant. To find  $D$ , we simply write the integral of  $G$  with respect to one of its variables: we want

$$0 = \int_{-\pi}^{\pi} \left( \frac{x^2 + y^2}{4\pi} - \frac{xy}{2\pi} - \frac{1}{2}|x - y| + D \right) dx$$

or

$$\begin{aligned}
2\pi D &= -\int_{-\pi}^{\pi} \frac{x^2}{4\pi} dx - \frac{y^2}{2} + \frac{1}{2} \int_{-\pi}^{\pi} |x-y| dx \\
&= -\frac{\pi^2}{6} - \frac{y^2}{2} + \frac{1}{2} \int_{-\pi}^y (y-x) dx + \frac{1}{2} \int_y^{\pi} (x-y) dx \\
&= -\frac{\pi^2}{6} - \frac{y^2}{2} + \frac{1}{2} \left( xy - \frac{x^2}{2} \right) \Big|_{-\pi}^y + \frac{1}{2} \left( \frac{x^2}{2} - xy \right) \Big|_y^{\pi} \\
&= -\frac{\pi^2}{6} - \frac{y^2}{2} + \frac{y^2}{2} - \frac{y^2}{4} + \frac{\pi y}{2} + \frac{\pi^2}{4} + \frac{\pi^2}{4} - \frac{\pi y}{2} - \frac{y^2}{4} + \frac{y^2}{2} \\
&= -\frac{\pi^2}{6} + \frac{\pi^2}{2} = \frac{\pi^2}{3}.
\end{aligned}$$

Therefore  $D = \frac{\pi}{6}$  and finally:

$$(B.1.3) \quad G(x, y) = \frac{1}{4\pi}(x^2 + y^2) - \frac{xy}{2\pi} - \frac{1}{2}|x-y| + \frac{\pi}{6} = \frac{1}{4\pi}(x-y)^2 - \frac{1}{2}|x-y| + \frac{\pi}{6}$$

and the Robin mass is  $m(x) \equiv \frac{\pi}{6}$ . Integrating this constant mass, we get  $\frac{\pi^2}{3} = 2\zeta(2)$ . It should also be noted that metrics of constant Robin mass are, in some sense, nicer; its constancy on round spheres of all dimensions is instrumental in showing that the round metric satisfies (yet another) extremal property.

**B.1.3 Example** (The Neumann Problem on the Interval). In the Neumann problem, we have the same singularity. In 1 dimension, the normal derivative at the boundary is just the ordinary derivative at the right endpoint, and the negative of the ordinary derivative at the left endpoint (since pointing outward for an interval is in the negative direction for the left endpoint as in Table 1.2e).

The “volume” of the interval  $[-\pi, \pi]$  is of course just  $2\pi$ . In other words, we are solving

$$\frac{\partial^2}{\partial x^2} G(x, y) = \frac{1}{2\pi} - \delta(x-y).$$

This is exactly the same situation as for the circle, except now we have to satisfy the condition  $-\frac{\partial}{\partial x}G(-\pi, y) = 0$  and  $\frac{\partial}{\partial x}G(\pi, y) = 0$ . Integrating twice gives us

$$G(x, y) = \frac{x^2}{4\pi} - (x - y)U(x - y) + B(y)x + C(y).$$

But of course actually we went a little far by integrating twice. What about just once?

We'll need it for the derivative at the endpoints:

$$\frac{\partial}{\partial x}G(x, y) = G_x(x, y) = \frac{x}{2\pi} - U(x - y) + B(y)$$

Since  $-\pi$  certainly is less than  $y \in [-\pi, \pi]$ , we have that  $G_x(-\pi, y) = -\frac{1}{2} + B(y) = 0$ . Therefore  $B(y) = \frac{1}{2}$ . For the other endpoint, since  $\pi \geq y$ ,  $U(\pi - y) = 1$ , so  $G_x(\pi, y) = \frac{1}{2} - 1 + B(y) = 0$  which again says  $B(y) = \frac{1}{2}$ ; this is good news, since it shows that the Neumann condition for  $G$  is self-consistent. Taking a cue from the calculation for a circle, we enforce symmetry by trying  $C(y) = \frac{y^2}{4\pi} - \frac{1}{2}y + D$ , and then determine  $D$  using the total integral.

Now this means

$$\begin{aligned} G(x, y) &= \frac{x^2 + y^2}{4\pi} - (x - y)U(x - y) + \frac{1}{2}(x - y) + D \\ &= \begin{cases} \frac{x^2 + y^2}{4\pi} + \frac{1}{2}(x - y) + D & \text{if } x < y \\ \frac{x^2 + y^2}{4\pi} - (x - y) + \frac{1}{2}(x - y) + D & \text{if } x > y \end{cases} \\ &= \frac{x^2 + y^2}{4\pi} - \frac{1}{2}|x - y| + D. \end{aligned}$$

which is almost like the circle case except it's lacking a  $xy$  term. To find  $D$ , we integrate. However recall in the calculation for the circle, that the integral of the  $xy$  term vanishes because it is an odd function of  $x$ , integrated over the origin-symmetric interval  $[-\pi, \pi]$ .

So this means the integral ends up being exactly the same, and  $D = \frac{\pi}{6}$ . Therefore,

$$G_{\mathcal{N}}(x, y) = \frac{x^2 + y^2}{4\pi} - \frac{1}{2}|x - y| + \frac{\pi}{6}$$

and the Robin mass is

$$m_{\mathcal{N}}(x) = \frac{x^2}{2\pi} + \frac{\pi}{6}$$

which, unlike the version on the circle, is not constant. Also note the squared term is positive instead of negative in the Dirichlet case. Also interesting is that, remembering that the Dirichlet Robin mass consists solely of an  $xy$  term plus a constant, whereas this Neumann Robin mass has only the sum of squares, and on the circle, both kinds of quadratic term appear.

Finally, as before, we see what happens when we integrate the mass:

$$\int_{-\pi}^{\pi} m(x) dx = \int_{-\pi}^{\pi} \frac{x^2}{2\pi} dx + \frac{\pi^2}{3} = \frac{\pi^2}{3} + \frac{\pi^2}{3} = \frac{2\pi^2}{3}.$$

which is identical to the Dirichlet case, equal to  $4\zeta(2)$ .

## B.2 Two-Dimensional Examples

In two dimensions, things are more complicated. One immensely important tool we have in two dimensions is complex analysis; we make liberal use of it in this section. The Dirichlet Green's Function for the Euclidean unit disk  $\mathbb{D}$  in  $\mathbb{R}^2$  actually is very nice, because we can use the techniques of complex analysis to compute it. But first, we should recall that harmonicity is invariant under conformal mappings, that is, if  $f : \mathbb{D} \rightarrow \mathbb{D}$  is bijective and holomorphic, then  $u : \mathbb{D} \rightarrow \mathbb{R}$  is harmonic if and only if  $u \circ f$  is. More generally, if  $f : \mathbb{D} \rightarrow \mathbb{D}$  is merely holomorphic (conformal but not bijective),  $u \circ f$  is harmonic whenever  $u$  is. This is easy to prove, especially in complex

coordinates (recall that

$$\Delta = -4 \frac{\partial^2}{\partial z \partial \bar{z}}$$

in complex coordinates). Namely, if  $u : \mathbb{D} \rightarrow \mathbb{R}$  is  $C^2$ , we have, writing  $w = f(z)$  for convenience, and recalling  $\frac{\partial f}{\partial \bar{z}} = \frac{\partial \bar{f}}{\partial z} = 0$  and  $\frac{\partial}{\partial z} \left( \frac{\partial f}{\partial z} \right) = 0$  by analyticity:

$$\begin{aligned} \Delta(u \circ f) &= -4 \frac{\partial^2}{\partial z \partial \bar{z}} (u \circ f) = -4 \frac{\partial}{\partial z} \left( \frac{\partial u}{\partial w} \frac{\partial f}{\partial \bar{z}} + \frac{\partial u}{\partial \bar{w}} \frac{\partial \bar{f}}{\partial z} \right) = -4 \frac{\partial}{\partial z} \left( \frac{\partial u}{\partial \bar{w}} \frac{\partial f}{\partial z} \right) \\ &= -4 \left( \left( \frac{\partial^2 u}{\partial \bar{w}^2} \frac{\partial \bar{f}}{\partial z} + \frac{\partial^2 u}{\partial w \partial \bar{w}} \frac{\partial f}{\partial z} \right) \frac{\partial \bar{f}}{\partial z} + \frac{\partial u}{\partial \bar{w}} \frac{\partial}{\partial z} \left( \frac{\partial f}{\partial z} \right) \right) = -4 \frac{\partial^2 u}{\partial w \partial \bar{w}} \left| \frac{\partial f}{\partial z} \right|^2 = |f'|^2 \Delta u. \end{aligned}$$

Actually we didn't use the fact that our domain was the disk  $\mathbb{D}$ , only that it was in the complex plane (with the Euclidean metric). The fundamental solution in  $\mathbb{R}^2$ , as derived in many a PDE text, is

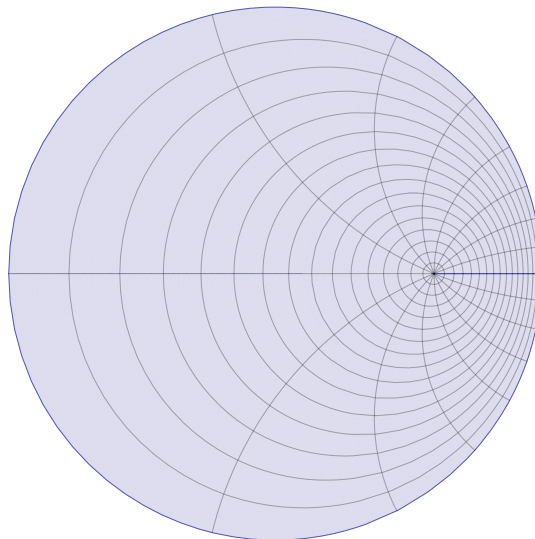
$$(B.2.1) \quad \Phi(z) = -\frac{1}{2\pi} \log|z|.$$

With these preliminary results we now are ready to begin doing more interesting things.

**B.2.1 Example** (The Euclidean Unit Disk  $\mathbb{D}$ ). With the Fundamental solution, since  $\log 1 = 0$ , we have already found the Dirichlet Green's function at 0, namely,

$$G(z, 0) = -\frac{1}{2\pi} \log|z|,$$

since  $\Delta_z G(z, 0) = 0$  in the punctured disk, and it is 0 on the boundary circle. Now since we are trying to solve the general equation  $\Delta_z G(z, w) = 0$  in  $\mathbb{D} \setminus \{w\}$  (this is the cheap way of getting around the use of distribution theory and the  $\delta$  function), and  $G(z, w) = 0$  for all  $z \in S^1 = \partial\mathbb{D}$ , what we could do is take advantage of conformal invariance:



**Figure B.3:** Transformation  $f_w$  for  $w \approx -0.6$  given by its action on a polar grid.

find a conformal map  $f_w : \mathbb{D} \rightarrow \mathbb{D}$  taking  $w$  to 0 and preserving the boundary  $S^1$ ; then defining

$$G(z, w) = G(f_w(z), 0) = -\frac{1}{2\pi} \log |f_w(z)|,$$

we have  $G(z, w)$  is also harmonic in  $z$ , and also is 0 on the boundary. Can we find a conformal map that does this? In fact, yes we can; this is just the much-heralded theory of the (conformal) automorphisms, or Möbius transformations, of the disk, which has prominent application in hyperbolic geometry (we shall also see what happens on the hyperbolic disk—and find lots of interesting stuff there, too!). The map is as follows:

$$(B.2.2) \quad f_w(z) = \frac{z - w}{1 - \bar{w}z}.$$

It turns out that all conformal (not necessarily bijective) self-maps of the disk are products of  $f_w$ 's for different  $w$ 's, possibly also with rotations. The function  $f_w$  is called a BLASCHKE FACTOR. See Figure B.3 for an example of what the transformation  $f_w$  does to a polar grid. Also see the figures on the next page illustrating the conformal

map on a very interesting planar subset of the disk (cf. VI. Arnol'd's "cat map" and the fact he uses a cat-like shape to demonstrate the effects of mappings):

The upshot of all that exploration is that now we can write the Green's function explicitly:

$$(B.2.3) \quad G(z, w) = -\frac{1}{2\pi} \log \left| \frac{z-w}{1-\bar{w}z} \right| = -\frac{1}{2\pi} \log |z-w| + \frac{1}{2\pi} \log |1-\bar{w}z|.$$

We rewrote the logarithmic term so we can see exactly where the fundamental solution comes in, and hence which term to cancel to find the Robin mass. Therefore

$$(B.2.4) \quad m(z) = \lim_{w \rightarrow z} \frac{1}{2\pi} \log |1-\bar{w}z| = \frac{1}{2\pi} \log (1-|z|^2).$$

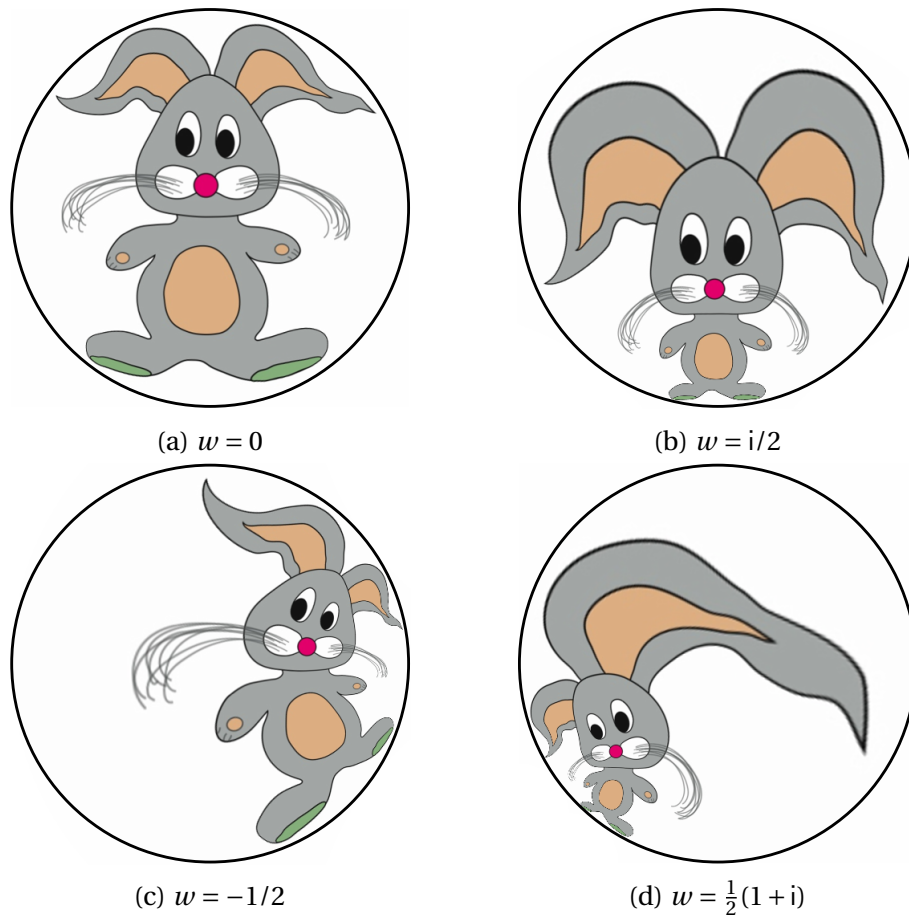
This is not constant; and in fact it blows up at the boundary.

Now before we proceed, we can derive some general formulæ involving conformal changes of metric on surfaces. This will enable us to calculate the Robin mass on the hyperbolic disk (in fact, we shall be led to it by asking how can we conformally change the metric on the disk to get a constant Robin mass!)

**B.2.2 Definition.** Recall that we say  $\tilde{g}$  is conformal to  $g$  if  $\tilde{g} = e^{2u}g$  for  $u \in C^\infty(M)$ . If  $(M, g)$  and  $(N, h)$  are manifolds and  $F : M \rightarrow N$  is a smooth map, we say  $F$  is a CONFORMAL TRANSFORMATION if  $F$  is a diffeomorphism and  $F^*h = e^{2u}g$  for some  $u \in C^\infty(M)$ . For example, if  $F : \Omega \rightarrow \Omega$  is a biholomorphism of a domain in the complex plane, then  $F^*dz = dF = F'(w)dw$  and  $F^*d\bar{z} = \overline{F'(w)}d\bar{w}$ . So  $F^*(dzd\bar{z}) = |F'(w)|^2dw d\bar{w}$ , i.e.  $F$  is a conformal transformation with respect to the Euclidean metric on  $\Omega$  (hence conformal mappings deserve their name).

Following now is a number of useful theorems. see what happens to conformal changes of metrics on surfaces.





**Figure B.4:** Visualizing the effects of the conformal mapping  $f_w$  on the disk, distorting the reference image (B.4a), Bubi.

**B.2.3 Theorem.** Let  $(M, g)$  be a Riemannian surface with boundary. Let  $u \in C^\infty(M)$  and  $\tilde{g} = e^{2u}g$ . We'll write tildes over all the corresponding quantities for  $\tilde{g}$ . Then the following transformation properties hold:

1.  $\tilde{\Delta} = e^{-2u}\Delta$ .
2.  $d\tilde{A} = e^{2u}dA$  (where  $dA$  is the area element).
3.  $\tilde{G}_\mathcal{D} = G_\mathcal{D}$  (where  $G_\mathcal{D}$  is the Dirichlet Green's function corresponding to  $\Delta_g$ ).
4.  $\tilde{m}_\mathcal{D} = \frac{u}{2\pi} + m_\mathcal{D}$ .
5.  $\tilde{K} = e^{-2u}(\Delta u + K) = \tilde{\Delta}u + e^{-2u}K$  where  $K$  is the Gauß curvature of  $g$  (note: in the

convention in the main body of this work, there is an extra minus sign for the Laplacian).

*Proof.* First note  $\sqrt{\tilde{g}} = \sqrt{\det(e^{2u}g_{ij})} = \sqrt{e^{4u}\det(g_{ij})} = e^{2u}\sqrt{g}$ . This already gives us (2). For (1), we just compute

$$\tilde{\Delta}f = \frac{1}{e^{2u}\sqrt{g}} \frac{\partial}{\partial x^i} \left( e^{2u}\sqrt{g}e^{-2u}g^{ij} \frac{\partial}{\partial x^j} f \right) = e^{-2u} \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left( \sqrt{g}g^{ij} \frac{\partial}{\partial x^j} f \right) = e^{-2u}\Delta f.$$

For (3), let  $\mathcal{H}$  be the (Sobolev) space of all functions vanishing at the boundary (the space suited for Dirichlet boundary conditions, possessing enough weak derivatives for the elliptic regularity theory to apply). We know that  $\Delta_h$  is invertible for any metric  $h$  on  $\mathcal{H}$ , and its inverse is given by using the Green's function as an integration kernel:

$$\Delta_h^{-1}f = \int_M G_h(x, y)f(y) dA_h(y).$$

Now since multiplying by a smooth positive function is also an invertible operation, we have

$$\tilde{\Delta}^{-1}f = (e^{-2u}\Delta)^{-1}f = \Delta^{-1}(e^{2u}f).$$

So on the one hand we have

$$(B.2.5) \quad \tilde{\Delta}^{-1}f = \int_M \tilde{G}(x, y)f(y) d\tilde{A}(y) = \int_M \tilde{G}(x, y)f(y)e^{2u(y)} dA(y),$$

and on the other hand, we have

$$(B.2.6) \quad \Delta^{-1}(e^{2u}f) = \int_M G(x, y)e^{2u(y)}f(y) dA(y).$$

Combining (B.2.5) and (B.2.6) we see that the integrals are identical in every respect for any function with vanishing boundary conditions, with the exception of the fact

that one involves  $G$  and the other,  $\tilde{G}$ . This implies  $G = \tilde{G}$ .

For (4), we are immensely assisted by (3). We have, since  $\tilde{G} = G$ ,

$$\tilde{m}(p) = \lim_{q \rightarrow p} \left( G(p, q) + \frac{1}{2\pi} \log(\tilde{d}(p, q)) \right).$$

so that

$$\tilde{m}(p) - m(p) = \lim_{q \rightarrow p} \frac{1}{2\pi} (\log(\tilde{d}(p, q)) - \log(d(p, q))) = \lim_{q \rightarrow p} \frac{1}{2\pi} \log\left(\frac{\tilde{d}(p, q)}{d(p, q)}\right)$$

So it remains to calculate the ratio of the two geodesic distances as  $q \rightarrow p$ . Heuristically, since  $g$  measures infinitesimal squared distance, in the infinitesimal limit, the ratio of the squared distance is  $e^{2u}$ . So the ratios of the non-squared distances is just  $e^u$ . For true proof of this fact (which is valid in any dimension), we recall the concept of the exponential map and normal coordinates: given a point  $p \in M$  in a metric  $h$ , there exist coordinates  $(x^i)$  such that  $p$  maps to 0,  $h_{ij}(p) = \delta_{ij}$ , and the first-order derivatives of  $h_{ij}$  also vanish (this can be guaranteed to happen only at  $p$ ; due to the fact we cannot eliminate second-order derivatives in general—the obstruction is curvature). This is in turn accomplished by mapping a tangent vector  $V$  based at  $p$  to the point in  $M$  arrived at by moving out along a geodesic, with initial velocity  $V$ , for unit time. The map that does this is called the EXPONENTIAL MAP. In a small enough neighborhood of the origin in the tangent space, the exponential map is a diffeomorphism, which gives us normal coordinates. The coordinates of the image point  $p$  are the vector components of  $V$ . The crucial observation to make is that a straight line with direction vector  $V$  through the origin in  $T_pM$  corresponds to a geodesic passing through  $p$  with tangent vector  $V$  (straight lines missing the origin do not necessarily correspond to geodesics). So let's prove the following

**B.2.4 Lemma.**  $\lim_{q \rightarrow p} \frac{\tilde{d}(p, q)}{d(p, q)} = e^{u(p)}$ .

Note that this proves the transformation formula (4) because  $\frac{1}{2\pi} \log(e^u) = \frac{u}{2\pi}$ .

*Proof of Lemma.* Consider normal coordinates  $(x^j)$  for  $g$  and  $(y^j)$  for  $\tilde{g}$  at  $p$ . Since we are only considering what happens when  $q$  approaches  $p$ , we may assume  $q$  is in the intersection of these two neighborhoods for which normal coordinates exist. In other words,  $q$  is close enough to  $p$  for there to be a minimizing geodesic between the two. Thus the  $g$ -geodesic through  $p$  and  $q$  in coordinates is a straight line from 0 to  $x$ , and the representation of the tangent vector is also given by the constant vector  $x$  (recall that the tangent vector is parallel-transported along a geodesic). We normalize the geodesics to be unit speed in their respective metrics, that is, use  $g$ - and  $\tilde{g}$ -unit vectors  $v, w$ , respectively. Let  $\alpha, \beta$  be those two geodesics, for  $g$  and  $e^{2u}g$  respectively. Then there exist parameters  $t(q), s(q)$  at which  $\alpha(t(q)) = q$  and  $\beta(s(q)) = q$  (in other words  $t, s$  are inverses of  $\alpha, \beta$  respectively). Note that  $t(q) = d(p, q)$  and  $s(q) = \tilde{d}(p, q)$  since the geodesics are unit speed. Now  $\beta$  is not necessarily a geodesic in the metric  $g$ , so in particular the  $g$ -length of  $\beta$  is at least the length of  $\alpha$ , by the minimality of  $\alpha$  (since both have the same endpoints). So, we have  $t(q) \leq \int_0^{s(q)} \|\beta'(\tau)\|_g d\tau$ , the length of the not-necessarily-geodesic  $\beta$  which gives us

$$\frac{\tilde{d}(p, q)}{d(p, q)} = \frac{s(q)}{t(q)} \geq \frac{s(q)}{\int_0^{s(q)} \|\beta'(\tau)\|_g d\tau}.$$

By the Mean Value Theorem, there is  $\xi(q)$  between 0 and  $s(q)$  such that

$$\int_0^{s(q)} \|\beta'(\tau)\|_g d\tau = s(q) \|\beta'(\xi(q))\|_g.$$

Therefore

$$\frac{\tilde{d}(p, q)}{d(p, q)} \geq \frac{s(q)}{s(q) \|\beta'(\xi(q))\|_g} = \frac{1}{\|\beta'(\xi(q))\|_g}$$

Letting  $q \rightarrow p$ , we have  $\xi(q) \rightarrow 0$ , so that since the norm is continuous and geodesics are smooth,  $\|\beta'(\xi(q))\|_g \rightarrow \|\beta'(0)\|_g = \|w\|_g$ . But  $\|w\|_g = e^{-u(p)} \|w\|_{\tilde{g}} = e^{-u(p)}$  by definition of conformal change and the fact that  $w$  is a unit vector for  $\tilde{g}$ . Therefore

$$\lim_{q \rightarrow p} \frac{\tilde{d}(p, q)}{d(p, q)} \geq e^{u(p)}.$$

Now we prove the other inequality by a symmetry argument: there's no reason why  $g$  should have been preferred, and in fact  $g = e^{-2u} \tilde{g}$ . So the exact same argument above, using  $-u$  in place of  $u$  and swapping the roles of  $d$  and  $\tilde{d}$  gives

$$\lim_{q \rightarrow p} \frac{d(p, q)}{\tilde{d}(p, q)} \geq e^{-u(p)}.$$

Inverting both sides, which reverses the inequality, gives

$$\lim_{q \rightarrow p} \frac{\tilde{d}(p, q)}{d(p, q)} \leq e^{u(p)}.$$

□

For (5) things are a bit more involved. We follow the argument in [18] using the method of moving frames. Let  $f_1, f_2$  be a frame field, orthonormal in the metric  $g$ , and  $e_1 = e^{-u} f_1, e_2 = e^{-u} f_2$ , which are orthonormal in the metric  $\tilde{g}$ . Consider their dual coframes  $\{\eta^i\}$  and  $\{\omega^i\}$ , respectively. Note that  $\omega^i = e^u \eta^i$ , for  $i = 1, 2$ . The coframes, of course, satisfy the inverse of the relationship satisfied by the frames. The connection 1-forms  $\eta_i^j$  and  $\omega_i^j$  for the metrics  $g$  and  $\tilde{g}$ , respectively, are implicitly defined by the relationships

$$\nabla_X f_i = f_j \eta_i^j(X)$$

$$\tilde{\nabla}_X e_i = e_j \omega_i^j(X).$$

Explicitly, with the metric, we have

$$g(\nabla_X f_i, f_k) = \eta_i^k(X)$$

$$\tilde{g}(\tilde{\nabla}_X e_i, e_k) = \omega_i^k(X)$$

Because the basis is orthonormal, by the product rule, we have that the connection 1-forms are antisymmetric, namely  $\omega_i^j = -\omega_j^i$  (or even if not orthonormal, then defining  $\omega_{ij} = g_{ik}\omega_j^k$ , we always have  $\omega_{ij} = -\omega_{ji}$ ). Similar considerations hold for the  $\eta$ 's. Also, the relationship between exterior derivatives and covariant derivatives gives us the relations

$$d\omega^i = -\omega_j^i \wedge \omega^j$$

$$d\eta^i = -\eta_j^i \wedge \eta^j.$$

Finally, we have the curvature forms

$$\widetilde{\text{Rm}}_i^j = d\omega_i^j + \omega_k^j \wedge \omega_i^k = d\omega_i^j$$

$$\text{Rm}_i^j = d\eta_i^j + \eta_k^j \wedge \eta_i^k = d\eta_i^j$$

where the wedged terms drop out because either a form is wedged with itself (giving 0), or the indices are equal, also giving 0 by antisymmetry. So our task is simply to calculate  $d\omega_i^j$  in terms of  $d\eta_i^j$  and other quantities associated to the  $\eta$ 's. Since the

forms are antisymmetric, we only need to calculate  $d\omega_1^2$ . But first, we calculate  $d\omega^i$ :

$$d\omega^1 = d\omega^1(e_1, e_2)\omega^1 \wedge \omega^2 = -\omega_2^1 \wedge \omega^2$$

$$d\omega^2 = d\omega^2(e_1, e_2)\omega^1 \wedge \omega^2 = -\omega_1^2 \wedge \omega^1$$

which says

$$\omega_1^2 = d\omega^1(e_1, e_2)\omega^1 + d\omega^2(e_1, e_2)\omega^2.$$

But then  $\omega^i = e^u \eta^i$ , so

$$d\omega^i = e^u du \wedge \eta^i + e^u d\eta^i = e^u (d\eta^i + du \wedge \eta^i).$$

Finally, we note that  $du = f_1[u]\eta^1 + f_2[u]\eta^2 = e_1[u]\omega^1 + e_2[u]\omega^2$ , where  $f_i[u]$  denotes the directional derivative (the component formula for exterior derivatives works even for non-coordinate frames). Plugging this in, we have

$$d\omega^1 = e^u (d\eta^1 + f_2[u]\eta^2 \wedge \eta^1) = e^u (-\eta_2^1 \wedge \eta^2 + f_2[u]\eta^2 \wedge \eta^1)$$

$$d\omega^2 = e^u (d\eta^2 + f_1[u]\eta^1 \wedge \eta^2) = e^u (-\eta_1^2 \wedge \eta^1 + f_1[u]\eta^1 \wedge \eta^2)$$

Now,

$$d\omega^1(e_1, e_2) = e^{-2u} d\omega^1(f_1, f_2)$$

$$= e^{-u} (-\eta_2^1(f_1)\eta^2(f_2) + \eta_2^1(f_2)\eta^2(f_1) - f_2[u]) = e^{-u} (-\eta_2^1(f_1) - f_2[u])$$

$$d\omega^2(e_1, e_2) = e^{-2u} d\omega^2(f_1, f_2)$$

$$= e^{-u} (-\eta_1^2(f_1)\eta^1(f_2) + \eta_1^2(f_2)\eta^1(f_1) + f_1[u]) = e^{-u} (\eta_1^2(f_2) + f_1[u]).$$

Therefore,

$$\omega_1^2 = \eta_1^2(f_1)\eta^1 + \eta_1^2(f_2)\eta^2 - f_2[u]\eta^1 + f_1[u]\eta^2 = \eta_1^2 - f_2[u]\eta^1 + f_1[u]\eta^2.$$

So,

$$\begin{aligned} \widetilde{\text{Rm}}_1^2 &= d\omega_1^2 = d\eta_1^2 - f_2[u]d\eta^1 + f_1[u]d\eta^2 \\ &\quad - (f_1[f_2[u]]\eta^1 + f_2[f_2[u]]\eta^2) \wedge \eta^1 + (f_1[f_1[u]]\eta^1 + f_2[f_1[u]]\eta^2) \wedge \eta^2 \\ &= d\eta_1^2 + f_2[u](\eta_2^1 \wedge \eta^2) - f_1[u](\eta_1^2 \wedge \eta^1) + (f_1[f_1[u]] + f_2[f_2[u]])\eta^1 \wedge \eta^2. \end{aligned}$$

Now by the definition of sectional curvature, and orthonormality, we have

$$K = g(\text{Rm}(f_1, f_2)f_2, f_1) = \text{Rm}_1^2(f_2, f_1)$$

and similarly for  $\tilde{g}$ . Plugging it in, we have

$$\begin{aligned} \tilde{K} &= d\omega_1^2(e_2, e_1) = e^{-2u}(d\eta_1^2(f_2, f_1) - f_2[u]\eta_2^1(f_1) - f_1[u]\eta_1^2(f_2) - f_1[f_1[u]] - f_2[f_2[u]]) \\ &= e^{-2u}(K + f_1[u]\eta_2^1(f_2) - f_1[f_1[u]] + f_2[u]\eta_1^2(f_1) - f_2[f_2[u]]). \end{aligned}$$

Finally, we have

$$\Delta u = \sum_i (\nabla_{f_i} f_i)[u] - f_i[f_i[u]]$$

and the last thing to calculate is what  $\nabla_{f_i} f_i$  is. Using the definition of the forms  $\eta_i^j$ , we have  $\nabla_{f_i} f_i = f_j \eta_i^j(f_i)$ . Making it act on  $u$  and comparing, we finally have the result:

$$\tilde{K} = e^{-2u}(\Delta u + K).$$

□



For closed manifolds there are analogous formulæ for the transformation of the Green's function and the Robin mass, but they are considerably more complicated. We'll pursue those formulæ in short order. It also is similar to the case for Neumann conditions. We'll look at that later however; first let's get back to our original goal in looking at the disk.

### B.3 Two-Dimensional Example: The Hyperbolic Disk

Can we find a metric on the unit disk conformal to the flat metric with constant Robin mass? Using the above, we want to find  $u$  such that

$$\frac{1}{2\pi} (u(z) + \log(1 - |z|^2)) = m_{e^{2u}g}(z) \equiv M.$$

for some constant  $M$ . This says

$$u(z) = 2\pi M - \log(1 - |z|^2)$$

So the conformal factor is

$$e^{2u(z)} = \frac{e^{4\pi M}}{(1 - |z|^2)^2}$$

so that

$$(B.3.1) \quad \tilde{g} = \frac{e^{4\pi M}}{(1 - |z|^2)^2} dzd\bar{z}.$$

However, notice that this is just the hyperbolic metric (up to a scale factor)! There is a small issue with the fact that this is not conformal to the Euclidean metric if the boundary is included, since the metric blows up there. Technically we should say  $\tilde{g}$  is only conformal to  $g$  on the interior, boundaryless manifold. Nevertheless, the

formulæ still hold because we can still speak of functions approaching the boundary in Dirichlet conditions.

Using (5) in the above theorem, we can express  $M$  rather elegantly in terms of the (constant negative) curvature  $K$  of the hyperbolic metric:

$$K = -4e^{-4\pi M}(1 - |z|^2)^2 \frac{\partial^2}{\partial z \partial \bar{z}} (2\pi M - \log(1 - |z|^2)) = 4e^{-4\pi M}(1 - |z|^2)^2 \frac{\partial^2}{\partial z \partial \bar{z}} \log(1 - |z|^2).$$

Now

$$\frac{\partial}{\partial \bar{z}} \log(1 - |z|^2) = \frac{-z}{1 - |z|^2}$$

because the usual product and chain rule work exactly the same way with complex coordinates, and  $|z|^2 = z\bar{z}$ . Differentiating this with respect to  $z$ ,

$$\frac{\partial}{\partial z} \left( \frac{-z}{1 - |z|^2} \right) = \frac{(1 - |z|^2)(-1) + z(-\bar{z})}{(1 - |z|^2)^2} = -\frac{1}{(1 - |z|^2)^2}.$$

Therefore

$$K = -4e^{-4\pi M}$$

or

$$(B.3.2) \quad m_{\tilde{g}} \equiv M = -\frac{1}{4\pi} \log \left| \frac{K}{4} \right| = \frac{1}{4\pi} \log(4) - \frac{1}{4\pi} \log |K|.$$

So for example if  $K = -1$ , we have  $M = \frac{1}{4\pi} \log(4) = \frac{1}{2\pi} \log(2)$ , and  $K = -4$  gives a Robin mass of zero, in other words the Robin mass varies proportionally to the negative log of the magnitude of the curvature.

Although we've already established the mass and we can say all is said and done, nevertheless we should review a bit of hyperbolic geometry to help get a feel for things.

First, we should note that biholomorphisms of the disk are actually *isometries* of the hyperbolic metric (they were merely conformal transformations for the Euclidean metric). This is just an application of the famous

**B.3.1 Schwarz's Lemma.** *Let  $f : \mathbb{D} \rightarrow \mathbb{D}$  be a holomorphic function such that  $f(0) = 0$ . Then  $|f'(0)| \leq 1$  and  $|f(z)| \leq |z|$  with equality if and only if  $f$  is a rotation (multiplication by  $e^{i\varphi}$  for some  $\varphi$ ).*

The proof is merely an application of the maximum principle. Using the fact that biholomorphisms of the disk consist entirely of rotations and single Blaschke factors, we can prove the following more symmetric (i.e. less dependent of being origin-specific), generalized version due to Pick:

**B.3.2 Pick's Lemma.** *Let  $f : \mathbb{D} \rightarrow \mathbb{D}$  be a holomorphic function. Then for all  $z, w \in \mathbb{D}$ , we have*

$$(B.3.3) \quad |f'(w)| \leq \frac{1 - |f(w)|^2}{1 - |w|^2}$$

and

$$(B.3.4) \quad \left| \frac{f(z) - f(w)}{1 - \overline{f(w)}f(z)} \right| \leq \left| \frac{z - w}{1 - \bar{w}z} \right|,$$

with equality if and only if  $f$  is a biholomorphism.

The proof simply uses conformal maps to reduce to the Schwarz Lemma.

*Proof.* Let  $w$  be given,

$$H(z) = \frac{z - w}{1 - \bar{w}z}, \text{ and}$$

$$G(\eta) = \frac{\eta - f(w)}{1 - \overline{f(w)}\eta}.$$

Then  $G \circ f \circ H^{-1} : \mathbb{D} \rightarrow \mathbb{D}$  is holomorphic and  $G(f(H^{-1}(0))) = 0$ , so by the usual Schwarz lemma,  $|(G \circ f \circ H^{-1})'(0)| \leq 1$  and  $|G(f(H^{-1}(\zeta)))| \leq |\zeta|$  for all  $\zeta$ , with equality if the total composition map is a rotation, that is, if and only if  $f$  is a biholomorphism (since  $G$  and  $H$  are biholomorphisms). Therefore,  $|G(f(z))| \leq |H(z)|$  for all  $z$ . But writing the definition of  $G$  and  $H$  out, this is just (B.3.4). Now observe  $H^{-1}(0) = w$  by definition, so by the Chain Rule,

$$|(G \circ f \circ H^{-1})'(0)| = |G'(f(w))f'(w)(H^{-1})'(0)| = \left| \frac{G'(f(w))f'(w)}{H'(w)} \right| \leq 1$$

Therefore,

$$|f'(w)| \leq \left| \frac{H'(w)}{G'(f(w))} \right|.$$

But

$$H'(w) = \frac{(1 - \bar{w}z) - (z - w)(-\bar{w})}{(1 - \bar{w}z)^2} \Big|_{z=w} = \frac{1 - |w|^2}{(1 - |w|^2)^2} = \frac{1}{1 - |w|^2}.$$

Because  $G$  is also a Blaschke factor, we have  $G'(f(w)) = \frac{1}{1 - |f(w)|^2}$  so that

$$|f'(w)| \leq \frac{|1 - |f(w)|^2|}{|1 - |w|^2|}.$$

This is (B.3.3), since both the numerator and denominator without the absolute values are real and positive. □

**B.3.3 Corollary.** *The biholomorphisms of the disk are hyperbolic isometries.*

*Proof.* We have that any hyperbolic metric on  $\mathbb{D}$  is given by

$$g = \frac{B}{(1 - |z|^2)^2} dzd\bar{z}$$

for a  $B > 0$  a constant (it is  $-4/K$  where  $K$  is the Gauß curvature, or  $e^{4\pi M}$  where  $M$  is the Robin mass). Let  $F$  be a biholomorphism and write  $\zeta$  for the range variable. Then as

noted before,  $F$  is a conformal transformation; specifically,  $F^*(d\zeta d\bar{\zeta}) = |F'(z)|^2 dz d\bar{z}$ .

By the Schwarz Lemma,  $|F'(z)| = \frac{1-|F(z)|^2}{1-|z|^2}$ . So

$$\begin{aligned} F^*g &= F^*\left(\frac{B}{(1-|\zeta|^2)^2} d\zeta d\bar{\zeta}\right) = F^*\left(\frac{B}{(1-|\zeta|^2)^2} |F'(z)|^2 dz d\bar{z}\right) \\ &= \frac{B}{(1-|F(z)|^2)^2} \left(\frac{1-|F(z)|^2}{1-|z|^2}\right)^2 dz d\bar{z} = \frac{B}{(1-|z|^2)^2} dz d\bar{z} = g. \end{aligned}$$

□

Note that this proof means that conformal maps are therefore isometries under any rescaling of the standard hyperbolic metric with curvature  $-4$  i.e. ( $B = 1$ ). This in turn means conformal mappings preserve geodesic distance (i.e. it is an isometry in the basic real analysis sense). Let's recall the following

**B.3.4 Theorem.** *In the standard hyperbolic metric on the disk  $\mathbb{D}$ , we have*

$$d(z, w) = \tanh^{-1} \left| \frac{z-w}{1-\bar{w}z} \right| = \frac{1}{2} \log \left( \frac{1 + \left| \frac{z-w}{1-\bar{w}z} \right|}{1 - \left| \frac{z-w}{1-\bar{w}z} \right|} \right)$$

*Proof.* It suffices to prove  $d(z, 0) = \tanh^{-1}(|z|)$  because  $d$  is invariant under biholomorphisms, i.e.  $d(f(\xi), f(\eta)) = d(\xi, \eta)$  for any  $\xi, \eta \in \mathbb{D}$ , so that using the same trick we used for the Green's function,  $d(z, w) = d(f_w(z), 0)$  (where  $f_w$  is that Blaschke factor). Notice that as  $z$  goes to the boundary,  $d$  blows up, i.e. the boundary circle is infinitely far away from any point, in hyperbolic geometry. Rotations about the origin are also hyperbolic isometries, so we may assume additionally that  $z$  is on the positive real line. Then, a geodesic from 0 to  $z$  is a (Euclidean) straight line,  $\gamma(t) = tz$ . Therefore,

$$d(z, 0) = \int_0^1 \left( \frac{z^2}{(1-t^2z^2)^2} \right)^{1/2} dt = \int_0^1 \frac{z}{1-z^2t^2} dt.$$

Now let  $t = (1/z) \tanh(u)$ , or  $u = \tanh^{-1}(zt)$ . Then  $dt = (1/z) \operatorname{sech}^2(u) du$ . But  $1 -$

$\tanh^2(u) = \operatorname{sech}^2(u)$ . Therefore

$$d(z, 0) = \int_0^{\tanh^{-1}(z)} du = \tanh^{-1}(z) = \tanh^{-1}(|z|).$$

□

In hyperbolic geometry, one starts to appreciate Blaschke factors a lot. Note that different authors have competing definitions of what it means to be a “standard” hyperbolic metric. Ours has constant Gaussian curvature  $-4$ , and our “standard” is that the conformal factor multiplying the Euclidean metric is 1 at the origin. What this means is that close to the origin, the hyperbolic distance is approximately the same as the Euclidean distance. Some books also take the curvature  $-1$  hyperbolic metric to be the “standard” because apparently it is more aesthetically pleasing to have curvatures be normalized. In that metric, hyperbolic distances near the origin look approximately double the Euclidean distance.

In summary, we can recompute the Robin mass directly from the Green’s function and log of the distance:

$$\begin{aligned} M &= \lim_{w \rightarrow z} \frac{1}{2\pi} \left( -\log \left| \frac{z-w}{1-\bar{w}z} \right| + \log \left( \tanh^{-1} \left| \frac{z-w}{1-\bar{w}z} \right| \right) \right) \\ &= \lim_{w \rightarrow z} \frac{1}{2\pi} \left( -\log \left| \frac{z-w}{1-\bar{w}z} \right| + \log \left( \left| \frac{z-w}{1-\bar{w}z} \right| + \frac{1}{3} \left| \frac{z-w}{1-\bar{w}z} \right|^3 + \frac{1}{5} \left| \frac{z-w}{1-\bar{w}z} \right|^5 + \dots \right) \right) \\ &= \lim_{w \rightarrow z} \frac{1}{2\pi} \left( \log \left( 1 + \frac{1}{3} \left| \frac{z-w}{1-\bar{w}z} \right|^2 + \frac{1}{5} \left| \frac{z-w}{1-\bar{w}z} \right|^4 + \dots \right) \right) = 0, \end{aligned}$$

directly confirming our previous calculation.

## B.4 Derivations for Neumann Boundary Conditions

Theorem B.2.3 above on the transformation properties of the Dirichlet Green's functions and Robin masses needs to be modified for the case of Neumann boundary conditions.<sup>1</sup> Since the kernel of  $\Delta$  (restricted to functions of vanishing normal derivative) is the constant functions, things are a little trickier to calculate, because we have to work in the orthogonal complement (in the Sobolev space) of those functions, and these orthogonal complements are different for different metrics! This makes it difficult to guess at what kinds of combinations of normalizations (i.e. choices of functions with vanishing total integral with respect to various volume elements) will make a suitable definition of  $\Delta^{-1}$ .

Instead, we follow the argument in [78], which calculates the transformation formula in the Green's function for conformal changes of metric by using properties of harmonic functions analogous to properties of holomorphic functions in the complex plane—namely that if they are bounded in any punctured neighborhood of a singularity, it in fact extends harmonically (i.e. the singularity is removable—Riemann's theorem), and if a function is defined and harmonic everywhere on a closed manifold (or on a manifold with boundary and has vanishing normal derivative at the boundary), then it is in fact constant (Liouville's Theorem).

Let us now add to Theorem B.2.3 on various transformation formulæ on surfaces:

**B.4.1 Theorem.** *Let  $M$  be a compact surface possibly with boundary. Then we have,*

---

<sup>1</sup>Which we will take from now on to mean either closed, i.e. compact with  $\partial M = \emptyset$ , or to have vanishing normal derivative, its original meaning. If we want to emphasize the original meaning, we'll say the "true" Neumann conditions, problem, etc.) This suggests that the Neumann condition is the more correct generalization of the closed manifold concept; indeed, if one considers a closed manifold with a small disk removed, and looks at what happens to the Neumann Green's function  $\tilde{G}$  as the radius of the excised disk tends to zero, one will see that it will approach the Green's function  $G$  for the closed manifold. Heuristically this is because the vanishing normal derivative allows the function to "close up" to yield a ( $C^2$ -) smooth solution (both  $G(\cdot, q)$  and  $\tilde{G}(\cdot, q)$  and their derivatives equal limits at the point).

for  $F \in C^\infty(M)$  a positive function (or  $u \in C^\infty(M)$  any smooth function and  $F = e^{2u}$ ), the following transformation formulæ for  $G_{\mathcal{N}}$  and  $m_{\mathcal{N}}$ :

$$G_{\mathcal{N},Fg}(p, q) = G_{\mathcal{N},g}(p, q) - \frac{1}{A_F}(\Delta_{\mathcal{N},g}^{-1}F)(q) - \frac{1}{A_F}(\Delta_{\mathcal{N},g}^{-1}F)(p) + \frac{1}{A_F^2} \int_M F \Delta_{\mathcal{N},g}^{-1}F dA_g$$

and

$$m_{\mathcal{N},Fg} = m_{\mathcal{N},g} + \frac{1}{4\pi} \log F - \frac{2}{A_F} \Delta_{\mathcal{N},g}^{-1}F + \frac{1}{A_F^2} \int_M F \Delta_{\mathcal{N},g}^{-1}F dA_g$$

where  $A_F = \int dA_{Fg} = \int F dA_g$  is the area in the  $Fg$  metric.

(For a comparison with the Dirichlet case, using  $F$  instead of  $e^{2u}$ , we have

$G_{\mathcal{D},Fg} = G_{\mathcal{D},g}$ , and

$$m_{\mathcal{D},Fg} = m_{\mathcal{D},g} + \frac{1}{4\pi} \log F,$$

which is significantly less complicated.)

*Proof.* Again, this is an adaptation of a proof for certain operators (the Paneitz operator) of general even order in [76]. For notational clarity, we drop all subscripts, and put tildes over all the metric-dependent quantities associated to  $Fg$  (so  $G$  is the Neumann Green's function for  $g$ , while  $\tilde{G}$  is the corresponding function for  $Fg$ ).

We write  $\Delta_q u(q, p_2, \dots, p_k)$  for the “partial” Laplacian with respect to the  $q$  variable, if  $u$  is a sufficiently smooth function on  $M^k$ . Consider the function

$$E(p, r, q) := G(p, q) - G(r, q),$$

and similarly  $\tilde{E}$  for the quantities in terms of  $\tilde{G}$ .  $E$  and  $\tilde{E}$  are smooth whenever  $q \notin \{p, r\}$ .

Thus  $\Delta_q E(p, r, q) = 0$  for  $q \notin \{p, r\}$ , and subtracting off the logarithmic singularities, we have that

$$E(p, r, q) + \frac{1}{2\pi} \log \left( \frac{d(p, q)}{d(r, q)} \right)$$



is bounded, and integrating over  $q$ , it is zero:

$$(B.4.1) \quad \int E(p, r, q) dA(q) = 0,$$

since the Green's functions are chosen to have vanishing integral in  $q$ . The same thing holds, of course, with tildes inserted over the relevant quantities. Now

$$\tilde{\Delta}_q \tilde{E}(p, r, q) = 0$$

also when  $q \notin \{p, r\}$ ; but we have that  $\tilde{\Delta}_q \tilde{E}(p, r, q) = F(q)^{-1} \Delta_q \tilde{E}(p, r, q)$ , so that in particular,  $\Delta_q \tilde{E}(p, r, q) = 0$  also. Therefore  $\Delta_q (\tilde{E}(p, r, q) - E(p, r, q)) = 0$  when  $q \notin \{p, r\}$ . However, for  $q$  in a sufficiently small neighborhood of  $p$  (with  $p \neq r$ ), adding and subtracting the logarithmic singularities appropriately,

$$\begin{aligned} \tilde{E}(p, r, q) - E(p, r, q) &= \tilde{G}(p, q) - \tilde{G}(r, q) - G(p, q) + G(r, q) \\ &= \left( \tilde{G}(p, q) + \frac{1}{2\pi} \log(\tilde{d}(p, q)) \right) - \left( G(p, q) + \frac{1}{2\pi} \log(d(p, q)) \right) \\ &\quad + G(r, q) - \tilde{G}(r, q) + \frac{1}{2\pi} \log\left(\frac{d(p, q)}{\tilde{d}(p, q)}\right) \end{aligned}$$

which is bounded, because  $q$  is in a neighborhood away from  $r$ , and  $\frac{1}{2\pi} \log\left(\frac{d(p, q)}{\tilde{d}(p, q)}\right)$  is, in the limit as  $q \rightarrow p$  is equal to  $1/\sqrt{F(p)}$ . Similarly, replacing  $d(p, q)$  with  $d(r, q)$  in the log singularities and putting them with the corresponding  $G(r, q)$ 's, the same calculation implies  $\tilde{E}(p, r, q) - E(p, r, q)$  is bounded for  $q$  in a neighborhood of  $r$ . If  $p = r$ , then trivially  $\tilde{E}(p, r, q) - E(p, r, q) = 0$  which is of course bounded. The upshot is:  $\tilde{E}(p, r, q) - E(p, r, q)$  is bounded for all  $p, q, r$ , and harmonic in  $q$  whenever  $q \notin \{p, r\}$ , i.e. harmonic in  $q$  on  $M \setminus \{p, r\}$ . But if a harmonic function is bounded in the neighborhood of a singularity, that singularity must be removable (Riemann's

theorem), so that  $\tilde{E}(p, r, q) - E(p, r, q)$  extends harmonically in  $q$  to all of  $M$ .

However,  $\tilde{E}(p, r, q) - E(p, r, q)$  satisfies the Neumann condition, since its normal derivative is the difference of the appropriate, all vanishing normal derivatives of the  $G$  and  $\tilde{G}$ , and is harmonic. Hence it must be constant. In the case that  $M$  is closed, all (global) harmonic functions are constant. In either case, we have

$$\tilde{E}(p, r, q) - E(p, r, q) \equiv C(p, r)$$

a constant independent of  $q$ . In the process of evaluating what  $C(p, r)$  is, we find the transformation formulas above. To do that, we simply average with respect to the  $Fg$  metric over the  $q$  variable (that is, integrate against  $d\tilde{A}(q) = F(q)dA(q)$  and divide by  $A_F$ ; note that averaging a constant leaves it alone):

$$\begin{aligned} C(p, r) &= \frac{1}{A_F} \int_M (\tilde{E}(p, r, q) - E(p, r, q)) F(q) dA(q) \\ &= \frac{1}{A_F} \int_M \tilde{E}(p, r, q) d\tilde{A}(q) - \frac{1}{A_F} \int_M E(p, r, q) F(q) dA(q) \\ &= -\frac{1}{A_F} \int_M E(p, r, q) F(q) dA(q) \end{aligned}$$

where the first integral goes away as observed in (B.4.1). But

$$\begin{aligned} -\frac{1}{A_F} \int_M E(p, r, q) F(q) dA(q) &= \frac{1}{A_F} \int_M G(r, q) F(q) dA(q) - \frac{1}{A_F} \int_M G(p, q) F(q) dA(q) \\ &= \frac{1}{A_F} (\Delta^{-1}F)(r) - \frac{1}{A_F} (\Delta^{-1}F)(p). \end{aligned}$$

This means

$$\tilde{E}(p, r, q) - E(p, r, q) = \tilde{G}(p, q) - \tilde{G}(r, q) - G(p, q) + G(r, q) = \frac{1}{A_F} (\Delta^{-1}F)(r) - \frac{1}{A_F} (\Delta^{-1}F)(p),$$

or, rearranging,

$$\tilde{G}(p, q) = \tilde{G}(r, q) + G(p, q) - G(r, q) + \frac{1}{A_F}(\Delta^{-1}F)(r) - \frac{1}{A_F}(\Delta^{-1}F)(p).$$

Now averaging with respect to  $F(r)dA(r)$ , we have that the first term on the RHS goes away (because it is integrating  $\tilde{G}$  against  $d\tilde{A}$ ), the second and last terms are unchanged because they are independent of  $r$ , and the third term becomes

$$-\frac{1}{A_F} \int_M G(r, q)F(r) dA(r)$$

which is just  $-\frac{1}{A_F}(\Delta^{-1}F)(q)$ . The fourth term multiplies the integrand by  $F$  and introduces an extra  $A_F$  in the denominator because of averaging. Therefore the first statement of the theorem,

$$\tilde{G}(p, q) = G(p, q) - \frac{1}{A_F}(\Delta^{-1}F)(q) - \frac{1}{A_F}(\Delta^{-1}F)(p) + \frac{1}{A_F^2} \int_M F\Delta^{-1}F dA$$

is proved. For the Robin mass, adding  $\frac{1}{2\pi} \log(\tilde{d}(p, q))$  to both sides, and rewriting it on the RHS as  $\frac{1}{2\pi} \log(d(p, q)) + \frac{1}{2\pi} \log\left(\frac{\tilde{d}(p, q)}{d(p, q)}\right)$ , we have, taking the limit as  $q \rightarrow p$ , which gives the  $\frac{1}{4\pi} \log F$  term:

$$\begin{aligned} m_{Fg}(p) &= m_g(p) + \lim_{q \rightarrow p} \frac{1}{2\pi} \log\left(\frac{\tilde{d}(p, q)}{d(p, q)}\right) \\ &\quad - \frac{1}{A_F}(\Delta^{-1}F)(p) - \lim_{q \rightarrow p} \frac{1}{A_F}(\Delta^{-1}F)(q) + \frac{1}{A_F^2} \int_M F\Delta^{-1}F dA \\ &= m_g(p) + \frac{1}{4\pi} \log F(p) - \frac{2}{A_F}(\Delta^{-1}F)(p) + \frac{1}{A_F^2} \int_M F\Delta^{-1}F dA. \end{aligned}$$

□

## B.5 The Finite Cylinder

We now give a more complicated example. We calculate the Robin mass on the finite cylinder  $C = S^1 \times [0, \pi]$ . The idea is simple: we calculate the Green's function for the infinite strip, and then *periodize* the Green's function by adding all the  $2\pi k$ -translates in the second variable. In physical terms, this means we are looking for the electric potential for the 2D-cross section of a field resulting from a large number of evenly spaced lines of charge, in the space between two parallel, grounded planes. The difficult issue here is whether the series converges, i.e. as the number of charged lines tends to infinity, the field remains finite. It is not hard to see, for example, if the grounded planes were *not* there, that the field would grow large very quickly, i.e. this trick does not work for an *infinite* cylinder  $S^1 \times \mathbb{R}$ .

To get the result on the strip, we use conformal mapping. We map the disk to the strip conformally, and pull the Green's function back; the result is in fact the Green's function for the strip, because of the conformal invariance of  $\Delta$ . The conformal map can be broken down as follows: first map the disk to the upper half-plane, using the mapping  $z \mapsto i \frac{1+z}{1-z}$ . In polar coordinates the upper half-plane has argument from 0 to  $\pi$ , but there are no restrictions on the radius. So after applying the appropriate branch of the logarithm (using the argument in range  $(0, \pi)$ ), this maps the upper-half plane to the strip  $S = \mathbb{R} \times (0, \pi)$ . In summary, the map is

$$F(z) = \log \left( i \frac{1+z}{1-z} \right)$$

and its inverse is

$$H(z) = F^{-1}(z) = \frac{ie^z + 1}{ie^z - 1}.$$

The Green's function is then

$$(B.5.1) \quad G_S(z, w) = G_{\mathbb{D}}(H(z), H(w)) = -\frac{1}{2\pi} \log \left| \frac{H(z) - H(w)}{1 - \overline{H(w)}H(z)} \right|.$$

To motivate finding the Green's function for the strip, we imagine now that  $w$  is the “source” term. Putting additional sources at every integer multiple of  $2\pi$  along the real axis, this periodizes  $G_S$  in the  $w$  variable:

$$G_C(z, w) = \sum_{k=-\infty}^{\infty} G_S(z, w + 2\pi k) = -\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \log \left| \frac{H(z) - H(w + 2\pi k)}{1 - \overline{H(w + 2\pi k)}H(z)} \right|.$$

Assuming convergence (when  $z$  is not  $w$  or any of its translates), this automatically periodizes  $G$  in the variable  $z$  as well, since translation of the strip by any real number (i.e. horizontal motion) sends the strip conformally into itself, so that

$$\begin{aligned} G_C(z + 2\pi n, w) &= \sum_{k=-\infty}^{\infty} G_S(z + 2\pi n, w + 2\pi k) \\ &= \sum_{k=-\infty}^{\infty} G_S(z, w + 2\pi(k - n)) = \sum_{j=-\infty}^{\infty} G_S(z, w + 2\pi j) = G_C(z, w). \end{aligned}$$

Conformal invariance of  $G_S$  under horizontal translations also allows us to see that this function is symmetric in  $z$  and  $w$ . To summarize, we have the following:

**B.5.1 Theorem.** *Consider the finite cylinder  $C$ . Then its Dirichlet Green's function is given by*

$$(B.5.2) \quad G_C(z, w) = \sum_{j=-\infty}^{\infty} G_S(z, w + 2\pi j) = -\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \log \left| \frac{H(z) - H(w + 2\pi k)}{1 - \overline{H(w + 2\pi k)}H(z)} \right|,$$

where the series converges absolutely and uniformly for  $z, w$  in the strip.

*Proof that the series converges.* To show convergence, we consider the function, for

$\alpha \in \mathbb{D}$  and  $x \in \mathbb{R}$ ,

$$\psi(\alpha, x) = \log \left| \frac{\alpha - H(w+x)}{1 - \overline{H(w+x)}\alpha} \right| = \log \left| \frac{\alpha - H(w+x)}{1 - \bar{\alpha}H(w+x)} \right|.$$

We contend that  $\psi$  behaves like  $C(\alpha)e^{-|x|}$  for some  $C > 0$ , for sufficiently large  $|x|$ , i.e. it decays exponentially in both directions. Then the series converges by the Integral Test (provided, of course, none of  $G_S$  or its translates are evaluated directly on the singularity).

We have, multiplying through by the denominator in the definition of  $H$ , we have

$$\begin{aligned} \left| \frac{\alpha - H(w+x)}{1 - \bar{\alpha}H(w+x)} \right| &= \left| \frac{\alpha(ie^w e^x - 1) - (ie^w e^x + 1)}{ie^w e^x - 1 - \bar{\alpha}(ie^w e^x + 1)} \right| = \left| \frac{(\alpha - 1)ie^w e^x - (\alpha + 1)}{(1 - \bar{\alpha})ie^w e^x - 1 - \bar{\alpha}} \right| \\ &= \left| \frac{(1 + \alpha) + (1 - \alpha)ie^w e^x}{(1 - \alpha)(-i)e^{\bar{w}} e^x - (1 + \alpha)} \right| = \left| \frac{(1 + \alpha) + i(1 - \alpha)e^w e^x}{(1 + \alpha) - i(1 - \alpha)(-e^{\bar{w}} e^x)} \right| = \left| \frac{1 + \sigma(\alpha)e^w e^x}{1 - \sigma(\alpha)(-e^{\bar{w}})e^x} \right|, \end{aligned}$$

where  $\sigma$  is the conformal map  $z \mapsto i\frac{1+z}{1-z}$  which takes the disk to the upper half-plane. Since  $\alpha \in \mathbb{D}$ ,  $\sigma(\alpha)$  is therefore in the upper half-plane. Write  $A = \sigma(\alpha)e^w$  and  $B = -\sigma(\alpha)e^{\bar{w}}$ . Note since  $e^{\bar{w}} = \overline{e^w}$ , we have that  $|A| = |B|$ .

We then have, by the preceding derivation,

$$\psi(\alpha, x) = \log \left| \frac{1 + Ae^x}{1 - Be^x} \right|.$$

For  $x$  sufficiently large and negative, then  $|Ae^x| = |Be^x| < 1$ , so that, by the triangle inequalities  $|a + b| \leq |a| + |b|$  and  $|a - b| \geq ||a| - |b||$ , we have

$$\psi(\alpha, x) = \log \left| \frac{1 + Ae^x}{1 - Be^x} \right| \leq \log \left( \frac{1 + |A|e^x}{1 - |B|e^x} \right) = 2 \tanh^{-1}(|A|e^x),$$

since  $|A| = |B|$ , for  $x$  large and negative. The power series expansion of  $\tanh^{-1}$  yields

$$\begin{aligned} 2 \tanh^{-1}(|A|e^x) &= |A|e^x + \frac{1}{3}|A|^3 e^{3x} + \frac{1}{5}|A|^5 e^{5x} + \dots \leq |A|e^x \sum_{k=0}^{\infty} (|A|e^x)^{2k} \\ &= \frac{|A|e^x}{1 - |A|^2 e^{2x}} \leq 2|A|e^x, \end{aligned}$$

again, for  $x$  large and negative (the last inequality follows because  $|A|^2 e^{-2x}$  is eventually less than  $\frac{1}{2}$ ).

Now for  $x$  large and positive, we still have, by the triangle inequalities above, regardless of  $x$ ,

$$\psi(\alpha, x) \leq \log \left| \frac{1 + |A|e^x}{1 - |B|e^x} \right|,$$

where we have not taken away the absolute value bars. However, we have, dividing through by  $|A|e^x = |B|e^x$  (which is not zero because  $\sigma(\alpha)$  and  $e^w$  are not zero),

$$\log \left| \frac{1 + |A|e^x}{1 - |B|e^x} \right| = \log \left| \frac{1 + |A|^{-1}e^{-x}}{1 - |B|^{-1}e^{-x}} \right|.$$

For sufficiently large positive  $x$ , we have that, this time,  $|A|^{-1}e^{-x} < 1$ , so we may remove the absolute values and obtain

$$\psi(\alpha, x) \leq 2 \tanh^{-1}(|A|^{-1}e^{-x}) \leq 2|A|^{-1}e^{-x}$$

by the same argument with the geometric series. Taking  $C(\alpha) = \max\{2|A|, 2|A|^{-1}\}$ , we have the result follows for  $|x|$  sufficiently large.

The proof that the series in the definition of the Green's function converges uniformly, we observe that, with our notation, that, setting  $\alpha = H(z) \in \mathbb{D}$ ,

$$G_C(z, w) = -\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \psi(\alpha, 2\pi k).$$

By the Weierstraß  $M$ -test, the convergence in  $z$  in compact subsets of  $S \setminus \{w + 2\pi k : k \in \mathbb{Z}\}$ . Actually, even at those particular points, the divergence of the series is caused by one bad term, not bad behavior of the terms in the tails of the series. This means, in particular, we can interchange integration and infinite summation, which allows us to check that this really is indeed the Green's function (i.e. it satisfies the Laplace equation and so forth).

Choosing coordinates such that  $z \in [-\pi, \pi] \times [0, \pi]$ , we have

$$\begin{aligned} \iint G_C(z, w) f(w) dA(w) &= \int_{-\pi}^{\pi} \int_0^{\pi} \sum_{k=-\infty}^{\infty} G_S(z + 2\pi k, \theta, \zeta) \Delta u(\theta, \zeta) d\zeta d\theta \\ &= \int_{-\pi}^{\pi} \int_0^{\pi} G_S(z, \theta, \zeta) \Delta u(\theta, \zeta) d\zeta d\theta + \int_{-\pi}^{\pi} \int_0^{\pi} \sum_{k \neq 0} G_S(z + 2\pi k, \theta, \zeta) \Delta u(\theta, \zeta) d\zeta d\theta \\ &= \int_{-\pi}^{\pi} \int_0^{\pi} G_S(z, \theta, \zeta) \Delta u(\theta, \zeta) d\zeta d\theta + \sum_{k \neq 0} \int_{-\pi}^{\pi} \int_0^{\pi} G_S(z + 2\pi k, \theta, \zeta) \Delta u(\theta, \zeta) d\zeta d\theta \end{aligned}$$

where the latter interchange is valid because of uniform convergence on compact sets, and we have isolated the possibly bad term. Heuristically, we use Green's identities with distributions, and Dirac  $\delta$ , and treat  $u$  as a periodic function on the strip. Applying Green's formulæ, we have a lot of boundary terms. However, the values at the top and bottom of the strip go away due to the zero Dirichlet boundary conditions on  $G_S$ , and the values on the sides give a telescoping sum due to the periodicity of  $u$ . What is left over is the delta function integrated against  $u$ , which should just give us  $u$  evaluated at the point. Because the questions are local (since only one term of the series has a problem), in the neighborhood of the singularity, a modified version of the proof for the ball (not using distributions), namely cutting out the singularity and taking a limit (as exemplified in Evans, [30, Ch. 2]) applies and gives us that it is indeed the Green's function. We now can calculate the Robin mass easily, using the coordinates as before, and isolating the bad term:



$$\begin{aligned}
m_C(z) &= \lim_{w \rightarrow z} \left( \frac{1}{2\pi} \log|z - w| + \sum_{k=-\infty}^{\infty} G_S(z, w) \right) = \frac{1}{2\pi} \log|1 - |H(z)|^2| \\
&\quad - \frac{1}{2\pi} \lim_{w \rightarrow z} (\log|H(z) - H(w)| - \log|z - w|) - \frac{1}{2\pi} \sum_{k \neq 0} \log \left| \frac{H(z) - H(z + 2\pi k)}{1 - \overline{H(z)}H(z + 2\pi k)} \right| \\
&= -\frac{1}{2\pi} \left( \log \left| \frac{H'(z)}{1 - |H(z)|^2} \right| + \sum_{k \neq 0} \log \left| \frac{H(z) - H(z + 2\pi k)}{1 - \overline{H(z)}H(z + 2\pi k)} \right| \right)
\end{aligned}$$

where the  $H'(z)$  term comes from the fact that

$$\lim_{w \rightarrow z} \log|H(z) - H(w)| - \log|z - w| = \lim_{w \rightarrow z} \log \left| \frac{H(z) - H(w)}{z - w} \right| = \log|H'(z)|.$$

from the definition of derivative. □

This actually gives us the Green's function for an annulus, because we may conformally map a cylinder to an annulus.

## B.6 Domains with Holes in the Plane and the Bergman Metric

For domains in the plane with finitely many holes, the situation is more complicated. First off, there is a UNIFORMIZATION THEOREM for such domains, similar in spirit to the RIEMANN MAPPING THEOREM:

**B.6.1 Riemann Mapping Theorem.** *Let  $\Omega \subsetneq \mathbb{C}$  be a simply connected domain which is not all of  $\mathbb{C}$ . Then there exists a conformal mapping of  $\Omega$  onto  $\mathbb{D}$ .*

Because of the conformal invariance of the Dirichlet Green's function we therefore can, in principle, calculate the Robin mass of all simply connected domains, by composing with the appropriate conformal map.

Note that a conformal mapping (or any smooth mapping) of  $\Omega$  to  $\mathbb{D}$  can be used to transport a metric via pullback, so we can also pull back the hyperbolic metric to get an invariant metric on  $\Omega$ . Again, this means the automorphism group (conformal self-maps of  $\Omega$ ) actually become isometries, or, more generally, for holomorphic self-maps of  $\Omega$ , hyperbolic distance-reducing (by Schwarz's Lemma).

For  $k$ -connected domains, we have the following

**B.6.2 Uniformization Theorem for  $k$ -Connected Domains.** *Let  $\Omega$  be a  $k$ -connected domain (i.e.  $\hat{\mathbb{C}} \setminus \Omega$  consists of  $k$  connected components  $A_1, \dots, A_k$ ), such that none of the connected components of the complement (with respect to the sphere) is a point. Then there exists a conformal mapping  $\Omega$  onto an annulus with  $k-2$  concentric circular arcs removed (concentric, with the same center as the boundary circles of the annulus as well).*

*Proof, a modernized adaptation of Ahlfors [3].* The first step is to transform the domain conformally until the boundaries become analytic. Let  $A_1, \dots, A_k$  be the connected components of the complement of  $\Omega$  (in the sphere). Let  $A_k$  be the component containing  $\infty$ . First we use the Riemann mapping theorem to map the complement of the unbounded component  $A_k$  (i.e. the domain with all the holes filled in) to the unit disk. This converts the outermost boundary, however irregular it may be (which is the amazing part of the RMT) to the unit circle, a perfectly regular curve. Removing the  $A_j$  for  $j = 1, \dots, k-1$  gives a conformal map of  $\Omega$  to a domain contained in the unit disk. The interior boundary cycles may still, of course, be irregular. We may thus assume that the unbounded component  $A_k$  is just the exterior of the unit disk, and just say that all the images under that transformation mapping are the  $A_j$  for  $j$  up to  $k-1$ .

Now here comes the slightly tricky part. In a particular bounded  $A_j$ , its complement  $A_j^c$  is an unbounded domain containing  $\Omega$ . We may map one of the interior points  $a_j$  to infinity (the mapping  $(z - a_j)^{-1}$  will do nicely) and this makes  $A_j^c$  map to a

simply connected domain. Since  $A_j$  consists of more than one point, this complement is simply connected domain which is not all of  $\mathbb{C}$ . So by RMT again, it maps to the unit disk. Thus we have rendered two possibly irregular curves to curves that are now regular. Repeating the process, now inverting with respect to points the other bounded components  $A_\ell$ , and using the RMT to smooth them out to the unit circle, we have ourselves more analytic boundary curves (since the other already regular curves must remain regular, now being affected by conformal maps in the interior). So we have that our  $k$ -connected domain is conformally equivalent to a  $k$ -connected domain with boundary consisting of analytic curves. A final complex inversion can be made to put the original outer boundary on the outside.

Therefore, the domain now satisfies the interior sphere condition (i.e., one can fit a sufficiently small sphere at the boundary point such that the whole sphere is contained in the domain; see [39, 46] for details) at every point of the boundary, and thus the Dirichlet problem may be solved for any continuous boundary values [39, 46]. We solve for  $k - 1$  harmonic functions  $\omega_j$  which vanish on all  $\partial A_\ell$  not equal to  $\partial A_j$  and equal to 1 on  $\partial A_j$  (the technique of harmonic measures). Each  $\omega_j$  satisfies  $0 < \omega_j(z) < 1$  for all  $z \in \Omega$ , by the Maximum Principle, and moreover, by the Schwarz reflection principle, we may assume that each  $\omega_i$  can be extended a little bit past those boundaries (because either  $\omega_i$  or  $1 - \omega_i$  vanishes on each of these (analytic!) boundaries which is precisely the condition for a Schwarz reflection to exist).

We consider the matrix of periods

$$\alpha_{ij} = \int_{\partial A_i} \star d\omega_j,$$

where  $\star d\omega_j$  is the Hodge dual of the differentials  $d\omega_j$ . This is closely related to the normal derivatives—the measure induced by each  $\star d\omega_j$  is just  $\frac{\partial \omega_j}{\partial n} ds$  where  $ds$

is the line element. The matrix entries are positive on the diagonal, since by the maximum principle, each  $\omega_j(z) \rightarrow 1$  from below as  $z \rightarrow \partial A_j$ , and negative off the diagonal because  $\omega_j(z) \rightarrow 0$  from above as  $z \rightarrow \partial A_i$  with  $i \neq j$ . We show that  $(\alpha_{ij})$  is invertible for  $i, j$  between 1 and  $k-1$ , or equivalently no linear combination  $\sum \lambda_j \omega_j$  has a harmonic conjugate. In the most modern terms, this says:

**B.6.3 Lemma.** *The cohomology classes of the differential forms  $\star d\omega_j$  are a basis for  $\mathfrak{H}_{dR}^1(\Omega)$ .*

*Proof.* To show linear independence, suppose that  $\sum \lambda_i [\star d\omega_i] = 0$  in cohomology. This says that  $\sum \lambda_i \star d\omega_i = d\psi$  for some  $\psi$ . Writing  $\varphi = \sum \lambda_i \omega_i$ , this says that  $\psi$  is a harmonic conjugate of  $\varphi$  and hence  $h = \varphi + i\psi$  is a holomorphic function on  $\Omega$ . By the Schwarz reflection principle, we can assume  $h$  extends holomorphically to (a neighborhood of)  $\bar{\Omega}$ . We claim that  $h$  is constant, hence so are  $\varphi$  and  $\psi$ . By definition of the  $\omega_j$ ,  $\varphi(z) = \sum \lambda_i \omega_i(z) = \lambda_j$  whenever  $z \in \partial A_j$  for  $j < k$ , and also  $\varphi(z) = 0$  when  $z \in \partial A_k$ . So, in particular,  $h$  maps every boundary curve of  $\partial A_j$  to a vertical segment in  $\mathbb{C}$ . However since  $\Omega$  is bounded (we can make all such  $k$ -connected domains bounded via an additional complex inversion), all the  $\partial A_j$  are compact, and so  $h(\partial A_j)$  are also compact (i.e. bounded and closed) vertical segments. Those vertical segments only determine one connected component in their complements, so if  $\tau$  is any point off one of these segments, their winding numbers about  $\tau$  must be zero (because that single connected component is necessarily unbounded):

$$\frac{1}{2\pi i} \oint_{\partial\Omega} \frac{h'(\zeta)}{h(\zeta) - \tau} d\zeta = \frac{1}{2\pi i} \oint_{h(\partial\Omega)} \frac{dw}{w - \tau} = \frac{1}{2\pi i} \sum_{i=1}^k \oint_{h(\partial A_i)} \frac{dw}{w - \tau} = 0.$$

In other words, the count, with multiplicity, of points  $z \in \Omega$  such that  $h(z) = \tau$  is zero, that is,  $h$  maps  $\bar{\Omega}$  nowhere except for the union of finite segments. By the Open Mapping Theorem, this is only possible if  $h$  is constant.

Now  $\varphi$  is constant, and we can evaluate what that constant actually is by evaluating it somewhere it is known: since  $\varphi(z) \rightarrow 0$  as  $z$  approaches  $\partial A_k$ , the outermost boundary, this shows that the constant must be 0. Approaching each  $\partial A_i$  we find, from the definition of  $\varphi$ , that  $\varphi(z) = \lambda_i$  on  $\partial A_i$ , for all  $i$ ,  $1 \leq i < k$ . Therefore all the  $\lambda_i$  are all zero. This shows, in particular, that the (square) period matrix consists of linearly independent columns, and is thus invertible.

To show that it spans, we suppose  $\xi$  is a closed 1-form on  $\Omega$ , and let  $\mu_i = \int_{\partial A_i} \xi$ . Let  $\lambda_i$  be the inverse of the period matrix applied to the coefficients  $\mu_i$ . We claim  $\xi - \sum_{i=1}^{k-1} \lambda_i \star d\omega_i$  is exact. Integrating the form over  $\partial A_j$  we get

$$\int_{\partial A_j} \xi - \sum \lambda_i \star d\omega_i = \mu_j - \sum_i \int_{\partial A_j} \star d\omega_i = \mu_j - \sum_i \alpha_{ji} \lambda_i = \mu_j - \mu_j = 0.$$

Thus  $\xi - \sum \lambda_i \star d\omega_i$  vanishes over all the boundaries which are a basis for the homology of  $\Omega$ . Therefore its integrals are independent of path and thus it is exact.  $\square$

We continue the proof of the theorem. Consider the closed but inexact differential form

$$\eta = \star d(\log|z - a|) = d\text{“arg}(z - a)\text{”}$$

where  $a \in A_1$ . Because the  $[\star d\lambda_j]$  are a basis in cohomology, there exist unique real scalars  $\lambda_i$  such that

$$[\eta] = \sum_{i=1}^{k-1} \lambda_i [\star d\omega_i]$$

or equivalently,  $\eta = \sum_i \lambda_i \star d\omega_i - d\psi$  for some exact differential  $d\psi$ . Moreover, since  $\int_{\partial A_j} \eta = 0$  when  $1 < j < k$  but is  $2\pi$  when  $j = 1$ , this shows that the  $\lambda_i$  are not all zero (the  $\lambda_i$  are then computed by applying the inverse of the period matrix to the vector  $2\pi(1, 0, \dots, 0)$ ), and thus  $u = \sum \lambda_i \omega_i$  has no harmonic conjugate. In classical terminology, we pretend that it does, and get *multivalued* harmonic conjugates  $v$

such that  $f = u + iv$  is a multivalued holomorphic function with period  $2\pi i$  along  $\partial A_1$ . Taking its exponential  $F = e^f$  gives a genuine holomorphic function, because it precisely kills off the  $2\pi i$  ambiguity about  $\partial A_1$  that  $f$  suffers.

In more modern terms, since arguments with multivalued functions are imprecise, we consider the exact differential  $d\psi = -\eta + \sum \lambda_i \star d\omega_i = \star du - \eta$ , and, writing

$$h(z) = u(z) - \log|z - a| + i\psi,$$

we see that  $h$  is holomorphic (and single-valued) and  $e^{h(z)} = (z - a)^{-1} e^{f(z)}$ . In other words, we can get  $F(z) = e^{f(z)}$  by more legitimate means by instead defining it to be  $F(z) = (z - a)e^{h(z)}$ .

We claim that  $F$  actually maps  $\Omega$  conformally onto the type of domain we are looking at. First,

$$|F(z)| = |z - a|e^{\operatorname{Re}(h(z))} = e^{u(z)}.$$

But we know that  $u = \sum_i \lambda_i \omega_i$  vanishes on the boundary  $\partial A_k$  and is equal to  $\lambda_j$  on each  $\partial A_j$  by virtue of the construction of the  $\omega_i$ . So  $F$  maps the outer boundary  $\partial A_k$  to the unit circle and each inner boundary  $\partial A_j$  to other arcs of circles centered about the origin. Note that  $F$  never vanishes in  $\Omega$  and so since  $\partial\Omega$  is homologous to 0 with respect to  $\Omega$  (i.e. a boundary!), by Stokes' Theorem, we have:

$$0 = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{F'(z)}{F(z)} dz = \frac{1}{2\pi i} \sum_{i=1}^k \oint_{\partial A_i} \frac{F'(z)}{F(z)} dz.$$

But

$$\frac{F'(z)}{F(z)} = \frac{(z - a)e^{h(z)} h'(z) + e^{h(z)}}{(z - a)e^{h(z)}} = h'(z) + \frac{1}{z - a}.$$

which is just  $f'(z)$  for  $f$  that ill-defined function.  $h'(z)$  is holomorphic and has a primitive  $h(z)$  (i.e.  $h'(z)dz$  is an exact differential) so its integral vanishes over all

cycles, not just those that are homologous to 0. So this eliminates most of the terms in the sum:

$$0 = \frac{1}{2\pi i} \sum_{i=1}^k \oint_{\partial A_i} \frac{F'(z)}{F(z)} dz = \frac{1}{2\pi i} \left( \oint_{\partial A_1} \frac{F'(z)}{F(z)} dz + \oint_{\partial A_k} \frac{F'(z)}{F(z)} dz \right) = 1 + \frac{1}{2\pi i} \oint_{\partial A_k} \frac{F'(z)}{F(z)} dz.$$

This says the winding numbers of each  $F(\partial A_j)$  about the origin is 0 (i.e. are not full circles), except for  $i = 1$  and  $i = k$  in which case they are 1 and  $-1$ , respectively (because we are keeping track of orientations). So the  $F(\partial A_1)$  and  $F(\partial A_k)$  fully wind around the origin (i.e. are full circles), showing us that indeed the image boundary curves yield something that looks like two bounding circles of an annulus, with  $k - 2$  concentric slits.

Now if  $\tau$  is any point in the annulus between the two bounding circles, but not on any of the other arcs, then

$$\frac{1}{2\pi i} \oint_{\partial \Omega} \frac{F'(z)}{F(z) - \tau} dz = \frac{1}{2\pi i} \oint_{\partial A_1 + \partial A_k} \frac{F'(z)}{F(z) - \tau} dz = \frac{1}{2\pi i} \oint_{F(\partial A_1) + F(\partial A_k)} \frac{dw}{w - \tau} = 1,$$

because  $\tau$  is in the unbounded component determined by the inner circle and the circular arcs, but is in the bounded component determined by the outer circle. This shows that  $\tau$  is taken on as a value once and exactly once in  $\Omega$ .

Similarly, if  $\tau$  is inside the inner circle

$$\frac{1}{2\pi i} \oint_{\partial \Omega} \frac{F'(z)}{F(z) - \tau} dz = 0$$

because  $\tau$  is in the same connected component as 0 which we saw is never taken as a value on  $\Omega$  (it is enclosed by both circles, but with opposite orientations, so they cancel). Finally, if  $\tau$  is outside the outermost circle, then the winding number is 0 for all the circles and arcs, hence it is 0 overall. So the value  $\tau$  is in the image  $F(\Omega)$  if and

only if  $\tau$  lies between the two bounding circles of the annulus and off any of the arcs. Thus  $F$  is a biholomorphism (on  $\Omega$ . If extended to  $\bar{\Omega}$ , there may be double points; this may be verified using Cauchy principal values).

We are done with the proof, but as a final note, we can check which circle is inner and which is outer. First, since  $u = \sum_{i=1}^{k-1} \lambda_i \omega_i$  is in fact the solution to the Dirichlet problem, it assumes boundary values  $\lambda_i$  on  $\partial A_i$  and 0 on  $\partial A_k$ . Since  $F(\partial A_k)$  and  $F(\partial A_1)$  are full circles, it follows that all the  $\lambda_j$  for  $1 < j < k$  cannot be the min or max (by connectivity of the domain and the fact that the arcs  $F(\partial A_j)$  are not full circles). Therefore either 0 or  $\lambda_1$  is the maximum. However, we have, by the above computations with the argument principle, that

$$-2\pi = \oint_{\partial A_k} \eta = \oint_{\partial A_k} \star du = \oint_{\partial A_k} \frac{\partial u}{\partial n} ds$$

where we take the outward pointing normal to  $\partial A_k$ . Since  $ds$  is a positive measure, this shows that  $\frac{\partial u}{\partial n} < 0$  somewhere on  $\partial A_k$ , or in a small enough neighborhood of such a point,  $u$  is *decreasing* to 0 as  $z$  approaches  $A_k$ . By the maximum principle, it follows that 0 must actually be the global minimum. Therefore, in particular,  $\lambda_1 > 0$ , and  $e^{\lambda_1} > 1$ , so that  $\partial A_1$  corresponds to the outer circle and  $\partial A_k$  corresponds to the inner circle (it is the unit circle). (this also shows that the conformal mapping here has an extra inversion. We could rectify this via another complex inversion (the genuine  $z \mapsto 1/z$  but this is unnecessary unless one wants to specify the mapping uniquely by saying, for example, that a certain point in the domain *must* map to a certain other point. □

So of course, it suffices to prove theorems on Green's function, etc. for annuli with slits removed. Again, it is interesting to not only look at the Euclidean case, but in the case of certain canonical metrics defined on such domains.



It turns out that the Euclidean Green's function can be used to construct an invariant metric similar to the hyperbolic metric on  $\mathbb{D}$ , called the POINCARÉ-BERGMAN METRIC. We refer to [60] for the following method of construction (the ideas date back to the work of Bergman).

**B.6.4 Definition.** Let  $\Omega$  be a domain and let

$$A^2(\Omega) = \{f \in \mathcal{L}^2(\Omega) : f = \text{a holomorphic function a.e.}\}.$$

Usually, defining subspaces of smooth functions in  $\mathcal{L}^2$  is not such a great thing to do, because they are usually not closed in the  $\mathcal{L}^2$  norm (i.e. not complete). It is true, however, in the case of holomorphic functions:

**B.6.5 Theorem.**  $A^2(\Omega)$  is a closed subspace of  $\mathcal{L}^2(\Omega)$ , and hence also a Hilbert space with the same inner product  $(f, g) = \int f \bar{g}$ . Moreover, for each compact  $K \subseteq \Omega$ , there exists a constant  $C_K$  depending only on  $K$  such that

$$\|f\|_K = \sup_{z \in K} |f(z)| \leq C_K \|f\|_{\mathcal{L}^2(\Omega)},$$

that is,  $\mathcal{L}^2$  convergent sequences of functions in  $A^2(\Omega)$  also converge uniformly on compact sets.

By elementary Hilbert space theory, it follows therefore that the Riesz Representation Theorem holds in  $A^2(\Omega)$  and it has an orthonormal basis.

**B.6.6 Definition.** The BERGMAN KERNEL is the function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  such for every  $f \in A^2(\Omega)$  and  $z \in \Omega$ ,

$$\int_{\Omega} K(z, \zeta) f(\zeta) dA(\zeta) = f(z).$$

In other words, it is the “identity matrix,” or represents the evaluation functional. The reason why we can actually represent it as such (in general, we need the  $\delta$  distribution

to do this for continuous functions!) is because the mapping  $e_z$  given by

$$e_z(f) = f(z)$$

is actually a bounded linear functional on  $\mathcal{A}^2$ :

$$|e_z(f)| = |f(z)| \leq C_{\{z\}} \|f\|_{\mathcal{L}^2(\Omega)}$$

where  $C_{\{z\}} < \infty$  is that constant on the compact set  $K = \{z\}$  in the lemma above (in fact just  $\frac{1}{\sqrt{\pi}\delta}$  works, as soon as  $\delta$  is small enough for  $B_\delta(z) \subseteq \Omega$ ). By the Riesz Representation Theorem, there exists  $k_z \in \mathcal{L}^2(\Omega)$  such that

$$f(z) = e_z(f) = (f, k_z)_{\mathcal{L}^2(\Omega)} = \int_{\Omega} f(\zeta) \overline{k_z(\zeta)} dA(\zeta).$$

We just define  $K(z, \zeta) = \overline{k_z(\zeta)}$ . It follows that  $\zeta \mapsto K(z, \zeta)$  is antiholomorphic.

**B.6.7 Theorem.** *The Bergman kernel is the unique function  $K : \Omega \times \Omega \rightarrow \mathbb{C}$  satisfying*

1.  $\int_{\Omega} K(z, \zeta) f(\zeta) dA(\zeta) = f(z)$  (called the REPRODUCING PROPERTY)
2.  $K(z, \zeta)$  is antiholomorphic in  $\zeta$ .
3.  $K(z, \zeta) = \overline{K(\zeta, z)}$  (and thus  $K$  is holomorphic in its first variable  $z$ ) (CONJUGATE SYMMETRY).

**B.6.8 Theorem.** *Let  $K$  be the Bergman kernel for  $\Omega$ . Then if  $(\phi_n)$  is any orthonormal basis for  $A^2(\Omega)$ , then*

$$K(z, \zeta) = \sum_{n=1}^{\infty} \phi_n(z) \overline{\phi_n(\zeta)}.$$

The proof is simply that we show it satisfies the 3 properties of a Bergman kernel. Thus  $K(z, z) \geq 0$  and by completeness of an orthonormal basis, never actually

is equal to 0. Thus  $\log(K(z, z))$  is well-defined.

**B.6.9 Theorem.** *Let  $\Omega$  be a domain with Bergman kernel  $K$ . Then*

$$F(z) = \frac{\partial^2}{\partial z \partial \bar{z}} \log K(z, z) = -\frac{1}{4} \Delta \log K(z, z).$$

*defines a conformal factor for the Euclidean metric  $g$  on  $\Omega$ . The metric  $Fg$  is called the POINCARÉ-BERGMAN METRIC on  $\Omega$ .*

It's not entirely obvious that  $F > 0$ , however.

**B.6.10 Theorem.** *The Bergman kernel for the disk is*

$$K(z, \zeta) = \frac{1}{\pi} \frac{1}{(1 - z\bar{\zeta})^2}.$$

*Proof.* The functions  $z^k$  for  $k \geq 0$  are square integrable, holomorphic functions on  $\mathbb{D}$ , and

$$\int_{\mathbb{D}} |z^k|^2 dA(z) = 2\pi \int_0^1 r^{2k+1} dr = \frac{2\pi}{2k+2} = \frac{\pi}{k+1}.$$

Therefore the functions

$$\sqrt{\frac{k+1}{\pi}} z^k$$

form an orthonormal set in  $\mathbb{D}$ . They must form an orthonormal basis in  $A^2(\mathbb{D})$  since all holomorphic functions on  $\mathbb{D}$  are expressible by power series with radius of convergence  $\geq 1$ , and so if  $(f, z^k)_{\mathcal{L}^2(\mathbb{D})} = 0$  for all  $k$ ,  $f \equiv 0$ . Therefore

$$\begin{aligned} \pi K(z, \zeta) &= \sum_{k=0}^{\infty} (k+1) z^k \bar{\zeta}^k = \sum_{k=0}^{\infty} (k+1) (z\bar{\zeta})^k = \sum_{k=1}^{\infty} k w^{k-1} \Big|_{w=z\bar{\zeta}} = \frac{d}{dw} \Big|_{w=z\bar{\zeta}} \sum_{k=1}^{\infty} w^k \\ &= \frac{d}{dw} \left( \frac{1}{1-w} \right) \Big|_{w=z\bar{\zeta}} = -\frac{-1}{(1-z\bar{\zeta})^2} = \frac{1}{(1-z\bar{\zeta})^2}. \end{aligned}$$

□

Let us then compute the Poincaré-Bergman metric for  $\mathbb{D}$ . We have that

$$K(z, z) = \frac{1}{\pi(1 - |z|^2)^2}$$

so that its logarithm is

$$\log K(z, z) = -2\log(1 - |z|^2) - \log \pi.$$

Taking the derivative with respect to  $\bar{z}$ :

$$\frac{\partial}{\partial \bar{z}} \log K(z, z) = \frac{2z}{1 - |z|^2},$$

and finally we have

$$F(z) = \frac{\partial^2}{\partial z \partial \bar{z}} \log K(z, z) = 2 \frac{(1 - z\bar{z})1 - z(-\bar{z})}{(1 - |z|^2)^2} = \frac{2}{(1 - |z|^2)^2}.$$

This differs by a factor of 2 from our usual hyperbolic metric (i.e. it is a hyperbolic disk with curvature  $-4/2 = -2$ ), and thus distances near the origin look like  $\sqrt{2}$  times Euclidean distance. Because the hyperbolic disk has rather nice properties, this shows that the hyperbolic metric is canonical yet in another sense: it is given by (a suitable rescaling) of the Poincaré-Bergman metric.

Now let  $A$  be an annulus  $\{z \in \mathbb{C} : r < |z| < R\}$ . Now we have that  $(z^k)$  are an orthogonal basis for all  $k \in \mathbb{Z}$  by Laurent expansions (as long as the annulus is not degenerate i.e. its inner radius  $r$  is positive). To normalize, we observe

$$\int_A |z^k|^2 dA(z) = 2\pi \int_r^R \rho^{2k+1} d\rho = \begin{cases} \frac{2\pi(R^{2k+2} - r^{2k+2})}{2k+2} & \text{if } k \neq -1 \\ 2\pi \log\left(\frac{R}{r}\right) & \text{if } k = -1. \end{cases}$$

Thus,

$$\frac{z^k \sqrt{2k+2}}{\sqrt{2\pi(R^{2k+2} - r^{2k+2})}}$$

for  $k \neq -1$  and

$$\frac{1}{z\sqrt{2\pi\log(R/r)}}$$

form an orthonormal basis. The Bergman kernel is thus

$$\frac{1}{2\pi\log(R/r)(z\bar{\zeta})} + \frac{1}{2\pi} \sum_{k \neq -1} \frac{(2k+2)z^k \bar{\zeta}^k}{R^{2k+2} - r^{2k+2}}$$

Evaluation of this sum is quite intractable without further information on  $R$  and  $r$ .

The goal here is to understand what the canonical Poincaré-Bergman metric looks like on annuli with slits removed, and see if the Robin mass of such domains is anything special, and to compare the Euclidean (Dirichlet and Neumann) Robin mass of such things with to the Robin mass of the canonical metric defined on them. We use the transformation formula: given the Bergman Kernel,

$$\tilde{m}(z) = \frac{1}{4\pi} \log \left( \frac{\partial^2}{\partial z \partial \bar{z}} \log K(z, z) \right) + m(z).$$

This is just an application of Theorem B.2.3, taking  $u = \frac{1}{2} \log F$ . I strongly suspect that it will be constant, especially given the following

**B.6.11 Theorem.** *Let  $K$  be the Bergman kernel for  $\Omega$ . Then if  $G$  is the Euclidean Green's function for  $\Omega$ ,*

$$K(z, \zeta) = 4 \frac{\partial^2}{\partial \bar{z} \partial \zeta} G(\zeta, z)$$

Clearly, since we are considering  $K(z, z)$ , we have to let  $\zeta$  tend to  $z$  in the above, this should be familiar from similar properties of the Robin mass.

## B.7 Conclusion and Future Work

We have explored some interesting geometrical concepts for domains in the plane, in particular, geometries associated with invariants, and various hyperbolic geometries given by Green's functions (the Bergman metric). As noted in Okikiolu's work [78, 77, 76] and others in closely related research [71, 70, 100, 101], interesting geometry arises by considering extremal problems for the mass and other related quantities. Namely, we wish to find critical metrics for various functionals involving the mass. For example, integrating the mass yields the  $\Delta$ -mass, which leads to the study of spectral zeta functions. Another interesting invariant is given by an infinite-dimensional generalization of the determinant of the Laplacian, also viewed as a function of the metric [75].

As variational problems, the concepts above, of course, lead to interesting, but difficult nonlinear differential equations. Variational formulations, as noted in previous chapters, are also suited to approximation by some form of finite element method. Again, one of the general goals for numerical solution to such partial differential equations is to gain a more intuitive understanding of the concepts and hopefully generate more conjectures. Attempting to visualize all these concepts is indeed what lead the author to numerical analysis in the first place. It is unfortunate, however, that we will not be able to achieve this original goal in this current work, as there is much more work to be done in nonlinear equations. However, with the frameworks presented in the previous chapters (and extensions proved), solid groundwork has been laid for future endeavors.

# Bibliography

- [1] R. Abraham and J. E. Marsden. *Foundations of Mechanics*. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1985.
- [2] R. A. Adams and J. F. Fournier. *Sobolev Spaces*. Academic Press, San Diego, CA, second edition, 2003.
- [3] L. V. Ahlfors. *Complex Analysis*. McGraw-Hill, 1979.
- [4] D. Arnold and H. Chen. Finite element exterior calculus for parabolic problems. *arXiv:1209.1142*, 2012.
- [5] D. Arnold, R. Falk, and R. Winther. Finite element exterior calculus, homological techniques, and applications. *Acta Numerica*, pages 1–155, 2006.
- [6] D. Arnold, R. Falk, and R. Winther. Finite element exterior calculus: from Hodge theory to numerical stability. *Bulletin of the American Mathematical Society*, 47(2):281–354, 2010.
- [7] I. Babuška. Error bounds for the finite element method. *Numerische Mathematik*, 16:322–333, 1971.
- [8] A. Bossavit. Whitney forms: a class of finite elements for three-dimensional computations in electromagnetism. *Science, Measurement and Technology, IEE Proceedings*, 135(8):493–500, Nov 1988.
- [9] R. Bott and L. W. Tu. *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics. Springer, New York, NY, 1982.
- [10] D. Braess. *Finite Elements*. Cambridge University Press, Cambridge, MA, 1997.
- [11] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, third edition, 2007.
- [12] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2002.

- [13] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, New York, NY, second edition, 2002.
- [14] J. Brüning and M. Lesch. Hilbert Complexes. *J. Funct. Anal.*, 108(1):88–132, August 1992.
- [15] W. L. Burke. *Applied Differential Geometry*. Cambridge University Press, Cambridge, UK, 1985.
- [16] J. W. Cahn, P. Fife, and O. Penrose. A phase field model for diffusion induced grain boundary motion. *Acta Mater.*, 45:4397–4413, 1997.
- [17] Y. Choquet-Bruhat and C. DeWitt-Morette. *Analysis, Manifolds and Physics*, volume I. North-Holland, Amsterdam, 2002.
- [18] B. Chow and D. Knopf. *The Ricci Flow: An Introduction*. American Mathematical Society, Providence, RI, 2004.
- [19] B. Chow, P. Lu, and L. Ni. *Hamilton's Ricci Flow*. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2006.
- [20] D. Christodoulou and S. Klainerman. *The global nonlinear stability of the Minkowski space*, volume 41 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1993.
- [21] G. de Rham. *Variétés Différentiables: Formes, Courants, Formes Harmoniques*. Hermann, Paris, 1973.
- [22] K. Deckelnick and G. Dziuk. Numerical approximation of mean curvature flow of graphs and level sets. In P. Colli and J. Rodrigues, editors, *Mathematical Aspects of Evolving Interfaces*, 2003.
- [23] K. Deckelnick, G. Dziuk, and C. M. Elliott. Computation of geometric partial differential equations and mean curvature flow. *Acta Numer.*, 14:139–232, 2005.
- [24] A. Demlow. Higher-order finite element methods and pointwise error estimates for elliptic problems on surfaces. *SIAM J. Numer. Anal.*, 47(2):805–827, 2009.
- [25] A. Demlow and G. Dziuk. An adaptive finite element method for the Laplace-Beltrami operator on surfaces. *SIAM J. Numer. Anal.*, 2006. to appear.
- [26] M. P. do Carmo. *Riemannian Geometry*. Birkhäuser Boston, 1992.
- [27] G. Dziuk. Finite elements for the Beltrami operator on arbitrary surfaces. In *Partial differential equations and calculus of variations*, pages 142–155, Berlin, 1988. Springer.



- [28] G. Dziuk and C. M. Elliott. Finite elements on evolving surfaces. *IMA J. Num. Anal.*, 27:262–292, 2007.
- [29] G. Dziuk and J. E. Hutchinson. Finite element approximations to surfaces of prescribed variable mean curvature. *Numer. Math.*, 102(4):611–648, 2006.
- [30] L. C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 1998.
- [31] FETK. The Finite Element ToolKit. <http://www.FETK.org>.
- [32] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics: Mechanics, radiation, and heat*, volume I. Addison Wesley, Commemorative Issue edition, 1989.
- [33] H. Flanders. *Differential Forms with Applications to the Physical Sciences*. Dover Publications, New York, NY, 1989.
- [34] G. B. Folland. *Real Analysis*. John Wiley & Sons, Inc., New York, NY, second edition, 1999.
- [35] E. A. Forgy. *Differential Geometry in Computational Electromagnetics*. PhD thesis, University of Illinois at Urbana-Champaign, 2002.
- [36] T. Frankel. *The Geometry of Physics*. Cambridge University Press, Cambridge, UK, 2004.
- [37] S. Fucik and A. Kufner. *Nonlinear Differential Equations*. Elsevier Scientific Publishing Company, New York, NY, 1980.
- [38] I. M. Gelfand and G. E. Shilov. *Generalized Functions*, volume 4. Academic Press, 1964.
- [39] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001.
- [40] A. Gillette and M. Holst. Finite element exterior calculus for evolution problems. Submitted for publication. Available as arXiv:1202.1573 [math.NA].
- [41] H. Goldstein. *Classical Mechanics*. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1980.
- [42] D. Griffiths. *Introduction to Electrodynamics*. Addison Wesley, 3rd edition, 1999.
- [43] R. Haberman. *Elementary Applied Partial Differential Equations*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1998.
- [44] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer-Verlag, Berlin, Germany, 2001.

- [45] R. S. Hamilton. Three-manifolds with positive Ricci curvature. *J. Diff. Geom.*, 17, 1982.
- [46] Q. Han and F. Lin. *Elliptic Partial Differential Equations*. Courant Lecture Notes. American Mathematical Society, Providence, RI, 2nd edition, 2011.
- [47] M. Hirsch, S. Smale, and R. Devaney. *Differential Equations, Dynamical Systems and an Introduction to Chaos*. Elsevier Scientific Publishing Company, New York, NY, 2004.
- [48] K. Hoffman and R. Kunze. *Linear Algebra*. Pearson, second edition, 1971.
- [49] M. Holst. MCLite: An adaptive multilevel finite element MATLAB package for scalar nonlinear elliptic equations in the plane. User's Guide to the MCLite software package.
- [50] M. Holst and A. Stern. Geometric variational crimes: Hilbert complexes, finite element exterior calculus, and problems on hypersurfaces. *Found. Comput. Math.*, 12(3):263–293, 2012. Available as arXiv:1005.4455 [math.NA].
- [51] M. Holst and A. Stern. Semilinear mixed problems on Hilbert complexes and their numerical approximation. *Found. Comput. Math.*, 12(3):363–387, 2012. Available as arXiv:1010.6127 [math.NA].
- [52] M. J. Holst. Mclite: An adaptive multilevel finite element matlab package for scalar nonlinear elliptic equations in the plane. Technical report, UCSD, 1997.
- [53] J. H. Hubbard and B. B. Hubbard. *Vector Calculus and Linear Algebra: A Differential Forms Approach*. Matrix Editions, 4th edition, 2011.
- [54] T. J. R. Hughes. *The Finite Element Method*. Dover Publications, New York, NY, 2000.
- [55] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, MA, 1996.
- [56] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, Hoboken, NJ, 1998.
- [57] C. Johnson and V. Thomée. Error estimates for some mixed finite element methods for parabolic type problems. *RAIRO Anal. Numér.*, 15(1):41–78, 1981.
- [58] J. Jost. *Riemannian Geometry and Geometric Analysis*. Universitext. Springer-Verlag, New York, NY, 6th edition, 2011.
- [59] J. L. Kazdan and F. W. Warner. Curvature functions for compact 2-manifolds. *Annals of Mathematics*, 99(1):14–47, 1974.

- [60] S. G. Krantz. *Geometric Function Theory*. Cornerstones. Birkhäuser, 2006.
- [61] S. Lang. *Differential and Riemannian Manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer, 3rd edition, 1995.
- [62] J. M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, second edition, 2012.
- [63] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge University Press, Cambridge, MA, 2004.
- [64] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. AMS, 1997.
- [65] A. Logg, K.-A. Mardal, and G. N. Wells. *The FEniCS Book*, volume 84 of *Lecture Notes in Computational Science and Engineering*. Springer, 2011.
- [66] C. Lubich and D. Mansour. Variational discretization of linear wave equations on evolving surfaces. *Math. Comp.*, 84:513–542, 2015.
- [67] J. E. Marsden and A. J. Tromba. *Vector Calculus*. Freeman, fourth edition, 1996.
- [68] U. F. Mayer and G. Simonnett. Classical solutions for diffusion induced grain boundary motion. *J. Math. Anal.*, 234(660-674), 1999.
- [69] C. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman & Co., 1973.
- [70] C. Morpurgo. Zeta functions on  $S^2$ . In J. R. Quine and P. Sarnak, editors, *Extremal Riemann Surfaces (San Francisco 1995)*, Contemporary Mathematics, pages 213–225. American Mathematical Society, 1997.
- [71] C. Morpurgo. Sharp inequalities for functional integrals and traces of conformally invariant operators. *Duke Math. J.*, 114:477–553, 2002.
- [72] J.-C. Nédélec. Mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 35(3):315–341, 1980.
- [73] J.-C. Nédélec. A new family of mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 50(1):57–81, 1986.
- [74] T. Needham. *Visual Complex Analysis*. Oxford University Press, 2000.
- [75] K. Okikiolu. Critical metrics for the determinant of the laplacian in odd dimensions. *Annals of Mathematics*, 153(2):471–531, 2001.
- [76] K. Okikiolu. Extremals for Logarithmic Hardy-Littlewood-Sobolev inequalities on compact manifolds. *Geometric and Functional Analysis*, 17:1655–1684, 2008.

- [77] K. Okikiolu. A negative mass theorem for surfaces of positive genus. Available as arXiv:0810.0724 [math.SP], Oct 2008.
- [78] K. Okikiolu. A negative mass theorem for the 2-torus. Available as arXiv:0711.3489 [math.SP], Jul 2008.
- [79] H.-O. Peitgen, H. Jürgens, and D. Saupe. *Chaos and Fractals: New Frontiers of Science*. Springer-Verlag, 1992.
- [80] G. Perelman. The entropy formula for the Ricci flow and its geometric applications. Available as arXiv:math.DG/0211159.
- [81] G. Perelman. Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. Available as arXiv:math/0307245v1.
- [82] G. Perelman. Ricci flow with surgery on three-manifolds. Available as arXiv:math.DG/0303109.
- [83] P. Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics. Springer-Verlag, New York, NY, 2nd edition, 2006.
- [84] R. Picard. An elementary proof for a compact imbedding result in generalized electromagnetic theory. *Mathematische Zeitschrift*, 187:151–164, 1984.
- [85] Z. Popović and B. D. Popović. *Introductory Electromagnetics*. Prentice Hall, Upper Saddle River, NJ, 2000.
- [86] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [87] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*, volume 37 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 2nd edition, 2006.
- [88] P. A. Raviart and J. Thomas. A mixed finite element method for 2nd order elliptic problems. In I. Galligani and E. Magenes, editors, *Mathematical aspects of the Finite Elements Method*, Lectures Notes in Math. 606, pages 292–315. Springer, Berlin, 1977.
- [89] M. Renardy and R. Rogers. *An introduction to partial differential equations*, volume 13. Springer Verlag, 2nd edition, 2004.
- [90] R. Resnick, D. Halliday, and K. S. Krane. *Physics*, volume One. John Wiley & Sons, 1992.
- [91] R. Resnick, D. Halliday, and K. S. Krane. *Physics*, volume Two. John Wiley & Sons, 1992.
- [92] W. Rudin. *Real & Complex Analysis*. McGraw-Hill, New York, NY, 1987.

- [93] H. M. Schey. *Div, Grad, Curl and all That: An Informal Text on Vector Calculus*. W. W. Norton & Company, 3rd edition, 1997.
- [94] R. Schoen. Conformal deformation of a Riemannian metric to constant scalar curvature. *J. Differential Geom*, 20(2):479–495, 1984.
- [95] M. D. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume II. Publish or Perish, Houston, TX, third edition, 1999.
- [96] I. Stakgold and M. Holst. *Boundary Value Problems: Theory and Applications*. John Wiley & Sons, Inc., New York, NY, 496 pages, October 2012. The preface and table of contents of the book are available at: <http://ccom.ucsd.edu/~mholst/pubs/dist/StHo2011b-preview.pdf>.
- [97] I. Stakgold and M. Holst. *Green's Functions and Boundary Value Problems*. John Wiley & Sons, Inc., New York, NY, third edition, 888 pages, February 2011. The preface and table of contents of the book are available at: <http://ccom.ucsd.edu/~mholst/pubs/dist/StHo2011a-preview.pdf>.
- [98] E. M. Stein and R. Shakarchi. *Fourier Analysis: An Introduction*, volume I of *Princeton Lectures in Analysis*. Princeton University Press, 2003.
- [99] E. M. Stein and R. Shakarchi. *Functional Analysis*, volume IV of *Princeton Lectures in Analysis*. Princeton University Press, 2011.
- [100] J. Steiner. *Green's Functions, Spectral Invariants, and a Positive Mass on Spheres*. PhD thesis, UCSD, 2003.
- [101] J. Steiner. A geometrical mass and its extremal properties for metrics on  $S^2$ . *Duke Math. J.*, 129:63–86, 2005.
- [102] G. Strang. *Linear Algebra and its Applications*. Saunders HBJ, 1988.
- [103] G. Strang and G. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, Englewood Cliffs, NJ, 1973.
- [104] W. Strauss. *Partial Differential Equations: An Introduction*. John Wiley & Sons, 1992.
- [105] M. E. Taylor. *Partial Differential Equations*, volume I. Springer-Verlag, New York, NY, 1996.
- [106] V. Thomée. *Galerkin finite element methods for parabolic problems*. Springer Verlag, 2006.
- [107] F. W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*, volume 94 of *Graduate Texts in Mathematics*. Springer, 1971.

- [108] G. Weinreich. *Geometrical Vectors*. Chicago Lectures in Physics. University of Chicago Press, Chicago, IL, 1998.
- [109] R. O. Wells. *Differential Analysis on Complex Manifolds*, volume 65 of *Graduate Texts in Mathematics*. Springer, 3rd edition, 2008.
- [110] M. Wheeler. A priori  $L^2$  error estimates for Galerkin approximations to parabolic partial differential equations. *SIAM J. Numer. Anal.*, pages 723–759, 1973.
- [111] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.
- [112] K. Yosida. *Functional Analysis*. Springer-Verlag, Berlin, Germany, 1980.

# Index

- affine invariant, **133**
- algebraic operator on sections, **17**
- assembly process (finite element), **125, 137**
- average scalar curvature, **149, 239**
- Bergman kernel, **314**
  - conjugate symmetry, **315**
  - reproducing property, **315**
- best approximation, **5, 120**
- bilinear form, **243**
- Blaschke factor, **280**
- Bochner mixed weak parabolic problem, **196**
- Bochner spaces, **89, 90, 193**
- boundary condition, **11, 59**
  - essential, **3, 48, 55, 77**
  - natural, **3, 55**
    - for the mixed weak Hodge Laplacian problem, **46**
- boundary value problem, **11**
  - Hodge Laplacian version, **43**
- bounded away from zero, **63**
- bounded cochain projection
  - for open subsets of  $\mathbb{R}^n$ , **217**
- canonical choices
  - none for orientations in a general vector space, **21**
- Cauchy problem, **11**
- Céa's Lemma, **121**
- chain homotopy, **133**
- classical solution, **43, 61**
- coboundary, **76**
  - in Hilbert complexes, **171**
- cochain property, **76, 171**
- cocycle, **171**
  - in Hilbert complexes, **78**
- codifferential, **234**
- coercive, **63**
- coercivity constant, **70**
- cohomology
  - in Hilbert complexes, **78, 171**
  - reduced, **78, 171**
- commutation formula for Hodge duals
  - and exterior derivatives, **29**
- compactness property, **86**
- completeness
  - of  $H\Omega$  spaces, **34**
- conformal factor equation, **237**
- conformal transformation, **281**
- conforming mesh, **102**
- constitutive relations, **29, 98**
  - defining geometry, **69**
- convention for conjugation, **26**
- corrector function, **262**
- current (linear functional on forms), *see* distribution
- $d$  commutes with pullback, **19**
- de Rham complex
  - with boundary conditions, **48**
- degrees of freedom, **135**
- determinants, **15**
- differential form, **14**
  - polynomial finite elements, **137, 216**
  - pullback, **17**
    - Sobolev spaces of, **30–35, 235**
- differential pseudoform, **22**
- directional derivative, **18**
- Dirichlet conditions, **11, 259**
- Dirichlet problem, **259**

- Dirichlet Robin mass, **263**
- discretization
  - of a domain, **102**
- distribution, 35
  - current, **29, 33**
  - tempered, **35, 92**
- distributional derivative, *see also* weak derivative
- divergence, *see also* exterior derivative, **235**
- divergence form, **63**
  - for nonlinear equations, **140**
- domain complex, **77**
- domain of a linear operator, **76, 171**
- DuHamel's Principle, **115**
- elementary  $k$ -forms, **14**
- elliptic, **63, 141**
  - differential operator, **64**
  - nonlinear differential operator, **141, 240**
- elliptic projection, **5, 168, 201**
- energy norms, **72**
- energy-norm estimates, **72, 122**
- error estimates
  - for the elliptic problem, 176
    - extension to handle nonzero harmonic, 183
  - for variational crimes, 178
  - for the elliptic projection, 221
  - for the parabolic Hodge Laplacian problem, 219
  - general interpolation, 119
  - generalities for functions, 122
  - main parabolic estimates theorem, 205
  - relation to best approximation, 121
- Euler method, **113**
  - backward or implicit, **114**
- Euler-Lagrange equations, 97
- evolutionary differential equation, **11**
- exponential map, **284**
- extending bounded operators (standard technique), 39
- exterior derivative, **18**
  - as part of abstract Hilbert complexes, 76
  - weak, **33**
- finite differencing, **95**
- finite element, **103**
- finite element method, **3, 95, 96**
- finite element spaces
  - for domains in  $\mathbb{R}^n$ , 218
  - for Riemannian manifolds, 219
- first-order method (for ODEs), **113**
- flux, 25
- Fourier transform, **35**
- Fréchet derivative, **91**
- frames and the postmultiplication convention, **20**
- Fredholm, **86**
- function spaces, *see also* Sobolev space
  - evolutionary equation as a curve in, 87
- fundamental solution, **261**
- Galärkin method, **96, 102, 243**
- Galärkin orthogonality, 122
- Gårding's inequality, **72**
- Gâteaux derivative, **18, 141**
- Gelfand triple, *see* rigged Hilbert space
- graph inner product, **33**
- Green's First Identity, 62
- Green's function
  - Dirichlet, **259**
- Green's Representation Formula, **262**
- harmonic conjugates, **52**
- harmonic form, **42**
  - in a Hilbert complex, **78, 171**
- heat equation, **2, 88**
- Hilbert complex, **4, 76, 76–87, 170, 171**
  - bounded, **77, 171**
  - closed, **76, 171**
  - dual complex, **78, 172**
- Hodge decomposition



- in Hilbert complexes, **79, 80, 172**
- Hodge dual operator, **28**
- Hodge heat equation, **3, 164**
- $H\Omega$  spaces, 33–41
- homogeneous (polynomial) forms, **133**
- inf-sup condition, **83, 174**
- initial value problem, **11**
- inner product
  - energy, **70**
  - graph inner product in Hilbert complexes, **77, 171**
  - $\mathcal{L}^2$ , 26
- integral
  - of a top degree form, **22**
- interior product, **16**
- interpolation, **119**
- interpretation of  $\Delta$  as mapping to the dual, 44
- invariant formula for the exterior derivative, 19
- Kantorovitch's Theorem, 158–159
- Koszul differential, **132, 217**
- $\mathcal{L}^2$  inner product, *see* inner product
- Laplace-Beltrami operator, *see* Laplacian
- Laplacian
  - Abstract Hodge, **82, 173**
  - Abstract Hodge problem, **82, 173**
  - in Riemannian geometry, **67, 67**
- least squares
  - for inverting the gradient, **48**
- Lebesgue integral convention, 19
- linear interpolation, **104**
- linearization, **141**
- linearized stiffness matrix, **153, 245**
- linearized weak form, **245**
- Lipschitz mappings, **14**
- main parabolic estimates theorem, **205**
- mass matrix, **246**
- master element, **105**
- mesh, **102**
- mesh parameter, **102**
- mesh size, **102**
- mixed abstract Hodge Laplacian problem, **82, 173**
- mixed formulation, **4, 44**
- mixed weak formulation
  - in spaces of differential forms, **45**
- Moore-Penrose pseudoinverse, 183
- morphism of Hilbert complexes, **79, 172**
- Neumann condition, **11**
- Newton's Method
  - generalities, 155–157
  - globalizing, 159–161
- nondivergence form, **64**
- norm
  - energy, **70, 72**
  - graph norm in  $H\Omega$  spaces, 33
  - essential sup or  $\mathcal{L}^\infty$ , 72
  - Sobolev, 31, 57
- normal projection, **214**
- normalized conformal factor equation, **239**
- normalized Ricci flow, **147**
- normalizing, **266**
- numerical Methods for ODES
  - Symplectic methods, 116
- numerical methods for ODES
  - Runge-Kutta methods, 116
- order of convergence, **3**
- orientable, **21**
- orientation
  - of a manifold, **22**
  - of a vector space, **21**
- Paneitz operator, **265**
- parity of a form, **22**
- partial differential equation
  - evolutionary, **87**
- Petrov-Galärkin method, **101**
- physical interpretation of general elliptic operators, 68

- Picard's theorem
  - for infinite-dimensional spaces, 87
  - insufficiency for heat equation, 88
- piecewise linear continuous approximation, **104**
- Poincaré-Bergman metric, **314, 316**
- Poisson kernel, **260**
- Poisson's equation, **259**
- product rule
  - for interior products, 17
- pullback, 17
  - naturality, 19
  - of pseudoforms, 25
- quantum mechanics
  - use of rigged Hilbert space theory in, 92
- quasi-best approximation, **120**
- quasi-optimality, **176**
- quasilinear, **142**
- Rayleigh-Ritz method, **96**
- Ricci flow, **236**
- Riemann Mapping Theorem, **306**
- Riemannian metric
  - on differential forms, 16
- Riemannian volume form, **26**
- rigged Hilbert space, **91**
- right-handed, **21**
- semi-discrete Bochner parabolic problem, **200**
- semilinear, **141**
- separation of variables, **109**
- shape function, **105**
- Sobolev embedding theorem
  - Trace theorem a generalization of, 37
- Sobolev space, **31**
  - fractional order, **36**
  - $H^k$ , 57
  - $H\Omega$  spaces, 33–41
  - negative order, **36**
- Sobolev-Orlicz spaces, **143**
- sparse matrix
  - in finite element methods, **108**
- stability constant, **83, 174, 175**
- star (Hodge operator), *see* Hodge dual operator
- stiffness matrix, **100**
- Stokes' Theorem, 25
- strong form, 98
- strong Hodge decomposition, **80**
- strong solution, **43, 61**
- Sturm-Liouville Problem, **62**
- superconverge, **158**
- Thomé's error equations, 204
- time-ignorant discrete problem, **201**
- timestepping, **112**
- trace (boundary restriction), **33**
- trace (boundary restriction)
  - extended to forms, 36–40
- transversely oriented, **23**
- triangulation, **102**
- Uniformization Theorem
  - for compact surfaces, **146**
  - for domains in the plane, **306**
- uniformly elliptic, **63**
- variational crimes, **130, 170**
- vector proxy fields, **50**
- weak derivative, 31
  - evolutionary, **90**
- weak exterior derivative, **235**
- weak formulation, **3, 60**
  - for general elliptic problem, 64
  - for nonlinear problems, 142
  - for Sturm-Liouville Problem, 62
  - for the conformal factor equation, 145
- weak solution, **43, 61**
  - Poisson's equation, **59**
  - to the Abstract Hodge problem, **82, 173**
- wedge product, **14**

why sections, not individual covectors,  
pull back, 17