Data assimilation problems in glaciology
Daniel Shapero

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

University of Washington
2017

Reading Committee:
Randall J. Leveque, Chair
Ian R. Joughin
Benjamin E. Smith
Anne Greenbaum

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

Data assimilation problems in glaciology
Daniel Shapero
Chair of the Supervisory Committee:
Dr. Randall J. Leveque
Department of Applied Mathematics

Rising sea levels due to mass loss from Greenland and Antarctica threaten to inundate coastal areas the world over. For the purposes of urban planning and hazard mitigation, policy makers would like to know how much sea-level rise can be anticipated in the next century. To make these predictions, glaciologists use mathematical models of ice sheet flow, together with remotely-sensed observations of the current state of the ice sheets. The quantities that are observable over large spatial scales are the ice surface elevation and speed, and the elevation of the underlying bedrock. There are other quantities, such as the viscosity within the ice and the friction coefficient for sliding over the bed, that are just as important in dictating how fast the glacier flows, but that are not observable at large scales using current methods. These quantities can be inferred from observations by using data assimilation methods, applied to a model of glacier flow. In this dissertation, I will describe my work on data assimilation problems in glaciology. My main contributions so far have been: computing the bed stress underneath the three biggest Greenland outlet glaciers; developing additional tools for glacier modelling and data assimilation in the form of the open-source library *icepack*; and improving the statistical methodology through the user of total variation priors.

Contents

Chapter 1

Introduction

In this dissertation, I will describe my work on data assimilation problems in glaciology. The chief motivation for studying this subject is for making predictions of sea-level rise in the coming century (Church et al., 2013). Changes in Earth's climate due to rising levels of atmospheric carbon dioxide may in turn affect the mechanical and thermodynamic state of Earth's glaciers and ice sheets, which could then melt into the oceans. The largest potential contributors to sea-level rise are Earth's two large ice sheets, Greenland and Antarctica. Greenland contains enough ice to raise global sea levels by 7 m, and Antarctica by 58 m (Vaughan et al., 2013). Fully melting either ice sheet in even the next several thousand years is highly unlikely, so the question becomes how much melting will occur. This depends on various human factors, for example, how successful efforts are to curb the use of CO_2-producing fossil fuels. The results of such efforts, depending as they do on human behavior, are hard to predict. For the purposes of informing policy makers, the best one can do is to consider a range of scenarios – complete cessation of fossil fuel use on one end, and business-as-usual increase in fossil fuel use on the other.

While Greenland is the smaller of the two ice sheets, it is at a lower latitude and thus experiences more surface melting. Antarctica, on the other hand, experiences very little surface melting at all, having a yearly average surface temperature of -60° C. It was once thought that, under a warming climate, Antarctica would grow because the hydrological cycle would accelerate, bringing more snow accumulation to the ice sheet. Losses from Greenland could still offset this growth in Antarctica, resulting in a net sea-level rise. This type of projection only takes into account the thermodynamics of ice sheets, and not their mechanics, i.e. how fast the ice is flowing into the ocean. In the 1990s and early 2000s, substantial acceleration in the flow of glaciers along the Amundsen Sea Embayment and in the Antarctic Peninsula bore witness to the necessity of studying glacier flow in order to understand the total volume loss from Earth's ice sheets. As a result, the Intergovernmental Panel on Climate Change (IPCC) has highlighted the dynamics of ice sheets and glaciers as one of the chief sources of uncertainty in predictions of sea-level rise (Church et al., 2013). One of the more alarming ramifications of glacier dynamics is the possibility that many areas of Earth's ice sheets are unstable. Due to a mechanism

known as *marine ice sheet instability*, an initial retreat of glaciers resting on reverse bed slopes may trigger even more rapid retreat.

In order to predict sea levels in the coming century and beyond, glaciologists often use numerical simulations of ice-sheet flow. Given measurements of the current state of the ice sheet and estimates of climate forcing, its future state can be simulated assuming that ice flow can be described using continuum mechanics and thermal physics. The simulated final state can then be taken as representative of a possible future state of the ice sheet (Seddik et al., 2012). By running several simulations with different climate forcing, one can establish a range of potential scenarios for how much mass the ice sheets may lose or gain in the next century or beyond. These simulations have been run with varying degrees of success and form the basis for the aggregated projections of the IPCC, one of which is shown in figure 1.1.

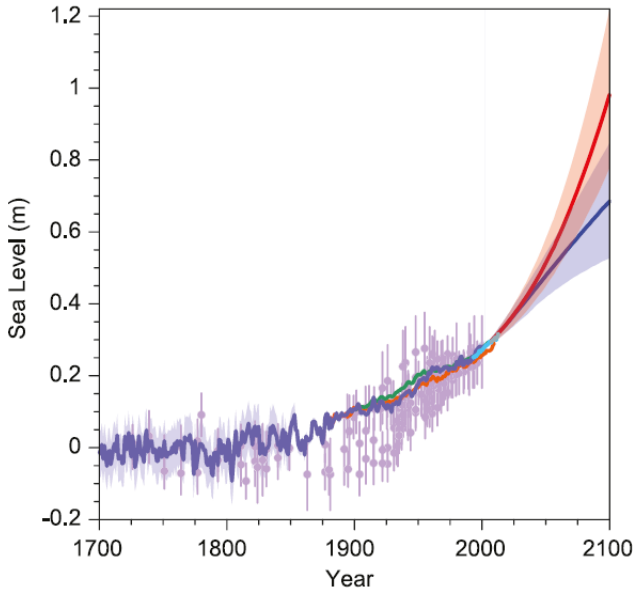


Figure 1.1: Projections of sea level in the 21st century relative to pre-industrial level. The blue curve represents a scenario where atmospheric CO_2 is curbed, the red curve where it is not. Reproduced from Church et al. (2013).

In order to run such a simulation in the first place, one needs to know the initial ice sheet state. As it turns out, we need to know more to initialize an ice sheet forecast than we can observe easily using common remote sensing techniques.

One way around this problem is to do "spin-up" experiments. A model spin-up consists of picking some reasonable initial conditions, propagating the model forward in time for several thousand years until the ice sheet is roughly in steady state, and taking this state as the start of the simulation (Martin et al., 2011). There

are several problems with this approach. First, there is no guarantee that the ice sheet state obtained at the end of a model spin-up is, in any way, representative of the current state. Second, the climate forcing used to spin up the model might not accurately capture the true climate forcing experienced by the real ice sheets. The introduction of a modern climate to an ice sheet spin up using some approximation of past climate may result in unphysical transients at the beginning of the true simulation. Finally, the computational cost of the spin-up may exceed the cost of the simulation we wish to do in the first place; several millenia of spin-up may be necessary for only a centurial-scale projection.

Instead, we can borrow a page from the meteorologists' book and try to leverage existing measurements to the greatest extent possible. While some fields may not be directly observable at large scales, presumably these quantities have some effect on the fields which are measurable. For example, we cannot directly measure the friction coefficient that dictates how much resistance a glacier encounters as it slips over the bed underneath it. Nonetheless, a glacier flowing over a very resistive bed will likely flow slower than a glacier flowing over a slippery bed, all other factors being equal. We can then ask which spatial distribution of bed friction is most consistent with the velocities we did observe at the ice surface. This idea is the essence of *inverse problems* or *data assimilation*.

1.1 Outline of the dissertation

In order to use data assimilation methods to infer the complete present state of the ice sheet, we must use the physics of how the various fields we would like to estimate relate to each other. For example, the velocities of ice flowing over crystalline bedrock, which has a very high friction coefficient, would tend to be lower than that of ice flowing over water-saturated sediment. Similarly, the mechanical hardness of warm ice is lower than that of cold ice, all factors being held equal, so the strain rates experienced by warm ice would tend to be higher. In chapter §2, I will give an overview of the physics of glacier flow. The velocity of a glacier can be modelled as the solution of an elliptic system of partial differential equations, the Stokes equations, which describe slow viscous fluid flow. Modelling glacier flow is complicated by the fact that ice is not a typical material, for which the viscosity is more or less a constant. The viscosity is a function of the strain rate, which makes the governing PDE system nonlinear. Most real glacier flows are predominantly two-dimensional, while the Stokes equations are posed in three space dimensions. In this chapter, I will also describe approximation schemes for simplifying the Stokes equations to a 2D system.

Knowing the current ice geometry, velocity, temperature, and boundary conditions, we could in principle make predictions of future ice flow. We can measure the ice geometry and velocity remotely from planes and satellites, but not so for the temperature and boundary conditions. In chapter §3, I will describe how we can use our knowledge of the physics, together with the available observations of ice geometry and velocity, to attempt to infer unobservable quantities such as temperature and bed friction. The main mathematical tool necessary for this subject is

PDE-constrained optimization, which I describe in some detail. These mathematical preliminaries are followed by a review of data assimilation problems in glaciology specifically. Finally, I will describe the work from my first paper (Shapero et al., 2016), on applying these methods to the three largest outlet glaciers in Greenland.

The partial differential equations describing the flow of glaciers are inhomogeneous and nonlinear, so they do not have exact solutions. Our only recourse in this case is to use numerical methods. While there are several existing software packages for modelling glacier flow, many of them are difficult to use by recent standards, having been initially written at a time when software engineering as a whole was much less advanced. I began developing a glacier modelling library called icepack in 2015, which I will describe in chapter 4. In particular, I will describe some of the software engineering choices I made and how I think they will make icepack easy to use for practitioners, and how I have tried to test and validate the code for fidelity to the underlying model physics. All the work in subsequent chapters uses icepack.

One of the conclusions of my first paper on Greenland outlet glaciers was to disagree with some of the findings in Sergienko et al. (2014) about the basal shear stress under Jakobshavn Isbrae. I attributed the difference between my work and theirs to issues relating to the methodology of how the observational data are used to constrain the inferred fields. Since we have only finitely many measurements, and noisy ones at that, the inferred field can only be recovered up to a certain degree of detail. Consequently, we must decide in advance what kinds of spatial variability we consider to be unlikely in the true field, and how much to penalize such variability. This procedure can be described using the language of Bayesian statistics, in which setting a resolution limit can be thought of as selecting a *prior distribution* on the inferred field. In chapter §5, I will review Bayesian inference and some of the considerations that go into selecting a prior distribution. In particular, I argue that the most common prior used in the literature is not the best choice for inferring the rheology of a floating ice shelf. Instead, I argue that a prior constrained by the *total variation* of the inferred field is the more physically appropriate prior distribution. The difference between a smoothness penalty and a total variation penalty is illustrated on a synthetic model problem.

Finally, in chapter §6 I apply these methods to inferring the rheology of the Ross Ice Shelf. My results obtained with icepack agree in broad strokes with previous work on the Ross Ice Shelf (Rommelaere and Macayeal, 1997). By using a total variation prior, however, I was able to recover the sharp transitions in rheological hardness that occur across rifts or shear margins in the ice shelf. These interfaces are not discernible with the most commonly-used prior distribution. This is ongoing work, with application to the Larsen C Ice Shelf in progress.

1.2 Quantities of interest

The main quantities of interest when studying glacier flow are the ice bed elevation, surface elevation, velocity, temperature, water content, impurity content, crystal fabric, and bed friction. Some of these quantities can be easily observed at large scales from planes and satellites. These are the ice surface elevation and surface

velocity, and, to a somewhat lesser extent, the ice bed elevation. The other quantities can only be observed at individual locations through drilling boreholes into the ice. Borehole drilling is expensive and often logistically challenging; it cannot be deployed over wide regions of the ice sheet.

Elevation

The ice surface elevation, bed elevation, and surface slope combine to determine the total stress driving the ice due to pressure imbalances between nearby ice columns. The fast-flowing glaciers that drain the ice sheets often coincide with areas where the bedrock is highly channelized. On short time scales, one can think of the topography determining where ice can flow fast, but on longer time scales, ice flow feeds back on the topography through erosion.

The elevation of an ice sheet can be measured using satellite altimetry (Howat et al., 2014). Altimetry works by measuring the time it takes for electromagnetic waves to reach the earth and return to the satellite; the surface elevation can then be obtained, knowing the satellite's orbit. Altimeters can function, in principle, at any wavelength not absorbed by earth's atmosphere; for example, the CryoSat and European Remote Sensing satellites used radar altimeters, while ICESat and its successor ICESat-2 use green and red lasers.

The bed elevation is typically measured with airborne radar, but there are many more complications than for measuring surface elevation (Gogineni et al., 2014). In very deep bed channels, such as the bed underneath Jakobshavn Isbrae in Greenland, there is much more ambiguity in determining how much of the radar echo comes from the glacier bed and how much from the side walls of the fjord. Additionally, water is a much brighter reflector of radio waves than rock, so a layer of englacial meltwater can be mistaken for the ice bed. Finally, airborne remote sensing over remote regions of Greenland and Antarctica is much more time-consuming and logistically difficult than satellite remote sensing. Until recently, very large regions of Antarctica were almost completely unmapped.

Velocity

The velocity of a glacier together with its thickness determines the total flux of mass out of the system. Additionally, the strain rate (the symmetrized gradient of the velocity field) in part determines how likely a glacier is to calve off icebergs into the ocean.

The earliest measurements of ice velocity (by no less than John Muir) used the movement of stakes planted into the glacier to record displacement. This technique is still used to validate remote sensing measurements, but it does not give wide coverage of ice velocity.

Instead, ice sheet velocity can be measured remotely using optical or radar instruments. The earliest method for remote sensing of ice velocity used surface features, such as crevasses or ice rumples, to measure displacement between pairs of optical images (Fahnestock et al., 2015). This technique depends on having several clear images, but cloud cover in the winter often obscures the ice surface to visible

instruments. Additionally, there must be many identifiable features on the ice surface to track, which is not always the case in the deep interior of the ice sheet. The launch of the Landsat-8 satellite in 2013 dramatically improved this situation; the sensor on Landsat-8 has a much higher radiometric resolution than its predecessors, allowing much smaller features to be tracked than before.

Using radio frequencies is advantageous in some respects because they can penetrate through cloud cover, so velocity measurements can be made even in winter (Joughin, 2002). In fast-flowing areas of the ice sheet, radar speckle noise from scatterers just under the ice surface can be coherent between images. This radar speckle can function as a trackable feature, much like in optical repeat-image feature tracking. An alternative not available for optical instruments is to use interferometry. If the radar platform is flown over the same point twice, differences in the phase of the returned electromagnetic wave can be used to infer the displacement of the ice surface toward the satellite. When supplemented with a map of the ice surface elevation, the displacements can then be used to determine the ice velocity. Interferometric synthetic aperture radar (InSAR) together with speckle-tracking has been used to map the velocity of the entire Greenland and Antarctic Ice Sheets.

While optical and radar remote sensing methods can be used to measure the ice surface velocity, the velocity within the ice column is not directly measurable other than through boreholes.

The following quantities can at present only be measured from boreholes.

Temperature

The temperature of a glacier determines its resistance to deformation. The important determining factors are the surface temperature, the geothermal heat flux, and internal heating from strain.

Bed friction

Ice can slide more readily over water-saturated sand and mud than it can over hard crystalline bedrock. On longer time scales, the ice can transport sediment and erode bedrock.

Meltwater either from surface runoff or generated directly at the bed from strain heating can lubricate the sliding of a glacier over its bed. In extreme cases, the meltwater is at such high pressure that it lifts the ice off the bed completely, a process known as “hydraulic jacking.”

Fabric

An individual ice crystal is not mechanically isotropic, i.e. deformation is easier along on axis (the c-axis) than in the other directions. Glacier ice is usually assumed to be polycrystalline – consisting of many randomly-oriented crystals, so that the response to an applied stress is effectively isotropic. However, at the base of a thick

ice sheet, sustained deformation can preferentially align the crystal c-axes so that the ice exhibits anisotropy at large scales.

Impurities

During the last ice age, earth's atmosphere was much dustier. This dust was deposited on the surface of the Greenland and Antarctic ice sheets and incorporated into the ice as more snow fell. This higher impurity content is partly responsible for making ice from the last ice age deform much more under the same applied stress than ice accumulated during the curent epoch.

Chapter 2

Glacier physics

In this chapter, I will describe the physical processes important for glacier flow. To a first approximation, glacier ice deforms like a viscous fluid, flowing downhill under its own weight. The local rates of snowfall and surface or basal melting dictate where material accumulates, and the ice then flows toward a steady-state geometry. The ice sheet can reach thicknesses upwards of 3 km in Greenland and 4 km in Antarctica at the highest ice divides. These regions provide the best locations for ice core drilling, since the slow flow ensures that the ice column undergoes little deformation that would disturb the stratigraphy. The surface elevation gradient induces a pressure difference within the ice, forcing it to flow downhill under its own weight. The bed topography contributes to channelling the flow of ice into separate drainage basins and, in turn, into the ocean through often narrow outlets. In Greenland, most outlet glacier flow is through highly channelized, incised fjords, carved into the bedrock through millennia of erosion. These dramatic features are not as prevalent in Antarctica, where the ice flow nonetheless channelizes into the less numerous but broader fast-flowing ice streams.

Terrestrial glaciers flow at speeds that span several orders of magnitude. In the deep interior of the Greenland and Antarctic ice sheets, flow speeds are on the order of 10-30 cm/year. The pressure differential between ice at higher elevations and at lower elevations causes the ice to flow downhill and accelerate. The total pressure gradient due to the surface elevation gradient ∇s and ice thickness h, integrated through the entire ice column, is called the *driving stress*:

$$\tau = -\rho g h \nabla s. \tag{2.1}$$

This is the chief forcing term for ice sheet flow. In the steepest regions of the Greenland Ice Sheet, around the glacier Jakobshavn Isbrae, the driving stress can reach 250 kPa (Truffer and Echelmeyer, 2003). By contrast, the ice streams feeding the Ross Ice Shelf can have driving stresses as low as 20 kPa (Truffer and Echelmeyer, 2003). The downhill flow is channelled through a select few small but fast *outlet glaciers*, where speeds reach 1-10 km/year. The high points on the ice sheet, where the flow speed is almost exactly zero, are called *ice divides*. The distance from the ice divide to the ocean is roughly on the order of 100-1000 km. We can then make

the rough approximation that the average strain rate is the change in speed divided by the distance, i.e. $\dot{\varepsilon} \approx 10\frac{\text{km}}{\text{a}}/1000\text{km} = 10^{-2}\text{a}^{-1}$.

While ice flow is viscous, the viscosity is not a material constant like most Newtonian fluids. Instead, the ice viscosity decreases as a power of the second invariant of the strain rate tensor, i.e. ice is a shear-thinning substance. This constitutive law makes the governing diagnostic equations for the ice velocity nonlinear.

A glacier that is in contact with ocean water is called a *tidewater glacier*, in contrast to a *land-terminating* glacier. The fastest-flowing glaciers in Greenland and Antarctica generally discharge ice directly to the ocean. The base these glaciers can either be resting on the underlying bedrock, or it can be floating. The transition between the grounded and floating parts of the glacier is called the *grounding line*. The floating parts of a glacier are called *ice shelves*. At the base of an ice shelf, the pressure from the overlying ice and the pressure from the surrounding ocean water must be equal. Otherwise, the ice would not be in buoyant equilibrium and would either sink deeper or float higher until equilibrium is achieved. Suppose that the ice thickness is h, and the elevation below sea level of the ice base is d. The densities of glacier ice and sea water are, respectively, $\rho_I = 917$ and $\rho_W = 1024\frac{\text{kg}}{\text{m}^3}$. Letting $g = 9.81\frac{\text{m}}{\text{s}^2}$ be the acceleration due to gravity, the pressure due to the weight of the ice column at the base is $P_I = \rho_I g h$. Likewise, the pressure of the surrounding sea water is $P_W = \rho_W g d$. Setting the two to be equal, the depth of the ice base below sea level is

$$d = \frac{\rho_I}{\rho_W} h. \tag{2.2}$$

Since $\rho_I / rho_W \approx 0.9$, roughly 90% of the ice shelf rests below sea level. One of the stark differences between Earth's ice sheets is that Antarctica possesses massive floating ice shelves, chief among which are the Filchner-Ronne and Ross ice shelves. The Ross and Filchner-Ronne ice shelves are each roughly the size of Texas. Ice shelves are thought to play an important role in regulating the speed of Antarctic outlet glaciers by *buttressing* the flow of its tributary outlet glaciers (Dupont and Alley, 2005). Greenland outlet glaciers, on the other hand, are mostly grounded at their termini.

Fast-flowing outlet glaciers typically have a roughly constant velocity profile with depth. In this plug-flow regime, the flux of ice through the grounding line is an increasing and highly nonlinear function of the ice thickness at the grounding line (Schoof, 2007). This nonlinearity gives rise to a potential instability mechanism for glaciers which rest on beds that slope downward toward the center of the ice sheet. Thwaites Glacier in Antarctica may already be experiencing a runaway retreat due to this *marine ice sheet instability* (Joughin et al., 2014b).

These topics should give some idea of the kinds of questions we would like to address in studying the physics of glacier flow. In this chapter, I will review all of the aspects that I consider relevant for the work in the next chapter on data assimilation. First and foremost is the relevant continuum mechanics of how glaciers achieve stress balance, but there are many other effects that also need to be accounted for. The ice constitutive relation is nonlinear, and the proportionality constant is called the *rheology coefficient*. The rheology coefficient is mainly determined by temperature, but the amount of water in the glacier that percolates down from

surface melt ponds, the impurity content in the ice from dust settling out of the atmosphere, and the ice crystal fabric all factor in to the rheology as well. At scales on the order of tens to hundreds of meters – far larger than an individual ice crystal but still smaller than a typical grid cell in a numerical simulation – ice can behave as a brittle material. These breaks or *crevasses* in the ice base and surface mechanically decouple the ice on either side, and at the bulk scale act to weaken the glacier. These crevasses can then expand under increasing longitudinal extension of the ice or under pressure from trapped meltwater, until they penetrate all the way through the glacier. This hydrofracture process can, in some cases, initiate calving of large icebergs into the ocean. The conditions at the boundary of a glacier also exert a significant control on the flow. The beds and side walls surrounding a glacier exert resistive stresses on the ice. The resistive stress depends on what material the ice is sliding past; deformation over wet sand or till is easier than over hard granite. Moreover, depending on the constitution of the basal material, the resistive stress may be a linear or nonlinear function of the ice sliding speed. Clark and Pollard (1998) suggest that the basal material constitution played a critical role in changing the flow regime of the Laurentide Ice Sheet one million years ago, in such a way that the periodicity of ice ages was irrevocably altered thereafter. In many cases, the ice at the glacier bed is at the pressure melting point. High subglacial water pressures can act to lubricate the sliding of the glacier over its bed. Meltwater can also make porous materials, such as glacial till, more porous and thus more deformable. The importance of subglacial hydrology can be seen in the stagnation of Ice Stream C 150 years ago (Anandakrishnan and Alley, 1997).

2.1 Stokes equations

The most general description of viscous fluid flow is the Navier-Stokes equations. Let ρ be the ice density, u_i the ice velocity, p the ice pressure and τ_{ij} the stress tensor. The full Navier-Stokes equations are

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x_j} u_j\right) \rho u_i = \frac{\partial}{\partial x_j} \tau_{ij} - \frac{\partial p}{\partial x_i} + \rho g_i \tag{2.3}$$

where g_i are the components of the gravitational acceleration vector in the chosen coordinate system. One can pick out which terms are important by examining the relevant length scales for the problem (Greve and Blatter, 2009). The pressure at the base of a 1 km-thick ice stream is

$$P = \rho g h \approx 1 \text{ MPa}. \tag{2.4}$$

A typical length scale for a narrow Greenland outlet glacier is $L \approx 10$ km. Meanwhile, for an ice velocity scale U on the order of 1 km/year and a time scale T of 10 years, the ratio of acceleration scale to pressure gradient scale is

$$\frac{\rho U L}{P T} \approx 10^{-15}. \tag{2.5}$$

This quantity is the *Froude number*; since it is so much smaller than 1 for this problem, dropping the acceleration terms from the left-hand side of (2.3) is a reasonable approximation. The remaining system of PDEs

$$0 = \frac{\partial}{\partial x_j}\tau_{ij} - \frac{\partial p}{\partial x_i} + \rho g_i \tag{2.6}$$

are called the Stokes equations, and they describe the net stress balance of the glacier. The approximation of using the Stokes equations instead of the full Navier-Stokes equations amounts to the fact that roughly all of the internal and external forces in the material are in balance.

These equations are supplemented with some equation of mass conservation. For simplicity's sake, we can ignore the densification of snow and firn in the upper few tens of meters of ice. In this case, the ice flow is divergence-free:

$$\frac{\partial u_i}{\partial x_i} = 0. \tag{2.7}$$

The stress and mass conservation equations (2.6), (2.7) are not a closed system of PDEs for the ice velocity. We need to describe how the stress tensor τ_{ij} is related to the ice velocity in some way. The *strain rate tensor* for the velocity field u_i is the rank-2 tensor

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{2.8}$$

which has units of inverse time. The strain rate tensor is symmetric and trace-free, since the velocity field is divergence-free. The strain rate tensor describes infinitesimal deformations of a material parcel along the coordinate axes, ignoring pure translations and solid-body rotations (Gonzalez and Stuart, 2008). For a typical Newtonian fluid, the stress tensor is linearly proportional to the strain rate tensor, and the proportionality constant is the material viscosity:

$$\tau_{ij} = 2\mu\dot{\varepsilon}_{ij}. \tag{2.9}$$

Ice, however, is a non-Newtonian substance; the viscosity depends on the strain rate.

The key experiments on the deformation of polycrystalline ice as a function of applied stress and temperature were performed by Glen (1955). Polycrystalline ice has several deformation regimes – primary, secondary and tertiary creep. These are distinguished by the total strain ε that the material experiences, as opposed to the strain rate. Secondary creep beings at strains of around 1%, while tertiary creep begins at strains on the order of 10-15% (Cuffey and Paterson, 2010). The most relevant regime for glacier flow is tertiary creep. In Glen's experiments, a block of ice was placed under uniform compression along the vertical axis while being confined along the y-axis. The stresses in Glen's experiments ranged from 100-1000 kPa, which is on the high end of the stresses that can be expected in real glaciers. The extension rate $\dot{\varepsilon}$ of the ice block along the x-axis is measured for several values of the compressive stress τ. By taking logarithms, a power-law

$$\dot{\varepsilon} = A\tau^n \tag{2.10}$$

can be fit to the data for the prefactor A and the exponent n. Glen found a best fit of $n = 3.17 \pm 0.2$. Most studies and textbooks on glacier modelling use $n = 3$; this relation is referred to as *Glen's flow law*. Laboratory experiments such as Glen's, however, do not achieve total strain values on the order of the 10% that typify tertiary creep, so these experiments are not necessarily representative of real glacier flows (Cuffey and Paterson, 2010).

To evaluate whether $n = 3$ is a good description, we would need to compare exact solutions of the Stokes equations with real field data. The relation (2.10) applies to uniaxial loads, where only one stress or strain component is non-zero. For a general description of glacier flow, we need a relation between the stress and strain rate tensors. However, we can describe the flow of an idealized glacier using only the scalar relation (2.10). Consider the flow of a glacier of uniform thickness h down an inclined slab of angle α. We will choose the x-axis as parallel to the slab and the z-axis as perpendicular to it, in which case the components of the gravitational acceleration vector along the x- and z-axes are $\rho g \sin\alpha$ and $-\rho g \cos\alpha$ respectively. If the length of the slab is sufficiently long, we can approximate it as infinite and take the flow as being translation-invariant along in the x-axis. The horizontal velocity u is the only flow component, and it varies only in the z-direction. The Stokes equations then simplify to

$$\frac{\partial}{\partial z}\tau_{xz} + \rho g \sin\alpha = 0 \tag{2.11}$$

since the pressure is independent of x. By integrating this expression from an arbitrary height z to the ice thickness h and using the fact that the shear stress at the ice surface is 0, we find that

$$-\tau_{xz} + \rho g (h - z) \sin\alpha = 0. \tag{2.12}$$

We can now use the Glen flow law (2.10) to write

$$\dot{\varepsilon}_{xz} = A\tau_{xz}^n = A(\rho g (h - z) \sin\alpha)^n . \tag{2.13}$$

We can then integrate this expression from the ice base to some elevation z to obtain the full vertical velocity profile:

$$u(z) = u(0) + \frac{2A}{n+1}(\rho g \sin\alpha)^n \left(h^{n+1} - (h - z)^{n+1}\right). \tag{2.14}$$

For a Newtonian fluid, $n = 1$ and the profile is a parabola; for higher values of n, i.e. more plastic flow, the profile becomes a higher-degree polynomial, with more deformation near the ice base than in the entire column. Although this situation is highly idealized, Raymond (1980) was able to compare it with real velocity profiles on valley glaciers to show that $n = 3$ was an acceptable fit to actual field data and that $n = 1$ was not.

For complex stress regimes, we need a relation

$$\dot{\varepsilon} = f(\tau) \tag{2.15}$$

between the stress and strain rate tensors that reduces to (2.10) in the uniaxial case. Such a function f is called the *constitutive relation* for the material. Provided that glacier ice is anisotropic, any such relation should be invariant to the choice of coordinate system. In other words, if Q is some 3×3 orthogonal matrix, then

$$f(Q^* \tau Q) = Q^* f(\tau) Q. \tag{2.16}$$

The condition that the constitutive relation is isotropic is a dramatic restriction on the form that it can take. For any rank-2 tensor r_{ij}, there are three quantities, called the *tensor invariants*, which do not change under a rotation of the underlying coordinate system:

$$r_I = \text{trace}(r)/3, \tag{2.17}$$
$$r_{II} = \text{trace}(r^* r)/2, \tag{2.18}$$
$$r_{III} = \det(r). \tag{2.19}$$

Since we are assuming glacier ice to be incompressible, the trace of the strain rate tensor is 0. Likewise, the trace of the deviatoric stress tensor is 0 by definition. The most general constitutive relation possible, consistent with both isotropy and incompressibility, is

$$\dot{\varepsilon}_{ij} = C_1(\tau_{II}, \tau_{III})\tau_{ij} + C_2(\tau_{II}, \tau_{III})\left(\tau_{ik}\tau_{kj} - \frac{2}{3}\tau_{II}\delta_{ij}\right) \tag{2.20}$$

but experiments find that $C_2 = 0$ and C_1 is independent of τ_{III} (Cuffey and Paterson, 2010). If C_2 were nonzero, the ice could experience nontrivial longitudinal stretching even under a pure shear stress, which is not observed either in the laboratory or in the field. In order to match the uniaxial case, we are then left with

$$\dot{\varepsilon}_{ij} = A|\tau|^2 \tau_{ij} \tag{2.21}$$

where we introduce here the shorthand that $|\tau| = \sqrt{\tau_{II}/2}$. For the purposes of closing the Stokes equations, however, it is better to have a relation describing the stress in terms of the strain rate than the other way around. To invert the relation (2.21), we can take the second invariant of both sides:

$$|\dot{\varepsilon}| = A|\tau|^3 \tag{2.22}$$

from which we get $|\tau| = (|\dot{\varepsilon}|/A)^{1/3}$. By defining the *rheology coefficient* $B = A^{-1/3}$, we then find that

$$\tau_{ij} = \frac{1}{A|\tau|^2}\dot{\varepsilon}_{ij} = B|\dot{\varepsilon}|^{-2/3}\dot{\varepsilon}_{ij}. \tag{2.23}$$

Comparison with the usual Newtonian constitutive relation $\tau = 2\mu\dot{\varepsilon}$ shows that the viscosity of ice is

$$\mu = \frac{B}{2}|\dot{\varepsilon}|^{-2/3}. \tag{2.24}$$

The negative power of $|\dot{\varepsilon}|$ shows that ice is a *shear-thinning* material, i.e. the viscosity decreases as the material is put under higher strain rates.

The flow rate parameter A depends on several factors, chief among which is the temperature; warmer ice is easier to deform. Dust and impurity content, often in conjunction with anisotropic crystal fabrics, can enhance the ice flow due to processes acting at small scales. At larger scales, damage to the ice from crevasses and rifts can be approximated as a bulk enhancement to the flow rate parameter. The various physical effects determining the rheology will be covered in §5.1.

For a partial differential equation to be well-posed, it must also be supplemented with a set of boundary conditions. The ice surface undergoes stresses from winds. While these stresses can redistribute snow on the ice surface and affect the local mass balance, they are negligible compared to the overall gravitational driving stress. Consequently, at the surface we impose a homogeneous Neumann boundary condition:

$$\tau_{ij} n_j - p n_i = -p_0 n_i \tag{2.25}$$

where n_i are the components of the unit outward normal vector to the surface and p_0 is the ambient atmospheric pressure. The stress that the ocean exerts on the underside of a floating ice shelf or the submerged part of the front of a tidewater glacier is also negligible. The Neumann boundary condition (2.25) is applicable under ice shelves as well, but p_0 is instead the ambient pressure of the surrounding ocean. The pressure imbalance at the terminus of a tidewater glacier between the base and the surface is a key factor in how icebergs calve off from the glacier front.

At the base of a grounded glacier, the ice cannot penetrate its bed, so the velocity must be perpendicular to the outward normal vector:

$$u_i n_i = 0. \tag{2.26}$$

Along the flow direction, the bed exerts a shear stress on the ice, the magnitude of which is a function of the sliding speed. The relation between basal shear stress and sliding speed is, in general, a power law:

$$\tau_{ij} n_j - p n_i = -C|u|^{\frac{1}{m}-1} u_i - p_0 n_i \tag{2.27}$$

where m is the *sliding rheology exponent* and C is the *friction coefficient*. For grounded ice, the pressure p_0 exerted by the external medium must be equal to the ice overburden pressure $\rho g h$. The last two equations constitute a mixed Dirichlet-Robin boundary condition.

The sliding rheology exponent and the friction coefficient are determined by two physical effects. First is the geological makeup of the underlying bedrock; ice slides more readily over deformable, granular materials such as clay or till than over hard granite bedrock. The second is the hydrological state of the bed. Basal meltwater at sufficient pressure can lift up the ice base, causing it to lose contact with the bed. Sedimentary beds are also a porous medium; meltwater infiltration at high pressure can increase the till porosity, making the till easier to deform (Cuffey and Paterson, 2010). In principle, the rheological exponent could be anything between $m = 1$, the linear viscous case, and $m = \infty$, the perfectly plastic case.

The solution u_i, p of the Stokes equations can be thought of as the extremum of a certain nonlinear functional E, called the *action* (Dukowicz et al., 2010). Action

principles are useful both in proving the existence and uniqueness of a solution and in obtaining a numerical approximation of the solution. For a general power-law fluid, the solution u_i of the Stokes equations is the minimizer of the functional

$$E(u) = \int_\Omega \left\{ \frac{2nB}{n+1} |\dot{\varepsilon}(u)|^{1+\frac{1}{n}} - \rho g_i u_i \right\} \mathrm{d}x \tag{2.28}$$

over all fields u_i such that $\partial u_i / \partial x_i = 0$. The function space over which this optimization problem is posed is a subspace of the Sobolev space $W_1^{1+1/n}(\Omega)$. Instead of imposing the constraint explicitly, we can introduce a Lagrange multiplier p and rephrase the problem as one of extremizing the augmented functional

$$E(u,p) = \int_\Omega \left\{ \frac{2nB}{n+1} |\dot{\varepsilon}(u)|^{1+\frac{1}{n}} - p \frac{\partial u_i}{\partial x_i} - \rho g_i u_i \right\} \mathrm{d}x. \tag{2.29}$$

The value of the Lagrange multiplier p at the extremum can be identified with the cryostatic pressure. In other words, the pressure is a Lagrange multiplier used to enforce the incompressibility of the velocity field.

2.2 Mass balance

The Stokes equations give a description of the ice velocity in terms of the current bed and surface elevation, as well as other parameters such as temperature or subglacial water pressure. They do not, however, tell us how the glacier changes in time, i.e. they are purely *diagnostic* equations. To derive a *prognostic* equation for the glacier geometry, we can think of the ice surface or bed as a manifold embedded in 3D space and derive an appropriate PDE for its evolution (Greve and Blatter, 2009). We start with an implicit definition of the glacier surface as the zero-level set of the function

$$F(x,y,z,t) = s(x,y,t) - z = 0. \tag{2.30}$$

The gradient of the level-set function F at the ice surface is the unit-outward normal vector n_i to the free surface, or

$$n_i = \frac{\partial F}{\partial x_i}. \tag{2.31}$$

So far, we have used u_i to denote the ice velocity, but the free surface may well move at a different velocity v_i. The difference between the velocity of the free surface and the ice velocity along the unit normal vector n_i is exactly equal to the rate $\dot{s}$ of accumulation or ablation of ice atop the surface, or

$$(v_i - u_i) n_i = \dot{s}. \tag{2.32}$$

By definition, the derivative of F does not change following the free surface:

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + v_i \frac{\partial F}{\partial x_i} = 0. \tag{2.33}$$

Using equations (2.31) and (2.32) in the last equation, we find that

$$\frac{\partial F}{\partial t} + u_i \frac{\partial F}{\partial x_i} = \dot{s}. \tag{2.34}$$

Since the surface elevation is a function only of the horizontal coordinates, we will commit a gross abuse of notation and write u_i $(i = 1, 2)$ for just the horizontal components of the ice velocity, letting w be a separate scalar field. We can then use the implicit definition (2.30) to get everything in terms of the surface elevation:

$$\frac{\partial s}{\partial t} + u_i \frac{\partial s}{\partial x_i} - w = \dot{s}. \tag{2.35}$$

In a similar fashion, we can derive an evolution equation for the height z_b of the ice base:

$$\frac{\partial z_b}{\partial t} + u_i \frac{\partial z_b}{\partial x_i} - w = \dot{b}, \tag{2.36}$$

where $\dot{b}$ denotes the rate of accumulation or ablation at the ice base. The ice base can either be over water, e.g. for a floating ice shelf, or resting on a solid bed such as till or bedrock. In the second case, the ice base cannot penetrate its substrate, in which case we are left with the condition that $w = \dot{b}$. In the glacier flow modelling library Elmer/Ice, the ice geometry evolution is modelled using equations (2.35) and (2.36), along with solving the contact problem to guarantee that the ice base z_b is greater than the bed elevation b (Gagliardini et al., 2013).

A simpler evolution equation for the ice geometry can be obtained by exploiting the 3D continuity equation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \tag{2.37}$$

Let $\bar{u}_i$ be the vertically-averaged horizontal components of the ice velocity. Using the Leibniz rule for differentiating under the integral sign,

$$\int_{z_b}^{s} \frac{\partial u}{\partial x} \, \mathrm{d}z = \frac{\partial}{\partial x} h\bar{u} - u \frac{\partial s}{\partial x} + u \frac{\partial z_b}{\partial x}, \tag{2.38}$$

and likewise for v and y. By integrating the continuity equation from the ice bed b to the ice surface h, we find that

$$\frac{\partial}{\partial x_i} h\bar{u}_i - u_i(z = s) \frac{\partial s}{\partial x_i} + u_i(z = b) \frac{\partial z_b}{\partial x_i} + w(z = s) - w(z = z_b) = 0. \tag{2.39}$$

If we add (2.35) and subtract (2.36) to the last equation, we are left with

$$\frac{\partial s}{\partial t} - \frac{\partial z_b}{\partial t} + \frac{\partial}{\partial x} h\bar{u}_i = \dot{s} - \dot{b}. \tag{2.40}$$

The difference $s - z_b$ is equal to the ice thickness h. Rather than distinguish between different sources of ice accumulation or ablation at the surface and base, we can

instead rewrite the net balance $\dot{s} - \dot{b}$ as the difference of the total accumulation $\dot{a}$ and the total melting $\dot{m}$. With these two changes we arrive at the the thickness evolution equation

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x_i} h\bar{u}_i = \dot{a} - \dot{m}. \tag{2.41}$$

Ignoring for now the dependence of the velocity on the thickness, this is a conservation law for thickness, where the sources are accumulation or ablation of ice at either the surface or bed.

2.3 Approximations

The nonlinear Stokes equations, together with the Glen flaw, are a closed system of PDEs for the ice velocity. Nonetheless, most glacier flow is predominantly two-dimensional. The thickness of a typical glacier is on the order of 1 km, while the typical length scales are often on the order of 10 ice thicknesses are more. The aspect ratio is $\delta = H/L$ is small enough that we can use a perturbative expansion of the Stokes equations in this parameter. The lowest-order term in this asymptotic series is much simpler than the full Stokes equations, while still providing an adequate description of the physics in many situations.

Following Greve and Blatter (2009), consider the z-component of the Stokes equations:

$$\frac{\partial}{\partial x}\tau_{zx} + \frac{\partial}{\partial y}\tau_{zy} + \frac{\partial}{\partial z}(\tau_{zz} - p) - \rho g = 0. \tag{2.42}$$

If we use non-dimensional coordinates x', y', z' such that $x = Lx'$, $y = Ly'$, $z = Hz'$, this equation becomes

$$\delta\left(\frac{\partial}{\partial x'}\tau_{zx} + \frac{\partial}{\partial y'}\tau_{zy}\right) + \frac{\partial}{\partial z'}(\tau_{zz} - p) - H\rho g = 0 \tag{2.43}$$

upon multiplying through by the thickness scale H. Since the aspect ratio δ is much smaller than 1, we can neglect the first term. In that case, we can replace the last equation with

$$\frac{\partial}{\partial z}(\tau_{zz} - p) - \rho g = 0. \tag{2.44}$$

The boundary condition at the ice surface is $\tau_{ij} n_j = 0$, but again in the limit of low aspect ratio the unit normal vector at the ice surface is approximately equal to the unit vector in the z direction. We can then integrate the last equation from an arbitrary height z up to the ice surface s to get

$$-\tau_{zz} + p - \rho g(s - z) = 0. \tag{2.45}$$

The deviatoric stress tensor is trace-free, so by adding $\tau_{xx} + \tau_{yy} + \tau_{zz}$ to the last equation we get

$$p = \rho g(s - z) - \tau_{xx} - \tau_{yy}. \tag{2.46}$$

We can now eliminate the pressure from the horizontal components of the Stokes equations:

$$\begin{aligned} \frac{\partial}{\partial x}\left(2\tau_{xx}+\tau_{yy}\right)+\frac{\partial}{\partial y}\tau_{xy} &+\frac{\partial}{\partial z}\tau_{xz}-\rho g\frac{\partial s}{\partial x}=0 \\ \frac{\partial}{\partial x}\tau_{yx} &+\frac{\partial}{\partial y}\left(\tau_{xx}+2\tau_{yy}\right)+\frac{\partial}{\partial z}\tau_{yz}-\rho g\frac{\partial s}{\partial y}=0 \end{aligned} \tag{2.47}$$

We can simplify this even further by defining the 2D tensor

$$T_{ij}=\tau_{ij}+\delta_{ij}\tau_{kk}, \tag{2.48}$$

and the vector $\tau_{iz}=(\tau_{xz},\tau_{yz})^\top$ of vertical shear stresses. With these definitions, the system above becomes

$$\frac{\partial}{\partial x_j}T_{ij}+\frac{\partial}{\partial z}\tau_{iz}-\rho g\frac{\partial s}{\partial x_i}=0 \tag{2.49}$$

when written in tensor form. Note that the indices run from $i=1,2$ now instead of $i=1,2,3$. This system of PDE is called the *first-order* or *Blatter-Pattyn* equations (Greve and Blatter, 2009). An even further simplification is possible by recognizing that the vertical shear strain is dominated by vertical changes in the horizontal velocities:

$$\dot{\varepsilon}_{xz}=\frac{1}{2}\left(\frac{\partial u_x}{\partial z}+\frac{\partial u_z}{\partial x}\right)\approx\frac{1}{2}\frac{\partial u_x}{\partial z}. \tag{2.50}$$

With this approximation, the Blatter-Pattyn equations are a closed system of equations for just the horizontal components $(u_x,u_y)^\top$ of the ice velocity. The pressure and vertical velocity have been eliminated entirely by virtue of the approximation (2.50) and the condition (2.46) that the vertical stress balance is roughly "cryostatic".

The Blatter-Pattyn equations can be thought of as a 2D conservation law for the horizontal stress tensor T_{ij}, with two forcing terms. First is the derivative of the vertical shear stress vector τ_{iz}. The vertical shear stress at the ice surface is zero, while the vertical shear stress at the bed has some finite value due to basal resistance. Consequently, the derivative of the vertical shear stress vector is nonzero somewhere in the ice column, and this acts to dissipate the horizontal extension of the ice flow. The second forcing term is the ice surface slope, $\partial s/\partial x_i$. Compared to the Stokes equations, which are forced everywhere by gravity, the tendency to flow downhill is most directly evident in the Blatter-Pattyn equations.

The Blatter-Pattyn equations require boundary conditions to have a unique solution; these can be derived from the boundary conditions for the full Stokes equations. The unit-outward normal vectors to the ice surface and bed can be written in terms of the gradients of the ice surface and bed elevations:

$$\hat{n}|_{z=s}=\left(-\frac{\partial s}{\partial x},-\frac{\partial s}{\partial y},+1\right)^\top/\sqrt{1+|\nabla s|^2},\qquad \hat{n}|_{z=b}=\left(\frac{\partial b}{\partial x},\frac{\partial b}{\partial y},-1\right)^\top/\sqrt{1+|\nabla b|^2}. \tag{2.51}$$

We can then express the boundary conditions for the Blatter-Pattyn equations as the tensor equations

$$-T_{ij}\frac{\partial s}{\partial x_j}+\tau_{iz}=0 \tag{2.52}$$

$$T_{ij}\frac{\partial b}{\partial x_j}-\tau_{iz}=-\sqrt{1+|\nabla b|^2}\cdot C|u|^{\frac{1}{m}-1}u_i, \tag{2.53}$$

which are evaluated respectively at $z=s$ and at $z=b$. Unlike the full Stokes equations, the basal boundary condition at $z=b$ includes the correction factor $\sqrt{1+|\nabla b|^2}$ for the basal surface slope. In practice, the bed slopes are usually small enough that this factor is ignorable or can be included in the friction coefficient C.

The solution of the Blatter-Pattyn equations can also be viewed as the extremizer of some nonlinear functional (Dukowicz et al., 2010). For the Blatter-Pattyn equations, the action is

$$E(u)=\int_\Omega\left\{\frac{2nB}{n+1}|\dot\varepsilon|^{1+\frac{1}{n}}-\rho g\frac{\partial s}{\partial x_i}u_i\right\}\mathrm{d}x \tag{2.54}$$

where now the effective strain rate is

$$|\dot\varepsilon|^2=\frac{1}{2}\left(\dot\varepsilon_{xx}^2+\dot\varepsilon_{yy}^2+(\dot\varepsilon_{xx}+\dot\varepsilon_{yy})^2+2\dot\varepsilon_{xy}^2+2\dot\varepsilon_{xz}^2+2\dot\varepsilon_{yz}^2\right). \tag{2.55}$$

The effective strain rate is different from that of the full Stokes equations because we have eliminated the vertical component $\dot\varepsilon_{zz}$. More substantively, however, is the fact that the action for the Blatter-Pattyn equations is positive-definite. This is in distinct contrast to the Stokes equations, the solution of which amounts to solving a saddle-point problem.

While the Blatter-Pattyn equations are much simpler than the full Stokes equations, they are still a three-dimensional partial differential equation for an essentially two-dimensional phenomenon. We can derive simpler equation sets by considering separately the cases where horizontal stresses dominate and where vertical stresses dominate.

In the deep interior of the ice sheet, the horizontal stresses τ_{xx}, τ_{yy}, τ_{xy} are negligible compared to the vertical shear stresses τ_{xz}, τ_{yz}. This regime is called the *shallow ice approximation* (Greve and Blatter, 2009). To derive the shallow ice approximation (SIA) it will be most convenient to switch over to vector instead of tensor notation; we will write τ_z for the vector $(\tau_{xz},\tau_{yz})^\top$ and likewise for $\dot\varepsilon_z$, and use the gradient operator ∇ instead of indexed partial derivatives, and so forth. Under these assumptions, the first term in (2.49) is negligible, leaving us with

$$\frac{\partial}{\partial z}\tau_z-\rho g\nabla s=0. \tag{2.56}$$

This equation can be integrated from an arbitrary elevation z to the ice surface s to give us the vertical shear stress at any height:

$$\tau_z=-\rho g(s-z)\nabla s. \tag{2.57}$$

The effective stress under this simplification is

$$|\tau| = \left(\tau_{xz}^2 + \tau_{yz}^2\right)^{\frac{1}{2}} = \rho g (s-z)|\nabla s|, \tag{2.58}$$

so that the constitutive relation gives us the vertical shear strain:

$$\frac{1}{2}\frac{\partial u}{\partial z} = \dot{\varepsilon}_z = A|\tau|^{n-1}\tau_z = -A(\rho g(s-z))^n |\nabla s|^{n-1}\nabla s. \tag{2.59}$$

To progress, we need to know the basal sliding velocity $u(z=b)$ and the full depth-dependence of the rate factor A. If we know both of these quantities, the last equation can be integrated from the bed to ascertain the velocity at an arbitrary elevation. In many regions of the interiors of large ice sheets, the ice is frozen to the bed, in which case $u(z=b)=0$ (MacGregor et al., 2016). We can make the further simplifying assumption that A is constant with depth to obtain the relation

$$\begin{aligned} u &= -2A(\rho g)^n \int_b^z (s-z')^n \, \mathrm{d}z' \cdot |\nabla s|^{n-1}\nabla s \\ &= -\frac{2A}{n+1}(\rho g)^n \left(h^{n+1} - (s-z)^{n+1}\right)|\nabla s|^{n-1}\nabla s. \end{aligned} \tag{2.60}$$

For the purposes of computing the flow of the ice sheet, we only need to know the flux of ice mass through some region, not the full depth-dependence of the ice velocity. We can simplify even further by integrating the last equation again to get the depth-averaged ice velocity $\bar{u}$:

$$\begin{aligned} \bar{u} &= \frac{1}{h}\int_b^s u(z)\,\mathrm{d}z \\ &= -\frac{2A}{n+1}(\rho g)^n \cdot \frac{1}{h}\int_b^s \left(h^{n+1} - (s-z)^{n+1}\right)\mathrm{d}z \cdot |\nabla s|^{n-1}\nabla s \\ &= -\frac{2A}{n+2}(\rho g)^n h^{n+1}|\nabla s|^{n-1}\nabla s. \end{aligned} \tag{2.61}$$

In the more general case where the rate factor is not constant with depth, we would need to integrate equation (2.59) twice to arrive at the true depth-averaged velocity. Nonetheless, several features of the shallow ice approximation velocity (2.61) are noteworthy.

First, the ice velocity is a purely local function of the ice thickness and surface slope. In every other model for glacier flow velocity – the full Stokes system, the Blatter-Pattyn equations, the shallow shelf approximation which we will define next, or other higher-order depth-averaged models – the ice velocity is obtained though the solution of some elliptic partial differential equation. In these formulations, the ice velocity is a non-local function of the glacier geometry. This difference has several consequences. When attempting to diagnose the depth-averaged ice velocity from measured data for the ice thickness and surface elevation, noisy data can grossly pollute the computed velocity. In addition to differentiating the true signal in s, we are also differentiating the measurement errors, which are often white

noise. Independent of this issue, the glacier surface often possesses undulations and hummocks due to flow over bumpy bedrock topography; the slope of these hillocks may locally be high enough to point opposite to the direction of flow. In the shallow ice approximation, the velocity would then point opposite to the true flow direction, since the velocity always points directly down-slope. In real glacier flows, the ice velocity can locally point in the opposite direction to the driving stress. The shallow ice approximation is typically applied at comparatively low grid resolutions on the order of ≈ 10 ice thicknesses (Greve, 1997).

Second, the ice velocity is a highly nonlinear function of the thickness and surface slope. Together with the mass conservation equation equation (2.41), the evolution of the ice surface elevation is a nonlinear, degenerate parabolic PDE:

$$\frac{\partial s}{\partial t} + \nabla \cdot \left(\frac{2A}{5} (\rho g)^3 h^5 |\nabla s|^2 \nabla s \right) = \dot{a} - \dot{m}. \tag{2.62}$$

For the case of a radially symmetric ice sheet of constant temperature, the steady state can be solved for analytically in some cases, for example the Vialov profile (Greve and Blatter, 2009).

The stress regime of floating ice shelves and outlet glaciers is the exact opposite of the interior of an ice sheet. In these regions, the vertical shear stress is much smaller than the longitudinal stresses. Since the vertical shear strain rate is proportional to the vertical shear stress, we find that $\dot{\varepsilon}_{xz}/\dot{\varepsilon}_{xx}$ is small too. Consequently, the ice velocity is roughly a *plug flow*, i.e. the velocity is approximately constant with depth. This suggests that we can depth-average the Blatter-Pattyn approximation to arrive at a simpler, 2D equation set. First, the Leibniz integral rule shows that

$$\frac{\partial}{\partial x_j} \int_b^s T_{ij} \, \mathrm{d}z = \int_b^s \frac{\partial}{\partial x_j} T_{ij} \, \mathrm{d}z + \frac{\partial s}{\partial x_j} T_{ij}|_{z=s} - \frac{\partial b}{\partial x_j} T_{ij}|_{z=b}. \tag{2.63}$$

We can now integrate the Blatter-Pattyn equations over the thickness of the glacier and use the previous equation to replace depth-averages of stress divergences with divergences of depth-averaged stresses:

$$0 = \int_b^s \left(\frac{\partial}{\partial x_j} T_{ij} + \frac{\partial}{\partial z} \tau_{iz} - \rho g \frac{\partial s}{\partial x_i} \right) \mathrm{d}z \tag{2.64}$$

$$= \frac{\partial}{\partial x_j} h \bar{T}_{ij} - \frac{\partial s}{\partial x_j} T_{ij}|_{z=s} + \frac{\partial b}{\partial x_j} T_{ij}|_{z=b} \tag{2.65}$$

$$+ \tau_{iz}|_{z=s} - \tau_{iz}|_{z=b} - \rho g h \frac{\partial s}{\partial x_i} \tag{2.66}$$

$$= \frac{\partial}{\partial x_j} h \bar{T}_{ij} - \sqrt{1 + |\nabla b|^2} \cdot C |u|^{\frac{1}{m}-1} u_i|_{z=b} - \rho g h \frac{\partial s}{\partial x_i} \tag{2.67}$$

This equation is not yet a closed system of PDE for the depth-averaged velocities $\bar{u}_i$ – it still includes the basal sliding velocity $u_i|_{z=b}$, and the depth-averaged stress is not necessarily equal to the stress of the depth-averaged velocities. Define the *membrane stress tensor*

$$M_{ij} = 2h \cdot \mu(\dot{\varepsilon}(\bar{u})) \cdot (\dot{\varepsilon}_{ij}(\bar{u}) + \delta_{ij} \dot{\varepsilon}_{kk}(\bar{u})), \tag{2.68}$$

which has units of stress × thickness. With the approximation that $\dot{\varepsilon}_{xz} \ll \dot{\varepsilon}_{xx}$, we have that $h\bar{T}_{ij} \approx M_{ij}$, and $u_i|_{z=b} \approx \bar{u}_i$. If we lump the geometric factor $\sqrt{1+|\nabla b|^2}$ into C, equation (2.67) then simplifies to

$$\frac{\partial}{\partial x_j} M_{ij} - C|\bar{u}|^{\frac{1}{m}-1}\bar{u}_i - \rho g h \frac{\partial s}{\partial x_i} = 0, \tag{2.69}$$

where M is the membrane stress tensor. These are the shallow stream equations. We can simplify them even further for the case of a floating ice shelf, in which case the basal friction coefficient C is 0. Moreover, for the ice to be floating, the cryostatic pressure at the ice base must be equal to the pressure exerted by the surrounding ocean. The depth z_b of the ice base is equal to $s-h$; at this depth, the ocean exerts a pressure $p_W = -\rho_W g(s-h)$. The ice pressure from a column of thickness h is $p_I = \rho_I g h$; setting these two quantities equal to each other gives

$$s = \left(1 - \frac{\rho_I}{\rho_w}\right) h \approx h/10. \tag{2.70}$$

If we let $\varrho = (1-\rho_I/\rho_W)\rho_I$ be the *reduced density* and use the fact that $h\frac{\partial h}{\partial x_i} = \frac{1}{2}\frac{\partial h^2}{\partial x_i}$, the shallow stream equations become

$$\frac{\partial}{\partial x_j} M_{ij} - \frac{1}{2}\varrho g \frac{\partial}{\partial x_i} h^2 = 0. \tag{2.71}$$

These are the *shallow shelf equations*; they are a conservation law for membrane stress. Both the shallow stream and shallow shelf equations are second-order, nonlinear, elliptic systems of PDE for the depth-averaged ice velocity $\bar{u}$. Unlike the shallow ice approximation, the ice velocity in the shallow stream/shelf approximation is a nonlocal function of the ice surface and bed elevation, since the velocity is obtained through the solution of an elliptic PDE.

The shallow ice approximation works well in areas where the vertical shear stress dominates, such as in the deep interior of a large ice sheet, while the shallow stream approximation works well in areas where longitudinal stresses dominate, such as fast-flowing ice streams or floating ice shelves. Defining the stress ratio parameter $\lambda = \dot{\varepsilon}_{xz}/\dot{\varepsilon}_{xx}$, these approximations can formally be described as perturbative expansions in the limits where $\lambda \gg 1$ and $\lambda \ll 1$ respectively. Of course, both extremes can occur throughout a real ice sheet, so if we wanted to model the evolution of an entire drainage basin of a glacier we would need either

1. a procedure for using different physical models in different regions; or
2. a more intricate physical model that can exhibit both shallow ice-like and shallow stream-like behavior.

A difficulty with using the first approach is that of deciding where the dividing line should be between the regions where each physical model is used. If the boundary is chosen poorly, the sudden change in the mathematical model could introduce an unphysical shock to the system. Alternatively, one could use the Blatter-Pattyn

or even full Stokes equations only in the transition region for the greatest physical acuity. A blunt approach to implementing the second approach would be to use the Blatter-Pattyn or full Stokes models everywhere; several studies have advocated precisely this approach (Larour et al., 2012; Gagliardini et al., 2013; Brinkerhoff and Johnson, 2013). While 3D physics are more accurate, they come with a high computational cost compared to depth-averaged models.

As a compromise between the two extremes, several studies have attempted to develop "hybrid" models of ice flow (Schoof and Hindmarsh, 2010; Goldberg, 2011; Perego et al., 2012). A hybrid model is a 2D model of glacier flow, but which is nonetheless still valid in both the fast-sliding and no-sliding regimes. Rather than assume that vertical shear strain rates are negligible, hybrid models instead use the SIA balance as a first guess at the vertical shear strain. This guess can then be used in the Blatter-Pattyn equations and depth-averaged to get a 2D PDE for the depth-averaged velocities. Schoof and Hindmarsh (2010) has shown that models derived in this way are formally accurate to 2nd order in perturbation theory. Goldberg (2011) and Perego et al. (2012) compared hybrid models to the Blatter-Pattyn and Stokes models, finding that the hybrid models generally agree well with 3D models at length scales greater than 20 km. At shorter length scales, the hybrid models sometimes reproduce the 3D model results and sometimes do not, depending on the particular surface and bed elevation.

A final approach to developing more accurate depth-averaged models is to use semi-discretization. In a semi-discretization approach, the governing partial differential equations are discretized in some, but not all, of the spatiotemporal variables. The resulting system of PDE for the semi-discretized fields can then be thought of as a new mathematical model. For example, Jouvet (2015) suggests first using a finite-difference discretization of the Blatter-Pattyn equations in the z variable. Semi-discretization then gives a coupled system of 2D elliptic PDE for each layer velocity. Using a successively greater number of layers gives a hierarchy of successively more accurate models. One problem with this approach is the need for a larger number of layers in the interior of the ice sheet than in the ice streams. Rather than use finite difference methods as the basis for semi-discretization, Bassis (2010) instead use Galerkin methods. The velocity is written as an expansion in a set of basis functions p_k:

$$u = \sum_k p_k(\zeta) u_k(x, y) \tag{2.72}$$

where $\zeta = (z - b)/h$ is a stretched vertical coordinate. This approach is reminiscent of the terrain-following coordinate systems commonly applied in ocean modelling, for example in the Regional Ocean Modelling System (Haidvogel et al., 2008). Since the ice bed elevation and thickness vary in space, the basis functions $p_k(\zeta)$ are also functions of x, y as well as the vertical coordinate z. In particular, the gradient of ζ can be written as

$$\nabla\zeta = -\frac{\nabla b}{h} + \zeta\frac{\nabla h}{h}. \tag{2.73}$$

This complication makes it difficult to explicitly write down the system of PDEs that the modes u_k satisfy upon substituting equation (2.72) into the Blatter-Pattyn

equations. Instead, Bassis (2010) rely on the variational formulation of the Blatter-Pattyn equations (eq. (2.54)). The strain rate tensor for u can be written explicitly as

$$\dot{\varepsilon}(u) = \sum_k \left(p_k(\zeta)\dot{\varepsilon}(u_k) + \frac{1}{2}\frac{\mathrm{d}p_k}{\mathrm{d}\zeta}(\nabla\zeta \otimes u_k + u_k \otimes \nabla\zeta) \right). \tag{2.74}$$

All other quantities needed to solve for the expansion coefficients u_k can be computed from this expression, together with equation (2.73) for $\nabla\zeta$. An obvious choice of the first basis function is $p_1 = 1$ to represent pure plug flow, but the remaining basis functions need not be orthogonal. The exact solution of the shallow ice approximation with constant temperature has a vertical dependence of $1-(1-\zeta)^4$ (see equation (2.60)); Bassis (2010) uses this expression as the second basis function p_2. Bassis then demonstrates that this semi-spectral model can exhibit a transition from shallow-ice to shallow-stream like flow for a simple flowline geometry.

2.4 Linearization

The diagnostic equations for ice velocity are nonlinear as a consequence of Glen's flow law. To solve a nonlinear system of equations $G(u) = 0$, we would like to be able to use Newton's method; given a guess u_k for the solution, an improved guess can be obtained through the iteration

$$u_{k+1} = u_k - \left(\frac{\mathrm{d}G}{\mathrm{d}u} \right)^{-1} G(u_k) \tag{2.75}$$

where the derivative $\mathrm{d}G/\mathrm{d}u$ is evaluated at u_k. For a smooth function G and sufficiently good initial guess, the iterates from Newton's method converge quadratically to the solution of the nonlinear system. In order to use Newton's method, we need to be able to linearize the model physics equations around some basic state u. In this section, we will compute the linearization of the Stokes and shallow shelf equations, since these are the two models used most throughout this work.

As we have shown in the last section, the Stokes equations and all of the approximate models we derived can be obtained as the Euler-Lagrange equations for the extremization of some functional E. In the case of the Stokes equations, extremization involved finding a saddle point, whereas for the Blatter-Pattyn and shallow shelf equations the energy functional is convex and positive-definite. The energy functional is a map from some function space X to $\mathbb{R}$, so the derivative $G = \mathrm{d}E/\mathrm{d}u$ is a map from X to the dual space X^* of X. The linearization of the model physics $\mathrm{d}G/\mathrm{d}u$ is actually the second derivative $\mathrm{d}^2E/\mathrm{d}u^2$ of the energy functional. The second derivative $\mathrm{d}^2E/\mathrm{d}u^2$ evaluated around some state u is a bilinear form on X, but Clairaut's theorem tells us that it is also a *symmetric* bilinear form. In other words, we should expect that the linearization of the model physics equations is not just any kind of differential operator, but a self-adjoint one. Moreover, the energy functionals for the Blatter-Pattyn and shallow shelf equations are both convex, so we should also find that $\mathrm{d}G/\mathrm{d}u$ is positive semi-definite.

Ignoring the pressure in the Stokes equations, the PDEs we are interested in have a special structure – they are *quasilinear*. The operator $G(u)$ can be written as

$$G(u) = \nabla \cdot K(\nabla u)\nabla u - C(u)u - f \tag{2.76}$$

where K and C take values in the spaces of positive rank-4 and rank-2 symmetric tensors respectively. If v is some perturbation to the flow field u, a first guess at the linearization of this operator around u might be to "freeze the coefficients" of G, i.e.

$$\frac{\partial G}{\partial u} v \approx \nabla \cdot K(\nabla u)\nabla v - C(u)v. \tag{2.77}$$

Using Picard's method to solve the nonlinear system $G(u) = 0$ is exactly equivalent to Newton's method with a frozen-coefficient approximation to the derivative. Freezing the coefficients is only an approximation, however, and the full linearization will include corrections to the coefficients K, C due to their explicit dependence on u.

We will need a few preliminaries on tensor algebra before we continue. For the derivations that follow, we will make a temporary digression into vector notation instead of the usual index notation traditionally used for computations with tensors. As a shorthand, we will write $\dot{\varepsilon}(u)$, $\dot{\varepsilon}(v)$ to distinguish between the strain rate tensors of the basic state velocity u and the perturbed field v. The set of symmetric rank-2 tensors in dimension d is a $d(d+1)/2$-dimensional vector space, which can be endowed with the inner product

$$\langle a, b \rangle = \mathrm{tr}\,(a^\top b) = a_{ij} b_{ji}. \tag{2.78}$$

In an arbitrary inner product space, the *outer product* of two elements a, b is the linear operator $a \otimes b$ that acts on another element c according to

$$(a \otimes b)c = \langle b, c \rangle \cdot a, \tag{2.79}$$

which is a linear operator in c by the properties of an inner product. The orthogonal projection onto an element a of this space can be concisely expressed as

$$P_a = \frac{a}{\|a\|} \otimes \frac{a}{\|a\|}. \tag{2.80}$$

For the space of rank-2 symmetric tensors, the linear operators on this space can be thought of as the set of all rank-4 tensors with the symmetry properties

$$C_{jikl} = C_{ijkl}, \qquad C_{ijlk} = C_{ijkl}. \tag{2.81}$$

The outer product of two symmetric rank-2 tensors can be expressed in index notation as

$$(a \otimes b)_{ijkl} = a_{ij} b_{kl}. \tag{2.82}$$

The outer product $a \otimes a$ of a tensor with itself also has the symmetry $C_{klij} = C_{ijkl}$, i.e. it is a self-adjoint operator on the space of rank-2 tensors. The tensor outer

product will be especially useful to us in the following for expressing the equivalent constitutive tensors for the linearization of the model physics equations.

First, we will tackle the Stokes equations. To compute the action of the linearized Stokes equations on some perturbation δv, we need to compute the $\mathcal{O}(\delta)$ term in the expansion of the perturbed stress tensor:

$$\begin{aligned}\tau(\dot\varepsilon(u+\delta v))-\tau(\dot\varepsilon(u)) &= 2\mu(\dot\varepsilon(u)+\delta\dot\varepsilon(v))\cdot\dot\varepsilon(u+\delta v)-2\mu(\dot\varepsilon(u))\dot\varepsilon(u)\\ &= \delta\left(\mu(\dot\varepsilon(u))\dot\varepsilon(v)+2\left\langle\frac{\mathrm{d}\mu}{\mathrm{d}\dot\varepsilon},\dot\varepsilon(v)\right\rangle\dot\varepsilon(u)\right)+\mathcal{O}(\delta^2). \quad (2.83)\end{aligned}$$

In the order-1 term of this Taylor expansion, the first part $\mu(\dot\varepsilon(u))\dot\varepsilon(v)$ is what we would expect to find by freezing the coefficients, while the second term is the correction accounting for the dependence of the viscosity on u.

In equation (2.21), we defined the constitutive relation between the stress and strain rate tensors in terms of the effective strain rate

$$|\dot\varepsilon(u)| = \sqrt{\langle\dot\varepsilon(u),\dot\varepsilon(u)\rangle/2}. \quad (2.84)$$

This factor of 2 means that the effective strain rate is not equal to the norm $\|\dot\varepsilon(u)\|$ of the strain rate tensor with respect to this inner product. Nonetheless, the effective strain rate is the quantity used to define the constitutive relation in most textbooks on glacier modelling, e.g. (Greve and Blatter, 2009). In any case, the perturbed value of the effective strain rate is

$$|\dot\varepsilon(u+\delta v)|^2 = |\dot\varepsilon(u)|^2+\delta\langle\dot\varepsilon(u),\dot\varepsilon(v)\rangle+\mathcal{O}(\delta^2). \quad (2.85)$$

Using the Taylor expansion

$$t^{\frac{1-n}{2n}} = t_0^{\frac{1-n}{2n}}+\frac{1-n}{2n}t_0^{\frac{1-3n}{2n}}(t-t_0)+\mathcal{O}(t-t_0)^2 \quad (2.86)$$

for the function $f(t)=t^{\frac{1-n}{2n}}$ and equation (2.85), the perturbed value of the ice viscosity around u is

$$\begin{aligned}\mu(\dot\varepsilon(u+\delta v)) &= \frac{B}{2}\left(|\dot\varepsilon(u+\delta v)|^2\right)^{\frac{1-n}{2n}} &(2.87)\\ &= \frac{B}{2}\left(|\dot\varepsilon(u)|^2+\delta\langle\dot\varepsilon(u),\dot\varepsilon(v)\rangle+\mathcal{O}(\delta^2)\right)^{\frac{1-n}{2n}} &(2.88)\\ &= \mu(\dot\varepsilon(u))+\frac{B}{2}\cdot\frac{1-n}{2n}|\dot\varepsilon(u)|^{\frac{1-3n}{n}}\langle\dot\varepsilon(u),\dot\varepsilon(v)\rangle\delta+\mathcal{O}(\delta^2) &(2.89)\\ &= \mu(\dot\varepsilon(u))\left(1+\frac{1-n}{2n}\frac{\langle\dot\varepsilon(u),\dot\varepsilon(v)\rangle}{|\dot\varepsilon(u)|^2}\delta\right)+\mathcal{O}(\delta^2). &(2.90)\end{aligned}$$

Putting together equations (2.83) and (2.90), we find that

$$\begin{aligned}\frac{\tau(\dot\varepsilon(u+\delta v))-\tau(\dot\varepsilon(u))}{\delta} &= 2\mu(\dot\varepsilon(u))\dot\varepsilon(v)\\ &\quad+2\mu(\dot\varepsilon(u))\cdot\frac{1-n}{2n}\frac{\langle\dot\varepsilon(u),\dot\varepsilon(v)\rangle}{|\dot\varepsilon(u)|^2}\dot\varepsilon(u)+\mathcal{O}(\delta). \quad (2.91)\end{aligned}$$

We can make this rather awful expression much nicer by recognizing that

$$\langle \dot{\varepsilon}(u), \dot{\varepsilon}(v)\rangle \dot{\varepsilon}(u) = (\dot{\varepsilon}(u) \otimes \dot{\varepsilon}(u))\dot{\varepsilon}(v) \tag{2.92}$$

using the definition of the outer product of two elements of a vector space. Let $\mathscr{I}$ be the identity operator on rank-2 tensors; using the previous expression, equation (2.91) can be rewritten as

$$\frac{\tau(\dot{\varepsilon}(u+\delta v)) - \tau(\dot{\varepsilon}(u))}{\delta} = 2\mu(\dot{\varepsilon}(u))\left(\mathscr{I} + \frac{1-n}{2n}\frac{\dot{\varepsilon}(u)\otimes\dot{\varepsilon}(u)}{|\dot{\varepsilon}(u)|^2}\right)\dot{\varepsilon}(v). \tag{2.93}$$

This is the final step we need; the action of the derivative of the Stokes operator $G(u,p)$ on a vector field v is

$$\frac{\partial G}{\partial u}v = \nabla\cdot 2\mu(\dot{\varepsilon}(u))\left(\mathscr{I} + \frac{1-n}{2n}\frac{\dot{\varepsilon}(u)\otimes\dot{\varepsilon}(u)}{|\dot{\varepsilon}(u)|^2}\right)\dot{\varepsilon}(v). \tag{2.94}$$

Several features of this operator are noteworthy. The nonlinear Stokes operator is isotropic, i.e. it is invariant to a rotation of the coordinate system. The linearized operator, however, is anisotropic; the viscosity can now be thought of as a rank-4 tensor acting on the strain rate $\dot{\varepsilon}(v)$ of the perturbation, rather than just a scalar multiple. The strain rate tensor $\dot{\varepsilon}(u)$ of the basic state u defines the favored anisotropic orientation. Since $n = 3$ for glacier flows, the quantity $(1-n)/2n$ is negative, so the effect of this anisotropy is to weaken the linearized medium to deformations the more aligned they are with $\dot{\varepsilon}(u)$. By contrast, if ice were a shear-thickening material, we would have $n < 1$, in which case the anisotropic correction would be a positive operator and the medium would instead be stronger to deformations aligned with $\dot{\varepsilon}(u)$.

At the beginning of this section, we argued that the linearized operator should be symmetric and positive-definite. Since the outer product of a tensor with itself is a symmetric rank-4 tensor, the anisotropic correction to the constitutive tensor is also symmetric. The symmetric of the constitutive tensor implies that the elliptic operator is self-adjoint by the usual trick with Green's identities. Although the anisotropic correction is overall a negative operator because $n > 1$, we still find that, for some arbitrary rank-2 tensor a,

$$\left\langle a, \left(\mathscr{I} + \frac{1-n}{2n}\frac{\dot{\varepsilon}(u)\otimes\dot{\varepsilon}(u)}{|\dot{\varepsilon}(u)|^2}\right)a\right\rangle = \|a\|^2 - \frac{n-1}{n}\frac{|\langle\dot{\varepsilon}(u), a\rangle|^2}{\|\dot{\varepsilon}(u)\|^2} \geq \|a\|^2/n \tag{2.95}$$

by the Cauchy-Schwarz inequality. Note that as the material grows more plastic, i.e. n grows larger and larger, this lower bound approaches 0.

The linearization of the shallow stream equations can derived in a similar fashion as the Stokes equations. Let $\mathscr{I}$ be the identity operator on rank-2 symmetric tensors, and I the Kronecker tensor; the shallow stream equations can be written as

$$\nabla\cdot 2\mu(\dot{\varepsilon}(u))(\mathscr{I} + I\otimes I)\dot{\varepsilon}(u) - C|u|^{\frac{1}{m}-1}u - \rho g h\nabla s = 0. \tag{2.96}$$

With the cryostatic and low-shear stress approximations, the effective strain rate for the shallow stream equations in components is

$$|\dot{\varepsilon}|^2 = \frac{1}{2}\left(\dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij} + (\dot{\varepsilon}_{kk})^2\right). \tag{2.97}$$

Unlike in 3D, the trace $\dot{\varepsilon}_{kk}$ of the depth-averaged strain rate is non-zero. The perturbed effective strain rate is

$$|\dot{\varepsilon}(u+\delta v)|^2 = |\dot{\varepsilon}(u)|^2 + \delta\left(\langle\dot{\varepsilon}(u),\dot{\varepsilon}(v)\rangle + \operatorname{tr}\dot{\varepsilon}(u)\cdot\operatorname{tr}\dot{\varepsilon}(v)\right) + \mathscr{O}(\delta^2). \tag{2.98}$$

Using the Taylor expansion for $t^{\frac{1-n}{2n}}$ again, we find that the perturbed viscosity is

$$\mu(\dot{\varepsilon}(u+\delta v)) = \frac{B}{2}\left(|\dot{\varepsilon}(u+\delta v)|^2\right)^{\frac{1-n}{2n}} \tag{2.99}$$

$$= \frac{B}{2}\left(|\dot{\varepsilon}(u)|^2 + \delta(\langle\dot{\varepsilon}(u),\dot{\varepsilon}(v)\rangle + \operatorname{tr}\dot{\varepsilon}(u)\cdot\operatorname{tr}\dot{\varepsilon}(v)) + \mathscr{O}(\delta^2)\right)^{\frac{1-n}{2n}} \tag{2.100}$$

$$= \mu(\dot{\varepsilon}(u))\left(1 + \frac{1-n}{2n}\frac{\langle\dot{\varepsilon}(u),\dot{\varepsilon}(v)\rangle + \operatorname{tr}\dot{\varepsilon}(u)\cdot\operatorname{tr}\dot{\varepsilon}(v)}{|\dot{\varepsilon}(u)|^2}\delta\right) + \mathscr{O}(\delta^2). \tag{2.101}$$

If we now define a dimensionless tensor

$$\gamma = \frac{\dot{\varepsilon}(u) + \operatorname{tr}\dot{\varepsilon}(u)\cdot I}{\sqrt{2}|\dot{\varepsilon}(u)|}, \tag{2.102}$$

then the perturbed membrane stress tensor is

$$\frac{M(\dot{\varepsilon}(u+\delta v)) - M(\dot{\varepsilon}(u))}{\delta} = 2\mu(\dot{\varepsilon}(u))\left(\mathscr{I} + I\otimes I + \frac{1-n}{2n}\gamma\otimes\gamma\right)\dot{\varepsilon}(v) + \mathscr{O}(\delta). \tag{2.103}$$

Much like the Stokes equations, the linearization of the shallow stream equations consists of the sum of the frozen-coefficient approximation and an anisotropic correction. The orientation of anisotropy is dictated by the tensor $\dot{\varepsilon}(u)+\operatorname{tr}\dot{\varepsilon}(u)I$ rather than the strain-rate tensor alone, as in the linearization of the Stokes equations.

The final component is the linearization of the boundary condition

$$\tau_b = -C|u|^{\frac{1}{m}-1}u. \tag{2.104}$$

The algebra is substantially simpler since we only have to deal with vectors instead of rank-2 tensors. Since $|u+\delta v|^2 = |u|^2 + 2\delta\langle u, v\rangle + \mathscr{O}(\delta^2)$, the Taylor expansion of the basal shear stress is

$$\tau_b(u+\delta v) = -C|u|^{\frac{1}{m}-1}\left(I + \frac{1-m}{m}\frac{u\otimes u}{|u|^2}\right)v. \tag{2.105}$$

Much like the stress tensors for the Stokes and shallow stream equations, the linearization of the basal shear stress is anisotropic, with the velocity u defining the direction of anisotropy. Since $m > 1$ in most cases of glaciological interest, the basal sliding resistance is weaker along the direction u for the linearized medium than across-flow. For the balance of internal stresses, almost every study in glaciology uses $n = 3$, so we needn't worry about degeneracy of the linearized problem. However, when it comes to the basal shear stress, $m = \infty$ is quite common in cases where all basal motion of the glacier is attributed to deformation within a perfectly plastic sedimentary substrate (Joughin et al., 2004b, 2009). The well-posedness of

the linearized PDE is not in jeopardy in the limit as $m \to \infty$, which would be the case if we were to take the same limit with the Glen flow law exponent n. However, the full linearization departs most dramatically from the frozen-coefficient approximation in this limit. It is most important to retain the anisotropic correction to the linearized equations in the limit of perfect plasticity.

Chapter 3

Data assimilation

This chapter focuses on how observations of glacier elevation, thickness, and velocity can be used to estimate quantities that cannot be observed directly, such as the rheology or basal shear stress. Douglas MacAyeal was the first to apply these methods in glaciology to examine the basal stress regimes of the ice streams feeding the Ross Ice Shelf. In the mid-1980s, seismic measurements upended the conventional wisdom that these ice streams flow over hard beds, revealing instead that they flow over deformable sediments. No observational method, however, was capable of estimating the spatial pattern of basal shear stress, i.e. the extent of where soft-bed sliding occurs. MacAyeal (1992) applied methods of PDE-constrained optimization to estimate the basal shear from observations. MacAyeal found that the region of soft-bed sliding is not homogeneous; instead, there are intermittent "sticky spots", which had been hypothesized to exist, of high bed resistance. Since then, data assimilation methods have been a valuable tool in exploring aspects of glacier physics that are not observable directly through remote sensing.

3.1 History

In the late 1970s, it was thought that most glaciers flow over hard, lithified beds (Boulton, 1979). Measurements of glacial erosion rates find typical values on the order of 1 mm/year. Given that Earth has had permanent ice sheets for the last million years, and large ice sheets in North America and Scandinavia during ice ages, the ice sheets can in principle have eroded a kilometer of material over the whole Pleistocene epoch. Glacial erosion will have scoured away any soft sediments underlying the permanently glaciated regions by now, with only hard bedrock remaining.

The dominant theory at this time for how glaciers interact with their beds was due to Weertman. Weertman (1957) proposed a mechanism of glacier sliding in which ice melts under high pressure at the upstream side of a bedrock bump and diffuses to the downstream side, where the water subsequently refreezes. The latent heat released in refreezing conducts through the bed material back upstream,

allowing further melting. This process is called *regelation*. Kamb and Lachapelle (1964) sought to confirm Weertman's theory through a field study at Blue Glacier in Washington State. They observed that in some regions, the basal water pressure was high enough to cause the ice sole to lose contact with the bed. Meltwater and the incipient loss of ice-bed contact were thus thought to be the primary factors responsible for rapid glacier sliding.

The fast flow of the ice streams feeding the Ross Ice Shelf, however, was considered something of an enigma (Murray, 1997). The Ross Ice Streams flow at speeds on the order of 1 km/year even though they are almost completely flat. Ice Stream B[1] is roughly 1 km thick, with a surface slope on the order of 0.002, giving a driving stress on the order of 2–20 kPa. The meltwater production rate necessary to sustain a water film of sufficient thickness to permit fast sliding was found to be too high for the observed driving stress and geothermal heat flux at Ice Stream B (Alley et al., 1986).

In 1986, Richard Alley, Don Blankenship and others undertook a series of seismic measurements at Ice Stream B as part of the ongoing field campaigns to explore the geophysics of the Ross Ice Streams. If the ice were flowing entirely over hard bedrock, the reflection from a seismic signal at the surface would consist of one return from the compressional P-wave, followed by a second return from the shear S-wave. Instead, what they observed was a P-wave return from the ice base, a "ghost" P-wave return, the S-wave return from the ice base, and finally a "ghost" S-wave return (Blankenship et al., 1986). Based on the travel times of ghost seismic signals, they conclude that the "ghost" return is due to propagation of the seismic signal through a secondary layer beneath the glacier itself. This layer has a thickness of roughly 5-8 m, a P-wave velocity of 1600 m/s and an S-wave velocity of 150 m/s. Such low shear-wave velocities are typically found only porous materials saturated by water at high pressure (Blankenship et al., 1986). From the P- and S-wave velocities, Blankenship estimates that the porosity of the till layer underneath the ice stream is $\phi \approx 40\%$ and the effective pressure, the difference between the ice overburden pressure and the water pressure in the till, is $N \approx 50 \pm 40$ kPa. Alley et al. (1986) argue that a porosity of 30% is typical of lodged, non-deforming tills, whereas a porosity as high as 40% is indicative of a deforming till. Moreover, the yield stress for till as a function of effective pressure is known from laboratory tests; with such low values of the effective pressure at Ice Stream B, the driving stress likely exceeds the calculated yield stress for till. Alley then concluded that nearly all of the flow of Ice Stream B, and possibly the other fast-flowing Ross Ice Streams, is due not to deformation within the ice or sliding over its bed, but to deformation of the bed itself.

These seismic studies demonstrated that the beds under the Ross Ice Streams are deforming in some locations, but they do not reveal the spatial extent of where the bed is weak. Where does the transition from hard to deformable beds occur? Alley et al. (1986) and several subsequent papers postulated that all fast flow in

[1]The ice streams feeding the Ross Ice Shelf were historically designated with letters as Ice Stream A, B, C, D, and E. These were later renamed respectively Mercer, Whillans, Kamb, Bindschadler, MacAyeal, and Echelmeyer Ice Streams. For brevity's sake, I will refer to them by their historical designations as Ice Stream A–E.

Antarctica is by bed deformation, but this idea cannot be tested experimentally without some means of large-scale mapping of basal shear stress.

Several years after the initial publications by Alley and Blankenship on soft-bed sliding, Kamb (1991) objected to the notion that all of the flow of the Ross ice streams is due to bed deformation. Based on a comparison with soil mechanics, Kamb argued that the constitutive relation for dilatant till is highly nonlinear, almost to the point of being perfectly plastic. In some parameter regimes, the coupled ice stream-sediment-hydrological system is unstable to small perturbations because enhanced till deformation permits faster ice flow, increasing the basal melt rate and water pressure, the latter of which only makes the till more deformable. Kamb offers a possible explanation for the apparent stability of the Ross Ice Streams: isolated "sticky spots" of high basal shear stress are actually controlling the ice flow. These sticky spots may consist of protrusions of the underlying hard bedrock through the till layer. When Kamb's paper was published in 1991, there was no existing observational evidence for or against these sticky spots.

MacAyeal (1992) was the first paper to propose a method for large-scale mapping of shear stress beneath ice streams. MacAyeal applied so-called *control methods* to infer the shear stress underneath a fast-flowing ice stream indirectly, using only measurements of the velocity, bed elevation, and surface elevation. Today, glaciologists are accustomed to a veritable deluge of remotely sensed measurements of glacier geometry and velocity. At the time of MacAyeal's work, estimating glacier velocity through repeat-image feature tracking of satellite optical imagery was a comparatively new technique, having been applied to Ice Stream E only the previous year (Bindschadler and Scambos, 1991). MacAyeal used his control method to infer the basal shear stress of Ice Stream E from these velocity measurements, finding kilometer-wide sticky spots exactly as Kamb (1991) had predicted.

3.2 The adjoint method

The basal friction and rheology of a glacier are spatially-varying coefficients in the diagnostic equations; normally, these quantities are thought of as inputs that we use to compute the ice velocity. In practice, however, we are able to observe the ice velocity, but the basal friction and rheology are unknown. This situation is an example of a PDE *inverse problem*. In general, the parameters θ are related to the field u that we observe through a map G, called the *forward map*, such that

$$u = G(\theta). \tag{3.1}$$

We are given a concrete observation u^o, and we would like to back out which value ϑ of the parameter field resulted in the field we observed. Suppose that the parameters θ live in some function space P, and the observations u live in some function space X; in that case, the forward map is a function $G : P \to X$. In principle, we might try to approximate the inverse map to infer ϑ, i.e. $\vartheta \approx G^{-1}(u^o)$. However, evaluating the inverse of the forward map may be ill-posed. For example, the observed field u^o consists of some finite number, N, of independent measurements, for example at several different spatial locations, while we may be representing the

parameters as a linear combination of M basis functions. Usually we use far fewer basis functions to represent the parameters than we have independent measurements, i.e. $M \ll N$. In that case, the inverse map is, for all intents and purposes, ill-defined.

Instead, we can pose the problem of inferring θ using optimization. Rather than try to approximate $G^{-1}(u^o)$, we try to find a value of θ for which $G(\theta) - u^o$ is small. The size of the model-data misfit is quantified through some *error functional* E. The error functional is assumed to be convex and positive, and that $E(0) = 0$. A common choice of error functional is the sum of squared errors:

$$E(u - u^o) = \sum_k \frac{|u(x_k) - u^o(x_k)|^2}{2\sigma_k^2}, \tag{3.2}$$

where $\{x_k\}$ are the physical locations where the measurements were taken and σ_k are the measurement error standard deviations.

Having chosen some error functional, our goal is to find the minimizer of the *objective functional*

$$J(\theta) = E(G(\theta) - u^o). \tag{3.3}$$

The forward map and the error metric are usually differentiable. If we can devise a practical procedure to compute the derivative of the objective functional, then we can use common optimization algorithms such as nonlinear conjugate gradients or BFGS to iteratively approximate the minimizer of J.

Note that J is a map from the parameter space P to the real numbers $\mathbb{R}$, so the derivative of J at a particular value θ of the parameters is a linear map from P to $\mathbb{R}$. In other words, $\mathrm{d}J/\mathrm{d}\theta$ is an element of the dual space P^*. Elements of the parameter space P are spatial fields that we can evaluate at a point in the physical domain Ω. Elements of the dual space, however, are not – they are defined solely through how they act on elements of the space P. For example, a delta function or a line source within a 2- or 3D domain are both perfectly acceptable elements of the dual space, but it does not make sense to evaluate them at a point.

Let ϕ be some element of the parameter space P. As a first step towards a procedure to compute the derivative of the objective functional, we can apply the chain rule to express the action of $\mathrm{d}J/\mathrm{d}\theta$ on ϕ in terms of $\mathrm{d}E/\mathrm{d}u$ and $\mathrm{d}G/\mathrm{d}\theta$:

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\theta}, \phi \right\rangle_P = \left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, \frac{\mathrm{d}G}{\mathrm{d}\theta}\phi \right\rangle_X. \tag{3.4}$$

As a sanity check we can verify that all the function spaces make sense. The increment direction ϕ is an element of the parameter space P, $\mathrm{d}G/\mathrm{d}\theta$ maps the parameter space P to the solution space X, so $\mathrm{d}G/\mathrm{d}\theta \cdot \phi$ is an element of X. Since $\mathrm{d}E/\mathrm{d}u$ is an element of X^*, the duality pairing in the last equation makes sense.

The error functional usually has some fairly simple expression, so we can compute the action $\langle \mathrm{d}E/\mathrm{d}u, v\rangle_X$ on an arbitrary v in X analytically. For example, if E is the squared error functional from equation (3.2), then the derivative of E is

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, v \right\rangle = \sum_k \frac{(u(x_k) - u^o(x_k)) \cdot v(x_k)}{\sigma_k^2}. \tag{3.5}$$

The forward map, on the other hand, is more complicated; in our case, evaluating the forward map will involve solving a nonlinear system of partial differential equations. How then are we to evaluate the derivative of this map between function spaces?

The most obvious way to approximate the derivative of the forward map is to use finite differences. Given a starting value for the parameters θ, we select a set $\{\phi_1, \ldots, \phi_M\}$ of basis functions and compute the difference quotients

$$\delta_m G = \frac{G(\theta + \delta\phi_m) - G(\theta)}{\delta}. \tag{3.6}$$

Approximating the directional derivative of the objective functional along each of the M directions $\{\phi_m\}$ requires M forward-model solves, which may be prohibitively expensive. Moreover, without some *a priori* knowledge of the magnitude of the gradient, it is not immediately clear how large or small a value of the increment δ is necessary.

The *adjoint method* is a systematic way to compute the derivative of a functional of the solution of a differential equation using only one forward model solve and one solve of the linearization of the forward model (Vogel, 2002). In equation (3.4), we were able to express the derivative of the objective functional in terms of the derivatives of the error metric E and the forward map G. The derivative of the error metric is easy to compute, but we are stuck at computing the derivative of the forward map. The adjoint method takes advantage of the fact that the forward map, which comes from the solution of a differential equation, is expressed in implicit form. In other words, to evaluate the forward map $u = G(\theta)$, what we really mean is: find a value of u such that

$$\langle F(u, \theta), v \rangle_X = 0 \quad \text{for all } v \text{ in } X. \tag{3.7}$$

This last equation is the general abstract weak form for a nonlinear PDE. The operator F could be the nonlinear Stokes equations, the shallow shelf equations, and so forth. The implicit equation (3.7) defines a mapping $G : P \to X$ such that

$$\langle F(G(\theta), \theta), v \rangle_X = 0 \quad \text{for all } v \text{ in } X, \tag{3.8}$$

but this abstract solution operator is much less natural to work with than the differential operator F, which is just the familiar PDE defining our model physics. By using the implicit function theorem in Banach spaces, we can come up with an expression for the derivative of G in terms of the governing differential operator F. The application of this linear operator is then "pushed over" onto the functional $\mathrm{d}E/\mathrm{d}u$ by exploiting the properties of the adjoint of a linear operator between Banach spaces. This simple rearrangement of terms results in an expression for $\mathrm{d}J/\mathrm{d}\theta$ which is relatively easy to evaluate in four stages, using only the ingredients that we need to solve the forward problem in the first place.

Let ϕ be some direction in the parameter space P. To express how the operator $\mathrm{d}G/\mathrm{d}\theta$ acts on ϕ in terms of the PDE F, our first step is take the total derivative of equation (3.8) with respect to θ:

$$0 = \left\langle \frac{\mathrm{d}F}{\mathrm{d}\theta}\phi, v \right\rangle_X = \left\langle \left(\frac{\partial F}{\partial u}\frac{\mathrm{d}G}{\mathrm{d}\theta} + \frac{\partial F}{\partial \theta} \right)\phi, v \right\rangle_X \tag{3.9}$$

for all v in X. The last equation implies that, as linear operators,

$$\frac{\mathrm{d}G}{\mathrm{d}\theta} = -\left(\frac{\partial F}{\partial u}\right)^{-1} \frac{\partial F}{\partial \theta}. \tag{3.10}$$

Note that $\partial F/\partial\theta \in \mathscr{L}(P,X^*)$ and $\partial F/\partial u \in \mathscr{L}(X,X^*)$, so the operator composition in the last expression is in the space $\mathscr{L}(P,X)$, as we expect. This is essentially an application of the implicit function theorem in Banach spaces.

Now we can put together our first step, the application of the chain rule in (3.4), with our use of the implicit function theorem in (3.10):

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\theta}, \phi \right\rangle_P = \left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, \frac{\mathrm{d}G}{\mathrm{d}\theta}\phi \right\rangle_X \qquad \text{(chain rule)} \tag{3.11}$$

$$= \left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, -\left(\frac{\partial F}{\partial u}\right)^{-1} \frac{\partial F}{\partial \theta}\phi \right\rangle_X \qquad \text{(implicit fn.)} \tag{3.12}$$

We can simplify this expression even more by using the properties of the adjoint of a linear map.

Given a linear operator $A : U \to V$ from some function space U to another function space V, the adjoint map of A is the linear operator $A^* : V^* \to U^*$ such that

$$\langle f, Au\rangle_V = \langle A^* f, u\rangle_U \tag{3.13}$$

for all $f \in V^*$ and $u \in U$. This generic definition of the adjoint of a linear operator is the infinite-dimensional analogue of taking the transpose of a matrix. In our case, the operator $(\partial F/\partial u)^{-1}$ is a map from X^* to X, so the adjoint of this operator is a map from X^* to X^{**}. Consequently, we can "flip" the action of the operator $(\partial F/\partial u)^{-1}$ over onto the other side of the duality pairing in (3.12):

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, \left(\frac{\partial F}{\partial u}\right)^{-1} \frac{\partial F}{\partial \theta}\phi \right\rangle_X = \left\langle \left(\frac{\partial F}{\partial u}\right)^{-*} \frac{\mathrm{d}E}{\mathrm{d}u}, \frac{\partial F}{\partial \theta}\phi \right\rangle_{X^*} \tag{3.14}$$

A duality pairing in X^* is a little exotic. We can simplify matters by noting that, if X is some Sobolev space W_k^p with $1 < p < \infty$, then X is a reflexive Banach space, i.e. X^{**} is isometrically isomorphic to X. So, we can again flip the duality pairing in the last equation to get

$$\left\langle \left(\frac{\partial F}{\partial u}\right)^{-*} \frac{\mathrm{d}E}{\mathrm{d}u}, \frac{\partial F}{\partial \theta}\phi \right\rangle_{X^*} = \left\langle \frac{\partial F}{\partial \theta}\phi, \left(\frac{\partial F}{\partial u}\right)^{-*} \frac{\mathrm{d}E}{\mathrm{d}u} \right\rangle_X. \tag{3.15}$$

since quantity $(\partial F/\partial u)^{-*}\,\mathrm{d}E/\,\mathrm{d}u$ is actually an element of X.

Let's put all of this together now:

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\theta}, \phi \right\rangle_P = \left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, \frac{\mathrm{d}G}{\mathrm{d}\theta}\phi \right\rangle_X \qquad \text{(chain rule)} \tag{3.16}$$

$$= -\left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, \left(\frac{\partial F}{\partial u}\right)^{-1} \frac{\partial F}{\partial \theta}\phi \right\rangle_X \qquad \text{(implicit fn.)} \tag{3.17}$$

$$= -\left\langle \left(\frac{\partial F}{\partial u}\right)^{-*} \frac{\mathrm{d}E}{\mathrm{d}u}, \frac{\partial F}{\partial \theta}\phi \right\rangle_{X^*} \qquad \text{(adjoint)} \tag{3.18}$$

$$= -\left\langle \frac{\partial F}{\partial \theta}\phi, \left(\frac{\partial F}{\partial u}\right)^{-*} \frac{\mathrm{d}E}{\mathrm{d}u} \right\rangle_X \qquad \text{(reflexive)} \tag{3.19}$$

The justification for each step is listed in the right-most column. Now, note that the quantity

$$\lambda \equiv -\left(\frac{\partial F}{\partial u}\right)^{-*} \frac{\mathrm{d}E}{\mathrm{d}u} \tag{3.20}$$

is an element of the solution space X. In particular, λ is a genuine spatial field, which we can evaluate at a point. This field λ is called the *adjoint state*. The linearization $\partial F/\partial u$ is a partial differential operator, so we can express the operator equation (3.20) defining λ in weak form as

$$\left\langle \frac{\partial F}{\partial u}^{*} \lambda, v \right\rangle_X = -\left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, v \right\rangle_X . \tag{3.21}$$

This weak form is more reflective of how we compute the adjoint state, i.e. through the solution of a PDE. With this definition in hand, we can very succinctly express the derivative of the objective functional as

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\theta}, \phi \right\rangle_P = \left\langle \frac{\partial F}{\partial \theta}\phi, \lambda \right\rangle_X \tag{3.22}$$

where λ is the solution of the PDE (3.21).

To summarize, say we are given a value θ of the parameters and some direction ϕ in parameter space. The action of $\mathrm{d}J/\mathrm{d}\theta$ on ϕ can be computed in four steps:

1. Solve the forward problem: find the value of u such that, for all v in X,

$$\langle F(u, \theta), v \rangle = 0.$$

2. Compute the derivative $\mathrm{d}E/\mathrm{d}u$ of the error metric at u.

3. Solve the adjoint problem: find the value of λ such that, for all v in X,

$$\left\langle \frac{\partial F}{\partial u}^{*} \lambda, v \right\rangle_X = -\left\langle \frac{\mathrm{d}E}{\mathrm{d}u}, v \right\rangle_X .$$

4. The action of $\mathrm{d}J/\mathrm{d}\theta$ is

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\theta}, \phi \right\rangle_P = \left\langle \frac{\partial F}{\partial \theta}\phi, \lambda \right\rangle_X .$$

In order to implement this procedure, we need to be able to differentiate the error metric E; usually the error metrics we use are simple enough that we can do this by hand. We also need to know enough about the PDE that we are solving so that we can linearize it around a particular value of the parameters θ and the putative solution u. Usually we need to be able to linearize the PDE anyway in order to apply Newton's method to solve the resulting nonlinear equations. The derivation of the adjoint method is complex and involves multiple nontrivial results of functional analysis and calculus in Banach spaces. Nonetheless, computing the derivative of the objective does not require much more than what we need to solve the forward problem in the first place.

If θ is very close to the true value ϑ of the parameters, then u should be close to the observations u^o, in which case $\mathrm{d}E/\mathrm{d}u \approx 0$. The derivative of the error metric is the right-hand side in the adjoint equation for λ, so assuming that the inverse of the adjoint operator is a bounded linear operator,

$$\|\lambda\|_X \leq \|(\partial F/\partial u)^{-1}\|_{X^* \to X} \|\,\mathrm{d}E/\,\mathrm{d}u\|_{X^*}. \tag{3.23}$$

In our case, $\partial F/\partial u$ is an elliptic partial differential operator, for which we can derive analytical bounds on the operator norm of $(\partial F/\partial u)^{-1}$ in terms of the coefficients (Trèves, 1975). When $\mathrm{d}E/\mathrm{d}u$ approaches 0, so too does the adjoint state and thus $\mathrm{d}J/\mathrm{d}\theta$ by equation (3.22).

To illustrate all of these concepts, we will use a simple model problem: determining the diffusivity coefficient k in the generalized Poisson problem

$$-\nabla \cdot k\nabla u = f \tag{3.24}$$

over some domain Ω, with Dirichlet boundary conditions $u|_{\partial\Omega} = 0$. If f is an element of the dual space H^{-1} to the Sobolev space H_0^1, $\partial\Omega$ is a Lipschitz manifold and k is bounded above and below, then a unique solution u exists in the Sobolev space H_0^1 (Trèves, 1975). In order to enforce the constraint that k is strictly positive, we will instead choose to parameterize it as

$$k = k_0 e^{\theta} \tag{3.25}$$

in terms of some dimensionless field θ, where k_0 is a constant with the right physical units for the problem.

In deriving the adjoint method, we assumed that the forward problem was posed in weak form, as in equation (3.7). The weak form for the Poisson equation is obtained by multiplying (3.24) by some arbitrary test function v in the Sobolev space H_0^1, to arrive at the condition

$$\int_\Omega k_0 e^{\theta} \nabla u \cdot \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x. \tag{3.26}$$

If the equation (3.26) holds for all v in H_0^1, then u is said to solve the Poisson equation weakly. The advantages of writing the PDE in this weak form is that the differentiability requirements for the right-hand side f, the diffusion coefficient k,

and the solution u are substantially relaxed over the strong form of the differential equation.

We will take as our error functional the integrated square misfit:

$$E(u-u^o) = \frac{1}{2}\int_\Omega (u-u^o)^2 \,\mathrm{d}x. \tag{3.27}$$

Using this error functional implies that we have dense enough measurements of u^o to interpolate it to some well-resolved field. This is not always the case, and we will discuss what happens when this assumption fails later. The advantage of such a simple error metric is that its derivative is particularly easy to compute:

$$\left\langle \frac{\partial E}{\partial u}, v \right\rangle = \int_\Omega (u-u^o)\cdot v \,\mathrm{d}x. \tag{3.28}$$

We now compute the adjoint state by solving the PDE for $(\partial F/\partial u)^*$. In our case, F is linear in u, and moreover it is a self-adjoint operator. Consequently, the adjoint PDE we need to solve is identical to the forward PDE for u, only it has a different right-hand side:

$$\int_\Omega k_0 e^\theta \nabla\lambda\cdot\nabla v \,\mathrm{d}x = \int_\Omega (u-u^o)\cdot v \,\mathrm{d}x. \tag{3.29}$$

The second-to-last step is to compute the derivative of the forward problem with respect to the parameters along some direction ϕ in parameter space. The forward problem is particularly easy to differentiate with respect to the parameters θ; since the PDE depends only on e^θ, and the derivative of the exponential function is itself, we get that

$$\left\langle \frac{\partial F}{\partial \theta}\phi, w \right\rangle_X = \int_\Omega \phi\cdot k_0 e^\theta \nabla u\cdot\nabla w \,\mathrm{d}x. \tag{3.30}$$

Finally, we can compute how the derivative of the objective functional acts on a search direction ϕ as

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\theta}, \phi \right\rangle_P = \left\langle \frac{\partial F}{\partial \theta}\phi, \lambda \right\rangle_X = \int_\Omega \phi\cdot k_0 e^\theta \nabla u\cdot\nabla\lambda \,\mathrm{d}x. \tag{3.31}$$

The hard part of this computation is obtaining u and λ by solving the Poisson equation; evaluating the derivative in several different search directions $\phi_1,\ldots,\phi_m$, on the other hand, is computationally cheap.

Some extensions to this problem are noteworthy for the discussion that follows. If we used some more generic parameterization $k = k(\theta)$, where $k:\mathbb{R}\to\mathbb{R}^+$ is a continuously differentiable function, then

$$\left\langle \frac{\partial F}{\partial \theta}\phi, w \right\rangle_X = \int_\Omega \phi\cdot \frac{\mathrm{d}k}{\mathrm{d}\theta}\nabla u\cdot\nabla w \,\mathrm{d}x. \tag{3.32}$$

The parameterization in terms of the log-conductivity $\log(k/k_0)$ is particularly advantageous because its derivative is so simple.

For the simple model, we used the square L^2-norm as our error metric, which is not always realistic. If we have some estimates σ of the error standard deviations but we still have fairly dense measurements, we might instead use

$$E(u-u^o) = \int_\Omega \frac{(u-u^o)^2}{2\sigma^2}\,\mathrm{d}x, \tag{3.33}$$

especially if the measurement errors tend to be spatially variable. Including the errors explicitly has the added advantage of making the error metric dimensionless. In all cases, we only have a finite set of measurements, and one could argue that spatially interpolating these measurements to form some field introduces a degree of arbitrariness in how the data are interpolated. A more sensible choice of error metric in this case is the sum

$$E(u-u^o) = \sum_k \frac{(u(x_k)-u^o(x_k))^2}{2\sigma_k^2}. \tag{3.34}$$

In this case, the derivative of E is a sum of δ-functions. The measurement process may not in fact directly sample the true u field at some measurement points x_k, but rather collect some convolution with a smoothing kernel in a neighborhood of x_k. Let $\mathring{u}$ be the true field, from which the measurements $u^o(x_k)$ are obtained by convolution with q:

$$u^o(x_k) = q * \mathring{u}(x_k) = \int q(x_k - x)\mathring{u}(x)\,\mathrm{d}x + \text{error}. \tag{3.35}$$

Then arguably an even better choice of error metric is

$$E(u-u^o) = \sum_k \frac{((q*u)(x_k)-u^o(x_k))^2}{2\sigma_k^2}. \tag{3.36}$$

Finally, the use of the squared error tends to place too much emphasis on outliers. An error metric that gives solutions which are more robust to outliers in the data is the L^1-norm misfit:

$$E(u-u^o) = \sum_k \frac{|(q*u)(x_k)-u^o(x_k)|}{\sqrt{2}\sigma_k}. \tag{3.37}$$

In short, the selection of an appropriate error metric is far from obvious. The statistical distribution of the measurement errors can serve as a valuable guide in determining how the error should be quantified in data assimilation problems.

Data assimilation for realistic glacier models is more complicated than inferring the log-conductivity in the Poisson equation because the forward model is no longer linear in the observed field u. The chief mathematical tool we need is the ability to compute the linearization $\partial F/\partial u$ of the forward model. We need the linearization of the forward model to be able to solve the adjoint equation (3.21) for λ. Luckily, we already needed the linearization to solve the forward model equations for u in

the first place; we computed this in section §2.4. The only remaining piece is to compute the derivative of the PDE with respect to the parameters. We can show this easily for the shallow shelf equations, for which the weak form is

$$\langle F(u,\theta), v\rangle = \int_\Omega \left\{ B(\theta) h |\dot\varepsilon(u)|^{\frac{1}{n}-1} \left(\dot\varepsilon_{ij}(u)\dot\varepsilon_{ij}(v) + \dot\varepsilon_{kk}(u)\dot\varepsilon_{kk}(v)\right) - \frac{1}{2}\varrho g h^2 \frac{\partial v_i}{\partial x_i} \right\} \mathrm{d}x \quad (3.38)$$

and the condition for u to be a weak solution is that $\langle F(u,\theta), v\rangle = 0$ for all v in the solution space X. By taking some finite increment δ along some direction ϕ in the parameter space P, we can show that the partial derivative of the weak form with respect to θ acts on ϕ by

$$\left\langle \frac{\partial F}{\partial \theta}\phi, v \right\rangle = \int_\Omega \frac{\mathrm{d}B}{\mathrm{d}\theta} h |\dot\varepsilon(u)|^{\frac{1}{n}-1} \left(\dot\varepsilon_{ij}(u)\dot\varepsilon_{ij}(v) + \dot\varepsilon_{kk}(u)\dot\varepsilon_{kk}(v)\right) \cdot \phi \, \mathrm{d}x. \quad (3.39)$$

We can use the same approach with the weak form of any other glacier flow model and with respect to other parameters, such as the basal friction coefficient.

In section §2.4, we showed that the linearization of the forward model could be approximated by freezing the coefficients of the PDE around the basic state u. We then derived, with no small amount of algebra, what the true linearization of the forward model is. This begs the question of whether the frozen-coefficient approximation is acceptable or not. Martin and Monnier (2014) compared how the two approaches perform on a synthetic inverse problem with the full Stokes model. They found that features with a wavelength of 20 ice thicknesses could be recovered from surface data when using the frozen-coefficient approximation, but that keeping all terms in the linearization improved the resolution to 10 ice thicknesses. This is a substantial improvement. Joughin et al. (2004b) report that the convergence rate for inverse methods with perfectly plastic basal deformation ($m = \infty$) is much worse than for linear-viscous sliding ($m = 1$); since the inferred basal shear stress is the same in both cases, they use linear viscous sliding purely for inferring the shear stress. However, the poorer convergence they find in the perfectly plastic case may be an artifact of using the frozen-coefficient approximation, since the departure of the full linearization from this approximation is most severe in the perfectly plastic case. As we shall review in §4, the engineering difficulty of adding the corrections for the full linearization in a glacier modelling code is minor.

Arthern and Gudmundsson (2010) have devised an alternative to the adjoint method that they call the Robin method. In the adjoint method, we use a candidate parameter field, compute a solution of the PDE, and compare this solution with the observed data. The Robin method leverages the fact that the surface stresses are also zero. Given a candidate parameter field θ, one could also solve the Stokes equations using Dirichlet boundary conditions at the surface, requiring the velocities to exactly match observations. The surface stresses might be non-zero, and to find the right value of θ we optimize the surface stresses down to 0. We can combine these approaches by computing both solutions, which we will refer to as u^N and u^D for the Neumann and Dirichlet solutions respectively. The objective functional for this

problem is

$$J = \int_{\Gamma_s} (u_i^N - u_i^D)(\tau_{ij}^N - \tau_{ij}^D) n_j \, \mathrm{d}\Gamma, \tag{3.40}$$

where Γ_s is the upper surface of the glacier. We can simplify this expression by noting that u_i^D must be equal to the observed velocity u_i^o, and that $\tau_{ij}^N n_j = 0$, but the form above is the most useful for analyzing computing the gradient of the objective functional. We can show that

$$J = \int_\Omega 2\mu |\dot{\varepsilon}(u^N) - \dot{\varepsilon}(u^D)|^2 \, \mathrm{d}x + \int_{\Gamma_b} \beta |u^N - u^D|^2 \, \mathrm{d}\Gamma. \tag{3.41}$$

using the divergence theorem (Arthern and Gudmundsson, 2010). This equivalent form demonstrates that equation (3.40) is positive-definite. Moreover, the functional

$$E(u) = \int_\Omega 2\mu |\dot{\varepsilon}(u)|^2 \, \mathrm{d}x + \int_{\Gamma_b} \beta |u|^2 \, \mathrm{d}\Gamma \tag{3.42}$$

is the rate of energy dissipation due to internal strain heating (first term) and frictional heating through contact with the bed (second term). Again using the divergence theorem, the derivative of the objective functional is

$$\left\langle \frac{\partial J}{\partial \mu}, \nu \right\rangle = 2 \int_\Omega \nu \left(|\dot{\varepsilon}(u^N)|^2 - |\dot{\varepsilon}(u^D)|^2 \right) \mathrm{d}x \tag{3.43}$$

$$\left\langle \frac{\partial J}{\partial \beta}, \gamma \right\rangle = \int_{\Gamma_b} \gamma (|u^N|^2 - |u^D|^2) \, \mathrm{d}\Gamma \tag{3.44}$$

provided we ignore the nonlinear dependence of the material coefficients on the ice velocity.

The Robin method has several advantages and disadvantages relative to the adjoint method. First, the norm in the Robin method is arguably more physically meaningful, but the method is much less flexibile than using an adjoint formulation. The Robin method minimizes the mismatch with observations in a stronger norm than the adjoint method, since it includes derivatives of the computed velocities as well. The norm used to measure the misfit, being related to the thermal energy dissipation rate, has a closer resemblance to the underlying physics than the L^2-norm misfit. On the other hand, there are many more choices one can make for the objective functional in the adjoint method; for example, weighting by the error standard deviations, using a logarithmic penalty, using the L^1-norm misfit, and so forth. With the adjoint method, one can use a finite sum over the observation points to define the error metric; or, if the measurements are dense enough, one can interpolate them to a continuous spatial field and define the error metric through an integral. With the Robin method, the measurements must be dense enough to interpolate to a continuous spatial field.

Second, each method has different computational demands. The Robin method only requires solving the forward problem, albeit with different boundary conditions. The adjoint method requires a routine to compute the linearization of the

forward model around the computed velocity field in order to solve the adjoint problem. While computing the linearization is not difficult in principle, if one is using off-the-shelf software tools such as COMSOL or Elmer, the software library in question may not have implemented public, user-facing routines for computing this operator. The Robin method does not require the ability to compute the linearization of the forward model, but this comes at the expense of having to perform two forward model solves instead of one, as in the adjoint method.

In any case, Gillet-Chaulet et al. (2012) compared both methods on real data for the Greenland Ice Sheet, finding little substantive difference in how well the two approaches could fit the observations. The differences between the two methods come more from ease of implementation than from accuracy.

3.3 Regularization

The method outlined in the last section for inferring parameters of an elliptic PDE works only when the data are known with complete accuracy. In the face of noisy observed data, a direct implementation of the adjoint method will tend to produce an inferred field polluted with spurious oscillations. These spurious oscillations are due to a phenomenon called *overfitting*, which occurs because the inverse problem is extremely ill-conditioned. Inverse coefficient problems for elliptic partial differential equations can be ill-conditioned because the map from the coefficients to the observed data essentially acts like a low-pass filter. Loosely speaking, since high-wavenumber modes are damped in the forward map, they are amplified in the inverse map. The observed data tend to have some amplitude in high-wavenumber modes due to the fact that the measurement errors are often spatially white. These errors then get amplified far beyond the amplitude of the true signal. The map from parameters to observations is nonlinear, and we have not proven that it really does act like a low-pass filter. In the following, I will describe overfitting in more detail and why it occurs, proceeding from finite-dimensional least-squares problems up to nonlinear PDE inverse problems.

The simplest problem that exhibits overfitting is linear least-squares, i.e. given some $m \times n$ matrix A and some data d, find the vector Θ such that

$$J(u) = \frac{1}{2}\|A\Theta - d\|^2 \tag{3.45}$$

is minimized. When discretized and linearized, the PDE inverse problems we will consider become finite-dimensional linear least squares problems like that of equation (3.45). Let U, Σ, V be the singular value decomposition of A, i.e. $A = U\Sigma V^*$, where U and V have orthonormal columns, and Σ is positive and diagonal. The solution of the least-squares problem can be written as

$$\Theta = \sum_k \sigma_k^{-1}(u_k^* d)v_k \tag{3.46}$$

where u_k, v_k are the columns of U and V respectively, and σ_k are the diagonal entries of Σ. For discretizations of elliptic PDE inverse problems, the singular values

decay to 0 and the eigenvectors v_k represent increasingly oscillatory spatial modes as k grows larger. The data d are the sum of a true signal f and a noise component ξ. The problem comes from the fact that, for spatially uncorrelated measurement errors, the component $u_k^*\xi$ of the measurement noise along the k-th left singular vector does not decay as k grows large. Since the weight σ_k^{-1} of the k-th mode keeps growing, eventually the amplitude of the noise in the solution swamps the amplitude of the true signal.

Real inverse problems are often posed in infinite dimensions, in terms of integral operators rather than matrices. Suppose that the forward model physics are described by a Fredholm integral operator

$$\mathscr{G}\theta = \int_\Omega G(x,y)\theta(y)\,\mathrm{d}y \tag{3.47}$$

for some *kernel function* G. The infinite-dimensional problem can be approximated by a finite-dimensional one by some appropriate discretization. Suppose that we approximate θ as belonging to the span of a set of basis functions $\{\varphi_1,\ldots,\varphi_n\}$. Let P be the projection onto the span of these basis functions, and Θ the vector of expansion coefficients for the true solution. Then the discretized inverse problem is a linear least-squares problem for the matrix

$$A = P^*\mathscr{G}P, \tag{3.48}$$

obtained by conjugating with the projection operator P. Moreover, if $\mathscr{G}$ is an integral operator with a reasonable kernel, then it has an infinite-dimensional version of the singular value decomposition, with singular values decaying to 0 (see Courant and Hilbert (1965), §3.10.10). The finite-dimensional projection A then shares these properties.

The last step is to connect the problems we are interested in – inverse coefficient problems or PDEs – to Fredholm integral operators. While the map from the right-hand side f of a PDE to the solution u is linear, the map from the coefficients θ to u is not linear, and moreover has no closed-form expression. The linearization of this map, however, is a Fredholm integral operator, and its kernel can be expressed in terms of the Green's function for the PDE. For the sake of concreteness, we will consider the PDE

$$\mathscr{L}(\beta)u = (-\nabla^2 + \beta)u = f \tag{3.49}$$

with Dirichlet boundary conditions $u|_{\partial\Omega} = 0$. This scalar linear PDE roughly mimics, in simplified form, the inverse problem for the basal friction coefficient under an ice stream. We will refer to β as a friction coefficient in the following. In practice, one needs to guarantee positivity; we could then write $\beta = \beta_0 e^\theta$, but this will be unnecessary for the explanation that follows. The parameters β live in L^∞, and the map $\mathscr{L}$ takes L^∞ to the space of bounded bilinear forms on the Sobolev space H_0^1.

First, we want to show that the map from the coefficient β to the field u for a fixed right-hand side is differentiable. A function u is a weak solution of $\mathscr{L}(\beta)u = f$ if, for all v in H_0^1,

$$\langle \mathscr{L}(\beta)u, v\rangle = \langle f, v\rangle. \tag{3.50}$$

If the conductivity β is bounded below by some positive constant, i.e. the operator is elliptic, then the PDE has a unique weak solution. The norm of the solution can be bounded by the norms of the reciprocal $1/\beta$ and the right-hand side:

$$\|u\|_{H^1} \leq c(1 + \|\beta^{-1}\|_{L^\infty})^{-1}\|f\|_{H^{-1}} \tag{3.51}$$

where c is the constant in the Poincaré inequality (Trèves, 1975). Since there is a unique weak solution, we can say that the inverse operator $\mathscr{L}(\beta)^{-1}$ exists and is bounded as a map from $H^{-1} \to H_0^1$. We can get a little more abstract and think not of a single coefficient β but of the mapping $\beta \mapsto \mathscr{L}(\beta)$ from all coefficient fields to the space of bounded linear operators from H_0^1 to H^{-1}. An elementary manipulation shows that

$$\frac{\mathrm{d}\mathscr{L}}{\mathrm{d}\beta}(\beta_0)\beta_1 = \beta_1, \tag{3.52}$$

i.e. the directional derivative of the map $\mathscr{L}$ around β_0 in the direction β_1 is the operator that multiplies a function by β_1. From the inequality

$$\langle \mathscr{L}(\beta)u, v\rangle \leq (1 + \|\beta\|_{L^\infty})\|u\|_{H^1}\|v\|_{H^1}, \tag{3.53}$$

we find that the operator norm of $\mathscr{L}(\beta)$ is also bounded by

$$\|\mathscr{L}(\beta)\| \leq 1 + \|\beta\|_{L^\infty}. \tag{3.54}$$

This inequality is true whether or not β is positive. If we restrict ourselves to the open, convex subset of L^∞ consisting of the set of all β such that $1/\beta$ is in L^∞ and $\|1/\beta\|_{L^\infty} < 1/\kappa$ for some constant κ, then equation (3.51) implies that

$$\|\mathscr{L}(\beta)^{-1}\| \leq c(1 + \|\beta^{-1}\|_{L^\infty})^{-1}. \tag{3.55}$$

Now suppose that we have some conductivity β_0 such that $\|\beta_0^{-1}\| > \kappa^{-1}$; we want to examine perturbations of the map $\beta \mapsto \mathscr{L}(\beta)^{-1}$ in a neighborhood of β_0. We can use the geometric series formula to write

$$\mathscr{L}(\beta_0 + \epsilon\beta_1)^{-1} = \left(\sum_{n=0}^{\infty}(-1)^n\epsilon^n\left(\mathscr{L}(\beta_0)^{-1}\frac{\mathrm{d}\mathscr{L}}{\mathrm{d}\beta}(\beta_0)\beta_1\right)^n\right)\mathscr{L}(\beta_0)^{-1} \tag{3.56}$$

$$= \left(\sum_{n=0}^{\infty}(-1)^n\epsilon^n(\mathscr{L}(\beta_0)^{-1}\beta_1)^n\right)\mathscr{L}(\beta_0)^{-1}. \tag{3.57}$$

Applying the bounds from equations (3.54), (3.55) to β_1 and β_0 respectively, the operator $\mathscr{L}(\beta_0)^{-1}\beta_1$ is bounded from $H_0^1 \to H_0^1$ and

$$\|\mathscr{L}(\beta_0)^{-1}\beta_1\| \leq c(1 + \|\beta_0^{-1}\|_{L^\infty})\|\beta_1\|_{L^\infty}, \tag{3.58}$$

which is an upper bound for the radius of convergence of the series in equation (3.57). The above considerations show that, when restricted to an appropriate subset of the Banach space L^∞, the map from the coefficient β to the operator $\mathscr{L}(\beta)^{-1}$ is differentiable. The derivative of this map at β_0 is the operator

$$\frac{\mathrm{d}\mathscr{L}^{-1}}{\mathrm{d}\beta}\Big|_{\beta_0}\beta_1 = -\mathscr{L}(\beta_0)^{-1}\beta_1\mathscr{L}(\beta_0)^{-1}, \tag{3.59}$$

which is linear in β_1.

So, we have shown that $\beta \mapsto \mathscr{L}(\beta)^{-1}$ is a differentiable map between Banach spaces, but it remains to show from (3.59) that the linearization of the map around some particular coefficient β_0 is an integral operator. Let G_0 be the Green's function for the operator $\mathscr{L}(\beta_0)$, i.e.

$$\mathscr{L}(\beta_0)^{-1} f = \int_\Omega G_0(x,y) f(y)\,\mathrm{d}y. \tag{3.60}$$

If we have some specific right-hand side f, let $u_0 = \mathscr{L}(\beta_0)^{-1} f$; then from (3.59), we find that

$$\left(\frac{\mathrm{d}\mathscr{L}^{-1}}{\mathrm{d}\beta}(\beta_0)\beta_1\right) f = \int_\Omega G_0(x,y) u_0(y) \beta_1(y)\,\mathrm{d}y. \tag{3.61}$$

From this last equation, we deduce that $\mathrm{d}\mathscr{L}^{-1}/\mathrm{d}\beta$ is a Fredholm integral operator acting on the perturbation β_1, and its kernel is $G_0(x,y)u_0(y)$. The kernel is not symmetric, but the important fact is that the singular values decay to 0. For other PDEs, we would have to consider the form of $\mathrm{d}\mathscr{L}/\mathrm{d}\beta$ to decide whether the kernel of the parameter-to-observation map is sufficiently well-behaved. In the concrete example we considered, the kernel is no more singular than the fundamental solution of the Poisson equation, but for more complicated problems or parameterizations, more care might be necessary.

In summary, finite-dimensional least-squares problems are susceptible to overfitting when the singular values decay to 0 but the component of the data along high-wavenumber modes does not decrease. The discretization of an inverse problem where the forward map is a linear integral operator gives a matrix with rapidly-decaying singular values. Finally, the linearization of a nonlinear inverse problem gives a Fredholm integral operator.

Two factors cause overfitting: the singular values of the forward map decay to 0 rapidly, and the data are noisy. Both of these conditions are necessary for overfitting to be a risk in solving the inverse problem. By way of contrast, consider the advection equation

$$\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) q = 0 \tag{3.62}$$

over the whole real line, subject to the initial condition $q|_{t=0} = q_0$. The solution of this PDE at time t is

$$q(x,t) = q_0(x - ct). \tag{3.63}$$

The map from q_0 to the solution q at time t is unitary in the Hilbert space L^2, so the singular values of the forward map are all equal to 1. With no contrast between the large and small singular values, the inverse problem of determining the initial state from measurements made at some final time does not exhibit unstable dependence on noisy modes as in the case where the forward map is a smoothing operator. The same behavior can be found in more complex problems such as the 3D wave equation, Maxwell's equations, the Schrödinger equation, the Dirac equation, etc., for which the solution operator is a unitary semigroup in some Hilbert space.

Since we cannot solve an ill-conditioned problem reliably, we might instead choose to solve a different, well-conditioned problem. The hope is that the solution of this well-conditioned problem will be less sensitive to noise in the input data than the original problem. This procedure is called *regularization*. Regularization can be justified heuristically on an appeal to Occam's razor. The numerical problems we aim to solve are practically always nonlinear, so often our only choice is to use some kind of iterative method. These iterative methods must be given some stopping criterion ϵ such that, when we find some parameters q for which the residual is less than ϵ, we can consider this solution to be good enough. The inference problems we are trying to solve are so ill-conditioned that there are many possible values of the solution q for which the residual is less than any reasonable convergence tolerance. The solution of the inference problem is effectively no longer unique, again as a consequence of the fact that the forward map is insensitive to short-scale oscillations. Among the many candidate solutions for which the residual is sufficiently small, Occam's razor suggests that we pick the "simplest" one. Since most of the spurious oscillations we find in unregularized inverse problems are due to excessive sensitivity of the inverse problem to noise in the data, the hope is that, by picking the simplest of the many candidate solutions, we obtain the estimate which is as independent as possible of the random sample of errors that we happened to draw when the measurements were made in the first place. If we can come up with some appropriate definition of simplicity, we can optimize both for model-data misfit and simplicity. This defines a new objective, the minimization of which is hopefully better conditioned than optimizing solely for model-data misfit.

This heuristic can be put into practice by choosing some functional R which quantifies the degree to which the inferred field θ fails to be simple, for some definition of simplicity. This extra functional R is then added to the objective:

$$J(\theta, \alpha) = E(G(\theta) - u^o) + \alpha^p R(\theta). \tag{3.64}$$

The regularization parameter α dictates how much the problem is regularized, i.e. how much we optimize for model-data misfit versus simplicity, and the exponent p is chosen to make the objective functional have the correct dimensions. The most common regularization functional in the literature on inverse problems is the *Dirichlet energy*

$$R(\theta) = \frac{1}{2|\Omega|} \int_\Omega |\nabla \theta|^2 \, \mathrm{d}x. \tag{3.65}$$

This choice of regularization functional equates simplicity with spatial smoothness. The Dirichlet energy has units of $[\theta]^2 \cdot [x]^{-2}$, so to get the units right we take $p = 2$. In this case, the regularization parameter can be interpreted as a smoothing length in a spatial filter.

There are many other possible choices of R which could also effectively filter out spurious oscillations. For example, one might instead optimize for an inferred field with small curvature by penalizing the square of the Laplacian:

$$R(\theta) = \frac{1}{2|\Omega|} \int_\Omega |\Delta \theta|^2 \, \mathrm{d}x. \tag{3.66}$$

The dimensions of this functional are $[\theta]^2 \cdot [x]^{-4}$, so to make the units match up we choose $p = 4$ in (3.64). We could also have chosen R to be an anisotropic functional if we had reason to believe that the solution is more spatially smooth in one direction than others. For example, we might argue that the basal friction coefficient is smoother along the flow direction u than across-flow, in which case we might choose a regularization functional

$$R(\theta;\gamma) = \frac{1}{2|\Omega|}\int_\Omega \nabla\theta^* \left((1-\gamma)I + \gamma\frac{uu^*}{|u|^2}\right)\nabla\theta \,\mathrm{d}x \tag{3.67}$$

where I is the identity matrix and γ is a parameter between 0 and 1 dictating the degree of anisotropy. When $\gamma = 1$, this functional is identical to the Dirichlet energy; when $\gamma = 1$, smoothing is only applied along-flow.

This discussion begs the question of how we should choose one regularization functional over another – what is the appropriate definition of "simplicity" in the sense of Occam's razor? In much of the literature on PDE inverse problems, this decision is made based on some combination of heuristics and what everyone else in the field does. Many PDE inverse problems, however, can be viewed as exactly analogous to estimating a set of unknown parameters of a probability distribution using Bayesian statistical inference. The choice of a regularization functional is equivalent to the selection of a prior distribution for the parameters. In Bayesian inference, the choice of prior distribution is guided by maximum entropy principles. This topic is the subject of §5.

Having chosen how to regularize the inference problem, we then have to decide how much to regularize it. This amounts to choosing a value of the parameter α. There are two extremes: when $\alpha = 0$, no regularization is applied at all and the solution may be overfit; when α is large, optimality is determined more by smoothness and the model-data misfit may be poor. What we need is some procedure which picks the best compromise between model-data misfit and solution complexity.

One heuristic to select the regularization parameter is the *L-curve* (Hansen, 1999). Given an inverse problem posed as the minimization of a regularized objective J as in equation (3.64), let θ_α be the solution obtained with regularization parameter α. The L-curve is a parametric plot of the model-data misfit $E(G(\theta_\alpha)-u^o)$ on one axis and the penalty $R(\theta_\alpha)$ on the other axis, parameterized over α from 0 to some very large value. Hansen and O'Leary observed that for many linear inverse problems, this curve has a characteristic shape. Starting from $\alpha = 0$, the smoothness penalty $R(\theta_\alpha)$ decreases monotonically with almost no change in the model-data misfit. In this regime, the entire curve segment of progressively smoother solutions all fit the data almost equally well, because the forward map is largely insensitive to the oscillations which the penalty functional is filtering out in this α regime. At a certain critical value of the regularization parameter, the curve turns sharply at an almost 90° angle as the solutions become very smooth but no longer fit the data. The critical value can be identified as the value α^* for which the smoothness/misfit curve has the greatest curvature in the sense of differential geometry. This value α^* is heuristically justified as the best compromise between model-data misfit and smoothness, because it represents the point at which no more smoothing can be

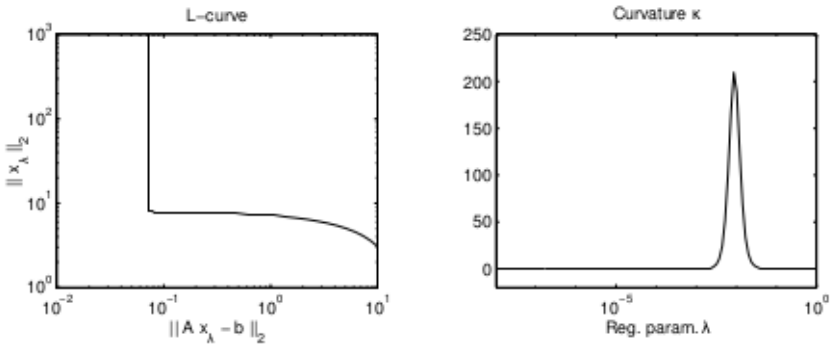


Figure 3.1: Left: the L-curve for regularized least-squares estimation, with the model-data misfit on the x-axis and the penalty on the y-axis. Right: the curvature of the L-curve as a function of the regularization parameter λ. Reproduced from (Hansen, 1999).

applied without impairing the fit to the data. The curve of model-data misfit versus smoothness as a function of the regularization parameter, with its characteristic bend at α^*, is often shaped like the letter "L", hence the name "L-curve" for this method of selecting the regularization parameter.

In most studies of regularized inverse problems, the regularization parameter is a scalar λ, the units of which are swept under the rug. I have elected to always use a regularization parameter α which has units of length, and weight the penalty functional by α^p in order to make sure the units are correct; the translation between the two is $\lambda = \alpha^p$. Fortunately, the differential-geometric curvature κ is invariant to reparameterizations of the underlying space curve, so using my dimensional parameterization and the more common parameterization in the literature will give equivalent values if optimized using the L-curve. In §3.5, I will describe my work on inferring the basal shear stress of three Greenland outlet glaciers. The computations for this work were performed using the glacier modelling software Elmer/Ice, which uses the more common parameterization in terms of λ. While writing this paper, I wanted to compare my work to that of other researchers, but was unable to ascertain anything meaningful from the values of the regularization parameter that they reported without substantial guesswork. Reporting instead a smoothing length α together with the physical units used will make it easier for researchers on inverse problems to compare their work, even when different software tools were used across studies.

Minimizing the objective functional is typically implemented using iterative optimization algorithms such as steepest descent, nonlinear conjugate gradients, or BFGS. An entirely different approach to preventing overfitting is to truncate the iterative method before full convergence (Habermann et al., 2012). This idea is based on the observation that the low-wavenumber components of the solution typically

converge first, with spurious oscillations only accruing at later iterations. By starting with a spatially smooth initial guess for the parameters, we can get a smooth solution by stopping the iteration after the low-order modes have converged but before the method overfits the data. We then need some criterion for when to stop the iteration in such a way that the solution is determined only by the true signal in the data and not the noise. One way is to use a recent improvement threshold; given some tolerance ϵ, the iteration is stopped if the error improves by less than ϵ from iteration k to iteration $k+\delta k$ for some fixed lag δk. Joughin et al. (2004b) used a recent improvement threshold to decide convergence in estimating the basal shear stress of all of the Ross Ice Streams. Alternatively, we can apply the discrepancy principle to truncated iterative methods in the same way as for Tikhonov regularization. If we have some a priori information on the magnitude $\|\xi\|$ of the measurement noise, we can stop the iteration when the model-data misfit is less than $\|\xi\|$. Habermann et al. (2012) argue that, when a priori information about the noise amplitude is available, the discrepancy principle should be preferred over a recent improvement threshold on the grounds that the solution will depend on whether steepest descent or nonlinear CG was used. Truncated iterative methods come with their own set of advantages and drawbacks over Tikhonov regularization. For one, we no longer need to make a possibly arbitrary choice of the regularization functional, relying instead on the smoothing properties of the forward map itself. On the other hand, when using rapidly converging algorithms such as BFGS, the error can blow right past the stopping threshold dictated by the discrepancy principle in a single iterate, resulting in an overfit solution (Habermann et al., 2012).

3.4 Data assimilation since MacAyeal

Since the initial work on glaciological data assimilation by MacAyeal, this technique has been used in several studies both to explore underconstrained aspects of glacier physics and to compute a sound initial state from which to make ice sheet forecasts. While MacAyeal's early work on data assimilation focused on ascertaining the shear stress underneath grounded ice streams, the scope of the technique was soon expanded to estimating the rheology of floating ice shelves (Rommelaere and Macayeal, 1997). More recent work has also the same methods to improve estimates of bedrock topography based on measurements of surface mass balance (Morlighem et al., 2011).

MacAyeal again examined the basal shear stress of Ice Stream E in MacAyeal et al. (1995). This paper was the first to apply ideas from statistical inference to formulate a criterion for when an inferred parameter field gives an adequate fit to the data. Since the measurement errors are normally distributed, MacAyeal argues that the model-data misfit should have a χ^2 distribution. Since there were three observed fields used (thickness, surface elevation, and velocity), MacAyeal argues that the number of degrees of freedom for the χ^2 distribution should be $\nu = 3$. This accounting ignores the fact that each of these observed fields is really just a finite collection of N point measurements, in which case the χ^2 distribution should have N degrees of freedom (Parker, 1994).

Subsequent use of inverse methods in glaciology depended on large-scale mapping of ice sheet velocity and elevation. More remote sensing data became available in the mid- to late-1990s with the launch of the European Remote Sensing (ERS-1/2) satellites in 1991 and 1995. Additionally, the RADARSAT-1 satellite was repositioned to gather data over the Antarctic Ice Sheet for 30 days in 1999. Joughin et al. (2002) mapped the velocities of all of the Ross ice streams using a combination of radar interferometry from RADARSAT-1 and feature-tracking from Landsat. Around the same time, Lythe and Vaughan (2001) collected data from hundreds of sources, including seismic sounding and both air- and ground-borne ice penetrating radar, to create a map of the bed elevation under large parts of the Antarctic Ice Sheet (BEDMAP). Shepherd et al. (2001) mapped the velocity and surface elevation change of Pine Island Glacier (PIG) in West Antarctica with ERS-1/2 measurements from 1992-1999, finding average thinning rates of 0.75 m/yr over the main trunk of the glacier, which extends inland by ∼80 km. The observed thinning rates are substantially greater than the interannual variability in accumulation at PIG, which is typically less than 0.1 m/yr; changes of this magnitude could only occur from glacier dynamics. These kinds of rapid changes can occur at PIG because of its close proximity to the continental shelf, where intrusions by warm, deep ocean waters melt the ice shelf from below (Jenkins et al., 2010). This situation is in distinct contrast to the Ross Ice Shelf and its tributaries, which are situated much farther back from the continental shelf. Pine Island and nearby Thwaites glacier are especially interesting because their bedrock configuration makes them possibly susceptible to unstable retreat through the marine ice sheet instability (Joughin et al., 2014b).

Joughin et al. (2004b) continued MacAyeal's early work to infer the basal shear stress of all of the Ross ice streams. MacAyeal (1992) demonstrated the presence of sticky spots under Ice Stream E; the results in Joughin et al. (2004b) showed that these features are present throughout all of the Ross Ice Streams. Additionally, they made several estimates of the basal shear stress under different assumptions about the rheology of basal sliding by changing the sliding exponent m. They found that, although the friction coefficient differs, the inferred basal shear stress is essentially the same when $m = 1$ and when $m = \infty$. While the exponent m cannot be determined from a single snapshot inversion, they go on to suggest that using a time series of inferred shear stress fields could constrain the sliding rheology.

Vieli and Payne (2003) used the surface velocity and elevation data from Shepherd et al. (2001) to infer the shear stress under Pine Island Glacier. At the time of their work, only two flight lines with ice-penetrating radar had been flown to map the bed elevation along the ice flow direction. For lack of sufficient measurements of the bed elevation under Pine Island Glacier, Vieli and Payne were only able to map the basal shear stress along these flight lines instead of throughout the entire region. They found that most of the driving stress was supported by basal resistance, rather than by drag at the side walls as in the case of the Ross ice streams.

During the austral summer of 2004-2005, additional airborne radio echo sounding measurements of the Pine Island and Thwaites glacier basins were conducted, giving much more accurate bed DEMs than were available from BEDMAP (Vaughan et al., 2006; Holt et al., 2006). With these new observations, Joughin et al. (2009) extended the work of Vieli and Payne (2003) by inferring the basal shear stress

under both glaciers using a full 2D plan-view model instead of a flowline model. They find mostly a strong bed underneath PIG near the grounding line, confirming the earlier work of Vieli and Payne, but large weak-bedded areas further upstream. Thwaites Glacier exhibits a similar spatial pattern of basal resistance as well. This study also took the new step of examining glacier response to retreat of the grounding line under different assumptions about the bed sliding rheology. They find that perturbations at the grounding line propagate furthest upstream in the same amount of time with plastic deformation of the bed than with linear viscous deformation.

Both Vieli and Payne (2003) and Joughin et al. (2009) find areas where the basal shear stress is comparable to the driving stress near the grounding line, which suggests that vertical deformation in this region is not negligible. In that case, the shallow shelf approximation might no longer be appropriate. Morlighem et al. (2010) estimated the shear stress under Pine Island Glacier using the shallow shelf, Blatter-Pattyn, and full Stokes models. This work was the first to examine whether the choice of physical model used to constrain the inference problem affects the solution enough to warrant more expensive 3D models. While the inferred shear stress is roughly similar throughout most of the glacier in all three models, the full Stokes model finds substantially weaker shear stresses in a neighborhood of the grounding line than either the shallow shelf or Blatter-Pattyn approximations. The authors attribute this discrepancy to the fact that the lower-order models ignore the "bridging effect", where vertical normal stresses under the grounded ice can exceed the overburden pressure in the immediate vicinity of the grounding line. It is not certain, however, that this interpretation is correct. The basal shear stress that Joughin et al. (2009) computed using the shallow shelf model is closer to the basal shear stress that Morlighem et al. (2010) computed using the full Stokes model than it is to the shallow shelf or Blatter-Pattyn model results. The discrepancy could also be due to differences in the temperature parameterization, the regularization or early stopping procedure, or the computational mesh.

3.5 Basal shear stress in Greenland

MacAyeal's introduction of control methods to glaciology was an outgrowth of the need to understand why the Ross ice streams flow as fast as they do, despite low driving stresses on the order of 10 kPa or less. Greenland outlet glaciers, on the other hand, are often extremely steep; for example, driving stresses at Jakobshavn Isbrae can exceed 250 kPa. One could easily imagine that a combination of high driving stresses and rheological weakening within the ice column are the cause of fast flow at Jakobshavn, even if the glacier is flowing over a hard bed. Echelmeyer and Harrison (1990) measured the surface velocity of Jakobshavn at several points from 1984-1986 and found almost no seasonal variation in the glacier speed, despite the fact that virtually all surface meltwater input occurs during summer. Blue Glacier in Washington State also exhibits relatively little seasonal variation; borehole measurements showed that its flow is due entirely to vertical deformation rather than sliding (Engelhardt et al., 1978). By analogy with Blue Glacier, Echelmeyer

and Harrison (1990) took the lack of seasonal variation at Jakobshavn as possible evidence that it too experiences relatively little sliding. They do offer an alternative explanation, that frictional heating at the base of Jakobshavn generates the meltwater necessary to maintain fast sliding. Meier and Post (1987) were more equivocal; they argue that vertical deformation within the ice column could be the main driver of fast flow, but that sliding could still be important depending on how water is routed through the glacial hydrological system. Borehole measurements conducted in subsequent field seasons revealed that the ice in the bottom few hundred meters of Jakobshavn undergoes substantial vertical shear strain, and is at the pressure-melting point (Iken et al., 1993). Since this bottom layer is at high temperature and is likely saturated with meltwater, the strain rate will be 50 times higher than ice at the column-wide minimum temperature of -22.5° C under the same applied stress. Iken et al. theorize that 3D flow effects further exaggerate the influence of this basal layer. The ice in this region converges horizontally on the deep bedrock trough in the last few tens of kilometers near the terminus; since ice is incompressible, this horizontal convergence means that the ice column is greatly extended in the vertical. Large vertical extension would then enhance the thickness of the weak temperate layer. Subsequent analysis of these borehole measurements quantified the degree of vertical extension, finding that the bottom 270 m of the ice column in the slower-flowing ice sheet are extended out to a 1700 m-thick layer in the faster-flowing regions downstream (Lüthi et al., 2002). They conclude that the fast flow of Jakobshavn can be explained by deformation within the thick, weak temperate basal layer, rather than through basal sliding or till deformation.

Earlier researchers drew certain conclusions about the flow regime of Jakobshavn based on the lack of a seasonal signal, but the story took an interesting turn when the glacier speed began to change dramatically in the late 1990s. From 1982 to 1992, Jakobshavn slowed from 6.7 km/yr to 5.7 km/year; between 1992 and 2003, the glacier accelerated from 5.7 km/yr to 12.6 km/yr (Joughin et al., 2004a). From 1997-2001, the floating ice tongue at the front of Jakobshavn thinned by roughly 80 m/yr (Thomas et al., 2003). In the following four years, the steady year-round speed of the glacier gave way to summer speedups (Joughin et al., 2008a). The seasonality became even more pronounced between 2008 and the present, reaching speeds of 14 km/yr in the summer of 2009 and 18 km/year in the summer of 2012 (Joughin et al., 2012, 2014a).

Several explanations were offered for these striking changes. Zwally et al. (2002) hypothesized that increasing surface melt from the Greenland Ice Sheet could initiate a runaway feedback. They argue that meltwater lubrication can facilitate faster outlet glacier flow, resulting in increased strain heating, more meltwater production and thus faster flow. Subsequent work found that, while seasonal meltwater lubrication did enhance the flow of the interior of the ice sheet, in some cases by as much as 100%, the relative changes in the fastest-flowing outlet glaciers were only 15% (Joughin et al., 2008a). While meltwater input can cause an outlet glacier to speed up, the acceleration likely levels off at high meltwater flux due to the development of an efficient, channelized subglacial drainage network. Instead, the loss of the floating ice tongue at the front of Jakobshavn in 1998 is regarded as the most likely explanation for the speedup and thinning that followed (Thomas,

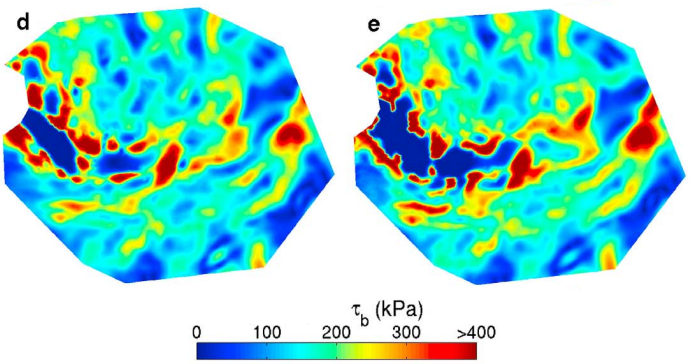


Figure 3.2: Basal shear stress of Jakobshavn Isbrae from the 1990s (d) and 2009 (e). Reproduced from Joughin et al. (2012).

2004; Joughin et al., 2004a). Through drag at the side walls and back-pressure at the terminus from the ice-choked fjord, the floating ice tongue restrained the flow of the glacier. From the 1950s, the glacier terminus held a relatively stable position, deviating by ±2.5 km from its mean; this is to be compared to the 15 km length of the entire ice tongue (Joughin et al., 2008b). The loss of buttressing that resulted from the disintegration of this ice tongue from 1998-2003 precipitated the dramatic changes that were to follow. The initial thinning of the ice tongue was triggered by sudden warming of subsurface ocean waters (Holland et al., 2008) and reduced sea ice in the fjord (Joughin et al., 2008b), which could have been enough to destabilize the ice tongue whether or not these oceanic conditions persisted in the following years (Motyka et al., 2011).

The basal shear stress under Jakobshavn was first computed using control methods in Joughin et al. (2012). Using velocity maps from the 1990s and from 2009, their results suggest that the bed resistance has decreased during this time period. The basal shear stress maps for each time period are reproduced in figure 3.2. The observed acceleration of Jakobshavn cannot be explained by advance or retreat of the terminus alone, so Joughin et al. argue that changes in bed resistance are also an important factor. The rapid acceleration that occurred at Jakobshavn over the past two decades was accompanied by substantial ice thinning, reducing the overburden pressure and thus the effective water pressure N of basal water. They hypothesize that the lower resistance at the ice bed is due to dependence of the basal shear stress on some power of the effective pressure, i.e.

$$\tau_b = -C(N/N_0)^q |u|^{\frac{1}{m}-1} u \tag{3.68}$$

where N_0 is some reference pressure. The observed changes in basal shear stress are consistent with a constant basal water pressure and $q = 3$. Nonetheless, this dependence is not conclusive and more time slices of basal shear would be necessary to constrain this exponent.

Habermann et al. (2013) performed several inversions for Jakobshavn Isbrae using data sets from 1985, 2000, 2005, 2006, and 2008, finding that the basal yield stress has decreased over this time period. The spatial pattern of weakening in their study aligns well with the change in height above flotation over this time period, which is consistent with basal resistance being a function of effective pressure at the ice bed. If increased summer melt were the controlling factor in the seasonal speedups at Jakobshavn observed since the early 2000s, the basal yield stress would be decreasing across the entire Jakobshavn bed.

Improvements in computational power in the past decade have made it possible to model whole ice sheets, rather than focusing on individual glaciers. Gillet-Chaulet et al. (2012) inferred the basal shear stress under the entire Greenland Ice Sheet using both the adjoint and Robin methods. They then used the initial state that they obtained for the Greenland Ice Sheet to model its evolution until 2100, finding a possible stabilization of the ice sheet size even under higher ablation rates. The rapid initial adjustment they find in the surface elevation highlights the fact that, while the initial conditions obtained from data assimilation methods can give a basal shear stress consistent with velocity measurements, the model may be far out of mass balance due to uncertainties in other data such as bed elevation. Due to the immense computational cost of a 3D simulation of the entire ice sheet, they were limited to a horizontal resolution of 1 km in the outlet glaciers. Helheim and Kangerdlugssuaq are roughly 10 km across at their termini, so this resolution is barely able to capture the outlet glaciers. As a consequence, the fit to the observational data had a relative error which approached 100% in some areas at the margins of the ice sheet. Brinkerhoff and Johnson (2013) developed a numerical ice sheet model called VarGlaS using both the full Stokes model and the Blatter-Pattyn approximation with the finite element modelling library FEniCS. Using the automatic differentiation functionality of FEniCS, the authors estimated the basal shear stress underneath the entire Greenland Ice Sheet at a maximum resolution of 500 m in the outlet glaciers. They used a smoothing length α equal to the ice thickness h to regularize the problem. To my knowledge, this is the only published work in the literature where the smoothing length is reported in such a way that other studies can reproduce their method. Nonetheless, a full L-curve is not reported, without which it is not clear whether their results are under- or over-regularized.

Sergienko and Hindmarsh (2013) find a spatial banding structure in the basal shear stress underneath Pine Island and Thwaites glaciers using the full Stokes model and high spatial resolution, which they associate with variations in the local hydraulic potential. The inferred basal shear they obtain for Pine Island Glacier differs substantially from previous work (Joughin et al., 2009; Morlighem et al., 2010). A comparison between the three is shown in figure 3.3. It could be argued that the discrepancy between the work of Sergienko et al. and Joughin et al. is due to the higher fidelity of the Stokes model used in the former work compared to the shallow stream model used in the latter, but Morlighem et al. also used the full Stokes model and found better agreement with Joughin than with Sergienko. Extending these results in Sergienko et al. (2014) to several glaciers in Greenland and Antarctica, including Jakobshavn Isbrae, they argue that these spatial band structures are the result of a pattern-forming instability (in the sense of Turing) related

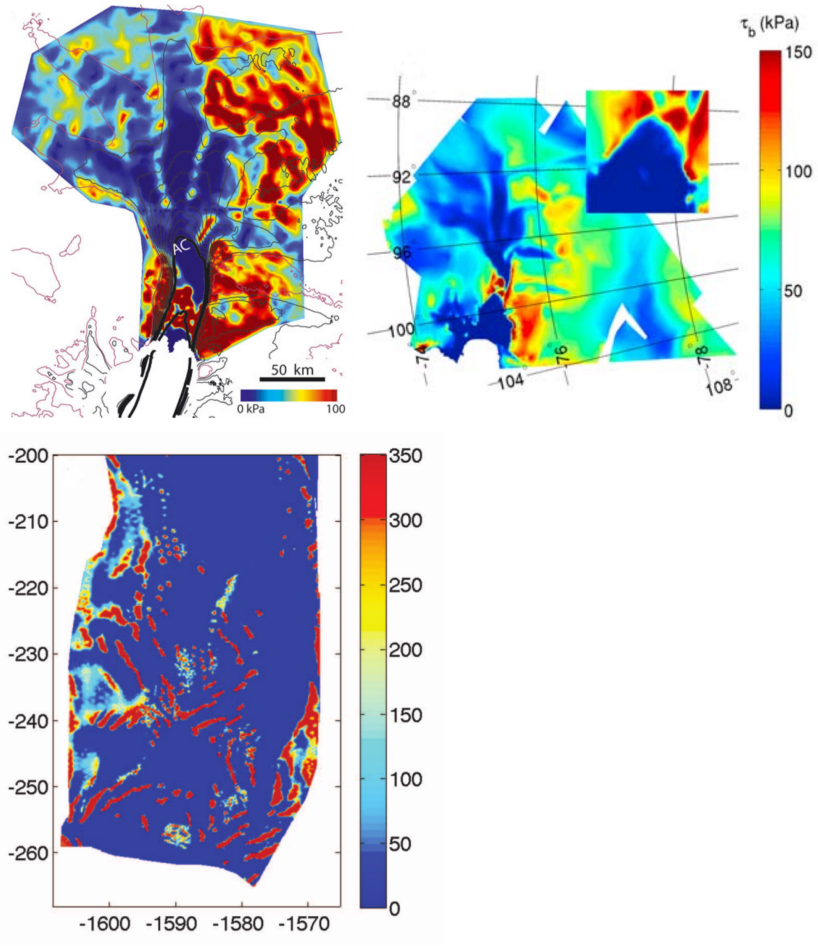


Figure 3.3: Basal shear stress of Pine Island Glacier from Joughin et al. (2009) (top left), Morlighem et al. (2010) (top right), and Sergienko and Hindmarsh (2013) (bottom).

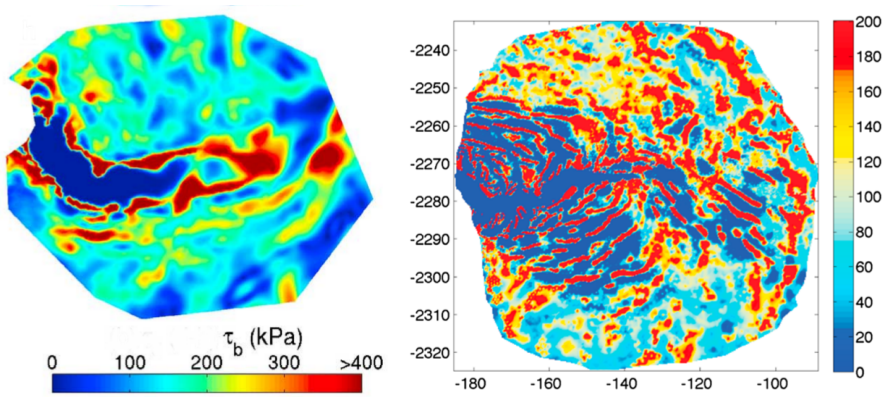


Figure 3.4: Basal shear stress of Jakobshavn Isbrae from Joughin et al. (2012) (left) and Sergienko et al. (2014) (right).

to subglacial hydrology. While some banding structure can be seen outside of the main trunk of Jakobshavn in the basal shear stress inferred in Joughin et al. (2012), the features that Sergienko et al. find are far more exaggerated (see figure 3.4 for a comparison). The small-scale features that Sergienko et al. find might instead be artifacts of overfitting. The peak-to-peak distance of these rib-like patterns varies from 5-20 km, while the width of a rib is often less than 1 km. The ice thicknesses in many of these regions are typically on the order of 1 km or less, so the spatial scales of interest here are approaching or below the limit of the scales that can be inferred from surface observations (Gudmundsson, 2003). In the supplementary material for Sergienko et al. (2014), the authors show a comparison of results obtained with three different values of the regularization parameter α. Many features of the ribbed patterns are present in the basal shear stress fields for every value of α they tested, but in some regions the peak-to-peak distances or rib widths change depending on the smoothing length. In any case, it cannot be determined whether their results are a result of overfitting or not without estimating the basal shear stress at many more values of the regularization parameter.

At the end of 2015, the only study in the literature that examined the basal shear stress of more than one Greenland outlet glacier besides Jakobshavn using both high spatial resolution and the full Stokes model was Sergienko et al. (2014). The spatial banding structure they find is not nearly as prevalent in previous work performed at comparable spatial resolution (Joughin et al., 2012; Habermann et al., 2013). The previous studies may have missed the basal traction ribs that Sergienko et al. find because they used the shallow stream approximation rather than the full Stokes model. Jakobshavn flows through a deep, narrow trench, roughly 10 km across at the terminus and 1 km thick. The aspect ratio of 1/10 could be large enough that the assumptions underlying depth-averaged models may no longer be valid.

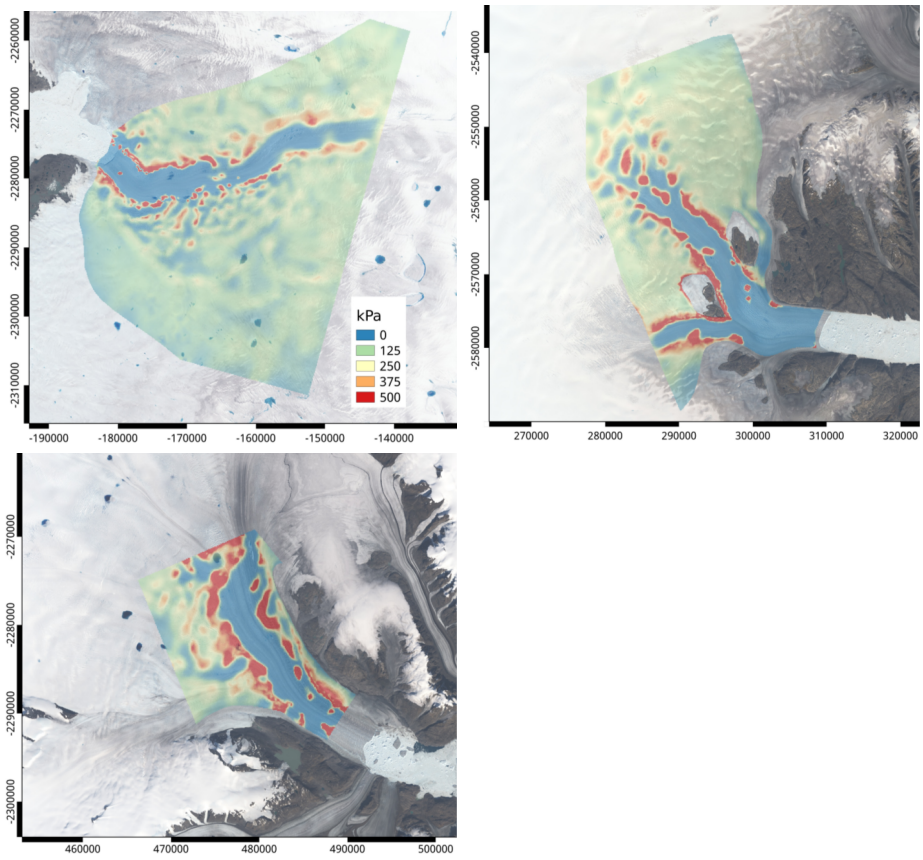


Figure 3.5: Basal shear stress of Jakobshavn, Helheim and Kangerdlugssuaq glaciers, from Shapero et al. (2016).

An additional outstanding question was whether Jakobshavn is an oddity or whether other glaciers in Greenland exhibit similar basal regimes. Sergienko et al. included studies of Petermann and Nioghalvfjerdsbrae glaciers in the northwest and northeast of Greenland respectively. Jakobshavn is much further south (69°N) than either Petermann (81°N) or Nioghalvfjerdsbrae (79°N), so it is likely to experience substantially different meltwater input and subglacial hydrological conditions. Helheim and Kangerdlugssuaq glaciers in east Greenland are at closer latitudes (66°N and 68°N respectively) to Jakobshavn. Moreover, Helheim and Kangerdlugssuaq discharge substantially more ice than either Petermann or Nioghalvfjerdsbrae. Jakobshavn, Helheim, and Kangerdlugssuaq are often referred to as the "big three" Greenland outlet glaciers; they account for a total of 40% of the dynamic mass loss from the Greenland Ice Sheet (Enderlin et al., 2014). How different are the basal states of these two glaciers from that of Jakobshavn?

In my first paper, I used a high-resolution full Stokes glacier model to infer the

basal shear stress under the big three Greenland outlet glaciers (Shapero et al., 2016). The goals of this study were:

1. to see whether Helheim and Kangerdlugssuaq flow vertical deformation as Iken et al. (1993); Van Der Veen et al. (2011) suggested Jakobshavn does, or whether they flow by sliding;

2. to evaluate how similar or different the basal regimes of these glaciers are, for example with respect to sliding ratio, size and distribution of basal sticky spots, and so forth;

3. and to independently confirm or deny the existence of the shear stress bands that Sergienko et al. (2014) found.

The basal shear stress maps from this paper are shown in figure 3.5, overlaid on Landsat 8 imagery.

One of the outstanding problems in determining the shear stress underneath a glacier is how to set the ice rheology coefficient. There is no unique solution for both shear stress and rheology from a single measurement (Arthern and Gudmundsson, 2010). Consequently, we must prescribe one field or the other. The rheology is in large part determined by the ice temperature, and we have some idea of the surface temperature and geothermal heat flux. Since we have some estimates of the relevant boundary conditions, we can try to set the ice rheology from a modelled englacial temperature. The meltwater content within a glacier also affects the rheology; if W is the meltwater fraction in percent, a standard parameterization for flow enhancement due to meltwater is

$$E = 1 + 1.8125W \tag{3.69}$$

which gives a net enhancement of 2.8125 if the meltwater fraction is at the theoretical maximum of 1% (Cuffey and Paterson, 2010). My officemate and co-author Kristin Poinar was developing a finite-difference model for heat and meltwater transport in glaciers at the time (Poinar et al., 2015). Her model computes the temperature along flowbands from the ice divide to the terminus of a glacier. Additionally, by solving the Stefan problem and tuning the results to fit measurements from Jakobshavn boreholes, her model calculates the height where the transition to temperate ice occurs and the meltwater fraction in the temperate ice. By using a dense set of flowbands, we can interpolate the results of the mode runs to the 3D finite element grid to set the ice rheology throughout the domain. While explicitly modelling the temperature improves on the approach used in Joughin et al. (2012) of trying a range of synthetic temperature fields, one potential weak point of this flowband temperature model is that it ignores lateral stresses. The largest discrepancies occur at the glacier shear margins, where strain heating is the strongest. Sergienko et al. (2014), on the other hand, used a 3D temperature model to set the ice rheology. To evaluate the sensitivity, I estimated the basal shear stress of Helheim glacier using a spatially constant temperature field. There were noticeable differences in the inferred stress, but not so much so that our conclusions and interpretation could depend on how the ice rheology was parameterized.

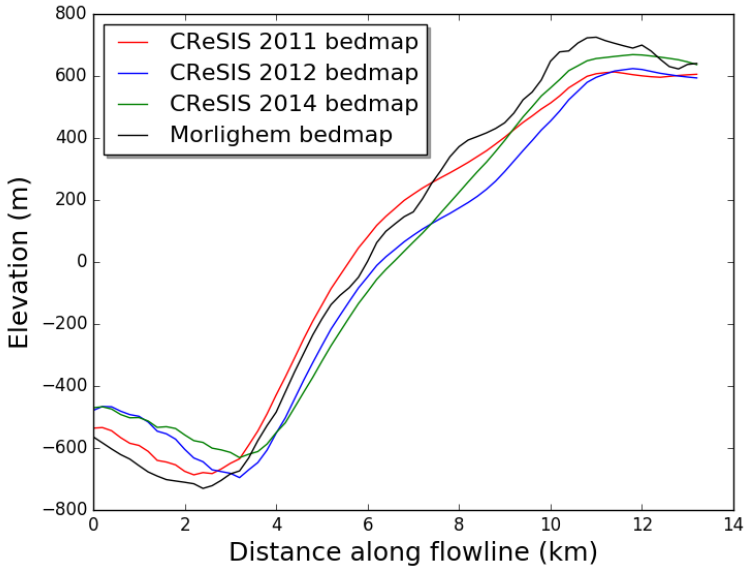


Figure 3.6: Comparison of 3 versions of CReSIS bed maps with Morlighem bed map along a flowline at Helheim Glacier.

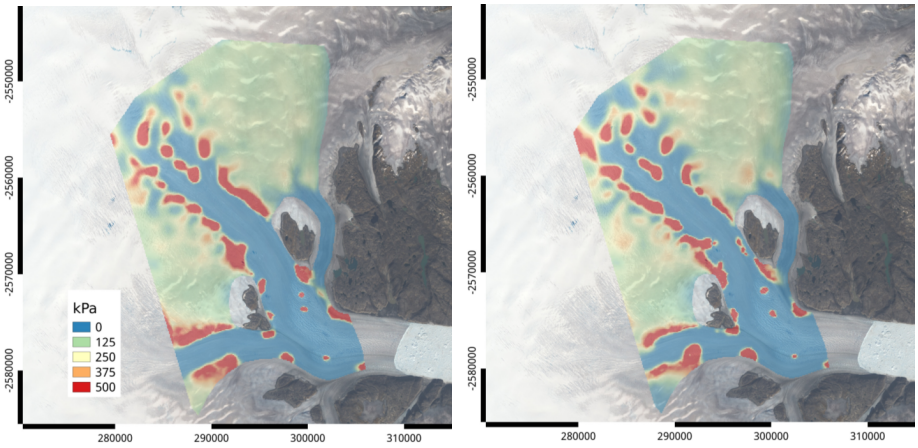


Figure 3.7: Shear stress under Helheim Glacier, using Morlighem bed map (left) and CReSIS bed map (right).

The surface topography and velocity of Greenland outlet glaciers have been measured at high accuracy from satellites, but the bed topography is much less constrained. The Center for Remote Sensing of Ice Sheets (CReSIS) has produced several bed DEMs of Jakobshavn Isbrae since 2005 using airborne ice penetrating radar (Gogineni et al., 2014). This technique is fraught with difficulties. Jakobshavn, Helheim, and Kangerdlugssuaq all flow through deeply-incised bedrock troughs. A radar return from the fjord side walls can easily be mistaken for an echo from a much deeper bed. Englacial water layers are a bright radar reflector, which can have the opposite effect of making the bed elevation appear higher than the true value. The ice thickness measured along a dense set of flight lines is then interpolated to a gridded data set through kriging, but the ice flux $\nabla \cdot hu$ computed from the resulting thickness map may depart substantially from measured rates of elevation change $\dot{a} - \dot{m} - \partial h / \partial t$. Rasmussen (1988) originally conceived of an approach to mapping bed topography with measured mass balance rates as an explicit constraint. Morlighem and others subsequently applied this technique to improving estimates of ice thickness in Greenland (Morlighem et al., 2011, 2014). The bed elevations along a flowline at Helheim Glacier from three different CReSIS data sets and from Morlighem's DEM are compared in 3.6. At some locations along the flowline, the bed DEMs differ by as much 20% of the ice thickness. In order to be sure that our results were not dependent on the particular choice of DEM, we compared the outcomes for the Morlighem bed map and the most recent (2014) CReSIS bed map; the inferred shear stresses are shown in figure 3.7. The basal shear stresses inferred from each DEM are, in broad strokes, very similar, with low bed resistance along the fastest-flowing parts of the glacier.

The shorter length-scale details, however, differ between the two bed maps, particularly in the shear margins. Especially noteworthy are the sticky spots in the glacier trunk. These sticky spots coincide with local bedrock highs, for example, the contour outline in grey in figure 3.8. The spatial coincidence of local bedrock maxima and basal sticky spots is consistent with the hypothesis that some of these intermittent patches of high bed resistance are caused by protrusions of the underlying bedrock into or through a layer of weak till (Alley, 1993). The sticky spots we identify under Helheim are roughly 1 km across. With the finest grid resolution we used in this study of 175 m, these spots are 6 grid cells across in our discretization. Assuming that these features are not artifacts, they may be less diffuse due to the regularization we applied. These features are largely missed in work that covers the entire Greenland Ice Sheet such as Gillet-Chaulet et al. (2012); Brinkerhoff and Johnson (2013) due to the necessity of lower spatial resolution.

To solve the diagnostic equations, we have to supply all the boundary conditions for the elliptic PDE, including the velocities along the inflow boundary of the domain. For the shallow stream equations, we can use the observed surface velocities as an approximation to the depth-averaged velocities at the inflow and side wall boundaries, interpolating as need be to the finite element grid points. For 3D flow models, such as the Stokes or Blatter-Pattyn equations, we have to also decide how ice velocity changes with depth at the inflow boundary of the domain. Provided we had some a priori guess for the vertical shear strain rate in the ice column, we could integrate this quantity from the ice surface to the bed to compute the depth-

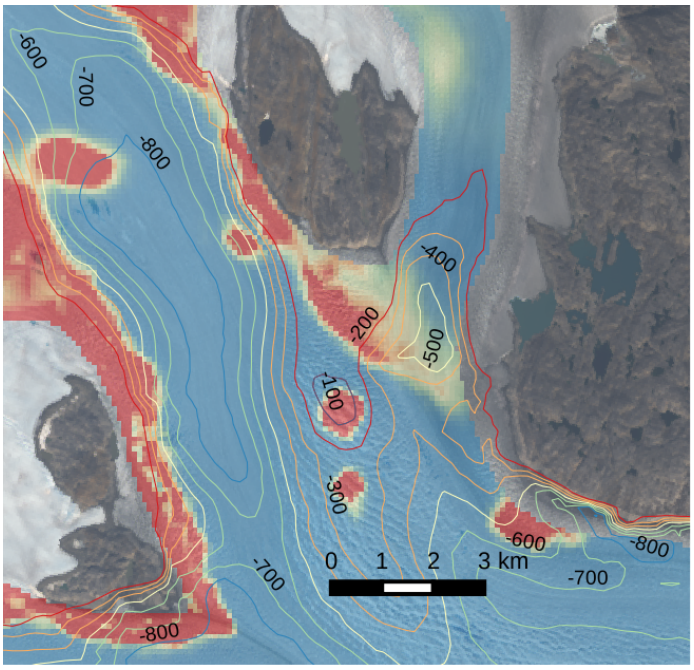


Figure 3.8: Bed elevation contour lines of Helheim Glacier overlaid on the inferred basal shear stress. A sticky spot coincides with the -100 m-elevation bedrock high, outlined in grey.

dependence of the inflow velocities. However, the vertical shear strain rate depends on precisely the quantity we are trying to infer in the first place, since the strain rate tensor is a power of the stress tensor. If we made the blunt approximation that the ice velocity is pure plug flow at the inflow boundary, we would overestimate the flux into the domain from the parts of the boundary that are predominantly in sheet flow. The inferred basal shear stress of Jakobshavn with plug flow along the inflow boundary is shown in figure 3.9. Along the boundaries, the basal shear stress is 0, since the Dirichlet boundary conditions force the vertical variation of velocities to be 0. The basal shear stress adjusts to the proper value over a 1 km-wide transition zone. We could instead use the shallow ice approximation and the surface DEM to estimate the vertical shear stress and, in turn, the vertical shear strain rate, assuming that all flow is due to vertical deformation along the boundary. The vertical shear strain rate estimate at the surface is then integrated down to the bed to estimate the side wall velocities. This approach also turns out to give poor results. In some regions, the surface slopes are so high that the sliding velocity computed from the shallow ice approximation points in the opposite direction as the surface velocity, which is clearly unphysical.

To make matters even worse, the optimization algorithm used to obtain β can

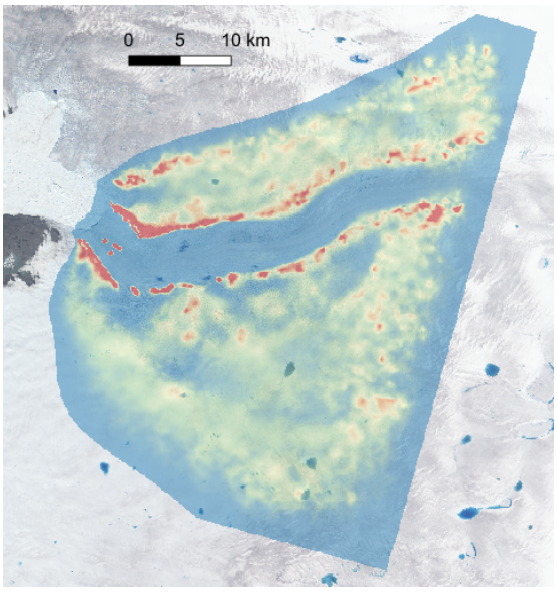


Figure 3.9: Basal shear stress of Jakobshavn Isbrae assuming that the inflow velocities are constant with depth.

hardly correct the inflow boundary conditions, if at all. When using the adjoint method, we are stuck with whatever initial guess we use for the inflow velocities because the boundary conditions for the adjoint state λ are homogeneous Dirichlet conditions, i.e. $\lambda_i|_{\partial\Omega} = 0$, wherever the velocities have Dirichlet boundary conditions. The gradient of the objective functional with respect to the basal friction parameter β is

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\beta}, \phi \right\rangle = \int_{\Gamma_b} \frac{\mathrm{d}C}{\mathrm{d}\beta} |u|^{\frac{1}{m}-1} u_i \lambda_i \cdot \phi \, \mathrm{d}\Gamma. \tag{3.70}$$

If λ is 0 in a neighborhood of the inflow boundary, then the gradient of the objective functional in a neighborhood of the inflow boundary is also 0. The situation is hardly better if we are using the Robin method, for which the gradient of the objective functional is

$$\left\langle \frac{\mathrm{d}J}{\mathrm{d}\beta}, \phi \right\rangle = \int_{\Gamma_b} \left(|u^D|^2 - |u^N|^2 \right) \phi \, \mathrm{d}\Gamma \tag{3.71}$$

where u^D and u^N are the Dirichlet and Neumann solutions, respectively. Presumably, we use the same inflow boundary conditions for both the Dirichlet and Neumann solution, so the gradient of the objective functional will be roughly 0 near the inflow boundary for the Robin method as well. In short, β will update slowly if at all along the inflow boundary for both the adjoint and Robin methods.

In order to combat this problem, I went with an intermediate approach between pure plug flow and vertical deformation. First, the ice surface elevation is spatially filtered; this step is often necessary when using the shallow ice approximation. The SIA is only appropriate at scales greater than 10 ice thicknesses. At smaller scales, large fluctuations in the surface slope are smoothed over by membrane stresses. Additionally, differentiating the surface elevation could be disastrous if the white noise component from the measurement errors is not removed in some way. Using this guess for the strain rate, we compute the SIA inflow velocity u_s, which may have unphysical retrograde sliding velocities. The inflow velocity is then set as a convex combination

$$u_{\text{inflow}} = (1-r)u_p + ru_s \tag{3.72}$$

of the SIA velocity and the plug flow velocity u_p. I tried several values of the SIA fraction r: 0.25, 0.5 and 0.75. For $r = 0.75$, isolated patches points on the inflow boundary of Jakobshavn Isbrae still have retrograde sliding velocities. For $r = 0.25$, the velocities still exhibited a 1-3 km-wide adjustment zone to the true vertical shear at the inflow boundaries. The intermediate value $r = 0.5$ gave an acceptable-looking velocity field.

In order to evaluate whether the results I obtained were polluted by edge effects from misspecification of the inflow boundary conditions, I computed the basal shear stress under Kangerdlugssuaq Glacier using four different meshes of the domain. The resulting basal shear stress maps are shown in figure 3.10. Some features of the basal shear stress depend on the particular choice of domain boundary, but the overall features in the interior are preserved. The basal shear stress is relatively unchanged along the inflow boundary for each domain, so we can be fairly confident that the ad-hoc parameterization used at least gives consistent results there. The side wall boundaries, however, are another story. Jakobshavn Isbrae is surrounded by slow-flowing ice on either side of its fast-flowing trunk, so if one were to do a poor job setting the inflow boundary velocities, the inferred basal shear stress might still be mostly correct in the interior of the domain. There is no slow-flowing ice on either side of the trunks of Helheim and Kangerdlugssuaq, only the rock faces of the fjords they have carved out over time. Additionally, the velocity maps for these glaciers have the greatest uncertainty in the immediate vicinity of the rock walls. It is difficult to discern from satellite measurements whether the glacier slips along the fjord walls, or whether there is instead a narrow and intense shear margin. In short, the problem of misspecifying the side wall velocities is arguably more severe than the inflow velocities for glaciers like Helheim and Kangerdlugssuaq that flow into deep fjords.

Several studies have argued that the full Stokes model is preferable for general glacier flow modelling and for data assimilation problems (Pattyn et al., 2008; Morlighem et al., 2010). The whole point of inferring the basal shear stress underneath an outlet glacier is so that we can set a boundary condition that we cannot otherwise measure. Unless the spatial domain one uses is the entire Greenland Ice Sheet, using the Stokes equations only introduces another under-constrained boundary condition, namely the inflow velocities. For all their supposed inaccuracy for modelling complex glacier flows due to the near-plug flow assumption, the shal-

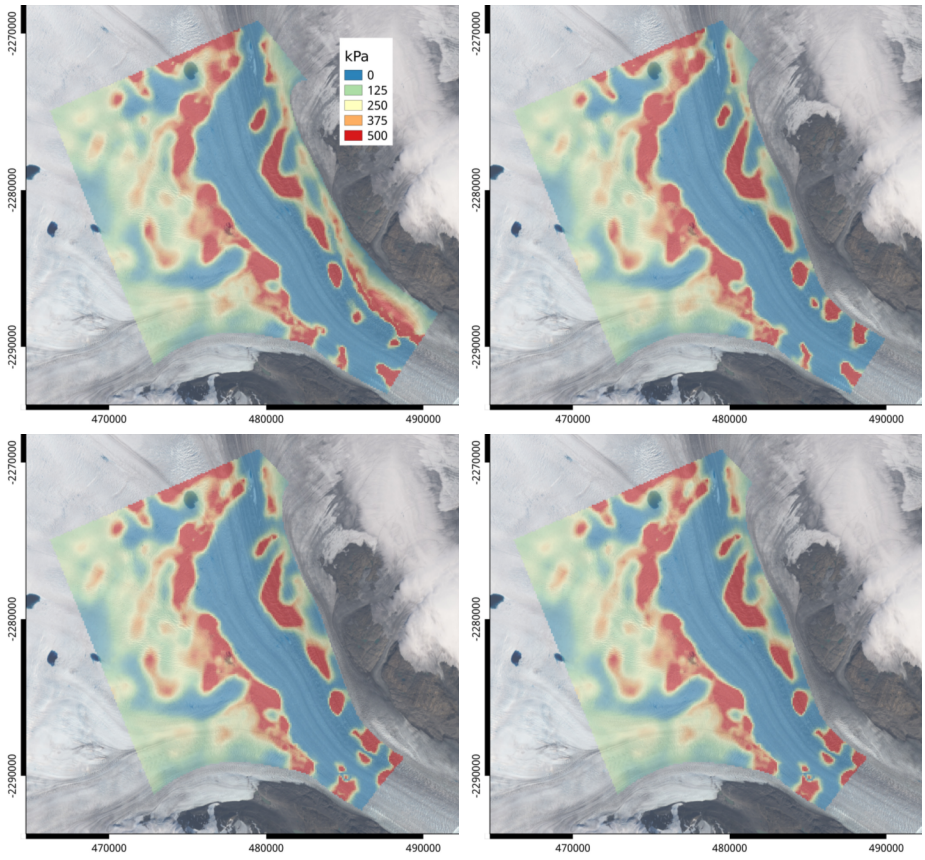


Figure 3.10: Basal shear stress under Kangerdlugssuaq Glacier, computed with several different spatial domains.

low stream equations are at least self-consistent. One way to resolve this dilemma would be to to infer the basal friction parameter using the shallow shelf equations as the physical model first, possibly on a larger spatial domain. The inferred parameter is then used to estimate the vertical shear strain, fixing the inflow velocities, and as an initial guess in a subsequent inversion with the Stokes model. To my knowledge, no one has attempted such a multi-fidelity modelling approach in glaciology.

Previous borehole studies of Jakobshavn Isbrae near its fast-flowing trunk observed high vertical shear strain rates, which would seem to suggest a strong bed and little sliding. We can reconcile this supposed inconsistency with our results by examining the locations of the boreholes that Iken et al. (1993) drilled in 1989 and the later measurements by Lüthi et al. (2002) from 1995-1996; these are shown in figure 3.11. We also find high vertical deformation rates near boreholes A and C, which are right at the boundary of the ice stream shear margins. At sites B and E,

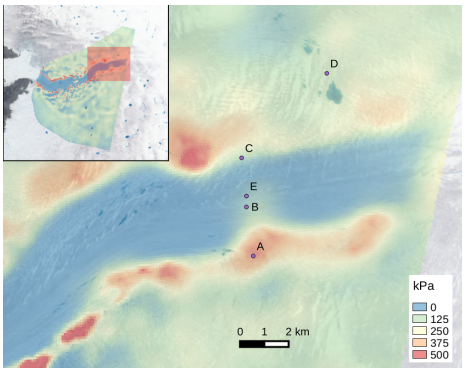
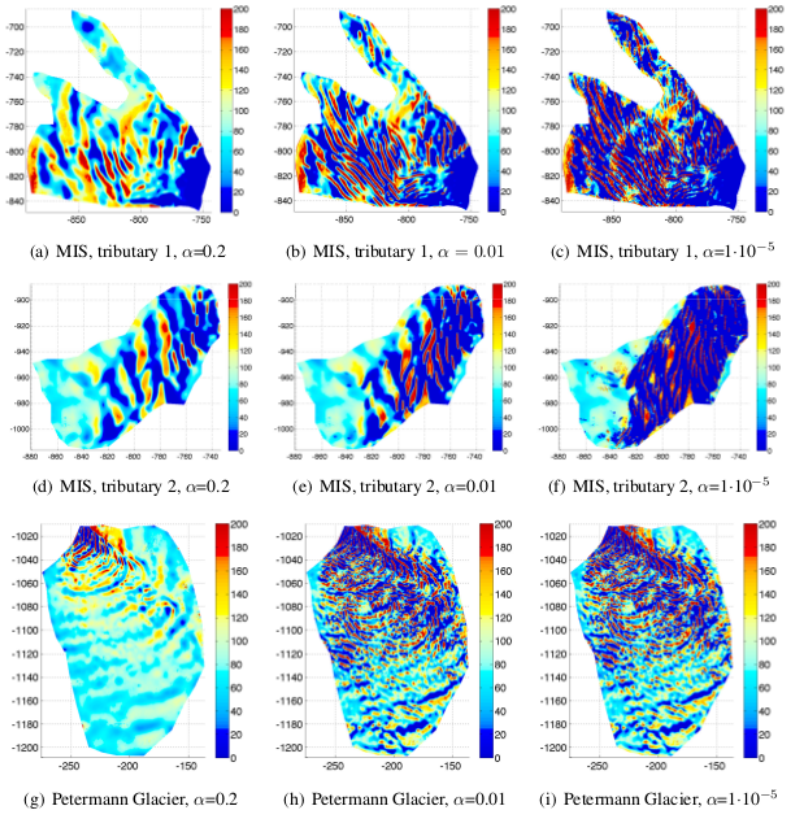


Figure 3.11: Locations of boreholes from Iken et al. (1993) (A-C) and Lüthi et al. (2002) (D-E).

we find a small promontory of higher basal shear stress (∼65 kPa) relative to the otherwise weak background (∼10 kPa). While the shear stress at these two boreholes is still appreciably smaller than the driving stress of ∼300 kPa, the slightly elevated bed resistance at this spot could explain some of the vertical deformation that was seen in both previous measurements. It is possible that both studies came to the conclusion that Jakobshavn flows by vertical deformation because they happened to drill at the one spot within what looks like the fast-flowing trunk that has a patch of anomalously high bed resistance, according to the basal shear stress we inferred. If a new borehole were drilled 2 km to the west of site B, our results suggest that it would show almost no vertical deformation. More recent field work has also shown that, even in slower-flowing areas of the Greenland Ice Sheet, sliding ratios can exceed 0.5 in the winter and 0.9 in the summer (Ryser et al., 2014).

Finally, I was not able to reproduce the results of Sergienko et al. (2014). In the last 10 km before the terminus, the southern shear margin of Jakobshavn Isbrae exhibits what could be called faint echoes of the traction ribs that Sergienko et al. find. Nonetheless, the ribbed patterns in their work are far more widespread and the amplitude of the oscillations they find are substantially larger. We chose the regularization parameter for all the results in this study using the L-curve, i.e. by computing the basal shear stress of Helheim Glacier with $\lambda = 8.0, 8.25, \ldots, 11.0$ and finding the point of maximum curvature along the plot of the model-data misfit vs. the integrated square gradient of the friction parameter (Hansen, 1999). We can thus be fairly confident that we have not missed any genuine features by smoothing the solution too much, and that the features we do find are not mere artifacts. In the supplementary information to their paper, Sergienko et al. show the basal shear stress of three glaciers computed with $\alpha = 10^{-5}$, 10^{-2}, 0.2. These are shown in figure 3.12. They argue that, since the basal shear stress bands are present for each value of the regularization parameter, these features are robust to the choice of α. On the other hand, the spatial pattern of basal shear stress changes appreciably. For example, while some oscillations with a wavelength of roughly



(a) MIS, tributary 1, α=0.2 (b) MIS, tributary 1, $\alpha = 0.01$ (c) MIS, tributary 1, α=1·10^{-5}

(d) MIS, tributary 2, α=0.2 (e) MIS, tributary 2, α=0.01 (f) MIS, tributary 2, α=1·10^{-5}

(g) Petermann Glacier, α=0.2 (h) Petermann Glacier, α=0.01 (i) Petermann Glacier, α=1·10^{-5}

Figure 3.12: Sensitivity of inverse method to regularization parameter α from supplementary material to Sergienko et al. (2014). The glaciers shown are two tributaries of MacAyeal Ice Stream in (a-c) and (d-f), and Petermann Glacier (g-i).

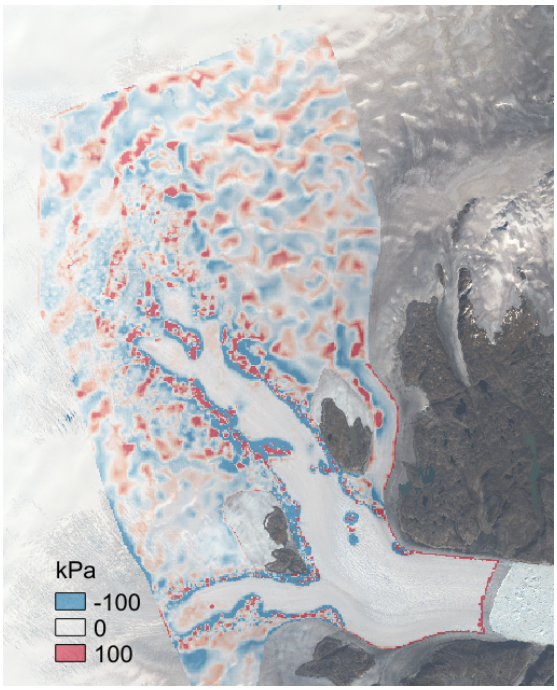


Figure 3.13: Difference of the basal shear stress maps computed under Helheim glacier for regularization parameter values of $\alpha = 10^{10}$ (near the corner of the L-curve) and $\alpha = 10^{8}$ (overfit).

20 km are discernible in the shear stress under Petermann Glacier for $\alpha = 0.2$ in panel (g), these give way to a completely different result for $\alpha = 10^{-2}$ in panel (h). The change from $\alpha = 10^{-2}$ to $\alpha = 10^{-5}$ from panels (e) to (f) for the second tributary of MacAyeal Ice Stream also shows a substantial change in the number of peaks and in the sharpness of the transition. It is not possible to ascertain whether a map of basal shear stress from any one particular value of α is in the overfit or underfit regime without a denser sampling over the range of parameter values they used. While it would be tempting to attribute their results entirely to overfitting, the basal shear stress maps I computed in the course of calibrating the right value of α through the L-curve do not exhibit the same spatial banding structure. A plot of the difference between the correctly fit and an overtfit basal shear stress for Helheim Glacier is shown in figure 3.13. Most of the differences between the two maps are spotty patches, rather than elongated features. It is interesting to note that many of the differences between the two maps correspond to visual surface features in the Landsat-8 imagery.

In sum, we argue that the banded features found in Sergienko et al. (2014) might have resulted from overfitting. We could be disproven on this point if further studies also found substantial banded structures, and showed that the regulariza-

tion parameter used to obtain these results was selected in accordance with the L-curve or the discrepancy principle. By arguing that the banded structures might be an artifact of overfitting, we are not asserting that they do not exist at all, only that our ability to ascertain their location or wavelengths from surface observations is limited. There are indications from geomorphology that regular wavelike patterns in bed topography occur naturally under real ice sheets. If so, the basal shear stress should then exhibit similar patterns at the same wavelengths, since it is partly controlled by topography. Richard Hindmarsh, a coauthor on both Sergienko and Hindmarsh (2013); Sergienko et al. (2014), did much of the foundational work on this topic in a series of papers from 1997-1999. To close this chapter, I will review some papers from the glacial geomorphology literature that have studied bedform development.

In Hindmarsh (1998a,b,c, 1999), Richard Hindmarsh elucidated the "instability theory" for the formation of drumlins and other glacial bedforms. This theory is predicated on his finding that the coupled ice-till-water system is linearly unstable in certain regions of parameter space. At small scales, the propagation speed of the instability is the same as the speed of the overlying ice, while at longer wavelengths the instabilities propagate back upstream; instabilities in the intermediate range are singled out as being the cause of drumlin formation (Hindmarsh, 1999). While this theory is suggestive of a possible generative mechanism, it uses simplified geometries, it assumes that till deformation is viscous at large scales, and it uses the shallow ice approximation. Consequently, its applicability is not obvious from the theory alone to fast-flowing outlet glaciers like Jakobshavn Isbrae, where Sergienko et al. (2014) claimed to find these ribbed features through inverse methods.

When Hindmarsh's papers on the subject were published, there was not enough observational data on the shape of bedforms such as drumlins or ribbed moraines to confirm or refute his theory. In order to evaluate this theory, Dunlop and Clark (2006) collected measurements on the shape of over 13,500 km of transects of ribbed moraines from regions formerly underneath the Laurentide, Irish, and Fenno-Scandian ice sheets. Fourier analysis of these landforms revealed wavelengths from 12 to 5800 m; the most common wavelength they found was 400 m. Dunlop et al. (2008) then compared this data set to the results of the Hindmarsh coupled till-ice flow model. The range of wavelengths they find across all parameter values roughly agrees with the range of wavelengths they find in the observational data. These studies serve as partial experimental confirmation of the hypothesis that ribbed moraines arise through an instability of the flat bed state of the coupled till-glacier system.

Stokes et al. (2016) sought to compare the traction ribs inferred in the work of Sergienko et al. with bedforms from palaeo-ice streams in western Canada. Stokes et al. remark on the general similarity between real landforms and inferred traction ribs, but the length scales between the two are appreciably different. The landforms they analyzed had lengths ranging from 1.2 to 3.6 km, which is well below the typical value of 6 km found in Sergienko et al. Sergienko and Hindmarsh (2013) point to the mega-ribs found in Greenwood and Kleman (2010) as a possible glacial landform that might correspond to the features they claim to have found, but the length scales of these features (20 km or more) are too large to match those of traction ribs.

Additionally, the typical amplitude of mega-ribs (∼5-10 m) is low enough that they were only discernible with recently available, high-resolution DEMs; Sergienko et al. question whether such low-amplitude features would produce enough of a disturbance to be detectable at the surface. The amplitudes of the rib-like landforms found in Stokes et al. (2016) were on the order of 10-20 m; since these could be large enough to produce an appreciable patch of basal resistance, they argue that these are likely the closer analogue to traction ribs than the mega-ribs found in Greenwood and Kleman.

Both Dunlop and Clark (2006) and Stokes et al. (2016) find ribbed bedforms under palaeo-ice streams, but in both studies the wavelengths are noticeably below those that Sergienko et al. find under Greenland and Antarctica. While the length scales of the traction ribs that Sergienko et al. find are generally above the ice thickness, the landforms found in observations are generally narrower than what the palaeo-ice stream thickness is thought to have been.

Finally, Ely et al. (2016) attempted to answer the question of whether subglacial bedforms span a continuum of shapes and orientations, or whether they can be grouped into distinct categories. They analyzed over 96,000 distinct bedforms from beneath palaeo-ice streams of the Laurentide, Cordilleran, British, and Fenno-Scandian ice sheets. This dataset was compiled from 16 previous studies that manually identified these bedforms from either photogrammetry or digital elevation models. Ely et al. then tabulated the dimensions of the bedforms in the along- and across-flow directions with respect to the palaeo-ice stream velocity. To confirm or deny this "continuum hypothesis", they used automatic clustering algorithms and manual cluster identification on the dataset of axis length and elongation ratio. They found a continuum of morphological characteristics among along-flow features, such as drumlins and mega-scale glacial lineations, which have in previous work been classified as distinct. Across-flow features, ranging from ribbed moraines up to mega-ribs, were likewise not distinguishable from each other using either automatic or manual clustering. The authors are more equivocal on whether the along- and across-flow features all belong to the same category. They point to a previously-unidentified class of nearly circular bedforms that constitute a bridge between the along- and across-flow bedforms in the length/elongation feature space. The along- and across-flow features and the quasi-circular forms are clustered either together or separately depending on small variations of the parameters used in the automatic clustering algorithm. In any case, the continuum nature of subglacial bedforms means that a procedure for inferring basal shear stress is likely to find some features that happen to have the same geometry as real subglacial bedforms by pure chance, even if the solution is overfit to the observational data.

Chapter 4

icepack

For my first paper, I used the glacier modelling library Elmer/Ice for all of the computations, which is built on the general-purpose finite element modelling library Elmer. The inputs to Elmer/Ice and the simulation to be conducted are specified through "solver-input files". The solver-input file describes whether the simulation is steady-state or transient; what kind of physical system is being solved (Stokes flow, the heat equation, etc.); what kind of linear solvers and preconditioners to use (CG, GMRES); convergence tolerances; what file format to output the results in; and so forth. The Elmer program then parses this input file, selects appropriate numerical modelling routines, and runs the simulation. While Elmer has a large library of different physical systems it can simulate, it is a large standalone program. This design choice dictates a certain workflow, and imposes a conceptual barrier to extending the core library functions for novice users. When the user only interacts with the software as a standalone program, but the core library is written in Fortran and C++, the learning curve to go from using the software to extending its functionality is much steeper than if the user is accustomed to using the software as a library in the language of implementation in the first place. The continued success of any open-source software project is contingent upon its ability to turn novice users into contributors; a project design that prevents this development should be viewed as deficient.

In this chapter, I will describe my work on the glacier flow modelling library icepack, available at `www.github.com/danshapero/icepack`. One of the key design considerations in writing icepack was to create a C++ library that could be called from programs that the user writes, rather than a large standalone program. Icepack is in turn built on the open-source finite element library deal.II (Differential Equations Analysis Library, version 2), available at `www.dealii.org`. My motivation for picking deal.II was largely based on the quality of its documentation. At present, deal.II has a suite of 56 tutorial programs that demonstrate its use for a host of physical problems, ranging from the Poisson equation to the coupled nonlinear Stokes and heat equations. The initial solvers in icepack were an adaptation of the step-8 tutorial program, which demonstrates the solution of the steady elasticity equations; ignoring nonlinearity, the shallow shelf equations are formally identical

to 2D steady elasticity. Many of the design choices in icepack are inspired by how the same conundrums were solved in deal.II.

This chapter assumes some familiarity with the C++ programming language and the C++ Standard Template Library, described in detail in Stroustrup (2013).

4.1 Key deal.II classes

To describe the design of icepack, we will first need to review some of the key classes in the deal.II finite element library, which icepack is built on (Bangerth et al., 2016). A similar design to deal.II is also employed in the FEniCS project and other finite element modelling libraries.

The most basic classes in deal.II represent points and direction vectors in Euclidean space. These class templates have the signature

```
template <int dim> class Point;
template <int rank, int dim> class Tensor;
```

The template arguments are the dimension `dim` of the underlying Euclidean space, and the rank in the case of tensors. For example, vectors (tensors of rank 1), matrices (tensors of rank 2), and higher-order tensors are all instantiations of the same class template. The ability to manipulate entire tensors rather than have to work with their individual components was especially useful in implementing solvers for the equations of motion of ice sheets in icepack. As demonstrated in section §2.4, the linearized constitutive relation is most naturally expressed at the level of tensors. A tensor contraction or outer product is much more concise to write in terms of the underlying mathematical operation than the loop it represents. The use of template metaprogramming techniques saves the developers from having to write and maintain several similar versions of the same class.

The first entry point into finite element modelling with deal.II is the class template used to represent the problem geometry. The `Triangulation` class template has the signature:

```
template <int dim, int spacedim> class Triangulation;
```

Rather confusingly, this class template represents unstructured quad- and hexahedral meshes and not triangular meshes. This class is templated on the intrinsic dimension `dim` and the dimension `spacedim` of the Eucliean space into which the geometry is embedded. These dimensions need not be the same; for example, to represent the boundary of a domain in 3D, we would use a Triangulation with `dim = 2` and `spacedim = 3`. In most scenarios, however, the spatial dimension defaults to be equal to the intrinsic dimension using template specialization. Much like with geometric primitives, the use of template metaprogramming allows for every possible dimension of geometry to be implemented in a single class template, rather than implemented several times with the incipient maintenance difficulty. The chief responsibility of this class is to provide iterators over the individual quadrangular or hexahedral cells of the geometry. This facility is used by other classes in deal.II for assembling finite element matrices and vectors.

Several other functionalities are required for modelling applications. These functionalities are implemented in distinct classes with the express goal of keeping their responsibilities as orthogonal as possible.

First, it is common practice to map the distorted cells of the geometry to the unit cube $[0,1]^d$, where d is the space dimension. Any functionality that requires evaluating a field at a point within the domain can instead be implemented in terms routines defined over the unit cell. For example, rather than compute the Gauss quadrature points within an arbitrary cell, the cell can be mapped to the unit cell and the Gauss quadrature points for the unit cell computed once. The ability to transform arbitrary cells to the unit cell is implemented in the class template `Mapping`. The default mapping class uses bilinear functions, but higher-order (biquadratic, bicubic, etc.) mappings can be used for higher accuracy. In order to integrate functions over arbitrary cells using a mapping to the unit cell, the `Mapping` class template defines routines to evaluate the determinant of the derivative of this mapping.

Assembly of finite element matrices requires evaluating integrals. In general, we cannot exactly evaluate these integrals exactly, so we must resort to numerical quadrature. The template

```
template <int dim> class Quadrature;
```

defines the interface for all quadrature rules. The main member functions that quadrature classes define are

```
const std::vector<Point<dim>>& get_points() const;
const std::vector<double>& get_weights() const;
```

which respectively return vectors of the quadrature points and weights. Concrete quadrature rules are implemented in classes such as `QGauss` or `QMidpoints`.

Finite element basis functions are defined as piecewise polynomials within each cell. Different Galerkin bases are defined by the polynomial degree or the continuity across cell boundaries. The class template

```
template <int dim, int spacedim> class FiniteElement;
```

defines the interface for all finite element objects. Two of the key responsibilities of these objects are to evaluate shape functions and their derivatives at points within the unit cell. These member functions have the signature

```
double
  shape_value(unsigned int i, const Point<dim>& p) const;

Tensor<1, dim>
  shape_gradient(unsigned int i, const Point<dim>& p) const;
```

where the first argument is the component of the shape function in the case of vector-valued elements. With the aid of a Mapping object, the finite element classes only need to be able to evaluate shape functions within the unit cell. The most common implementation of the finite element interface used in icepack is the `FE_Q`

class, which implements the standard bilinear, biquadratic, etc. functions on quadrangles and hexahedra. A slew of other finite elements, for example, the bubble elements that are common in stable discretizations of mixed problems, are implemented in deal.II. Vector-valued finite elements can be constructed using the class `FESystem`, which aggregates together several `FiniteElement` objects for each dimension of the vector. In the most common case, one would use several copies of the same finite element, but one could also use a single `FE_Q` object of degree p and d `FE_Q` objects of degree $p + 1$ to represent the pressure and velocity for the Taylor-Hood element.

The triangulation class describes the geometry, and the finite element describes shape functions within a single cell of the geometry. The class `DoFHandler` uses both of these objects to define the entire set of Galerkin basis functions by associating to each geometric primitive a set of integer *degrees of freedom* based on the finite element. For example, when using bilinear elements on quads, each vertex of the geometry is associated to one degree-of-freedom. When using biquadratic elements, there is one degree-of-freedom for each vertex and one for each edge. When using vector-valued bilinear elements, there are m degrees of freedom for each vertex, where m is the number of fields in the vector-valued element. There is no natural, best way to assign each of these degrees of freedom an index from which to build an array representing the Galerkin expansion coefficients. For example, consider a vector-valued PDE in two space dimensions, such as the shallow shelf equations, and say that there are N vertices in the geometry. Should the x- and y-components of the ice velocity at vertex `i` be assigned the indices `2 * i, 2 * i + 1`, or should they be assigned to indices `i, N + i`? Likewise, if we use biquadratic elements, should the vertex and edge degrees of freedom be interleaved, or should all the edges be put at the end of the array and all the vertices at the beginning? In both examples, an argument could be made for either ordering depending on the exigencies of the problem at hand and, ideally, the library should provide both options so that the user can compare the performance. Once this choice has been made, however, it should not "leak out" to other parts of the program – they should be able to use a `DoFHandler` object under the assumption that it has defined an injective ordering without knowing anything more. The `DoFHandler` provides iterators over all the degrees of freedom within each cell of the geometry; in so doing, it acts as a surrogate for the actual geometry. The main loop of finite element matrix and vector assembly usually consists of iterating over all cells of a `DoFHandler`.

The final ingredient is the class template `FEValues`. The triangulation describes the geometry; the finite element describes shape functions in the unit cube; the quadrature object describes how to integrate functions within the unit cube; the mapping object transforms the unit cube to the physical cube; and the degree-of-freedom handler describes the mapping of geometric primitives to array indices. The `FEValues` class uses all of these objects to evaluate shape functions and their derivatives at quadrature points within the physical cells of the mesh.

The point of all of these classes is to be able to assemble finite element matrices and vectors. Wrappers for the Trilinos and PETSc libraries are included in deal.II, as well as its own vector and sparse matrix classes. Rather than describe the linear

algebra classes in detail, we refer the reader to the user guides for the PETSc and Trilinos (Balay et al., 2004; Heroux et al., 2005).

4.2 Design of icepack

Many of the building blocks for constructing larger modelling applications are provided in deal.II, but there were many more abstractions that I wanted to provide in icepack.

Every problem introduced in the deal.II tutorials uses a triangulation, a finite element, a degree-of-freedom handler, and a mass matrix. Rather than have to pass all these objects separately to the various functions that need them, I defined a class template `Discretization` that aggregates all of them together. Many functions in icepack involves synchronously stepping through a pair of `DoFHandler` iterators, one for scalar fields and the other for vectors fields. The `Discretization` class defines its own iterator that aggregates both the scalar and vector field degree-of-freedom iterators for this purpose.

The most fundamental objects in icepack are scalar and vector fields; these are the inputs and return types of all PDE solvers. Fields have a dimension and a tensor rank; the rank of a scalar field is 0, and the rank of a vector field is 1. I have not needed to implement fields of higher tensor rank so far. In the future, it may be useful to represent the strain rate field of some flow, which is a symmetric tensor field of rank 2. In any case, all ranks are implemented as a single class template:

```
template <int rank, int dim, /* other arguments */>
class FieldType : /* parent classes */
{
  /* methods */

protected:
  SmartPointer<const Discretization<dim>> discretization;
  Vector<double> coefficients;
};
```

This class template has one additional template argument and a superclass, which we have left off for now but will discuss next. A field consists of a pointer to a discretization, and the vector of its Galerkin expansion coefficients. The main responsibility of a field is to provide read access to the underlying discretization, and to provide general access to the expansion coefficients. Rather than have to write out `FieldType` for every use, icepack provides some more familiar aliases:

```
template <int dim> using Field = FieldType<0, dim>;
template <int dim> using VectorField = FieldType<1, dim>;
```

Finally, several convenience functions are provided for interpolating gridded data to a finite element representation; these functions return fields or vector fields depending on the input.

Several fields in finite element modelling are computed by integrating against shape functions. For example, when discretizing the prognostic equation

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x_i} u_i\right) h = \dot{a} - \dot{m}, \tag{4.1}$$

the n-th component of the numerical flux is

$$F_n = \int_\Omega \frac{\partial \phi_n}{\partial x_i} u_i h \, \mathrm{d}x, \tag{4.2}$$

where ϕ_n is the n-th shape function. If we use a forward Euler discretization for the time-dependence, the finite element expansion of H at time $k+1$ can be written as

$$MH^{k+1} = MH^k + \delta t \cdot F^k \tag{4.3}$$

where δt is the timestep and M is the finite element mass matrix. Note the extra factor of the mass matrix M – to obtain the ice thickness at the next timestep, we need to solve a linear system defined by the mass matrix. In an earlier version of icepack, I implemented a routine to compute the numerical flux with the following signature:

```
Field<2>
DepthAveragedModel::m_dh_dt(
  const Field<2>& h0,
  const Field<2>& a,
  const VectorField<2>& u
) const;
```

The prefix `m_` is to indicate that the resulting field is implicitly multiplied by the mass matrix, and to get the true flux the user would have to solve a linear system. This design, which relies on a naming convention, turned out to be error-prone. When the first unit test I wrote for the prognostic solve routine did not give the correct answer for an exactly-solvable problem, I spent several days trying to debug the flux computation when, in fact, I had forgotten to multiply by M^{-1}.

To solve this problem, I chose to make a distinction at the level of the type system that would catch these kinds of mistakes at compile-time. The invariant that can be encoded into the type system is the distinction between elements of Sobolev spaces and elements of their dual spaces. In finite element analysis, we think of fields and vector fields that we can evaluate pointwise within a domain Ω as elements of some Sobolev space $W_k^p(\Omega)$. Differential operators are mappings between this Sobolev space and its dual space $W_k^p(\Omega)^*$. For example, evaluating the Laplacian of a piecewise linear function in the strong sense is not defined, because a piecewise linear function does not have two continuous derivatives everywhere. On the other hand, for any v in the Sobolev space H_0^1, we can evaluate the dual pairing $\langle -\Delta u, v \rangle$ through the relation

$$\langle -\Delta u, v \rangle = \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x \tag{4.4}$$

since the gradients of u and v are in L^2. In other words, the weak Laplacian of a piecewise linear function is a perfectly well-defined element of the Sobolev space H^{-1} which is dual to H_0^1.

In order to enforce this distinction at the level of types, I first defined an enumerated type

```
enum Duality {primal, dual};
```

that takes only the values `primal` and `dual`. A variable of type `Duality` is the final template argument to the general `FieldType` class template:

```
template <int rank, int dim, Duality duality = primal>
class FieldType : /* parent classes */;
```

The `Duality` argument defaults to `primal`; most of the fields we are interested in are primal fields that we can evaluate pointwise. As a shorthand, I introduced some aliases:

```
template <int dim> using Field = FieldType<0, dim, primal>;
template <int dim> using DualField = FieldType<1, dim, dual>;
```

and likewise on for vector fields/dual vector fields. Any operation that involves taking derivatives of a field returns a dual field; any operation that involves solving a PDE takes in a dual field and returns a field. The L^2-inner product of two fields u, v is, in terms of their Galerkin basis expansions U, V,

$$\int_\Omega u \cdot v \, \mathrm{d}x = U^* M V. \tag{4.5}$$

This relation shows that, to convert a field to a dual field, one multiplies by the mass matrix; given a dual field f with Galerkin representation F, the dual pairing with v is

$$\langle f, v \rangle = F^* V. \tag{4.6}$$

Likewise, to convert a dual field back to a primal field, one multiplies by the inverse of a mass matrix. The conversion operations are implemented in the function `transpose`.

Once we have a convenient data type to represent fields and vector fields, it would be useful to also be able to manipulate them algebraically; for example, one ought to be able to write

```
Field<2> h = h0 + dt * dh_dt;
```

with the `+` and `*` operators overloaded to implement field addition and scalar multiplication. These operations on fields reduce to the equivalent vector space operations on their Galerkin expansion coefficients. If these operations are implemented naively, however, they result in gross inefficiencies. To see why, observe that, if the return type of `dt * dh_dt` is another object of type `Field<2>`, the line of code above will effectively be expanded to the following:

```
Field<2> dh = dt * dh_dt;
Field<2> h = h0 + dh;
```

Creating the temporary field dh will require allocating memory for its vector of Galerkin expansion coefficients, which might be large. Moreover, as the algebraic operation grows more complex, the number of unnecessary temporary objects will grow. We could instead explicitly write out all the looping code in every case to avoid a gross performance degradation, but in so doing we lose any shred of syntactic elegance. The C++ metaprogramming technique of *expression templates* allows one to write code that looks like the underlying algebra, but which compiles to the efficient looping code (Veldhuizen, 1995). No unnecessary temporary objects are created. The basic idea of expression templates is to encode the abstract syntax tree for an algebraic expression into nested C++ templates, which are then expanded at compile-time. The general type of expressions that evaluate to a field is a class template with the following signature:

```
template <int rank, int dim, Duality duality, class Expr>
class FieldExpr
{
  double coefficient(size_t i) const
  {
    return static_cast<const Expr&>(*this).coefficient(i);
  }

  /* other methods */
};
```

To encode the fact that any concrete field is a field expression, the FieldType class template inherits from FieldExpr, with itself as one of the template arguments:

```
template <int rank, int dim, Duality duality>
class FieldType :
  public FieldExpr<rank, dim, duality,
                   FieldType<rank, dim, duality>>
{
public:
  double coefficient(size_t i) const
  {
    return coefficients(i);
  }

protected:
  Vector<double> coefficients;

  /* other methods */
};
```

This design pattern is called the *curiously recurring template pattern*; its most common use is to implement static polymorphism (Coplien, 1995). The remaining im-

plementation of expression templates for fields closely mirrors the same implementation for vector and matrix algebra in other libraries; a description of this technique can be found in Vandevoorde and Josuttis (2003) and Iglberger et al. (2012).

The classes I have described so far are used to represent fields and vector fields, which are the input and output data to the model physics for all of our problems. The physics models themselves are implemented in classes that have methods for prognostic and diagnostic PDE solves. The prognostic solve method is similar for every depth-averaged model, so we can put this functionality into its own class `DepthAveragedModel`. Both the `IceShelf` and `IceStream` classes inherit from `DepthAveragedModel` to use its prognostic solve functionality without rewriting the same function in two different places. Since the diagnostic solve method for grounded ice streams requires more arguments than for ice shelves, it is implemented as a separate class. The signatures for these methods are all shown in figure 4.1. Each model object delegates evaluating the viscosity of the glacier to a class `Rheology`, which has methods for evaluating the rheology coefficient B given the ice temperature. The rheology class defaults to using the standard Glen flow exponent $n = 3$, but can nonetheless be customized to use any value of n. Likewise, the user can also set the nonlinearity in the equivalent class for basal friction. In future, I plan to generalize these classes so that the user can plug in his or her own physics models. For example, the user should be able to add an additional variable for basal water pressure to the physics determining basal friction.

For every physics model, the methods for solving the PDEs of glacier flow are pure functions that take in and return field objects. These field objects can then be compared with each other, manipulated algebraically, or otherwise analyzed in whatever way the user sees fit.

4.3 Testing PDE solvers

Writing software is extremely error-prone, so engineers rely on suites of automated tests to check for the validity of software libraries. Tests are usually classified into one of two categories: unit tests and integration tests. Unit tests check for the correctness of the smallest discrete unit of a complex software system – an individual function or class method. An example of a unit test might be to check that a trigonometric polynomial can be evaluated accurately for a range of values of the argument. Integration tests check for the correctness of the program or library as a whole. An example of an integration test might be to check that a library for spectral methods accurately solves a differential equation. To use spectral methods to solve differential equations, one must be able to accurately evaluate trigonometric polynomials, so the aforementioned unit test is a useful component of the overall software validation.

Given that software bugs are inevitable, a well-designed set of tests allows engineers to do a form of differential diagnosis akin to how a medical doctor assesses the likely cause of a disease. If a change to a software library breaks existing unit tests, we know the implementation is faulty right away. If the low-level unit tests are comprehensive and they all pass after a change, but an integration test fails,

```
class DepthAveragedModel
{
  public:
  Field<2> prognostic_solve(
    const double dt,
    const Field<2>& thickness,
    const Field<2>& accumulation,
    const VectorField<2>& velocity
  ) const;

  /* other methods */
};

class IceShelf : public DepthAveragedModel
{
public:
  VectorField<2> diagnostic_solve(
    const Field<2>& thickness,
    const Field<2>& rheology,
    const VectorField<2>& u_boundary
  ) const;

  /* other methods */
};


class IceStream : public DepthAveragedModel
{
public:
  VectorField<2> diagnostic_solve(
    const Field<2>& surface,
    const Field<2>& thickness,
    const Field<2>& temperature,
    const Field<2>& friction,
    const VectorField<2>& u_boundary
  ) const;

  /* other methods */
};
```

Figure 4.1: Signatures for several classes that implement depth-averaged glacier flow models in icepack.

we know that the change did not break any individual components but broke how two or more of these components interact. Without a hierarchy of tests, it is not immediately possible to diagnose where a failure occurred. For example, say that we had a suite of integration tests for a library that implements spectral methods but no function-level unit tests, and one of the integration tests fails. The failure might be because: the routines to evaluate trigonometric polynomials are faulty due to floating-point error; the linear algebra routines used to solve linear systems are faulty; or the both are correct, but the routines to assemble the linear system in the first place are wrong.

Several tests are included in icepack in order to check for failures of the various solvers. The most common way to test for the failure of a PDE solver is to find a geometry and input data for which the problem is exactly solvable, and compare the exact solution with the numerical one. For example, the tests for the diagnostic solvers in icepack compare the solution obtained from the diagnostic solve procedure with the exact solution of the shallow shelf equations in 1D, with constant rheology coefficient and linearly decreasing ice thickness. The exact and numerical solutions can be compared, but this only begs the question of how far apart the two must be before we decide that there is an error in the implementation. The error of the solution obtained from a correctly-implemented finite element method tends to 0 asymptotically as $\mathcal{O}(\delta x^2)$, where δx is the mesh spacing. In order to verify this behavior directly, we need to know what constant is hidden in this big-$\mathcal{O}$. We rarely know this constant even for linear PDE, much less for nonlinear problems like the shallow shelf equations. Instead, we could compute a sequence of numerical solutions with successively smaller values of δx, fit the errors to a parabola, and check that the linear term is close to 0. Implementing this idea is difficult in its own right, because the quadratic convergence is only asymptotic; for large enough values of δx, the curve may look more linear, spoiling the overall fit. Additionally, with too low a value of δx, we run into truncation errors. All in all, there is no principle we can use to guide us in what range of δx values to test, other than common sense.

These integration tests can also pass even if the individual units are implemented wrong. For example, my initial test for the inversion routines checked that they reproduce a known input rheology tolerably well, which they did. When I attempted to extend this analysis beyond synthetic problems to the Ross and Larsen C ice shelves, the optimization procedures often converged slowly, if at all. To diagnose what was wrong, I wrote a unit test to check that the gradient of the objective functional was being computed correctly by comparing with a finite difference approximation. Again, using a test like this this requires that one make an ultimately arbitrary decision of how far the computed gradient has to depart from the finite difference approximation before deciding that the method is wrong. This test revealed an error in my implementation, fixed in commit `f2f67f1c71`. Nonetheless, the incorrect gradient that I was computing was still a descent direction, so the optimization procedure would still converge for simple problems.

In all, testing numerical codes presents a host of difficulties that are not present in testing other kinds of software. Numerical codes do not fail discretely. For most programs, there are a set of mathematical invariants we can check about the code that must be satisfied at all times. If an invariant is broken at the end of program

execution, we can step through the program one instruction at a time in a debugger to find the precise moment when one of the invariants was violated. Not so for numerical codes. For example, the Ritz vectors computed in each stage of the Lanczos algorithm are orthonormal in infinite-precision arithmetic, but in floating-point arithmetic this invariant does not hold. After several iterations, the Ritz vectors can become linearly dependent and the algorithm breaks down, but there is no precise moment when the code goes wrong – errors are accumulated silently and slowly. Many errors in numerical codes are a result of an incorrect translation of a mathemtical principle into a programming language, and, short of a computer or a language type system that understands complex mathematics better than humans do, there is no way to statically verify the correctness of this translation. Some progress is being made in this direction, however, through the use of symbolic computing methods in finite element libraries such as FEniCS (Logg et al., 2012). Finally, many numerical codes are written to approximate some quantity, and rigorous error bounds on the approximation procedure are obtainable only for problems that are so trivial that numerical methods are unnecessary in the first place. A computer program, being limited to operating for some finite time, cannot rigorously verify whether an approximation procedure really does produce the limiting behavior we expect.

The fact that icepack can reproduce the exact solutions for simple glacier geometries gives some confidence in its correctness for more realistic cases. In §5, I will use icepack to explore the statistical methodology in data assimilation problems by inferring the rheology of ice shelves on simplified geometries. In §6, I will use icepack to study real ice shelves.

Chapter 5

Priors and regularization

In §3, we focused on inferring the shear stress underneath a grounded ice stream. Inferring the basal shear stress is complicated by the fact that we must also supply a rheology coefficient in order to solve the forward model. It is not possible in general to infer both the basal shear stress and the rheology coefficient; instead, there is a whole family of solutions lying along some curve in parameter space (Arthern and Gudmundsson, 2010). Inferring both fields is only well-posed with substantial regularization (Arthern et al., 2015). Instead, one can estimate the ice temperature with a model of heat flow through the glacier, which gives a baseline estimate of the ice rheology. A range of enhancement factors can then be tested to account for other factors that influence the rheology. Joughin et al. (2004b, 2003) used this approach to prescribe the rheology to infer the basal shear stress of the Ross ice streams; likewise, I used a temperature model from Poinar et al. (2015) to set the rheology in my work on Greenland outlet glaciers (Shapero et al., 2016). This approach is not entirely satisfactory because we have little way of knowing whether some feature of the basal shear stress is not in fact due to misspecification of the rheology field. Moreover, the ice temperature at the base depends on shear heating, which is what we were trying to estimate in the first place.

The experience of writing my first paper showed me that regularization of inference problems is a more important and complex issue than it is usually given credit for. Much of this paper was devoted to explaining the contrast between my own findings and those of Sergienko et al. (2014). We both used the full Stokes model, and the spatial resolution of the computational meshes in both cases were comparable. I was using the Robin inverse method, while Sergienko used the adjoint method, but Gillet-Chaulet et al. (2012) compared the two approaches and found the differences to be minor. Short of some unknown difference in the model inversion procedures used, the only remaining explanation for the discrepancy between my work and that of Sergienko is in how the problem was regularized. At the same time, Arthern et al. (2015) upended the conventional wisdom that held that the basal shear stress and rheology of a glacier cannot be inferred simultaneously, showing that both fields can be recovered if they are adequately regularized. Finally, Arthern also elucidated how the selection of a penalty functional can be

guided by maximum entropy principles, using ideas from Bayesian statistical inference (Arthern, 2015).

In this chapter, I will describe my work on inferring the rheology of a floating ice shelf. One of the chief goals of this work is to explore some of the methodological issues that arise in data assimilation problems. For example, most studies using Tikhonov regularization to infer basal shear stress all use the same penalty functional, namely the integrated square gradient of the unknown field. Is this penalty functional the best choice, or would some other penalty be more appropriate? Likewise, many published studies (my first paper among them) use the L-curve to select the regularization parameter. The L-curve is not the only way to select the smoothing length; for example, Habermann et al. (2012) advocate using the discrepancy principle. For the purposes of exploring methodological issues, inferring the rheology of an ice shelf is a much more straightforward problem than inferring the basal shear stress of a grounded glacier, since the rheology is now the only unknown field. Ice shelves are also fascinating physical systems in their own right. Drag at the side walls and from intermittent patches of grounding act as a mechanical buttress against the acceleration of the tributary glaciers that flow into the shelf (Dupont and Alley, 2005). Additionally, flow stripes and surface features in optical satellite imagery of ice shelves provide evidence for past variations in ice stream flow (Hulbe and Fahnestock, 2007). Data assimilation methods are also playing a crucial role in our evolving understanding of how crevassing affects the flow and buttressing effect of ice shelves (Borstad et al., 2016).

All of the simulations shown in this chapter used the library icepack, described in §4.

5.1 Ice shelf rheology

In chapter §2, we described how the strain rate $\dot{\varepsilon}_{ij}$ of a glacier depends on the flow rate parameter A and the deviatoric stress tensor τ_{ij} through the Glen flow law

$$\dot{\varepsilon}_{ij} = A\tau^2\tau_{ij}. \tag{5.1}$$

We now describe the many factors that influence the flow rate parameter. In Glen's original experiments on ice creep, he found an Arrhenius relationship between the flow rate parameter and the ice temperature T:

$$A(T) = A_0 e^{-Q/RT} \tag{5.2}$$

where R is the universal gas constant, and A_0 is an experimentally-determined constant with units of $\text{kPa}^{-3}\ \text{yr}^{-1}$, and Q is an experimentally-determined activation energy (Glen, 1955). (At temperatures below $-10°$ C, the activation energy is the energy per mole stored in two hydrogen bonds between neighboring water molecules in an ice crystal.) In ice shelves or fast-flowing ice streams, the controlling factor is the depth-averaged rheology coefficient

$$\bar{B} = \int_b^s B(T(z))\,\mathrm{d}z = \int_b^s A(T(z))^{-1/n}\,\mathrm{d}z. \tag{5.3}$$

The rheology B is a highly nonlinear and convex function of the ice temperature, so approximating the average of B by $B(\bar{T})$ would consistently under-estimate the depth-averaged rheology by the Jensen inequality. The temperature within the ice shelf depends on the temperature of the ice flowing into it, strain heating, and the exchange of heat with the atmosphere and oceans. When seawater freezes to the bottom of an ice shelf, forming marine ice, it releases latent heat to the ice; conversely, latent heat is absorbed from the shelf in areas of submarine melting. Rommelaere and Macayeal (1997) argue that low values of the inferred viscosity downstream of Crary Ice Rise are the signature of a strong temperature effect due to marine ice formation in this region.

During the last ice age, atmospheric dust concentrations were much higher than those of the present day. This particulate matter settled onto the surface of the Greenland and Antarctic ice sheets and was incorporated into the ice fabric as more snow accumulated. The rate factor A can be enhanced by a factor of roughly 2.5 in ice from the last ice age relative to ice from the present day (Paterson, 1991). The effect of impurities is to impede the growth of large ice crystals. Both impurities and anisotropy due to fabric development are thought to be important in explaining higher strain rates in ice accumulated during the last ice age (Thorsteinsson et al., 1999).

The Glen flow law that we have used so far is isotropic. At the scale of an individual ice crystal, however, the flow parameter is higher along the crystal c-axis (Cuffey and Paterson, 2010). In a larger ice block consisting of many crystals, the c-axes are often randomly aligned, giving a net isotropic response to applied stress. Sustained deformation along one axis for a sufficiently long enough time can make the c-axes align perpendicular to the principal axis of deformation, giving a net anisotropic flow; this has been observed deep in ice cores (Thorsteinsson et al., 1997). Nonetheless, there is no widely agreed-upon mathematical model for the rate at which anisotropic crystal fabrics develop, and how they influence the large-scale glacier flow. Numerical ice sheet models have begun to incorporate anisotropic effects relatively recently (Gagliardini et al., 2013).

When a crevasse forms, the ice on either side of the fracture becomes mechanically decoupled. At large scales, the net effect of fractures is to induce a rheological weakening of the glacier, but there is currently no widely-agreed upon mathematical model for precisely how crevasses and rifts affect the overall glacier flow. The relevant length scales for an individual crevasse are typically on the order of tens of meters (Van der Veen, 1998), while a typical mesh spacing in numerical glacier flow models is on the order of 250 m or more. This contrast of scales means that accurately modelling fracture initiation and growth is not generally feasible. If we could instead use a coarse-graining or homogenization approach, the net effect of crevasses could be summed up in some scalar field D that can be represented accurately at the 250-m scales of our discretization rather than the 10-m scales of an individual crevasse. Several studies have examined what such a coarse-grained model might look like using the theory of *continuum damage mechanics* (CDM) (Borstad et al., 2012; Albrecht and Levermann, 2012; Krug et al., 2014). The *damage field* takes values between 0 and 1, with 0 indicating intact ice and 1 indicating completely broken ice with no mechanical cohesion. Lumping all other sources of flow

enhancement (impurities, anisotropy, etc.) into a single factor E_0, the enhancement factor for a given damage level D is

$$E = \frac{E_0}{(1-D)^n}. \tag{5.4}$$

Since $B = (E \cdot A(T))^{-1/n}$, the coarse-grained rheology field is

$$B(T, D) = (1-D) \cdot E_0^{-1/n} \cdot A(T)^{-1/n}. \tag{5.5}$$

Without additional constraints on the temperature of the ice shelf, it is not possible to determine the damage field from the rheology alone. This situation is reminiscent of extricating the influences of hydrology and geology when inferring the basal shear stress.

The rheology of a glacier can be inferred from observational data using the same adjoint method formalism as for the basal friction coefficient. Since the rheology is chiefly determined by the ice temperature and damage state, spatial variations in the inferred rheology of an ice shelf can tell us about the various outside forces at play. Rommelaere and Macayeal (1997) first used control methods to ascertain the viscosity of Ross Ice Shelf from velocity data collected during the RIGGS campaign (Bentley, 1990). Rommelaere and MacAyeal observed that the viscosity of the shelf was much lower than the spatial average in the wake of Crary Ice Rise. Since the viscosity decreases at higher ice temperatures, they took this low viscosity to be an indication of latent heat release from the formation of *marine ice* – ice formed on the bottom of the shelf from refrozen ocean waters. Additionally, they found a region of high viscosity in the center of the ice shelf. Since this anomaly is far from any coastal feature such as Crary Ice Rise, they argue that it is most likely a relic of ice stream variability 1000 years ago, when this ice was entering the shelf from Ice Stream A or B. The idea of using features of an ice shelf to explore the history of its tributary ice streams was later explored in Fahnestock et al. (2000) and Hulbe and Fahnestock (2007).

Larour et al. (2005) computed the rheology of the Ronne Ice Shelf, improving on Rommelaere's technique by inferring the rheology coefficient $\bar{B}$ rather than the viscosity. The inferred rheology from their paper is shown in figure 5.1. Larour et al. find variations variations in the rheology from 300 – 900 kPa $\text{yr}^{1/3}$. They note that, if effects from damage are ignored, this range of $\bar{B}$ suggests column-averaged temperatures ranging from -6.7° C down to -31° C. The influences of temperature and damage on ice rheology were both considered in this study. For example, they argue that the strong ice between Korff and Henry ice rises (location A on figure 5.1) is due to both low surface temperatures and marine ice formation in this region, while the weak ice in the wakes of these ice rises (locations B and C) is attributed to substantial crevassing discernible from satellite imagery. Reduced ice stiffness can also be seen in the southern margins of Evans and Rutford ice streams; Larour et al. attribute this weakening to strain heating.

From January to April 2002, Larsen B Ice Shelf on the Antarctica Peninsula disintegrated almost in its entirety (Rack and Rott, 2004; Scambos et al., 2004). During the austral summer of 2001/2002, Larsen B was subjected to an unusually

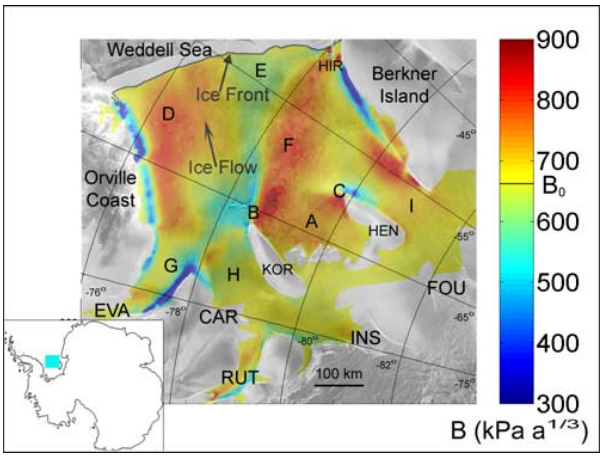


Figure 5.1: Rheology of the Ronne Ice Shelf from Larour et al. (2005). Abbreviations are: Evans Ice Stream (EVA), Carlson Inlet (CAR), Rutford Ice Stream (RUT), Foundation Ice Stream (FOU), Institute Ice Stream (INS), Hemmen Ice Rise (HIR), Korff Ice Rise (KOR), Henry Ice Rise (HEN).

large degree of surface melting; these melt ponds are thought to have caused hydrofractures which ultimately resulted in the collapse of the ice shelf (Scambos et al., 2003). These events prompted numerous studies of the stress state of Larsen B from before its breakup using data assimilation methods, the first of which were Vieli et al. (2006, 2007). Vieli and Payne found that observed changes in the velocity of Larsen B from 1995 to 1999 can only be explained by significant rheological weakening of the ice shelf, the relative changes of which reached 50% in some locations. They argue that such dramatic changes over the course of only four years could result only from mechanical decoupling through the formation of crevasses and rifts, rather than warming of the ice through meltwater ponding. From a methodological standpoint, the two studies by Vieli and Payne are remarkable for several reasons. First, they do not make the usual conceit that the observational data are dense enough to constitute a field which can be evaluated at any point in space; rather, their objective functional consists only of a discrete sum of misfits with observations. Additionally, most studies assume that we know both components of the velocity vector. Two of the three velocity data sets used in this study were obtained with just one descending satellite orbit, so the ice velocity is only resolved in the satellite look-direction. This single component of the ice velocity can still be compared with the projection of the modelled ice velocity onto the look-direction; in doing so, Vieli and Payne were able to constrain the inference problem using what might otherwise be considered irredeemably bad data. Finding that the data are too sparse around the boundary of the ice shelf, they take the unusual step of optimizing for the ice inflow velocities that give the best match to observations in the interior as well as the ice rheology. Finally, they penalized values of the rheology

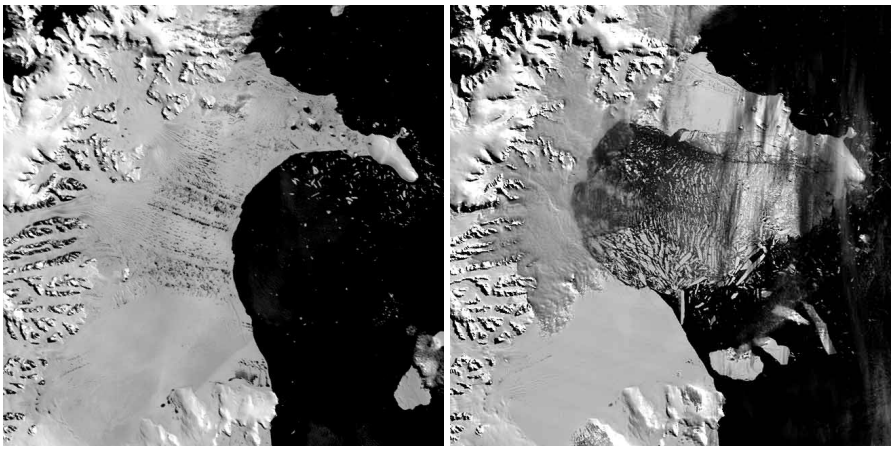


Figure 5.2: Images of Larsen B Ice Shelf on 31 January 2002 (left) and 5 March (right) from the NASA Moderate Imaging Resolution Spectroradiometer (MODIS) satellite. Reproduced from the National Snow and Ice Data Center, https://nsidc.org/news/newsroom/larsen_B/2002.html.

field that would correspond to unrealistically cold ice (below -18° C) in addition to the usual Dirichlet energy used to regularize the inference problem.

Khazendar et al. (2007) applied the same methods to Larsen B, but using the RADARSAT-1 InSAR velocity maps collected in 2000 (Jezek, 2002). Vieli et al. (2006) consider the rheology fields they infer in the center of Larsen B to be poorly constrained due to the high errors (40 m/yr) of the InSAR velocity maps they used. The much lower velocity errors (5-10 m/a) of the RADARSAT velocity maps permitted substantially more accurate estimates of the rheology throughout the whole ice shelf. Khazendar et al. found that, but for one exception, the locations where the ice shelf disintegrated agree well with the weakest regions of the ice shelf. This study is also remarkable for the effort they put in to validating the rheology field they infer through forward modelling. Given some guess for the rheology, one could model the ice shelf evolution for the next several years to see how well this agrees with observations. Khazendar et al. used the inferred rheology from 2000 and a spatially-constant rheology field to initialize a forward model run from 1996 velocity and thickness data. The agreement with observed ice speed changes was substantially better for the ice rheology they inferred using control methods. The demise of Larsen B prompted concern that its larger neighbor, Larsen C, would meet a similar fate. Khazendar et al. (2011) sought to evaluate the risk of catastrophic collapse of Larsen C using similar methods to their previous studies of Larsen B. While they found consistently weaker ice downstream of promontories and ice rises, Larsen C does not exhibit the same degree of mechanical decoupling across the suture zones between distinct flow units that Larsen B did before its collapse.

Data assimilation methods play a crucial role in Borstad et al. (2012) and Borstad

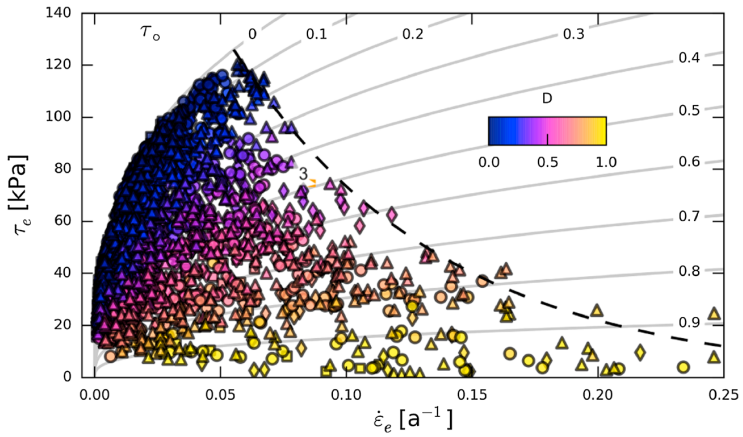


Figure 5.3: Inferred effective stress vs. strain rate at several points throughout the Larsen B Ice Shelf, Antarctica, reproduced from Borstad et al. (2016). The colors indicate the value of the damage field D.

et al. (2016). Their work adopts the CDM formalism to describe the combined effects of temperature and damage on the ice rheology. Without a priori knowledge of either the temperature or damage field, it is not possible to distinguish one field from the other by assimilating a single set of observational data. We do, however, have observations of the temperature of the atmosphere and ocean around Antarctica; by using regional climate and ocean circulation models, we can estimate the temperature at the surface and base of an ice shelf. These form the boundary conditions we need to solve the heat equation within the ice shelf. Borstad et al. (2012) used surface temperatures from RACMO and ocean temperatures from MITgcm and solved the heat equation analytically in vertical columns, with a correction applied to account for the advection of cold ice from upstream. The uncertainties in their prescription of the ice temperature lead to an uncertainty in the ice rheology from 7-17%. Let $\bar{B}(T)$ be the rheology coefficient computed from this modelled temperature field, and B the rheology inferred from observational data. Assuming that any part of the inferred rheology that is not explained by temperature is due to damage, one can then calculate the damage field as

$$D = 1 - B/\bar{B}(T). \tag{5.6}$$

They used this technique to infer the damage field of Larsen B Ice Shelf, finding that the areas of highest damage coincide with the area of catastrophic ice shelf breakup.

This analysis was extended in Borstad et al. (2016). No matter how damage and temperature are distinguished from each other in a data assimilation procedure for the ice rheology, the stress τ and strain rates $\dot{\varepsilon}$ can be estimated from the output of the model. Borstad et al. found that, when plotted against each other, these

quantities are bounded by an envelope at high strain rates; this is shown in figure 5.3. In effect, Glen's flow law – that the ice stress τ is proportional to $\dot{\varepsilon}^{1/3}$ – is valid only up to a certain threshold stress, after which the effects of fracture take over. From the shape of the envelope shown in figure 5.3, they assume that the threshold stress curve for a given value of the strain rate has the shape

$$\tau_e = \tau_0 \exp\left(-\frac{\dot{\varepsilon}_e/\dot{\varepsilon}_0 - 1}{\kappa - 1}\right) \tag{5.7}$$

where the free parameters τ_0 and κ can be fit from observations. The catastrophic breakup of Larsen B and the incipient extremes of stress and strain rate make it an especially attractive site to study this problem.

The approach of Borstad et al. has some limitations. The ability to diagnose the damage field is contingent upon having a good estimate of the ice temperature, without which the rheology could be mistakenly attributed to one effect when the other is responsible. For example, marine ice formation in suture zones on the ice shelf is thought to be important in maintaining mechanical stability of Larsen C (Holland et al., 2009). The accretion of marine ice onto the bottom of the shelf will fill and heal bottom crevasses, reducing the damage field while warming the ice column. For certain combinations of parameters, the increase in temperature of the ice shelf could exactly match the reduction in the damage state in such a way that the rheology is constant. Borstad et al. (2012) hypothesize that calving occurs when the damage field is above a threshold of ≈ 0.6, so a misspecification of the temperature in the suture zones could be the difference between predicting a large calving event or not.

The stress/strain-rate envelope has far-reaching implications for how the effect of damage should be incorporated into models of ice shelf flow, but the shape of this envelope also depends on the temperature used to initialize the inverse method. In Borstad et al. (2016), the envelope shape is fit using an inversion with a constant ice temperature of -7.5°C. The same plot as figure 5.3 is shown in the supplementary information to this paper using the temperature field estimated from regional climate and ocean models, as in Borstad et al. (2012). This plot is reproduced in figure 5.4. The shape of the stress/strain envelope is noticeably different when using the temperature field obtained by solving the heat equation. Additionally, in both the 2012 and 2016 papers, the depth-averaged rheology is written in terms of the depth-averaged temperature, i.e. it is assumed that $B(\bar{T}) \approx \bar{B}(T)$. They argue that the error associated with this approximation is negligible compared to other sources of error. On the other hand, the function $B(T)$ is both convex and highly nonlinear, so this approximation introduces a systematic bias: $B(\bar{T}) < \bar{B}(T)$. In that case, the part of the ice rheology due to temperature will be consistently under-estimated, and the value of the damage field will then be overestimated.

The capacity to estimate the temperature in the first place is contingent on the assumption that the ice velocity and temperature have been roughly in steady state since the time that the ice at the shelf terminus passed the grounding line. The Ross Ice Shelf has gone through appreciable changes over the past millennium as a consequence of the periodic stagnation and reactivation of its ice streams (Hulbe and Fahnestock, 2007). The first study of the viscosity of the Ross Ice Shelf using

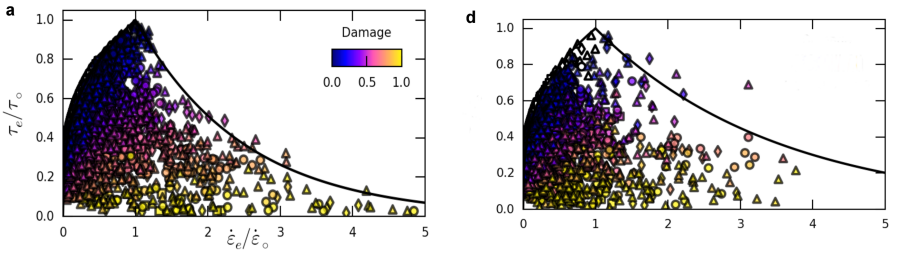


Figure 5.4: Figure from the supplementary information of Borstad et al. (2016) showing the stress/strain obtained with a constant temperature (a) and a spatially varying temperature (b).

inverse methods found a region of hard ice that the authors attribute to variability in the flow of ice streams A and B (Rommelaere and Macayeal, 1997). The steady-state assumption might then be a poor fit for the Ross shelf. Less is known about the flow history of Larsen B, beyond the fact that it has been stable during the current geological epoch (Domack et al., 2005). Nonetheless, it may be impossible to unambiguously distinguish between the damage field and temperature anomalies caused by variations in ice shelf flow from centuries past. Borstad et al. (2012) apply an ad-hoc correction to the temperature to account for advection of cold ice from upstream. One could instead assume that only the ice velocities are in steady state and attempt a spin-up of the shelf temperature. Arguably, this approach introduces even more under-constrained quantities into the problem than we had before.

Finally, Borstad's work provides a way for diagnosing the damage state of a floating ice shelf, but not for predicting how the damage will evolve. Borstad's results establish what the stress/strain-rate/damage state space looks like, but they do not prescribe how a parcel of ice will traverse this state space. Nonetheless, the envelope provides a valuable constraint on models of damage evolution. In Borstad et al. (2016), the authors argue that the evolutionary model in Albrecht and Levermann (2012) is ad-hoc and lacks a physical basis. Any putative model for the time-dependence of the damage field should, in Lagrangian coordinates, reproduce the exponential stress/strain-rate envelope.

5.2 Temperature and damage evolution

The temperature of an ice shelf is transported mainly by advection rather than diffusion, and the proposed models for damage evolution also share this feature. The characteristics of this transport are the main motivation for the use of total variation regularization, which we will describe in section §5.3.

Ignoring melting and freezing effects, the 3D heat equation is

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x_i} u_i\right) \rho c_p T = \frac{\partial}{\partial x_i} k \frac{\partial T}{\partial x_i} + \dot{\varepsilon}_{ij} \tau_{ij}, \tag{5.8}$$

where k is the thermal conductivity. The first term on the right-hand side of this equation is the transport of heat via diffusion, and the second term is the amount of strain heating due to deformation within the ice. The diffusivity $\kappa = k/\rho c_p$ of ice is roughly 35 m^2 yr^{-1}. The Ross Ice Shelf is on the order of 1000 km across, with ice speeds on the order of 1 km/yr. This gives a residence time of roughly 1000 years, for which the diffusive length scale is

$$L_{\text{diff}} = \sqrt{\kappa \cdot L_{\text{adv}}/U} \approx 200 \text{ m}. \tag{5.9}$$

This length scale is minuscule compared to the advective length scale of 1000 km, so the effect of horizontal diffusion can be safely ignored. For comparison, a typical grid cell diameter numerical simulations of ice shelf flow is on the order of 5 km or more (Rommelaere and Macayeal, 1997; Larour et al., 2005). On the other hand, the thickness of the Ross Ice Shelf throughout much of its central region is around 400 m, which is comparable to the diffusive length scale.

Diffusive transport tends to smooth out sharp temperature gradients, while advective transport mostly leaves these gradients intact. Consequently, we can expect that, aside from forcing due to heat exchange with the ocean, any variations in temperature across the ice grounding line are largely preserved as the ice flows toward the shelf calving front.

The proposed models for the evolution of the damage field are all completely advective, with no diffusion whatsoever (Albrecht and Levermann, 2012; Krug et al., 2014). The damage field is passively transported by advection, with source terms for rift opening and closing:

$$\left(\frac{\partial}{\partial t} + u_i \frac{\partial}{\partial x_i}\right) D = r - c. \tag{5.10}$$

Let τ_+, τ_- be the large and small eigenvalues of the stress tensor, and likewise for the strain rate. Rifts open in the glacier when the von Mises stress $\tau = \sqrt{\tau_+^2 + \tau_-^2 - \tau_+\tau_-}$ exceeds some threshold stress τ_r:

$$r = \begin{cases} \gamma_r \dot{\varepsilon}_+ (1 - D) & \tau \geq \tau_r \\ 0 & \tau < \tau_r \end{cases} \tag{5.11}$$

where γ_r is some constant. Similarly, damage healing occurs when the principal strain rate drops below a threshold $\dot{\varepsilon}_c$:

$$c = \begin{cases} 0 & \dot{\varepsilon}_+ \geq \dot{\varepsilon}_c \\ \gamma_c (\dot{\varepsilon}_c - \dot{\varepsilon}_+) & \dot{\varepsilon}_+ < \dot{\varepsilon}_c \end{cases} \tag{5.12}$$

Not only is the transport of damage entirely through advection, the rifting source term is a discontinuous function of the ice stress. These models for damage evolution are not based on fundamental physical principles as with the heat equation, but rather on how the macroscopic fracture density field might evolve when viewed at some coarse scale Krug et al. (2014). They have not been rigorously tested against

long-term observational time series of ice shelf flow to evaluate whether they reproduce measurements under some assignation of the free parameters. Nonetheless, individual fractures are transported along with the ice flow, so it is likely that any large-scale damage evolution model will be predominantly advective, even if the form of the source terms is found to be different from those shown in equations (5.11), (5.12).

5.3 Total variation

The preceding discussion strongly suggests that the temperature and damage fields in an ice shelf both have very sharp features. These might be outright jump discontinuities, or steep gradients that cannot be distinguished from a jump at the resolution of a numerical model. Nevertheless, in most of the literature on estimating ice shelf rheology from observations, the data assimilation methods heavily favor smooth solutions. Several studies use a spatially constant initial guess for the rheology and an early-stopping criterion rather than Tikhonov regularization (Rommelaere and Macayeal, 1997; Larour et al., 2005; Khazendar et al., 2007, 2009, 2011). Borstad et al. (2013) also use an early-stopping criterion for inferring the damage field of the Larsen C ice shelf, but find a better fit to the data when using a spatially variable initial guess. While early-stopping criteria have many advantages, dependence of the results on the initial guess introduces a degree of arbitrariness into the procedure; using a smooth initial guess will likely yield a smooth solution. The same group of authors changed their methodology in more recent work, using the Dirichlet energy as a regularization functional and selecting the smoothing length with the L-curve (Borstad et al., 2016). This paper even goes so far as to explicitly state, "...we want to capture sharp gradients in flow."

To test whether sharp gradients can be captured, we will use a synthetic inverse problem posed on a rectangular ice shelf of length L and width W. The ice thickness decreases linearly from h_0 at the inflow boundary $x = 0$ to $h_0 - \delta h$ at outflow. The ice temperature is set to

$$T^{\text{true}} = T_0 + \begin{cases} \delta T & L/4 < x < 3L/4, \quad W/4 < y < 3W/4 \\ 0 & \text{otherwise} \end{cases} \tag{5.13}$$

where T_0 is the background temperature and δT is the temperature jump over the interface. The rheology is computed from this temperature using the usual parameterization, from which we compute the ideal velocity u^{true} by solving the diagnostic equations. A plot of the temperature and velocity fields is shown in figure 5.5. In real problems, we of course do not get to see the true velocity but some observation $u^o = u^{\text{true}} + \xi$ for some noise component ξ. For InSAR measurements of real glaciers, the standard deviation σ of the measurement errors is on the order of 10 m/a or less, which we will use to set the noise level. We can then evaluate the efficacy of an inverse method using the Dirichlet energy as regularization functional by treating the observed field u^o as the observations and comparing the inferred rheology or the incipient temperature T with the true value. In order to make

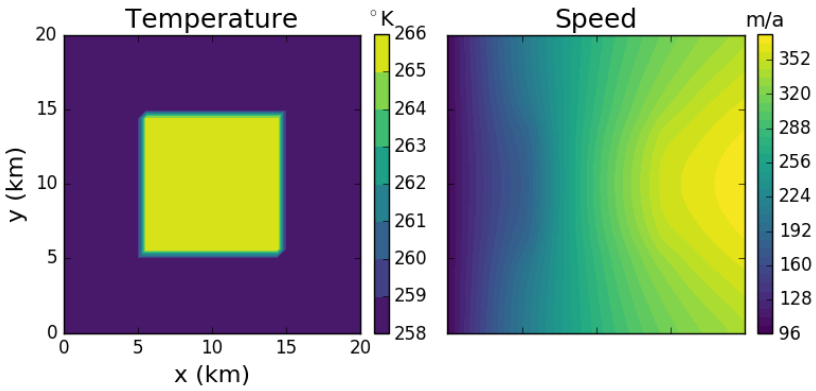


Figure 5.5: Exact temperature and ice speed for the synthetic ice shelf rheology inference problem.

the synthetic problem as close as possible to the real one, we will also evaluate the best choice of the regularization parameter α using the L-curve. The results of this synthetic problem are shown in figure 5.3. The best value of the smoothing length was roughly 3.25 km. While much of the broad-scale features of the true temperature field are recovered, the sharp edge is smoothed over even in the under-regularized case where the smoothing length was 950 m.

This example shows that, when penalizing the rheology inverse problem using the Dirichlet energy, sharp gradients cannot be accurately recovered. The natural question to ask now is whether we can devise a different functional that still penalizes highly oscillatory fields, but that does not penalize the occasional jump discontinuity. As we will show in the following, the *total variation* (TV) functional is one possible choice of penalty which fulfills these criteria.

The first use of total variation regularization was in Rudin et al. (1992) for image denoising problems. In image denoising, one is given the pixel values u^o_{ij} of some image which has been corrupted by measurement noise, and we would like to filter out the instrument noise while preserving as much of the true image as possible. The objective functional for this inverse problem is

$$J(u) = \|u - u^o\|^2 + \alpha^p R(u) \tag{5.14}$$

where R is the regularization functional. In this case, the forward map G from parameters to observations is just the identity map. Image denoising is an important preprocessing step in many computer vision applications, for example, in optical character recognition – reconstructing the sequence of characters of a written work from an image of the printed text. An obvious choice is to take R to be the Dirichlet energy, in which case the objective functional is quadratic and the resulting linear problem to be solved is a discretization of an elliptic PDE. Real images, however, can have sharp contrasts in hue or luminance across pixel boundaries. For example,

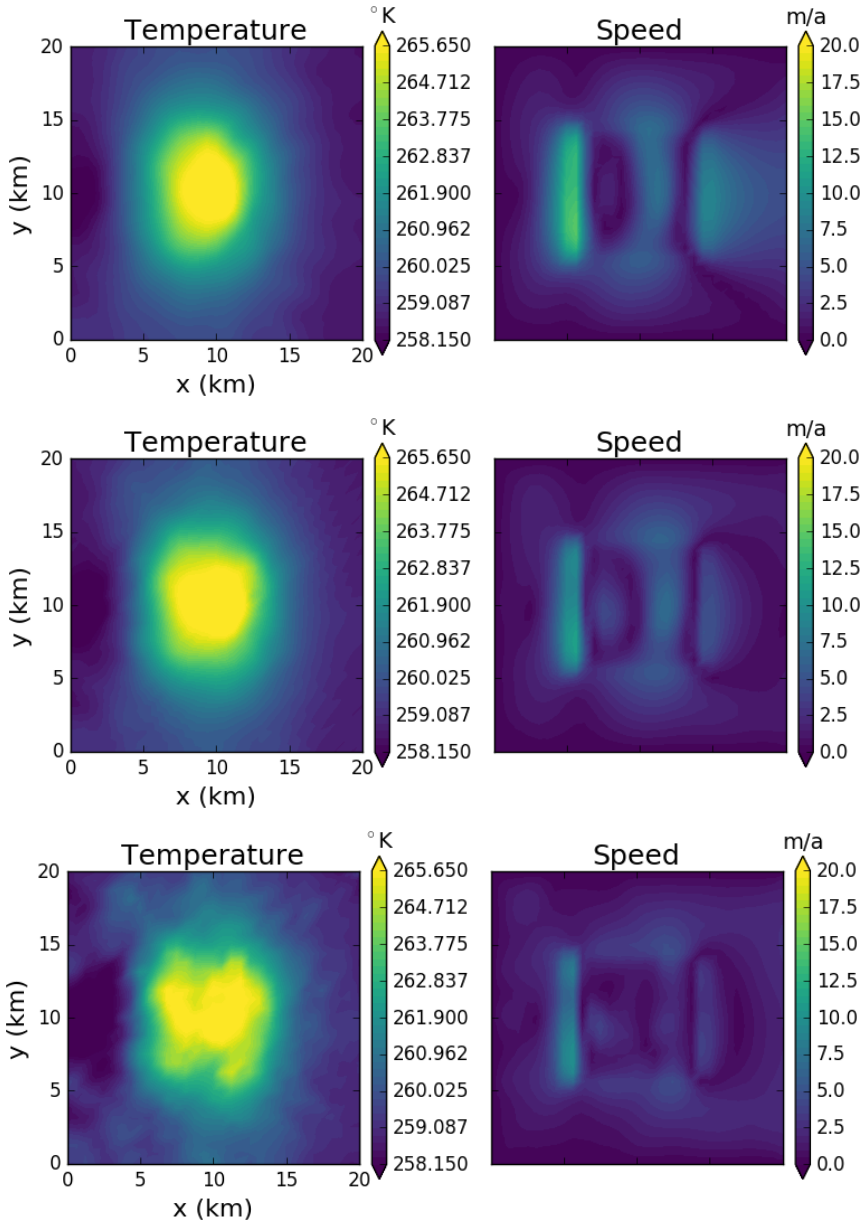


Figure 5.6: Inferred temperature and velocity residuals obtained with smoothing lengths of 6.5 km (top), 3.25 km (middle), and 950 m (bottom).

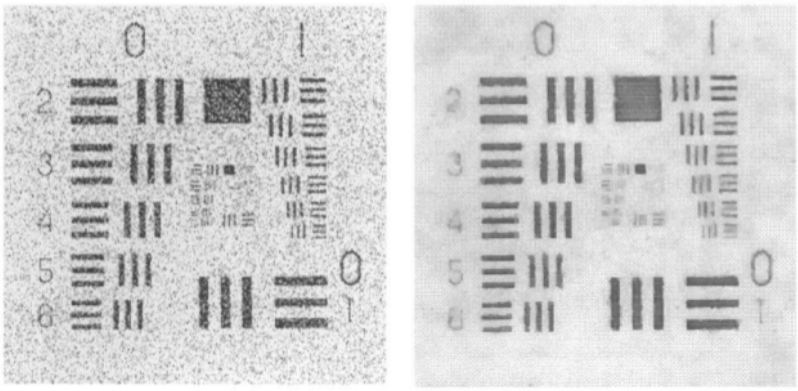


Figure 5.7: Noisy image (left) and total-variation reconstruction of the image (right), reproduced from Rudin et al. (1992)

in optical character recognition, there is a sharp contrast between the white background and black characters. Using the Dirichlet energy as a penalty functional will eliminate the noise, but it will have the undesirable consequence of also smoothing out the contrast between genuinely distinct regions of the image. Rudin et al. (1992) found that the total variation functional

$$R(u) = \int_\Omega |\nabla u| \, \mathrm{d}x \tag{5.15}$$

is also successful at penalizing measurement noise in imaging problems, but with the convenient property that sharp edges were left intact.

The condition that the total variation (TV) of a field is finite is strictly weaker than the condition that the Dirichlet energy is finite by the Cauchy-Schwarz inequality. We can get a better feel for the properties of the total variation functional by examining the simple case of the characteristic function of some smooth domain. Let γ be a closed curve in 2D, and let

$$u(x) = \begin{cases} a & x \text{ inside } \gamma \\ 0 & \text{otherwise} \end{cases}. \tag{5.16}$$

Then the total variation of this field is

$$R(u) = a \cdot \text{length}(\gamma). \tag{5.17}$$

By contrast, the Dirichlet energy of this field is infinite due to the jump discontinuity.

More generally, let u be a Lipschitz-continuous function; given some real number s, let γ_s be the s-level set of u:

$$\gamma_s = \text{boundary}(\{x : u(x) > s\}). \tag{5.18}$$

Then the total variation of u can be expressed as

$$\int_\Omega |\nabla u| \, \mathrm{d}x = \int_{-\infty}^{\infty} \text{length}(\gamma_s) \, \mathrm{d}s. \tag{5.19}$$

This is the famous *coarea formula* (Federer, 1959). Given some function, we can draw all of its contour lines in the plane; the coarea formula states that the total variation of a field is the integral of the lengths of all of these contour lines. The coarea formula shows that the total variation functional does not penalize sharp transitions as such. Rather, it penalizes two qualities, the combination of which makes it an effective regularization functional when the measurement errors are white noise. To illustrate these qualities, I have plotted a Gaussian distribution in 2D in figure 5.8, which is slightly elongated along one axis, with the contour lines shown below the surface plot. First, a large jump is more TV-expensive than a small jump, irrespective of the width of the interface over which this jump occurs. The middle panel shows a perturbation of the Gaussian distribution that has smaller TV because the peak height has been shrunk. Second, a jump of fixed height a across a curve γ is more TV-expensive the larger the length of γ. In the bottom panel of figure 5.8, I have shown a perturbation of the Gaussian distribution where the area inside every contour line has been preserved, but the contour lines have been relaxed to be more circular than elliptical. Since the area inside each contour line is the same, the total volume under the surface is the same; but the length of each contour line has been reduced, so the total variation of this function is smaller by the coarea formula.

Another way of understanding why TV regularization works is through its expression in terms of the *wavelet* expansion of the field. If ϕ is expanded in terms of the Haar wavelet ψ as

$$\phi(x,y) = \sum_{k,l\geq 0} \sum_{m=0}^{2^k-1} \sum_{n=0}^{2^l-1} 2^{\frac{k+l}{2}} \phi_{kl,mn} \psi(2^k x - m) \cdot \psi(2^l y - n), \tag{5.20}$$

then the total variation of ϕ is the ℓ^1-norm of its wavelet coefficients:

$$\int_\Omega |\nabla \phi| \, \mathrm{d}x = \sum |\phi_{kl,mn}| \tag{5.21}$$

(Daubechies et al., 2004). Penalizing the ℓ^1-norm of the solution of regression problems favors sparsity, i.e. the computed regression coefficients are mostly 0 (Aster et al., 2013). Equation (5.21) shows that TV regularization promotes sparsity in the wavelet basis. Real glaciers, however, have irregular shapes, so coming up with a basis of spatial wavelets for our domain is impractical. Consequently, the expression of TV in terms of wavelets is not directly applicable to our inference problems.

The total variation functional is only Fréchet-differentiable in the neighborhood of a field ϕ for which $|\nabla \phi|$ is bounded away from 0. We can circumvent this problem by using the convex conjugate (Vogel, 2002). The total variation of a field ϕ can

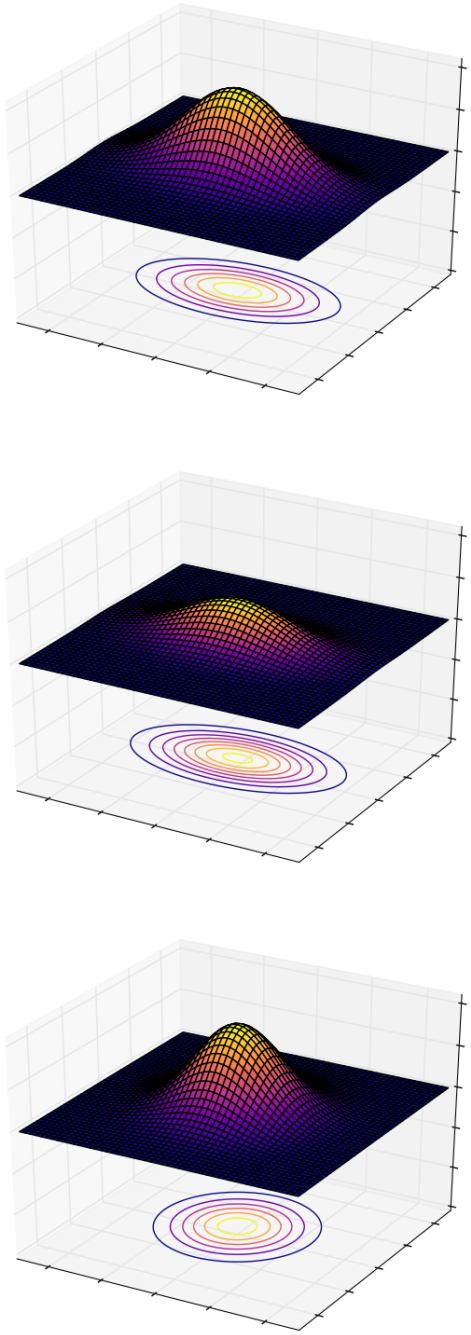


Figure 5.8: A Gaussian distribution in 2D (top), with two perturbations that reduce its total variation; shrinking the peak height (middle) and making the contour lines more circular (bottom).

be expressed as the solution of an extremization problem over the set of all vector fields v such that $|v(x)| \leq 1$ for all x:

$$R(\phi) = \int_\Omega |\nabla \phi| \, \mathrm{d}x = \max_{\|v\|_\infty \leq 1} \int_\Omega \phi \nabla \cdot v \, \mathrm{d}x. \tag{5.22}$$

Introducing the dual variable v, our non-differentiable minimization problem for ϕ is turned into a differentiable problem for the pair of variables ϕ, v. The extra cost is that the problem has the non-holonomic constraint $\|v\|_\infty \leq 1$, and that we are now seeking a saddle point instead of a minimum.

Alternatively, we can introduce a parameter γ with which to construct a smooth approximation of the true TV:

$$R(\phi; \gamma) = \int_\Omega \left\{ \sqrt{|\nabla \phi|^2 + \gamma^2} - \gamma \right\} \mathrm{d}x. \tag{5.23}$$

In robust statistics, one often uses such smoothed functionals to approximate the ℓ^1-norm of a vector; this is called the *Huberized* loss function, after the statistician Peter Huber. By analogy with the loss function in statistics, we will refer to the functional in equation (5.23) as the Huber total variation (HTV). The HTV functional is differentiable for any ϕ, and its derivative is

$$\left\langle \frac{\mathrm{d}R}{\mathrm{d}\phi}, \psi \right\rangle = \int_\Omega \frac{\nabla \phi \cdot \nabla \psi}{\sqrt{|\nabla \phi|^2 + \gamma^2}} \, \mathrm{d}x. \tag{5.24}$$

In the limit as $\gamma \to 0$, this smoothed functional approaches the true total variation. Unlike the true TV, for fields ϕ such that $|\nabla \phi| \ll \gamma$, the HTV of ϕ is approximately equal to the Dirichlet energy. This approximation is pictured in figure 5.9.

The temperature field inferred with HTV regularization is shown in figure 5.10. While the sharp transition is blurred somewhat due to the extra factor γ, the overall boxy nature of the transition is preserved better with the HTV functional than with the Dirichlet energy, as shown in figure 5.3. That said, the HTV-penalized solution captures the upstream side of the interface far better than it does the downstream or along-flow sides. The HTV functional may give better results than the Dirichlet energy, but there is still room for improvement; it is no silver bullet.

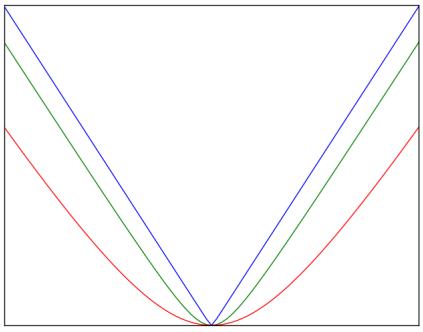


Figure 5.9: The true total variation in blue, and two Huberized approximations in green and red.

One feature especially worth noting is that the discrepancy principle selects a much smaller value of the regularization parameter ($\alpha = 1.4$ km) for HTV regularization than for the Dirichlet energy ($\alpha = 3.25$ km). In both cases, the velocity

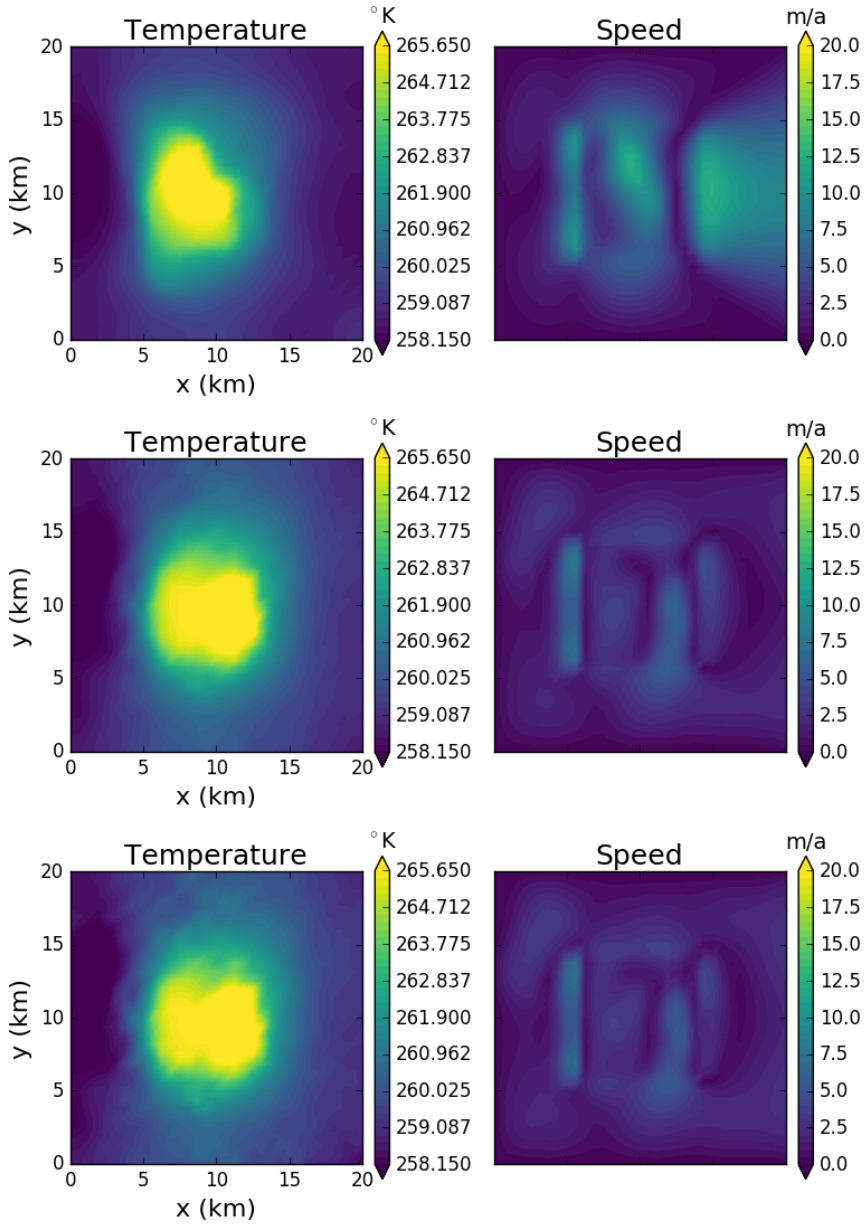


Figure 5.10: Inferred temperature and velocity residuals obtained using the Huber total variation functional and smoothing lengths of 6.5 km (top), 1.4 km (middle), and 950 m (bottom).

residuals were roughly the same – about equal to the synthetic measurement errors of 10 m/a. The solution with $\alpha = 950$ m has some evident spurious oscillations, especially around the boundary of the physical domain; these are much less evident in the solution with $\alpha = 1.4$ km. Additionally, the interface between the two regions of different temperature is visibly rougher for the under-regularized solution. Despite the fact that the total variation functional permits jump discontinuities, the contrast between the properly- and under-regularized solutions demonstrate that it is also effective at filtering out spurious oscillations.

For the results shown in figure 5.10, the Huber TV functional was used to regularize the inference problem because it is much more computationally tractable to minimize than the true TV using convex duality. Smoothing over the true TV functional may also be genuinely desirable on physical grounds. I have argued that the Dirichlet energy is not the most appropriate penalty because the fields we are inferring can have sharp changes over the distance of a single cell of the computational mesh. Nonetheless, the temperature field does not exhibit jump discontinuities down to the smallest scales because ultimately thermal diffusion would smooth over any jumps. Evolutionary models for damage evolution such as in Albrecht and Levermann (2012) do feature genuine jump discontinuities. These models have not been subjected to the degree of validation against real data that, say, the shallow shelf equations have in (MacAyeal et al., 1996). The possibility that the true damage field exhibits sharp transitions that are nonetheless continuous is well within the range of model uncertainty in this case.

Using the HTV functional does introduce another difficulty in the form of the parameter γ. The parameter γ has the units of $\nabla\theta$. In the limit as $\gamma^{-1}|\nabla\theta| \ll 1$, the HTV functional is approximately equal to $\gamma^{-1}|\nabla\theta|^2/2$; when $\gamma^{-1}|\nabla\theta| \gg 1$, the HTV functional is asymptotic to $|\nabla\theta|$. In other words, the asymptotic growth of the penalty as a function of $|\nabla\theta|$ changes between the large- and small-gradient regimes, and the gradient scale γ dictates where this transition occurs. We can argue that, if Θ is some range of the smallest to largest parameter values we expect and L is the diameter of the spatial domain, then $\gamma = \Theta/L$ is a reasonable choice. This value of γ was used for the results shown in figure 5.10. For larger values of γ, the HTV-regularized results begin to resembles the results using the Dirichlet energy. The arbitrariness in selecting the secondary scale γ is a more serious concern; we will address the topic of *hyperparameter selection* in §5.5.

Another disadvantage of the HTV functional is its computational expense. The Dirichlet energy is a quadratic functional, and its second derivative d^2R is the Laplace operator. The HTV functional is not quadratic, so a Newton update, which involves solving a linear system for $I + \alpha^p\,\mathrm{d}^2R$, is a nonlinear elliptic PDE that might require several iterations. In a software implementation, the sparse matrix that discretizes the Laplace operator can be assembled once, whereas the second derivative of the HTV functional has to be recalculated several times during the course of an iterative minimization algorithm. In practice, however, the additional computational expense of HTV regularization is comparatively minor.

Overall, the Huber total variation functional is more specific than the Dirichlet energy, which unduly penalizes spatial variations that we consider likely. Several studies of ice shelf rheology have the explicit goal of locating sharp interfaces

that modellers hypothesize are present on real ice shelves (Khazendar et al., 2011; Borstad et al., 2016). Theories like continuum damage mechanics that predict these sharp features might appear to be inadequate when compared against a smooth inferred rheology. For example, Larour et al. (2014) have argued that CDM is not ideal for representing the effect of rifts on ice shelf flow because it is inherently diffusive. This diffusivity may not in fact be inherent, but rather a consequence of the the choice of regularization functional. The arguments of this section show that the HTV functional is a better fit for the physics and for the goals of these types of study than the Dirichlet energy. We do not claim, however, that HTV is perfect. The fact that the sharp transition at the upstream side of the interface is recovered with much greater fidelity than the along-flow transition in the synthetic inference problem shown in figure 5.10 suggests that there is still some room for improvement. For example, anisotropic functionals aligned with the flow field might give still better results.

5.4 Bayesian inference

In the last section, we argued that total variation regularization is superior to other methods for inferring fields that can have jump discontinuities or sharp interfaces. The TV functional specifically penalizes the kinds of spatial variability we consider unlikely of the temperature and damage fields; the Dirichlet energy, on the other hand, is not specific enough in that it smooths over sharp gradients. This justification, however, is largely heuristic. How should we apply this lesson to other inference problems with different physics?

All of the inverse problems that we have considered so far can be viewed through the lens of statistical inference. In this section, I will describe the correspondence between deterministic inverse problems and Bayesian nonlinear regression. The statistics lore suggests that some of the methodology in glaciological inference problems should be reevaluated in light of this correspondence.

In Bayesian inference, the corresponding problem to the selection of a regularization functional is how to choose a prior distribution. This problem has been studied in the statistical literature at least since Jaynes (1957); the statistical community has behind it many decades of collective wisdom on prior selection. These ideas, in particular the application of maximum entropy principles, can suggest new ways of regularizing glaciological inference problems. Once the modeller has chosen how to regularize a problem, it is then incumbent on him or her to decide how much to regularize it by picking a regularization parameter. The analogous problem in Bayesian inference is the selection of *hyperparameters*, i.e. parameters of the prior distribution. With some knowledge of the statistical distribution of the measurement errors, we can sometimes make more informed choices for the regularization parameter than what the L-curve would suggest by using the discrepancy principle or Bayesian information criteria.

Finally, we have so far been concerned with deriving a single estimate for the unknown parameters, which in Bayesian statistics corresponds to finding the maximum of the posterior distribution. We could instead try to sample from the posterior

distribution directly using Monte Carlo techniques. Sampling is certainly not cheap, but may achieve some goals, e.g. quantifying uncertainty in predictions of future sea-level rise, more directly than computing a single estimate of either basal shear or rheology.

The statistical analogue of deterministic inverse problems falls under the general aegis of nonlinear regression. In a frequentist nonlinear regression problem, we are given an M-dimensional vector of measurements d from which we would like to infer an N-dimensional vector of parameters θ. The true parameters θ_{true} and the observations are related through some map G such that

$$d = G(\theta_{\text{true}}) + \xi \tag{5.25}$$

for some random vector ξ of measurement noise. We will assume that the probability distribution P_ξ of the measurement noise is known.

One way to estimate the parameters is *maximum likelihood estimation* (MLE). The maximum likelihood estimate $\hat{\theta}$ of the parameters is the solution of the optimization problem

$$\hat{\theta} = \operatorname{argmax}_\theta P_\xi(G(\theta) - d). \tag{5.26}$$

Provided we have analytical expressions for the forward map and measurement error density P_ξ, we can differentiate them to obtain a nonlinear system that $\hat{\theta}$ must satisfy and approximate the solution numerically. One favorable property that MLE has in general is that, under certain mild assumptions on P_ξ and G, the maximum likelihood estimates converge to the true value of the parameters θ_{true} as the number of observations grows large (Van der Vaart, 2000). While there are other ways to estimate a set of unknown parameters, such as finding the minimum variance unbiased estimator, we will only consider maximum likelihood estimation and its Bayesian counterpart in the following.

A simple and illustrative example is multivariate linear regression. Suppose we know that the map G is linear, and that the measurement errors are normally distributed with mean 0 and covariance matrix C; the likelihood is

$$L(d;\theta) = (2\pi)^{-M/2} |\det C|^{-1/2} \exp\left(-\frac{1}{2}(G\theta - d)^* C^{-1}(G\theta - d)\right). \tag{5.27}$$

Instead of maximizing the likelihood, we can take the negative logarithm and minimize instead. Since the function $f(p) = -\log p$ is convex and decreasing, the two optimization problems are equivalent. In the normal case, taking the negative logarithm makes the optimization problem quadratic:

$$-\log L(d;\theta) = \frac{1}{2}(G\theta - d)^* C^{-1}(G\theta - d) + \dots \tag{5.28}$$

where we have dropped terms that do not depend on θ. If we define the norm $\|y\|^2_{C^{-1}} = y^* C^{-1} y$, then we wish to minimize the quadratic function

$$J(\theta) = \frac{1}{2}\|G\theta - d\|^2_{C^{-1}}. \tag{5.29}$$

The condition that $\hat{\theta}$ is a minimum of the negative log-likelihood is

$$G^* C^{-1} G\hat{\theta} = G^* C^{-1} d. \tag{5.30}$$

Rather than compute the inverse of the covariance matrix explicitly, we can compute the Cholesky factorization $C = \Gamma\Gamma^*$ where Γ is lower triangular. The estimate $\hat{\theta}$ is normally distributed, since it is a map applied to the normal random variable d. Consequently, the probability distribution of the ML estimate is characterized entirely by its mean and covariance matrix. The mean of the ML estimate is equal to the true parameters:

$$\mathsf{E}(\hat{\theta}) = \theta_{\text{true}}. \tag{5.31}$$

The covariance matrix is

$$\mathsf{E}((\widehat{\theta} - \theta_{\text{true}})(\hat{\theta} - \theta_{\text{true}})^*) = (G^* C^{-1} G)^{-1}. \tag{5.32}$$

Knowing the distribution of the ML estimate, we can compute confidence intervals and other useful quantities. Most importantly, the residual $\|G\hat{\theta} - d\|^2_{C^{-1}}$ has a χ^2-distribution with $M - N$ degrees of freedom, where M is the dimension of d and N the dimension of θ (Aster et al., 2013). The mean of a χ^2-random variable with ν degrees of freedom is ν. This fact gives us some indication of what size residuals we can expect when we solve an inverse problem; $2J(\theta)/(M-N)$ will be of order 1 due to random errors in the measurements, and we cannot expect to do better.

Linear regression provides a more general viewpoint on why overfitting occurs in some PDE inverse problems. The solution of a linear regression problem can be expressed in terms of the generalized singular value decomposition of the map G. The generalized SVD of G is the matrix factorization

$$G = U\Sigma V^*. \tag{5.33}$$

The matrix Σ is diagonal:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix} \tag{5.34}$$

The entries σ_k are the *singular values* of G, assumed to be in decreasing order, i.e. $\sigma_1 \geq \ldots \geq \sigma_n$. The columns of U and V are called the left- and right-singular vectors. The left- and right-singular vectors are orthonormal in the C^{-1}-inner product and the Euclidean inner product, respectively, i.e.

$$u_j^* C^{-1} u_k = \delta_{jk}, \qquad v_j^* v_k = \delta_{jk}. \tag{5.35}$$

The solution of the least-squares problem is

$$\hat{\theta} = \sum_k \sigma_k^{-1} \langle u_k, d\rangle v_k. \tag{5.36}$$

The singular values are, by definition, decreasing as k grows larger. For many real problems, the ratio between the largest and the smallest singular value is extremely high, i.e. $\sigma_1/\sigma_n \gg 1$, and the decay is exponential. The solution of the linear least-squares problem will then be extremely sensitive to the component $\langle u_k, d\rangle$ of the input data along higher-mode left singular vectors, since the weight σ_k^{-1} grows very large. Moreover, many measurement techniques are prone to errors that are equidistributed among all singular vectors; if ξ is the noise in our measurements, then the variance of $\langle u_k, \xi\rangle$ is roughly equal for all values of k. The solution to the least-squares problem is then dominated by fitting to noise rather than to the true signal.

In short, overfitting is a risk when

1. the singular values of the forward map decay rapidly;
2. and the components of the measurement errors along the high-k singular vectors are just as larges as along the low-k singular vectors.

For the estimation of geophysical fields, the singular vectors corresponding to higher modes often represent spatial variability at finer and finer scales. The decay of the singular values is a consequence of the fact that the field we observe is insensitive to high-wavenumber modes of the parameters we wish to estimate. For many common measurement techniques, however, the measurement errors are white noise – the measurement errors at two distinct points are uncorrelated. White noise has equal spectral power at all frequencies, in which case we can expect that $\langle u_k, d\rangle$ is bounded below by the noise amplitude as k gets larger.

Overfitting of statistical models is a more general phenomenon than just in the estimation of geophysical parameters. By a "statistical model", we mean the choice of space in which the problem parameters are assumed to live and the representation of a particular parameter value in that space. For example, when we use Galerkin's method to express the basal friction or rheology coefficients in terms of a set of spatial basis functions, our statistical model consists of the linear span of the set of Galerkin basis functions. The risk of overfitting is greatest when we use more and more sophisticated, high-dimensional statistical models; one can fit practically any data whatsoever with a model of sufficient complexity. This consideration suggests that, to prevent overfitting, we might instead use only as many basis functions as we need. A perfect example of this in the glaciological literature is the use in MacAyeal (1992) of Galerkin basis functions which are only capable of resolving variations in basal shear stress on the scale of 4 ice thicknesses or greater.

On the other hand, coming up with just the right set of basis functions requires a degree of knowledge about the problem that we might not have. Instead, we might use a high-dimensional statistical model and penalize the solution after the fact. This idea is the intuition behind *ridge regression* in linear regression problems. Ridge regression is formally identical to Tikhonov regularization; it provides a bridge between frequentist and Bayesian inference. Rather than minimize $\|G\theta - d\|^2_{C^{-1}}/2$, the ridge regression solution is the minimizer of

$$J(\theta) = \frac{1}{2}(G\theta - d)^* C^{-1}(G\theta - d) + \frac{1}{2}\theta^* L\theta, \tag{5.37}$$

where L is some positive semi-definite matrix. In most applications of ridge regression, L is a multiple of the identity matrix, i.e. $L = \alpha^2 I$. The solution of the ridge regression problem can be written in terms of the singular value decomposition as

$$\hat{\theta} = \sum_k \frac{\sigma_k}{\sigma_k^2 + \alpha^2} \langle u_k, d \rangle v_k. \tag{5.38}$$

A mode k such that $\sigma_k \ll \alpha$ will be filtered out, where in the unregularized problem the weight of this mode in the solution would be σ_k^{-1} (Vogel, 2002).

If the parameters θ were themselves normally distributed with mean 0 and covariance matrix L^{-1}, then the solution of the ridge regression problem would be the maximizer of the conditional density of θ given the data d. The idea that the parameters can have a distribution of their own, however, is antithetical to the frequentist interpretation of probabilities. To this way of thinking, the parameters have one definite, non-random value; the estimator $\hat{\theta}$ is a random variable only through the randomness of the data d. The jumping-off point for Bayesian methods is to view probabilities as a way of quantifying a degree of uncertainty, rather than as a limit of frequencies of events. This distinction can be illustrated with a simple thought experiment. A friend of yours is six months pregnant; what is the probability that her child is a girl? In the absence of any other information such as an ultrasound, it's reasonable to argue that the probability is $1/2$. However, to a strict frequentist, the probability is either 0 or 1; at six months, the sex of the child is already determined, we just don't know what it is. To a Bayesian, however, the probabilistic statement is a perfectly valid assessment of our uncertainty, as is the assignation of a probability distribution to the unknown parameters (Gelman et al., 2014).

In frequentist inference, we attempt to estimate the parameters θ from the data d alone. Bayesian inference, by contrast, starts by assuming that we know something about the parameters in the form of a *prior distribution* P_θ. The prior distribution encodes knowledge that we have of the parameters independent of any of the measurements. The chief object of study in Bayesian inference is the *posterior distribution*. Given a set of observations u^o, the posterior distribution is the conditional density of the parameters on the measurements, i.e. $P(\theta|u^o)$. Using Bayes's rule, we can rewrite this conditional density:

$$P(\theta|u^o) = \frac{P_\xi(u^o|\theta)P_\theta(\theta)}{P(u^o)}, \tag{5.39}$$

where P_ξ is the likelihood of observing u^o assuming θ are the true parameters, P_θ is the prior density for the parameters, and $P(u^o)$ is the probability of having observed u^o over all possible values of the parameters (Jaynes, 2003). As we have shown before, the likelihood is determined by the forward map and the statistical distribution of the measurement errors ξ. We will show in the remainder of this section that the prior distribution is determined by a combination of what we know about the parameters from the governing physics and from maximum entropy principles. The total probability can be written as

$$P(u^o) = \int P_\xi(u^o|\theta)P_\theta(\theta)\,\mathrm{d}\theta, \tag{5.40}$$

i.e. the integral of the likelihood weighted by the prior. In frequentist inference, our goal was to compute an estimate of the parameters by maximizing the likelihood. The end goal of a Bayesian inference problem is to evaluate how our initial state of uncertainty about the parameters, as quantified by the prior distribution, changes when we get a set of concrete observations u^o.

These notions are all very abstract, and to turn them into procedures that we can leverage for applied problems there are several equations that need answering. First of all, what is the prior distribution? Assuming we knew this distribution, how can we practically go about computing the posterior? Evaluating the probability $P(u^o)$ of observing u^o as in equation (5.40) involves a high-dimensional integral; if we use, say, 10,000 finite element basis functions to represent θ, then this will be a 10,000-dimensional integral. Evaluating such an integral will be effectively impossible for anything except a linear forward map and normal errors and priors. Even if we had a computationally tractable procedure for evaluating the posterior distribution, what useful information can we gain out of it?

Let us answer the second two questions first, assuming that we know the prior density on θ. For simple probability distributions, we can often calculate exactly or approximate the cumulative distribution function; we can then sample directly from the original density if we can generate uniformly-distributed random numbers. For more complex distributions, evaluating the cumulative distribution is intractable. Instead, we can sample from a complex posterior distribution using *Markov-chain Monte Carlo* (MCMC) sampling. Given two values of the parameters θ and θ', we can evaluate which of the two is more likely by computing the ratio of the two probabilities:

$$\frac{P(\theta'|u^o)}{P(\theta|u^o)} = \frac{P_\xi(u^o|\theta')P_\theta(\theta')}{P_\xi(u^o|\theta)P_\theta(\theta)}. \tag{5.41}$$

Note that the normalization factor $P(u^o)$ does not appear in the last equation; we can still compare parameters without normalizing the whole distribution. The Markov chain $\{\theta_n\}$ that we will use to sample from the posterior is defined as follows:

1. Pick an initial parameter field θ_0.
2. Given some value θ_n of the parameters, randomly compute a new value θ', for example through a random walk.
3. Compute the ratio $r = P(\theta'|u^o)/P(\theta_n|u^o)$ of the posterior probability of θ' to that of θ_n.
4. If $r > 1$, the candidate θ' is more likely than θ_n, so we immediately *accept* it and set $\theta_{n+1} = \theta'$.

   If $r < 1$, choose a random number z in $[0, 1]$. If $z < r$, then we accept the guess and set $\theta_{n+1} = \theta'$; otherwise, we *reject* the guess and take $\theta_{n+1} = \theta_n$.

The resulting sequence $\{\theta_n\}$ is a Markov chain, and one can prove that the limiting distribution of this stochastic process is the posterior distribution (Gelman et al., 2014).

We can approximate the conditional mean of a function f of the parameters by computing a weighted sum over the MCMC samples:

$$\int f(\theta)P(\theta|u^o)\,\mathrm{d}\theta \approx \frac{1}{N}\sum_{n=1}^{N} f(\theta_n). \tag{5.42}$$

Suppose that θ is the basal friction parameter for some glacier. An example of f could be to compute the current velocity, solve the diagnostic and prognostic equations forward 100 years from now, compute the new glacier volume, and subtract this from the original to see how big the sea-level contribution will be from this glacier. In that case, computing a conditional mean of this function tells us how much sea-level rise we can expect from this glacier on average. Having approximated this conditional mean, we can go back and sample again to compute the conditional variance of the estimate, or whatever other statistic is of interest. Sampling from the posterior distribution is exactly the tool we need if we want to make general probabilistic forecasts. Berliner et al. (2008) used MCMC sampling with the shallow ice approximation as the forward model to compute the conditional mean and variance of the basal sliding speed under the Northeast Greenland Ice Stream.

There are a number of practical difficulties that have to be addressed in order to apply MCMC sampling. The distribution of the Markov Chain converges to the posterior distribution given enough time, but the convergence rate may be so slow that the algorithm is impractical. Even if the MCMC samples converge in a reasonable amount of time, they may still be highly correlated over several timesteps, so the samples are not truly statistically independent. These issues can be addressed by using the right strategy to pick the candidate parameters θ' in step 2 of the MCMC algorithm. In any case, we will not use MCMC sampling in the following; more details on how to design rapidly-convergent sampling strategies can be found in Gelman et al. (2014).

By analogy with the frequentist method of maximum likelihood estimation, we could instead seek the value of θ which maximizes the posterior distribution. This value of θ is called the *maximum a posteriori* (MAP) estimator of the parameters. Much like frequentist maximum likelihood estimation, we can take the negative logarithm of the posterior and minimize instead, giving us the objective functional

$$J(\theta) = -\log P(\theta|u^o) = -\log P_\xi(u^o|\theta) - \log P_\theta(\theta) + \log P(u^o). \tag{5.43}$$

Since the total probability $P(u^o)$ does not depend on θ, we are relieved of the problem of how to evaluate this high-dimensional integral when computing the MAP estimate also. If we define the two functionals

$$E(\theta) = -\log P_\xi(u^o|\theta), \tag{5.44}$$

$$R(\theta) = -\log P_\theta(\theta) \tag{5.45}$$

then the procedure for computing the MAP estimate is formally identical to that of solving a regularized inverse problem with objective functional $J = E + R$. Raymond and Gudmundsson (2009) was the first paper to approach glaciological inference using maximum a posteriori estimation. Owing to the similarity with existing

methodology, this approach was applied in several subsequent papers (Pralong and Gudmundsson, 2011; Isaac et al., 2015; Arthern, 2015). Petra et al. (2014) combined the two approaches by first computing the MAP estimate for the basal shear stress, then using a Gaussian approximation around this estimate with MCMC sampling.

If we are not concerned with obtaining information about the model spread around our prediction, the MAP estimate could be used to initialize a prognostic model of glacier evolution. Joughin et al. (2012, 2014b) used this approach for Jakobshavn and Thwaites glaciers, respectively. Brinkerhoff and Johnson (2013) and Gillet-Chaulet et al. (2012) simulated the evolution of the entire Greenland Ice Sheet, applying the same principle. Compared to MCMC sampling of the future ice sheet state over the posterior distribution of parameters, a single forward model run initialized with the MAP estimate is much cheaper computationally. We can get away with cutting corners when certain assumptions are satisfied. First, the posterior distribution must be unimodal and have no skewness. In this case, the MAP estimate coincides with the posterior mean. Additionally, if the parameter-to-state and prognostic models are linear, then the conditional mean of the prediction that we would obtain from MCMC sampling is the same as the prediction based on the MAP estimate. For glaciological problems, neither of these assumptions is satisfied. Petra et al. (2014) and Isaac et al. (2015) attempt to construct a Gaussian approximation of the posterior, assuming the prior distribution on the basal sliding coefficient is also Gaussian. To approximate the posterior with a Gaussian distribution, they compute the first and second derivatives of the negative logarithm of the posterior. The second derivative operator $\mathscr{H} = \mathrm{d}^2 J / \mathrm{d}\theta^2$ is taken to be the inverse covariance of the approximate posterior. The size of this problem is still unmanageably large, so they use a low-rank approximation of $\mathscr{H}$ to reduce the dimensionality. This approach, however, requires that the prior distribution is Gaussian, which may not be the case for two reasons. The prior may be intrinsically non-normal, as will be the case for the distribution that represents fields with small total variation. The coefficient we seek – rheology or basal friction – is also defined through some convenient function that may have no inherent physical meaning. For example, if we parameterize the basal friction coefficient C as $C = C_0\beta^2$ or $C = C_0 e^{\beta}$, then a Gaussian distribution on the quadratic parameterization gives a non-Gaussian distribution on the equivalent exponential parameterization and vice versa. Even if the prior is Gaussian, the forward model may be highly nonlinear around some parameter value β, in which case the posterior distribution will be far from the Gaussian approximation.

In Bayesian inference, the prior distribution encapsulates information about the problem that comes from sources other than the data at hand. We often have such additional information in the form of conservation laws, physical principles, or previous studies of the same phenomenon. The use of regularization in deterministic inverse problems, on the other hand, is typically justified on the grounds that it makes the problem well-conditioned. This begs the question of why computing the maximizer of the posteriori probability should be a well-conditioned problem at all. One can easily imagine a posterior distribution that has a broad, flat plateau, a region where all parameter values are roughly equally likely and there is no unique

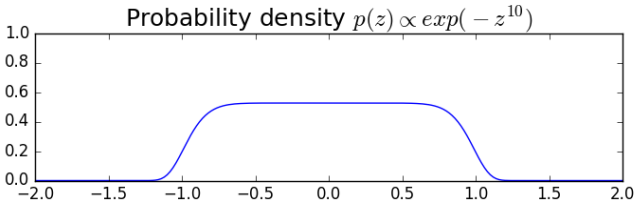


Figure 5.11: A probability distribution on the real line with a nearly flat plateau around its mode at $z = 0$.

maximizer. One such probability distribution is shown in figure 5.11. This distribution has a maximum at $z = 0$, but is so flat in the neighborhood of the MAP estimate that minimizing it numerically is terribly ill-conditioned. Many inference problems do have a well-defined MAP estimate, but there is no fundamental principle that guarantees a sharply-peaked posterior distribution with a data set of finite size. Nonetheless, we can still characterize the degree of uncertainty in our estimates or predictions by sampling from the posterior.

5.5 Prior selection

Given a prior distribution and some characterization of the measurement error statistics, we can compute the maximum a posteriori estimate or sample from the posterior distribution. We have not shown how to come up with a sensible prior distribution in the first place. The prior is supposed to encode pre-existing information about the parameters. How are we supposed to quantify the notion of information in the first place, and how do we go from information to a probability distribution? There are many distinct probability distributions that are all consistent with the same set of prior information, so isn't the choice of one over the other completely subjective and arbitrary? Ideally, among the many probability distributions that are consistent with what we know, we would like to pick the one that is otherwise least informative. In other words, we want our prior distribution to include what we know, but we don't want it to also imply all sorts of other things that we did not intend; we want the *most uninformative* distribution. These ideals are all well and good, but they only raise another question: how do we quantify how uninformative a probability distribution is?

The answer is provided by the notion of *entropy*, denoted H. We will first consider discrete probability distributions; let p_k be the probability that event A_k occurs, where the events $\{A_1, \ldots, A_n\}$ are mutually disjoint. There are three properties which characterize the entropy of a discrete probability distribution.

1. Fix some value n, and consider the set of all probability distributions on n outcomes. The entropy should take its maximum on the distribution $p_k = 1/n$ for $k = 1, \ldots, n$.

2. Suppose that $\{B_1, \ldots, B_m\}$ is another mutually disjoint partitioning of the sample space, and q is a discrete probability distribution on this partitioning. Let r_{jk} be the joint distribution of q and p, i.e. the probability of both B_j and A_k. The quantity r_{jk}/p_k is the conditional probability of B_j on the event A_k; one can check that this also fulfills the axioms for a probability distribution. We can then define the conditional entropy $H(q|p_k)$ of q on p_k as the entropy of the conditional distribution. In the case, we expect that the entropy of the joint distribution of both p and q is equal to the entropy of p, plus a weighted some of the conditional entropies:

$$H(p,q) = H(p) + \sum_k p_k H(q|p_k). \tag{5.46}$$

3. Suppose we add another event A_{n+1} to the sample space and define a new distribution p', such that $p'_k = p_k$ if $k \leq n$ and $p'_{n+1} = 0$. Then

$$H(p') = H(p). \tag{5.47}$$

Any function H on the set of all discrete probability distributions which satisfies these three properties is of the form

$$H(p) = -\lambda \sum_k p_k \log p_k \tag{5.48}$$

for some constant λ (Khinchin, 1957); we can take $\lambda = 1$ for definiteness. Since these properties uniquely characterize the entropy up to a multiplicative constant, we can say that the entropy functional H is *the only* way to characterize how uninformative a probability distribution is.

In Shannon's original work on information theory, he took the entropy of a continuous probability distribution $P(\theta)$ to be

$$H(P) = -\int P(\theta) \log P(\theta)\, \mathrm{d}\theta \tag{5.49}$$

by analogy with the definition for discrete probability distributions (Khinchin, 1957). This quantity is called the *differential entropy*. However, this definition runs into difficulties in that it is defined only up to an additive constant. One can show this with an elementary argument based on dimensional analysis. Unlike a discrete probability distribution, the probability density P has units of $[\,\mathrm{d}\theta\,]^{-1}$. For example, if P is a probability distribution defined over three-dimensional space with lengths measured in meters, then P has units of meters^{-3}. However, one can only evaluate transcendental functions such as exp, log, sin, cos on dimensionless quantities. Instead, one can take a large but finite number of points θ_n, and compute the entropy of the discretized distribution obtained by computing the probability that θ is in $[\theta_n, \theta_{n+1})$. Provided that these discrete points θ_n are distributed according to some density $\pi(\theta)$ in the limit as the number N of discrete points grows large, then the entropy of this limiting distribution of discrete points is asymptotic to

$$H(P|\pi) = -\int P(\theta) \log\left(\frac{P(\theta)}{\pi(\theta)}\right) \mathrm{d}\theta. \tag{5.50}$$

This quantity is the *Kullback-Leibler divergence* of P with respect to the *invariant distribution* π. For the case of continuous probability densities, one must make some choice of invariant distribution, and entropies can only be measured with respect to some invariant distribution (Jaynes, 1963). The Kullback-Leibler divergence does not suffer from the drawbacks of the differential entropy, i.e. changing the units of the problem does not affect the definition of the entropy. In many cases, we can take the invariant measure to be a uniform distribution on the range of values we expect.

The preceding discussion suggests that, given a set of n expected values

$$\mathsf{E}(f_k(\theta)) = \alpha_k \tag{5.51}$$

for $k = 1, \ldots, n$, the prior distribution we should use maximizes the entropy

$$H(P|\pi) = -\int P(\theta) \log\left(\frac{P(\theta)}{\pi(\theta)}\right) \mathrm{d}\theta \tag{5.52}$$

where the integral is taken over all of parameter space. This constrained optimization problem can be turned into an unconstrained problem by enforcing the constraints in equation (5.51) with a set of Lagrange multipliers λ_k. There is also one more constraint to enforce, namely that the probability distribution should integrate to 1. We then need to extremize the functional

$$H = -\int P(\theta) \log\left(\frac{P(\theta)}{\pi(\theta)}\right) \mathrm{d}\theta - \sum_k \lambda_k \int f_k(\theta) P(\theta)\, \mathrm{d}\theta + \alpha \int P(\theta)\, \mathrm{d}\theta. \tag{5.53}$$

Using the usual argument of the calculus of variations, the condition for an extremum is that

$$-\log(P(\theta)/\pi(\theta)) - \sum_k \lambda_k f_k(\theta) + \alpha - 1 = 0. \tag{5.54}$$

If we set $Z = e^{\alpha - 1}$, rearranging some terms and taking exponentials gives

$$P(\theta) = Z^{-1} \pi(\theta) \exp\left(-\sum_k \lambda_k f_k(\theta)\right). \tag{5.55}$$

The quantity Z is a normalization constant to make sure that the probability density integrates to 1, i.e.

$$Z(\lambda_1, \ldots, \lambda_n) = \int \pi(\theta) \exp\left(-\sum_k \lambda_k f_k(\theta)\right) \mathrm{d}\theta. \tag{5.56}$$

In statistical mechanics, Z is called the *partition function*. By differentiating under the integral sign, one can show that

$$-\frac{\partial}{\partial \lambda_j} \log Z = \mathsf{E}(f_j(\theta)). \tag{5.57}$$

So if we can compute the partition function, we can invert it to find the right values of the Lagrange multiplier λ_k for the constraints in equation (5.51) to hold.

To return to the glaciological problem, the probability density for the prior distribution with a total variation constraint is

$$P(\theta) = Z(\lambda)^{-1}\pi(\theta)\exp\left(-\lambda\int_\Omega |\nabla\theta|\,\mathrm{d}x\right). \tag{5.58}$$

This probability density is defined over the Sobolev space $W_1^1(\Omega)$, which is infinite-dimensional. The measure-theoretic issues of defining integration over a Banach space are well beyond the scope of this work. Instead, we can assume that the parameters θ are represented in terms of some Galerkin basis functions $\{\phi_k\}$, i.e.

$$\theta = \sum_k \theta_k\phi_k. \tag{5.59}$$

Substituting this expansion into equation (5.58), we get a probability density over the finite-dimensional vector space of Galerkin expansion coefficients.

The principle of maximum entropy tells us what is the least informative probability distribution consistent with what we already know about our problem, but it is still incumbent on us to accurately quantify what it is we know. In the previous section, we argued that using a total variation penalty is a better fit to the problem of inferring ice shelf rheology than penalizing the Dirichlet energy because we expect based on the properties of temperature and damage transport that the rheology might have steep gradients. Viewed in the light of Bayesian inference, penalizing the Dirichlet energy amounts to using a prior distribution which assumes more about the parameters than the physics of the problem justify.

The properties of advective transport motivated our choice of a total variation prior, but other sources of information can also be used to constrain the problem even further. For example, Echelmeyer et al. (1994) measured the velocity and strain rate of a cross-flow profile on Ice Stream B and found that the rheology coefficient in the shear margins of the ice stream was about 10 times higher than in the center. This measurement suggests a lower bound on the values that the rheology coefficient can take. This prior information can be encoded into the requirement that the expectation of some function of the parameters is 0. Let f be a function such that $f(\theta) = 0$ for $\theta < 0$, and f is convex and rapidly increasing. The condition that the field θ is bounded below by θ_0 can be expressed by the condition that

$$\mathsf{E}\left(\int_\Omega f(\theta_0 - \theta)\,\mathrm{d}x\right) = 0 \tag{5.60}$$

with respect to the prior distribution. Similarly, an upper bound θ_1 on the parameters can be enforced by instead using the integral of $f(\theta - \theta_1)$. A function f with these properties is referred to as a *barrier function*. The only study in the literature to use such an approach is Vieli et al. (2006). They used the barrier function

$$f(B) = \begin{cases} (B - B_1)^4 & B > B_1 \\ 0 & B \le B_1 \end{cases} \tag{5.61}$$

to enforce the upper bound $B < B_1$ on the ice rheology, in addition to a penalty for smoothness. The upper bound on the ice rheology is chosen to correspond to a temperature of -18° C with no flow enhancement. The barrier function that Vieli and Payne used is rapidly increasing, but arbitrarily large values of the rheology are still theoretically possible. A harder constraint can be enforced by using a barrier function such as

$$f(\theta) = -\ln|\theta| \tag{5.62}$$

which goes to infinity at $\theta = 0$. In any case, the asymptotic growth of the barrier function can be tuned to reflect the degree of belief in the bounds being enforced.

Many of the prior distributions described in this chapter are non-Gaussian. For example, when we incorporate prior information about the boundedness of the parameters or that the total variation of the parameters is not too large, the maximum-entropy distributions are not Gaussian. The forward map from parameters to velocities is also highly nonlinear, so even if the error metric used is quadratic, the likelihood is also not Gaussian. While the Gaussian distribution is very convenient for computational purposes, we argue that assuming Gaussian priors and constructing Gaussian approximations of the posterior, as in Petra et al. (2014); Isaac et al. (2015), will miss the true statistical properties of the posterior.

5.6 Hyperparameter selection

In the preceding section, we showed how to construct the maximum entropy distribution for a given set of constraints $\mathsf{E}(f_k) = \alpha_k$. This assumes that we know the right values of either the constraints α_k or the parameters λ_k in equation (5.55). In general, we only know that the prior belongs to some family of distributions parametrized by $\lambda_1, \ldots, \lambda_n$. Parameters of the prior distribution are called *hyperparameters*, and selecting the right value of the hyperparameters is the analogue of choosing a regularization parameter in a deterministic inverse problem.

In section §3.3, we reviewed the L-curve method for selecting a single hyperparameter in a data assimilation problem. Hansen (1999) justified the use of the L-curve based on two properties of the error-smoothness curve found for synthetic linear inverse problems: the model-data misfit increases with the regularization parameter, and the error-smoothness curve exhibits a sharply-defined point of maximum curvature. These properties are not nearly as robust for nonlinear problems with real data. In Gillet-Chaulet et al. (2012); Shapero et al. (2016), and several of the synthetic inverse problems presented in this chapter, the corner of the L-curve is not nearly as well-defined as in Hansen's work. The model-data misfit actually decreases initially as more smoothing is applied in Gillet-Chaulet et al. (2012). They attribute this anomaly to the fact that the Robin inverse method does not furnish an exact gradient for nonlinear ice rheology, positing that regularization improves the convergence rate of the algorithm. One of the criticisms of the L-curve is that it does not take advantage of knowledge we have about the signal-to-noise ratio of the measurements used to constrain the inference problem (Habermann et al., 2012). Finally, it is not immediately obvious how to generalize the L-curve the case where

there are multiple regularization functionals. One could argue that the right value is the point of maximum curvature of the surface parameterized by model-data misfit and the values of all of the regularization functionals. In higher dimensions, this begs the question of whether one should use the mean curvature or the Gaussian curvature, to say nothing of the computational expense of approximating either. The convenient properties that Hansen used to justify the L-cruve for linear inverse problems with a single regularization functional may no longer even hold with more prior information.

Chapter 6

Ice shelves

In the previous chapter, we discussed common methods for inferring the rheology of an ice shelf, and how these methods might be improved by using ideas from statistics. The correspondence between deterministic PDE inverse problems and Bayesian statistical inference suggests that the issues of (1) how to choose a regularization functional and (2) how much to regularize an inverse problem be revisited. Data assimilation methods are a key component in recent work on how large-scale damage from crevasses affect the flow of ice shelves (Borstad et al., 2016). To validate a proposed model for damage evolution against real data, one needs to infer the rheology of the glacier. The choice of statistical methodology could, in principle, be the difference between rejecting such a model or not. In this chapter, I will describe my current work on applying data assimilation methods to real ice shelves. In particular, I am interested in the degree to which established flow features, such as visible rifts or flow stripes in satellite imagery, are reflected in the rheology of an ice shelf. All of the modelling shown in this chapter was carried out using the library icepack, described in §4.

6.1 The Ross Ice Shelf

The Ross Ice Shelf is the largest floating ice shelf in the world, with an area roughly equal to that of Texas (Bentley, 1990). In the recent past, the ice streams that flow into the Ross Ice Shelf vary on centurial time scales, the most notable example of which is the stagnation of Ice Stream C around 150 years ago (Anandakrishnan and Alley, 1997). Going further back in time, the grounding line of the West Antarctic Ice Sheet was much further advanced during previous ice ages than during the current epoch, reaching at least up to the current ice shelf terminus and possibly up to the continental shelf break (Denton and Hughes, 2002). The West Antarctic Ice Sheet has largely collapsed in previous interglacials (Hein et al., 2016), and may be in the process of doing so again (Joughin et al., 2014b). Understanding the flow history of Antarctic ice shelves is valuable in predicting the rapidity of this potential collapse, since ice shelves are important in determining the net mass flux of the ice

streams that feed them (Dupont and Alley, 2005).

Some of the early work in this vein was published by Casassa. In satellite imagery from MODIS, curvilinear features are visible on the surface of the Ross Ice Shelf, extending uninterrupted from the grounding line to the calving front. The residence time for the ice shelf is on the order of 1000 years, so these surface features can be thought of as recording the history of the ice shelf and the ice streams that flow into it over the past millenium. Casassa (1991); Casassa and Brecher (1993) examined possible causes for these flow stripes. They rule out several possible generative mechanisms. The brightness contrast is largely constant throughout the entire flow stripe, which rules out formation from either winds or variable surface topography. The flow stripes often deviate significantly from ice streamlines, so they cannot be simply a result of ice flow; moreover, the ice velocity is continuous across a flow stripe observed on Byrd Glacier, so they do not arise from the merger of two distinct tributaries. The two remaining causes are that the stripes are a surface expression of disturbances in the basal topography or friction (Gudmundsson et al., 1998), and as a signature of the sudden reactivation of previously stagnant ice streams.

Fahnestock et al. (2000) extend this analysis to construct a history of the Ross Ice Shelf over the past 1000 years. The flow stripes emerging from the glaciers in the Transantarctic Mountains (Skelton, Byrd, Starshot, Nimrod, Lenox-King) are mostly aligned with the ice velocities, suggesting that these glaciers have been in equilibrium over this time period. By contrast, the flow stripes originating from ice streams A, B, and C are not all aligned with the direction of flow. Ice streams A and B once flowed to the west of Crary Ice Rise, and Ice Stream C to the east. Ongoing migration of the margin of Ice Stream B over the last $\sim$700 years, together with the shutdown of Ice Stream C $\sim$150 years ago, prompted part of Ice Stream B to flow east of Crary Ice Rise. To estimate the timing of these events from present-day velocities, one has to assume that the ice shelf is roughly in steady state, which is exactly the opposite of the conclusion of their study. Hulbe and Fahnestock (2004) and Hulbe and Fahnestock (2007) instead use numerical models of ice shelf flow to spin up to the current state. They found that the scenario described in Fahnestock et al. (2000) is plausible, but under a fairly constrained set of timings for the onset of faster flow from Ice Stream A, migration of the Ice Stream B margin, and shutdown of Ice Stream C.

Rommelaere and Macayeal (1997) was the first study to use control methods to infer the viscosity of the Ross Ice Shelf, using velocity data from the RIGGS campaign. They find a highly viscous patch of ice in the shelf that would have originated from Ice Stream A roughly 1000 years ago. This finding supports the hypothesis from Casassa (1991) that the formation of a substantial ice raft at this time is responsible for a substantial increase in the speed of Ice Stream A.

6.2 Inferred rheology

Rommelaere and Macayeal (1997) used control methods to infer the viscosity of the Ross Ice Shelf, ignoring nonlinearity due to Glen's flow law. Their analysis focused

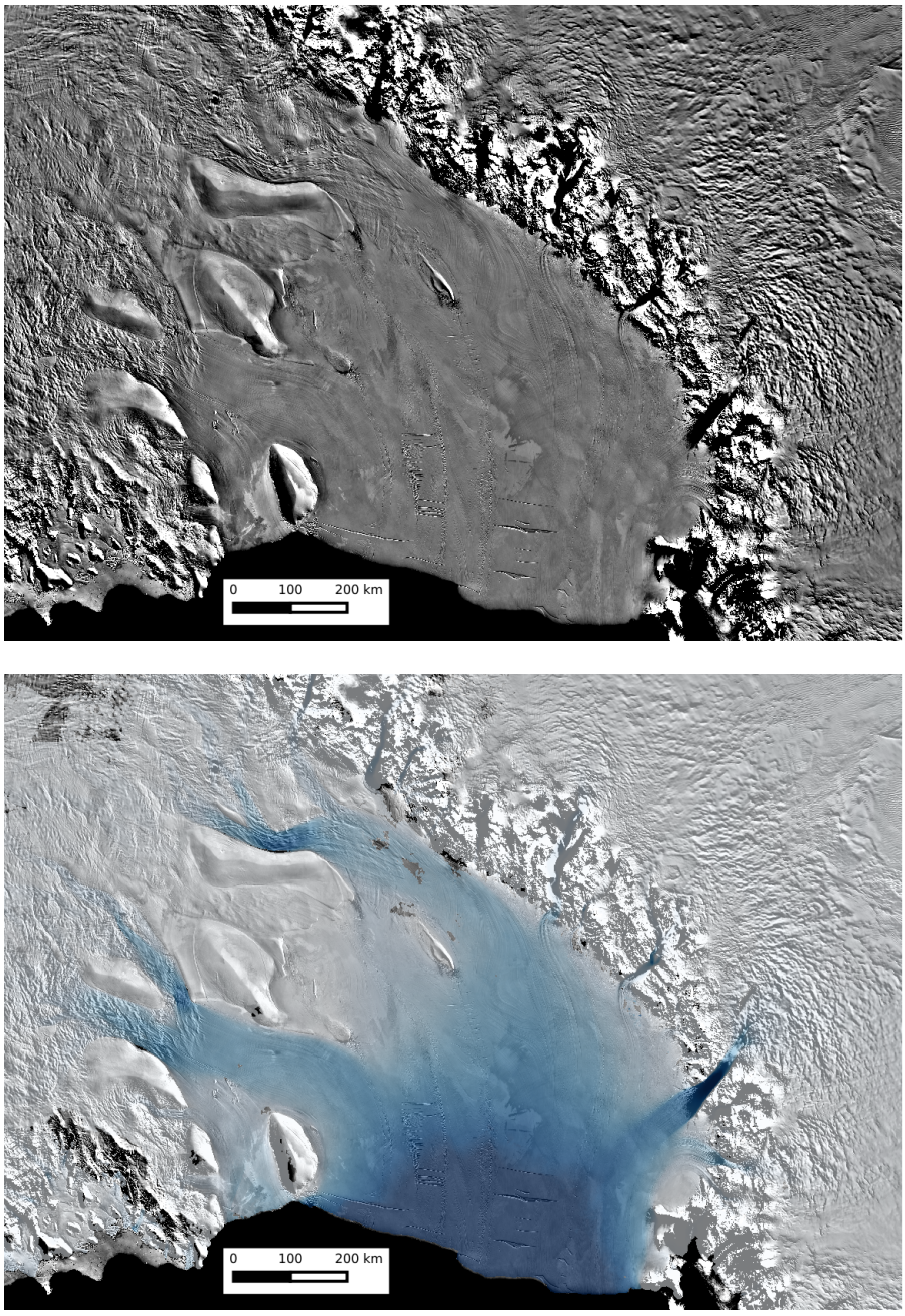


Figure 6.1: Mosaic image of the Ross Ice Shelf from MODIS (top) with ice speeds from 0-800 m/year overlay (bottom).

on the central part of the shelf, which extended from Steershead Ice Rise in the west to 4° of longitude east. This captures some of ice that flows into the shelf through the Trans-Antarctic Mountains. Figure 6.2 shows the rheology of the Ross Ice Shelf inferred using icepack with no regularization, a penalty for smoothness, and a penalty for total variation. The domain size is extended from Rommelaere's original work to include the flow of Byrd and Skelton glaciers.

For comparison, the viscosity computed in Rommelaere and Macayeal (1997) is reproduced in figure 6.3. Rather than use a nonlinear forward model and infer the depth-averaged rheology coefficient $\bar{B}$, they infer the depth-averaged viscosity $\bar{\mu}$ (recall that the two are related through $\bar{\mu} = \bar{B}|\dot{\varepsilon}|^{1/n-1}$). Nonetheless, a number of similarities are observable between their viscosity and the rheology shown in figure 6.2. My results reproduce the cold patch 400 km downstream of the grounding line that Rommelaere attributed to past changes in the outflow of Ice Stream A. I used a synthetic glacier as a first check that the data assimilation routines in icepack work correctly; these results for the Ross shelf show that icepack agrees with previously published work for real problems as well.

Several finer-scale features are discernible in my work that are not obvious in Rommelaere's inferred viscosity. The shear margins of Ice Stream A, Ice Stream D, and Byrd Glacier are visible as stripes of weaker ice. Similarly, large rifts just south of the Ice Stream A cold patch and just upstream of the calving front also appear as lines of weaker material rheology. These features can be resolved in my results due to higher resolution of the numerical model and of the InSAR velocities used to constrain the model, compared to the RIGGS velocities used in Rommelaere's work.

The rheology field inferred with no regularization is highly oscillatory in some areas. While smoothness regularization filters out these oscillations, the sharp interfaces at the boundaries of rifts and shear margins are diffused out. The final panel of figure 6.2 shows the result obtained with total variation regularization, in which much more of the detail is preserved, but most of the unphysical oscillations are still filtered out.

Inferring the rheology of an ice shelf near a poorly-known grounding line is much more difficult than in the center of the shelf. The position of an ice sheet grounding line is uncertain due to errors in the bed elevation and thickness maps. Additionally, an ice shelf can become grounded locally on ice rises or ice rumples. Ice rumples are often visible on satellite imagery as initiating flow stripes, but they need not have a noticeable expression at the surface. Suppose that we were to mistakenly treat some location where the glacier is grounded as floating instead. The ice velocity at this point is a function of the membrane, driving, and basal shear stresses. If we were to try and infer the ice rheology from these measurements under the mistaken assumption that the ice is floating, all of the resistance from the glacier bed will instead be attributed to the rheology. The inferred rheology will then have a positive bias. For example, in my first attempt at inferring the rheology of the Ross Ice Shelf, the temperature necessary to reproduce the shelf rheology downstream of Byrd Glacier would have been $-160°C$, which is completely unrealistic. Such a large value of the rheology coefficient is clearly unphysical, but the bias to the rheology might escape notice if the basal shear stress near the grounding line is small compared to the driving stress.

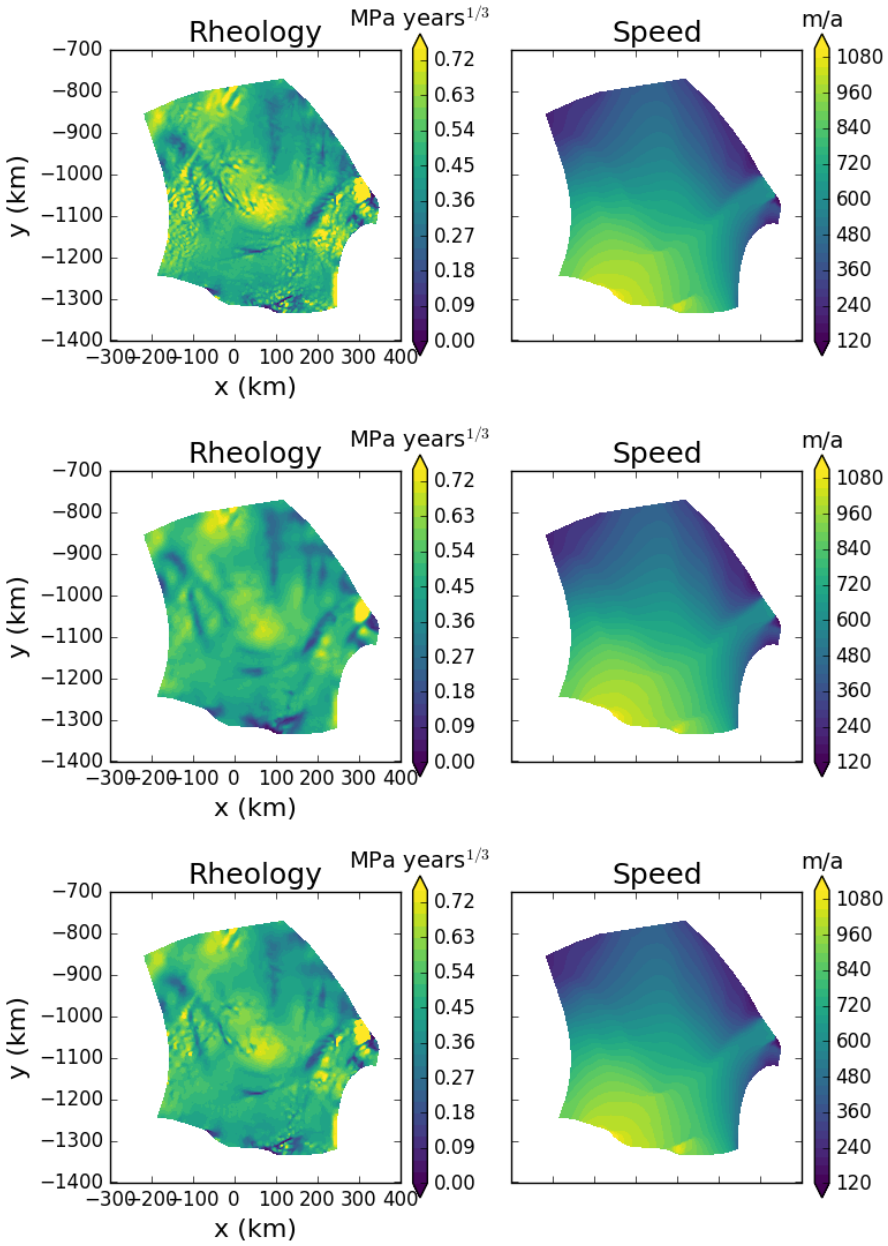


Figure 6.2: Rheology and speed of the Ross Ice Shelf with minimal regularization (top), smoothness regularization (middle) and TV regularization (bottom).

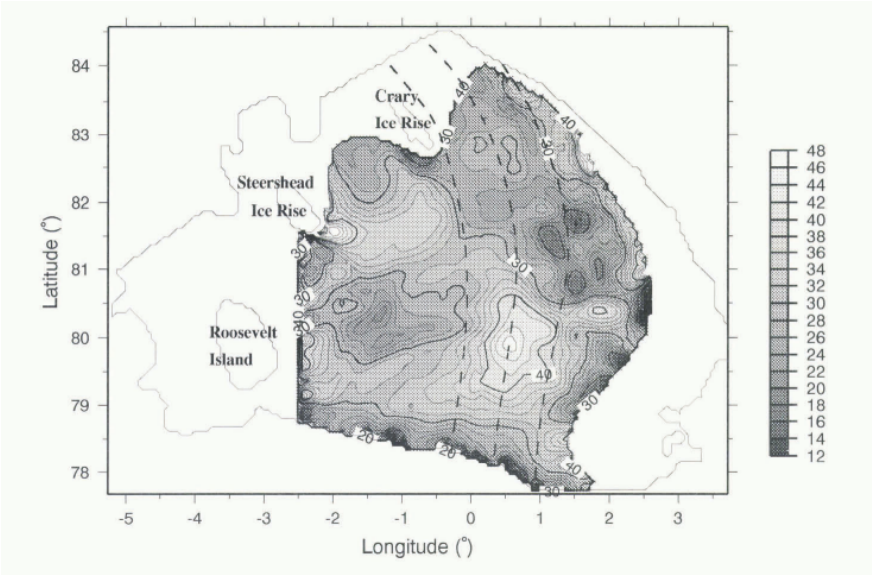


Figure 6.3: Viscosity (MPa × years) of the Ross Ice Shelf inferred from RIGGS data. Reproduced from Rommelaere and Macayeal (1997).

Even if the rheology is obviously biased, it is not clear what percent of this anomalously high value of the rheology is due to advection of cold ice from the grounded part of the ice sheet rather than to an under-constrained grounding line. For example, the rheology field that I infer captures the cold spot in the center of the shelf that Rommelaere and Macayeal (1997) found, which has a similar stiffness to the ice entering the shelf from Byrd Glacier. How much of the high rheology that I found downstream of Byrd Glacier is due to advection of cold ice, and how much is due to bias?

Conventional wisdom has held that simultaneously inferring the rheology and basal shear stress of a glacier is ill-posed, i.e. the minimizer of the objective functional is non-unique (Arthern and Gudmundsson, 2010). More recent work has found that this limitation is not as stark as previously thought, provided that one uses a sufficiently informative prior (Arthern et al., 2015). For example, in the old way of thinking, one has to provide a complete rheology field in order to infer the basal shear stress. This was exactly the approach I took in Shapero et al. (2016). Using the temperature and meltwater model output from Poinar et al. (2015) to set the ice rheology, rather than rely on an ad-hoc value, was one of the key novelties that set this work apart from its predecessors. Arthern et al. (2015) instead found that, when inferring the basal shear stress under the Antarctic ice sheet, optimizing for changes to the rheology parameter also is well-posed, provided we assume that these perturbations are relatively small. Arthern et al. use damped iterative methods to ensure that the rheology field changes by less than an order of magni-

tude. This assumption could also be enforced through a regularization functional that penalizes departure from the initial guess B_0 for the rheology:

$$R(B) = \int_\Omega \log(B/B_0)^2 \, \mathrm{d}x. \tag{6.1}$$

Other penalties might also be applied, such as for smoothness or small total variation of the inferred rheology. The optimization algorithm in this case is generally searching for the basal friction parameter that best explains the observations. However, if some amount of model-data misfit can be explained more easily with a slight change in the initial rheology field, then this explanation will be favored over a highly improbable change to the basal friction parameter, for example one that would induce large spatial variations.

The results of Arthern et al. (2015) suggest that, in the face of poorly-mapped grounding lines, we might attempt to infer both rheology and basal shear stress, subject to the prior information that there is no basal shear stress throughout most of the region of interest. Fürst et al. (2015) used this approach to identify possible unmapped pinning points on several ice shelves throughout Antarctica. This approach might be improved by using ℓ^1-regularization methods. Suppose we represent the basal friction parameter β in some finite element basis $\{\phi_k\}$ as a linear combination $\beta = \sum_k \beta_k \phi_k$. The constraint that the basal shear stress is 0 throughout most of the domain could be enforced through an ℓ^1-penalty on the expansion coefficients:

$$R(\beta) = \sum_k |\beta_k|. \tag{6.2}$$

Penalties of this form tend to make most of the coefficients 0, i.e. they favor sparse solutions (Aster et al., 2013). Since pinning points are one of the crucial elements of the buttressing effect (Dupont and Alley, 2005), locating unobserved ice shelf pinning points is especially important to making predictions of future ice sheet flow. One of my goals for future work is to explore the use of ℓ^1 penalties, which have yet to be explored in the literature on glaciological inverse problems.

In summary, ice shelves are important in dictating the stability of Antarctic glaciers (Dupont and Alley, 2005). Recent research has raised the prospect that the loss of ice shelves could lead to the large-scale collapse of the West Antarctic Ice Sheet on much shorter time scales than previously imaged, due to the brittle failure of tall ice cliffs (Pollard et al., 2015). We have already witnessed the collapse of Larsen B Ice Shelf (Rack and Rott, 2004). The nearby Larsen C Ice Shelf appears to be stable at present (Jansen et al., 2010), but may meet the same fate in the coming decades. In addition to my work on the purely methodological aspects of inferring the rheology of an ice shelf, in my future work I plan to extend this analysis to other ice shelves in Antarctica including Larsen C, and the Pine Island and Thwaites ice shelves.

Chapter 7

Conclusion

In this dissertation, I have described the work completed during my doctoral degree on data assimilation problems in glaciology. While remote sensing of ice sheets and glaciers provides a wealth of data on the flow of ice sheets and glaciers, there are many other unobservable quantities that are necessary for making predictions of future glacier flow. Data assimilation methods have proved useful in ascertaining these unobservable fields from the available measured data. I view my work as having made three main contributions to this subject area.

First is my work on the basal shear stress of the three big Greenland outlet glaciers. Up to this point, most of the existing work on this problem had focused on inferring the basal shear stress of either Jakobshavn Isbrae at high resolution, or of the entire Greenland Ice Sheet at lower resolution. These kinds of studies leave unanswered the question of how similar or different the subglacial environments of other fast-flowing outlet glaciers such as Kangerdlugssuaq and Helheim are to Jakobshavn. Additionally, there were still some unanswered questions about the spatial variability of basal resistance under Jakobshavn and other glaciers. Previous studies had used simplified models of glacier flow that assume a low aspect ratio and a high sliding ratio. Subsequent work by Sergienko and others using the full Stokes equations found strikingly different results in the form of banded structures. They argue that these banded structures are the hallmark of an instability of the coupled glacier-hydrology-sediment system studied by Hindmarsh in the decade previous. In my work, I sought to confirm or deny whether these banded structures could be inferred reliably from surface observations. In contrast to Sergienko's findings, the basal shear stress maps that I found for Jakobshavn agreed with the earlier work that had a completely weak bed under the main glacier trunk, interspersed with the occasional sticky spot. The basal shear stress under the two other big Greenland glaciers, Helheim and Kangerdlugssuaq, exhibited similar features. I attributed the discrepancy between my results and those of Sergienko to the underlying statistical methodology. In any inverse problem, one has to decide a lower bound on the length scale of features one expects to infer; for example, to decide whether features smaller than 1 km can be resolved faithfully. In the inverse problems literature, the most common technique for deciding on this smoothing length is called the L-curve;

in my work, I used the L-curve, whereas Sergienko did not. The features they find could be the result of overfitting, especially given that their initial guess, that the basal shear stress is a multiple of the driving stress, also has banded features.

I came to the realization through working on this paper that these subtle methodological issues, such as selecting a smoothing length, have a big impact on the conclusions one can or cannot draw from the results of data assimilation methods. For the work that followed, I wanted to find a way of highlighting these issues in as striking a way as possible for other practitioners in the field for whom considerations of hyperparameter selection can easily seem esoteric. Inferring the rheology of a floating ice shelf is an ideal model problem because it is as free of additional unknown fields as one can get in glaciology; by contrast, it is difficult to simultaneously infer the basal shear stress and rheology of a grounded glacier. By chance, at the same time I attended a lecture on total-variation methods for image reconstruction. TV methods use a different definition of smoothness to set the resolution limit for inference problems. Rather than require that the inferred field is smooth, one instead uses a different penalty functional that allows the occasional jump discontinuity, but which nonetheless filters out highly oscillatory modes. Similarly, there can be sharp contrasts in the rheological stiffness of an ice shelf over fairly short length scales through highly localized crevassing and damage. In this work, I used synthetic inversions to demonstrate that sharp interfaces in the rheology field can be resolved better with total variation regularization than with the usual methods. My current work is to apply these methods to the Ross Ice Shelf and to Larsen C.

Finally, the existing software tools for modelling the flow of ice sheets and glaciers are beginning to show their age, as new tools for finite element modelling have cropped up in the past five years and as software engineering has evolved generally. The existing tools usually enforce a batch-processing workflow, where the user feeds some input data into a large stand-alone program with some delimited range of possible simulations. With the library icepack, I have attempted to instead create a library for glacier modelling, providing a set of routines for solving for the flow of glaciers and allowing users to construct arbitrarily complex simulations from these building blocks. The development of icepack is ongoing, and my hope is that it will soon become useful for the glacier modelling community at large.

While I feel that these are valuable contributions to the field of inverse methods in glaciology, there are many possibilities still left unexplored at present. Currently, inverse methods mostly rely on assimilating a single snapshot of the ice geometry and velocity. They have yet to reach the advanced state of similar methods in numerical weather prediction, which use ensemble Kalman filtering to assimilate the entire time series of observations. This technique promises greater accuracy to the existing measurements and the possibility of confirming or rejecting putative models for various underconstrained aspects of glacier physics, such as iceberg calving. Many of the design decisions I have made in icepack are geared towards the possibility of time-dependent data assimilation in glaciology, and I intend to pursue this avenue of research in my postdoc and beyond.

Bibliography

T. Albrecht and A. Levermann. Fracture field for large-scale ice dynamics. *Journal of Glaciology*, 58(207):165–176, 2012.

R. B. Alley. In search of ice-stream sticky spots. *Journal of Glaciology*, 39(133): 447–454, 1993.

R. B. Alley, D. D. Blankenship, C. R. Bentley, and S. T. Rooney. Deformation of till beneath ice stream B, West Antarctica. *Nature*, 322(6074):57–59, 3 July 1986.

S. Anandakrishnan and R. B. Alley. Stagnation of ice stream C, West Antarctica by water piracy. *Geophysical Research Letters*, 24(3):265–268, 1997.

R. J. Arthern. Exploring the use of transformation group priors and the method of maximum relative entropy for Bayesian glaciological inversions. *Journal of Glaciology*, 61(229):947–962, 2015.

R. J. Arthern and G. H. Gudmundsson. Initialization of ice-sheet forecasts viewed as an inverse Robin problem. *Journal of Glaciology*, 56(197):527–533, 2010.

R. J. Arthern, R. C. Hindmarsh, and C. R. Williams. Flow speed within the Antarctic ice sheet and its controls inferred from satellite observations. *Journal of Geophysical Research: Earth Surface*, 120(7):1171–1188, 2015.

R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter estimation and inverse problems*. Academic Press, Waltham, MA, USA, 2013.

S. Balay, K. Buschelman, W. D. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang. Portable, extensible toolkit for scientific computation (PETSc). *Argonne National Laboratory*, 2004.

W. Bangerth, D. Davydov, T. Heister, L. Heltai, G. Kanschat, M. Kronbichler, M. Maier, B. Turcksin, and D. Wells. The deal. II library, version 8.4. *Journal of Numerical Mathematics*, 2016.

J. Bassis. Hamilton-type principles applied to ice-sheet dynamics: new approximations for large-scale ice-sheet flow. *Journal of Glaciology*, 56(197):497–513, 2010.

C. R. Bentley. The Ross Ice Shelf Geophysical and Glaciological Survey (RIGGS): introduction and summary of measurements performed. *The Ross Ice Shelf: Glaciology and Geophysics*, pages 1–20, 1990.

L. M. Berliner, K. Jezek, N. Cressie, Y. Kim, C. Lam, and C. J. van der Veen. Modeling dynamic controls on ice streams: a Bayesian statistical approach. *Journal of Glaciology*, 54(187):705–714, 2008.

R. A. Bindschadler and T. A. Scambos. Satellite-image-derived velocity field of an Antarctic ice stream. *Science*, 252(5003):242–246, 12 Apr. 1991.

D. D. Blankenship, C. R. Bentley, S. T. Rooney, and R. B. Alley. Seismic measurements reveal a saturated porous layer beneath an active Antarctic ice stream. *Nature*, 322(6074):54–57, 3 July 1986.

C. Borstad, A. Khazendar, E. Larour, M. Morlighem, E. Rignot, M. Schodlok, and H. Seroussi. A damage mechanics assessment of the Larsen B ice shelf prior to collapse: Toward a physically-based calving law. *Geophysical Research Letters*, 39 (18), 2012.

C. Borstad, E. Rignot, J. Mouginot, and M. Schodlok. Creep deformation and buttressing capacity of damaged ice shelves: theory and application to Larsen C ice shelf. *The Cryosphere*, 7(6), 2013.

C. Borstad, A. Khazendar, B. Scheuchl, M. Morlighem, E. Larour, and E. Rignot. A constitutive framework for predicting weakening and reduced buttressing of ice shelves based on observations of the progressive deterioration of the remnant Larsen B Ice Shelf. *Geophysical Research Letters*, 2016.

G. S. Boulton. Processes of Glacier Erosion on Different Substrata. *J. Glaciol.*, 23 (89):15–38, 1 Jan. 1979.

D. J. Brinkerhoff and J. Johnson. Data assimilation and prognostic whole ice sheet modelling with the variationally derived, higher order, open source, and fully parallel ice sheet model VarGlaS. *The Cryosphere*, 7:1161, 2013.

G. Casassa. Relict flow stripes on the Ross Ice Shelf. *Annals of Glaciology*, 15(1): 132–139, 1991.

G. Casassa and H. Brecher. Relief and decay of flow stripes on Byrd Glacier, Antarctica. *Annals of Glaciology*, 17(1):255–261, 1993.

J. Church, P. Clark, A. Cazenave, J. Gregory, S. Jevrejeva, A. Levermann, M. Merrifield, G. Milne, R. Nerem, P. Nunn, A. Payne, W. Pfeffer, D. Stammer, and A. Unnikrishnan. *Sea Level Change*, book section 13, pages 1137 – 1216. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN ISBN 978-1-107-66182-0. doi: 10.1017/CBO9781107415324.026. URL www.climatechange2013.org.

P. U. Clark and D. Pollard. Origin of the middle Pleistocene transition by ice sheet erosion of regolith. *Paleoceanography*, 13(1):1–9, 1998.

J. O. Coplien. Curiously recurring template patterns. *C++ Report*, 7(2):24–27, 1995.

R. Courant and D. Hilbert. *Methods of mathematical physics*, volume 1. CUP Archive, 1965.

K. M. Cuffey and W. S. B. Paterson. *The physics of glaciers*. Academic Press, 2010.

I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

G. H. Denton and T. J. Hughes. Reconstructing the Antarctic ice sheet at the Last Glacial Maximum. *Quaternary Science Reviews*, 21(1):193–202, 2002.

E. Domack, D. Duran, A. Leventer, S. Ishman, S. Doane, S. McCallum, D. Amblas, J. Ring, R. Gilbert, and M. Prentice. Stability of the Larsen B ice shelf on the Antarctic Peninsula during the Holocene epoch. *Nature*, 436(7051):681–685, 2005.

J. K. Dukowicz, S. F. Price, and W. H. Lipscomb. Consistent approximations and boundary conditions for ice-sheet dynamics from a principle of least action. *Journal of Glaciology*, 56(197):480–496, 2010.

P. Dunlop and C. D. Clark. The morphological characteristics of ribbed moraine. *Quaternary Science Reviews*, 25(13):1668–1691, 2006.

P. Dunlop, C. D. Clark, and R. C. Hindmarsh. Bed ribbing instability explanation: testing a numerical model of ribbed moraine formation arising from coupled flow of ice and subglacial sediment. *Journal of Geophysical Research: Earth Surface*, 113(F3), 2008.

T. Dupont and R. Alley. Assessment of the importance of ice-shelf buttressing to ice-sheet flow. *Geophysical Research Letters*, 32(4), 2005.

K. Echelmeyer and W. D. Harrison. Jakobshavns Isbrae, West Greenland: Seasonal Variations in Velocity – or Lack Thereof. *Journal of Glaciology*, 36(122):82–88, 1990.

K. Echelmeyer, W. Harrison, C. Larsen, and J. Mitchell. The role of the margins in the dynamics of an active ice stream. *Journal of Glaciology*, 40(136):527–538, 1994.

J. C. Ely, C. D. Clark, M. Spagnolo, C. R. Stokes, S. L. Greenwood, A. L. Hughes, P. Dunlop, and D. Hess. Do subglacial bedforms comprise a size and shape continuum? *Geomorphology*, 257:108–119, 2016.

E. M. Enderlin, I. M. Howat, S. Jeong, M.-J. Noh, J. H. Angelen, and M. R. Broeke. An improved mass budget for the Greenland ice sheet. *Geophysical Research Letters*, 41(3):866–872, 2014.

H. Engelhardt, W. Harrison, and B. Kamb. Basal sliding and conditions at the glacier bed as revealed by bore-hole photography. *Journal of Glaciology*, 20(84):469–508, 1978.

M. Fahnestock, T. Scambos, R. Bindschadler, and G. Kvaran. A millennium of variable ice flow recorded by the Ross Ice Shelf, Antarctica. *Journal of Glaciology*, 46 (155):652–664, 2000.

M. Fahnestock, T. Scambos, T. Moon, A. Gardner, T. Haran, and M. Klinger. Rapid large-area mapping of ice flow using Landsat 8. *Remote Sensing of Environment*, 2015.

H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

J. Fürst, G. Durand, F. Gillet-Chaulet, N. Merino, L. Tavard, J. Mouginot, N. Gourmelen, and O. Gagliardini. Assimilation of Antarctic velocity observations provides evidence for uncharted pinning points. *The Cryosphere*, 9(4):1427–1443, 2015.

O. Gagliardini, T. Zwinger, F. Gillet-Chaulet, G. Durand, L. Favier, B. De Fleurian, R. Greve, M. Malinen, C. Martín, P. Råback, et al. Capabilities and performance of Elmer/Ice, a new-generation ice sheet model. *Geoscientific Model Development*, 6(4):1299–1318, 2013.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.

F. Gillet-Chaulet, O. Gagliardini, H. Seddik, M. Nodet, G. Durand, C. Ritz, T. Zwinger, R. Greve, and D. G. Vaughan. Greenland ice sheet contribution to sea-level rise from a new-generation ice-sheet model. *The Cryosphere*, 6(6):1561–1576, 2012.

J. W. Glen. The creep of polycrystalline ice. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 228, pages 519–538. The Royal Society, 1955.

S. Gogineni, J.-B. Yan, J. Paden, C. Leuschen, J. Li, F. Rodriguez-Morales, D. Braaten, K. Purdon, Z. Wang, W. Liu, et al. Bed topography of Jakobshavn Isbræ, Greenland, and Byrd Glacier, Antarctica. *Journal of Glaciology*, 60(223):813–833, 2014.

D. N. Goldberg. A variationally derived, depth-integrated approximation to a higher-order glaciological flow model. *Journal of Glaciology*, 57(201):157–170, 2011.

O. Gonzalez and A. M. Stuart. *A first course in continuum mechanics*. Cambridge University Press, 2008.

S. L. Greenwood and J. Kleman. Glacial landforms of extreme size in the Keewatin sector of the Laurentide Ice Sheet. *Quaternary Science Reviews*, 29(15):1894–1910, 2010.

R. Greve. Application of a polythermal three-dimensional ice sheet model to the Greenland ice sheet: response to steady-state and transient climate scenarios. *Journal of Climate*, 10(5):901–918, 1997.

R. Greve and H. Blatter. *Dynamics of ice sheets and glaciers*. Springer Science & Business Media, 2009.

G. H. Gudmundsson. Transmission of basal variability to a glacier surface. *J. Geophys. Res. [Solid Earth]*, 108(B5), 2003.

G. H. Gudmundsson, C. F. Raymond, and R. Bindschadler. The origin and longevity of flow stripes on Antarctic ice streams. *Annals of Glaciology*, 27(1):145–152, 1998.

M. Habermann, D. Maxwell, and M. Truffer. Reconstruction of basal properties in ice sheets using iterative inverse methods. *J. Glaciol.*, 58(210):795–807, 1 Aug. 2012.

M. Habermann, M. Truffer, and D. Maxwell. Changing basal conditions during the speed-up of Jakobshavn Isbræ, Greenland. *The Cryosphere*, 7(6):1679–1692, 2013.

D. B. Haidvogel, H. Arango, W. P. Budgell, B. D. Cornuelle, E. Curchitser, E. Di Lorenzo, K. Fennel, W. R. Geyer, A. J. Hermann, L. Lanerolle, et al. Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. *Journal of Computational Physics*, 227(7): 3595–3624, 2008.

P. C. Hansen. *The L-curve and its use in the numerical treatment of inverse problems*. IMM, Department of Mathematical Modelling, Technical University of Denmark, 1999.

A. S. Hein, J. Woodward, S. M. Marrero, S. A. Dunning, E. J. Steig, S. P. Freeman, F. M. Stuart, K. Winter, M. J. Westoby, and D. E. Sugden. Evidence for the stability of the West Antarctic Ice Sheet divide for 1.4 million years. *Nature communications*, 7, 2016.

M. A. Heroux, R. A. Bartlett, V. E. Howle, R. J. Hoekstra, J. J. Hu, T. G. Kolda, R. B. Lehoucq, K. R. Long, R. P. Pawlowski, E. T. Phipps, et al. An overview of the Trilinos project. *ACM Transactions on Mathematical Software (TOMS)*, 31(3): 397–423, 2005.

R. C. Hindmarsh. Drumlinization and drumlin-forming instabilities: viscous till mechanisms. *Journal of Glaciology*, 44(147):293–314, 1998a.

R. C. Hindmarsh. Ice-stream surface texture, sticky spots, waves and breathers: the coupled flow of ice, till and water. *Journal of Glaciology*, 44(148):589–614, 1998b.

R. C. Hindmarsh. The stability of a viscous till sheet coupled with ice flow, considered at wavelengths less than the ice thickness. *Journal of Glaciology*, 44(147): 285–292, 1998c.

R. C. Hindmarsh. Coupled ice–till dynamics and the seeding of drumlins and bedrock forms. *Annals of Glaciology*, 28(1):221–230, 1999.

D. M. Holland, R. H. Thomas, B. De Young, M. H. Ribergaard, and B. Lyberth. Acceleration of Jakobshavn Isbrae triggered by warm subsurface ocean waters. *Nature geoscience*, 1(10):659–664, 2008.

P. R. Holland, H. F. Corr, D. G. Vaughan, A. Jenkins, and P. Skvarca. Marine ice in Larsen ice shelf. *Geophysical Research Letters*, 36(11), 2009.

J. W. Holt, D. D. Blankenship, D. L. Morse, D. A. Young, M. E. Peters, S. D. Kempf, T. G. Richter, D. G. Vaughan, and H. F. Corr. New boundary conditions for the West Antarctic Ice Sheet: Subglacial topography of the Thwaites and Smith glacier catchments. *Geophysical Research Letters*, 33(9), 2006.

I. Howat, A. Negrete, and B. Smith. The Greenland Ice Mapping Project (GIMP) land classification and surface elevation data sets. *The Cryosphere*, 8(4):1509–1518, 2014.

C. Hulbe and M. Fahnestock. Century-scale discharge stagnation and reactivation of the Ross ice streams, West Antarctica. *J. Geophys. Res.*, 2007.

C. L. Hulbe and M. A. Fahnestock. West Antarctic ice-stream discharge variability: mechanism, controls and pattern of grounding-line retreat. *Journal of Glaciology*, 50(171):471–484, 2004.

K. Iglberger, G. Hager, J. Treibig, and U. Rüde. Expression templates revisited: a performance analysis of current methodologies. *SIAM Journal on Scientific Computing*, 34(2):C42–C69, 2012.

A. Iken, K. Echelmeyer, W. Harrison, and M. Funk. Mechanisms of fast flow in Jakobshavns Isbræ, West Greenland. I: Measurements of temperature and water level in deep boreholes. *J. Glac.*, 39(131):15–25, 1993.

T. Isaac, N. Petra, G. Stadler, and O. Ghattas. Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *Journal of Computational Physics*, 296:348–368, 2015.

D. Jansen, B. Kulessa, P. Sammonds, A. Luckman, E. King, and N. F. Glasser. Present stability of the Larsen C ice shelf, Antarctic Peninsula. *Journal of Glaciology*, 56 (198):593–600, 2010.

E. T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4): 620, 1957.

E. T. Jaynes. Information theory and statistical mechanics (Notes by the lecturer). In *Statistical Physics 3*, volume 1, pages 181–218, 1963.

E. T. Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

A. Jenkins, P. Dutrieux, S. S. Jacobs, S. D. McPhail, J. R. Perrett, A. T. Webb, and D. White. Observations beneath Pine Island Glacier in West Antarctica and implications for its retreat. *Nature Geoscience*, 3(7):468–472, 2010.

K. C. Jezek. RADARSAT-1 Antarctic Mapping Project: change-detection and surface velocity campaign. *Annals of Glaciology*, 34(1):263–268, 2002.

I. Joughin. Ice-sheet velocity mapping: a combined interferometric and speckle-tracking approach. *Annals of Glaciology*, 34(1):195–201, 2002.

I. Joughin, S. Tulaczyk, R. Bindschadler, and S. F. Price. Changes in west Antarctic ice stream velocities: Observation and analysis. *J. Geophys. Res.*, 107(B11):2289, 1 Nov. 2002.

I. Joughin, W. Abdalati, and M. Fahnestock. Large fluctuations in speed on Greenland's Jakobshavn Isbrae glacier. *Nature*, 432(7017):608–610, 2004a.

I. Joughin, D. R. MacAyeal, and S. Tulaczyk. Basal shear stress of the Ross ice streams from control method inversions. *J. Geophys. Res. [Solid Earth]*, 109(B9), 2004b.

I. Joughin, S. B. Das, M. A. King, B. E. Smith, I. M. Howat, and T. Moon. Seasonal speedup along the western flank of the Greenland Ice Sheet. *Science*, 320(5877): 781–783, 2008a.

I. Joughin, I. M. Howat, M. Fahnestock, B. Smith, W. Krabill, R. B. Alley, H. Stern, and M. Truffer. Continued evolution of Jakobshavn Isbrae following its rapid speedup. *Journal of Geophysical Research: Earth Surface*, 113(F4), 2008b.

I. Joughin, S. Tulaczyk, J. L. Bamber, D. Blankenship, J. W. Holt, T. Scambos, and D. G. Vaughan. Basal conditions for Pine Island and Thwaites Glaciers, West Antarctica, determined using satellite and airborne data. *Journal of Glaciology*, 55(190):245–257, 2009.

I. Joughin, B. E. Smith, I. M. Howat, D. Floricioiu, R. B. Alley, M. Truffer, and M. Fahnestock. Seasonal to decadal scale variations in the surface velocity of Jakobshavn Isbrae, Greenland: Observation and model-based analysis. *Journal of Geophysical Research: Earth Surface*, 117(F2), 2012.

I. Joughin, B. Smith, D. Shean, and D. Floricioiu. Brief communication: Further summer speedup of Jakobshavn Isbræ. *The Cryosphere*, 8:209–214, 2014a.

I. Joughin, B. E. Smith, and B. Medley. Marine ice sheet collapse potentially under way for the Thwaites Glacier Basin, West Antarctica. *Science*, 344(6185):735–738, 2014b.

I. R. Joughin, S. Tulaczyk, and H. F. Engelhardt. Basal melt beneath Whillans ice stream and ice streams A and C, West Antarctica. *Annals of Glaciology*, 36(1): 257–262, 2003.

G. Jouvet. A multilayer ice-flow model generalising the shallow shelf approximation. *Journal of Fluid Mechanics*, 764:26–51, 2015.

B. Kamb. Rheological nonlinearity and flow instability in the deforming bed mechanism of ice stream motion. *J. Geophys. Res.*, 96(B10):16585–16595, 10 Sept. 1991.

B. Kamb and E. Lachapelle. Direct observation of the mechanism of glacier sliding over bedrock. *J. Glaciol.*, 5:159–172, 1964.

A. Khazendar, E. Rignot, and E. Larour. Larsen B Ice Shelf rheology preceding its disintegration inferred by a control method. *Geophysical Research Letters*, 34(19), 2007.

A. Khazendar, E. Rignot, and E. Larour. Roles of marine ice, rheology, and fracture in the flow and stability of the Brunt/Stancomb-Wills Ice Shelf. *Journal of Geophysical Research: Earth Surface*, 114(F4), 2009.

A. Khazendar, E. Rignot, and E. Larour. Acceleration and spatial rheology of Larsen C ice shelf, Antarctic Peninsula. *Geophysical Research Letters*, 38(9), 2011.

A. I. Khinchin. *Mathematical Foundations of Information Theory*, volume 434. Courier Corporation, 1957.

J. Krug, J. Weiss, O. Gagliardini, and G. Durand. Combining damage and fracture mechanics to model calving. *The Cryosphere*, 8(6):2101–2117, 2014.

E. Larour, E. Rignot, I. Joughin, and D. Aubry. Rheology of the Ronne Ice Shelf, Antarctica, inferred from satellite radar interferometry data using an inverse control method. *Geophysical Research Letters*, 32(5), 2005.

E. Larour, H. Seroussi, M. Morlighem, and E. Rignot. Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM). *Journal of Geophysical Research: Earth Surface*, 117(F1), 2012.

E. Larour, A. Khazendar, C. Borstad, H. Seroussi, M. Morlighem, and E. Rignot. Representation of sharp rifts and faults mechanics in modeling ice shelf flow dynamics: Application to Brunt/Stancomb-Wills Ice Shelf, Antarctica. *Journal of Geophysical Research: Earth Surface*, 119(9):1918–1935, 2014.

A. Logg, K.-A. Mardal, and G. Wells. *Automated solution of differential equations by the finite element method: The FEniCS book*, volume 84. Springer Science & Business Media, 2012.

M. Lüthi, M. Funk, A. Iken, S. Gogineni, and M. Truffer. Mechanisms of fast flow in Jakobshavn Isbrae, West Greenland: Part III. Measurements of ice deformation, temperature and cross-borehole conductivity in boreholes to the bedrock. *Journal of Glaciology*, 48(162):369–385, 2002.

M. B. Lythe and D. G. Vaughan. BEDMAP: A new ice thickness and subglacial topographic model of Antarctica. *J. Geophys. Res.*, 106(B6):11335–11351, 10 June 2001.

D. R. MacAyeal. The basal stress distribution of Ice Stream E, Antarctica, inferred by control methods. *Journal of Geophysical Research: Solid Earth*, 97(B1):595–603, 1992.

D. R. MacAyeal, R. A. Bindschadler, and T. A. Scambos. Basal friction of Ice Stream E, West Antarctica. *J. Glaciol.*, 41(138):247–262, 1 Jan. 1995.

D. R. MacAyeal, V. Rommelaere, P. Huybrechts, C. L. Hulbe, J. Determann, and C. Ritz. An ice-shelf model test based on the Ross Ice Shelf, Antarctica. *Annals of Glaciology*, 23(1):46–51, 1996.

J. A. MacGregor, M. A. Fahnestock, G. A. Catania, A. Aschwanden, G. D. Clow, W. T. Colgan, S. P. Gogineni, M. Morlighem, S. M. Nowicki, J. D. Paden, et al. A synthesis of the basal thermal state of the Greenland Ice Sheet. *Journal of Geophysical Research: Earth Surface*, 121(7):1328–1350, 2016.

M. Martin, R. Winkelmann, M. Haseloff, T. Albrecht, E. Bueler, C. Khroulev, and A. Levermann. The Potsdam Parallel Ice Sheet Model (PISM-PIK)–Part 2: Dynamic equilibrium simulation of the Antarctic ice sheet. *The Cryosphere*, 5(3): 727–740, 2011.

N. Martin and J. Monnier. Adjoint accuracy for the full-Stokes ice flow model: limits to the transmission of basal friction variability to the surface. *The Cryosphere*, 8: 721–741, 2014.

M. Meier and A. Post. Fast tidewater glaciers. *Journal of Geophysical Research: Solid Earth*, 92(B9):9051–9058, 1987.

M. Morlighem, E. Rignot, H. Seroussi, E. Larour, H. Ben Dhia, and D. Aubry. Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica. *Geophysical Research Letters*, 37(14), 2010.

M. Morlighem, E. Rignot, H. Seroussi, E. Larour, H. Ben Dhia, and D. Aubry. A mass conservation approach for mapping glacier ice thickness. *Geophysical Research Letters*, 38(19), 2011.

M. Morlighem, E. Rignot, J. Mouginot, H. Seroussi, and E. Larour. Deeply incised submarine glacial valleys beneath the Greenland ice sheet. *Nature Geoscience*, 7 (6), 2014.

R. J. Motyka, M. Truffer, M. Fahnestock, J. Mortensen, S. Rysgaard, and I. Howat. Submarine melting of the 1985 Jakobshavn Isbræ floating tongue and the triggering of the current retreat. *Journal of Geophysical Research: Earth Surface*, 116 (F1), 2011.

T. Murray. Assessing the paradigm shift: Deformable glacier beds. *Quat. Sci. Rev.*, 16(9):995–1016, 1 Jan. 1997.

R. L. Parker. *Geophysical Inverse Theory*. Princeton University Press, 1994.

W. Paterson. Why ice-age ice is sometimes "soft". *Cold Regions Science and Technology*, 20(1):75–98, 1991.

F. Pattyn, L. Perichon, A. Aschwanden, B. Breuer, B. De Smedt, O. Gagliardini, G. H. Gudmundsson, R. Hindmarsh, A. Hubbard, J. V. Johnson, et al. Benchmark experiments for higher-order and full Stokes ice sheet models (ISMIP-HOM). *The Cryosphere Discussions*, 2(1):111–151, 2008.

M. Perego, M. Gunzburger, and J. Burkardt. Parallel finite-element implementation for higher-order ice-sheet models. *Journal of Glaciology*, 58(207):76–88, 2012.

N. Petra, J. Martin, G. Stadler, and O. Ghattas. A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1525–A1555, 2014.

K. Poinar, I. Joughin, S. B. Das, M. D. Behn, J. Lenaerts, and M. R. Broeke. Limits to future expansion of surface-melt-enhanced ice flow into the interior of western Greenland. *Geophysical Research Letters*, 42(6):1800–1807, 2015.

D. Pollard, R. M. DeConto, and R. B. Alley. Potential Antarctic Ice Sheet retreat driven by hydrofracturing and ice cliff failure. *Earth and Planetary Science Letters*, 412:112–121, 2015.

M. R. Pralong and G. H. Gudmundsson. Bayesian estimation of basal conditions on Rutford Ice Stream, West Antarctica, from surface data. *Journal of Glaciology*, 57 (202):315–324, 2011.

W. Rack and H. Rott. Pattern of retreat and disintegration of the Larsen B ice shelf, Antarctic Peninsula. *Annals of Glaciology*, 39(1):505–510, 2004.

L. Rasmussen. Bed topography and mass-balance distribution of Columbia Glacier, Alaska, USA, determined from sequential aerial photography. *Journal of Glaciology*, 34(117):208–216, 1988.

C. Raymond. Temperate valley glaciers. *Dynamics of snow and ice masses*, pages 79–139, 1980.

M. J. Raymond and G. H. Gudmundsson. Estimating basal properties of ice streams from surface measurements: a non-linear Bayesian inverse approach applied to synthetic data. *Cryosphere*, 3(2):265–278, 2009.

V. Rommelaere and D. R. Macayeal. Large-scale rheology of the Ross Ice Shelf, Antarctica, computed by a control method. *Ann. Glaciol.*, 24:43–48, 1997.

L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

C. Ryser, M. P. Luethi, L. C. Andrews, M. J. Hoffman, G. A. Catania, R. L. Hawley, T. A. Neumann, and S. S. Kristensen. Sustained high basal motion of the Greenland ice sheet revealed by borehole deformation. *Journal of Glaciology*, 60(222):647–660, 2014.

T. Scambos, C. Hulbe, and M. Fahnestock. Climate-induced ice shelf disintegration in the Antarctic peninsula. *Antarctic Peninsula Climate Variability: Historical and Paleoenvironmental Perspectives*, pages 79–92, 2003.

T. A. Scambos, J. Bohlander, C. u. Shuman, and P. Skvarca. Glacier acceleration and thinning after ice shelf collapse in the Larsen B embayment, Antarctica. *Geophysical Research Letters*, 31(18), 2004.

C. Schoof. Ice sheet grounding line dynamics: Steady states, stability, and hysteresis. *Journal of Geophysical Research: Earth Surface*, 112(F3), 2007.

C. Schoof and R. C. Hindmarsh. Thin-film flows with wall slip: an asymptotic analysis of higher order glacier flow models. *QJ Mech. Appl. Math*, 63(1):73–114, 2010.

H. Seddik, R. Greve, T. Zwinger, F. Gillet-Chaulet, and O. Gagliardini. Simulations of the Greenland ice sheet 100 years into the future with the full Stokes model Elmer/Ice. *Journal of Glaciology*, 58(209):427–440, 2012.

O. Sergienko, T. T. Creyts, and R. Hindmarsh. Similarity of organized patterns in driving and basal stresses of Antarctic and Greenland ice sheets beneath extensive areas of basal sliding. *Geophysical Research Letters*, 41(11):3925–3932, 2014.

O. V. Sergienko and R. C. Hindmarsh. Regular patterns in frictional resistance of ice-stream beds seen by surface data inversion. *Science*, 342(6162):1086–1089, 2013.

D. R. Shapero, I. R. Joughin, K. Poinar, M. Morlighem, and F. Gillet-Chaulet. Basal resistance for three of the largest Greenland outlet glaciers. *Journal of Geophysical Research: Earth Surface*, 121(1):168–180, 2016.

A. Shepherd, D. J. Wingham, J. A. Mansley, and H. F. Corr. Inland thinning of Pine Island Glacier, West Antarctica. *Science*, 291(5505):862–864, 2 Feb. 2001.

C. R. Stokes, M. Margold, and T. T. Creyts. Ribbed bedforms on palaeo-ice stream beds resemble regular patterns of basal shear stress ('traction ribs') inferred from modern ice streams. *Journal of Glaciology*, 62(234):696–713, 2016.

B. Stroustrup. *The C++ programming language*. Pearson Education, 2013.

R. H. Thomas. Force-perturbation analysis of recent thinning and acceleration of Jakobshavn Isbrae, Greenland. *Journal of Glaciology*, 50(168):57–66, 2004.

R. H. Thomas, W. Abdalati, E. Frederick, W. B. Krabill, S. Manizade, and K. Steffen. Investigation of surface melting and dynamic thinning on Jakobshavn Isbrae, Greenland. *Journal of Glaciology*, 49(165):231–239, 2003.

T. Thorsteinsson, J. Kipfstuhl, and H. Miller. Textures and fabrics in the GRIP ice core. *Journal of Geophysical Research: Oceans*, 102(C12):26583–26599, 1997.

T. Thorsteinsson, E. Waddington, K. Taylor, R. Alley, and D. Blankenship. Strain-rate enhancement at Dye 3, Greenland. *Journal of Glaciology*, 45(150):338–345, 1999.

F. Trèves. *Basic linear partial differential equations*, volume 62. Academic Press, 1975.

M. Truffer and K. A. Echelmeyer. Of isbrae and ice streams. *Annals of Glaciology*, 36(1):66–72, 2003.

A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 2000.

C. Van der Veen. Fracture mechanics approach to penetration of surface crevasses on glaciers. *Cold Regions Science and Technology*, 27(1):31–47, 1998.

C. J. Van Der Veen, J. Plummer, and L. Stearns. Controls on the recent speed-up of Jakobshavn Isbræ, West Greenland. *Journal of Glaciology*, 57(204):770–782, 2011.

D. Vandevoorde and N. M. Josuttis. *C++ Templates: The Complete Guide*. Addison-Wesley Professional, 2003.

D. Vaughan, J. Comiso, I. Allison, J. Carrasco, G. Kaser, R. Kwok, P. Mote, T. Murray, F. Paul, J. Ren, E. Rignot, O. Solomina, K. Steffen, and T. Zhang. *Observations: Cryosphere*, book section 4, pages 317 – 382. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN ISBN 978-1-107-66182-0. doi: 10.1017/CBO9781107415324.012. URL www.climatechange2013.org.

D. G. Vaughan, H. F. Corr, F. Ferraccioli, N. Frearson, A. O'Hare, D. Mach, J. W. Holt, D. D. Blankenship, D. L. Morse, and D. A. Young. New boundary conditions for the West Antarctic ice sheet: Subglacial topography beneath Pine Island Glacier. *Geophysical Research Letters*, 33(9), 2006.

T. Veldhuizen. Expression templates. *C++ Report*, 7(5):26–31, 1995.

A. Vieli and A. J. Payne. Application of control methods for modelling the flow of Pine Island Glacier, West Antarctica. *Ann. Glaciol.*, 36(1):197–204, 1 Jan. 2003.

A. Vieli, A. J. Payne, Z. Du, and A. Shepherd. Numerical modelling and data assimilation of the Larsen B ice shelf, Antarctic Peninsula. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 364(1844):1815–1839, 2006.

A. Vieli, A. Payne, A. Shepherd, and Z. Du. Causes of pre-collapse changes of the Larsen B ice shelf: Numerical modelling and assimilation of satellite observations. *Earth and Planetary Science Letters*, 259(3):297–306, 2007.

C. R. Vogel. *Computational methods for inverse problems*, volume 23. SIAM, 2002.

J. Weertman. On the sliding of glaciers. *J. Glaciol.*, 1957.

H. J. Zwally, W. Abdalati, T. Herring, K. Larson, J. Saba, and K. Steffen. Surface melt-induced acceleration of Greenland ice-sheet flow. *Science*, 297(5579):218–222, 2002.

Acknowledgements

To my parents, and to Joyce and Sharon: for their love and support.
To Zack and Amir: I can't believe I've been fooled this long!
To Arthur and Shujatte: Tssssss...
To Marcus: For your strong opinions.
To Omar: For the surprise of a lifetime.
To Andrew and Ister: For being very bon.
To Chris: For all the beatboxing in the office.
To Mau, McCoy, Miguel, Rodrigo: For the party that lasted years and years.
To Danny: For the inspiration in climbing, and many other things.
To my grandmother, Lillian Ravdin, who showed us by her example what real love is: defending her husband from an armed robber with a lamp.