Applications of Mathematics

Erik Bängtsson; Maya Neytcheva

Numerical simulations of glacial rebound using preconditioned iterative solution methods

Applications of Mathematics, Vol. 50 (2005), No. 3, 183--201

Persistent URL: http://dml.cz/dmlcz/134602

50 (2005) APPLICATIONS OF MATHEMATICS No. 3, 183–201

NUMERICAL SIMULATIONS OF GLACIAL REBOUND USING PRECONDITIONED ITERATIVE SOLUTION METHODS

Erik Bängtsson, Maya Neytcheva, Uppsala

Abstract. This paper discusses finite element discretization and preconditioning strategies for the iterative solution of nonsymmetric indefinite linear algebraic systems of equations arising in modelling of glacial rebound processes. Some numerical experiments for the purely elastic model setting are provided. Comparisons of the performance of the iterative solution method with a direct solution method are included as well.

Keywords: elasticity, advection, FEM, error estimates, saddle point problem, iterative methods

MSC 2000: 65F10, 86-08

1. Introduction

This work deals with numerical simulations of the response of the solid Earth to large-scale glacial advance and recession and the resulting crustal stress state. Very large ongoing glacioisostatic recovery is registered in Central Scandinavia. The amount of residual rebound, coupled with low and declining rates of recovery, imply that the lithosphere may not completely reach equilibrium within itself before the next glacial period. All evidence at hand indicates that the effect of the rebound processes on crustal stability has to be analysed and taken into account in the context of other problems, such as various ground and underground excavating/building activities near the regions of rebound, and, in particular, for the safety assessment of radioactive waste repositories.

Two main mathematical frameworks to model the dynamics of the lithosphere during glacial advance and recession are commonly used.

Isostatic model. Isostasy adopts the concept that the elevation of the Earth's surface seeks a balance between the weight of lithospheric rocks and the buoyancy of asthenospheric "fluid" (nearly-molten rock). The model describes the geophysical

problem in terms of a system of partial differential equations (PDEs) which describe the equilibrium state of a pre-stressed visco-elastic material body, subject to surface and body forces. It includes a first-order term representing *advection of pre-stress*, the incorporation of which has proven to be crucial for the successful modelling of the underlying processes, see [22].

Coupled plate tectonic hypothesis and (thermo-)visco-plastic rheology. In this second model the stresses in the lithosphere are modelled as being the result of glaciation, superimposed on the plate tectonic processes. The problem is then formulated as a contact problem with visco-plastic friction. In this setting the advection stress corresponds to a convective term in the equation of motion. The latter model allows for a more complex analysis of the interplay between the various processes occurring in the different layers of the Earth mantle.

In this paper we deal with numerical simulations of the isostasy model only. This model is broadly used by geo-scientists for numerical simulations of glaciation effects. The models are discretized using the finite element method (FEM) and solved by a commercial FE package at hand. It is observed, see [23], that most such packages are not well suited for geophysical applications related to deformations of a viscoelastic Earth. In particular, it is not possible in a straightforward manner to include the pre-stress advection term and some modifications in the packages are necessary. Another potential difficulty when using a standard commercial FE package is encountered when modelling incompressible materials. Most such packages allow treatment of almost incompressible materials only.

In this paper we address the following issues:

(i) the pre-stress advection term is included in the isostatic model and we study its influence on the properties of the resulting FE linear system;
(ii) the original system of PDEs is reformulated as a saddle point problem in order to treat compressible and incompressible materials in a unified manner;
(iii) we apply a preconditioned iterative solution method, in contrast to the commonly used direct solution methods, and discuss some preconditioning strategies;
(iv) we present some numerical tests and comparisons with results obtained when solving the same problem using a commercial FE package.

2. Description and analysis of the isostatic model

2.1. Governing equations

Consider a pre-stressed visco-elastic material body which occupies a domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ with a (piecewise) smooth boundary $\partial\Omega$. The governing balance

equation describing the equilibrium state of a solid body subject to surface and body forces (cf. [14]) is of the form

$$\underbrace{\underline{\nabla} \cdot \underline{\sigma}}_{\text{(A)}} + \underbrace{\nabla(\mathbf{u} \cdot \nabla p^{(0)})}_{\text{(B)}} - \underbrace{\varrho^{(0)} g^{(0)} \mathbf{e}_d \nabla \cdot \mathbf{u}}_{\text{(C)}} - \underbrace{\varrho^{(0)} \underline{\nabla} \varphi^{(\Delta)}}_{\text{(D)}} = \mathbf{0} \quad \text{in } \Omega. \tag{1}$$

In (1), the term (A) describes the force from spatial gradients in stress, $\mathbf{u} = [u_i]_{i=1}^d \in \Omega$ is the displacement vector and ϱ is the density. When a large elastic solid is put in a gravitational field, it becomes gravitationally pre-stressed with pressure $p^{(0)}$. This pressure can be regarded as an initial condition imposed on the problem and does not cause deformations. The term (B) represents the advection of this pre-stress and describes how it is carried by the moving material. The terms (C) and (D) describe perturbations of the gravitational force and gravitational acceleration due to changes of density. In the present study, a non-selfgravitating (flat) Earth model is used, which implies that the term (D) vanishes because of the constant gravitational field.

In its full complexity, the model includes viscoelastic constitutive relations. However, in this paper we discuss a purely elastic material behavior only, as analyzed in [14], for instance. Although simplified, this model serves as a reference benchmark for the study of the above listed tasks (i)–(iv).

R e m a r k 2.1. It can be seen (cf. [21]) that incorporating viscoelastic constitutive relations with particular stress-relaxation functions (of Dirichlet-Prony type) does not lead to significant changes in the properties of the linear algebraic systems to be solved when computing the deformation of the viscoelastic solid through time. This topic falls, however, out of the scope of the present paper.

Under the assumptions that the displacements are small and the solid is linear and isotropic, the Lamé-Navier formulation of Equation (1) takes the form

$$\begin{cases} \underbrace{-2\mu\Delta\mathbf{u} - \mu\nabla \times (\nabla \times \mathbf{u})}_{\text{(A1)}} - \underbrace{\underline{\nabla}(\mathbf{u} \cdot \mathbf{b})}_{\text{(B)}} + \underbrace{\mathbf{c}(\nabla \cdot \mathbf{u})}_{\text{(C)}} - \underbrace{\mu\underline{\nabla} p}_{\text{(A2)}} = \mathbf{0}, \\ \mu\nabla \cdot \mathbf{u} - \dfrac{\mu^2}{\lambda} p = \mathbf{0}, \end{cases} \tag{2}$$

where $p = \lambda\mu^{-1}\nabla \cdot \mathbf{u}$ is the kinematic pressure, introduced as an extra variable, and λ and μ are the Lamé coefficients. In this way the term (A) in (1) is split into (A1)+(A2).

Equation (2) generalizes (1), since $\nabla p^{(0)}$ and $\varrho^{(0)} g^{(0)} \mathbf{e}_d$ are replaced by arbitrary vectors $\mathbf{b}$ and $\mathbf{c}$. In this way, the analysis of Equation (2) is valid for a broader class of problems. If $\mathbf{c} = \mathbf{0}$, $\mathbf{b}$ is constant, and $\mathbf{u}$ is curl-free, we obtain a system of PDEs

which is similar to the linearized Navier-Stokes (NS) equations. For reasons of comparison we include the general form of the steady-state (linearized) incompressible NS equations (3) and compressible NS equations (4):

$$\begin{cases} \underbrace{-\mu\Delta\mathbf{u}}_{\text{(NA1)}} - \underbrace{(\mathbf{b}\cdot\underline{\nabla})\mathbf{u}}_{\text{(NB)}} + \underbrace{\underline{\nabla}p}_{\text{(NA2)}} = \mathbf{f} \text{ in } \Omega, \\ \nabla\cdot\mathbf{u} = 0 \text{ in } \Omega, \end{cases} \tag{3}$$

$$\begin{cases} \underbrace{-\mu\Delta\mathbf{u}}_{\text{(NA1)}} + \underbrace{\varrho(p)(\mathbf{b}\cdot\nabla)\mathbf{u}}_{\text{(NB)}} - \underbrace{\nu\underline{\nabla}(\nabla\cdot\mathbf{u})}_{\text{(NC)}} + \underbrace{\underline{\nabla}p}_{\text{(NA2)}} = \mathbf{f} \text{ in } \Omega, \\ \nabla\cdot(\varrho\mathbf{u}) = g \text{ in } \Omega. \end{cases} \tag{4}$$

In (3), (4) $\mathbf{u}$ is the velocity, μ is the viscosity constant, p is the pressure, $\varrho(p)$ is the density and ν is the so-called bulk viscosity constant. The vector $\mathbf{b}$ in (3), (4) replaces $\mathbf{u}$ in the linearized form of the term $(\mathbf{u}\cdot\nabla)\mathbf{u}$ in the nonlinear NS equations. The boundary conditions are omitted for simplicity.

Comparing (2) with (3), three major differences can be pointed out. First, (2) is of more general form, which can be seen from the following relation between vector operators:

$$\underline{\nabla}(\mathbf{u}\cdot\mathbf{b}) = (\mathbf{u}\cdot\nabla)\mathbf{b} + (\mathbf{b}\cdot\nabla)\mathbf{u} + \mathbf{b}\times(\nabla\times\mathbf{u}) + \mathbf{u}\times(\nabla\times\mathbf{b}).$$

Thus, in general, the term (NB) is only one part of the term (B). Second, μ in (2) is a material parameter which is positive and bounded away from zero, while in (3) and (4) μ is the (scaled) inverse of the Reynolds number (Re) and, thus, becomes very small for large Re. The latter means that (3), (4) can be seen as singularly perturbed adection-diffusion problems, which does not hold for (2). Third, it is known (see, for instance, [16] for the compressible NS) that the principal bilinear form in the variational formulation of (3) and (4) is always coercive. We show in Section 2.2 that for the corresponding principal bilinear form in (2) coercivity does not hold in general. To summarize, the problem arising in post-glacial rebound does not reduce to NS and the known numerical solution methods and preconditioning techniques for NS are not directly applicable. There are certain similarities, however, which allow us to use some well-known and studied solution techniques for saddle point problems when performing numerical simulations of the problem in question.

2.2. Variational formulation

The variational formulation of Equation (2) is defined in terms of the Sobolev spaces $\mathbf{V} = (H_0^1(\Omega))^d$, $d = 2, 3$ equipped with the norm $\|\cdot\|_{\mathbf{V}} = \|\cdot\|_1$, and $P =$

$\{p \in L^2(\Omega);\ \int_\Omega \mu p \,\mathrm{d}\Omega = 0\}$, equipped with the norm $\|\cdot\|_P = \|\cdot\|_0$. It leads to the following mixed variable problem: Find $\mathbf{u} \in \mathbf{V}$ and $p \in P$ such that

$$\begin{cases} a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) = \mathbf{0} & \forall\, \mathbf{v} \in \mathbf{V}, \\ b(\mathbf{u},q) - c(p,q) = 0 & \forall\, q \in P, \end{cases} \tag{5}$$

where the bilinear forms $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ and $c(\cdot,\cdot)$ are defined as follows

$$\begin{aligned} a(\mathbf{u},\mathbf{v}) = \int_\Omega \Big[& 2\mu \sum_{k=1}^{d} (\underline{\nabla} u_k)\cdot(\underline{\nabla} v_k) - \mu(\underline{\nabla}\times\mathbf{u})\cdot(\underline{\nabla}\times\mathbf{v}) \\ & - \underline{\nabla}(\mathbf{u}\cdot\mathbf{b})\cdot\mathbf{v} + (\nabla\cdot\mathbf{u})(\mathbf{c}\cdot\mathbf{v}) \Big] \,\mathrm{d}\Omega, \\ b(\mathbf{u},q) = \int_\Omega \mu(\underline{\nabla}\cdot\mathbf{u}) q \,\mathrm{d}\Omega \quad & \text{and} \quad c(p,q) = \int_\Omega \frac{\mu^2}{\lambda} pq \,\mathrm{d}\Omega. \end{aligned}$$

The usual way to ensure existence and uniqueness of the solution of a variational problem is to show boundedness and coercivity of $a(\mathbf{u},\mathbf{v})$ and then apply the Lax-Milgram lemma. Below we show that the bilinear form $a(\mathbf{u},\mathbf{v})$ in (5) is not coercive in general and only weaker estimates hold.

To address in more detail the boundedness and coercivity of the bilinear form $a(\cdot,\cdot)$, we introduce auxiliary bilinear forms $\hat{a}(\cdot,\cdot)$, $\tilde{a}(\cdot,\cdot)$ and $\tilde{\tilde{a}}(\cdot,\cdot)$ as

$$\begin{aligned} \hat{a}(\mathbf{u},\mathbf{v}) &= \int_\Omega 2\mu \sum_{k=1}^{d} (\underline{\nabla} u_k)\cdot(\underline{\nabla} v_k) \,\mathrm{d}\Omega, \\ \tilde{a}(\mathbf{u},\mathbf{v}) &= \int_\Omega \Big[2\mu \sum_{k=1}^{d} (\underline{\nabla} u_k)\cdot(\underline{\nabla} v_k) - \mu(\underline{\nabla}\times\mathbf{u})\cdot(\underline{\nabla}\times\mathbf{v}) \Big] \,\mathrm{d}\Omega, \\ \tilde{\tilde{a}}(\mathbf{u},\mathbf{v}) &= \int_\Omega \Big(-\underline{\nabla}(\mathbf{u}\cdot\mathbf{b})\mathbf{v} + (\nabla\cdot\mathbf{u})(\mathbf{c}\cdot\mathbf{v}) \Big) \,\mathrm{d}\Omega. \end{aligned}$$

By construction, $a(\mathbf{u},\mathbf{v}) = \tilde{a}(\mathbf{u},\mathbf{v}) + \tilde{\tilde{a}}(\mathbf{u},\mathbf{v})$. We now make the following assumptions on the coefficient vectors $\mathbf{b}(\mathbf{x}) \in \mathbb{R}^d$ and $\mathbf{c}(\mathbf{x}) \in \mathbb{R}^d$: there exist constants α_1, α_2 and β, independent on $\mathbf{u}$ and $\mathbf{v}$, such that

$$|b_i(\mathbf{x})| \leqslant \alpha_1, \quad i = 1,\dots,d, \tag{6}$$

$$|\nabla\cdot\mathbf{b}| \leqslant \alpha_2, \tag{7}$$

$$|\mathbf{c}| \leqslant \beta. \tag{8}$$

We also assume that the problem (5) possesses a solution, and want to show coercivity (in some weak form) and to derive quasi-optimal error bounds for the corresponding Galerkin finite element method.

First we note that the dominating part $\tilde{a}(\mathbf{u},\mathbf{v})$ of $a(\mathbf{u},\mathbf{v})$ is bounded and coercive. The latter can be seen by using the well-known Korn inequality. It is known that there exists a constant $K(\Omega)$, which may depend only on the domain Ω and on the boundary conditions, such that

$$K(\Omega)\hat{a}(\mathbf{u},\mathbf{u}) \leqslant \tilde{a}(\mathbf{u},\mathbf{u}) \leqslant 2\hat{a}(\mathbf{u},\mathbf{u}) \quad \forall\, \mathbf{u} \in \mathbf{V}. \tag{9}$$

Therefore, there exist constants K_1 and K_2, which depend only on Ω and on the boundary conditions, such that

$$|\tilde{a}(\mathbf{u},\mathbf{v})| \leqslant K_1\|\mathbf{u}\|_1\|\mathbf{v}\|_1 \quad \text{and} \quad \tilde{a}(\mathbf{u},\mathbf{u}) \geqslant K_2\|\mathbf{u}\|_1^2. \tag{10}$$

We consider now the first-order terms in $\tilde{\tilde{a}}(\mathbf{u},\mathbf{v})$. The following estimates hold:

$$\begin{aligned} \left|\int_\Omega \underline{\nabla}(\mathbf{u}\cdot\mathbf{b})\mathbf{v}\,\mathrm{d}\Omega\right| &\leqslant \sum_{k=1}^{d}\int_\Omega \left|\frac{\partial}{\partial x_k}(u_k b_k)v_k\right|\mathrm{d}\Omega \\ &\leqslant \sum_{k=1}^{d}\int_\Omega \left|\frac{\partial u_k}{\partial x_k}b_k v_k\right|\mathrm{d}\Omega + \sum_{k=1}^{d}\int_\Omega \left|u_k\frac{\partial b_k}{\partial x_k}v_k\right|\mathrm{d}\Omega \\ &\leqslant \alpha_1 d\|\mathbf{u}\|_1\|\mathbf{v}\|_0 + \alpha_2 d\|\mathbf{u}\|_0\|\mathbf{v}\|_0, \end{aligned} \tag{11}$$

$$\left|\int_\Omega (\underline{\nabla}\cdot\mathbf{u})(\mathbf{c}\cdot\mathbf{v})\,\mathrm{d}\Omega\right| \leqslant \sum_{k=1}^{d}\int_\Omega |(\underline{\nabla}\cdot\mathbf{u})c_k v_k|\,\mathrm{d}\Omega \leqslant \beta d\|\mathbf{u}\|_1\|\mathbf{v}\|_0. \tag{12}$$

Therefore,

$$\begin{aligned} |\tilde{\tilde{a}}(\mathbf{u},\mathbf{v})| &\leqslant d(\alpha_1\|\mathbf{u}\|_1 + \alpha_2\|\mathbf{u}\|_0 + \beta\|\mathbf{u}\|_1)\|\mathbf{v}\|_0 \\ &= d(\alpha_1+\beta)\|\mathbf{u}\|_1\|\mathbf{v}\|_0 + d\alpha_2\|\mathbf{u}\|_0\|\mathbf{v}\|_0 \\ &\leqslant \sigma\|\mathbf{u}\|_1\|\mathbf{v}\|_0, \end{aligned} \tag{13}$$

where $\sigma = d(\alpha_1+\alpha_2+\beta)$. We now use Young's inequality ($ab \leqslant \frac{1}{2}\varepsilon|a|^2 + \frac{1}{2}\varepsilon^{-1}|b|^2$) as follows

$$\|\mathbf{u}\|_1\|\mathbf{v}\|_0 \leqslant \frac{\varepsilon}{2}\|\mathbf{u}\|_1^2 + \frac{1}{2\varepsilon}\|\mathbf{v}\|_0^2 \quad \forall\, \varepsilon > 0. \tag{14}$$

Combining (13) and (14) we obtain

$$|\tilde{\tilde{a}}(\mathbf{u},\mathbf{v})| \leqslant \sigma\frac{\varepsilon}{2}\|\mathbf{u}\|_1^2 + \sigma\frac{1}{2\varepsilon}\|\mathbf{v}\|_0^2. \tag{15}$$

We want to prove a relation of the type (Gårding inequality) $a(\mathbf{u},\mathbf{u}) \geqslant C^{(1)}\|\mathbf{u}\|_1^2 - C^{(2)}\|\mathbf{u}\|_0^2$, where $C^{(1)} > 0$ holds for all $\mathbf{u} \in \mathbf{V}$. We show below that this is true for a particular choice of the parameter ε in (15),

$$
\begin{aligned}
a(\mathbf{u},\mathbf{v}) &= \tilde{a}(\mathbf{u},\mathbf{v}) + \tilde{\tilde{a}}(\mathbf{u},\mathbf{v}) \\
&\geqslant K_2\|\mathbf{u}\|_1^2 - \left|\tilde{\tilde{a}}(\mathbf{u},\mathbf{v})\right| \\
&\geqslant K_2\|\mathbf{u}\|_1^2 - \sigma\frac{\varepsilon}{2}\|\mathbf{u}\|_1^2 - \sigma\frac{1}{2\varepsilon}\|\mathbf{v}\|_0^2 \\
&= \Big(K_2 - \sigma\frac{\varepsilon}{2}\Big)\|\mathbf{u}\|_1^2 - \sigma\frac{1}{2\varepsilon}\|\mathbf{v}\|_0^2.
\end{aligned} \tag{16}
$$

We choose now ε such that $C^{(1)} \equiv K_2 - \sigma \cdot \frac{1}{2}\varepsilon > 0$, i.e., $0 < \varepsilon < 2K_2/\sigma$. For the latter choice of ε, $C^{(1)} > 0$ and $C^{(2)} \equiv \sigma/2\varepsilon > 0$.

Thus, we obtain that for all $\mathbf{u} \in \mathbf{V}$ there holds

$$
a(\mathbf{u},\mathbf{u}) \geqslant C^{(1)}\|\mathbf{u}\|_1^2 - C^{(2)}\|\mathbf{u}\|_0^2, \tag{17}
$$

where $C^{(1)} > 0$ and $C^{(2)} > 0$ do not depend on $\mathbf{u}$.

To show boundedness of $a(\mathbf{u},\mathbf{v})$, we denote $C^{(a)} = 2K_1 + \sigma$ and using the relation $\|\cdot\|_0 \leqslant \|\cdot\|_1$, we find

$$
\begin{aligned}
|a(\mathbf{u},\mathbf{v})| &\leqslant 2|\hat{a}(\mathbf{u},\mathbf{v})| + \sigma\|\mathbf{u}\|_1\|\mathbf{v}\|_0 \\
&\leqslant 2K_1\|\mathbf{u}\|_1\|\mathbf{v}\|_1 + \sigma\|\mathbf{u}\|_1\|\mathbf{v}\|_0 \\
&\leqslant C^{(a)}\|\mathbf{u}\|_1\|\mathbf{v}\|_1.
\end{aligned} \tag{18}
$$

R e m a r k 2.2. For incompressible materials, the term $\mathbf{c}\nabla\cdot\mathbf{u}$ becomes zero. When in addition $\mathbf{b} = \mathbf{e}_d$, then $\nabla\cdot\mathbf{b} = 0$ and we see that in this case the bilinear form $a(\mathbf{u},\mathbf{v})$ is coercive.

2.3. Finite element discretizations and error estimates

Let $\mathbf{V}^h$ and P^h be finite element subspaces of $\mathbf{V}$ and P, respectively, and $\mathbf{u}_h$, $\mathbf{v}_h$, p_h and q_h the discrete counterparts of $\mathbf{u}$, $\mathbf{v}$, p and q. The discrete formulation of (5) reads then as follows: Find $\mathbf{u}_h$ and p_h such that relations (19) hold for all $\mathbf{v}_h \in \mathbf{V}^h$ and for all $q_h \in P^h$.

$$
\begin{aligned}
a(\mathbf{u}_h,\mathbf{v}_h) + b(\mathbf{v}_h,p_h) &= (\mathbf{f}_h,\mathbf{v}_h) \quad \forall\,\mathbf{v}_h \in \mathbf{V}^h, \\
b(\mathbf{u}_h,q_h) - c(p_h,q_h) &= 0 \qquad\qquad \forall\,q_h \in P^h.
\end{aligned} \tag{19}
$$

As is well known, in order to obtain a stable discrete formulation, the finite element spaces $\mathbf{V}^h$ and P^h cannot be arbitrarily chosen. They have to form a stable pair, i.e., such that the discrete analog of the inf-sup condition holds, namely,

$$\sup_{\mathbf{u}_h \in \mathbf{V}^h} \frac{b(\mathbf{u}_h, p_h)}{\|\mathbf{u}_h\|_{\mathbf{V}^h}} \geqslant \gamma_h \|p_h\|_{P^h} \geqslant \gamma_0 \|p_h\|_{P^h} \quad \forall\, p_h \in P^h, \tag{20}$$

for some positive constant $\gamma_0 > 0$, which for practical purposes should not be very small.

The interpretation of the discrete LBB condition (20) is that if the LBB constant γ_0 is independent on the discretization parameter h, then the rate of convergence of the FE solution $\mathbf{u}_h$ to the solution of the continuous variational problem is bounded uniformly with respect to the problem parameters E and ν.

There exists a variety of stable finite element pairs. A preferred choice is $\mathbf{u}_h \in \pi_2^h$ (componentwise) and $p_h \in \pi_1^h$, where π_2^h and π_1^h are the spaces of piecewise quadratic and piecewise linear polynomials. The discretization error for $\mathbf{u}$ and p is shown to be

$$\|\mathbf{u} - \mathbf{u}_h\|_V + \|p - p_h\|_P \leqslant \text{const}\Big(\inf_{\mathbf{v}_h \in \mathbf{V}^h} \|\mathbf{u} - \mathbf{v}_h\|_V + \inf_{q_h \in P^h} \|p - q_h\|_P \Big),$$

for any elements $\mathbf{v}_h \in \mathbf{V}^h$ and $q_h \in P^h$.

2.3.1. Error estimates. As shown above, the bilinear form $a(\mathbf{u}, \mathbf{v})$ is not coercive in general. Following [1], we derive quasi-optimal error bounds for the Galerkin method, applied to the problem under consideration.

Assume that the following assumptions hold.

(A1) $|a(\mathbf{u}, \mathbf{v})| \leqslant C\|\mathbf{u}\|_1 \|\mathbf{v}\|_1$.

(A2) Let $\mathbf{V}^{h_N} \subset \mathbf{V}$, $h_N = 1/(N^d)$, $N = 1, 2, \ldots$, be a sequence of finite dimensional subspaces of $\mathbf{V}$. (For notational simplicity we omit the subscript N.) Let there exist a sequence of positive numbers $\{\delta_h\}_{N=1}^{\infty}$ such that $\lim\limits_{N \to \infty} \delta_h = 0$ and that for every $\mathbf{e} \in \mathbf{V}^h$ and $z \in P^h$, satisfying

$$\begin{aligned} a(\mathbf{e}, \mathbf{v}) + b(\mathbf{v}, z) &= \mathbf{0} \quad \forall\, \mathbf{v} \in \mathbf{V}^h, \\ b(\mathbf{e}, q) - c(z, q) &= 0 \quad \forall\, q \in P^h, \end{aligned}$$

there holds

$$\|\mathbf{e}\|_0 \leqslant \delta_N \|\mathbf{e}\|_1.$$

Let now $\mathbf{u} \in \mathbf{V}$ and $p \in P$ be given. Let $\mathbf{V}^h$ and P^h be finite-dimensional subspaces of $\mathbf{V}$ and P, respectively, and assume that there exist $\mathbf{u}_h^* \in \mathbf{V}^h$ and $p_h^* \in P^h$ such

that

$$
\begin{aligned}
a(\mathbf{u}-\mathbf{u}_h^*,\mathbf{v}_h)+b(\mathbf{v}_h,p-p_h^*)&=\mathbf{0}\quad\forall\,\mathbf{v}_h\in\mathbf{V}_h,\\
b(\mathbf{u}-\mathbf{u}_h^*,q_h)-c(p-p_h^*,q_h)&=0\quad\forall\,q_h\in P^h.
\end{aligned}
\tag{21}
$$

Then, choosing $\mathbf{v}=\mathbf{u}-\mathbf{u}_h^*=\mathbf{u}-\mathbf{u}_h+\mathbf{u}_h-\mathbf{u}_h^*$ and $q_h=p-p_h^*=p-p_h+p_h-p_h^*$ and subtracting the two equations in (21), we arrive at

$$
\begin{aligned}
a(\mathbf{u}_h-\mathbf{u}_h^*,\mathbf{u}_h-\mathbf{u}_h^*)+c(p_h-p_h^*,p_h-p_h^*)&\\
=a(\mathbf{u}-\mathbf{u}_h,\mathbf{u}_h-\mathbf{u}_h^*)+b(\mathbf{u}_h-\mathbf{u}_h^*,p-p_h)&\\
-b(\mathbf{u}-\mathbf{u}_h,p_h-p_h^*)+c(p-p_h,p_h-p_h^*)&
\end{aligned}
\tag{22}
$$

which holds for any $\mathbf{u}_h\in\mathbf{V}^h$ and any $p_h\in P^h$. We assume now that either

$$
\gamma\|p_h\|_0\leqslant\sup_{\mathbf{u}_h}\frac{b(\mathbf{u}_h,p_h)}{\|\mathbf{u}_h\|_1}\quad\text{or}\quad\alpha\|p_h\|_0^2\leqslant c(p_h,p_h),\quad\alpha>0,
$$

is satisfied. The former option holds true for LBB-stable discretizations and the latter option holds true for stabilized discretizations.

We combine the assumption (A2) with $\mathbf{e}=\mathbf{u}-\mathbf{u}_h^*$ and (17):

$$
\begin{aligned}
a(\mathbf{u}_h-\mathbf{u}_h^*,\mathbf{u}_h-\mathbf{u}_h^*)+c(p-p_h^*,p-p_h^*)&\geqslant a(\mathbf{u}_h-\mathbf{u}_h^*,\mathbf{u}_h-\mathbf{u}_h^*)\\
&\geqslant C^{(1)}\|\mathbf{u}_h-\mathbf{u}_h^*\|_1^2-C^{(2)}\|\mathbf{u}_h-\mathbf{u}_h^*\|_0^2\\
&\geqslant (C^{(1)}-C^{(2)}\delta_h^2)\|\mathbf{u}_h-\mathbf{u}_h^*\|_1^2.
\end{aligned}
\tag{23}
$$

Next we assume that a discrete LBB condition holds true (given in (20)) and obtain the following relations:

$$
\begin{aligned}
\|p-p_h^*\|_0&\leqslant\|p-p_h\|_0+\|p_h-p_h^*\|_0\\
&\leqslant\|p-p_h\|_0+\frac{1}{\gamma}\sup_{\mathbf{v}_h}\frac{b(\mathbf{v}_h,p_h-p_h^*)}{\|\mathbf{v}_h\|_1}\\
&\leqslant\|p-p_h\|_0+\frac{1}{\gamma}\sup_{\mathbf{v}_h}\frac{b(\mathbf{v}_h,p_h-p)}{\|\mathbf{v}_h\|_1}+\frac{1}{\gamma}\sup_{\mathbf{v}_h}\frac{b(\mathbf{v}_h,p-p_h^*)}{\|\mathbf{v}_h\|_1}\\
&\leqslant\|p-p_h\|_0+\frac{C^{(b)}}{\gamma}\|p-p_h\|_0+\frac{1}{\gamma}\sup_{\mathbf{v}_h}\frac{a(\mathbf{u}-\mathbf{u}_h^*,\mathbf{v}_h)}{\|\mathbf{v}_h\|_1}\\
&\leqslant\Big(1+\frac{C^{(b)}}{\gamma}\Big)\|p-p_h\|_0+\frac{C^{(a)}}{\gamma}\|\mathbf{u}-\mathbf{u}_h^*\|_1,
\end{aligned}
\tag{24}
$$

where we have used the first part of (21). From the boundedness estimates, applied to (22) and from (24) we obtain

$$
\begin{aligned}
(25)\ & a(\mathbf{u}_h - \mathbf{u}_h^*, \mathbf{u}_h - \mathbf{u}_h^*) + c(p_h - p_h^*, p_h - p_h^*) \\
&\leqslant C^{(a)}\|\mathbf{u}_h - \mathbf{u}_h^*\|_1\|\mathbf{u} - \mathbf{u}_h\|_1 + C^{(b)}\|\mathbf{u} - \mathbf{u}_h\|_1\|p_h - p_h^*\|_0 \\
&\quad + C^{(b)}\|\mathbf{u}_h - \mathbf{u}_h^*\|_1\|p - p_h\|_0 + C^{(c)}\|p - p_h^*\|_0\|p_h - p_h\|_0 \\
&\leqslant C^{(a)}\|\mathbf{u} - \mathbf{u}_h\|_1^2 + C^{(a)}\|\mathbf{u} - \mathbf{u}_h^*\|_1\|\mathbf{u} - \mathbf{u}_h\|_1 \\
&\quad + 2C^{(b)}\|\mathbf{u} - \mathbf{u}_h\|_1\|p - p_h\|_0 + C^{(b)}\|\mathbf{u} - \mathbf{u}_h\|_1\|p - p_h^*\|_0 \\
&\quad + C^{(b)}\|\mathbf{u} - \mathbf{u}_h^*\|_1\|p - p_h\|_0 + C^{(c)}\|p - p_h\|_0^2 \\
&\qquad\qquad + C^{(c)}\|p - p_h\|_0\|p - p_h^*\|_0 \\
&\leqslant \left[\alpha_1 C^{(a)} + \alpha_2 C^{(b)} + \alpha_3\frac{C^{(a)}C^{(b)}}{\gamma} + a_4\frac{C^{(a)}C^{(c)}}{\gamma}\right]\|\mathbf{u} - \mathbf{u}_h^*\|_1^2 \\
&\quad + \left[C^{(a)} + \frac{C^{(a)}}{4\alpha_1} + C^{(b)} + \frac{1}{2}C^{(b)}\Big(1 + \frac{C^{(b)}}{\gamma}\Big) + \frac{C^{(a)}C^{(b)}}{4\alpha_3\gamma}\right]\|\mathbf{u} - \mathbf{u}_h\|_1^2 \\
&\quad + \left[C^{(b)} + C^{(c)} + \frac{C^{(b)}}{4\alpha_2} + \frac{1}{2}(C^{(b)} + 2C^{(c)})\Big(1 + \frac{C^{(b)}}{\gamma}\Big)\right. \\
&\qquad\qquad \left. + \frac{C^{(a)}C^{(c)}}{4\alpha_4\gamma}\right]\|p - p_h\|_0^2.
\end{aligned}
$$

Combining (23) and (25) we obtain

$$
\begin{aligned}
(26)\quad & \left[C^{(1)} - C^{(2)}\delta_h^2 - \alpha_1 C^{(a)} - \alpha_2 C^{(b)} - \alpha_3\frac{C^{(a)}C^{(b)}}{\gamma} - \alpha_4\frac{C^{(a)}C^{(c)}}{\gamma}\right]\|\mathbf{u} - \mathbf{u}_h^*\|_1^2 \\
&\leqslant \left[C^{(a)} + \frac{C^{(a)}}{4\alpha_1} + C^{(b)} + \frac{1}{2}C^{(b)}\Big(1 + \frac{C^{(b)}}{\gamma}\Big) + \frac{C^{(a)}C^{(b)}}{4\alpha_3\gamma}\right]\|\mathbf{u} - \mathbf{u}_h\|_1^2 \\
&\quad + \left[C^{(b)} + C^{(c)} + \frac{C^{(b)}}{4\alpha_2} + \frac{1}{2}\big(C^{(b)} + 2C^{(c)}\big)\Big(1 + \frac{C^{(b)}}{\gamma}\Big) + \frac{C^{(a)}C^{(c)}}{4\alpha_4\gamma}\right] \\
&\qquad\qquad \times \|p - p_h\|_0^2
\end{aligned}
$$

Here $\alpha_1 \dots \alpha_4$ are arbitrary positive constants. We see that if we choose

$$
\alpha_1 = \frac{C^{(2)}\delta_h^2}{4C^{(a)}},\ \alpha_2 = \frac{C^{(2)}\delta_h^2}{4C^{(b)}},\ \alpha_3 = \frac{C^{(2)}\delta_h^2\gamma}{4C^{(a)}C^{(b)}} \text{ and } \alpha_4 = \frac{C^{(2)}\delta_h^2\gamma}{4C^{(a)}C^{(c)}}
$$

we obtain

$$
\begin{aligned}
(27)\quad & (C^{(1)} - 2C^{(2)}\delta_h^2)\|\mathbf{u} - \mathbf{u}_h^*\|_1^2 \\
&\leqslant \Big[C^{(1)} + C^{(a)} + \frac{{C^{(a)}}^2}{C^{(2)}\delta_h^2} + \frac{C^{(b)}}{2}\Big(3 + \frac{C^{(b)}}{\gamma}\Big) + \frac{{C^{(a)}}^2}{C^{(2)}\delta_h^2}\Big(1 + \frac{{C^{(b)}}^2}{\gamma^2}\Big)\Big] \\
&\qquad\qquad \times \|\mathbf{u} - \mathbf{u}_h\|_1^2 \\
&\quad + \Big[C^{(b)} + \frac{1}{2}\big(C^{(b)} + 2C^{(c)}\big)\Big(1 + \frac{C^{(b)}}{\gamma}\Big) \\
&\qquad\qquad + \frac{1}{C^{(2)}\delta_h^2}\Big({C^{(b)}}^2 + \frac{{C^{(a)}}^2{C^{(c)}}^2}{\gamma^2}\Big)\Big] \times \|p - p_h\|_0^2.
\end{aligned}
$$

We finally obtain the following error estimate, which holds for all $\mathbf{u}_h \in \mathbf{V}_h$, $p_h \in P$ and $N > N_0$:

$$
(28)\qquad \|\mathbf{u} - \mathbf{u}_h^*\|_1 + \|p - p_h^*\|_0 \leqslant C_1\|\mathbf{u} - \mathbf{u}_h\|_1 + C_2\|p - p_h\|_0
$$

and the constants C_1, C_2 do not depend on the discretization parameter h.

Further (see [1], for instance), it can be shown that the discrete solution $\mathbf{u}_h^*$, p_h^* exists and is unique for a sufficiently large N (small h), $N > N_0$.

R e m a r k 2.3. For small values of γ and δ_h we see from the latter derivations that the constants C_1 and C_2 can become large and the resulting error estimates become quite pessimistic. In practice, for a particular solution the constraints can take more favourable values when the solution is smooth and/or is not near incompressibility.

2.3.2. Equal order discretization for displacements and pressure. A known approach to circumvent the use of stable pairs of finite element spaces for saddle point problems and retain stable and locking-free solutions is to use equal order finite elements for displacements and pressure, and some stabilized version of the discrete problem. For example, a stabilized and consistent formulation of (19) can be obtained in the following manner. We take divergence of the first equation in (2), use the fact that divergence of curl of any vector function is equal to zero, and add the resulting equation to the second equation in (2), multiplied by a stabilization parameter σ_h. Formally we have the following sequence of transformations.

$$
\begin{aligned}
-2\mu\nabla\cdot\Delta\mathbf{u} - \mu\nabla\cdot(\nabla\times(\nabla\times\mathbf{u})) - \xi(\mathbf{u},\mathbf{b},\mathbf{c}) - \mu\nabla\cdot\underline{\nabla}p &= \nabla\cdot\mathbf{f}, \\
-2\mu\Delta(\nabla\cdot\mathbf{u}) - \mu\Delta p - \xi(\mathbf{u},\mathbf{b},\mathbf{c}) &= \nabla\cdot\mathbf{f}, \\
-\mu\Big(1 + \frac{2\mu}{\lambda}\Big)\Delta p &= \nabla\cdot\mathbf{f} + \xi(\mathbf{u},\mathbf{b},\mathbf{c}), \\
\mu\nabla\cdot\mathbf{u} - \sigma_h\mu\Big(1 + \frac{2\mu}{\lambda}\Big)\Delta p - \frac{\mu^2}{\lambda}p &= \sigma_h\nabla\cdot\mathbf{f} + \sigma_h\xi(\mathbf{u},\mathbf{b},\mathbf{c}),
\end{aligned}
$$

where $\xi(\mathbf{u},\mathbf{b},\mathbf{c}) = \nabla\cdot(\underline{\nabla}\mathbf{u}\cdot\mathbf{b}) - \nabla\cdot\mathbf{c}\nabla\cdot\mathbf{u}$. Then we consider the problem

$$
\begin{aligned}
(29)\quad -2\mu\Delta\mathbf{u} - \mu\nabla\times(\nabla\times\mathbf{u}) - \underline{\nabla}\mathbf{u}\cdot\mathbf{b} + \mathbf{c}\nabla\cdot\mathbf{u} - \mu\underline{\nabla}p &= \mathbf{f},\\
\mu\nabla\cdot\mathbf{u} - \sigma_h\mu\Big(1+\frac{\mu}{\lambda}\Big)\Delta p - \frac{\mu^2}{\lambda}p &= \sigma_h\nabla\cdot\mathbf{f} + \sigma_h\xi(\mathbf{u},\mathbf{b},\mathbf{c}).
\end{aligned}
$$

By using a similar technique as in [2] one can show that discrete LBB condition holds for problem (29) discretized by standard piecewise linear finite elements. The choice of the stabilization parameter $\sigma_h = O(h^2)$ can be justified as in [3]. There, a defect-correction algorithm is described in order to handle the term $\sigma_h\xi(\mathbf{u},\mathbf{b},\mathbf{c})$.

3. Preconditioning strategies

The finite element discretization of (19) leads to a linear algebraic system

$$
(30)\qquad \mathcal{A}\begin{bmatrix}\mathbf{u}_h\\ p_h\end{bmatrix} \equiv \begin{bmatrix}M & B^T\\ B & -C\end{bmatrix}\begin{bmatrix}\mathbf{u}_h\\ p_h\end{bmatrix} = \begin{bmatrix}\mathbf{r}_h\\ \mathbf{s}_h\end{bmatrix}.
$$

The system matrix $\mathcal{A}$ admits a saddle point form and is nonsymmetric indefinite. The nonsymmetricity is due to the discretized first order (advection) terms in the block M.

Much research has been devoted to iterative solution methods and preconditioning for saddle point problems. A recent comprehensive survey of numerical methods for saddle point problems is found in [5]. To summarize briefly, almost all preconditioners suggested for matrices of saddle point form utilize the block structure of the matrix and originate from the exact factorization

$$
(31)\quad \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix} = \begin{bmatrix}I & 0\\ A_{21}A_{11}^{-1} & I\end{bmatrix}\begin{bmatrix}A_{11} & 0\\ 0 & S\end{bmatrix}\begin{bmatrix}I & A_{11}^{-1}A_{12}\\ 0 & I\end{bmatrix}
$$

or

$$
(32)\quad \begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix} = \begin{bmatrix}A_{11} & 0\\ A_{21} & S\end{bmatrix}\begin{bmatrix}I & A_{11}^{-1}A_{12}\\ 0 & I\end{bmatrix} = \begin{bmatrix}I & 0\\ A_{21}A_{11}^{-1} & I\end{bmatrix}\begin{bmatrix}A_{11} & A_{12}\\ 0 & S\end{bmatrix},
$$

where S is the exact Schur complement with respect to A_{11}, $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

The most successful preconditioners are block-diagonal, block upper- or lower-triangular and indefinite (sometimes of a factorized form as in [2], for example). The particular nonsymmetric case ($A_{12} = A_{21}^T$, $A_{11} \neq A_{11}^T$) has been studied mostly in the context of incompressible NS, see for instance [7], [9], [8], [13], [11], [17] and more references in [5]. Less experience is available for compressible NS (e.g. [16]).

To precondition $\mathcal{A}$ in (30), most often one of the following two-by-two block preconditioners have been used:

$$\mathcal{B}_0 = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \quad \mathcal{B}_1 = \begin{bmatrix} D_1 & 0 \\ B & -D_2 \end{bmatrix}, \tag{33}$$
$$\mathcal{B}_2 = \begin{bmatrix} D_1 & B^T \\ B & -R \end{bmatrix} = \begin{bmatrix} D_1 & 0 \\ B & N_1 \end{bmatrix} \begin{bmatrix} I_1 & D_1^{-1}B^T \\ 0 & -N_2 \end{bmatrix},$$

where D_1, D_2, R and N_1, N_2 are some approximation matrices to be constructed. From the factorizations (31) and (32) it is easily seen that to obtain a good preconditioner to $\mathcal{A}$, D_1, D_2 and R have to be approximations of M, $C + BD_1^{-1}B^T$ and C, respectively. In the extreme case when $D_1 = M$ and $D_2 = C + BM^{-1}B^T$, all eigenvalues of the preconditioned matrix $\mathcal{B}_1^{-1}\mathcal{A}$ are equal to 1 and a preconditioned iterative solution method with $\mathcal{B}_2$ will converge within one or two iterations. In the special case when $D_1 = M$ and $D_2 = (BB^T)(BMB^T)^{-1}(BB^T)$, under a certain additional condition, all eigenvalues of $\mathcal{B}_1^{-1}\mathcal{A}$ are equal to one and no Jordan block has order higher than two (cf. [7]). Since to choose $D_1 = M$ and to construct S explicitly is not feasible in practice, a suitable approximation of these has to be used. We note here that some of the proposed approximations of S are symmetric while the true Schur complement is not, and these have shown to work well only in certain cases.

An observation has been made (see [2] for symmetric problems) that the convergence of an iterative method using the block-preconditioners in (33) is more sensitive to the quality of the block D_1 than to that of the block D_2. If D_1 and D_2 are optimal order preconditioners to M and S with respect to discretization and problem parameters, then the block preconditioners will also be of optimal order. The question how to construct such optimal D_1 and in particular D_2 is answered in some special cases. Optimal approximations for S are derived for the Stokes, the generalized Stokes and the Oseen problems (see [5] and the references therein). For the linear elasticity we refer to [12] and [4]. As noted in [5], incomplete factorizations of M in the nonsymmetric case remain a feasible option. Multigrid preconditioners for the block M are also used, after a fair tuning of the multigrid ingredients (see for instance [19]). To use an indefinite preconditioner of a factorized form, as in [2], is appealing only if we can easily (cheaply) compute explicitly a good approximation for S.

Remark 3.1. Denote $M = K + N$, where K is symmetric positive definite and corresponds to the terms (A1) and (NA1), respectively, N is the nonsymmetric part in the stiffness matrix, due to the first order terms (B)+(C) and (NB), respectively. In NS, the analysis of the corresponding rate of convergence utilizes the fact that N is skew-symmetric. We note that in our case this does not hold. However, N is 'component-wise' skew-symmetric. Namely, in separate displacement ordering (when

the displacements in the 'x'-direction are ordered first), the following holds:

$$N = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}, \quad N^T = \begin{bmatrix} \mathcal{D}_1 - N_{11} & -N_{21} \\ -N_{12} & \mathcal{D}_2 - N_{22} \end{bmatrix},$$

where $\mathcal{D}_i$, $i = 1, 2$, are diagonal.

4. Numerical experiments

In the numerical tests below we apply the preconditioners $\mathcal{B}_1$ and $\mathcal{B}_2$ to the following realistic benchmark problem, used for comparison with a commercial finite element package.

Problem 4.1. A 2D flat Earth model, which is symmetric with respect to $x = 0$, is subjected to a Heaviside load of a 1,000 km wide and 2 km thick ice sheet. The size of the domain is 10,000 km width and 4,000 km depth, and the boundary conditions are the homogenous Dirichlet conditions on the boundary $y = -4{,}000$ km and the symmetry conditions on the boundary $y = 0$, and the homogenous Neumann conditions on the boundary $x = 10{,}000$ km and on the boundary segment $y = 0$, $x > 1{,}000$ km. The Young modulus of the solid is 400 GPa, the Poisson ratio is 0.5 (the material is incompressible), and its density is $3{,}000\,\mathrm{kg\,m^{-3}}$. The density of the ice is $981\,\mathrm{kg\,m^{-3}}$.

For this problem, the corresponding bilinear form $a(\mathbf{u}, \mathbf{v})$ in (5) is coercive, thus the discrete solution exists and is unique.

The domain is discretized with quasi-regular quadrilateral finite elements and bilinear basis functions. For the finite element discretization we use the C++ package `deal.II` [6], and for the numerical linear algebra the package `PETSc` [24]. The tests were performed on a Sun Ultra-Sparc III 900 MHz processor running under Sun Solaris 9.

In the experiments we use GMRES as an iterative scheme, preconditioned by either $\mathcal{B}_1$ or $\mathcal{B}_2$ in (33). The iterations are terminated when the residual norm is decreased by six orders of magnitude compared to the initial residual.

Two approximations for the (negative) Schur complement matrix S are tested, one symmetric and one nonsymmetric. The symmetric approximation (S_m) is chosen as $S_m = C + M_p$, where M_p is the pressure mass matrix. To form a nonsymmetric approximation for S we assemble a matrix S_a, obtained from the exact Schur complement of the local element stiffness matrices. This approach was first used in [15] to construct coarse level matrices in the context of an algebraic multilevel preconditioner. The construction is computationally cheap and numerical tests show that it produces a relatively good approximation for S.

D_1 and D_2 are formed as incomplete LU factorizations of M and S_i, $i = m, a$, employing the `ILUT`-routine from `PETSc`. In this routine, the memory consumption of the L- and U-factors is controlled via a limit n on the number of fill-in elements and a column pivot tolerance, beside the drop tolerance τ. In these experiments, we have chosen n related to the size of D_1 and D_2, and the number of fill-in elements was 1.8 % of the number of columns/rows in M for D_1 and 0.9 % of the number of columns/rows in S for D_2. The column pivot tolerance was chosen as 0.01. In the tables of the next subsection, ILUT(A, τ) denotes `ILUT`-factorization of the matrix A with drop-tolerance τ, and $\tau = 0$ corresponds to a complete factorization.

4.1. Iteration counts

The dependence of the iteration counts on the choice of the drop tolerances q for D_1 and t for D_2 is presented in Tab. 1. Here S is replaced by S_m. The figures correspond to one solve in the defect-correction algorithm for (29).

	$D_1 = \text{ILUT}(M, q), \quad D_2 = \text{ILUT}(S_a, t)$					
	$t = 0.01$		$t = 0.001$		$t = 0.0001$	
	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_1$	$\mathcal{B}_2$
	$N = 8019$					
$q = 0.01$	49	38	49	38	49	38
$q = 0.001$	21	16	21	16	21	16
$q = 0.0001$	18	14	18	14	18	14
	$N = 31395$					
$q = 0.01$	206	156	208	153	207	153
$q = 0.001$	38	28	38	28	38	28
$q = 0.0001$	23	18	23	18	23	18
	$N = 124227$					
$q = 0.01$	857	587	856	767	863	823
$q = 0.001$	131	118	131	118	131	118
$q = 0.0001$	29	24	29	24	29	24

Table 1. Iteration counts for GMRES preconditioned with $\mathcal{B}_1$ and $\mathcal{B}_2$.

It is evident from Tab. 1 that the quality of the factorization of the approximated Schur complement does not affect the iteration count significantly (the numbers of iterations required for $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively, are nearly constant for different t). The choice of the accuracy of the factorization for M (the size of q) is more crucial, and this observation is similar to that in [2].

The observed growth in the iteration counts with increasing problem size is due to the choice of D_1 and D_2 as incomplete factorizations of M and S. The increase of the number of iterations can be stabilised with a better choice of the preconditioner for M and S (of multilevel or (algebraic) multigrid type).

	$D_1 = \mathrm{ILUT}(M,q), \quad D_2 = \mathrm{ILUT}(S,t)$											
	$S = S_m$						$S = S_a$					
	$t = 0.001$		$t = 0.0001$		$t = 0$		$t = 0.001$		$t = 0.0001$		$t = 0$	
	$N = 8019$											
$q = 0.001$	21	16	21	16	21	16	21	16	21	16	21	16
$q = 0.0001$	18	14	18	14	18	14	18	14	18	14	18	14
$q = 0$	9	9	9	9	9	9	9	9	9	9	9	9
	$N_e = 31395$											
$q = 0.001$	38	27	38	27	38	27	38	27	38	27	38	27
$q = 0.0001$	23	18	23	18	23	18	23	18	23	18	23	18
$q = 0$	9	9	9	9	9	9	9	9	9	9	9	9

Table 2. Iteration counts for Problem 4.1 with different Schur complement approximations.

Tab. 2 shows iteration counts for $\mathcal{B}_1$ and $\mathcal{B}_2$ with different Schur complement approximations. The iteration counts do not change for the different approximations of S (symmetric S_m and nonsymmetric S_a), which is due to the fact that that the nonsymmetric part of M is relatively small. The properties of the approximations of S are also illustrated in Fig. 1, where the real and imaginary parts of the spectra of the exact Schur complement S, S_a and S_m are shown (for a small matrix of size 567). The imaginary part of the spectrum of S_m is zero and is not plotted for clarity.

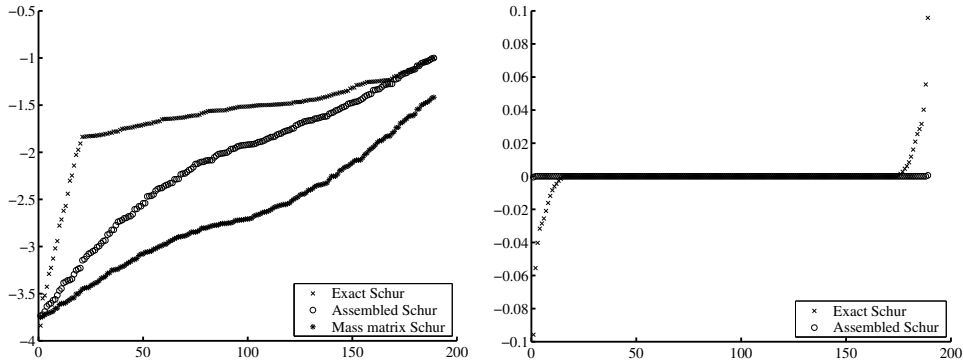

Figure 1. Real and imaginary parts of the spectrum of S, S_m and S_a.

Tab. 3 illustrates the behaviour of the methods for different values of the Poisson number ν. The iterations grow slowly with increasing ν. The slight discrepancy between Tab. 1 and Tab. 3 for $\nu = 0.5$ and $q = 0.001$, 0.0001, and $t = 0.001$, 0.0001 is explained by a difference in the parameters in (2). For the results in Tab. 1, $\mathbf{c} = 0$ since the material is incompressible and the term (C) can be neglected in the continuous equations. In the experiments presented in Tab. 3, $\mathbf{c} = [0 \quad \varrho^{(0)} g^{(0)}]^T$.

	$D_1 = \text{ILUT}(M,q)$, $D_2 = \text{ILUT}(S_a,t)$									
	$t = 0.001$		$t = 0.0001$		$t = 0$		$t = 0.001$		$t = 0.0001$		$t = 0$	
	$\nu = 0.2$						$\nu = 0.3$					
	$N = 8019$											
$q = 0.001$	15	13	15	13	15	13	17	14	17	14	17	14
$q = 0.0001$	13	12	13	12	13	12	15	12	15	12	15	12
$q = 0$	6	5	6	5	6	5	7	6	7	6	7	6
	$N = 31395$											
$q = 0.001$	23	21	23	22	23	22	26	23	26	23	26	23
$q = 0.0001$	15	14	15	14	15	14	18	15	18	15	18	15
$q = 0$	6	5	6	5	6	5	7	7	7	6	7	6
	$\nu = 0.4$						$\nu = 0.5$					
	$N = 8019$											
$q = 0.001$	19	15	19	15	19	15	20	16	20	16	20	16
$q = 0.0001$	16	13	16	13	16	13	18	15	18	15	18	15
$q = 0$	8	8	8	8	8	8	10	9	10	9	10	9
	$N = 31395$											
$q = 0.001$	29	25	29	25	29	25	38	27	38	27	38	27
$q = 0.0001$	20	16	20	16	20	16	22	18	22	18	22	18
$q = 0$	8	8	8	8	8	8	10	9	10	9	10	9

Table 3. Iteration counts for Problem 4.1 with different values of ν.

4.2. CPU time comparison

Tab. 4 is a comparison between the performance of our code and that of a commercial FEM package, solving Problem 4.1 with identical geometry, mesh and physical parameters. The only slight difference is in the boundary conditions. On the far boundaries ($x = 10{,}000$ and $y = -4{,}000$) the package imposes bilinear, infinite elements while we use standard homogeneous Neumann and Dirichlet conditions. The package was run on two different systems, an AMD Athlon 2.5 GHz processor, and a dual Itanium 1.5 GHz processor.

	PETSc implementation	Commercial FE package	
Size	Sun Ultra-Sparc III 900 MHz	Itanium 1.5 GHz	AMD Athlon 2.5 GHz
12675	6	20	22
49923	53	77	82

Table 4. Time comparison with a commercial FEM package solving Problem 4.1 (sec.).

It is important to note that the commercial package implements the problem (2) in the Lamé-Navier formulation, instead of the Stokes formulation we use. Due to the latter, the size of the problem solved by the commercial package is (two-thirds) smaller than the size in the saddle-point formulation. Nevertheless, the benefit from

using an appropriately preconditioned iterative method instead of a direct solver is clearly seen from the timing results.

5. Conclusions

We present an analysis of a stabilized saddle-point formulation of the discretized isostatic model of glacial rebound and numerical simulations using a preconditioned iterative solution method. It is shown that the bilinear form $a(\cdot,\cdot)$, associated with the displacements, is not coercive in general. The numerical experiments illustrate that the known block-triangular and indefinite preconditioners exhibit a robust behaviour, provided that good approximations for the pivot block M and the (negative) Schur complement matrix S can be computed. A cheap nonsymmetric approximation for S has been constructed. The question how to precondition M efficiently, utilizing the properties of the particular problem, is not fully answered. However, efficient preconditioners both for the elasticity and the scalar advection-diffusion problems have been devised (including multilevel and multigrid techniques) and using such will provide us with an optimal order (problem parameters- and mesh-independent) method for the class of problems considered in this work.

Acknowledgement. The authors thank the anonymous referee for the valuable comments.

References

[1] *O. Axelsson, V. A. Barker*: Finite Element Solution of Boundary Value Problems. Theory and Computation. Academic Press, Orlando, 1984.
[2] *O. Axelsson, M. Neytcheva*: Preconditioning methods for constrained optimization problems. Numer. Linear Algebra Appl. *10* (2003), 3–31.
[3] *O. Axelsson, V. A. Barker, M. Neytcheva, and B. Polman*: Solving the Stokes problem on a massively parallel computer. Math. Model. Anal. *6* (2001), 7–27.
[4] *O. Axelsson, A. Padiy*: On a robust and scalable linear elasticity solver based on a saddle point formulation. Int. J. Numer. Methods Eng. *44* (1999), 801–818.
[5] *M. Benzi, G. Golub, and J. Liesen*: Numerical solution of saddle point problems. Acta Numer. To appear.
[6] *W. Bangerth, R. Hartmann, and G. Kanschat*: `deal.II` Differential Equations Analysis Library, Technical Reference, IWR, http://www.dealii.org.
[7] *H. C. Elman*: Preconditioning for the steady-state Navier-Stokes equations with low viscosity. SIAM J. Sci. Comput. *20* (1999), 1299–1316.
[8] *H. C. Elman, D. Loghin, A. J. Wathen*: Preconditioning techniques for Newton's method for the incompressible Navier-Stokes equations. BIT *43* (2003), 961–974.
[9] *H. C. Elman, D. Silvester*: Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations. SIAM J. Sci. Comput. *17* (1996), 33–46.
[10] *H. Elman, D. Silvester, A. J. Wathen*: Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations. Numer. Math. *90* (2002), 665–688.

[11] *I. Ipsen*: A note on preconditioning nonsymmetric matrices. SIAM J. Sci. Comput. *23* (2001), 1050–1051.
[12] *A. Klawonn*: An optimal preconditioners for a class of saddle point problems with a penalty term. SIAM J. Sci. Comput *19* (1998), 540–552.
[13] *A. Klawonn, G. Starke*: Block triangular preconditioners for nonsymmetric saddle point problems: Field-of-values analysis. Numer. Math. *81* (1999), 577–594.
[14] *V. Klemann, P. Wu, and D. Wolf*: Compressible viscoelasticity: stability of solutions for homogeneous plane-Earth models. Geophys. J. *153* (2003), 569–585.
[15] *J. K. Kraus*: Algebraic multilevel preconditioning of finite element matrices using local Schur complements. Submitted.
[16] *B. Liu, R. B. Kellogg*: Discontinuous solutions of linearized steady state viscous compressible flows. J. Math. Anal. Appl. *180* (1993), 469–497.
[17] *D. Loghin, A. J. Wathen*: Analysis of preconditioners for saddle-point problems. SIAM J. Sci. Comput. *25* (2004), 2029–2049.
[18] *J. Nedoma*: Numerical Modelling in Applied Geodynamics. John Wiley & Sons, New York, 2000.
[19] *A. A. Ramage*: A multigrid preconditioner for stabilised discretisations of advection-diffusion problems. J. Comput. Appl. Math. *110* (1999), 187–203.
[20] *Y. Saad*: SPARSKIT: A basic tool-kit for sparse matrix computations. Technical Documentation, http://www-users.cs.umn.edu/∼saad/software/SPARSKIT/sparskit.html.
[21] *S. Shaw, M. K. Warby, J. R. Whiteman, C. Dawson, and M. F. Wheeler*: Numerical techniques for the treatment of quasistatic viscoelastic stress problems in linear isotropic solids. Comput. Methods Appl. Mech. Eng. *118* (1994), 211–237.
[22] *P. Wu*: Viscoelastic versus viscous deformation and the advection of pre-stress. Internat. J. Geophys. *108* (1992), 136–142.
[23] *P. Wu*: Using commercial finite element packages for the study of earth deformations, sea levels and the state of stress. Internat. J. Geophys. *158* (2004), 401–408.
[24] Portable, Extensible Toolkit for Scientific computation (PETSc) suite. Mathematics and Computer Science Division, Argonne Natinal Laboratory, www-unix.mcs.anl.gov/petsc/.

Authors' address: *E. Bängtsson, M. Neytcheva*, Department of Information Technology, Uppsala University, Box 337, SE-751 05 Uppsala, Sweden, e-mails: `erikba@it.uu.se`, `Maya.Neytcheva@it.uu.se`.