

# An inexact Gauss–Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model

Noemi PETRA,<sup>1</sup> Hongyu ZHU,<sup>1</sup> Georg STADLER,<sup>1</sup> Thomas J.R. HUGHES,<sup>1,2</sup>  
Omar GHATTAS<sup>1,3,4</sup>

<sup>1</sup>*Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX, USA*  
E-mail: noemi@ices.utexas.edu

<sup>2</sup>*Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX, USA*

<sup>3</sup>*Jackson School of Geosciences, University of Texas at Austin, Austin, TX, USA*

<sup>4</sup>*Department of Mechanical Engineering, University of Texas at Austin, Austin, TX, USA*

**ABSTRACT.** We propose an infinite-dimensional adjoint-based inexact Gauss–Newton method for the solution of inverse problems governed by Stokes models of ice sheet flow with nonlinear rheology and sliding law. The method is applied to infer the basal sliding coefficient and the rheological exponent parameter fields from surface velocities. The inverse problem is formulated as a nonlinear least-squares optimization problem whose cost functional is the misfit between surface velocity observations and model predictions. A Tikhonov regularization term is added to the cost functional to render the problem well-posed and account for observational error. Our findings show that the inexact Newton method is significantly more efficient than the nonlinear conjugate gradient method and that the number of Stokes solutions required to solve the inverse problem is insensitive to the number of inversion parameters. The results also show that the reconstructions of the basal sliding coefficient converge to the exact sliding coefficient as the observation error (here, the noise added to synthetic observations) decreases, and that a nonlinear rheology makes the reconstruction of the basal sliding coefficient more difficult. For the inversion of the rheology exponent field, we find that horizontally constant or smoothly varying parameter fields can be reconstructed satisfactorily from noisy observations.

## 1. INTRODUCTION

Here we consider the problem of estimating unknown parameters characterizing an ice sheet flow model governed by the nonlinear full-Stokes equations from surface flow observations. The central aims of the paper are (1) to present an efficient method for the solution of ice sheet flow inverse problems that is suitable for high-dimensional parameter spaces and (2) to employ this method to study how well finite-amplitude perturbation of a sliding coefficient in the basal boundary condition, and rheological parameters in the constitutive equation, can be recovered from surface observations that contain some degree of error.

Ice sheet flow models usually contain parameters that are unknown, due either to our inability to directly observe them, or their role as phenomenological parameters that must be constrained by data. The parameter field that relates the basal traction to the rate of basal slip arguably presents the largest uncertainty in determining the rate of ice flow. The basal sliding coefficient cannot be obtained from direct observations; instead, we may seek to infer it from surface flow data. Rheological parameters can be estimated in the laboratory, but their values may not adequately characterize field ice. By treating them as uncertain parameters to be inferred from data, we may be able to improve models of field ice sheets. Additional unknown or uncertain inputs to ice sheet models include the geothermal heat flux, the basal topography and the ice thickness.

Our goal is to devise inversion methods that target three-dimensional (3-D), large-scale inverse ice sheet problems with linear or nonlinear rheology, handle arbitrary geometry and boundary conditions, account for errors in the observations, and are capable of inverting for basal as well

as rheology parameters. We formulate the inverse problem as an infinite-dimensional optimization problem governed by the nonlinear Stokes equations. The cost functional we minimize is the sum of the squared misfit between observed and predicted surface velocities and a regularization term that makes the ill-posed inverse problem well-posed. The basal sliding and Glen's flow law exponent coefficients constitute the inversion parameter fields.

Discretization of this infinite-dimensional inverse problem results in a large-scale numerical optimization. Gradient-based methods offer the only hope of solving such high-dimensional optimization problems. Here we advocate (inexact, Gauss–)Newton methods, which employ Hessian information (i.e. second derivatives or curvature) to greatly speed up convergence relative to gradient-only methods (Dennis and Schnabel, 1996; Kelley, 1999; Nocedal and Wright, 2006). In many cases, convergence is obtained in a number of iterations independent of problem size (Deuffhard, 2004). We provide systematic derivations of the gradient and Hessian for both basal and rheology parameters, which invoke the solution of associated adjoint Stokes problems. Explicitly constructing the Hessian matrix at each Newton iteration is usually intractable for inverse problems governed by expensive forward simulations, since each column of the Hessian (which corresponds to a distinct parameter) requires solution of a linearized forward problem. Here, however, we do not explicitly construct the Hessian matrix; instead, we solve the linear system arising at each Newton iteration by the (linear) conjugate gradient (CG) method, which requires only the application of the Hessian matrix to a vector at each CG iteration. This requires solution of a pair of forward/adjoint linearized Stokes problems, and

makes the method well suited to large-scale problems when the number of CG iterations is small, as is often the case for inverse problems.

Several studies have focused on inverse problems for the basal boundary conditions in ice sheet models. There are three approaches commonly used to compute derivatives in this case. The first approach, which is related to the methods presented in this paper, uses adjoint equations to compute first derivatives with respect to basal sliding parameters. The approach was introduced in the context of glaciology problems by MacAyeal (1993), who employed the framework of optimal control theory. Applications of this method to inverse ice sheet flow problems governed by simplifications of the full-Stokes equations are presented by Vieli and Payne (2003), Joughin and others (2004), Larour and others (2005), Morlighem and others (2010) and Goldberg and Sergienko (2011). These contributions make use of the steepest descent or the nonlinear conjugate gradient method to solve the optimization, both of which use gradient information only and are thus slower than Newton-type methods.

A second approach to computing derivatives uses analytical transfer functions to describe the effects of two-dimensional, small-amplitude perturbations in basal conditions on the ice surface velocity (Gudmundsson, 2003; Thorsteinsson and others, 2003; Raymond and Gudmundsson, 2005; Gudmundsson and Raymond, 2008). Raymond (2007), Raymond and Gudmundsson (2009) and Pralong and Gudmundsson (2011) use the full-Stokes equations to model ice flow, but the derivatives needed in the inversion procedure are approximated through analytical transfer functions, which limits the accuracy of the gradient.

A third method uses gradients of a different cost functional, which compares the flow fields of two Stokes problems, one with Neumann and the other with Dirichlet surface boundary conditions (Arthern and Gudmundsson, 2010). The Dirichlet problem uses the observation data as boundary condition, while the Neumann problem satisfies given (zero) traction conditions. A gradient-based minimization algorithm adjusts the inversion parameters until the difference between the Dirichlet and the Neumann problem is small. Note that an appropriate termination of this iteration is important, especially since the observations always contain some noise. An attempt to overcome this problem by introducing a regularization term was recently made by Jay-Allemand and others (2011). This Dirichlet–Neumann problem approach, which has similarities to the method proposed by Maxwell and others (2008), only requires a Stokes solver that implements different types of boundary conditions. However, the method is derived for linear rheology and a basal sliding law, in which case the adjoint Stokes equations, as required in the first approach described above, also reduce to the usual linear Stokes equations. This method has also been applied successfully to inverse problems with nonlinear rheology (Arthern and Gudmundsson, 2010; Jay-Allemand and others, 2011), but it is not guaranteed to work for these cases.

To the best of our knowledge, inversion for the exponent parameter in Glen's flow law, and in particular its spatial variability, has not been addressed previously. Here we wish to develop the capability to construct a more realistic description of the ice viscosity as influenced by impurities and/or weakening and damage of the ice, which may not be accounted for by constant rheological parameters. Note that a first step towards addressing this problem has been taken by Arthern and Gudmundsson (2010), where they invert for

the spatially varying effective ice viscosity,  $\eta$ , using a linear rheology (employing the third approach described above). We go beyond this prior work by inverting for the Glen's flow law exponent,  $n$ , under a nonlinear rheology (using the fast Gauss–Newton method advocated here). Alternatively, one could also invert for parameters in the rate factor in Glen's flow law, which depends on the ice temperature, using the methodology presented here.

The remaining sections of this paper are organized as follows. We begin (Section 2) by describing the forward ice sheet problem and the corresponding inverse problem for the basal sliding coefficient field and for the flow law exponent parameter. In Section 3, we present the adjoint-based inexact Gauss–Newton method for solving the inversion problem.

Inversion and performance results are given in Section 4, where we compare the performance of the proposed method with the nonlinear conjugate gradient (NCG) method. Our results show that compared with NCG, the number of iterations (and thus the number of Stokes solutions) required by the Newton method is significantly smaller. We also find that the number of Stokes solutions for the Newton method is insensitive to the number of inversion parameters, a crucial requirement for the efficient solution of large-scale problems. Next, we study the quality of basal inversions for linear and nonlinear rheology and for different levels of observational error. The results show that the quality of the reconstruction of the basal coefficient depends on the wavelength of the variation in the true basal sliding coefficient, and on the noise level. Moreover, the reconstruction is less accurate for nonlinear rheology than for linear, in agreement with earlier work based on analytical transfer functions (Raymond, 2007; Gudmundsson and Raymond, 2008). Next, we invert for the exponent in Glen's flow law. Although the problem is highly underdetermined, due to the ratio of surface observations to volume parameters, we find good reconstructions for the location of viscosity anomalies and for smoothly varying exponent fields. Section 5 summarizes our findings. Finally, in the Appendices we give systematic derivation of the adjoint and incremental equations used to compute the gradient and the Hessian, and discuss discretization and Newton's method for the nonlinear Stokes equations.

## 2. PROBLEM FORMULATION

### 2.1. The forward problem

We model the flow of ice as a non-Newtonian, viscous, incompressible, isothermal fluid. The balance of mass and linear momentum state that (Hutter, 1983; Paterson, 1994; Marshall, 2005)

$$\nabla \cdot \mathbf{u} = 0 \quad (1a)$$

$$-\nabla \cdot \boldsymbol{\sigma}_u = \rho \mathbf{g}, \quad (1b)$$

where  $\mathbf{u} = (u_1, u_2, u_3)^T$  denotes the velocity vector,  $\boldsymbol{\sigma}_u$  the stress tensor,  $\rho$  the density of the ice and  $\mathbf{g}$  gravity. The stress,  $\boldsymbol{\sigma}_u$ , can be decomposed as  $\boldsymbol{\sigma}_u = \boldsymbol{\tau}_u - \mathbf{I}p$ , where  $\boldsymbol{\tau}_u$  is the deviatoric stress tensor,  $p$  the pressure and  $\mathbf{I}$  the unit tensor. We employ a constitutive law for ice that relates stress and strain-rate tensors by Glen's flow law (Glen, 1955),

$$\boldsymbol{\tau}_u = 2\eta(\mathbf{u}, n)\dot{\boldsymbol{\epsilon}}_u, \quad \text{with } \eta(\mathbf{u}, n) = \frac{1}{2}A^{-\frac{1}{n}}\dot{\boldsymbol{\epsilon}}_u^{\frac{1-n}{2n}}, \quad (2)$$

where  $\eta$  is the effective viscosity,  $\dot{\mathbf{e}}_{\mathbf{u}} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$  the strain-rate tensor,  $\dot{\mathbf{e}}_{\mathbf{1}} = \frac{1}{2}\text{tr}(\dot{\mathbf{e}}_{\mathbf{u}}^2)$  its second invariant,  $n$  Glen's flow law exponent field and  $A$  the temperature-dependent flow rate factor (here taken as constant in isothermal ice).

Without loss of generality, we choose the model domain to be a 3-D rectangular section of an ice sheet with the following boundary conditions. On the top surface,  $\Gamma_t$ , we impose a traction-free boundary condition, while periodic boundary conditions are invoked along the lateral boundaries, which are denoted  $\Gamma_p$ . The boundary condition at the base of the ice sheet,  $\Gamma_b$ , is given by a Weertman-type nonlinear sliding law with normal and tangential components given by (Paterson, 1994)

$$\mathbf{u} \cdot \mathbf{n} = 0, \quad \mathbf{T} \boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n} + \beta |\mathbf{T} \mathbf{u}|^{m-1} \mathbf{T} \mathbf{u} = \mathbf{0}, \quad (3)$$

where  $m$  is the basal sliding exponent,  $\beta$  the basal sliding coefficient field defined on  $\Gamma_b$ ,  $\boldsymbol{\sigma}_{\mathbf{u}}$  the stress tensor defined previously and  $\mathbf{T} = \mathbf{I} - \mathbf{n} \otimes \mathbf{n}$  the tangential operator. Here, ' $\otimes$ ' represents the tensor (or outer) product,  $\mathbf{n}$  is the outward normal vector and  $\mathbf{I}$  is the second-order unit tensor. Note that  $\beta$ , which relates tangential velocity to tangential traction, subsumes several physical phenomena and thus does not itself represent a physical parameter. It depends on the frictional behavior of the ice sheet, on the roughness of the bedrock and, if present, on the depth of a plastically deforming layer of till between the ice sheet and the bedrock.

In summary, the (forward) nonlinear Stokes ice sheet model we consider in this paper is given by

$$\nabla \cdot \mathbf{u} = 0 \quad \text{in } \Omega, \quad (4a)$$

$$-\nabla \cdot [\eta(\mathbf{u}, n) (\nabla \mathbf{u} + \nabla \mathbf{u}^T) - \mathbf{l}p] = \rho \mathbf{g} \quad \text{in } \Omega, \quad (4b)$$

$$\mathbf{u}|_{\Gamma_1} = \mathbf{u}|_{\Gamma_r}, \text{ and } \boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n}|_{\Gamma_1} = \boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n}|_{\Gamma_r}, \quad \text{on } \Gamma_p, \quad (4c)$$

$$\boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n} = \mathbf{0} \quad \text{on } \Gamma_t, \quad (4d)$$

$$\mathbf{u} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_b, \quad (4e)$$

$$\mathbf{T} \boldsymbol{\sigma}_{\mathbf{u}} \mathbf{n} + \beta |\mathbf{T} \mathbf{u}|^{m-1} \mathbf{T} \mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_b, \quad (4f)$$

where the effective viscosity is given by Eqn (2) and the stress is given by  $\boldsymbol{\sigma}_{\mathbf{u}} = \eta(\mathbf{u}, n)(\nabla \mathbf{u} + \nabla \mathbf{u}^T) - \mathbf{l}p$ . In the expressions above, pairs of opposing boundaries on  $\Gamma_p$  on which periodic conditions are imposed are denoted generically by  $\Gamma_1$  and  $\Gamma_r$ . The choice of periodic boundary conditions for the lateral boundaries is, of course, arbitrary; the expressions for infinite-dimensional gradients and Hessians given in Section 3 can be derived for other boundary conditions using the same variational approach.

### 2.2. The inverse problem

The inverse problem is formulated as follows: given (possibly noisy) observations  $\mathbf{u}^{\text{obs}}$  of the velocity,  $\mathbf{u}$ , on the surface,  $\Gamma_t$ , of an ice sheet occupying a domain,  $\Omega$ , we wish to infer the sliding coefficient field,  $\beta$ , defined on the base of the ice sheet, and the exponent parameter field,  $n$ , in Glen's flow law defined within the ice volume, that best reproduce the observed velocity. This can be formulated as the following nonlinear least-squares minimization problem,

$$\min_{\beta, n} \mathcal{J}(\beta, n) = \frac{1}{2} \int_{\Gamma_t} |\mathbf{u}(\beta, n) - \mathbf{u}^{\text{obs}}|^2 \, ds + \mathcal{R}(\beta, n), \quad (5)$$

where  $\mathbf{u}$  depends on  $(\beta, n)$  through the solution of the nonlinear Stokes problem (Eqn (4)).

The first term in the cost functional,  $\mathcal{J}(\beta, n)$ , represents the misfit between the observed velocity field,  $\mathbf{u}^{\text{obs}}$ , and that predicted by the nonlinear Stokes model,  $\mathbf{u}$ . The regularization term,  $\mathcal{R}(\beta, n)$ , imposes regularity on the two inversion fields, such as smoothness in an appropriate norm. Often this reflects prior knowledge of the model parameters. In the absence of such a term, the inverse problem is ill-posed, i.e. its solution is not unique and is highly sensitive to errors in the observations (Engl and others, 1996; Vogel, 2002). As is common in inverse problems (and as is discussed in Section 4), small-wavelength components of the parameter fields  $\beta$  or  $n$  cannot be identified from surface observations; as a result, they can appear as arbitrary noise in the reconstructed parameter fields. As a remedy, we impose a Tikhonov regularization, which penalizes oscillatory components of  $\beta$  and  $n$ , thus restricting the solution to smoothly varying fields:

$$\mathcal{R}(\beta, n) = \frac{\gamma_{\beta}}{2} \int_{\Gamma_b} |\mathbf{T} \nabla \beta|^2 \, ds + \frac{\gamma_n}{2} \int_{\Omega} \nabla n \cdot \nabla n \, dx. \quad (6)$$

Here  $\gamma_{\beta} > 0$  and  $\gamma_n > 0$  are regularization parameters that control the strength of the imposed smoothness relative to the data misfit. Large values for  $\gamma_{\beta}$  and  $\gamma_n$  in Eqn (6) put the emphasis in the minimization problem (Eqn (5)) on  $\mathcal{R}$ , and thus the solution  $(\beta, n)$  has small gradients (i.e. is smooth). For small or vanishing regularization parameters,  $\mathcal{R}$  has little or no effect and the inverse problem is ill-posed (noise in the data can result in oscillations in the solution of the inverse problem).

### 3. ADJOINT-BASED INEXACT GAUSS-NEWTON METHOD FOR SOLUTION OF THE INVERSE PROBLEM

We now present an adjoint-based inexact (Gauss-)Newton method for solving the nonlinear least-squares optimization problem, Eqn (5). Starting with an initial guess for the parameter fields  $(\beta, n)$ , Newton's method iteratively updates these parameters based on successive quadratic approximations of the cost functional,  $\mathcal{J}$ , using gradient (i.e. first-derivative) and Hessian (i.e. second-derivative) information with respect to  $(\beta, n)$ . That is, the parameters are updated by

$$(\beta, n)_{\text{new}} = (\beta, n) + \alpha (\tilde{\beta}, \tilde{n}),$$

where  $(\beta, n)$  are the current model parameters and the Newton 'direction',  $(\tilde{\beta}, \tilde{n})$ , is obtained by minimizing the quadratic approximation, or equivalently solving the linear system

$$\mathcal{H}(\beta, n) (\tilde{\beta}, \tilde{n}) = -\mathcal{G}(\beta, n). \quad (7)$$

Here  $\mathcal{G}$  is the gradient of the regularized data misfit functional,  $\mathcal{J}$ , in Eqn (5), and  $\mathcal{H}$  is its Hessian operator. To provide robustness, the new values of the parameters are found by 'damping' the Newton direction, i.e. by choosing a step length,  $\alpha$ , such that the cost functional in Eqn (5) is sufficiently decreased at each iteration. The details of carrying out this so-called line search are described later in this section.

The gradient and Hessian in the Newton step (Eqn (7)) are in fact Fréchet derivatives, obtained at the infinite-dimensional level using variational calculus. Since evaluating the functional,  $\mathcal{J}$ , requires the solution of the forward Stokes ice sheet model (Eqn (4)), which depends on  $(\beta, n)$ , derivatives of  $\mathcal{J}$  with respect to  $(\beta, n)$  need to

take into account this implicit dependence as well. The main part of this section is devoted to the presentation of expressions for computing the gradient and Hessian using so-called adjoint methods. Here we simply state the resulting expressions; their derivations are given in Appendix A. Readers wanting more details on procedures for finding gradients and Hessians of cost functionals defined by solutions of partial differential equations (PDEs) may consult Borzi and Schulz (2012), Gunzburger (2003) and Tröltzsch (2010). All expressions in this section are given in infinite-dimensional form, which provides a natural and ‘clean’ path to computing the derivatives. These expressions can then be discretized using standard finite-element methods, as discussed in Appendix B.

To facilitate the computation of infinite-dimensional gradient and Hessians, we introduce a so-called ‘Lagrangian functional’:

$$\begin{aligned} \mathcal{L}(\mathbf{u}, p; \mathbf{v}, q; \beta, n) = & \mathcal{J}(\beta, n) + \int_{\Omega} 2\eta(\mathbf{u}, n) \dot{\boldsymbol{\epsilon}}_{\mathbf{u}} : \dot{\boldsymbol{\epsilon}}_{\mathbf{v}} \, dx \\ & - \int_{\Omega} (p \nabla \cdot \mathbf{v} + q \nabla \cdot \mathbf{u} + f \cdot \mathbf{v}) \, dx \\ & + \int_{\Gamma_b} \beta |\mathbf{T}\mathbf{u}|^{m-1} \mathbf{T}\mathbf{u} \cdot \mathbf{T}\mathbf{v} \, ds, \end{aligned}$$

which augments the regularized data misfit functional,  $\mathcal{J}$ , with additional terms comprising the weak form of the forward nonlinear Stokes problem (Eqn (4)). This weak form is obtained by multiplying the nonlinear Stokes system (Eqn (4)) with arbitrary test functions  $\mathbf{v}$  and  $q$ , integrating over the domain  $\Omega$ , and using integration by parts (Oden and Reddy, 1976; Braess, 1997; Gockenbach, 2006). The test functions establishing the weak form are known as Lagrange multipliers and, since they end up being identified with solutions of so-called adjoint (Stokes) problems, are also known as the ‘adjoint velocity’  $\mathbf{v}$  and the ‘adjoint pressure’  $q$ . Finally, in the above equation,  $\dot{\boldsymbol{\epsilon}}_{\mathbf{v}} = \frac{1}{2}(\nabla \mathbf{v} + \nabla \mathbf{v}^T)$  is the adjoint strain rate tensor and ‘:’ represents the scalar product of two tensors.

The gradient of  $\mathcal{J}$  can be found by requiring that variations of the Lagrangian,  $\mathcal{L}$ , with respect to the forward velocity and pressure  $(\mathbf{u}, p)$  and the adjoint velocity and pressure  $(\mathbf{v}, q)$  vanish. Variations with respect to  $(\beta, n)$  then result in the following strong form of the gradient,  $\mathcal{G}$ :

$$\mathcal{G}(\beta, n) = \begin{cases} -\nabla \cdot (\gamma_{\beta} \mathbf{T} \nabla \beta) + |\mathbf{T}\mathbf{u}|^{m-1} \mathbf{T}\mathbf{u} \cdot \mathbf{T}\mathbf{v} & \text{on } \Gamma_b, \\ (\gamma_{\beta} \mathbf{T} \nabla \beta) \cdot \bar{\mathbf{n}} & \text{on } \partial \Gamma_b, \\ -\nabla \cdot (\gamma_n \nabla n) + 2 \frac{d\eta}{dn}(\mathbf{u}, n) \dot{\boldsymbol{\epsilon}}_{\mathbf{u}} : \dot{\boldsymbol{\epsilon}}_{\mathbf{v}} & \text{in } \Omega, \\ (\gamma_n \nabla n) \cdot \mathbf{n} & \text{on } \partial \Omega. \end{cases} \quad (8)$$

Here  $\partial \Gamma_b$  is the boundary of the basal surface,  $\bar{\mathbf{n}}$  is the outward normal vector on  $\partial \Gamma_b$  and  $\frac{d\eta}{dn}(\mathbf{u}, n)$  is given by

$$\frac{d\eta}{dn}(\mathbf{u}, n) = \ln \left( A \frac{1}{n^2} \dot{\boldsymbol{\epsilon}}_{\mathbf{I}}^{-\frac{1}{2n^2}} \right) \eta(\mathbf{u}, n). \quad (9)$$

The velocity,  $\mathbf{u}$ , in Eqn (8) is obtained by solving the *forward Stokes problem* (Eqn (4)) for given  $(\beta, n)$ , and the adjoint velocity,  $\mathbf{v}$ , is obtained by solving the following *adjoint*

*Stokes problem* for given  $(\beta, n)$  and for  $\mathbf{u}$  satisfying Eqn (4):

$$\nabla \cdot \mathbf{v} = 0 \quad \text{in } \Omega, \quad (10a)$$

$$-\nabla \cdot \boldsymbol{\sigma}_{\mathbf{v}} = \mathbf{0} \quad \text{in } \Omega, \quad (10b)$$

$$\mathbf{v}|_{\Gamma_t} = \mathbf{v}|_{\Gamma_r} \text{ and } \boldsymbol{\sigma}_{\mathbf{v}} \mathbf{n}|_{\Gamma_t} = \boldsymbol{\sigma}_{\mathbf{v}} \mathbf{n}|_{\Gamma_r}, \quad \text{on } \Gamma_p, \quad (10c)$$

$$\boldsymbol{\sigma}_{\mathbf{v}} \mathbf{n} = \mathbf{u}^{\text{obs}} - \mathbf{u} \quad \text{on } \Gamma_t, \quad (10d)$$

$$\mathbf{v} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_b, \quad (10e)$$

$$\begin{aligned} & \mathbf{T} \boldsymbol{\sigma}_{\mathbf{v}} \mathbf{n} + \beta |\mathbf{T}\mathbf{u}|^{m-1} \mathbf{T}\mathbf{v} \\ & + \beta(m-1) |\mathbf{T}\mathbf{u}|^{m-3} (\mathbf{T}\mathbf{u} \otimes \mathbf{T}\mathbf{u}) \mathbf{T}\mathbf{v} = \mathbf{0} \quad \text{on } \Gamma_b. \end{aligned} \quad (10f)$$

Here the adjoint stress,  $\boldsymbol{\sigma}_{\mathbf{v}}$ , is given by

$$\boldsymbol{\sigma}_{\mathbf{v}} = 2\eta(\mathbf{u}, n) \left( \mathbf{I} + \frac{1-n}{n} \frac{\dot{\boldsymbol{\epsilon}}_{\mathbf{u}} \otimes \dot{\boldsymbol{\epsilon}}_{\mathbf{u}}}{\dot{\boldsymbol{\epsilon}}_{\mathbf{u}} : \dot{\boldsymbol{\epsilon}}_{\mathbf{u}}} \right) \dot{\boldsymbol{\epsilon}}_{\mathbf{v}} - \mathbf{I}q,$$

where  $\mathbf{I}$  is the fourth-order identity tensor and ‘ $\otimes$ ’ denotes the outer product between second-order tensors. As can be seen from Eqn (10), the adjoint Stokes problem is driven by the misfit between observed and predicted surface velocity on the top boundary, given by Eqn (10d); all other source terms are zero. Note that while the forward problem is a nonlinear Stokes problem, the adjoint Stokes problem (Eqn (10)) is linear in the adjoint velocity and pressure, and is characterized by a linearized Stokes operator with an effective anisotropy that depends on the forward velocity,  $\mathbf{u}$ , as well as a linearized basal boundary condition with coefficient depending on  $\mathbf{u}$ . The effective anisotropy of the adjoint Stokes operator can be seen in the expression for the adjoint stress,  $\boldsymbol{\sigma}_{\mathbf{v}}$ , above: the isotropic action of the unit tensor,  $\mathbf{I}$ , is reduced (since  $n \geq 1$ ) by the second term in the viscosity tensor, which acts only in the direction of the forward strain rate,  $\dot{\boldsymbol{\epsilon}}_{\mathbf{u}}$ . Note also that the adjoint Stokes operator is the same operator as the Jacobian that arises in Newton’s method for the forward (nonlinear) Stokes problem. This is because the forward problem can be derived from a (constrained) variational problem (Appendix C), and thus the Jacobian of the forward problem is self-adjoint. This means that any forward nonlinear Stokes ice sheet code based on a Newton solver is already equipped with the operator needed to solve the adjoint Stokes problem.

The computation of the gradient,  $\mathcal{G}$ , for a given  $(\beta, n)$  iterate proceeds as follows. First, given the current estimate of the parameter fields  $(\beta, n)$ , the forward nonlinear Stokes problem (Eqn (4)) is solved using Newton’s method, with a step size appropriately chosen such that the corresponding energy functional is decreased (Appendix C). The resulting forward velocity,  $\mathbf{u}$ , is then used to construct the effective anisotropic viscosity tensor, the basal boundary condition coefficient and the data misfit ‘source’ term for the adjoint Stokes problem (Eqn (10)). This adjoint Stokes problem is then solved (using the same linear solver used for the forward (linearized) Stokes problem; Section 4.1) to yield the adjoint velocity,  $\mathbf{v}$  (and adjoint pressure,  $q$ ). The forward and adjoint velocities, along with the current parameter field iterates  $(\beta, n)$ , then enter into the evaluation of the gradient  $\mathcal{G}$  in Eqn (8). The role of the forward and adjoint Stokes solutions in the Newton iteration is summarized in Algorithm 1.

The gradient computation (Eqn (8)) is inexpensive relative to the forward and adjoint Stokes solutions, since it does not require the solution of any additional linear systems. Solution of the adjoint Stokes problem (Eqn (10)), in turn, is inexpensive relative to the solution of the forward Stokes problem (Eqn (4)), since the latter is nonlinear and its solution

typically requires five to ten linear solutions, while the former is linear. Thus, the gradient can be found using this adjoint approach at minimal additional cost beyond solving the forward Stokes problem, independent of the number of (discretized) inversion parameters. The same cannot be said about a ‘direct’ method for computing the gradient (Hinze and others, 2009). Direct methods compute the sensitivity with respect to each parameter by solving a linear Stokes problem, and thus require as many additional linear Stokes solutions as there are discretized inversion parameters.

Now that the gradient computation forming the right-hand side of the Newton system (Eqn (7)) has been described, we present the computation of the Hessian operator,  $\mathcal{H}$ , on the left-hand side of the Newton equation. It is well known that the Newton direction is a descent direction only if the Hessian is positive definite. Unfortunately, even when the inverse problem is well posed,  $\mathcal{H}$  is guaranteed to be positive definite only close to a minimum (Nocedal and Wright, 2006). The remedy we apply here is to neglect terms in the Hessian expression that involve the adjoint variable (Appendix B). This leads to the so-called Gauss–Newton approximation of the Hessian, which (with appropriate regularization) is guaranteed to be positive definite. Since the adjoint system (Eqn (10)) is driven only by the data misfit,  $(\mathbf{u}^{\text{obs}} - \mathbf{u})$ , on the top boundary,  $\Gamma_t$ , the adjoint velocity is expected to be small when the data misfit is small, which occurs close to the solution of the inverse problem, provided the model or observational errors are not large. The Gauss–Newton Hessian is thus often a good approximation of the full Hessian. In such cases, even though one loses the strict quadratic convergence guarantee of Newton’s method, one can still obtain fast, superlinear convergence, as well as independence of the number of iterations from the number of inversion parameters, as will be seen in the numerical results in Section 4. In the remainder of this paper, we employ the Gauss–Newton approximation of the Hessian, but for simplicity we occasionally drop the term ‘Gauss’ when referring to the method.

Formally, when the Hessian operator in the Newton system (Eqn (7)) is discretized, it gives rise to a dense Hessian matrix of dimension equal to the number of inversion parameters (Appendix B). Computing each column of the Hessian requires solution of a linearized forward problem (Hinze and others, 2009). Therefore, explicitly forming and storing this matrix is not an option. Instead, we solve the Newton system (Eqn (7)) using the linear CG method, which does not require the explicit Hessian, instead requiring only the action of the Hessian on a vector at each CG iteration. Next, we present expressions for this Hessian action in terms of the solution of a pair of linearized forward and adjoint problems. These expressions are simply stated here; their derivation is presented in Appendix A (in the case of a linear rheology, and invoking the Gauss–Newton approximation).

The action of the Gauss–Newton Hessian operator in a given CG direction,  $(\tilde{\beta}, \tilde{n})$ , evaluated at the current Newton iterate,  $(\beta, n)$ , can be expressed as

$$\mathcal{H}(\beta, n)(\tilde{\beta}, \tilde{n}) = \begin{cases} -\nabla \cdot (\gamma_\beta \mathbf{T} \nabla \tilde{\beta}) + |\mathbf{T}\mathbf{u}|^{m-1} \mathbf{T}\mathbf{u} \cdot \mathbf{T}\tilde{\mathbf{v}} & \text{on } \Gamma_b, \\ (\gamma_\beta \mathbf{T} \nabla \tilde{\beta}) \cdot \tilde{\mathbf{n}} & \text{on } \partial\Gamma_b, \\ -\nabla \cdot (\gamma_n \nabla \tilde{n}) + 2 \frac{d\eta}{dn}(\mathbf{u}, n) \dot{\boldsymbol{\varepsilon}}_{\mathbf{u}} : \dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{v}}} & \text{in } \Omega, \\ (\gamma_n \nabla \tilde{n}) \cdot \mathbf{n} & \text{on } \partial\Omega, \end{cases} \quad (11)$$

where the *incremental forward velocity/pressure*  $(\tilde{\mathbf{u}}, \tilde{p})$  satisfy the *incremental forward Stokes problem*,

$$\nabla \cdot \tilde{\mathbf{u}} = 0 \quad \text{in } \Omega, \quad (12a)$$

$$-\nabla \cdot \boldsymbol{\sigma}_{\tilde{\mathbf{u}}} = \nabla \cdot \boldsymbol{\tau}_{\mathbf{u}} \quad \text{in } \Omega, \quad (12b)$$

$$\tilde{\mathbf{u}}|_{\Gamma_t} = \tilde{\mathbf{u}}|_{\Gamma_r}, \text{ and } \boldsymbol{\sigma}_{\tilde{\mathbf{u}}}\mathbf{n}|_{\Gamma_t} = \boldsymbol{\sigma}_{\tilde{\mathbf{u}}}\mathbf{n}|_{\Gamma_r} \quad \text{on } \Gamma_p, \quad (12c)$$

$$\boldsymbol{\sigma}_{\tilde{\mathbf{u}}}\mathbf{n} = -\boldsymbol{\tau}_{\mathbf{u}}\mathbf{n} \quad \text{on } \Gamma_t, \quad (12d)$$

$$\tilde{\mathbf{u}} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_b, \quad (12e)$$

$$\begin{aligned} & \mathbf{T}\boldsymbol{\sigma}_{\tilde{\mathbf{u}}}\mathbf{n} + \beta|\mathbf{T}\mathbf{u}|^{m-1}\mathbf{T}\tilde{\mathbf{u}} \\ & + \beta(m-1)|\mathbf{T}\mathbf{u}|^{m-3}(\mathbf{T}\mathbf{u} \otimes \mathbf{T}\mathbf{u})\mathbf{T}\tilde{\mathbf{u}} \\ & = -\tilde{\beta}|\mathbf{T}\mathbf{u}|^{m-1}\mathbf{T}\mathbf{u} - \mathbf{T}\boldsymbol{\tau}_{\mathbf{u}}\mathbf{n} \quad \text{on } \Gamma_b, \end{aligned} \quad (12f)$$

with  $\boldsymbol{\sigma}_{\tilde{\mathbf{u}}} = 2\eta(\mathbf{u}, n) \left(1 + \frac{1-n}{n} \frac{\dot{\boldsymbol{\varepsilon}}_{\mathbf{u}} \otimes \dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{u}}}}{\dot{\boldsymbol{\varepsilon}}_{\mathbf{u}} : \dot{\boldsymbol{\varepsilon}}_{\mathbf{u}}}\right) \dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{u}}} - \mathbf{I}\tilde{p}$ , and  $\boldsymbol{\tau}_{\mathbf{u}} = 2 \frac{d\eta}{dn}(\mathbf{u}, n) \tilde{n} \dot{\boldsymbol{\varepsilon}}_{\mathbf{u}}$ , and the *incremental adjoint velocity/pressure*  $(\tilde{\mathbf{v}}, \tilde{q})$  satisfy the *incremental adjoint Stokes problem*

$$\nabla \cdot \tilde{\mathbf{v}} = 0 \quad \text{in } \Omega, \quad (13a)$$

$$-\nabla \cdot \boldsymbol{\sigma}_{\tilde{\mathbf{v}}} = \mathbf{0} \quad \text{in } \Omega, \quad (13b)$$

$$\tilde{\mathbf{v}}|_{\Gamma_t} = \tilde{\mathbf{v}}|_{\Gamma_r}, \text{ and } \boldsymbol{\sigma}_{\tilde{\mathbf{v}}}\mathbf{n}|_{\Gamma_t} = \boldsymbol{\sigma}_{\tilde{\mathbf{v}}}\mathbf{n}|_{\Gamma_r} \quad \text{on } \Gamma_p, \quad (13c)$$

$$\boldsymbol{\sigma}_{\tilde{\mathbf{v}}}\mathbf{n} = -\tilde{\mathbf{u}} \quad \text{on } \Gamma_t, \quad (13d)$$

$$\tilde{\mathbf{v}} \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_b, \quad (13e)$$

$$\begin{aligned} & \mathbf{T}\boldsymbol{\sigma}_{\tilde{\mathbf{v}}}\mathbf{n} + \beta|\mathbf{T}\mathbf{u}|^{m-1}\mathbf{T}\tilde{\mathbf{v}} \\ & + \beta(m-1)|\mathbf{T}\mathbf{u}|^{m-3}(\mathbf{T}\mathbf{u} \otimes \mathbf{T}\mathbf{u})\mathbf{T}\tilde{\mathbf{v}} = \mathbf{0} \quad \text{on } \Gamma_b, \end{aligned} \quad (13f)$$

with  $\boldsymbol{\sigma}_{\tilde{\mathbf{v}}} = 2\eta(\mathbf{u}, n) \left(1 + \frac{1-n}{n} \frac{\dot{\boldsymbol{\varepsilon}}_{\mathbf{u}} \otimes \dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{v}}}}{\dot{\boldsymbol{\varepsilon}}_{\mathbf{u}} : \dot{\boldsymbol{\varepsilon}}_{\mathbf{u}}}\right) \dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{v}}} - \mathbf{I}\tilde{q}$ . In these expressions,  $\dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{u}}}$  and  $\dot{\boldsymbol{\varepsilon}}_{\tilde{\mathbf{v}}}$  are defined in the same manner as  $\dot{\boldsymbol{\varepsilon}}_{\mathbf{u}}$  and  $\dot{\boldsymbol{\varepsilon}}_{\mathbf{v}}$ . Note that the only source term in the incremental adjoint problem (Eqn (13)) is provided by the incremental forward velocity on the right-hand side of Eqn (13d), which results from a linearization of the residual of the adjoint Stokes problem (Eqn (10)). The only source terms in the incremental forward Stokes problem are given, in turn, by the forward velocity, which again stems from the linearization of the residual of the forward problem.

In summary, the ‘workflow’ for computing the application of the Hessian to a vector,  $(\tilde{\beta}, \tilde{n})$ , proceeds as follows. First, in evaluating the gradient for each Newton iteration, we have already solved the forward nonlinear Stokes problem (Eqn (4)) to obtain the forward velocity/pressure,  $(\mathbf{u}, p)$ , and the adjoint Stokes problem (Eqn (10)) to obtain the adjoint velocity/pressure,  $(\mathbf{v}, q)$ . Next, the right-hand side of the incremental forward Stokes problem (Eqn (12)) is constructed from the forward velocity, and this Stokes problem is solved to obtain the incremental forward velocity/pressure,  $(\tilde{\mathbf{u}}, \tilde{p})$ . Then, the incremental forward velocity forms the right-hand side of the incremental adjoint Stokes problem (Eqn (13)), which is subsequently solved to yield the incremental adjoint velocity/pressure,  $(\tilde{\mathbf{v}}, \tilde{q})$ . Finally, the incremental adjoint velocity (along with the forward velocity) is used to evaluate Eqn (11), the application of the Gauss–Newton Hessian to  $(\tilde{\beta}, \tilde{n})$ . Algorithm 2 summarizes the role of incremental Stokes solutions in the CG iteration.

There are several observations to make about these expressions. First, the Hessian action (Eqn (11)) has the same form as the gradient (Eqn (8)). Second, the incremental forward and adjoint Stokes problems (Eqns (12)) and ((13)) have the same operator as the adjoint Stokes problem (Eqn (10)), and differ only in the domain and boundary source terms. Like the adjoint Stokes problem, both incremental

Initialize/define variables  $\mathbf{m}_1 = (\beta_1, n_1), \alpha, \mu, c, \varepsilon_{\text{tol}}$   
**for**  $k = 1, \dots$  **do**  
 ( $\mathbf{u}_k, p_k$ )  $\leftarrow$  solve the forward equation with  $\mathbf{m}_k$   
 ( $\mathbf{v}_k, q_k$ )  $\leftarrow$  solve the adjoint equation with  $\mathbf{m}_k$  and ( $\mathbf{u}_k, p_k$ )  
 $\mathbf{g}_k \leftarrow$  compute the gradient given by the discretization of (8)  
 Perform inexact CG iterations (Algorithm 2) to compute  $\tilde{\mathbf{m}}_k$   
 $\alpha \leftarrow 1$   
**while** no descent **do**  
 $\mathbf{m}_{k+1} \leftarrow \mathbf{m}_k + \alpha \tilde{\mathbf{m}}_k$   
 Solve the forward equation with  $\mathbf{m}_{k+1}$   
**if**  $\mathcal{J}(\mathbf{m}_{k+1}) \leq \mathcal{J}(\mathbf{m}_k) + c\alpha \mathbf{g}_k^T \tilde{\mathbf{m}}_k$  **then**  
 $\mathcal{J}(\mathbf{m}_k) \leftarrow \mathcal{J}(\mathbf{m}_{k+1})$   
**else**  
 $\alpha \leftarrow \mu\alpha$   
**end if**  
**end while**  
**if**  $\|\mathbf{g}_k\| < \varepsilon_{\text{tol}}$  **then**  
 converged  
**end if**  
**end for**

**Algorithm 1.** Adjoint-based inexact Newton method

systems are linearized Stokes equations with anisotropic effective viscosity tensors that depend on the solution of the forward Stokes problem. Since the operator for the adjoint Stokes problem (Eqn (10)) is identical to the Jacobian of the nonlinear forward Stokes problem (Eqn (4)), the tools needed to solve the incremental systems are already available in any Newton-based forward Stokes solver, as they are for the gradient.

Each CG iteration entails forming the Hessian action (Eqn (11)) (which, when discretized, leads to a Hessian matrix-vector product), which, in turn, requires solution of the incremental forward and adjoint systems (Eqns (12) and (13)). If a direct linear solver based on a matrix factorization is feasible, the cost of the Stokes matrix factorization can be amortized across the incremental system solutions in all CG iterations needed for each Newton iteration; in this case, only triangular solutions are required at each CG iteration. If a direct solver is not feasible, then at least the preconditioner can be reused for all of the incremental solutions during the CG iterations. In any case, it is critical to reduce the number of CG iterations as far as possible.

Moreover, since accurate solution of the Newton system (Eqn (7)) is needed only close to the minimum of the regularized data misfit functional,  $\mathcal{J}$ , to retain superlinear convergence (Nocedal and Wright, 2006), we terminate the CG iterations early for iterates that are far from the converged solution. This significantly reduces the number of required linearized forward/adjoint solutions. One form of this *inexact Newton* method terminates the CG iterations when the norm of the residual of the linear system (Eqn (7)) drops below a tolerance that is proportional to the norm of the gradient (Algorithm 2). Far from the minimum – when the relative gradient is large – the tolerance is also large, and the CG iterations are terminated early. As the minimum is approached, the norm of the gradient decreases, thereby enforcing an increasingly more accurate solution of the Newton system (Eqn (7)). The criterion above is able to significantly reduce the number of CG iterations (and thus the required number of linearized forward/adjoint Stokes solutions), while still maintaining superlinear convergence. Compared to a more accurate computation of the Newton direction, this inexactness can result in a slightly larger

$\tilde{\mathbf{m}}_k^0 \leftarrow 0, i \leftarrow 0$   
 $\mathbf{r}_k^0 \leftarrow \mathbf{g}_k$   
 $\mathbf{d}_k^0 \leftarrow -\mathbf{r}_k^0$   
 $\nu \leftarrow \min\left(0.5, \sqrt{\frac{\|\mathbf{g}_k\|}{\|\mathbf{g}_0\|}}\right)$   
**while**  $\|\mathbf{r}_k^i\| > \nu \|\mathbf{g}_k\|$  **do**  
 ( $\tilde{\mathbf{u}}_k, \tilde{p}_k$ )  $\leftarrow$  solve incremental forward equation with  $\mathbf{d}_k^i$   
 ( $\tilde{\mathbf{v}}_k, \tilde{q}_k$ )  $\leftarrow$  solve incremental adjoint equation with ( $\tilde{\mathbf{u}}_k, \tilde{p}_k$ )  
 $\mathbf{H}_k \mathbf{d}_k^i \leftarrow$  evaluate the Hessian-vector product as given by the discretization of (11)  
 $\alpha^i \leftarrow \frac{\mathbf{r}_k^{iT} \mathbf{r}_k^i}{\mathbf{d}_k^{iT} \mathbf{H}_k \mathbf{d}_k^i}$   
 $\tilde{\mathbf{m}}_k^{i+1} \leftarrow \tilde{\mathbf{m}}_k^i + \alpha^i \mathbf{d}_k^i$   
 $\mathbf{r}_k^{i+1} \leftarrow \mathbf{r}_k^i + \alpha^i \mathbf{H}_k \mathbf{d}_k^i$   
 $\beta^{i+1} \leftarrow \frac{\mathbf{r}_k^{i+1T} \mathbf{r}_k^{i+1}}{\mathbf{r}_k^{iT} \mathbf{r}_k^i}$   
 $\mathbf{d}_k^{i+1} \leftarrow -\mathbf{r}_k^{i+1} + \beta^{i+1} \mathbf{d}_k^i$   
 $i \leftarrow i + 1$   
**end while**

**Algorithm 2.** CG algorithm for solving  $\mathbf{H}_k \tilde{\mathbf{m}}_k = -\mathbf{g}_k$

number of Newton iterations, but significantly reduces the overall number of CG iterations, and thus the overall number of Stokes-like solutions.

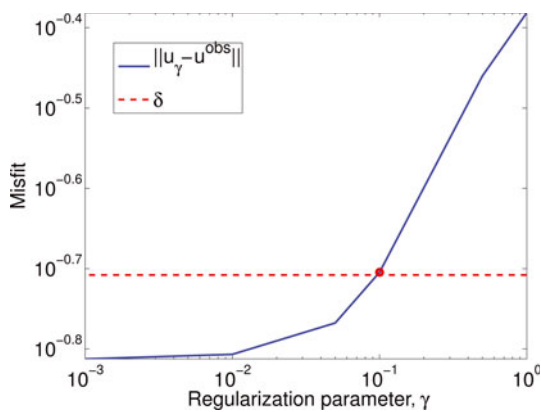
Once the Newton direction is computed by inexact solution of Eqn (7), we must guarantee that sufficient decrease in  $\mathcal{J}$  is obtained in that direction, so convergence of the iterations can be assured. This is achieved by a *line search* that finds a step size,  $\alpha$ , satisfying the so-called Armijo condition (Nocedal and Wright, 2006), which has the attractive property that it requires only cost function evaluations (entailing a forward Stokes solution), and not gradient information. The Newton iterations are repeated until the norm of the gradient of  $\mathcal{J}$  is sufficiently small. The inexact Newton method is summarized in Algorithms 1 and 2.

Typical values for the line search parameters in Algorithm 1 are  $\mu = 0.5$  and  $c = 10^{-4}$ . Note that the CG iteration in Algorithm 2 is initialized with  $\tilde{\mathbf{m}}_k^0 = 0$ , and thus the first CG iterate,  $\tilde{\mathbf{m}}_k^1$ , is the negative gradient direction, scaled by  $\alpha^0$ .

Finally we comment on the choice of the regularization parameters,  $\gamma_\beta$  and  $\gamma_n$ , which depend on (an estimate of) the observation error, also called the noise level. This error can, for instance, be due to measurement error or model uncertainties (Tarantola, 2005). We use the Morozov discrepancy criterion (Vogel, 2002), i.e. we find regularization parameters such that  $\|\mathbf{u} - \mathbf{u}^{\text{obs}}\| \approx \delta$ , where  $\delta$  is the noise level and  $\mathbf{u} = \mathbf{u}_\gamma$  is the solution computed with the regularization parameters (Fig. 1). Generally, the quality of the reconstruction is not very sensitive to the choice of the regularization parameters. However, choosing parameters that are too large will lead to an overly smoothed reconstruction, while a regularization parameter that is too small for the error in the observations results in instability in the inverse problem, manifesting as noise in the inverse solution.

#### 4. RESULTS FOR MODEL TEST PROBLEMS

In this section, we describe numerical test problems to study the performance of the optimization algorithm and the ability of the inversion procedure to recover the unknown parameter fields. Tests are performed for inversion of the sliding coefficient,  $\beta$ , in problems with linear and nonlinear



**Fig. 1.** Illustration of the detection of an optimal regularization parameter,  $\gamma$  (which can either be  $\gamma_\beta$  or  $\gamma_n$ ), using Morozov’s discrepancy principle. The dashed horizontal line depicts the noise level,  $\delta$ , and the solid curve shows the misfits  $\|\mathbf{u}_\gamma - \mathbf{u}^{obs}\|$  for different values of  $\gamma$ . The value of  $\gamma$  corresponding to the misfit marked by the dot is a near-optimal regularization parameter according to Morozov’s discrepancy principle. This criterion for choosing the regularization parameter requires the solution of several inverse problems with different values of  $\gamma$ . The plot shows the parameter selection for model 1 described in Section 4.1.

rheology, and a linear sliding law. Then we show inversion results for Glen’s flow law exponent parameter,  $n$ .

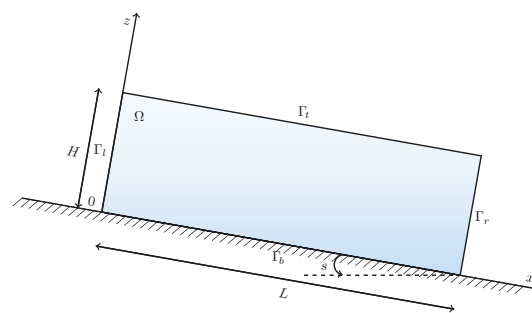
### 4.1. Model problems

We begin by presenting five model problems we use to assess performance and efficacy of the proposed inversion method. The forward problems are based on the Ice Sheet Model Intercomparison Project for Higher-Order Models (ISMIP-HOM) benchmark study by Pattyn and others (2008). For all test problems, we consider a 3-D hexahedral ice slab of thickness  $H = 1$  km on an inclined plane with slope  $s = 0.1^\circ$  (Fig. 2). The driving force in the Stokes equations (Eqn (4)) is the gravity,  $\rho\mathbf{g} = (\rho g \sin \theta, 0, -\rho g \cos \theta)$ , where  $\rho = 910 \text{ kg m}^{-3}$  is the ice density,  $g = 9.81 \text{ m s}^{-2}$  is the gravitational constant, and  $\theta = 0.1\pi/180$  is the slope in radians. The basal sliding coefficient field is defined as

$$\beta(x, y) = 1000 + 1000 \sin(\omega x) \sin(\omega y), \quad (14)$$

where  $(x, y) \in [0, L] \times [0, L]$  with  $L$  the length of the ice sheet, taken to be 5 km unless otherwise specified, and  $\omega = 2\pi/L$ . Thus, the wavelength of the basal variation is  $L$ . Unless otherwise specified, we use a linear basal sliding law (i.e.  $m = 1$  in Eqn (4f)) with the sliding coefficient given by Eqn (14), periodic boundary conditions for the lateral boundaries and zero traction on the top surface. The sliding and rheology laws considered here are given by Eqns (2) and (3). For linear rheology (i.e.  $n = 1$ ), the ice flow parameter,  $A$ , is  $2.140373 \times 10^{-7} \text{ Pa}^{-1} \text{ a}^{-1}$ , and for nonlinear rheology  $A = 10^{-16} \text{ Pa}^{-n} \text{ a}^{-1}$  (Pattyn and others, 2008).

For all numerical experiments, surface velocities extracted from forward solution fields are used as synthetic observations in the inverse problem. Random Gaussian noise is added to these observational data to lessen the ‘inverse crime’, which occurs when the same numerical method is used to both synthesize the data and drive the inverse solution (e.g. Kaipio and Somersalo, 2005). We choose a regularization parameter that approximately



**Fig. 2.** Coordinate system and cross section through a 3-D slab of ice, as used in the computational experiments.

satisfies Morozov’s discrepancy principle. We discretize the domain,  $\Omega$ , into hexahedra, and, for the forward and adjoint Stokes problems as well as their incremental counterparts, we employ Taylor–Hood finite elements to approximate the velocity–pressure pairs, i.e. trilinear elements for pressure and triquadratic elements for velocity components. For the unknown inversion fields,  $\beta$  and  $n$ , bi- and trilinear elements, respectively, are used. All Stokes-like systems (which share a coefficient matrix given by the Jacobian of the forward problem) are solved by MATLAB’s backslash operator, i.e. using a direct factorization.

The performance of the inexact Newton method presented in Section 3 is compared with two commonly used (e.g. Morlighem and others, 2010; Goldberg and Sergienko, 2011) versions of the NCG method, namely the Fletcher–Reeves and Polak–Ribière nonlinear CG methods with a line search such that the step length satisfies the strong Wolfe condition (Nocedal and Wright, 2006, p. 60). To speed up the convergence of the Fletcher–Reeves algorithm, a restart is performed whenever two consecutive gradients are far from orthogonal, as measured by the test  $|\mathbf{g}_k^T \mathbf{g}_{k-1}| / \|\mathbf{g}_k\|^2 \geq \nu$ , where  $\mathbf{g}_{k-1}$  and  $\mathbf{g}_k$  are the (discrete) gradients at iterations  $k - 1$  and  $k$ , respectively. A typical value for the parameter  $\nu$  is 0.1 (Nocedal and Wright, 2006, p. 125).

We now describe the five model problems used below. The general set-up of these problems is as previously described. Below we focus only on specifying the ‘true’ inversion fields and other specific problem features.

1. The first model problem considers inversion for the basal sliding coefficient,  $\beta$ , with linear rheology as given by Eqn (2), with  $n = 1$ , and linear sliding law, given by Eqn (3) with  $m = 1$ . The true  $\beta$  field is given by Eqn (14). This model is used for assessing the performance of the proposed inexact Newton method.
2. This problem is as model 1, but with  $L = 10$  km. Additionally, we add a small-wavelength variation to the sliding coefficient, i.e. we replace Eqn (14) by

$$\beta_1(x, y) = \beta(x, y) + 100 + 100 \sin(8\omega x) \sin(8\omega y).$$

3. This is the same as model 1, but in addition to the linear rheology we also consider a nonlinear rheology law ( $n = 3$ ) and various sizes of the domain, namely  $L = 5, 10, 20$  and 40 km. By choosing these different horizontal dimensions, we are able to assess the ability to invert for several different wavenumber variations in the true  $\beta$  field from surface observations.

**Table 1.** Number of iterations (#iter) and the number of Stokes factorizations (#Stokes) for the NCG and inexact Newton methods for an inverse problem with nonlinear rheology. The first column (Mesh) shows the number of elements used to discretize the variables in the two horizontal and the vertical directions; the second column (#dof) indicates the total number of velocity and pressure variables and the third column (#par) indicates the number of inversion ( $\beta$ ) parameters. The fourth and fifth columns report the number of NCG iterations and the number of Stokes factorizations for the Fletcher–Reeves (FR) and Polak–Ribière (PR) variants of the NCG method. The sixth column shows the number of Newton iterations, and in parentheses the overall number of CG iterations. The last column reports the number of factorizations needed by the inexact Newton method. The iterations are terminated when the norm of the gradient is decreased by a factor of  $10^5$ . This table shows that the cost of solving the inverse problem by the inexact Newton method measured as number of Stokes factorizations is roughly independent of the number of inversion parameters. This is not the case for the NCG method, i.e. as we increase the mesh resolution, the numbers of iterations and Stokes factorizations increase significantly. In addition, the comparison shows that the inexact Newton method is  $\sim 50$  times faster (measured in the number of Stokes factorizations) than the commonly used NCG method

Mesh	#dof	#par	#iter	NCG (FR/PR)		Inexact Newton	
				#Stokes	#iter	#Stokes	
$10 \times 10 \times 2$	6978	121	75/56	1933/1592	9 (34)	50	
$20 \times 20 \times 2$	26 538	141	83/65	2268/2373	9 (33)	57	
$40 \times 40 \times 2$	103 458	1681	137/103	2841/2841	9 (32)	57	
$80 \times 80 \times 2$	408 498	6561	164/128	3759/2812	9 (33)	64	

4. This model considers inversion for a smoothly varying Glen's flow law exponent parameter,  $n$ , in the nonlinear Stokes equations (Eqn (4)). The basic geometry and model set-up is as previously described, except that the boundary conditions on both lateral x-z-plane boundaries are set to no-slip conditions. The purpose of this model problem is to study whether variations in the volume viscosity field can be reconstructed from surface velocity observations. For this purpose, we propose the following true  $n$  field:

$$n(x, y, z) = 3 + \frac{1}{3} [\sin(\omega_1 x) + \sin(\omega_1 y) + \sin(\omega_2 z)], \quad (15)$$

where  $\omega_1 = 2\pi/L$  and  $\omega_2 = \pi H/4$ .

5. The geometry and problem set-up for this model are the same as for model 4, but we aim to reconstruct the  $n$  field:

$$n(x, y, z) = 3 + 2 \max(n_1, n_2), \quad (16)$$

where

$$n_1(x, y, z) = \exp\left[-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}\right],$$

$$n_2(x, y, z) = \begin{cases} \exp\left[-\frac{(y-y_0)^2}{2\sigma^2}\right], & x < x_0, \\ 0, & \text{else.} \end{cases}$$

Here we choose the mean and the standard deviation of the above Gaussians as  $x_0 = L/2$ ,  $y_0 = 3L/4$  and  $\sigma = 10\sqrt{L/5}$ . The goal of this problem is to study the ability of the inversion method to reconstruct strong local anomalies in the rheology. Such anomalies could indicate that Glen's flow law cannot be used to describe the mechanical properties of the ice in that region sufficiently well.

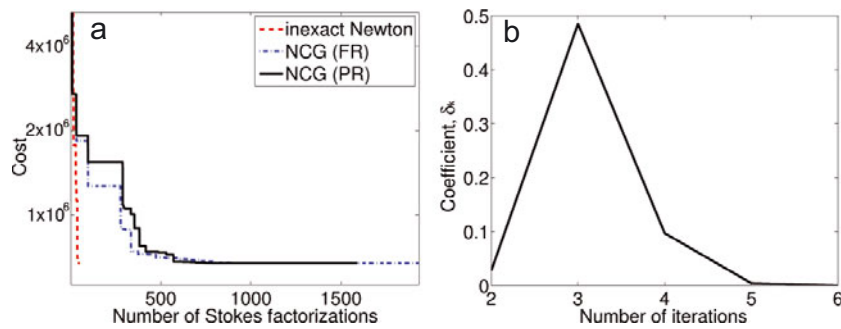
#### 4.2. Performance of the inexact Newton method

Now we consider the behavior of the inexact Newton method described in Section 3 and compare its performance with the NCG method, using model 1 but with a nonlinear rheology ( $n = 3$ ) as a test problem. To motivate our performance metric, note that the solution of linear or linearized Stokes systems dominates the computational cost for both inversion methods since, as explained in Section 3,

this is at the heart of cost function, gradient and Hessian vector product evaluation. Here, we use a direct method based on a matrix factorization to solve these Stokes systems. Relative to computing a factorization, the cost of the subsequent triangular substitutions is negligible, especially for large problems. Thus, we characterize the cost of the two inversion methods by the number of matrix factorizations needed to achieve convergence for a given tolerance. This allows conclusions about the performance of the methods to be independent of the specific direct solver employed. Note that the number of factorizations needed by both the inexact Newton and the NCG methods is equal to the number of distinct linear systems that must be factored (and thus, each CG iteration within the inexact Newton iteration does not require a factorization).

Table 1 compares the number of iterations and Stokes factorizations required for convergence of the NCG and inexact Newton methods for an inverse problem with nonlinear rheology. Since realistic inverse ice sheet problems are very high-dimensional, we are particularly interested in the behavior of these optimization methods as the size of the inverse problem grows. Thus, we use a sequence of increasingly finer meshes to assess the performance. As can be seen, for all four meshes the number of iterations needed by the inexact Newton method is significantly smaller than the number of iterations needed by the NCG method. For example, on a mesh of  $40 \times 40 \times 2$  elements, the Newton method takes 9 iterations with 32 CG iterations overall (i.e.  $\sim 4$  CG iterations per Newton step), whereas the NCG method takes  $>100$  iterations. When comparing the number of factorizations, the difference between the two methods is even more significant, i.e. the NCG method requires  $\sim 50$  times more Stokes factorizations than does the inexact Newton method. One reason for this is that CG (i.e. inner) iterations of Newton do not incur any additional factorizations, since the linear systems that are being solved have the same coefficient matrix. Moreover, even though computing a descent direction for the NCG method is theoretically cheaper (since it requires only gradient information), the difficulty of finding an appropriate step length that satisfies the Wolfe condition amounts to a large number of linearized forward solutions (i.e. number of factorizations). Finally, the results show that the number





**Fig. 3.** (a) The cost function value versus the number of Stokes factorizations for the inexact Newton (red) and the Fletcher–Reeves (FR) (blue) and Polak–Ribière (PR) (black) variants of the NCG method. The basal sliding coefficient,  $\beta$ , was reconstructed with nonlinear rheology, and for SNR = 500 and  $L = 10$ . This plot corresponds to the coarsest mesh ( $10 \times 10 \times 2$ ) case in Table 1. The inexact Newton method is seen to be significantly more efficient than the NCG methods. (b) The coefficient  $\delta_k = \|\beta_{k+1} - \beta^*\|_{L^2} / \|\beta_k - \beta^*\|_{L^2}$  (with  $\beta_k$  denoting the  $k$ th iterate and  $\beta^*$  the inverse solution) plotted against the iteration number. Since  $\delta_k \rightarrow 0$ , the method converges at a superlinear rate.

of Newton iterations is insensitive to the inverse problem size, in accordance with the theory (Heinkenschloss, 1993). This is not the case with the NCG method, which for the nonlinear rheology case shows a significant increase in the number of iterations and Stokes factorizations with problem size. The ability of Newton methods to scale independently of inverse problem size is critical to the prospects for solving large-scale ice sheet inverse problems, as will be encountered in full-continental inversions for basal sliding and rheology coefficients.

To visualize the differences in performance between the inexact Newton and NCG methods, in Figure 3 we plot (left) the cost function value vs the number of Stokes factorizations for the inexact Newton and for the Fletcher–Reeves (FR) and Polak–Ribière (PR) variants of the NCG method. These results are for model 1 with nonlinear rheology ( $n = 3$ ). This figure further illustrates the significant improvement in efficiency of the inexact Newton method over the NCG methods. In Figure 3b, we plot the convergence coefficient of the inexact Newton method, defined by  $\delta_k = \|\beta_{k+1} - \beta^*\|_{L^2} / \|\beta_k - \beta^*\|_{L^2}$ , where  $\beta_k$  is the  $k$ th iterate, and  $\beta^*$  is the inverse solution. This result shows that  $\lim_{k \rightarrow \infty} \delta_k = 0$ , and thus the convergence rate is superlinear (Kelley, 1999; Nocedal and Wright, 2006).

### 4.3. Inversion results

#### *Inversion for the basal sliding coefficient, $\beta$*

Here we study how well variations in the basal sliding coefficient field,  $\beta$ , can be reconstructed from synthetic observations. We use a linear sliding law ( $m = 1$ ) and both linear ( $n = 1$ ) and nonlinear ( $n = 3$ ) rheology. The goal is to study the limits of invertibility for the sliding coefficient,  $\beta$ , as a function of wavelength of the basal variation and the noise level in the synthetic velocity observations. Related studies based on approximate derivatives found from analytical transfer functions are given by Raymond and Gudmundsson (2005), Raymond (2007), Gudmundsson and Raymond (2008) and Raymond and Gudmundsson (2009). To specify the noise level, we use the signal-to-noise ratio (SNR), which we define as the ratio between the average surface velocity and the standard deviation of the added noise,  $\sigma_{\text{noise}}$ , i.e.

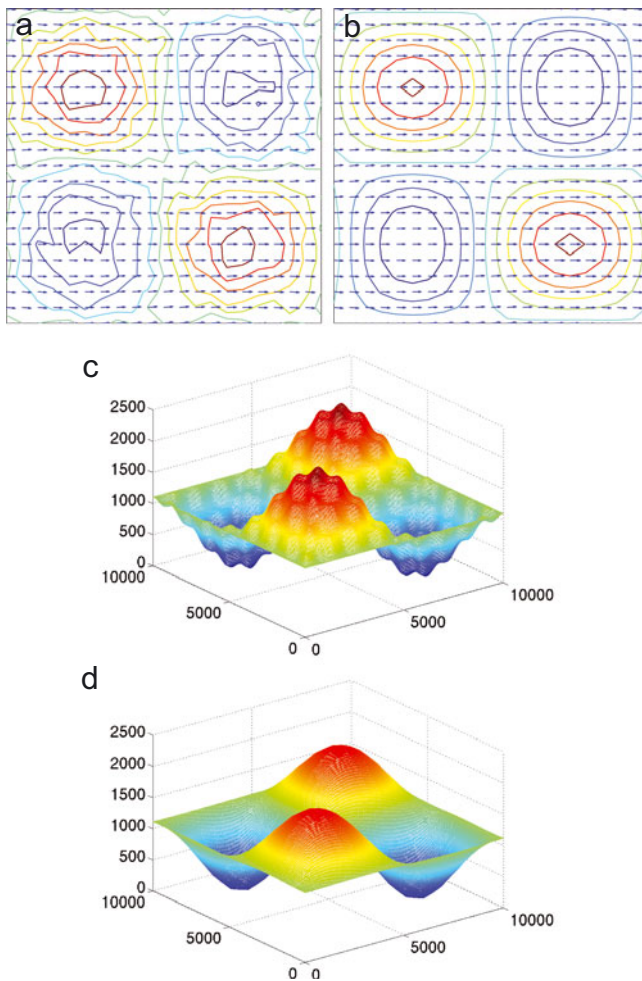
$$\text{SNR} = \frac{\bar{u}}{\sigma_{\text{noise}}}, \tag{17}$$

where  $\bar{u}$  is the average surface velocity based on some  $L^p$  norm (we use  $\bar{u} = \|u\|_{L^2(\Gamma_t)} / |\Gamma_t|^{1/2}$  for models 1–3, and  $\bar{u} = \|u\|_{L^\infty(\Gamma_t)}$  for models 4 and 5).

We start with the (linear) model 2, where the true  $\beta$  coefficient field is given by a sum of small and large wavelength variations (Fig. 4,c). Due to the presence of noise in the data (SNR = 100), the basal sliding coefficient,  $\beta$ , can be reconstructed only approximately (Fig. 4d). While the large-wavelength component is well recovered, the small-wavelength variations cannot be reconstructed by the inversion method. This can be explained by the fact that the surface velocity response to these small-wavelength variations in  $\beta$  is small due to the smoothing property of the parameter-to-observable map; effectively, the solution of the Stokes problem acts as a low-pass filter. These small observed velocities are overwhelmed by the noise in the data, making the reconstruction of the small-wavelength component impossible.

We continue with a more systematic study of the interplay between length scales in the basal variation, the SNR, the nonlinearity of the rheology and the quality of the reconstruction, using the problem setting for model 3. First, we solve the Stokes equations for the true basal sliding coefficient for different wavelengths,  $L$ , of the sliding coefficient. Then we add noise with a given SNR to the resulting surface velocities and use these synthetic observations to reconstruct the basal sliding coefficient. Results for different  $\beta$  noise levels and for linear and nonlinear rheology are summarized in Table 2. The regularization parameters are computed according to the Morozov discrepancy principle (Table 2a). The relative errors between the true and the reconstructed  $\beta$  coefficient field, i.e.  $\|\beta_{\text{true}} - \beta\|_{L^2} / \|\beta_{\text{true}}\|_{L^2}$ , are reported in Table 2b. Besides the expected result that more noise makes the reconstruction of the true sliding coefficient more difficult, we make the following observations:

1. As the noise level decreases, for both linear and nonlinear rheology, the basal sliding coefficient,  $\beta$ , reconstructed from the noisy surface observations converges to the true  $\beta$  (i.e. the one that was used for the synthetic observations).
2. The larger the wavelength,  $L$ , of the variation of the basal sliding coefficient, the better it can be reconstructed. This result is in agreement with findings reported by



**Fig. 4.** Inversion for a rough basal sliding coefficient,  $\beta$ , with a linear rheology law and a domain of  $10 \text{ km} \times 10 \text{ km} \times 1 \text{ km}$ . (a, b) Observations of (a) surface velocity with  $\text{SNR} = 100$ ; (b) surface velocity based on reconstructed basal sliding coefficient. (c) True basal sliding coefficient; (d) inverted  $\beta$  field based on noisy observations. The regularization parameter,  $\gamma$ , is chosen according to Morozov's discrepancy principle. Even though the observations are fitted to within the noise (a, b), only the smooth components of  $\beta$  can be reconstructed (d).

Gudmundsson and Raymond (2008), which are based on analytical transfer functions along flowlines and, thus, describe the effects of small-amplitude basal variations on the surface velocity. Gudmundsson and Raymond (2008) find that for 1% noise (i.e. for  $\text{SNR} = 100$ ), the variation in a basal sliding coefficient with a mean value of 500 can be reconstructed accurately only if its wavelength is  $\sim 50$  times the ice thickness, which is equivalent to  $L = 50$  in our set-up. Our finite-amplitude results (Table 2) suggest that, for a mean sliding coefficient of 1000, a reasonable reconstruction can be achieved already for a smaller wavelength-to-ice-thickness ratio of  $\sim 10$  for linear rheology and 20 for nonlinear rheology.

3. The reconstruction of  $\beta$  is significantly more difficult for nonlinear than for linear rheology. To illustrate this, in Figure 5 we show the basal sliding coefficient,  $\beta$ , reconstructed from synthetic data with  $\text{SNR} = 100$  for both linear and nonlinear rheology. For a smaller wavelength in the  $\beta$  field (of  $L = 10 \text{ km}$ ),  $\beta$  can be identified well using a linear rheology (Fig. 5b). However,

**Table 2.** (a) The optimal regularization parameter computed from the discrepancy principle. ('-' indicates cases for which no finite regularization parameter exists that satisfies the discrepancy principle, since the noise level is larger than the variance of the surface velocity. In these cases, the noise dominates the data and the sliding coefficient cannot be reconstructed from the data.) (b) The error with linear and nonlinear rheology for signal-to-noise ratios  $\text{SNR} = 500, 100, 20$  and  $10$

L	1. Linear rheology				2. Nonlinear rheology			
	SNR				SNR			
	500	100	20	10	500	100	20	10
(a) The optimal regularization parameter, $\gamma$								
5	0.015	0.1	-	-	-	-	-	-
10	0.2	1.2	8	35	0.03	0.15	-	-
20	1.5	10	70	200	0.8	5	25	55
40	8	60	550	1600	6	40	300	800
(b) The relative error, $\frac{\ \beta - \beta_{\text{true}}\ _2}{\ \beta_{\text{true}}\ _2}$								
5	0.031	0.112	-	-	-	-	-	-
10	0.022	0.049	0.157	0.310	0.032	0.165	-	-
20	0.017	0.041	0.109	0.177	0.026	0.045	0.118	0.285
40	0.013	0.039	0.125	0.191	0.015	0.036	0.097	0.149

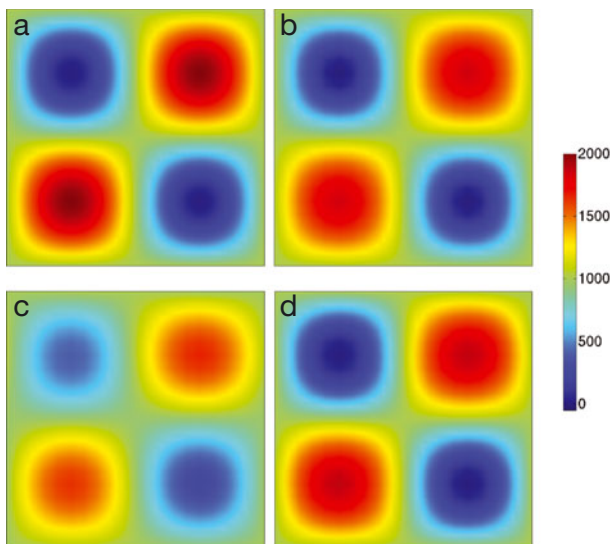
nonlinear rheology for the same  $L$  weakens the ability to reconstruct  $\beta$  (Fig. 5c). However, the larger wavelength in the  $\beta$  field (of  $L = 40 \text{ km}$ ) can be reconstructed well even for nonlinear rheology (Fig. 5d).

4. To further interpret the findings summarized in points 2 and 3, Figure 6 shows sections through the surface flow velocity field for small- ( $L = 10$ ) and large- ( $L = 40$ ) wavelength  $\beta$  fields, and for linear ( $n = 1$ ) and nonlinear ( $n = 3$ ) rheology. Note that the deviation from a constant flow at the top surface grows with increasing  $L$ , and decreases with the degree of nonlinearity ( $n$ ) of the rheology. This departure from a uniform flow is what allows the reconstruction of a spatially varying basal sliding coefficient, through the solution of an inverse problem; larger deviations from a constant are not as easily overwhelmed by noise. This explains why the reconstruction of the sliding coefficient in the presence of a certain noise level is better for  $L = 40$  than for  $L = 10$  (see 2 above) and for linear than for nonlinear rheology (see 3 above).

#### Inversion for Glen's flow law exponent, $n$

Here we present inversion results for models 4 and 5. We find that the reconstruction of a spatially varying Glen's flow law exponent field,  $n$  (i.e. the rheology parameter field), from surface observations is possible, and works particularly well when  $n$  varies smoothly. For both cases the inversion was started with a spatially uniform  $n$  field, namely  $n = 3$ .

Figures 7 and 8 present results of inversion for models 4 and 5, respectively. For each of these models, the top panels show the noisy observations of the surface velocity (left) and the velocity field obtained solving the Stokes equations with the reconstructed  $n$  field (right). These images demonstrate that the inversion matches the observations to within the noise. Figures 7c and 8c show the true  $n$  rheology parameter fields (used to generate the synthetic observations) and their reconstructed values. Slices through the same true and



**Fig. 5.** Reconstruction of basal sliding coefficient,  $\beta$ , from noisy synthetic observations (SNR = 100) from model 3. Shown are the true  $\beta$  field (a) and the reconstruction for  $L = 10$  with linear rheology (b) and with nonlinear rheology (c). (d) Reconstruction for a nonlinear rheology with  $L = 40$ .

reconstructed  $n$  fields for models 4 and 5 are shown in Figures 7d and 8d.

The  $n$  parameter fields, which are either smoothly varying (model 4) or highly localized (model 5), can be well reconstructed from the surface observations. The

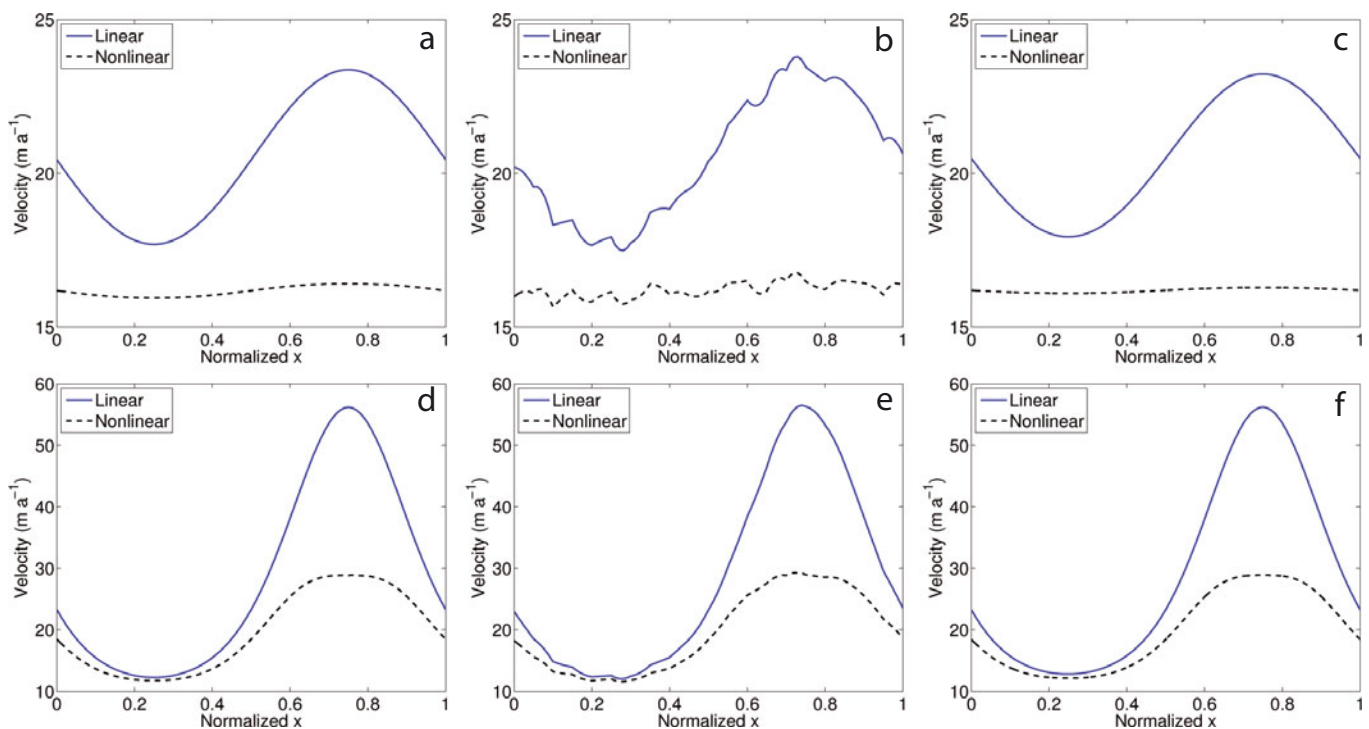
surface-to-volume ratio of observations to parameters makes the inversion of a spatially varying  $n$  parameter field an underdetermined problem and, thus, regularization is essential in the noise-free case, as well as the noisy case.

Despite this ill-posedness, the sharp horizontal variation in the  $n$  parameter field of model 5, which could originate from, for instance, spatially varying crystal fabric orientation, impurity content or fracture, is well reconstructed. Note that in the upper part of the slab the reconstructions are close to the true  $n$  fields, but they degrade in the deeper parts of the slab.

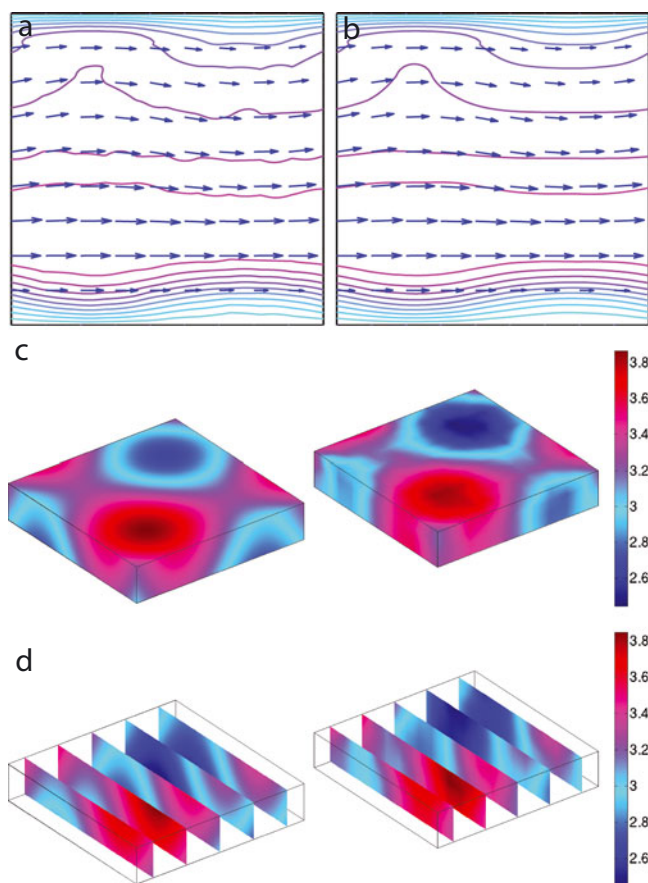
### 5. CONCLUSIONS

We have presented an adjoint-based inexact Newton method for estimating uncertain (or unknown) parameters in ice sheet models that are governed by the nonlinear full-Stokes equations. We have applied this method to the inverse problem of inferring the basal sliding coefficient and the exponent in Glen's flow law from observations of ice surface velocity. To address the ill-posedness of the inverse problem, a Tikhonov-type regularization is used. This penalizes oscillatory components in the unknown parameter fields, which cannot be reconstructed reliably in the inversion.

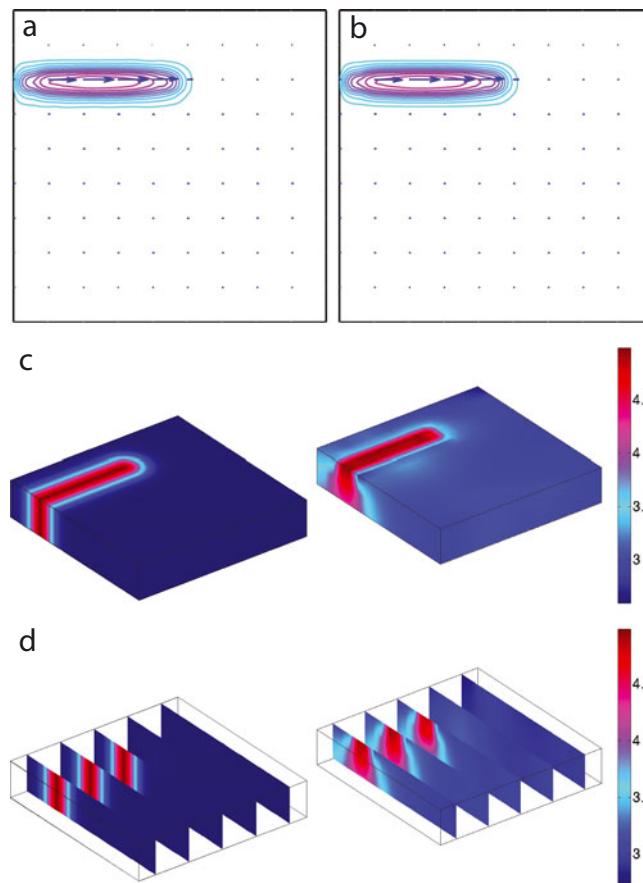
We have studied the performance of the proposed method in comparison with the NCG method, which has been applied previously to glaciology inverse problems (Vieli and Payne, 2003; Morlighem and others, 2010; Goldberg and Sergienko, 2011). We conclude that the Newton method is significantly (of the order of six times) more efficient than the



**Fig. 6.** Surface velocity response for Stokes flow problem described in model 3 for linear and nonlinear rheology. The plots show  $x$ -component velocities at  $y = L/4$  for variations in  $\beta$  with wavelengths  $L = 10$  km (a–c) and  $L = 40$  km (d–f). (a, d) depict surface velocities based on the true basal sliding coefficient. (b, e) show these synthetic observations with added noise (SNR = 100). (c, f) display surface velocities based on the reconstructed  $\beta$  field. Note that the noise in the data for  $L = 40$  km (e) appears smaller than for  $L = 10$  km (b) due to the plotting range for the  $y$ -axis, which is chosen according to the velocity variation. The deviation from a constant in the surface flow velocity (i.e. the observations for the inverse problem) decreases with  $L$  and with increasing nonlinearity in the rheology, which makes the reconstruction of  $\beta$  more difficult (cf. Fig. 5d).



**Fig. 7.** Reconstruction of  $n$  exponent field in Glen's flow law in model 4. Noisy synthetic observations of the surface velocity (a) are contrasted with the velocity field corresponding to the reconstructed  $n$  field (b). (c, d) The true (c) and reconstructed (d)  $n$  parameter fields.



**Fig. 8.** Same as Figure 7, but for model 5.

NCG method and that the cost of solving the inverse problem measured in number of Stokes solutions is insensitive to the number of inversion parameters. This finding suggests that the inexact Newton method is desirable for large-scale ice sheet inverse problems. Moreover, the additional linear system solutions entailed by the inexact Newton method are characterized by linearized Stokes operators that are identical to that of the adjoint system (which, in turn, is identical to that of a Newton solver for the nonlinear forward problem); they differ only in the source terms. Thus, implementation of the inexact Newton method for the inverse problem is conceptually straightforward and builds on existing Newton-based solvers for the nonlinear Stokes forward problem.

We have formulated and solved five model problems to study the invertibility of the basal sliding coefficient,  $\beta$ , and Glen's flow law exponent parameter,  $n$ . First, we focused on inversion for  $\beta$  fields characterized by wavelengths of oscillation that vary from smooth to rough, with linear and nonlinear rheology, and for different SNRs. We found that the reconstructions converge to the true basal sliding coefficient as the noise in the synthetic observations decreases, and that nonlinear rheology makes the reconstruction of the basal sliding coefficient more difficult. We attribute this difficulty associated with nonlinear rheology to the fact that the surface velocity from the nonlinear model is less sensitive to the basal coefficient field, and hence contains less information for the inversion. For the  $n$  inversion, the goal was to study how well

a spatially (volumetrically) varying Glen's flow law exponent field can be reconstructed from surface observations. We found that the reconstruction of the (volumetric) rheological parameter – although highly underdetermined due to the surface-to-volume ratio of observations to parameters, and in spite of the fact that the target  $n$  fields induced a high degree of nonlinearity in the Stokes equations – is very good at the surface and deteriorates with depth.

To date, we have employed synthetic observations on idealized geometries to study the performance of the proposed method, as well as to probe the limits of invertibility for basal and rheological parameters. In future work we intend to apply this inversion framework to continental-scale ice flow inverse problems with field observations. A step in this direction has been made by Price and others (2011), who used estimates of the basal sliding coefficient for the Greenland ice sheet to predict its contribution to sea-level rise by the year 2100. In practice it may be necessary to impose bound constraints on inversion parameters (e.g. to keep the sliding coefficient positive, or to restrict the flow law exponent parameter to an acceptable range). Therefore, in future work we also plan to incorporate a priori knowledge about the inversion parameters by extending our inversion method to handle inequality bound constraints. To conclude, we believe the proposed adjoint-based inexact Newton method provides an efficient, scalable and robust framework for solving large-scale ice sheet inverse problems.

## ACKNOWLEDGEMENTS

We thank Steve Price for providing very helpful comments on an earlier version of the paper, and Charles Jackson, Don Blankenship and Ginny Catania for input and encouragement of this research. We also thank the anonymous reviewers for thorough reading of the manuscript, and for valuable comments and suggestions that improved the quality of the paper. This work was partially supported by US National Science Foundation (NSF) grant OPP-0941678 and US Department of Energy grants DE-SC0002710, DE-FG02-08ER25860 and DEFC02-06ER25782. N.P. also acknowledges partial funding through the ICES Postdoctoral Fellowship.

## REFERENCES

- Arthern RJ and Gudmundsson GH (2010) Initialization of ice-sheet forecasts viewed as an inverse Robin problem. *J. Glaciol.*, **56**(197), 527–533 (doi: 10.3189/002214310792447699)
- Borzi A and Schulz V (2012) *Computational optimization of systems governed by partial differential equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA
- Braess D (1997) *Finite elements: theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, Cambridge
- Dennis JE and Schnabel RB (1996) *Numerical methods for unconstrained optimization and nonlinear equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA
- Deuflhard P (2004) *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*. Springer-Verlag, Berlin
- Dukowicz JK, Price SF and Lipscomb WH (2010) Consistent approximations and boundary conditions for ice-sheet dynamics from a principle of least action. *J. Glaciol.*, **56**(197), 480–496 (doi: 10.3189/002214310792447851)
- Elman HC, Silvester DJ and Wathen AJ (2005) *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Oxford University Press, Oxford
- Engl HW, Hanke M and Neubauer A (1996) *Regularization of inverse problems*. Kluwer Academic, Dordrecht
- Evans LC (1998) *Partial differential equations*. American Mathematical Society, Providence, RI
- Glen JW (1955) The creep of polycrystalline ice. *Proc. R. Soc. London, Ser. A*, **228**(1175), 519–538 (doi: 10.1098/rspa.1955.0066)
- Gockenbach MS (2006) *Understanding and implementing the finite element method*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA
- Goldberg DN and Sergienko OV (2011) Data assimilation using a hybrid ice flow model. *Cryosphere*, **5**(2), 315–327 (doi: 10.5194/tc-5-315-2011)
- Gudmundsson GH (2003) Transmission of basal variability to a glacier surface. *J. Geophys. Res.*, **108**(B5), 2253 (doi: 10.1029/2002JB0022107)
- Gudmundsson GH and Raymond M (2008) On the limit to resolution and information on basal properties obtainable from surface data on ice streams. *Cryosphere*, **2**(2), 167–178 (doi: 10.5194/tc-2-167-2008)
- Gunzburger MD (2003) *Perspectives in flow control and optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA
- Heinkenschloss M (1993) Mesh independence for nonlinear least squares problems with norm constraints. *SIAM J. Optimiz.*, **3**(1), 81–117
- Hinze M, Pinnau R, Ulbrich M and Ulbrich S (2009) *Optimization with PDE constraints*. Springer, Heidelberg
- Hutter K (1983) *Theoretical glaciology; material science of ice and the mechanics of glaciers and ice sheets*. D Reidel, Dordrecht/Terra Scientific, Tokyo
- Ito K and Kunisch K (2008) *Lagrange multiplier approach to variational problems and applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA
- Jay-Allemand M, Gillet-Chaulet F, Gagliardini O and Nodet M (2011) Investigating changes in basal conditions of Variegated Glacier prior to and during its 1982–1983 surge. *Cryosphere*, **5**(3), 659–672 (doi: 10.5194/tc-5-659-2011)
- Joughin I, MacAyeal DR and Tulaczyk S (2004) Basal shear stress of the Ross ice streams from control method inversions. *J. Geophys. Res.*, **109**(B9), B09405 (doi: 10.1029/2003JB002960)
- Kaipio JP and Somersalo E (2005) *Statistical and computational inverse problems*. Springer-Verlag, New York
- Kelley CT (1999) *Iterative methods for optimization*. Society for Industrial Mathematics (SIAM), Philadelphia, PA
- Larour E, Rignot E, Joughin I and Aubry D (2005) Rheology of the Ronne Ice Shelf, Antarctica, inferred from satellite radar interferometry data using an inverse control method. *Geophys. Res. Lett.*, **32**(5), L05503 (doi: 10.1029/2004GL021693)
- MacAyeal DR (1993) A tutorial on the use of control methods in ice-sheet modeling. *J. Glaciol.*, **39**(131), 91–98
- Marshall SJ (2005) Recent advances in understanding ice sheet dynamics. *Earth Planet. Sci. Lett.*, **240**(2), 191–204 (doi: 10.1016/j.epsl.2005.08.016)
- Maxwell D, Truffer M, Avdonin S and Stuefer M (2008) An iterative scheme for determining glacier velocities and stresses. *J. Glaciol.*, **54**(188), 888–898 (doi: 10.3189/002214308787779889)
- Morlighem M, Rignot E, Seroussi H, Larour E, Ben Dhia H and Aubry D (2010) Spatial patterns of basal drag inferred using control methods from a full-Stokes and simpler models for Pine Island Glacier, West Antarctica. *Geophys. Res. Lett.*, **37**(14), L14502 (doi: 10.1029/2010GL043853)
- Nocedal J and Wright SJ (2006) *Numerical optimization*, 2nd edn. Springer, New York
- Oden JT and Reddy JN (1976) *An introduction to the mathematical theory of finite elements*. Wiley, New York
- Paterson WSB (1994) *The physics of glaciers*, 3rd edn. Elsevier, Oxford
- Pattyn F and 20 others (2008) Benchmark experiments for higher-order and full-Stokes ice sheet models (ISMIP-HOM). *Cryosphere*, **2**(2), 95–108 (doi: 10.5194/tc-2-95-2008)
- Petra N and Stadler G (2011) *Model variational inverse problems governed by partial differential equations*. Institute for Computational Engineering and Sciences, University of Texas, Austin, TX (ICES Report 11-05)
- Pralong MR and Gudmundsson GH (2011) Bayesian estimation of basal conditions on Rutford Ice Stream, West Antarctica, from surface data. *J. Glaciol.*, **57**(202), 315–324 (doi: 10.3189/002214311796406004)
- Price SF, Payne AJ, Howat IM and Smith BE (2011) Committed sea-level rise for the next century from Greenland ice sheet dynamics during the past decade. *Proc. Natl Acad. Sci. USA (PNAS)*, **108**(22), 8978–8983 (doi: 10.1073/pnas.1017313108)
- Raymond MJ (2007) Estimating basal properties of glaciers and ice streams from surface measurements. (PhD thesis, ETH Zürich)
- Raymond MJ and Gudmundsson GH (2005) On the relationship between surface and basal properties on glaciers, ice sheets, and ice streams. *J. Geophys. Res.*, **110**(B8), B08411 (doi: 10.1029/2005JB003681)
- Raymond MJ and Gudmundsson GH (2009) Estimating basal properties of ice streams from surface measurements: a nonlinear Bayesian inverse approach applied to synthetic data. *Cryosphere*, **3**(2), 265–278 (doi: 10.5194/tc-3-265-2009)
- Tarantola A (2005) *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA
- Thorsteinsson T, Raymond CF, Gudmundsson GH, Bindshadler RA, Vornberger P and Joughin I (2003) Bed topography and lubrication inferred from surface measurements on fast-flowing ice streams. *J. Glaciol.*, **49**(167), 481–490 (doi: 10.3189/172756503781830502)

Tröltzsch F (2010) *Optimal control of partial differential equations: theory, methods and applications*. American Mathematical Society, Providence, RI  
 Vieli A and Payne AJ (2003) Application of control methods for modelling the flow of Pine Island Glacier, Antarctica. *Ann. Glaciol.*, **36** 197–204 (doi: 10.3189/172756403781816338)  
 Vogel CR (2002) *Computational methods for inverse problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA

**APPENDIX A. ADJOINT-BASED COMPUTATION OF DERIVATIVES**

Here we derive expressions for the gradient of the regularized least-squares functional (Eqn (5)), which involves solution of the Stokes equations (Eqn (4)). Gradients of functionals defined by solutions of PDEs can be derived systematically using the method of Lagrange multipliers (Gunzburger, 2003; Ito and Kunisch, 2008; Tröltzsch, 2010; Borzi and Schulz, 2012); for an illustrative application of this approach to model problems see Petra and Stadler (2011). Variations of the Lagrangian functional, which is a sum of the cost functional and the weak form of the PDE, can be used to derive the weak forms of the adjoint and incremental equations, as well as expressions for the gradient and Hessian. For an introduction to the mathematical theory of the finite-element method (and, in particular, strong and weak forms of PDEs), refer to Oden and Reddy (1976), Braess (1997) and Gockenbach (2006). The weak forms of the forward and adjoint equations can then be transformed to strong forms by application of Green’s identity (a multidimensional analogue of integration by parts; Evans, 1998; Gockenbach, 2006), which, for the inverse problem considered in this paper, results in the equations presented in Section 3. For simplicity of the presentation, in the derivation below we focus on the particular case described in model 1, i.e. inversion for the basal sliding coefficient in the linear Stokes equations.

Let  $L^2(\Omega)$  denote the space of square-integrable functions, and let  $\mathbf{H}^1(\Omega) = (H^1(\Omega))^3$ , where  $H^1(\Omega)$  denotes the subspace of  $L^2(\Omega)$  consisting of functions whose first derivatives also belong to  $L^2(\Omega)$ , and the domain  $\Omega$  is a 3-D rectangular section of an ice sheet. In addition, to impose the necessary boundary conditions, we define

$$\tilde{\mathbf{H}}(\Omega) = \{ \mathbf{u} \in \mathbf{H}^1(\Omega) : \mathbf{u}|_{\Gamma_l} = \mathbf{u}|_{\Gamma_r}, \mathbf{u} \cdot \mathbf{n}|_{\Gamma_b} = 0 \},$$

where  $\Gamma_l$ ,  $\Gamma_r$  and  $\Gamma_b$  are the pair of opposing boundaries and the boundary at the base (as described in Section 2). For  $\mathbf{u}, \mathbf{v} \in \tilde{\mathbf{H}}(\Omega)$  and  $p, q \in L^2(\Omega)$ , we introduce the bilinear forms  $A(\cdot, \cdot) : \tilde{\mathbf{H}}(\Omega) \times \tilde{\mathbf{H}}(\Omega) \rightarrow R$  and  $B(\cdot, \cdot) : \tilde{\mathbf{H}}(\Omega) \times L^2(\Omega) \rightarrow R$ , the trilinear form  $S(\cdot, \cdot, \cdot) : \tilde{\mathbf{H}}(\Omega) \times \tilde{\mathbf{H}}(\Omega) \times H^1(\Gamma_b) \rightarrow R$ , and the linear form  $F(\cdot) : \tilde{\mathbf{H}}(\Omega) \rightarrow R$ , defined by

$$A(\mathbf{u}, \mathbf{v}) = \int_{\Omega} 2\eta \dot{\epsilon}_{\mathbf{u}} : \dot{\epsilon}_{\mathbf{v}} \, dx, \quad B(\mathbf{u}, p) = \int_{\Omega} -p \nabla \cdot \mathbf{u} \, dx,$$

$$S(\mathbf{u}, \mathbf{v}; \beta) = \int_{\Gamma_b} \beta T \mathbf{u} \cdot T \mathbf{v} \, ds, \quad F(\mathbf{u}) = \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, dx.$$

Furthermore, let us define

$$Q(\mathbf{u}, p; \mathbf{v}, q; \beta) = A(\mathbf{u}, \mathbf{v}) + B(\mathbf{v}, p) + B(\mathbf{u}, q) + S(\mathbf{u}, \mathbf{v}; \beta).$$

Then the weak form of the Stokes equation is: find  $(\mathbf{u}, p) \in \mathcal{U} = \tilde{\mathbf{H}}(\Omega) \times L^2(\Omega)$  such that

$$Q(\mathbf{u}, p; \mathbf{v}, q; \beta) = F(\mathbf{v}), \text{ for all } (\mathbf{v}, q) \in \mathcal{U}. \quad (\text{A1})$$

The Lagrangian functional,  $\mathcal{L} : \mathcal{U} \times \mathcal{U} \times H^1(\Gamma_b) \rightarrow R$ , which we use to derive the expressions for the optimality system, is given by

$$\mathcal{L}(\mathbf{u}, p; \mathbf{v}, q; \beta) = \mathcal{J}(\beta) + Q(\mathbf{u}, p; \mathbf{v}, q; \beta) - F(\mathbf{v}), \quad (\text{A2})$$

where  $\mathcal{J}(\beta)$  is the cost functional corresponding to the  $\beta$  inversion given by

$$\mathcal{J}(\beta) = \frac{1}{2} \int_{\Gamma_t} |\mathbf{u}(\beta) - \mathbf{u}^{\text{obs}}|^2 \, ds + \frac{\gamma_{\beta}}{2} \int_{\Gamma_b} |T \nabla \beta|^2 \, ds. \quad (\text{A3})$$

Lagrange function theory states that variations of the Lagrangian functional with respect to  $\beta$  are equivalent to the gradient of  $\mathcal{J}$  given in Eqn (5), i.e.

$$J_{\beta}(\hat{\beta}) = \mathcal{L}_{\beta}(\mathbf{u}, p; \mathbf{v}, q; \beta)(\hat{\beta}) = \mathcal{R}_{\beta}(\beta)[\hat{\beta}] + S(\mathbf{u}, \mathbf{v}, \hat{\beta}), \quad (\text{A4})$$

provided variations with respect to  $(\mathbf{u}, p; \mathbf{v}, q)$  in all directions  $(\hat{\mathbf{u}}, \hat{p}; \hat{\mathbf{v}}, \hat{q})$  vanish, i.e.

$$\mathcal{L}_{(\mathbf{v}, q)}(\mathbf{u}, p; \mathbf{v}, q; \beta)(\hat{\mathbf{v}}, \hat{q}) = Q(\mathbf{u}, p; \hat{\mathbf{v}}, \hat{q}; \beta) - F(\hat{\mathbf{v}}) = 0, \quad (\text{A5})$$

$$\mathcal{L}_{(\mathbf{u}, p)}(\mathbf{u}, p; \mathbf{v}, q; \beta)(\hat{\mathbf{u}}, \hat{p}) = W_{\mathbf{u}}(\mathbf{u})[\hat{\mathbf{u}}] + Q(\hat{\mathbf{u}}, \hat{p}; \mathbf{v}, q; \beta) = 0, \quad (\text{A6})$$

where the variations  $(\hat{\mathbf{u}}, \hat{p}; \hat{\mathbf{v}}, \hat{q}; \hat{\beta})$  are taken from the same spaces as  $(\mathbf{u}, p; \mathbf{v}, q; \beta)$ , and

$$W_{\mathbf{u}}(\mathbf{u})[\hat{\mathbf{u}}] = \int_{\Gamma_t} (\mathbf{u} - \mathbf{u}^{\text{obs}}) \hat{\mathbf{u}} \, dx,$$

$$\mathcal{R}_{\beta}(\beta)[\hat{\beta}] = \gamma_{\beta} \int_{\Gamma_b} T \nabla \beta \cdot T \nabla \hat{\beta} \, ds.$$

In the above equations, the subscripts for the operators denote the variables with respect to which we take variations. Note that Eqn (A5) is the weak form of the forward equation and Eqn (A6) is the weak form of the adjoint equation. The solutions of the forward and adjoint equations are used to evaluate the gradient of  $\mathcal{J}(\beta)$  given in weak form by Eqn (A4). In order to recover the strong form of the forward, adjoint and gradient equations, one applies Green’s identity.

Next, we derive the second-order variational derivatives of the Lagrangian, which are needed for solution of the optimality system by Newton’s method. In abstract form, Newton’s method computes an update direction  $(\tilde{\mathbf{u}}, \tilde{p}; \tilde{\mathbf{v}}, \tilde{q}; \tilde{\beta})$  from the following Newton step:

$$W_{\mathbf{u}\mathbf{u}}(\hat{\mathbf{u}})[\tilde{\mathbf{u}}] + S(\hat{\mathbf{u}}, \mathbf{v}; \tilde{\beta}) + Q(\hat{\mathbf{u}}, \hat{p}; \tilde{\mathbf{v}}, \tilde{q}; \beta) = -\mathcal{L}_{(\mathbf{u}, p)}, \quad (\text{A7})$$

$$S(\tilde{\mathbf{u}}, \mathbf{v}; \tilde{\beta}) + \mathcal{R}_{\beta}(\beta)[\tilde{\beta}] + S(\mathbf{u}, \tilde{\mathbf{v}}; \tilde{\beta}) = -\mathcal{L}_{\beta}, \quad (\text{A8})$$

$$Q(\tilde{\mathbf{u}}, \tilde{p}; \tilde{\mathbf{v}}, \tilde{q}; \beta) + S(\mathbf{u}, \tilde{\mathbf{v}}; \tilde{\beta}) = -\mathcal{L}_{(\mathbf{v}, q)}, \quad (\text{A9})$$

for all variations  $(\hat{\mathbf{u}}, \hat{p}; \hat{\mathbf{v}}, \hat{q}; \hat{\beta})$ , where  $\mathcal{L}_{\beta}$ ,  $\mathcal{L}_{(\mathbf{u}, p)}$  and  $\mathcal{L}_{(\mathbf{v}, q)}$  denote the first variations of the Lagrangian, given by Eqns (A4–A6), respectively. Here Eqns (A7) and (A9) are the weak forms of the incremental adjoint and incremental forward equations, respectively, and

$$W_{\mathbf{u}\mathbf{u}}(\hat{\mathbf{u}})[\tilde{\mathbf{u}}] = \int_{\Gamma_t} \tilde{\mathbf{u}} \hat{\mathbf{u}} \, dx,$$

$$\mathcal{R}_{\beta}(\beta)[\tilde{\beta}] = \gamma_{\beta} \int_{\Gamma_b} T \nabla \tilde{\beta} \cdot T \nabla \hat{\beta} \, ds.$$

Since we assume that  $(\mathbf{u}, p)$  and  $(\mathbf{v}, q)$  satisfy the forward and the adjoint equations, the right-hand sides of Eqns (A7) and (A9) vanish. Then we apply block elimination to

eliminate the incremental forward variables  $(\tilde{\mathbf{u}}, \tilde{p})$  from the incremental forward equation,

$$Q(\tilde{\mathbf{u}}, \tilde{p}; \hat{\mathbf{v}}, \hat{q}; \beta) = -S(\mathbf{u}, \hat{\mathbf{v}}; \tilde{\beta}), \tag{A10}$$

and the incremental adjoint variables  $(\tilde{\mathbf{v}}, \tilde{q})$  from the incremental adjoint equation,

$$Q(\hat{\mathbf{u}}, \hat{p}; \tilde{\mathbf{v}}, \tilde{q}; \beta) = -W_{uu}(\hat{\mathbf{u}})[\tilde{\mathbf{u}}] - S(\hat{\mathbf{u}}, \mathbf{v}; \tilde{\beta}). \tag{A11}$$

This results in the linear system for the Newton step:

$$\mathcal{H}(\beta)[\tilde{\beta}, \hat{\beta}] = -\mathcal{G}(\beta)(\hat{\beta}),$$

where the gradient,  $\mathcal{G}$ , is given by the first variation of the Lagrangian with respect to  $\beta$  (given by the left-hand side in Eqn (A4)), and the Hessian,  $\mathcal{H}$ , is given by

$$\mathcal{H}(\beta)[\tilde{\beta}, \hat{\beta}] = \mathcal{R}_{\beta\beta}(\hat{\beta})[\tilde{\beta}] + S(\tilde{\mathbf{u}}, \mathbf{v}; \hat{\beta}) + S(\mathbf{u}, \tilde{\mathbf{v}}; \hat{\beta}),$$

where  $(\tilde{\mathbf{u}}, \tilde{p})$ , and  $(\tilde{\mathbf{v}}, \tilde{q})$  satisfy the incremental forward (Eqn (A10)) and incremental adjoint (Eqn (A11)) equations, respectively. We note that to obtain the Gauss–Newton approximation of the Hessian, we ignore the terms proportional to the adjoint variables (i.e.  $S(\tilde{\mathbf{u}}, \mathbf{v}; \hat{\beta})$ ).

### APPENDIX B. DISCRETIZATION

We discretize the weak form of the forward and adjoint Stokes equations (Eqns (A5) and (A6)) using the finite-element method with a hexahedral mesh and employ Taylor–Hood finite elements for the velocity–pressure solution, which satisfies the inf–sup stability condition required for numerical stability (Elman and others, 2005). Bilinear elements for  $\beta$ , the unknown inversion field in the basal boundary conditions of the Stokes model, are used.

Let us denote by  $\tilde{\mathbf{u}}, \tilde{\mathbf{v}}, \tilde{p}, \tilde{q}, \tilde{\beta}$  the vectors of discrete unknowns corresponding to the functions  $\tilde{\mathbf{u}}, \tilde{\mathbf{v}}, \tilde{p}, \tilde{q}, \tilde{\beta}$ . The discretized Newton step (at iteration  $k$ ) is given by the linear system:

$$\begin{bmatrix} \mathbf{W}_{uu} & \mathbf{W}_{u\beta} & \mathbf{A}^T \\ \mathbf{W}_{\beta u} & \mathbf{R} & \mathbf{C}^T \\ \mathbf{A} & \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{u}}_k \\ \tilde{\beta}_k \\ \tilde{\mathbf{v}}_k \end{bmatrix} = - \begin{bmatrix} \mathbf{g}_{(\mathbf{u},p)} \\ \mathbf{g}_\beta \\ \mathbf{g}_{(\mathbf{v},q)} \end{bmatrix}, \tag{B1}$$

where  $\mathbf{W}_{uu}, \mathbf{W}_{u\beta}, \mathbf{W}_{\beta u}$  and  $\mathbf{R}$  are the components of the Hessian matrix of the Lagrangian,  $\mathbf{A}$  is the discrete forward operator,  $\mathbf{C}$  is the Jacobian of the forward equation with respect to the optimization variable,  $\beta$ , and  $\mathbf{g}_{(\mathbf{u},p)}, \mathbf{g}_\beta$  and  $\mathbf{g}_{(\mathbf{v},q)}$  are the discrete gradients of the Lagrangian with respect to  $(\mathbf{u}, p), \beta$  and  $(\mathbf{v}, q)$ , respectively. As before, we assume that  $(\mathbf{u}_k, p_k)$  and  $(\mathbf{v}_k, q_k)$  satisfy the forward and the adjoint equations such that  $\mathbf{g}_{(\mathbf{u},p)} = \mathbf{g}_{(\mathbf{v},q)} = \mathbf{0}$ , and that terms proportional to the adjoint variable, i.e.  $\mathbf{W}_{u\beta}$  and  $\mathbf{W}_{\beta u}$ , can be neglected. We recall that the latter assumption gives rise to the Gauss–Newton approximation of the Hessian.

The incremental forward and adjoint variables,  $\tilde{\mathbf{u}}_k$  and  $\tilde{\mathbf{v}}_k$ , are computed from the first and last equations in Eqn (B1), namely

$$\begin{aligned} \tilde{\mathbf{u}}_k &= -\mathbf{A}^{-1} \mathbf{C} \tilde{\beta}_k, \\ \tilde{\mathbf{v}}_k &= -\mathbf{A}^{-T} \mathbf{W}_{uu} \tilde{\mathbf{u}}_k, \end{aligned}$$

where  $\mathbf{A}^{-T}$  is the adjoint operator. Finally, the discretized Newton system (for the inverse problem)

$$\mathbf{H} \tilde{\beta}_k = -\mathbf{g}_\beta, \quad \text{with } \mathbf{H} = \mathbf{R} + \mathbf{C}^T \mathbf{A}^{-T} \mathbf{W}_{uu} \mathbf{A}^{-1} \mathbf{C}$$

is solved iteratively using an inexact preconditioned CG method with preconditioning by the inverse of the regularization operator,  $\mathbf{R}$ .

### APPENDIX C. NEWTON'S METHOD FOR THE NONLINEAR STOKES EQUATIONS

Here we consider the solution of the nonlinear Stokes problem (Eqn (4)) using Newton's method. We state the energy minimization problem that can be used to implement a line search for finding the appropriate step length for the Newton direction. It is known that the Stokes problem can be written as the following optimization problem (e.g. Dukowicz and others, 2010):

$$\min_{\mathbf{u}} \mathcal{J}(\mathbf{u}) = \int_{\Omega} (\Phi - \mathbf{f} \cdot \mathbf{u}) \, dx + \frac{1}{m+1} \int_{\Gamma_b} \beta |\mathbf{T}\mathbf{u}|^{m+1} \, ds, \tag{C1}$$

subject to

$$\nabla \cdot \mathbf{u} = 0, \tag{C2}$$

where

$$\Phi(\mathbf{u}) = \frac{2n}{n+1} A^{-\frac{1}{n}} \hat{\epsilon}_{\parallel}^{\frac{1+n}{2n}}.$$

The corresponding Lagrangian is

$$\mathcal{L}(\mathbf{u}, p) = \mathcal{J}(\mathbf{u}) + (\nabla \cdot \mathbf{u}, p)_{\Omega},$$

where the pressure acts as a Lagrange multiplier. The optimality conditions for the minimization problem given by Eqn (C1) are computed as before, by taking variations of the Lagrangian with respect to  $\mathbf{u}$  and  $p$ , and by using the fact that at the solution these variations with respect to all variables must vanish. It can be seen that the variation of  $\Phi$  with respect to  $\mathbf{u}$  in the direction of  $\mathbf{v}$  is  $2\eta(\mathbf{u})\hat{\epsilon}_{\mathbf{u}} : \hat{\epsilon}_{\mathbf{v}}$ , and hence, after applying Green's identity one recovers the strong form of the nonlinear Stokes equations. Note that since the current iterate as well as the update satisfy the incompressibility condition (Eqn (C2)), using the linearity property of the divergence operator, we conclude that the incompressibility condition is satisfied for all iterates as well.

*MS received 28 August 2011 and accepted in revised form 18 May 2012*