An Optimal Iterative Solver for Symmetric Indefinite Systems stemming from Mixed Approximation

DAVID J. SILVESTER
University of Manchester
and
VALERIA SIMONCINI
Università di Bologna

We discuss the design and implementation of a suite of functions for solving symmetric indefinite linear systems associated with mixed approximation of systems of PDEs. The novel feature of our iterative solver is the incorporation of error control in the natural "energy" norm in combination with an a posteriori estimator for the PDE approximation error. This leads to a robust and optimally efficient stopping criterion: the iteration is terminated as soon as the algebraic error is insignificant compared to the approximation error. We describe a "proof of concept" MATLAB implementation of this algorithm, which we call EST_MINRES, and we illustrate its effectiveness when integrated into the Incompressible Flow Iterative Solution Software (IFISS) package (cf. ACM Transactions on Mathematical Software 33, Article 14, 2007).

Categories and Subject Descriptors: G.1.3 [**Numerical Linear Algebra**]: Linear systems (direct and iterative methods); G.1.8 [**Partial Differential Equations**]: Elliptic equations; Finite element methods; Iterative solution techniques; Multigrid and multilevel methods; G.4 [**Mathematical Software**]: MATLAB ; J.2 [**Computer Applications**]: Physical Sciences and Engineering

General Terms: Algorithms, Design

Additional Key Words and Phrases: Finite elements, incompressible flow, iterative solvers, stopping criteria, EST_MINRES

1. INTRODUCTION

This paper describes a novel algorithm for solving symmetric linear systems associated with mixed approximation of systems of PDEs. Our approach has three key ingredients: first, a block preconditioning strategy that engenders convergence with a rate that is *independent* of the problem parameters; second, an effective adaptation of the MINRES algorithm of Paige and Saunders [1975] which enables convergence in a computable monotonically decreasing norm that is equivalent to

Authors' addresses: D.J. Silvester, School of Mathematics, University of Manchester, Manchester M13 9PL, UK; email: d.silvester@manchester.ac.uk. V.Simoncini, Dipartimento di Matematica, Università di Bolgna, I-40127 Bologna, Italy; email: valeria@dm.unibo.it

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.
© 20YY ACM 0098-3500/20YY/1200-0001 $5.00

the natural norm for error estimation of the discrete solution; and third, the incorporation of a posteriori error estimation functionality which enables us to formulate a precise stopping criterion for the linear solver so as to balance the algebraic error with the PDE approximation error. In this sense our iterative method is *optimal*.

To put this work into context, significant efforts have been put into the derivation of estimates for the error norm for Krylov subspace solvers applied to symmetric and positive definite matrices, see, e.g., Strakoš and Tichý [2002], Meurant [2005], Meurant and Strakoš [2006], Golub and Meurant [1997], which have in turn led to the development of stopping criteria in the Conjugate Gradient method (CG) based on the algebraic error norm. We note, however, that in these works there is the necessity of estimating the *unobservable* quantity—the energy error—that is actually minimized by the CG method. In the case of MINRES, the quantity that is minimized is the residual Euclidean norm—which is readily available. Thus from an algebraic point of view no further estimates need to be determined. As we will see later, it is only when we look into the origin of the algebraic system that the relevance of a "natural" norm becomes apparent. Further motivation for this philosophy can be found in the work of Wathen [2007] and, specifically for the class of problems considered here, in the work of Mardal and Winther [2010]. The prominent role of the discretization error in the determination of the algebraic stopping tolerance has also been recognised by other researchers; see, e.g., Arioli et al. [2005], Arioli and Loghin [2008], Jiránek et al. [2010], and Arioli [2010]. Our paper extends previous work in that the norm-equivalence constants that arise in the stopping test are dynamically estimated in our algorithm. In contrast, constants need to be estimated a priori in the approach of Arioli and Loghin [2008].

The remainder of the paper is organised as follows. Section 2 sets up the governing PDE framework. Two representative saddle point problems are identified, and approximation and error estimation are discussed. The block diagonal preconditioning framework that is at the heart of the solver methodology is reviewed in Section 3. The design of the stopping criterion is described in Section 4, and practical implementation issues are discussed in Section 5.

2. SADDLE POINT PROBLEMS

Our aim is to design an *optimal solver* for discretized saddle point problems. These are *symmetric indefinite* linear algebra systems

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix}, \tag{1}$$

that are associated with the finite-dimensional approximation of the following variational problem: find $(u, p) \in V \times Q$ such that

$$a(u, v) + b(v, p) = f(v) \qquad \forall v \in V, \tag{2}$$

$$b(u, q) = g(q) \qquad \forall q \in Q. \tag{3}$$

Here, V and Q represent Hilbert spaces; $a : V \times V \to \mathbb{R}$ is a symmetric bounded bilinear form, $b : V \times Q \to \mathbb{R}$ is also a bounded bilinear form and $f : V \to \mathbb{R}$, $g : Q \to \mathbb{R}$ are linear functionals. Note that the fact that the two spaces V and Q are approximated independently leads to the nomenclature *mixed approximation*.

Energy arguments lead to a natural norm for measuring the quality of approximation for functions in the space $V \times Q$,

$$\|(u,p)\|_{V\times Q} = \|u\|_V + \|p\|_Q \,.$$

This measure is referred to as the *energy norm* in the sequel.

In section 3 we will follow the philosophy of Mardal and Winther [2010] for constructing a generic *preconditioner* for the saddle point system (1). To this end, we define the dual spaces V^* and Q^* respectively, and introduce the duality pairing $\langle\cdot,\cdot\rangle$. Then, if we associate the bilinear forms a and b with operators $\mathcal{A}: V \to V^*$ and $\mathcal{B}: V \to Q^*$ so that

$$\langle \mathcal{A}u, v\rangle = a(u,v) = \langle u, \mathcal{A}v\rangle \quad \text{and} \quad \langle \mathcal{B}u, q\rangle = b(u,q) = \langle u, \mathcal{B}^* q\rangle,$$

the problem (2)–(3) can be expressed in the "saddle point" form

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^* \\ \mathcal{B} & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}. \tag{4}$$

A suitable preconditioner for (1) can then be defined by first identifying a preconditioner for the associated continuous problem, and second, by ensuring that the discretization of the continuous problem is *stable*. Details are given later.

Systems of the form (2)–(3) arise when modelling elliptic or parabolic PDE problems that are associated with constraints. Examples include linear elasticity (Navier-Lamé equations), steady fluid flow (Stokes equations) and electromagnetism (Maxwell equations). See Brezzi and Fortin [1991] for a thorough overview of the approximation aspects and Benzi et al. [2005, pp. 9–20] for a detailed discussion of the properties of the discretized system (1). The two PDE problems considered below are naturally self-adjoint, and so give rise to symmetric linear systems if appropriately discretized. More generally, saddle point problems can also be found in optimal control when *minimizing* a cost functional with a non-selfadjoint PDE problem (e.g., the Navier-Stokes equations) as a constraint. This is a frontier topic of numerical analysis and we hope that our "optimal solver" strategy will help stimulate research in this rapidly developing field.

2.1 The Stokes equations

The Stokes equations, given by

$$-\nabla^2 \vec{u} + \nabla p = \vec{0}, \tag{5}$$

$$\nabla \cdot \vec{u} = 0, \tag{6}$$

on some domain $\Omega \subset \mathbb{R}^n$, together with boundary conditions

$$\vec{u} = \vec{w} \text{ on } \partial\Omega_D, \quad \frac{\partial \vec{u}}{\partial n} - \vec{n}p = \vec{0} \text{ on } \partial\Omega_N, \tag{7}$$

are a fundamental model for steady-state viscous flow. The variable $\vec{u}$ is a vector-valued function representing the velocity of the fluid, and the scalar function p represents the pressure. The equation (5) represents conservation of the momentum of the fluid and equation (6) enforces conservation of mass. The crucial modelling assumption is that the flow is low speed, so that convection effects can be neglected.

Introducing vector-valued velocity functions $\vec{v} \in V := (H_0^1(\Omega))^d$ and scalar pressure functions $q \in Q := L^2(\Omega)$ a variational formulation of (5)–(7) is given by

$$(\nabla \vec{u}, \nabla \vec{v}) - (p, \nabla \cdot \vec{v}) = f(\vec{v}) \qquad \forall \vec{v} \in V, \tag{8}$$

$$(q, \nabla \cdot \vec{v}) = g(q) \qquad \forall q \in Q, \tag{9}$$

where f, g incorporate the nonhomogeneous boundary data $\vec{w}$ on $\partial\Omega_D$, and $(\cdot,\cdot)$ represents the standard (either scalar or vector-valued) L^2–inner product.[1]

To get to the representation (4), we identify the dual spaces $V^* := (H^{-1}(\Omega))^d$ and $Q^* := L^2(\Omega)$ respectively, and define operators $\mathcal{A} : V \to V^*$ and $\mathcal{B} : V \to Q^*$ so that

$$\langle \mathcal{A}\vec{u}, \vec{v}\rangle = (\nabla \vec{u}, \nabla \vec{v}) \quad \text{and} \quad \langle \mathcal{B}\vec{u}, q\rangle = -(\nabla \cdot \vec{u}, q).$$

With these definitions the problem (5)–(7) can be expressed in the form (4). Moreover the coefficient matrix in (4),

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^* \\ \mathcal{B} & 0 \end{pmatrix} = \begin{pmatrix} -\nabla^2 & -\nabla \\ \nabla\cdot & 0 \end{pmatrix}$$

represents a mapping from $V \times Q$ onto $V^* \times Q^*$.

The discrete version of the Stokes problem (8)–(9) is immediate: given approximation spaces $V_h \subset V$ and $Q_h \subset Q$ our aim is to compute $(\vec{u}_h, p_h) \in V_h \times Q_h$ satisfying

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^* \\ \mathcal{B} & 0 \end{pmatrix} \begin{pmatrix} \vec{u}_h \\ p_h \end{pmatrix} = \begin{pmatrix} f_h \\ g_h \end{pmatrix}. \tag{10}$$

Mixed finite element approximation entails defining appropriate bases for the velocity and pressure finite element spaces V_h and Q_h respectively, and constructing the associated linear algebra system (1) for the coefficients in the basis expansion. This system will have dimension n_u+n_p where n_u and n_p are the numbers of velocity and pressure basis functions, respectively. We also note that in the linear algebra system (1) the matrix A is a $d \times d$ block diagonal matrix with scalar Laplacian matrices defining the diagonal blocks; the matrix B is an $n_p \times n_u$ rectangular matrix that represents a discrete (negative–)divergence operator.

If the discretized problem (10), or equivalently (1), is to properly represent a continuous Stokes problem, then the component approximation spaces V_h and Q_h need to be *compatible*. For example, if there are more pressure basis functions than velocity basis functions then the associated linear algebra problem is necessarily singular! It is well known that once the velocity approximation is fixed, then the validity of a corresponding set of pressure basis functions is determined by whether or not an *inf-sup* stability condition can be established.[2] Low order approximation schemes are generally unstable. The stable methods that are built into IFISS are $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ approximation which has local degrees of freedom shown in Fig. 1 and the *Taylor-Hood* $\boldsymbol{Q}_2$–$\boldsymbol{Q}_1$ approximation which is shown in Fig. 2. The $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ approximation combines continuous biquadratic approximation ($\boldsymbol{Q}_2$) for velocity

[1] In our notation the space $H_0^1(\Omega)$ consists of functions whose trace is zero on $\partial\Omega_D$. We implicitly assume that $\int_{\partial\Omega_N} ds > 0$ so that p satisfying (5)–(7) is uniquely defined.

[2] For an accessible discussion of inf-sup stability see Elman et al. [2005, Section 5.3.1].

together with a *discontinuous* linear ($\boldsymbol{P}_{-1}$) pressure, and is widely regarded as being the most cost-effective discretization approach for solving (Navier–)Stokes problems in $\mathbb{R}^2$. The $\boldsymbol{Q}_2$–$\boldsymbol{Q}_1$ method is stable but is less accurate than $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$: local mass conservation is compromised because of the C^0 pressure approximation.

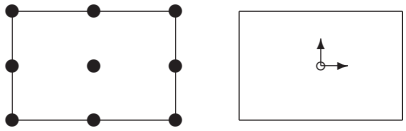

Fig. 1. $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ element (• velocity node; ∘ pressure; $\overset{\uparrow}{\rightarrow}$ pressure derivative).

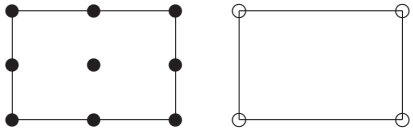

Fig. 2. $\boldsymbol{Q}_2$–$\boldsymbol{Q}_1$ element (• two velocity components; ∘ pressure).

An important ingredient in our optimal solution algorithm is the need to compute an *a posteriori* error estimate for discrete solutions in the energy norm. That is, given a candidate solution $(\vec{u}_h, p_h) \in V_h \times Q_h$ (not necessarily the Galerkin solution satisfying (10)), we want to compute an estimate η which is equivalent to the exact error in the sense that

$$c\,\eta \leq \|\nabla(\vec{u} - \vec{u}_h)\| + \|p - p_h\| \leq C\,\eta, \tag{11}$$

with $C/c \sim O(1)$. There are a number of possible approaches. The specific strategy that is built into the IFISS package is based on solving local Poisson problems for each velocity component over a suitably enlarged approximation space $\vec{\mathcal{Q}}_T$. The approach is discussed in detail in Elman et al. [2005, Section 5.4.2]. Having computed element interior residuals $\vec{R}_T := \{\nabla^2 \vec{u}_h - \nabla p_h\}|_T$ and $R_T := \{\nabla \cdot \vec{u}_h\}|_T$, and edge residuals (equidistributed stress jumps) $\vec{R}_E := \frac{1}{2}[\![\nabla \vec{u}_h - p_h \vec{\mathbf{I}}]\!]$, a velocity error estimate $\vec{e}_T \in \vec{\mathcal{Q}}_T$ is computed satisfying

$$(\nabla \vec{e}_T, \nabla \vec{v})_T = (\vec{R}_T, \vec{v})_T - \sum_{E \in \mathcal{E}(T)} \langle \vec{R}_E, \vec{v} \rangle_E \qquad \forall\, \vec{v} \in \vec{\mathcal{Q}}_T, \tag{12}$$

for every element in the grid. A local error estimator is then given by the combination of the 'energy norm' of the velocity error and the L^2 norm of the element divergence error, that is,

$$\eta_T^2 := \|\nabla \vec{e}_T\|_T^2 + \|R_T\|_T^2\,. \tag{13}$$

The global error estimator is $\eta := \left(\sum_{T\in\mathcal{T}_h} \eta_T^2\right)^{1/2}$. This style of error estimation was introduced for the lowest order stabilized $\boldsymbol{P}_1$–$\boldsymbol{P}_0$ approximation by Kay and Silvester [1999] and is a refinement of the methodology introduced by Ainsworth and Oden [1997]. Regarding the effectiveness of the estimator in the sense of (11), numerical results obtained in Liao and Silvester [2010] using $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ approximation suggest that the equivalence is tight if $(\vec{u}_h, p_h)$ also solves (10). Note that Galerkin orthogonality is not used in the analysis in Liao and Silvester [2010]—this means that the bound (11) is valid for any function in the approximation space. We will come back to this issue in Section 4.2.

2.2 Potential Flow equations

Our second example is an idealized model of incompressible flow:

$$-\vec{u} + \nabla p = \vec{0}, \tag{14}$$

$$\nabla \cdot \vec{u} = 0, \tag{15}$$

on some domain $\Omega \subset \mathbb{R}^n$, together with boundary conditions

$$\vec{u} \cdot \vec{n} = 0 \text{ on } \partial\Omega_N, \quad p = s \text{ on } \partial\Omega_D. \tag{16}$$

The additional modelling assumption in this case is that the flow is irrotational $\nabla \times \vec{u} = \vec{0}$, so that viscous effects can be neglected. Introducing vector-valued velocity functions $\vec{v} \in V := H_0(\mathrm{div}, \Omega)$ and scalar pressure functions $q \in Q := L^2(\Omega)$ a standard variational formulation of (14)–(16) is given by

$$(\vec{u}, \vec{v}) + (p, \nabla \cdot \vec{v}) = f(\vec{v}) \qquad \forall \vec{v} \in V, \tag{17}$$

$$(q, \nabla \cdot \vec{v}) = 0 \qquad \forall q \in Q, \tag{18}$$

where f incorporates the nonhomogeneous boundary data on $\partial\Omega_D$. The space $H_0(\mathrm{div}, \Omega)$ consists of all vector fields in $(L^2(\Omega))^d$ with divergence in $L^2(\Omega)$ and which have vanishing normal component on $\partial\Omega_N$.

To get to the representation (4), we identify the dual spaces $V^* := H_0(\mathrm{div}, \Omega)^*$ and $Q^* := L^2(\Omega)$ respectively, and define operators $\mathcal{A} : V \to V^*$ and $\mathcal{B} : V \to Q^*$ so that

$$\langle \mathcal{A}\vec{u}, \vec{v}\rangle = (\vec{u}, \vec{v}) \quad \text{and} \quad \langle \mathcal{B}\vec{u}, q\rangle = (\nabla \cdot \vec{u}, q).$$

In this case $V \times Q$ is mapped onto $V^* \times Q^*$ by the matrix operator

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^* \\ \mathcal{B} & 0 \end{pmatrix} = \begin{pmatrix} I & -\nabla \\ \nabla\cdot & 0 \end{pmatrix}. \tag{19}$$

A different representation of the original problem can be obtained by integrating the pressure terms in (17)–(18) by parts. Appropriate solution spaces are then $\vec{v} \in V := (L^2(\Omega))^d$ and $Q := H_0^1(\Omega)$ and the variational formulation of (14)–(16) is given by

$$(\vec{u}, \vec{v}) - (\nabla p, \vec{v}) = f(\vec{v}) \qquad \forall \vec{v} \in V, \tag{20}$$

$$(\nabla q, \vec{v}) = 0 \qquad \forall q \in Q, \tag{21}$$

where f again incorporates the nonhomogeneous boundary data on $\partial\Omega_D$. In this case we can define operators $\mathcal{A} : V \to V^*$ and $\mathcal{B} : V \to Q^*$ so that

$$\langle \mathcal{A}\vec{u}, \vec{v}\rangle = (\vec{u}, \vec{v}) \quad \text{and} \quad \langle \mathcal{B}^* p, \vec{v}\rangle = -(\nabla p, \vec{v}) = \langle p, \mathcal{B}\vec{v}\rangle,$$

and we see that the alternatively defined space $V \times Q$ is also mapped onto its dual $V^* \times Q^*$ by the very same matrix operator as in (19).

As in the Stokes case, mixed finite element approximation entails defining appropriate bases for the velocity and pressure finite element spaces V_h and Q_h respectively, and then constructing the system (1) for the coefficients in the basis expansion. The simplest choice of basis functions which leads to a stable approximation is the *Raviart-Thomas* flux approximation (normal components of velocity defined on the edges of triangles or rectangles in $\mathbb{R}^2$) together with a piecewise constant pressure. The local degrees of freedom for a triangular $\boldsymbol{RT}_0$ element are shown in Fig. 3.

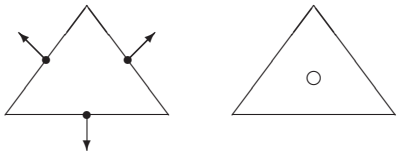

Fig. 3. $\boldsymbol{RT}_0$ element (• normal velocity component; ∘ pressure).

Using this mixed approximation the linear algebra system (1) will have dimension $n_u + n_p$, where n_u is the number of element edges (excluding $\partial\Omega_N$) and n_p is the number of elements. Morever the matrix A is just a Grammian matrix (approximating the identity) and the matrix B is an $n_p \times n_u$ rectangular matrix that represents a discrete (positive–)divergence operator.

3. PRECONDITIONING FRAMEWORK

Working in the framework of Mardal and Winther [2010] generic preconditioners for (1) naturally arise if we have a suitable preconditioner for the associated continuous problem (4). Specifically, having identified a suitable function space setting, then a *canonical* preconditioner for the saddle point problem (4) is the 2×2 block diagonal matrix operator that maps the dual space $V^* \times Q^*$ back into the original space $V \times Q$. Running through our examples:

—*Stokes preconditioning.* In this case we require a block matrix operator taking $(H^{-1}(\Omega))^d$ to $(H_0^1(\Omega))^d$ and a scalar operator that takes $L^2(\Omega)$ to itself. The canonical preconditioning operator is thus

$$\mathcal{M} = \begin{pmatrix} (-\nabla^2)^{-1} & 0 \\ 0 & I^{-1} \end{pmatrix}. \tag{22}$$

This approach was originally suggested by Rusten and Winther [1992] and developed by Silvester and Wathen [1994]. An alternative justification is given in Elman et al. [2005, Section 6.2].

—*Potential flow preconditioning.* In this case the matrix coefficient operator maps $H_0(\mathrm{div}, \Omega) \times L^2(\Omega)$ onto its dual space. The canonical preconditioning operator turns out to be

$$\mathcal{M} = \begin{pmatrix} (I - \mathrm{grad}\,\mathrm{div})^{-1} & 0 \\ 0 & I^{-1} \end{pmatrix}. \tag{23}$$

This approach was introduced by Arnold et al. [1997].

—*Alternative potential flow preconditioning.* As discussed above, the matrix coefficient operator also maps $V := (L^2(\Omega))^d$ and $Q := H_0^1(\Omega)$ onto their dual spaces. The generic preconditioning operator is thus

$$\mathcal{M} = \begin{pmatrix} I^{-1} & 0 \\ 0 & (-\nabla^2)^{-1} \end{pmatrix}. \tag{24}$$

For a practical solution algorithm, if the linear algebra system (1) to be solved has dimension n, then the action of a discrete version of $\mathcal{M}$ needs to be effected in $O(n)$ work. That is the component blocks appearing in (22)–(24) must be replaced by cost effective operators with equivalent mapping properties:

—*Mass matrix preconditioning (I^{-1} operator)*;

—*Negative Laplacian preconditioning ($(-\nabla^2)^{-1}$ operator)*;

—*H(div) preconditioning ($(I - \text{grad div})^{-1}$ operator).*

Implementation of the first two of these operators is (essentially) independent of the spatial discretization. We discuss these components in more detail below. The H(div) operator is more difficult to implement in a "black-box" setting since standard (elliptic–) multigrid algorithms are not applicable. Typically, special smoothers are needed; possibly in combination with a geometric grid hierarchy. Efficient algorithms are given by Arnold et al. [2000] and Hiptmair and Xu [2007].

3.1 Mass matrix preconditioning

Using a discontinuous finite element approximation space, e.g. $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ approximation for Stokes flow, so that $Q_h = \text{span}\{\phi_k\}_{k=1}^{n_p}$ the associated Grammian matrix $\mathbb{I}_{ij} := (\phi_j, \phi_i)$ is *diagonal*. In such cases the action of $\mathbb{I}^{-1}$ can be explicitly computed with n_p division operations. If the approximation is C^0 however, e.g. using $\boldsymbol{Q}_2$–$\boldsymbol{Q}_1$ approximation for Stokes flow (or $\boldsymbol{RT}_0$ velocity approximation) then the best strategy is to perform a fixed (and small) number of Jacobi iterations with Chebyshev acceleration, see Wathen and Rees [2009].[3] The quality of the approximation is determined by `its`, the number of Chebyshev iterations performed. This is shown by the spectral bounds in Table I. Here θ and Θ are the extremal eigenvalues satisfying

$$\theta \leq \frac{\mathbf{p}^T \mathbb{I}\, \mathbf{p}}{\mathbf{p}^T \mathbb{I}_*\, \mathbf{p}} \leq \Theta, \tag{25}$$

where $\mathbb{I}$ is the $\boldsymbol{Q}_1$ mass matrix and $\mathbb{I}_*$ is the inverse of the matrix operator computed by applying our Chebyshev semi-iteration successively to the canonical vectors $\mathbf{e}_1, \ldots, \mathbf{e}_{n_p}$. In all cases the domain Ω is $(-1,1)\times(-1,1)$. The 64×64 stretched grid is refined next to the sides of the square and the element aspect ratios vary from 1:1 at the corners to 18:1 at the mid-sides. Looking at these results, it is evident that the action of $\mathbb{I}^{-1}$ is efficiently computed by our preconditioner, independently of the grid resolution and the stretching of the grid.

[3]Our implementation of this algorithm is the function `m_masscheb.m` in release 3.1 of IFISS.

its	5		10		20	
grid	θ	Θ	θ	Θ	θ	Θ
uniform 16×16	0.883	1.234	0.986	1.003	1.000	1.000
uniform 64×64	0.883	1.234	0.986	1.003	1.000	1.000
stretched 64×64	0.883	1.234	0.986	1.003	1.000	1.000

Table I. Spectral bounds for $\boldsymbol{Q}_1$ mass matrix preconditioner

Similarly encouraging results for $\boldsymbol{RT}_0$ approximation are given in Table II. Here θ and Θ are the extremal eigenvalues satisfying (25) where $\mathbb{I}$ is the $\boldsymbol{RT}_0$ velocity approximation mass matrix. Results are given for two nonuniform triangular meshes. These are associated with discretization of problems $P1$ and $P3$ from the PIFISS toolbox, and are described in Silvester and Powell [2007].

its	3		5		10	
# elments	θ	Θ	θ	Θ	θ	Θ
585 triangles	0.852	1.118	0.971	1.034	0.999	1.001
980 triangles	0.852	1.118	0.971	1.034	0.999	1.000

Table II. Spectral bounds for $\boldsymbol{RT}_0$ mass matrix preconditioner

3.2 Negative Laplacian preconditioning

A standard "black-box" approach is to approximate the inverse Laplacian by a fixed number of algebraic multigrid (AMG) V–cycles.[4] The quality of the approximation is typically determined by `nv`, the number of V-cycles performed. This is shown by the spectral bounds in Table III. Here λ and Λ are the extremal eigenvalues satisfying the Rayleigh quotient bounds

$$\lambda \leq \frac{\mathbf{u}^T \mathbb{A} \mathbf{u}}{\mathbf{u}^T \mathbb{A}_* \mathbf{u}} \leq \Lambda, \tag{26}$$

where $\mathbb{A}$ is the scalar $\boldsymbol{Q}_2$ stiffness matrix and $\mathbb{A}_*$ is the AMG approximation to the inverse of the matrix operator. For comparison, the $\boldsymbol{Q}_2$ grids have the same number of degrees of freedom as the $\boldsymbol{Q}_1$ grids in Table I. For the 64×64 stretched grid the element aspect ratios vary from 1:1 at the corners to 16:1 at the mid-sides. Athough the AMG effectiveness does deteriorate with increasing aspect ratio, we are happy to see that the spectral bounds remain independent of the grid dimension if the aspect ratio is kept fixed under refinement.

nv	1		2		4	
grid	λ	Λ	λ	Λ	λ	Λ
uniform 8×8	0.864	1.000	0.981	1.000	1.000	1.000
uniform 32×32	0.831	1.000	0.971	1.000	0.999	1.000
stretched 32×32	0.447	1.000	0.694	1.000	0.906	1.000

Table III. Spectral bounds for negative Laplacian matrix preconditioner

[4]Our implementation of this algorithm is the function `m_amgzz.m` in release 3.1 of IFISS.

4. OPTIMAL STOPPING CRITERIA

Our target saddle point system (1) is to be solved using EST_MINRES, a specially tailored version of MINRES.[5] There are two novel features in the algorithm: the estimation of the algebraic error, and the inclusion of an estimate of the discretization error. These two features are discussed in turn.

4.1 Algebraic error estimation

Expressing the target system (1) in the standard form $K\mathbf{x} = \mathbf{b}$ and given a zero initial vector $\mathbf{x}^{(0)} = \mathbf{0}$, MINRES computes a sequence of iterates $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \ldots$ with the property that the ℓ_2-norm of the mth *residual*

$$\|\mathbf{r}^{(m)}\| = \|\mathbf{b} - K\mathbf{x}^{(m)}\| = \|K(\mathbf{x} - \mathbf{x}^{(m)})\|$$

is minimised over the Krylov space

$$\mathcal{K}_m(K, \mathbf{b}) = \text{span}\,\{\mathbf{b}, K\mathbf{b}, \ldots, K^{m-1}\mathbf{b}\}.$$

If the iteration is preconditioned by a positive definite and symmetric matrix operator $M = H^T H$ (corresponding to a discrete version of $\mathcal{M}$ above) then the preconditioned residual norm

$$\|H\mathbf{r}^{(m)}\| = \|\mathbf{r}^{(m)}\|_M \tag{27}$$

is correspondingly minimized over the Krylov space $H\,\mathcal{K}_m(KM, \mathbf{b})$. This means that the reduction of the residual error in the EST_MINRES algorithm is with respect to a discrete norm that explicitly involves the preconditioner M. This reinforces the point that the choice of preconditioner is crucially important.

If the mininimization process (27) is interpreted in the underlying function space setting, then the motive for choosing M to be a discrete version of the operator $\mathcal{M}$ in Section 3 is clear. Since residuals from the dual space $V^* \times Q^*$ are mapped by $\mathcal{M}$ into approximations in the original space, we anticipate that a monotonic reduction of residual errors in the range of the preconditioner will lead to monotonic convergence in the error norm associated with $V \times Q$. In terms of linear algebra, a more precise characterization is the following: given an error vector $\mathbf{e}^{(m)} = \mathbf{x} - \mathbf{x}^{(m)}$ with an associated residual vector $\mathbf{r}^{(m)} = K\mathbf{e}^{(m)}$, we hope to find constants c and C (independent of the dimension of the linear algebra system) such that

$$c\,\|\mathbf{e}^{(m)}\|_E \le \|\mathbf{r}^{(m)}\|_M \le C\,\|\mathbf{e}^{(m)}\|_E, \tag{28}$$

where $M = E^{-1}$, with E the block diagonal matrix representing the norms associated with the underlying space $V \times Q$, see Mardal and Winther [2010, Section 6]. This characterization will inevitably be application specific—for example, the upper bound C is inherited from the boundedness of the bilinear forms a and b appearing in the underlying variational formulation. The lower bound c in (28) is even harder to pin down since it typically depends on the inf-sup stability constant(s) associated with the discretization of V and Q. Estimation of such stability constants is discussed by Elman et al. [2005, Chapter 6] and Powell and Silvester [2004] for Stokes flow and potential flow, respectively. We summarise the key results below.

[5] Our implementation is encoded in the function `est_minres.m` in release 3.1 of IFISS.

We discuss Stokes flow first. To connect with the notation in Elman et al. [2005] we let $\mathbf{A}$ represent the $d \times d$ discrete vector Laplacian with diagonal block $\mathbb{A}$, and let Q represent the pressure mass matrix $\mathbb{I}$. Thus our matrix operators take the form

$$K = \begin{bmatrix} \mathbf{A} & B^T \\ B & 0 \end{bmatrix}, \quad E = \begin{bmatrix} \mathbf{A} & 0 \\ 0 & Q \end{bmatrix} \quad \text{and } M = \begin{bmatrix} \mathbf{A}^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix}. \tag{29}$$

The inf-sup stability of the Stokes mixed approximation is associated with the pressure Schur complement equivalence

$$\gamma^2 \leq \frac{\mathbf{q}^T B \mathbf{A}^{-1} B^T \mathbf{q}}{\mathbf{q}^T Q \mathbf{q}} \leq \Gamma^2 \leq d, \tag{30}$$

where γ is the inf-sup constant and d is the spatial dimension. If we assume that the discretization is stable in the sense of (30), then bounds (28) can be readily established (Elman et al. [2005, Theorem 6.9] with $\delta = \Delta = 1$ and $\theta = \Theta = 1$) with constants given by

$$c^2 = \gamma^2 \left(1 + \tfrac{1}{2}\gamma^2 - \sqrt{1 + \tfrac{1}{4}\gamma^4}\right) \text{ and } C^2 = \max\left\{2 + \Gamma^2, 2\Gamma^2\right\}.$$

Making the asymptotic simplification $(1+x)^{1/2} = 1 + \frac{1}{2}x$ gives $c^2 \sim \frac{1}{2}\gamma^4$, and inverting (28) then leads to the heuristic

$$\|\mathbf{e}^{(m)}\|_E \leq \frac{\sqrt{2}}{\gamma^2} \|\mathbf{r}^{(m)}\|_M, \tag{31}$$

that gives us a criterion for stopping the EST_MINRES iteration, see later. To illustrate the utility of the heuristic (31), Fig. 4 shows the evolution of the errors $\|\mathbf{r}\|_M$ and $\|\mathbf{e}\|_E$ when ideally preconditioned MINRES is applied to a representative flow problem[6] discretized using $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ mixed approximation on a uniform square grid. The upper bound in (31) is also tracked ($\gamma^2 \approx \mathbf{0.0247}$ is found by computing the minimum eigenvalue of the pressure Schur complement problem associated with (30)) and can be seen to provide a reliable bound for the algebraic error $\|\mathbf{x} - \mathbf{x}^{(m)}\|_E$. We note that the energy error is sandwiched between the preconditioned residual error and the upper bound estimate throughout the iteration process.

There are two issues that need to be addressed if a practical implementation is to be developed. The first issue is that in practice, as discussed in Section 3, the *ideal* preconditioner $M = E^{-1}$ is replaced by a spectrally equivalent operator M_* which leads to convergence of the residual in the M_* norm rather than the M norm. Whilst we could take account of this by estimating the equivalence constants associated with the approximation $M_* \sim M$, we prefer to keep things simple. (Our implementation provides a built-in estimate for λ in (26), but we defer discussion until later.) Thus, for a given accuracy tolerance `tol` we stop the EST_MINRES iteration at the first iteration k that satisfies the simple stopping test:

$$\frac{\sqrt{2}}{\gamma^2} \|\mathbf{r}^{(k)}\|_{M_*} \leq \texttt{tol}, \quad \text{with} \quad \|\mathbf{e}^{(k)}\|_E \sim \frac{\sqrt{2}}{\gamma^2} \|\mathbf{r}^{(k)}\|_{M_*}. \tag{32}$$

[6]IFISS problem S2: flow over a step—the convergence curves in Fig. 4 are for $\ell = 4$, but visually identical convergence curves are obtained if the grid resolution parameter is increased.

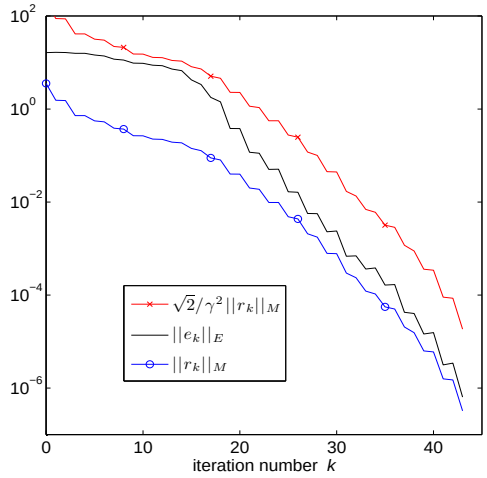

Fig. 4. Optimally preconditioned MINRES for Stokes flow: errors vs. iteration number.

Taking this approach it is important that the approximation of M by M_* be sufficiently accurate. For the discretized flow problem in Fig. 4, the results in Section 3.2 suggest that if sufficiently many AMG V-cycles are taken then the impact on convergence will be minimal.[7] Comparing the convergence curves shown in Fig. 5 with the exact case analogues in Fig. 4 we see that there is very close agreement, especially in the right-hand plot. Moreover, independent of the number of V-cycles that are performed, the energy error remains sandwiched between the preconditioned residual error and the quantity that is used to stop the iteration in (32).

The second practical issue associated with (32) is the need to explicitly compute the inf-sup constant. Finding the minimum eigenvalue of the pressure Schur complement problem (30) is not viable in general. We note that if a flow problem is solved on a nested mesh sequence then a more cost effective strategy is to precompute γ^2 by simply extrapolating estimates obtained by solving the eigenvalue problem associated with (30) on coarse meshes. Herein we propose an alternative approach that does not require coarse mesh estimates. Our idea is to compute estimates for γ^2 on the fly by exploiting the connection between the MINRES iteration and the Lanczos estimates of the eigenvalues of the preconditioned matrix. The key ingredient is the spectral analysis in Elman et al. [2005, Theorem 6.6] which gives bounds for the largest negative eigenvalue λ_- and the smallest positive eigenvalue λ_+ of the matrix M_*K:

$$\lambda_- \le \tfrac{1}{2}\left(\delta - \sqrt{\delta^2 + 4\delta\gamma^2}\right) \quad \text{and} \quad \delta < \lambda_+. \tag{33}$$

If we assume that these two bounds are tight and invert (33) then we get the

[7] Looking at Table III, we see that taking just one or two V-cycles will suffice, especially in the case of uniform grids.

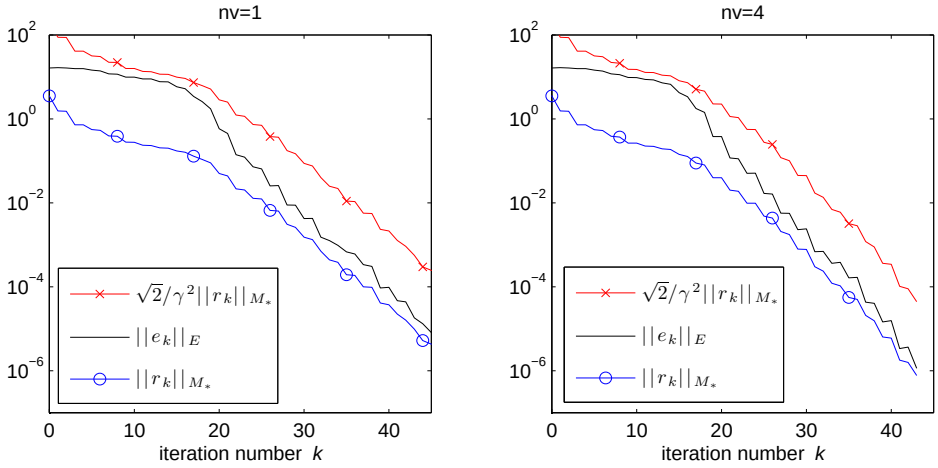

Fig. 5. Preconditioned MINRES for Stokes flow: Negative Laplacian preconditioning with one V-cycle (left) and four V-cycles (right).

estimate

$$\gamma_k^2 = \left(\lambda_-^2 - \lambda_-\lambda_+\right)/\lambda_+, \tag{34}$$

which can be computed at every step of EST_MINRES from the associated Lanczos estimates for λ_- and λ_+ at that iteration. The details are given in Section 5. Results obtained when γ^2 is replaced by γ_k^2 in the upper bound estimate for our test problem are shown in Fig. 6. Comparing the upper bounds with those in Fig. 5 we see very close agreement as the iteration proceeds (the red lines are indistinguishable for $k > 20$). This suggests that the bounds in (33) are tight. It also shows that the Lanczos convergence to the eigenvalues closest to zero is rapid enough for this strategy to be of practical use. There is another bonus—the computed estimates of λ_+ converge to the lower bound λ in (26). This means that we are given a free estimate of the accuracy of the approximation $M_* \sim M$ as the EST_MINRES iteration proceeds.

Turning to potential flow, we will see that analogous issues arise. We restrict our attention to ideal H(div) preconditioning herein—a more complete discussion of MINRES preconditioning using discrete versions of (23) and (24) is given by Powell and Silvester [2004]. Thus, if we let $\mathbb{I}$ represent the $\boldsymbol{RT}_0$ mass matrix, and let Q represent the diagonal pressure mass matrix then the analogue of the stability bound (30) is the pressure Schur complement equivalence (cf. (30)) :

$$\beta^2 \leq \frac{\mathbf{q}^T B(\mathbb{I}+D)^{-1}B^T\mathbf{q}}{\mathbf{q}^T Q\mathbf{q}} \leq 1, \tag{35}$$

where β plays the role of the inf-sup constant and the $n_u \times n_u$ matrix $D = B^T Q^{-1} B$ is a representation of the L^2-norm of the divergence operator. Our stopping criterion in this case is based on the heuristic:

$$\|\mathbf{e}^{(m)}\|_E \leq \frac{1}{\beta^2}\,\|\mathbf{r}^{(m)}\|_M. \tag{36}$$

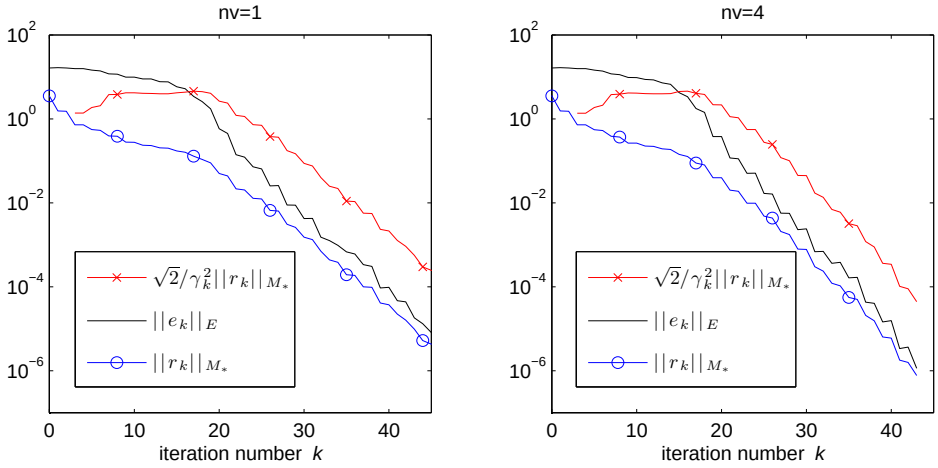

Fig. 6. Preconditioned MINRES for Stokes flow with dynamic estimation of γ^2: Negative Laplacian preconditioning with one V-cycle (left) and four V-cycles (right).

Next, from Powell and Silvester [2004, Corollary 2.4] we quote some simple eigenvalue bounds for the matrix M_*K :

$$\lambda_- \leq -\beta^2 \quad \text{and} \quad 1 = \lambda_+. \tag{37}$$

So we see that, mirroring the Stokes case, the lower bound λ_- gives a mechanism for estimating β^2 on the fly as

$$\beta_k = \sqrt{-\lambda_-}, \tag{38}$$

where λ_- is an estimate of the largest (closest to zero) negative eigenvalue of the coefficient matrix by the Lanczos recurrence.

Fig. 7 shows the evolution of the errors $\|\mathbf{r}\|_M$ and $\|\mathbf{e}\|_E$ when ideally preconditioned MINRES is applied to a representative flow problem from Silvester and Powell [2007][8] discretised using $\boldsymbol{RT}_0$ mixed approximation on a mesh of 9360 triangles. The upper bound in (36) is also tracked with β_k^2 estimated at every step via (37). Note that $\beta^2 \approx \mathbf{0.664}$. MINRES convergence is rapid—much faster than for Stokes flow—and the three curves plotted are not easily distinguished. Looking closely we see that the energy error stays below the upper bound after the third iteration. Refined eigenvalue bounds in the case of inexact H(div) preconditioning are developed by Powell [2005].

4.2 Discretization error estimation

If our algorithm is to be run in a "black-box" fashion, then we need to connect the absolute tolerance in the stopping test (32) with the PDE discretization error $\|\vec{u} - \vec{u}_h\|_V + \|p - p_h\|_M$. If we have an a posteriori estimator for the error, and if it is applicable to any function $(\vec{u}_h, q_h) \in V_h \times Q_h$ as in (13), then one possibility

[8]PIFISS problem P3: flow around a cylinder—the convergence curves in Fig. 7 do not change if the grid resolution is increased.

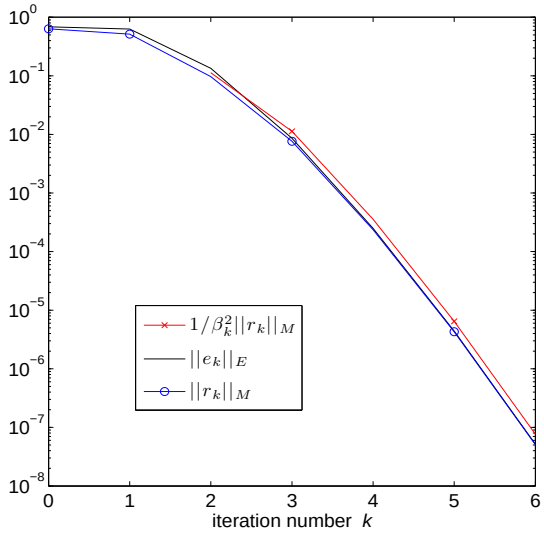

Fig. 7. H(div) preconditioned MINRES for potential flow: dynamic estimation of β^2.

is to estimate the error at every iteration (the iterates $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \ldots$ can be associated with a sequence of functions from $V_h \times Q_h$). For example, for Stokes flow, we know from (11) that

$$c\,\eta^{(m)} \leq \|\nabla(\,\vec{u} - \vec{u}_h^{(m)})\| + \|p - p_h^{(m)}\| \leq C\,\eta^{(m)}, \quad m = 1, 2, 3 \ldots,$$

so a simple strategy is to stop the EST_MINRES iteration when the algebraic error is comparable with the estimate of the discretization error $\eta^{(m)}$, that is, as soon as

$$\frac{\sqrt{2}}{\gamma^2}\,\|\mathbf{r}^{(m)}\|_{M_*} \leq \eta^{(m)}, \tag{39}$$

$$\text{with} \quad \|\mathbf{e}^{(m)}\|_E \sim \frac{\sqrt{2}}{\gamma^2}\,\|\mathbf{r}^{(m)}\|_{M_*} \quad \text{and} \quad \eta^{(m)} \sim \|\,\vec{u} - \vec{u}_h^{(m)}\|_V + \|p - p_h^{(m)}\|_M. \tag{40}$$

A rigorous justification for this choice is given by Jiránek et al. [2010, Theorem 6.3]. Looking at the specific Stokes flow test problem (flow over a step) used in Section 4.1, the results in Fig. 8 illustrate why we might consider adopting the dynamic stopping test (39). The evolution of errors shown in Fig. 8 are directly comparable with those in Fig. 6; for clarity we have simply removed the algebraic errors $\|\mathbf{x} - \mathbf{x}^{(m)}\|_E$ and plotted the evolution of the approximation error estimate $\eta^{(m)}$ instead. Note that as the iteration proceeds the algebraic error becomes insignificant relative to the approximation error. The termination point associated with the refined stopping test (39) is marked with a "$*$". Thus, if the stopping test (39) is "hard-wired" into EST_MINRES then the iteration is terminated after only 28 steps if we take one V-cycle of AMG. If the four V-cycle AMG preconditioner is used instead, then EST_MINRES stops after 26 steps.

In the example above the value of η (that is, the discretization error estimate achieved when the MINRES iteration converges) is relatively large $\eta \approx \mathbf{0.269}$. This

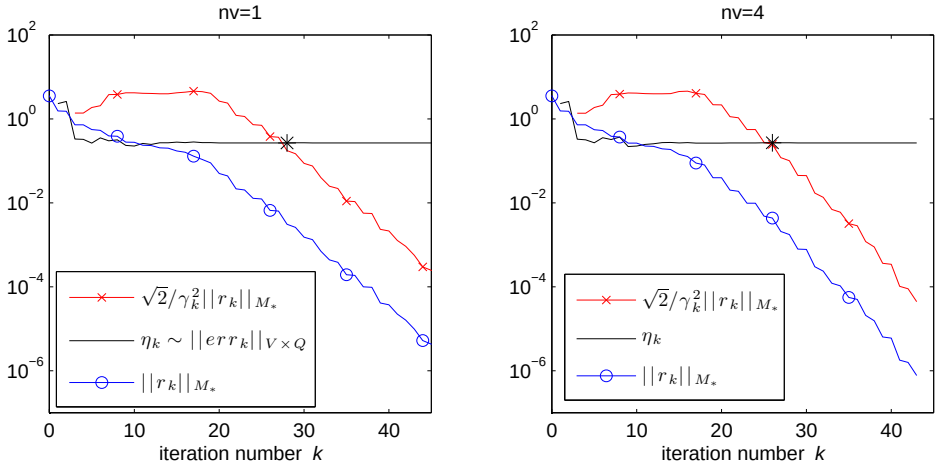

Fig. 8. Preconditioned MINRES for Stokes flow over a step with dynamic estimation of η: Negative Laplacian preconditioning with one V-cycle (left) and four V-cycles (right).

is bigger than might be anticipated because the flow over step problem solved in Fig. 8 is *singular* at the re-entrant corner—severely impinging on the attainable solution accuracy. Thus, as the grid is refined, the convergence in energy is slow: behaving like $O(h^{2/3})$ independent of the order of the mixed approximation.

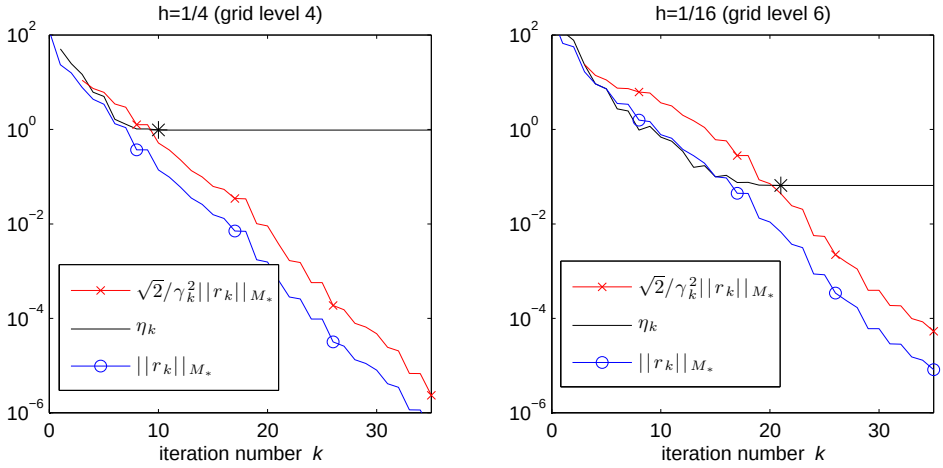

Fig. 9. Preconditioned MINRES for Stokes flow in a square domain with dynamic estimation of η: Negative Laplacian preconditioning with one V-cycle: 8×8 grid (left) and 32×32 grid (right).

Analogous results obtained for a more regular (enclosed–) flow problem[9] are shown in Fig. 9. In this case the spatial convergence is much more rapid, which

[9]IFISS problem S4: smooth colliding flow with a quartic polynomial velocity solution.

explains the very different position of "∗" in the left and right plots. For the coarse grid computation the automatic stopping test leads to termination after 10 iterations. The spatial approximation is sixteen times smaller for the fine grid computation—thus more MINRES iterations must be taken to reduce the algebraic error to a level that is commensurate with the approximation error. The stopping test suggests that 21 iterations would suffice in this case.

grid	k_*	η	$\|\nabla \cdot \vec{u}_h\|$
uniform 8×8	10	9.71×10^{-1}	2.97×10^{-2}
uniform 16×16	17	2.54×10^{-1}	3.66×10^{-3}
uniform 32×32	21	6.51×10^{-2}	4.56×10^{-4}
uniform 64×64	24	1.64×10^{-2}	5.69×10^{-5}

Table IV. Variation of estimated spatial accuracy with increased grid refinement for a smooth solution

Results reported in Table IV show the variation in "optimal" iteration count k_* when this flow problem is approximated with increasing spatial resolution. We know from a priori error analysis, see for example Elman et al. [2005, Section 5.4.1], that the spatial accuracy of $\boldsymbol{Q}_2$–$\boldsymbol{P}_{-1}$ approximation is $O(h^2)$ in the energy norm if the flow solution is sufficiently smooth. This behaviour is clearly evident in the tabulated values of η. We also note that the divergence residual error is converging at a faster rate ($O(h^3)$) which means that the estimated error η is increasingly dominated by the velocity error component in (13) as h is reduced.

5. SOFTWARE DESIGN AND IMPLEMENTATION ASPECTS

To set up notation for this section, consider solving the linear system $\hat{K}\hat{\mathbf{x}} = \hat{\mathbf{b}}$ with $\hat{K}$ representing the symmetric and indefinite preconditioned matrix. As already mentioned, MINRES generates a sequence of approximations $\hat{\mathbf{x}}^{(m)}$, $m = 1, 2, \ldots$, with $\hat{\mathbf{x}}^{(m)} \in K_m(\hat{K}, \hat{\mathbf{b}})$, such that the residual $\hat{\mathbf{r}}^{(m)} = \hat{\mathbf{b}} - \hat{K}\hat{\mathbf{x}}^{(m)}$ is minimized. Let $\{\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(m)}\}$ be a set of orthonormal vectors spanning $K_m(\hat{K}, \hat{\mathbf{b}})$, with $\mathbf{w}^{(1)} = \hat{\mathbf{b}}/\|\hat{\mathbf{b}}\|$, and let $W_m = [\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(m)}]$. These vectors can be generated iteratively by means of the following well-known recurrence:

$$\hat{K}W_m = W_m T_m + t_{m+1,m}\mathbf{w}^{(m+1)}\mathbf{e}_m^T =: W_{m+1}\underline{T}_m, \tag{41}$$

where $\mathbf{e}_m$ is the mth vector of the canonical basis and T_m is a tridiagonal symmetric matrix containing the orthogonalization coefficients; see Greenbaum [1997, Section 2.5] or Saad [2003, Section 6.6] for full details. Using (41) we readily obtain the following relationship for the approximation $\hat{\mathbf{x}}^{(m)} = W_m\mathbf{y}^{(m)}$:

$$\hat{\mathbf{r}}^{(m)} = \hat{\mathbf{b}} - \hat{K}\hat{\mathbf{x}}^{(m)} = V_{m+1}\left(\mathbf{e}_1\|\hat{\mathbf{b}}\| - \underline{T}_m\mathbf{y}^{(m)}\right).$$

The minimizing solution $\hat{\mathbf{x}}^{(m)}$ is thus obtained by solving the least squares problem $\min_{\mathbf{y}} \|\mathbf{e}_1\|\hat{\mathbf{b}}\| - \underline{T}_m\mathbf{y}\|$. Thanks to the tridiagonal form of $\underline{T}_m$ the least squares solution can be updated without explicitly solving the $(m+1) \times m$ problem at each iteration. This can be efficiently done on the fly by means of Givens rotations; see Greenbaum [1997, Algorithm 4] and the actual implementation shown in Figure 10.

The Lanczos relation in (41) can be used to show that $T_m = W_m^T \hat{K} W_m$, so that the eigenvalues of T_m, also referred to as Ritz values, provide approximations for the eigenvalues of $\hat{K}$. For a moderate space dimension m, the work of Parlett [1998, Chapter 13] shows that the extreme eigenvalues can give very accurate estimates for the corresponding eigenvalues of $\hat{K}$. In our context, $\hat{K}$ is indefinite, and we are interested in good approximations to the interior (closest to zero) eigenvalues, so as to obtain estimates such as (34) for γ_k and (38) for β_k. Unfortunately, Ritz values do not, in general, approximate interior eigenvalues accurately. Fortunately, the matrices generated within the Lanczos process also allow us to compute so-called *harmonic* Ritz values: $\theta_1, \ldots, \theta_m$, which are the m roots of the residual polynomial ϕ_m, defined as $\mathbf{r}^{(m)} = \phi_m(\hat{K})\hat{\mathbf{b}}$, with

$$\phi_m(\theta) = \frac{1}{\hat{\phi}_m(0)}\hat{\phi}(\theta), \quad \text{and} \quad \hat{\phi}(\theta) = \prod_{j=1}^{m}(\theta - \theta_j).$$

Using the Lanczos relation in (41) and its connection to polynomial recurrences, the harmonic Ritz values can be computed by solving the following generalized eigenvalue problem (see Morgan [1991] for an early implementation)

$$\underline{T}_m^T \, \underline{T}_m \mathbf{u} = \theta \, T_m \mathbf{u}.$$

We refer to Freund [1992], and to Paige et al. [1995] for a proof that these eigenvalues are indeed the roots of the residual polynomial. Morgan [1991] observed that the harmonic Ritz values tend to approximate first the interior eigenvalues, and he used these values to minimize the Rayleigh quotient for $\hat{K}^{-1}$. Paige et al. [1995] showed that the inverses of the harmonic Ritz values are weighted means of the inverses of the eigenvalues of $\hat{K}^{-1}$. Both arguments suggest that harmonic Ritz values are tightly related to a spectral approximation procedure for $\hat{K}^{-1}$, which is exactly what we want in our algorithm.

Another useful feature of harmonic Ritz values is that the smallest positive such value approximates the smallest positive eigenvalue of $\hat{K}$ from above, whilst the largest negative harmonic Ritz value approximates the largest negative eigenvalue of $\hat{K}$ from below. Therefore, any interval containing zero that is free of $\hat{K}$'s eigenvalues is also free of harmonic Ritz values. It was also experimentally observed by Paige et al. [1995] that "the minimum residual process converges much faster after the smallest positive eigenvalue and the largest negative one have been approximated sufficiently well" by the harmonic Ritz values. This behaviour can be observed in Fig. 6, where the convergence markedly improves at the point where the approximation of γ by the interior harmonic Ritz estimation process settles down. This phenomenon is more generally associated with the convergence of the extremal eigenvalues of positive definite matrices. In the indefinite case, however, the convergence behaviour is strongly influenced by the eigenvalues closest to zero, see Greenbaum [1997, Section 3.1] and thus it is not surprising that accurate approximations to these eigenvalues speed up convergence. In our context, such behaviour implies that an accurate computation of the error estimate is to be expected as soon as the underlying MINRES iteration starts converging.

A template of the MINRES algorithm which has our built-in stopping test is given in Fig. 10. The two preconditioner constructions in Section 4.1 are differen-

tiated by the value assigned to the switch `prob_type`. The algorithm calls three external functions. The first of these is the preconditioning function PREC which returns a vector $\mathbf{z}$ which effects the result of multiplying the input vector by the preconditioning matrix M_* as discussed in Section 3. The second external function PARAM_EST is described in Fig. 11. It returns the parameter that is used to scale the residual norm in the stopping criteria (34) and (38). The third external function is ERROR_EST. If an energy error estimator is available as discussed in Section 4.2 then this function generates the error estimate $\eta^{(m)}$ that is associated with the current MINRES solution iterate. If a discretization error estimator is not available, as in Section 4.2, then the value of $\eta^{(m)}$ that is returned by ERROR_EST can simply be set to a fixed algebraic tolerance `tol` that is less than the anticipated discretization error.

Algorithm : `EST_MINRES`.
Given $\mathbf{b}, K, \mathbf{x}^{(0)}$ and `prob_type`
Set $\mathbf{r}^{(0)} = \mathbf{b} - K\mathbf{x}^{(0)}$, $\hat{\mathbf{r}}^{(0)} = \texttt{prec}(\mathbf{r}^{(0)})$, $\rho_0 = \sqrt{(\mathbf{r}^{(0)})^T \hat{\mathbf{r}}^{(0)}}$
Initialize basis vectors: $\mathbf{w} = \hat{\mathbf{r}}^{(0)}/\rho_0$, $\mathbf{p}^{(-1)} = \mathbf{0}$, $\mathbf{p}^{(0)} = \mathbf{r}^{(0)}/\rho_0$
Initialize auxiliary vectors: $\mathbf{d}^{(-1)} = \mathbf{0}$, $\mathbf{d}^{(0)} = \mathbf{0}$
Initialize projected right-hand side: $f = \rho_0$
`for` $m = 1, 2, \ldots$ `until` convergence `do`
  Generate new basis and auxiliary vectors: $\mathbf{p}^{(m)} = K\mathbf{w}, \quad \mathbf{d}^{(m)} = \mathbf{w}$
    `if` m>1,
    $t_{m-1,m} = t_{m,m-1}$
    $\mathbf{p}^{(m)} = \mathbf{p}^{(m)} - \mathbf{p}^{(m-1)} t_{m-1,m}$
  $t_{m,m} = \mathbf{w}^T \mathbf{p}^{(m)}$
  $\mathbf{p}^{(m)} = \mathbf{p}^{(m)} - \mathbf{p}^{(m-1)} t_{m,m}$
  Compute preconditioned basis vector: $\mathbf{w} = \texttt{prec}(\mathbf{p}^{(m)})$
  $t_{m+1,m} = \sqrt{\mathbf{w}^T \mathbf{p}^{(m)}}$
  $\mathbf{p}^{(m)} = \mathbf{p}^{(m)}/t_{m+1,m}$, $\mathbf{w} = \mathbf{w}/t_{m+1,m}$
  Compute parameter for stopping test: `coef` = `param_est` ($\underline{T}_m$, `prob_type`)
  Apply previous rotations:
    `if` m>2, $\rho_{1:2} = G_{m-2} t_{m-2:m-1,m}$, $\rho_{2:3} = G_{m-1}[\rho_2; t_{m,m}]$
    `elseif` m=2, $\rho_{2:3} = G_{m-1} t_{1:2,2}$
    `elseif` m=1, $\rho_3 = t_{1,1}$
  Compute new rotations:
    $\delta = \sqrt{\rho_3^2 + t_{m+1,m}^2}$, $c = |\rho_3|/\delta$, $s = \text{sign}(\rho_3) t_{m+1,m}/\delta$
  Apply new rotations: $\rho_3 = c\rho_3 + s t_{m+1,m}$, $\hat{f} = -sf$, $f = cf$, $G_m = [c\ s; -s\ c]$
  Update auxiliary vector: $\mathbf{d}^{(m)} = (\mathbf{d}^{(m)} - \mathbf{d}^{(m-1)}\rho_1 - \mathbf{d}^{(m-2)}\rho_2)/\rho_3$
  Update solution: $\mathbf{x}^{(m)} = \mathbf{x}^{(m-1)} + \mathbf{d}^{(m)} \hat{f}$
  Compute discretization error estimate : $\eta^{(m)} = \texttt{error\_est}\ (\mathbf{x}^{(m)})$
  Stopping test: `if coef`$\cdot|\hat{f}| \leq \eta^{(m)}$, convergence
  Update residual norm: $f = \hat{f}$
`enddo`

Fig. 10. The EST_MINRES algorithm

A more efficient procedure to determine the harmonic Ritz values consists in recasting the generalized eigenvalue problem as a standard eigenvalue problem with a rank-one modification of T_m; see Paige et al. [1995]. The use of an effective preconditioner generally guarantees that a large number of iterations will not be

function `coef` = `param_est` ($\underline{T}_m$, `prob_type`)
Compute the eigenvalues $\theta_1 \leq \cdots \leq \theta_m$ of $(\underline{T}_m^T \underline{T}_m, T_m)$
Compute $\lambda_- = \max_{\theta_j<0} \theta_j$, $\lambda_+ = \min_{\theta_j>0} \theta_j$
if $\lambda_- = \emptyset$, set $\lambda_- = -\lambda_+$
if $\lambda_+ = \emptyset$, set $\lambda_+ = -\lambda_-$
if `prob_type` = 'Stokes flow', Compute γ_k^2 as in (34); set `coef` $= \sqrt{2}/\gamma_k^2$
if `prob_type` = 'Potential flow', Compute β_k as in (38); set `coef` $= 1/\beta_k^2$
endfunction

Fig. 11. Specification of associated function PARAM_EST

performed. Nonetheless, if this did happen, then one could consider computing only the required interior harmonic Ritz values.

To conclude this discussion, we reproduce a MATLAB session below that shows the utility of the EST_MINRES implementation within the IFISS software package, a full description of which can be found in the algorithm paper of Elman et al. [2007]. The specific Stokes flow problem under consideration is the analytic problem (S4) that features in the right-hand plot in Fig. 9. The discrete saddle point problem is assembled at the start of the session by calling the driver `stokes_testproblem`. For the chosen mixed approximation and grid parameter the saddle point system is of the form (10) with $n_u = 8,450$ and $n_p = 3,072$. The "exact" discrete solution `xst` is computed via the built-in ("backslash") sparse solver. Note that the ill-conditioning warning is generated because the discrete Stokes system has a one-dimensional null space in exact arithmetic; see Elman et al. [2005, pp. 224–229] for a full discussion. The call to the error estimator function `stokespost` postprocesses the solution vector `xst` and outputs the computed values of η and $\|\nabla \cdot \vec{u}_h\|$ given in Table IV.

```
>> stokes_testproblem, stokespost
specification of reference Stokes problem.
choose specific example (default is cavity)
     1  Channel domain
     2  Flow over a backward facing step
     3  Lid driven cavity
     4  Colliding flow
 : 4

Grid generation for cavity domain.
grid parameter: 3 for underlying 8x8 grid (default is 16x16) : 6
uniform/stretched grid (1/2) (default is uniform) :  1
Q1-Q1/Q1-P0/Q2-Q1/Q2-P1: 1/2/3/4? (default Q1-P0) : 4
setting up Q2-P1 matrices...  done
system matrices saved in square_stokes_nobc.mat ...
imposing boundary conditions and solving system ...
Warning: Matrix is close to singular or badly scaled.
         Results may be inaccurate. RCOND = 3.834113e-17.
This should not cause difficulty for enclosed flow problems.
 ...
FAST a posteriori error estimation
estimated velocity divergence error:  4.558566e-04
Estimated energy error is 6.5133e-02
```

The driver `itsolve_stokes` provides the interface to the `est_minres` function, which is called after the preconditioning strategy (one V-cycle in this instance) has been determined. The computed values of $\eta^{(m)}$, $\frac{\sqrt{2}}{\gamma^2}\|\mathbf{r}^{(m)}\|_{M_*}$ and $\|\mathbf{r}^{(m)}\|_{M_*}$ are output at each iteration. These are the values that are plotted in Fig. 9. The MINRES solution `xest` is returned as soon as the stopping criterion $\texttt{coef}\cdot|\hat{f}| \leq \eta^{(m)}$ is satisfied, namely after 21 iterations—corresponding to the asterisk in the plot. The evolution of the successive estimates of γ_k and λ_+ is reported before exiting. Note that the "Final estimated error" that is also displayed agrees to four decimal places with the exact error estimate η computed earlier. The final calculation that is made shows that the two velocity solution vectors agree to four decimal places in all components. As might be expected, the plots of the flow solution generated from `xst` and `xest` cannot be distinguished from each other.

```
>>  itsolve_stokes
Inexact AMG block preconditioning ..
number of V-Cycles? (default 1) : 1
AMG grid coarsening ...  8 grid levels constructed.
AMG with point damped Gauss-Seidel smoothing ..
Call to EST_MINRES with built in error control ..

   k    Estimated-Error   Algebraic-Bound   Residual-Error
   1    1.2035e+02                          6.6773e+01
   2    7.7520e+01                          5.6539e+01
   3    2.3430e+01                          1.6600e+01
   4    9.2256e+00                          9.2119e+00
   5    7.2940e+00                          7.2152e+00
   6    2.7405e+00        7.5056e+00        3.5500e+00
   7    2.4702e+00        7.3139e+00        3.4225e+00
   8    9.6702e-01        6.2312e+00        1.5778e+00
   9    1.1626e+00        5.9204e+00        1.4661e+00
  10    6.8898e-01        3.6656e+00        7.7561e-01
  11    5.6553e-01        3.1566e+00        6.5119e-01
  12    3.4578e-01        1.9966e+00        3.7893e-01
  13    1.5762e-01        1.5262e+00        2.7801e-01
  14    1.7042e-01        1.1055e+00        1.9291e-01
  15    9.9190e-02        6.0673e-01        9.9876e-02
  16    1.0663e-01        5.7272e-01        9.3967e-02
  17    7.5535e-02        2.8017e-01        4.4350e-02
  18    7.5720e-02        2.7952e-01        4.4243e-02
  19    6.6334e-02        8.6494e-02        1.3315e-02
  20    6.5934e-02        7.1054e-02        1.0922e-02
  21    6.5445e-02        4.3954e-02        6.7394e-03    Bingo!

 Eigenvalue convergence
    k       infsup       lambda
    4       0.9807       1.0212
    5       0.9340       1.0065
    6       0.9147       1.0024
    7       0.6689       0.9896
    8       0.6618       0.9887
```

```
    9       0.3581       0.9705
   10       0.3502       0.9671
   11       0.2992       0.9305
   12       0.2917       0.9218
   13       0.2684       0.9071
   14       0.2576       0.9031
   15       0.2468       0.9000
   16       0.2328       0.8973
   17       0.2320       0.8971
   18       0.2239       0.8945
   19       0.2238       0.8945
   20       0.2177       0.8904
   21       0.2174       0.8897

 Final estimated error is 6.5445e-02
 Optimality in  21 iterations

>> [np,nu]=size(Bst);  xdiff=norm(xest(1:nu)-xst(1:nu),inf);
>> fprintf('velocity solution difference is %7.3e\n',xdiff)
velocity solution difference is 6.888e-04
```

6. CONCLUSIONS

This article describes the design and implementation of EST_MINRES, an algorithm for solving symmetric saddle-point systems. It is argued that consideration of the PDE origins of such systems is essential if uniformly efficient preconditioning is to be achieved. It is also demonstrated that if an (energy–) a posteriori error estimation routine is available then an optimally efficient stopping criterion can be realized. An important point is that our solver methodology is very general in scope—although the emphasis in this article is on discretized problems arising in the modelling of incompressible fluid flow—the EST_MINRES algorithm (and our MATLAB implementation) is applicable to any saddle point system that arises from mixed approximation.

REFERENCES

Ainsworth, M. and Oden, J. 1997. A posteriori error estimates for Stokes' and Oseen's equations. *SIAM J. Numer. Anal 34*, 228–245.

Arioli, M. 2010. Generalized Golub-Kahan bidiagonalization and stopping criteria. *Tech. Report, RAL-TR-2010-008.*

Arioli, M. and Loghin, D. 2008. Stopping criteria for mixed finite element problems. *Electron. Trans. Numer. Anal. 29*, 178–192.

Arioli, M., Loghin, D., and Wathen, A. 2005. Stopping criteria for iterations in finite-element methods. *Numer. Math. 99*, 381–410.

Arnold, D., Falk, R., and Winther, R. 1997. Preconditioning in H(div) and applications. *Math. Comp. 66*, 957–984.

Arnold, D., Falk, R., and Winther, R. 2000. Multigrid in H(div) and H(curl). *Numer. Math. 85*, 197–217.

Benzi, M., Golub, G. H., and Liesen, J. 2005. Numerical solution of saddle point problems. *Acta Numerica 14*, 1–137.

Brezzi, F. and Fortin, M. 1991. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, New York.

Elman, H., Ramage, A., and Silvester, D. 2007. Algorithm 866: IFISS, a Matlab toolbox for modelling incompressible flow. *ACM Trans. Math. Soft. 33*, 2–14.

Elman, H., Silvester, D., and Wathen, A. 2005. *Finite Elements and Fast Iterative Solvers with Applications in Incompressible Fluid Dynamics*. Oxford University Press, Oxford, UK.

Freund, R. W. 1992. Quasi-kernel polynomials and convergence results for quasi-minimal residual iterations. In *Numerical Methods in Approximation Theory, Vol. 9*, D. Braess and L. L. Schumaker, Eds. Birkhäuser Verlag, Basel, 77–95.

Golub, G. H. and Meurant, G. A. 1997. Matrices, moments and quadrature II; how to compute the norm of the error in iterative methods. *BIT 37*, 687–705.

Greenbaum, A. 1997. *Iterative Methods for Solving Linear Systems*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Hiptmair, R. and Xu, J. 2007. Nodal auxiliary space preconditioning in H(curl) and H(div) spaces. *SIAM J. Numer. Anal. 45*, 2483–2509.

Jiránek, P., Strakoš, Z., and Vohralík, M. 2010. A posteriori error estimates including algebraic error and stopping criteria for iterative solvers. *SIAM J. Sci. Comput. 32*, 1567–1590.

Kay, D. and Silvester, D. 1999. A posteriori error estimation for stabilized mixed approximations of the Stokes equations. *SIAM J. Sci. Comput. 21*, 1321–1336.

Liao, Q. and Silvester, D. 2010. A simple yet effective a posteriori error estimator for classical mixed approximation of Stokes equations. *Applied Numerical Mathematics*. doi:10.1016/j.apnum.2010.05.003.

Mardal, K.-A. and Winther, R. 2010. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*. doi: 10.1002/nla.716.

Meurant, G. 2005. Estimates of the l_2 norm of the error in the conjugate gradient algorithm. *Numerical Algorithms 40*, 157–169.

Meurant, G. and Strakoš, Z. 2006. The Lanczos and conjugate gradient algorithms in finite precision arithmetic. *Acta Numerica 15*, 471–542.

Morgan, R. B. 1991. Computing interior eigenvalues of large matrices. *Lin. Alg. Appl. 154–156*, 289–309.

Paige, C., Parlett, B., and van der Vorst, H. 1995. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numer. Linear Algebra Appl. 2*, 115–134.

Paige, C. and Saunders, M. 1975. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal. 12*, 617–629.

Parlett, B. N. 1998. *The Symmetric Eigenvalue Problem*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Powell, C. 2005. Parameter-free H(div) preconditioning for mixed finite element formulation of diffusion problems. *IMA J. Numer. Anal. 25*, 783–796.

Powell, C. and Silvester, D. 2004. Optimal preconditioning for Raviart-Thomas mixed formulation of second-order elliptic problems. *SIAM J. Matrix Anal. Appl. 25*, 718–738.

Rusten, T. and Winther, R. 1992. A preconditioned iterative method for saddle-point problems. *SIAM J. Matrix Anal. Appl. 13*, 887–904.

Saad, Y. 2003. *Iterative Methods for Sparse Linear Systems*, 2nd ed. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA.

Silvester, D. and Powell, C. 2007. Potential (Incompressible) Flow and Iterative Solution Software Guide (PIFISS), version 1.0. Tech. Rep. 2007.14, MIMS, Manchester. `http://eprints.ma.man.ac.uk/700/`.

Silvester, D. and Wathen, A. 1994. Fast iterative solution of stabilised Stokes systems. Part II: Using general block preconditioners. *SIAM J. Numer. Anal. 31*, 1352–1367.

Strakoš, Z. and Tichý, P. 2002. On error estimation in the Conjugate Gradient method and why it works in finite precision computations. *Electron. Trans. Numer. Anal. 13*, 56–80.

Wathen, A. 2007. Preconditioning and convergence in the right norm. *Int. J. Comput. Math. 84*, 1199–1209.

Wathen, A. and Rees, T. 2009. Chebyshev semi-iteration in preconditioning for problems including the mass matrix. *Elect. Trans. Numer. Anal. 34*, 125–135.