Technical Report
RAL-TR-2006-010

Stopping criteria for mixed finite element problems

Mario Arioli and D. Loghin

October 29, 2006

Council for the Central Laboratory of the Research Councils

Library and Information Services
CCLRC Rutherford Appleton Laboratory
Chilton Didcot
Oxfordshire OX11 0QX
UK
Tel: +44 (0)1235 445384
Fax: +44(0)1235 446403
Email: library@rl.ac.uk

CCLRC reports are available online at:
http://www.clrc.ac.uk/Activity/ACTIVITY=Publications;SECTION=225;

ISSN 1358-6254

RAL-TR-2006-010

Stopping criteria for mixed finite element problems

M. Arioli[1] and D. Loghin[2]

ABSTRACT
We study stopping criteria that are suitable in the solution by Krylov space based methods of linear and non linear systems of equations arising from the mixed and the mixed-hybrid finite-element approximation of saddle point problems. Our approach is based on the equivalence between the Babuška and Brezzi conditions of stability which allows us to apply some of the results obtained in Arioli, Loghin and Wathen (2005). Our proposed criterion involves evaluating the residual in a norm defined on the discrete dual of the space where we seek a solution. We illustrate our approach using standard iterative methods such as MINRES and GMRES. We test our criteria on Stokes and Navier-Stokes problems both in a linear and nonlinear context.

Keywords: augmented systems, sparse matrices, GMRES, MINRES, Navier-Stokes equation, stopping criteria.

AMS(MOS) subject classifications: 65F05

Current reports available by anonymous ftp to ftp.numerical.rl.ac.uk in directory pub/reports.
[1] m.arioli@rl.ac.uk, Rutherford Appleton Laboratory,
[2] loghind@for.mat.bham.ac.uk University of Birmingham Watson Building Birmingham B15 2TT, UK
The work of first author was supported by EPSRC grant GR/S42170/01.

Computational Science and Engineering Department
Atlas Centre
Rutherford Appleton Laboratory
Oxon OX11 0QX

October 29, 2006

Contents

1 Introduction

Mixed and mixed-hybrid finite-element methods form a class of popular discretization methods designed to approximate systems of partial differential equations of saddle-point type arising in the modeling of a variety of physical phenomena in areas such as fluid-dynamics or linear elasticity. They generally give rise to large, nonsymmetric, indefinite linear and nonlinear systems for which the solution is typically sought via iterative approaches. An essential feature of such methods is the stopping criteria employed. This work aims to describe how to devise a suitable stopping procedure, given the well-defined theoretical context of variational formulation of partial differential equations and in particular the mixed finite element theory.
The outline of the paper is as follows. In Section 2, we introduce the abstract formulation of a generic saddle-point problem as a system based on bilinear forms. Then, we describe a general framework in which we can formulate a stopping criterion based on the energy norm of the error between the exact solution of the continuous problem and the solution computed by an iterative method. Section 3 generalizes the stopping criterion derived in (Arioli et al. 2005) to the case of mixed finite element formulations, discussing both the linear symmetric and nonsymmetric cases. We also propose a strategy for the extension of the stopping criteria to the nonlinear case. Finally, in Section 4, we present our class of test problems together with the convergence behaviour of some iterative algorithms showing the beneficial effect of our stopping criteria.

2 Mixed variational formulation

We start by summarizing the theoretical setting necessary to describe our problem. A comprehensive and exhaustive introduction can be found in the book of Brezzi and Fortin (1991).
Let $\mathcal{V}, \mathcal{Q}$ be Hilbert spaces with norms $\|\cdot\|_{\mathcal{V}}, \|\cdot\|_{\mathcal{Q}}$ and duals $\mathcal{V}^*, \mathcal{Q}^*$, respectively. Consider the two real-valued bilinear forms $a(\cdot,\cdot) : \mathcal{V}\times\mathcal{V}$, $b(\cdot,\cdot) : \mathcal{V}\times\mathcal{Q}$ and the two linear functionals $f(\cdot) \in \mathcal{V}^*, g(\cdot) \in \mathcal{Q}^*$. We are interested in the following abstract variational formulation

$$(\mathcal{SP})\begin{cases} \text{Find } (u,p)\in\mathcal{V}\times\mathcal{Q} \text{ such that for all } (v,q)\in\mathcal{V}\times\mathcal{Q} \\ a(u,v)+b(v,p)=f(v), \\ b(u,q) \qquad\qquad\;\; = g(q). \end{cases}$$

In the nonlinear case the bilinear form $a(\cdot,\cdot)$ is replaced by the nonlinear operator $F : \mathcal{V}\to\mathcal{V}^*$, as, for example, in the Navier-Stokes case. The variational formulation in this case reads

$$(\mathcal{NSP})\begin{cases} \text{Find } (u,p)\in\mathcal{V}\times\mathcal{Q} \text{ such that for all } (v,q)\in\mathcal{V}\times\mathcal{Q} \\ \langle F(u),v\rangle_{(\mathcal{V}^*,\mathcal{V})}+b(v,p)=f(v) \\ b(u,q) \qquad\qquad\qquad\qquad\;\; = g(q). \end{cases}$$

Following Hughes, Franca and Balestra (1986), Demkowicz (2006), and Xu and Zikatanov (2003), we introduce the Hilbert space $\mathcal{H} = \mathcal{V}\times\mathcal{Q}$ with the norm graph:

$$\begin{cases} \mathcal{H}\ni\mathfrak{w} &= \{u,q\} \\ \|\mathfrak{w}\|^2_{\mathcal{H}} &= \|v\|^2_{\mathcal{V}}+\|q\|^2_{\mathcal{Q}}, \end{cases}$$

the bilinear form $\mathfrak{K} : \mathcal{H}\times\mathcal{H}\to\mathbb{R}$ and the linear functional $\mathfrak{f} : \mathcal{H}\to\mathbb{R}$, $\mathfrak{f}\in\mathcal{H}^*$:

$$\begin{aligned} \mathfrak{K}(u,p;v,q) &= a(u,v)+b(v,p)+b(u,q), \\ \mathfrak{f}(u,q) &= f(v)+g(q), \end{aligned} \tag{1}$$

where we equip $\mathcal{H}^*$ with the norm $\|\cdot\|_{\mathcal{H}^*}$ given by

$$\|\mathfrak{f}\|^2_{\mathcal{H}^*} = \|f\|^2_{\mathcal{V}^*}+\|g\|^2_{\mathcal{Q}^*}.$$

Problem $\mathcal{SP}$ can be reformulated as

$$\left\{ \begin{array}{c} \text{Find } \ \mathfrak{u} \in \mathcal{H} \ \text{ such that for all } \mathfrak{v} \in \mathcal{H} \\ \mathfrak{K}(\mathfrak{u}, \mathfrak{v}) = \mathfrak{f}(\mathfrak{v}). \end{array} \right. \tag{2}$$

Existence and uniqueness of solutions to problems of type (2) is guaranteed provided the following conditions hold for all $\mathfrak{u}, \mathfrak{v} \in \mathcal{H}$

$$\mathfrak{K}(\mathfrak{w}, \mathfrak{v}) \leq C_1 \|\mathfrak{w}\|_{\mathcal{H}} \|\mathfrak{v}\|_{\mathcal{H}} \tag{3a}$$

$$\sup_{\mathfrak{v} \in \mathcal{H} \setminus \{0\}} \frac{\mathfrak{K}(\mathfrak{w}, \mathfrak{v})}{\|\mathfrak{v}\|_{\mathcal{H}}} \geq C_2 \|\mathfrak{w}\|_{\mathcal{H}}, \tag{3b}$$

$$\sup_{\mathfrak{w} \in \mathcal{H} \setminus \{0\}} \frac{\mathfrak{K}(\mathfrak{w}, \mathfrak{v})}{\|\mathfrak{w}\|_{\mathcal{H}}} \geq C_2 \|\mathfrak{v}\|_{\mathcal{H}}; \tag{3c}$$

for some positive constants C_1, C_2.

Remark 2.1. *Requirements* (3) *are known as the Babuška conditions and can be shown to be equivalent to the Brezzi conditions which essentially are (i) continuity conditions (of type* (3a)*) on $a(\cdot,\cdot), b(\cdot,\cdot)$, (ii) a condition of type* (3b) *for $b(\cdot,\cdot)$ and (iii) a coercivity condition on $a(\cdot,\cdot)$ (Xu and Zikatanov 2003, Demkowicz 2006). In the following we find it convenient to work with the Babuška conditions.*

Consider now the finite dimensional spaces $\mathcal{V}_h \subset \mathcal{V}$ and $\mathcal{Q}_h \subset \mathcal{Q}$ with bases $\{\psi_i\}_{1 \leq i \leq n}$ and $\{\phi_j\}_{1 \leq j \leq m}$, respectively. Moreover, we denote by $\mathcal{H}_h$ and its dual $\mathcal{H}_h^*$ the spaces

$$\mathcal{H}_h = \mathcal{V}_h \times \mathcal{Q}_h, \quad \mathcal{H}_h^* = \mathcal{V}_h^* \times \mathcal{Q}_h^* \ .$$

Variational formulation (2) restricted to the finite dimensional space $\mathcal{H}_h$ reads

$$\left\{ \begin{array}{c} \text{Find } \ \mathfrak{u}_h \in \mathcal{H}_h \ \text{ such that for all } \mathfrak{v}_h \in \mathcal{H}_h \\ \mathfrak{K}_h(\mathfrak{u}_h, \mathfrak{v}_h) = \mathfrak{f}_h(\mathfrak{v}_h). \end{array} \right. \tag{4}$$

where $\mathfrak{K}_h(\cdot,\cdot)$ is a bilinear form on $\mathcal{H}_h \times \mathcal{H}_h$ and $f_h(\cdot)$ is a continuous linear form on $\mathcal{H}_h$.
In the following we assume that the Babuška conditions (3) hold for the bilinear form $\mathfrak{K}_h(\cdot,\cdot)$. This allows us to derive the *a priori* error estimate

$$\|\mathfrak{u} - \mathfrak{u}_h\|_{\mathcal{H}} \leq \left(1 + \frac{C_1}{C_2}\right) \min_{\mathfrak{v}_h \in V_h} \|\mathfrak{u} - \mathfrak{v}_h\|_{\mathcal{H}}. \tag{5}$$

Remark 2.2. *We shall be assuming that the variational formulations introduced above are weak formulations of a system of partial differential equations defined on some open subset Ω of $\mathbb{R}^d$. Then the Hilbert spaces are spaces of real-valued functions defined on Ω, while $\mathcal{V}_h, \mathcal{Q}_h$ are finite element spaces, spanned by basis functions defined on a subdivision Ω_h of Ω. Replacing $\mathfrak{v}_h$ by the interpolant of $\mathfrak{u}$ on Ω_h and using standard interpolation error estimates we can derive a priori bounds of the form*

$$\|\mathfrak{u} - \mathfrak{u}_h\|_{\mathcal{H}} \leq C(\mathfrak{u}) C(h),$$

which are very useful in informing our approach to designing stopping criteria.

For the choice (1), the weak formulation (4) gives rise to a linear system of equations

$$K\mathbf{u} = \mathbf{f},$$

where the matrix K has the 2-by-2 block structure

$$K = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}$$

with

$$A_{ij} = a(\psi_j, \psi_i), \quad B_{kj} = b(\psi_j, \phi_k), \qquad i, j = 1 \cdots n, k = 1 \cdots m.$$

Let us examine the discrete setting further. Note first that there is an isomorphism Π_h between $\mathbb{R}^{n+m}$ and $\mathcal{H}_h$ defined via

$$\Pi_h \mathbf{w} = \Pi_h \begin{pmatrix} \mathbf{v} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \mathbf{v}_i \psi_i \\ \sum_{j=1}^m \mathbf{q}_j \phi_j \end{pmatrix} = \begin{pmatrix} v_h \\ q_h \end{pmatrix} = \mathfrak{w}_h.$$

In particular, since

$$\|v_h\|^2_{\mathcal{V}_h} = \mathbf{v}^T V \mathbf{v} = \|\mathbf{v}\|^2_V, \quad \|q_h\|^2_{\mathcal{Q}_h} = \mathbf{q}^t Q \mathbf{q} = \|\mathbf{q}\|^2_Q,$$

where $V \in \mathbb{R}^{n \times n}$ and $Q \in \mathbb{R}^{m \times m}$, the finite dimensional Hilbert spaces $(\mathcal{V}_h, \|\cdot\|_{\mathcal{V}_h})$, $(\mathcal{Q}_h, \|\cdot\|_{\mathcal{Q}_h})$ are represented, respectively, by $(\mathbb{R}^n, \|\cdot\|_V)$, $(\mathbb{R}^m, \|\cdot\|_Q)$. Therefore, the space $\mathcal{H}_h$ can be represented by $\mathbb{R}^{n+m}$ with norm $\|\cdot\|_H$ where $H \in \mathbb{R}^{(n+m)\times(n+m)}$ is given by

$$H = \begin{bmatrix} V & 0 \\ 0 & Q \end{bmatrix}.$$

The dual space $\mathcal{H}_h^*$ can be shown to be represented by $\mathbb{R}^{n+m}$ with norm $\|\cdot\|_{H^{-1}}$.
Finally, we have the following discrete representation

$$\mathfrak{K}_h(\mathfrak{u}_h, \mathfrak{v}_h) = \mathbf{v}^T K \mathbf{u} \quad \forall \mathfrak{u}_h, \mathfrak{v}_h \in \mathcal{H}_h,$$

which allows us to write the continuous stability conditions (3) as

$$\max_{\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T K \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \leq C_1 \tag{6a}$$

$$\min_{\mathbf{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \max_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\mathbf{w}^T K \mathbf{v}}{\|\mathbf{w}\|_H \|\mathbf{v}\|_H} \geq C_2 \tag{6b}$$

which is equivalent to uniform conditioning of K with respect to the norm induced by H:

$$\|K\|_{H,H^{-1}} \leq C_1, \quad \|K^{-1}\|_{H^{-1},H} \leq C_2^{-1},$$

or, $\kappa_H(K) \leq C_1/C_2$. We point out that both C_1 and C_2 are constants independent of h and, thus, independent of n and m.

3 Stopping criteria

Conditions (6) are sufficient for the main theorem in (Arioli et al. 2005) to apply:

Theorem 3.1. *Let* $\mathfrak{u}$ *be the solution of the weak formulation* (2) *and let* $\mathbf{u}, \mathfrak{u}_h = \Pi_h \mathbf{u}$ *satisfy*

$$K\mathbf{u} = \mathbf{f}; \qquad \frac{\|\mathfrak{u} - \mathfrak{u}_h\|_{\mathcal{H}}}{\|\mathfrak{u}_h\|_{\mathcal{H}}} \leq C(h).$$

Then $\tilde{\mathfrak{u}}_h = \Pi_h \tilde{\mathbf{u}}$ *satisfies*

$$\frac{\|\mathfrak{u} - \tilde{\mathfrak{u}}_h\|_{\mathcal{H}}}{\|\tilde{\mathfrak{u}}_h\|_{\mathcal{H}}} \leq \tilde{C}(h) = O(C(h))$$

if

$$\frac{\|\mathbf{f} - K\tilde{\mathbf{u}}\|_{H^{-1}}}{\|\tilde{\mathbf{u}}\|_H} \leq \eta C(h) C_2, \tag{7}$$

for some $\eta \in (0,1)$.

Remark 3.1. *This result means that one may replace the finite element solution* $\mathfrak{u}_h$ *by an approximation* $\tilde{\mathfrak{u}}_h$ *constructed by an iterative method provided the* H^{-1}*-norm of the residual* $\mathbf{r} = \mathbf{f} - K\tilde{\mathbf{u}}$ *is of the same order as the finite element error.*

In the following, we consider in greater detail the application of the above criterion to saddle-point systems, both in a linear and nonlinear setting.

3.1 The linear case

Unlike the positive-definite case considered in (Arioli et al. 2005), there is no obvious solution, or iterative method, that would allow for the approximation of $\|\mathbf{r}\|_{H^{-1}}$ in an indefinite context. In fact, it appears that this may have to be computed by solving a linear system with coefficient matrix H. Fortunately, this is a procedure that is included already in some preconditioned iterative methods.

3.1.1 Symmetric indefinite problems

It is an established fact that symmetric saddle-point problems arising from the stable finite element discretization of a system of partial differential equations are rather amenable to iterative treatment in the sense that they come equipped with optimal preconditioners. We quote here a general result from (Loghin and Wathen 2004) that expresses this fact.

Theorem 3.2. *Let* (6) *hold. Then*

$$\|H^{-1}K\|_H = \|KH^{-1}\|_{H^{-1}} \leq C_1, \tag{8a}$$
$$\|K^{-1}H\|_H = \|HK^{-1}\|_{H^{-1}} \leq C_2^{-1}. \tag{8b}$$

While the form of (8) is useful when we consider the nonsymmetric case, we note here that one can write the above bounds as a bound on the 2-norm condition number of K preconditioned centrally by the norm

$$\kappa_2(H^{-1/2}KH^{-1/2}) \leq \frac{C_1}{C_2}.$$

This suggests that an iterative method such as the Minimum Residual method (MINRES) will converge in a number of steps independent of the size of the problem. Furthermore, the residual computed by this method is in fact measured in the right norm: $\|\cdot\|_{H^{-1}}$. Hence, one can easily incorporate in this approach bound (7). This we carry out in our numerics section.
We note here that there is a significant amount of research devoted to the analysis of norm-based preconditioners for symmetric saddle-point problems and derivation of bounds of type (8). Some of the problems considered come from ground-water flow applications (Bramble and Pasciak 1988), (Bramble and Pasciak 1997), (Chen, Ewing and Lazarov 1996), (Glowinski and Wheeler 1988), (Rusten and Winther 1993), (Vassilevski and Wang 1992), Stokes flow (Cahouet and Chabard 1988), (Wathen and Silvester 1993), (Silvester and Wathen 1994), (Fischer, Wathen and

Silvester 1998), (Chen and Strikwerda 1999), elasticity (Glowinski and Pironneau 1979),(Arnold, Falk and Winther 1997), (Klawonn 1998), (Brown, Jimack and Mihajlovic 2000), (Mihajlovic and Silvester 2002), magnetostatics (Perugia and Simoncini 1998), (Perugia, Simoncini and Arioli 1999) etc.

3.1.2 Nonsymmetric indefinite problems

While convergence of iterative methods for nonsymmetric problems is not fully understood, bounds such as (8) are clearly attractive in a preconditioning context. They guarantee that for $H^{-1/2}KH^{-1/2}$ both the singular values and the absolute values of the eigenvalues are bounded from below and above. This means that the use of the norm as a preconditioner can be recommended also in the nonsymmetric case. The general approach, as suggested by the form of (8) is to employ an iterative solver in the H-inner product with left preconditioner H. The resulting algorithm is equivalent to employing an Euclidean inner-product and system matrix $H^{-1/2}KH^{-1/2}$ and output a residual measured in the norm $\|\cdot\|_{H^{-1}}$ which is what we want to monitor. We carry out this kind of procedure in the case of the Generalized Minimum Residual method (GMRES).

3.2 The nonlinear case

In the nonlinear case the approximation of the solutions by mixed or mixed-hybrid methods in combination with the linearization of the operator by a Newton method or a Picard approach, yields a sequence of finite dimensional problems of type (4), generally nonsymmetric, each of which satisfy the stability conditions (6). Writing the approximation of problem $\mathcal{NSP}$ as

$$\mathcal{F}(\mathbf{w}) = 0$$

after linearization, we want to solve

$$K_k \mathbf{w}_{k+1} = \mathbf{g}_k, \tag{9}$$

where

$$\mathbf{w}_{k+1} = \begin{pmatrix} \mathbf{u}_{k+1} \\ \mathbf{p}_{k+1} \end{pmatrix} \quad \text{and} \quad \mathbf{g}_k = \begin{pmatrix} \mathbf{f}_k \\ 0 \end{pmatrix}.$$

In practice, an inner-outer iteration process is the method of choice for large problems, with an inner linear solve, typically an iterative process, and an outer nonlinear update. A popular approach is the Newton-Krylov procedure, where the outer Newton iteration uses an inner iterative procedure of Krylov type to solve the linear system (9). The efficiency of such methods relies on the choice of inner stopping criteria. This was recognized by Dembo et al. (Dembo, Eisenstat and Steihaug 1982). We review briefly their result and adapt it to the finite element context.

3.2.1 Nonlinear stopping criteria

Let us assume that we seek the solution of problem (9) via an iterative routine in which at every step j we compute or estimate a residual

$$\mathbf{r}_j = \mathbf{g}_k - K_k \mathbf{w}_{k+1}.$$

Since initially system (9) approximates poorly the equation $\mathcal{F}(\mathbf{w}) = 0$ one can compute only a coarse approximation to the solution $\mathbf{w}_{k+1}$. As convergence of the outer iteration improves one would want to improve the quality of $\mathbf{w}_{k+1}$. One way of achieving this is through the use of of the following inner stopping criterion (Dembo et al. 1982)

$$\frac{\|\mathbf{r}_j\|}{\|\mathcal{F}(\mathbf{w}_k)\|} \leq c\|\mathcal{F}(\mathbf{w}_k)\|^q. \tag{10}$$

The norm employed in (10) is general, though in practice the standard Euclidean norm is employed. This is wasteful in a finite element context. In particular, given the result of Thm 3.1, we propose to evaluate the above criterion in the relevant norm $\|\cdot\|_{H^{-1}}$

$$\frac{\|\mathbf{r}_j\|_{H^{-1}}}{\|\mathcal{F}(\mathbf{w}_k)\|_{H^{-1}}} \leq c\|\mathcal{F}(\mathbf{w}_k)\|^q_{H^{-1}}, \tag{11}$$

Criterion (11) is to be combined with criterion (7). Thus, while (7) is not satisfied, one employs at each nonlinear step an iterative method with criterion (11). Moreover, the choice of c, q in (11) needs to be related to $C(h)$ in (7). Thus, if the problem is large, one can compute a satisfactory solution in just a few iterations, without the need to attain a very small order for the nonlinear residual, which in a finite element context has no relevant meaning. A typical algorithm for solving (9) is outlined below:

$k = 0$, choose $\mathbf{w}_k$, $\mathbf{r}_k = -\mathcal{F}_k(\mathbf{w}_k)$ tol_out $:= \eta C(h)C_2$

while $\|\mathbf{r}_k\|_{H^{-1}}/\|\mathbf{w}_k\|_H \geq$ tol_out

  $\mathbf{w}^0 = \mathbf{w}_n$, $\mathbf{r}^0 = \mathbf{r}_k$, tol_in $:= c\|\mathbf{r}^0\|^q_{H^{-1}}$

  $\mathbf{w}^j = \mathrm{GMRES}(K_k, \mathbf{g}_k, \mathbf{w}^0, \text{tol\_in}, H)$

  $k = k+1, \ \ \mathbf{w}_k = \mathbf{w}^j, \ \mathbf{r}_k = -\mathcal{F}(\mathbf{w}_k)$

end while

where the iterative routine $\mathrm{GMRES}(E, \mathbf{b}, \mathbf{x}^0, \text{tol}, H)$ applied to a matrix E computes an approximate solution $\mathbf{x}^k$ such that

$$\frac{\|\mathbf{b} - E\mathbf{x}^k\|_{H^{-1}}}{\|\mathbf{b} - E\mathbf{x}^0\|_{H^{-1}}} \leq \text{tol}.$$

A GMRES routine which uses the H^{-1}-norm in its stopping criterion is a GMRES iteration in the H-inner product preconditioned from the left by H (cf. section 3.1.2)

3.2.2 3-term GMRES

It was shown in the positive-definite case in (Arioli et al. 2005) that the GMRES method in the H-inner product with left preconditioner H is a three-term recurrence provided H is the symmetric part of K. In this case, the preconditioned system matrix is a normal matrix

$$H^{-1/2}KH^{-1/2} = I + S$$

where S is a skew-symmetric matrix. Such an implementation of GMRES is storage-free, a desirable feature in an iterative solver. One would naturally want to extend this to the indefinite case, particularly for the case where we have to solve a long sequence of problems of type (9). We show how this can be achieved for a class of nonlinear saddle-point problems.

Let K_k in (9) have the form

$$K_k = \begin{pmatrix} A_k & B_k^T \\ B_k & 0 \end{pmatrix}$$

where A_k are nonsymmetric positive-definite for all k, a standard assumption for a great variety of problems. Let us replace the sequence of problems (9) with the following sequence

$$\begin{pmatrix} A_k & B_k^T \\ B_k & -\tau M \end{pmatrix} \begin{pmatrix} \mathbf{u}_{k+1} \\ \mathbf{p}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_k \\ -\tau M \mathbf{p}_k \end{pmatrix}. \tag{12}$$

It is easy to see that this sequence converges to the same solution provided τ is sufficiently small (in fact, we require $\tau \leq \rho(M^{-1}BA_k^{-1}B^T)$).
Multiplying the second set of equations by minus one, equation (12) becomes

$$\begin{pmatrix} A_k & B_k^T \\ -B_k & \tau M \end{pmatrix} \begin{pmatrix} \mathbf{u}_{k+1} \\ \mathbf{p}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_k \\ \tau M \mathbf{p}_k \end{pmatrix} \tag{13}$$

and thus one can split the system matrix into a symmetric (positive-definite) and anti-symmetric part

$$\begin{pmatrix} A_k & B_k^T \\ -B_k & \tau M \end{pmatrix} = \begin{pmatrix} L_k & 0 \\ 0 & \tau M \end{pmatrix} + \begin{pmatrix} S_k & B_k^T \\ -B_k & 0 \end{pmatrix}.$$

It is clear now that the three-term GMRES method devised for the scalar case (Arioli et al. 2005) will work also in this case, provided we precondition the above matrix centrally with the hermitian part H_k of the modified matrix K_k

$$H_k = \begin{pmatrix} L_k & 0 \\ 0 & \tau M \end{pmatrix}.$$

Residual convergence will then be automatically measured in the norm

$$\|\mathbf{v}\|^2_{H_k^{-1}} = \|\mathbf{v}_1\|^2_{L_k^{-1}} + \tau^{-1}\|\mathbf{v}_2\|^2_{M^{-1}}.$$

It is clear that M will be chosen to be Q, while L_k will be in general equivalent to V, when not identically equal to it. Thus, during the Arnoldi process, one can monitor strictly the H^{-1}-norm of the residual. Numerical experiments indicate that the method does not affect the convergence rate of the outer iteration for τ sufficiently small, the advantage being that one can employ a short-term recurrence for the inner iteration.

4 Experiments

4.1 Test problems

We used two test problems suggested in (Berrone 2001). The first is Stokes flow in the unit square while the second is 2D Navier-Stokes flow in a cavity, which tries to mimic the behaviour of the driven-cavity flow. The problems we solved are

$$\begin{aligned} -\Delta\vec{u} + \nabla p &= \mathbf{f} && \text{in } \Omega && (14a) \\ \operatorname{div}\vec{u} &= 0 && \text{in } \Omega && (14b) \\ \vec{u}(\mathbf{x}) &= \vec{u}^*(\mathbf{x}) && \text{on } \Gamma, && (14c) \end{aligned}$$

and

$$\begin{aligned} -\varepsilon\Delta\vec{u} + (\vec{u}\cdot\nabla)\vec{u} + \nabla p &= \mathbf{f} && \text{in } \Omega && (15a) \\ \operatorname{div}\vec{u} &= 0 && \text{in } \Omega && (15b) \\ \vec{u}(\mathbf{x}) &= \vec{u}^*(\mathbf{x}) && \text{on } \Gamma, && (15c) \end{aligned}$$

both of which have the exact solution $(\vec{u}^*, p) = (u^*, v^*, p^*)$ given by

$$\begin{aligned}
u^*(x,y) &= -\frac{R_2}{2\pi} q_0(R_2,y)\,(1-\cos 2\pi q(R_1,x))\sin 2\pi q(R_2,y) \\
v^*(x,y) &= \frac{R_1}{2\pi} q_0(R_1,x)\,(1-\cos 2\pi q(R_2,y))\sin 2\pi q(R_1,x) \\
p^*(x,y) &= R_1 R_2 q_0(R_1,x) q_0(R_2,y)\sin 2\pi q(R_1,x)\sin 2\pi q(R_2,y).
\end{aligned}$$

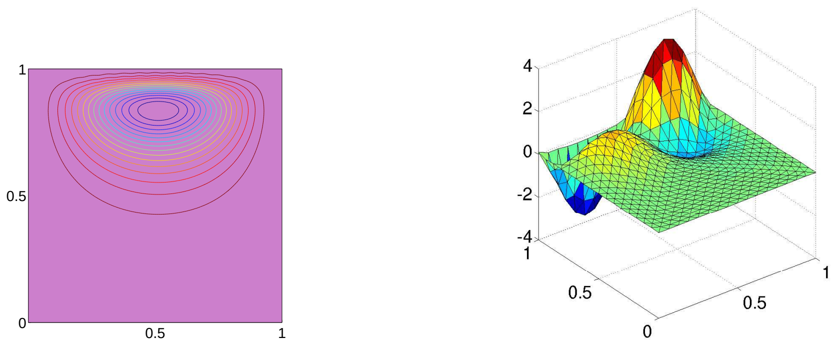
where

$$q(R,t) = \frac{e^{Rt}-1}{e^R-1}, \quad q_0(R,t) = \frac{e^{Rt}}{e^R-1}$$

and R_1, R_2 are two real constants that can be used to modify the flow behaviour. The pressure satisfies

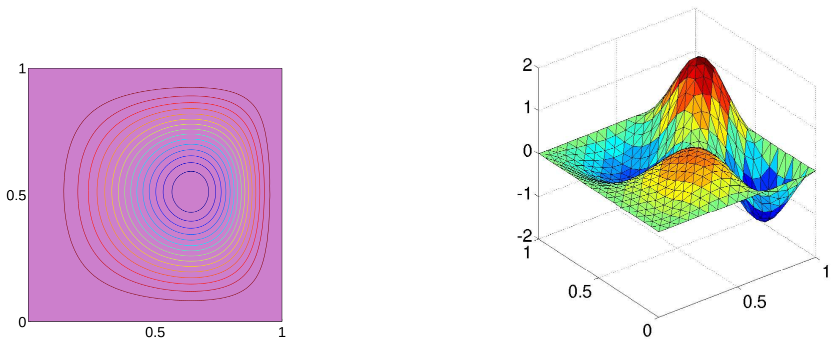
$$\int_\Omega p^* \mathrm{d}\mathbf{x} = 0, \tag{16}$$

and this is the type of condition one can use to ensure that equations (15) have a unique solution.



(a) Streamlines (b) Pressure

Figure 1: *Exact solution for* $R_1 = 0.1, R_2 = 4.2$.



(a) Streamlines (b) Pressure

Figure 2: *Exact solution for* $R_1 = 1.2, R_2 = 0.1$.

Streamlines and pressure plots for various values of R_1, R_2 are given in Figs 1, 2. The problem used in our tests corresponds to the choice $R_1 = 0.1, R_2 = 4.2$.

We solved (14), (15) using the discrete mixed formulation (4) where the space $\mathcal{H}_h \subset \mathcal{H}$ with $\mathcal{H} = H_0^1(\Omega) \times L^2(\Omega)$. In particular, we chose to work with the norm

$$\|\mathfrak{v}\|_{\mathcal{H}}^2 = \varepsilon|\vec{v}|_1^2 + \|q\|_{L^2(\Omega)}^2$$

where $\mathfrak{v} = (\vec{v}, q)$ and

$$|\vec{v}(\mathbf{x})|_{H_0^1(\Omega)} = |\vec{v}(\mathbf{x})|_1 = \left(\sum_{|\alpha|=1} \int_\Omega |D^\alpha \vec{v}(\mathbf{x})|^2 \,\mathrm{d}\mathbf{x}\right)^{1/2}.$$

Our discrete spaces V_h, Q_h were finite element spaces spanned by quadratic basis functions in the case of the velocity space and linear basis functions in the case of the pressure space.

4.2 The linear case

We chose to compare the stopping criterion (7) with both the exact finite element error and interpolation error measured in the norms inherited by the problem. We computed

(i) FE: the exact relative errors between the solution at step k and the exact continuous solution of either (14) or (15)

$$FE := \frac{\|\mathfrak{u} - \mathfrak{u}_h^k\|_{\mathcal{H}}}{\|\mathfrak{u}_h^k\|_{\mathcal{H}}};$$

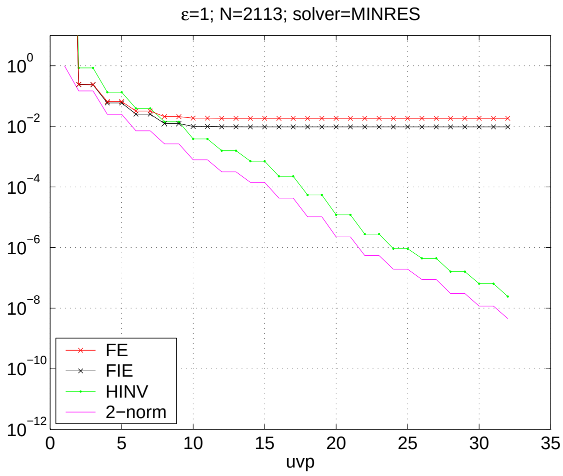
(ii) FIE: the exact relative interpolation errors

$$FIE := \frac{\|\mathfrak{u}^I - \mathfrak{u}_h^k\|_{\mathcal{H}}}{\|\mathfrak{u}_h^k\|_{\mathcal{H}}};$$

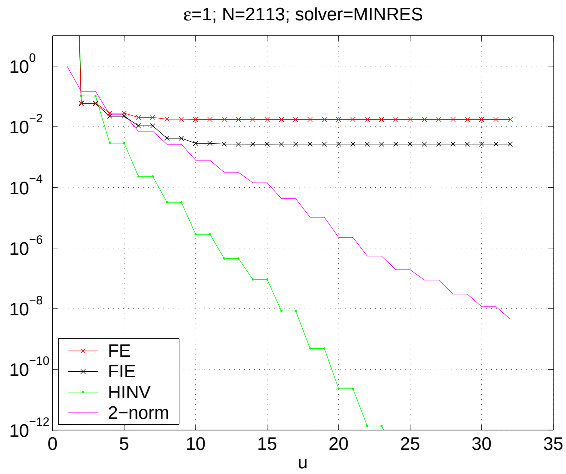
(iii) HINV: the exact H^{-1}-norm criterion (7) with C_2 estimated on a coarse mesh;

(iv) the standard 2-norm stopping criterion $\|\mathbf{r}^k\|/\|\mathbf{r}^0\|$.

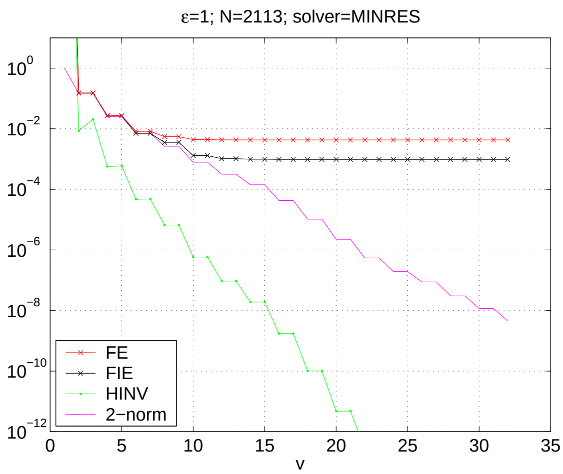
We first display in Fig. 3 the results for MINRES preconditioned with the norm in the case of the Stokes problem. In Fig. 3 (a), we plot the value of the global error while in Fig. 3 (b), (c), and (d), we plot the values of FE, FIE, HINV, and 2-norm of the residual for each one of the components of the velocity and the pressure, which in this case appear distributed unevenly. The pressure component provides the largest error, while the velocity components appear to converge faster. This will not be the case for the nonsymmetric example. Moreover, the interpolation error seems to be higher than the energy in the case of the pressure and this is also reflected globally. Our guess is that this is to do with imposing condition (16) numerically. We remark here that Theorem 3.1 is only applicable to the global solution $\mathfrak{w} = (u, v, p)$ and there is no reason to expect that the criterion should work componentwise.
We also examined the convergence of the symmetrically norm-preconditioned GMRES applied to the Navier-Stokes problem. More precisely, we computed the solution to the nonlinear problem using a Picard iteration, but displayed only the results corresponding to the last linear system solve. As before, the convergence can be examined globally or separately for the different components of the solution. These curves are shown in Figs 4, 5 for two values of the diffusion parameter: $\varepsilon = 0.1, \varepsilon = 0.01$. Again, the criterion works fine; moreover, it seems that in this case, the component convergence can be described by components of our criterion, a feature which did not work for MINRES. More precisely, plots (b), (c), (d) seem to indicate that velocity and pressure residuals can provide respective bounds for the velocity and pressure forward errors.



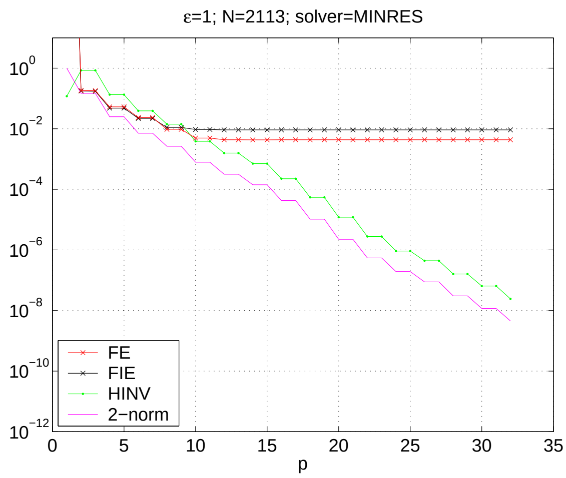
(a) Global (uvp) convergence



(b) Component u convergence



(c) Component v



(d) Component p

Figure 3: *Convergence criteria for preconditioned MINRES.*

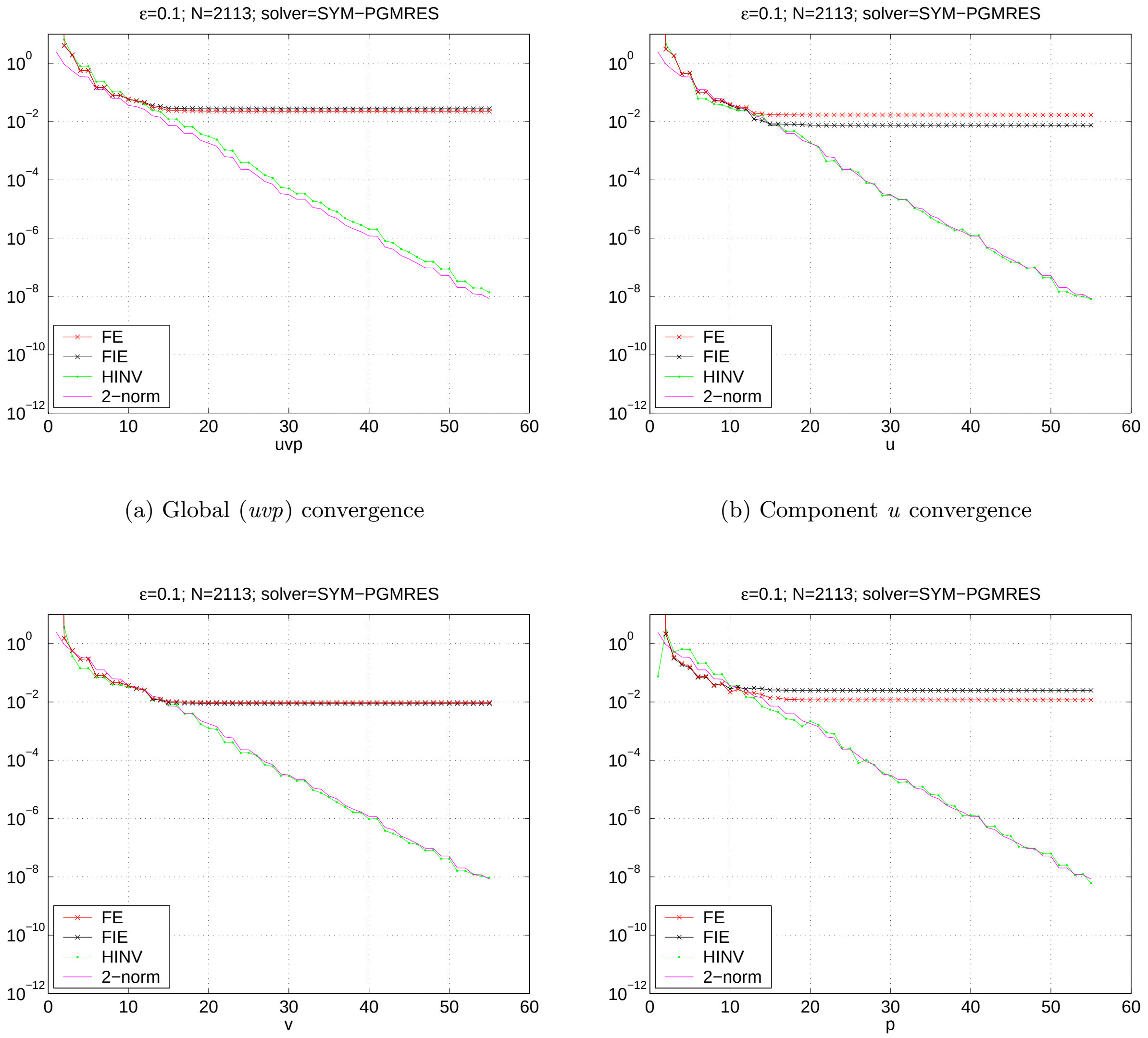
4.3 The nonlinear case

In this section, we present the results for the fully nonlinear Navier-Stokes problem (15).

First, we show the simple run of unpreconditioned GMRES as a solver for the Picard iteration applied to the nonlinear problem (15). The Picard iteration takes the form (9), but we used the modified iteration (13), in order to both exhibit its convergence properties and highlight the relevance of our stopping criteria.

The choice $c = 10^{-1}, q = 1/4$ in (10) does not affect the number of nonlinear iterations in our tests. Note that the norm in (10) is the Euclidean one. The purpose of this criterion, as described in (Dembo et al. 1982) is to make GMRES work hard only when it matters (i.e., when the residual is sufficiently small).

The nonlinear convergence history is displayed in Fig. 6. More precisely, the 2-norm of the residuals $\mathbf{r}_k$, concatenated from all nonlinear iterations (7 in this case) is plotted together with



(a) Global (uvp) convergence

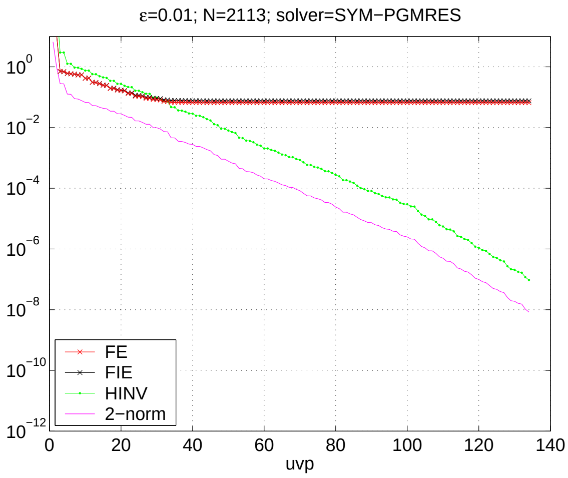
(b) Component u convergence

(c) Component v

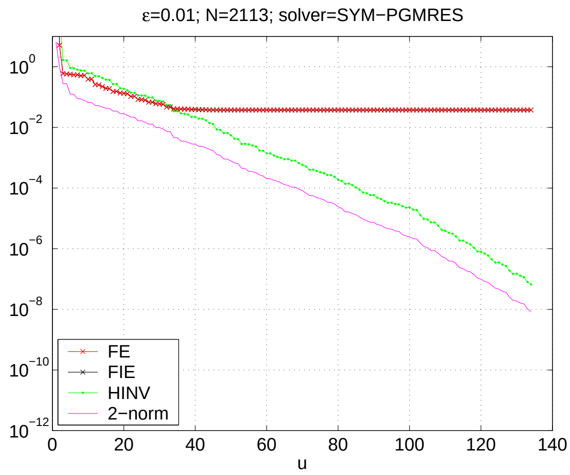
(d) Component p

Figure 4: *Convergence criteria for symmetrically preconditioned GMRES for* $\varepsilon = 0.1$.

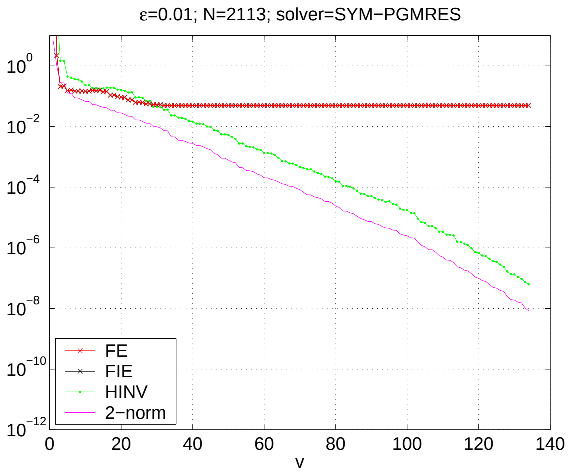
the energy-error and the H^{-1}-norm of the residual, which is computed exactly for illustration purposes. Of course, in the case of unpreconditioned GMRES it is not clear how one could derive an approximation for $\|\mathbf{r}\|_{H^{-1}}$, except through direct computation.



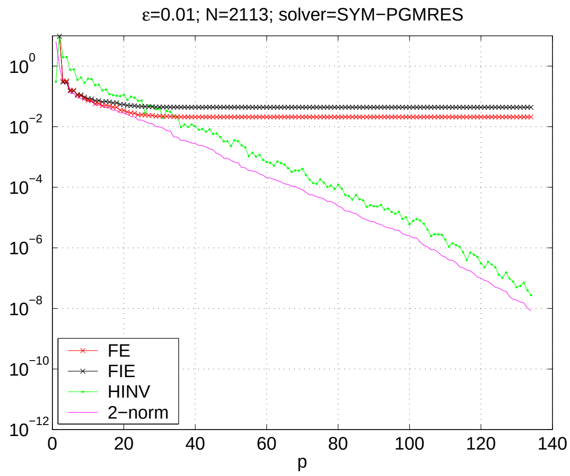
(a) Global (uvp) convergence



(b) Component u convergence

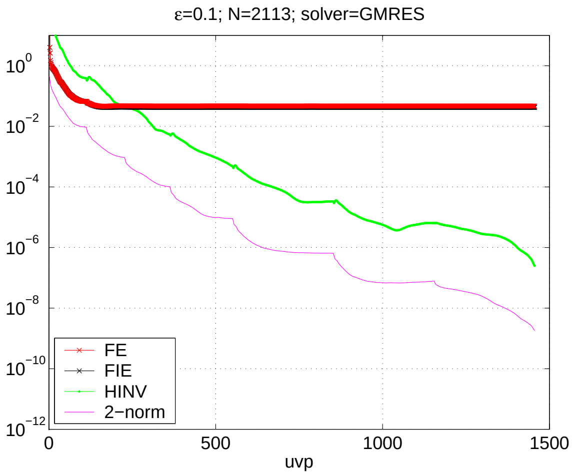


(c) Component v

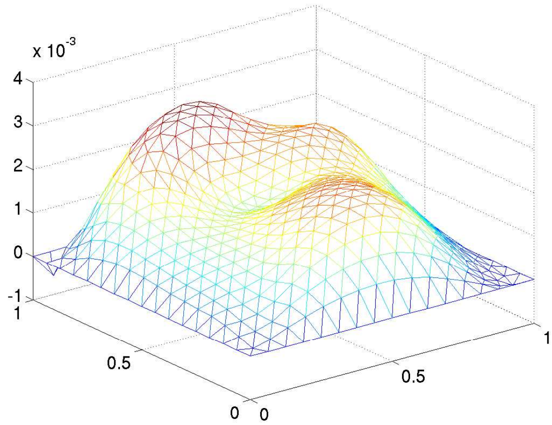


(d) Component p

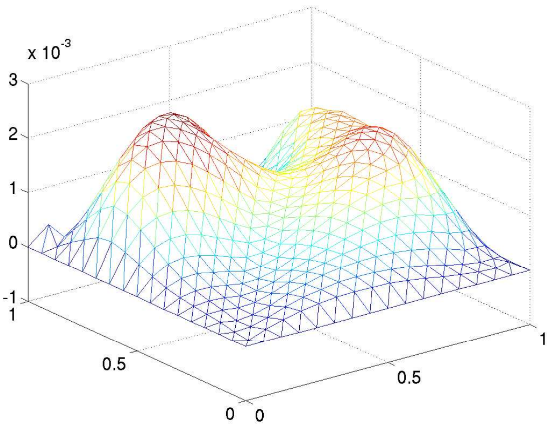
Figure 5: *Convergence criteria for symmetrically preconditioned GMRES for* $\varepsilon = 0.01$.



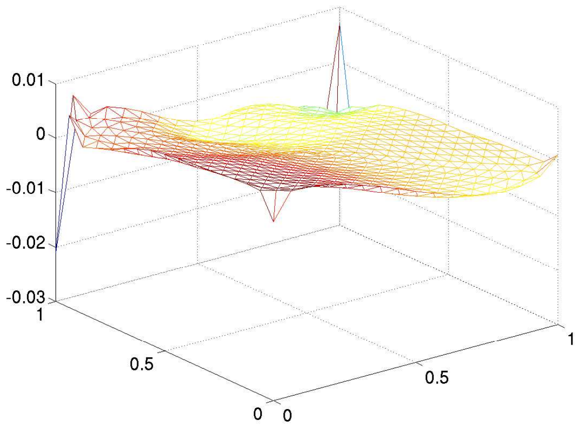
(a) Nonlinear convergence and our stopping criterion



(b) u difference



(c) v difference



(d) p difference

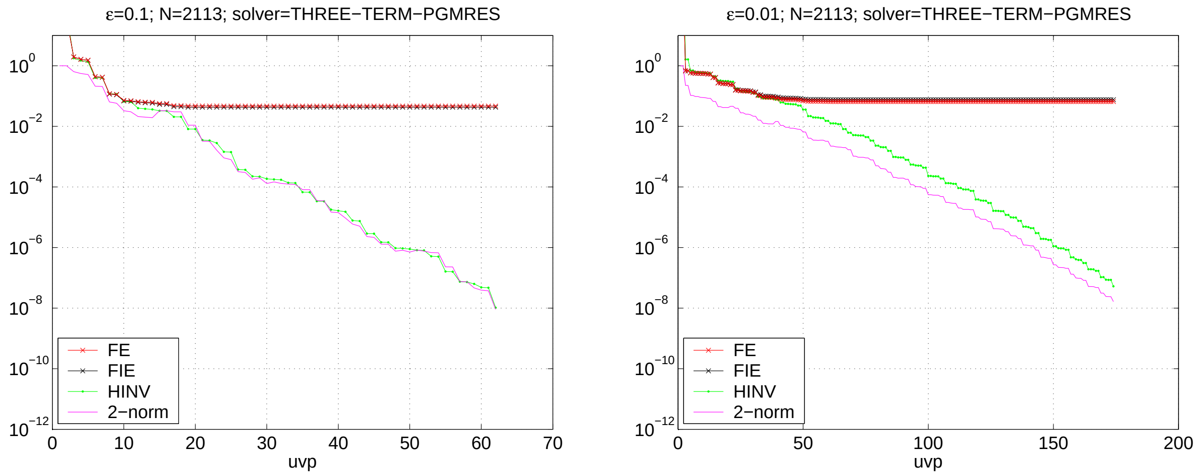
Figure 6: *Convergence criteria for unpreconditioned GMRES for* $\varepsilon = 0.1$ *and the difference between the solution converged using* (10) *and using our proposed stopping criterion* (11)*.*

We plotted also in Fig. 6 the difference between the final solution obtained using criterion (10) and that obtained by employing the H^{-1}-norm of the residual. The errors are of the order 10^{-3} for the velocities and 10^{-2} for the pressure (given solutions of order one), which indeed are of the order of the FEM error.

In Fig 7 we display the convergence of the 3-term GMRES method in the last Picard step of the nonlinear iteration of type (13) using the hermitian part to symmetrically precondition the system. We point out that the choice $\tau = 10^{-2}$ in (12) did not change the number of nonlinear iterations.

Finally, we present the results obtained using the 3-term GMRES algorithm suggested in Section 3.2.2 for solving the full nonlinear problem.

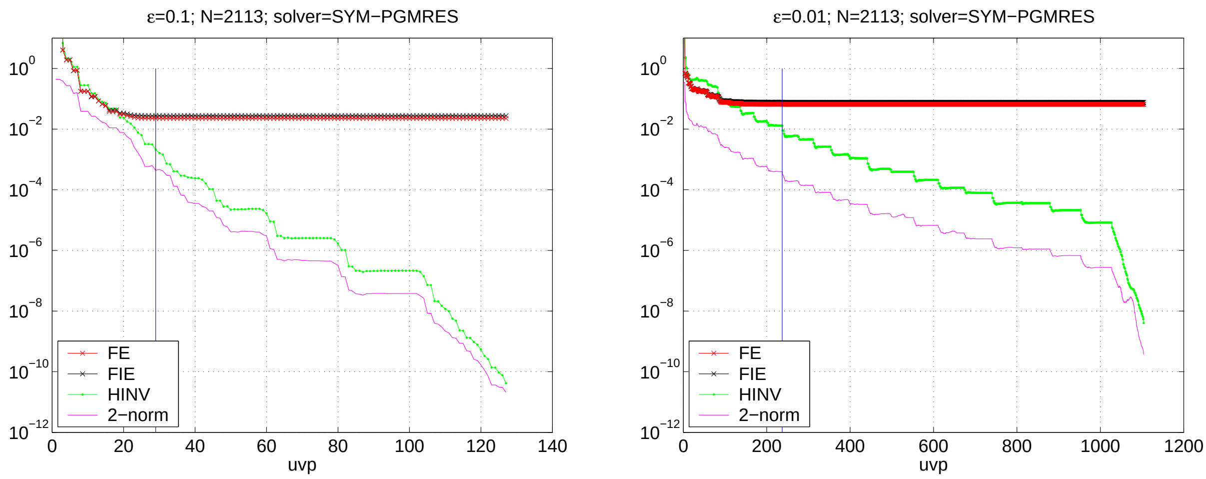
First, we note that different choices of c, q will lead to different nonlinear convergence curves. We present a typical example in Fig 8 (with $c = 1, q = 0.5$) and highlight convergence properties for several choices of parameters in Tables 1, 2. In particular, we chose to work with $c = c(h)$, for



(a) Global convergence for $\varepsilon = 0.1$

(b) Global convergence for $\varepsilon = 0.01$

Figure 7: *Convergence criteria for symmetrically preconditioned 3-term GMRES.*



(a) $\varepsilon = 0.1$

(b) $\varepsilon = 0.01$

Figure 8: *Convergence criteria for the full nonlinear problem using symmetrically-preconditioned GMRES-Picard for* $\varepsilon = 0.1, 0.01$

three values of q. We set tol_out$= 10^{-6}$ and highlighted the number of iterations needed for this 'classic' criterion compared with that suggested in the algorithm above where tol_out$= \eta C(h) C_2$. We worked with $\eta = C_2 = 1$ and $C(h) = h^2$ for $\varepsilon = 0.1$ and $C(h) = h^{3/2}$ for $\varepsilon = 0.01$; we note that this leads to a robust stopping criterion which we highlight in the vertical lines across the convergence curves in Fig 8. As expected, a small value of $c = c(h)$ leads to best performance in the H^{-1}-norm. In particular, the GMRES-Picard algorithm is most wasteful when we attempt to solve each iteration to the full FEM error level, i.e., when $c(h) = h^2$. On the other hand, when $c(h) = h^{1/2}$, the convergence is relatively robust with respect to the q parameter.

	$q = 0.25$		$q = 0.5$		$q = 0.75$	
$c(h) =$	dual	classic	dual	classic	dual	classic
$h^{1/2}$	26	83	30	125	31	175
h	31	112	33	155	37	199
h^2	45	170	49	215	52	257

Table 1: *Total number of preconditioned GMRES iterations for the full nonlinear solution of our test problem with* $\varepsilon = 0.1$ *using both the classic (l2) and dual stopping criteria*

	$q = 0.25$		$q = 0.5$		$q = 0.75$	
$c(h) =$	dual	classic	dual	classic	dual	classic
$h^{1/2}$	229	914	309	1440	405	1928
h	317	1261	405	1754	495	2205
h^2	544	1949	635	2385	722	2747

Table 2: *Total number of preconditioned GMRES iterations for the full nonlinear solution of our test problem with* $\varepsilon = 0.01$ *using both the classic (l2) and dual stopping criteria*

5 Conclusion

We showed how the results described in (Arioli et al. 2005) can be used and extended in the framework of mixed and mixed-hybrid finite-element approximation of partial differential equation systems in saddle-point form.

Moreover, we described how the dual norm of the residual can be easily used within classical Krylov methods to obtain reliable and efficient stopping criteria.

Finally, we described how to generalize these techniques to the nonlinear case, thus obtaining a considerable gain in efficiency. In particular, we showed how the use of the dual norm of the residual (essentially, the energy norm of the error) can be successfully combined with a short term recurrence GMRES in order to solve nonlinear saddle-point problems such as Navier-Stokes equations.

References

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

Arnold, D. N., Falk, R. S. and Winther, R. (1997), 'Preconditioning discrete approximations of the Reissner-Mindlin plate model', *RAIRO Model. Math. Anal. Numer.* **31**(4), 517–557.

Berrone, S. (2001), 'Adaptive discretization of stationary and incompressible navier–stokes equations by stabilized finite element methods', *Comput. Meth. Appl. Engrg.* **190**, 4435–4455.

Bramble, J. H. and Pasciak, J. E. (1988), 'A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems', *Math. Comp.* **50**, 1–17.

Bramble, J. H. and Pasciak, J. E. (1997), 'Iterative techniques for time dependent Stokes problems', *Computers Math. Applic.* **33**(1-2), 13–30.

Brezzi, F. and Fortin, M. (1991), *Mixed and Hybrid Finite Element Methods*, Vol. 15, Springer-Verlag, Berlin.

Brown, B. M., Jimack, P. K. and Mihajlovic, M. D. (2000), 'An efficient direct solver for a class of mixed finite element problems', *Applied Numerical Mathematics* **38**, 1–20.

Cahouet, J. and Chabard, J. P. (1988), 'Some fast 3D finite element solvers for the generalized Stokes problem', *Internat. J. Numer. Methods Fluids* **8**(8), 869–895.

Chen, H. H. and Strikwerda, J. C. (1999), 'Preconditioning for regular elliptic systems', *SIAM J. Num. Anal.* **37**(1), 131–151.

Chen, Z., Ewing, R. E. and Lazarov, R. (1996), 'Domain decomposition algorithms for mixed methods for second order elliptic problems', *Math. Comp.* **65**, 467–490.

Dembo, R. S., Eisenstat, S. C. and Steihaug, T. (1982), 'Inexact Newton methods', *SIAM J. Numer. Anal.* **19**(2), 400–408.

Demkowicz, L. (2006), Babuška ↔ Brezzi ??, Technical Report 06-08, ICES University of Texas at Austin. http://www.ices.utexas.edu/research/reports/2006/0609.pdf.

Fischer, B., Wathen, A. J. and Silvester, D. J. (1998), The convergence of iterative solution methods for symmetric and indefinite linear systems, *in* 'Numerical analysis 1997 (Dundee)', Longman, Harlow, pp. 230–243.

Glowinski, R. and Pironneau, O. (1979), 'Numerical methods for the first biharmonic equation and for the two-dimensional Stokes problem', *SIAM Rev.* **21**(2), 167–212.

Glowinski, R. and Wheeler, M. (1988), Domain decompostion and mixed finite element methods for elliptic problems, *in* R. G. et. al., ed., 'First International Symposium on Domain Decomposition Methods for Partial Differential Equations', SIAM, Philadelphia, pp. 144–172.

Hughes, T. J. R., Franca, L. and Balestra, M. (1986), 'A new finite element formulation for computational fluid dynamics: V. circumventing the Babuška-Brezzi condition: a stable Petrov-Galerkin formulation of the Stokes problem accommodating equal-order interpolations', *Comput. methods Appl. Mech. Engrg.* **59**, 85–99.

Klawonn, A. (1998), 'Block-triangular preconditioners for saddle-point problems with a penalty term', *SIAM J. Sci. Comput.* **19**(1), 172–184.

Loghin, D. and Wathen, A. J. (2004), 'Analysis of preconditioners for saddle-point problems', *SIAM J. Sci. Comput.* **25**(6), 2029–2049 (electronic).

Mihajlovic, M. D. and Silvester, D. J. (2002), A black-box multigrid preconditioner for the biharmonic equation, Technical Report 362, UMIST.

Perugia, I. and Simoncini, V. (1998), Optimal and quasi-optimal preconditioners for certain mixed finite element approximations, Technical Report 1098, Istituto di Analisi Numerica del C.N.R., Pavia.

Perugia, I., Simoncini, V. and Arioli, M. (1999), 'Linear algebra methods in a mixed approximation of magnetostatic problems', *SIAM J. Sci. Comput.* **21**(3), 1085–1101.

Rusten, T. and Winther, R. (1993), 'Substructure preconditioners for elliptic saddle point problems', *Math. Comp.* **60**(201), 23–48.

Silvester, D. J. and Wathen, A. J. (1994), 'Fast iterative solution of stabilised Stokes systems. II. Using general block preconditioners.', *SIAM J. Num. Anal.* **31**(5), 1352–1367.

Vassilevski, P. S. and Wang, J. (1992), 'Multilevel iterative methods for mixed finite element discretizations of elliptic problems', *Numer. Math.* **63**, 503–520.

Wathen, A. J. and Silvester, D. J. (1993), 'Fast iterative solution of stabilised Stokes systems. I. Using simple diagonal preconditioners.', *SIAM J. Num. Anal.* **30**(3), 630–649.

Xu, J. and Zikatanov, L. (2003), 'Some observations on Babuška and Brezzi theories', *Numerische Mathematik* **94**(1), 195–202.