OCCAM

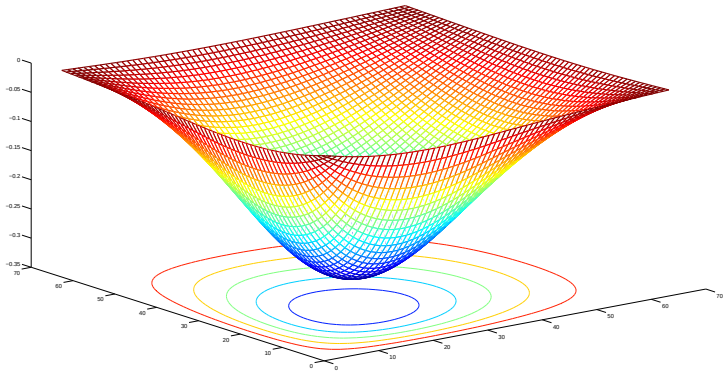
OXFORD CENTRE FOR COLLABORATIVE APPLIED MATHEMATICS

Report Number 09/15

Block triangular preconditioners for PDE constrained optimization

by

Tyrone Rees, Martin Stoll



Oxford Centre for Collaborative Applied Mathematics
Mathematical Institute
24 - 29 St Giles'
Oxford
OX1 3LB
England

Block triangular preconditioners for PDE constrained optimization

Tyrone Rees[1], Martin Stoll[2,*]

[1] *Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, OX1 3QD, United Kingdom, (tyrone.rees@comlab.ox.ac.uk)*
[2] *Oxford Centre for Collaborative Applied Mathematics, Mathematical Institute, 24 - 29 St Giles', Oxford OX1 3LB (martin.stoll@maths.ox.ac.uk)*

SUMMARY

In this paper we investigate the possibility of using a block triangular preconditioner for saddle point problems arising in PDE constrained optimization. In particular we focus on a conjugate gradient-type method introduced by Bramble and Pasciak which uses self adjointness of the preconditioned system in a non-standard inner product. We show when the Chebyshev semi-iteration is used as a preconditioner for the relevant matrix blocks involving the finite element mass matrix that the main drawback of the Bramble-Pasciak method – the appropriate scaling of the preconditioners – is easily overcome. We present an eigenvalue analysis for the block triangular preconditioners which gives convergence bounds in the non-standard inner product and illustrate their competitiveness on a number of computed examples.

KEY WORDS: Saddle point problems, linear systems, Krylov subspaces, preconditioning, PDE constrained optimization

1. Introduction

At the heart of many applications lies the solution of a linear system in saddle point form

$$\underbrace{\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}}_{\mathcal{A}} \quad x = b \tag{1}$$

where $A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{m,n}$. The properties of the blocks of $\mathcal{A}$ vary with the underlying application – for example, when looking at saddle point problems arising in the mixed finite

*Correspondence to: Oxford Centre for Collaborative Applied Mathematics, Mathematical Institute, 24 - 29 St Giles', Oxford OX1 3LB (martin.stoll@maths.ox.ac.uk)

Contract/grant sponsor: This publication is partially based on work supported by Award No. KUK-C1-013-04 , made by King Abdullah University of Science and Technology (KAUST).Publishing Arts Research Council; contract/grant number: KUK-C1-013-04

element treatment of Stokes flow problems the matrix A is symmetric and positive definite and the matrix C positive semi-definite – often $C = 0$. A comprehensive survey describing methods for solving saddle point problems – along with a list of applications that lead to such systems – is given by Benzi, Golub and Liesen in [1]. In this paper we focus on the situation where the matrix $\mathcal{A}$ has a symmetric and positive definite block A and $C = 0$, and in particular on an important class of saddle point problems arising in optimization problems with PDE constraints. It is well known that under the assumption that $B \in \mathbb{R}^{m,n}$ is of full rank the invertibility of $\mathcal{A}$ is guaranteed. The resulting system is symmetric and indefinite and hence MINRES [2] is applicable as an iterative solution algorithm. Unfortunately, the conditioning of the system matrix $\mathcal{A}$ does not allow for an iterative process to be effectively applied without using a preconditioner. Depending on the iterative method of choice one needs to design the preconditioner $\mathcal{P}$ in such a way that the preconditioned matrix $\widehat{\mathcal{A}} = \mathcal{P}^{-1}\mathcal{A}$ satisfies any required properties needed for the numerical scheme – for MINRES this means that the preconditioner has to be symmetric and positive definite in order to preserve symmetry in the preconditioned system. In this paper we consider using MINRES with a block diagonal preconditioner,

$$\mathcal{P} = \begin{bmatrix} A_0 & 0 \\ 0 & S_0 \end{bmatrix}. \tag{2}$$

We will discuss preconditioners of this type constructed for problems from PDE constrained optimization in more detail in Section 4.

The most popular Krylov subspace iterative method for symmetric matrices is the conjugate gradient method (CG) proposed by Hestenes and Stiefel in [3]. This method can, in its basic form, only be applied to systems $\mathcal{A}$ that are symmetric and positive definite, and in addition needs the preconditioner $\mathcal{P}$ also to be symmetric [4]. Various modifications to CG were proposed to make it applicable for systems of the form (1) (see [5] for a comprehensive survey). One of the most popular variants is the so-called Bramble-Pasciak CG method. This method uses a block triangular preconditioner and a non-standard inner product in which the preconditioned matrix is symmetric and, under certain conditions, positive definite. In Section 3, we discuss this method in more detail. The main emphasis in Section 3 is to show that the Bramble-Pasciak CG method is well suited for saddle point problems arising in PDE constrained optimization – we introduce these problems in Section 2.

Finally, in Section 5 we compare numerical results of MINRES with block-diagonal preconditioner to the Bramble-Pasciak CG with block-triangular preconditioner.

2. The optimal control problem

One of the application areas that leads to linear systems in saddle point form is the field of PDE constrained optimization (see [6, 7] for introductions to the area). The main aim here is to minimize a functional $J(y, u)$ subject to a partial differential equation as the constraint. In our case, the functional to be minimized over a domain $\Omega \in \mathbb{R}^d$ with $d = 2, 3$ is given by

$$J(y, u) := \frac{1}{2} \|y - \bar{y}\|^2_{L^2(\Omega)} + \frac{\beta}{2} \|u\|^2_{L^2(\Omega)}. \tag{3}$$

The state y and the control u are linked via a partial differential equation – we consider the the Poisson equation here

$$-\triangle y = f + u \text{ in } \Omega, \tag{4}$$

where f and $\bar{y}$ are some given functions, and β is a regularization constant. Note that although our theory and computations as presented here are based on this equation the ideas presented here can be applied to any elliptic PDE. In some applications the control can be bounded by so-called *box constraints*

$$u_a(x) \leq u(x) \leq u_b(x) \text{ a.e in } \Omega. \tag{5}$$

Note that we will neglect the box constraints for the rest of this paper and only talk about the unconstrained problem. Nevertheless, we want to emphasize that the techniques presented in this paper are useful when solving problems with box constraints (see [8]). We can summarize the setup in the following optimality system

$$\begin{cases} \min \frac{1}{2} \|y - \bar{y}\|^2_{L^2(\Omega)} + \frac{\beta}{2} \|u\|^2_{L^2(\Omega)} & \text{s.t.} \\ -\triangle y = f + u & \text{in } \Omega \\ y = \bar{y} & \text{on } \Gamma. \end{cases} \tag{6}$$

The discretization of the optimality system (6) can be done via the finite element method [9]. As a result we obtain a discrete optimality system given by

$$\begin{cases} \min \frac{1}{2} (y_h - \bar{y}_h)^T M (y_h - \bar{y}_h) + \frac{\beta}{2} u_h^T M u_h \quad \text{s.t.} \\ K y_h = M f_h + M u_h \end{cases} \tag{7}$$

where $M \in \mathbb{R}^{n,n}$ is a mass matrix and $K \in \mathbb{R}^{n,n}$ represents the stiffness matrix of the Poisson equation. With a Lagrange multiplier approach we obtain the following Lagrange function

$$\begin{aligned} L(y_h, u_h, \lambda_h) \quad &= \quad \frac{1}{2} (y_h - \bar{y}_h)^T M (y_h - \bar{y}_h) + \frac{\beta}{2} u_h^T M u_h \\ & \qquad -\lambda_h^T (K y_h - M f_h - M u_h). \end{aligned} \tag{8}$$

Using the standard techniques for (8), we see that the optimal solution $(y_h^*, u_h^*, \lambda_h^*)$ of the system (7) satisfies the linear system

$$\underbrace{\begin{bmatrix} M & 0 & -K^T \\ 0 & \beta M & M \\ -K & M & 0 \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} y_h^* \\ u_h^* \\ \lambda_h^* \end{bmatrix} = \begin{bmatrix} M\bar{y}_h \\ 0 \\ -M f_h \end{bmatrix} \tag{9}$$

where the saddle point matrix $\mathcal{A}$ is now a symmetric matrix with a 3×3 block structure. The system (7) needs to be solved efficiently and in the remainder of this paper we discuss methods that are well suited for this purpose. For more details on the derivation of (9) we refer to [10]. Note that for notational convenience we will switch between the 3×3 and 2×2 block structure of the saddle point matrix in (1). Typically we will introduce general statements in the 2×2 block form and properties specific to the matrix from the optimal control problem will be explained in the 3×3 block structure.

3. Bramble-Pasciak CG

The Bramble-Pasciak CG method was introduced in [11] as a solver for problems coming from the simulation of fluid flow problems and is a widely used tool in the finite element community [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. We first introduce the method based on the 2×2 saddle point structure for $\mathcal{A}$ (cf. (1)). The Bramble-Pasciak CG is based on the application of a block triangular preconditioner of the form

$$\mathcal{P} = \begin{bmatrix} A_0 & 0 \\ B & -S_0 \end{bmatrix}, \tag{10}$$

where A_0 is an approximation to the $(1,1)$-block of $\mathcal{A}$ and S_0 approximates the corresponding Schur-complement $BA^{-1}B^T$. The preconditioning blocks A_0 and S_0 are generally chosen such that a good clustering of the eigenvalues is achieved. The application of the block-triangular preconditioner is only slightly more expensive than the block-diagonal preconditioner (2) as it requires one more multiplication with B [23, 24, 11]. When applying the left preconditioner $\mathcal{P}$ to the system matrix $\mathcal{A}$ the preconditioned matrix $\widehat{\mathcal{A}} = \mathcal{P}^{-1}\mathcal{A}$ will be non-symmetric. On first sight the question arises whether such a preconditioner might be useful in practice since the symmetric system is transformed into a non-symmetric one. Amazingly, Bramble and Pasciak showed that the matrix $\widehat{\mathcal{A}}$ is symmetric and also positive definite in $\langle u, v\rangle_{\mathcal{H}} = u^T\mathcal{H}v$ with

$$\mathcal{H} = \begin{bmatrix} A - A_0 & 0 \\ 0 & S_0 \end{bmatrix} \tag{11}$$

whenever $\mathcal{H}$ defines an inner product. To see that the preconditioned matrix $\widehat{\mathcal{A}}$ is symmetric and positive definite in $\mathcal{H}$ one has to show that

$$\langle \widehat{\mathcal{A}}x, y\rangle_{\mathcal{H}} = \langle x, \widehat{\mathcal{A}}y\rangle_{\widehat{\mathcal{A}}} \Leftrightarrow \widehat{\mathcal{A}}^T\mathcal{H} = \mathcal{H}\widehat{\mathcal{A}} \qquad \forall x, y \in \mathbb{R}^n$$

holds and by using a splitting technique we can prove the positivity condition

$$\langle \widehat{\mathcal{A}}x, x\rangle_{\mathcal{H}} > 0 \Leftrightarrow \mathcal{H}\widehat{\mathcal{A}} > 0. \tag{12}$$

The details can be found in [11, 23, 25, 26, 15]. Note that it is also possible to consider right preconditioning but since the difference is typically marginal we employ left preconditioning and refer to [26, 27, 28] for more information. The implementation of a CG method based on this is given in Algorithm 1 (see [23] for details).

One of the questions that arises when implementing the Bramble-Pasciak CG is whether we can use preconditioners A_0 and S_0 where these matrices are never given explicitly. This is the case, for example, with multigrid based preconditioners, where we apply a linear process that could be written as the inverse of a matrix, but we don't generally have that matrix. It can be shown that even in this case the method is applicable since only the action of the inverse of A_0 or S_0 is needed when evaluating the inner products with $\mathcal{H}$ – i.e. the application of a fixed number of multigrid cycles is sufficient. We will demonstrate this only on one of the quantities involved and refer the interested reader to [23, 11] for further details. The quantity to compute is given by $\langle \mathcal{P}^{-1}\mathcal{A}r^{(k+1)}, p^{(k)}\rangle_{\mathcal{H}}$ which reduces to evaluating $\mathcal{H}\mathcal{P}^{-1}\mathcal{A}r^{(k+1)}$ without using A_0

Given $x^{(0)} = 0$, set $r^{(0)} = \mathcal{P}^{-1}\left(b - \mathcal{A}x^{(0)}\right)$ and $p^{(0)} = r^{(0)}$
for $k = 0, 1, \ldots$ **do**
$\alpha = \frac{\langle r^{(k)}, p^{(k)} \rangle_{\mathcal{H}}}{\langle \mathcal{P}^{-1}\mathcal{A}p^{(k)}, p^{(k)} \rangle_{\mathcal{H}}}$
$x^{(k+1)} = x^{(k)} + \alpha p^{(k)}$
$r^{(k+1)} = r^{(k)} - \alpha \mathcal{P}^{-1}\mathcal{A}p^{(k)}$
$\beta = \frac{\langle \mathcal{P}^{-1}\mathcal{A}r^{(k+1)}, p^{(k)} \rangle_{\mathcal{H}}}{\langle \mathcal{P}^{-1}\mathcal{A}p^{(k)}, p^{(k)} \rangle_{\mathcal{H}}}$
$p^{(k+1)} = r^{(k+1)} - \beta p^{(k)}$
end for

Algorithm 1: Bramble and Pasciak CG

and S_0. In more detail, using $Ar^{(k+1)} = \left[(\hat{r}_1^{(k+1)})^T \ (\hat{r}_2^{(k+1)})^T\right]^T$ we get

$$\mathcal{H}\mathcal{P}^{-1}\mathcal{A}r^{(k+1)} = \begin{bmatrix} A - A_0 & 0 \\ 0 & S_0 \end{bmatrix} \begin{bmatrix} A_0^{-1}\hat{r}_1^{(k+1)} \\ S_0^{-1}(BA_0^{-1}\hat{r}_1^{(k+1)} - \hat{r}_2^{(k+1)}) \end{bmatrix}$$

$$= \begin{bmatrix} AA_0^{-1}\hat{r}_1^{(k+1)} - \hat{r}_1^{(k+1)} \\ BA_0^{-1}\hat{r}_1^{(k+1)} - \hat{r}_2^{(k+1)} \end{bmatrix}$$

which is seen to involve only A_0^{-1}, not A_0. It can also be shown that for the evaluation of the inner product with $\mathcal{H}$ the Schur complement preconditioner S_0 does not need to be evaluated [23]. In fact the preconditioner $\mathcal{P}$ has to be applied only once per iteration and in order to evaluate the inner product with $\mathcal{H}$ we only need multiplications with blocks of the saddle point matrix (see [23] for details).

So far we have not talked about the drawbacks of the Bramble-Pasciak CG. For the algorithm to work we require $A - A_0$ to be positive definite – recall that $\mathcal{H}$ needs to define an inner product – which can only be achieved when A_0 is scaled appropriately. The scaling typically involves solving an eigenvalue estimation problem for the matrix $A_0^{-1}A$ and can only be avoided when a good knowledge of the eigenvalues of $A_0^{-1}A$ is at hand.

We now return to the 3×3 system derived in Section 2, i.e.,

$$\underbrace{\begin{bmatrix} M & 0 & -K^T \\ 0 & \beta M & M \\ -K & M & 0 \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} y_h^* \\ u_h^* \\ \lambda_h^* \end{bmatrix} = \begin{bmatrix} M\bar{y}_h \\ 0 \\ -Mf_h \end{bmatrix}.$$

The preconditioner in this case is given by

$$\mathcal{P} = \begin{bmatrix} A_0 & 0 & 0 \\ 0 & A_1 & 0 \\ -K & M & -S_0 \end{bmatrix} \tag{13}$$

and the inner product is defined by

$$\mathcal{H} = \begin{bmatrix} M - A_0 & 0 & 0 \\ 0 & \beta M - A_1 & 0 \\ 0 & 0 & S_0 \end{bmatrix}. \tag{14}$$

In this case, the condition for the applicability of the Bramble-Pasciak CG is the positivity of $M - A_0$ and $\beta M - A_1$, along with the positive definiteness of S_0. If we can find scalings for A_0 and A_1 without having to approximate an eigenvalue of $A_0^{-1}M$ or $\beta A_1^{-1}M$ the Bramble-Pasciak CG represents a well-suited method for PDE constrained optimization problems: we address this question in Section 3.1 below.

The Schur-complement of the saddle point problem (9) is given by

$$\frac{1}{\beta}M + KM^{-1}K^T. \tag{15}$$

Here, we want S_0 to represent a good approximation to the Schur complement. A strategy used by Rees, Dollar and Wathen [10] is for S_0 to be an approximation to $KM^{-1}K^T$ which means that the term $\frac{1}{\beta}M$ in (15) is neglected. The Schur-complement preconditioner is now given by $S_0 = \tilde{K}M^{-1}\tilde{K}^T$, where $\tilde{K}$ represents a fixed number of algebraic multigrid V-cycles using the `HSL` package `HSL_MI20` [29].

It is well known that the classical CG method [3] minimizes the $\mathcal{A}$-norm of the error, i.e. $\|e_k\|_{\mathcal{A}}$ where $e_k = x - x_k$, even in its preconditioned version, PCG; the method of Bramble and Pasciak with the non-standard inner product defined by $\mathcal{H}$ thus minimizes the error in the norm defined by $\mathcal{H}\widehat{\mathcal{A}}$, i.e. $\|e_k\|_{\mathcal{H}\widehat{\mathcal{A}}}$. One estimate for the error in the $\mathcal{H}\widehat{\mathcal{A}}$-norm was introduced in [30] and is given by

$$\left(\kappa_{\mathcal{H}\widehat{\mathcal{A}}}(\widehat{\mathcal{A}})^{-1}\left|\frac{\langle \mathcal{P}^T\mathcal{H}\widehat{\mathcal{A}}e_k, r_k\rangle}{\langle \mathcal{P}^T\mathcal{H}\widehat{\mathcal{A}}e_k, r_k\rangle}\right|\right)^{1/2} \leq \frac{\|e_k\|_{\mathcal{H}\widehat{\mathcal{A}}}}{\|x\|_{\mathcal{H}\widehat{\mathcal{A}}}} \leq \left(\kappa_{\mathcal{H}\widehat{\mathcal{A}}}(\widehat{\mathcal{A}})\left|\frac{\langle \mathcal{P}^T\mathcal{H}\widehat{\mathcal{A}}e_k, r_k\rangle}{\langle \mathcal{P}^T\mathcal{H}\widehat{\mathcal{A}}e_k, r_k\rangle}\right|\right)^{1/2}$$

where $\kappa_{\mathcal{H}\widehat{\mathcal{A}}}(\widehat{\mathcal{A}})$ only depends on the eigenvalues of $\mathcal{H}\widehat{\mathcal{A}}$.

3.1. Chebyshev semi-iteration

As we mentioned earlier the downside of the Bramble-Pasciak CG is that the preconditioners A_0 and A_1 need to be scaled to guarantee that $\mathcal{H}$ defines an inner product. In the case of the optimal control problem (7) we have the blocks M and βM, a mass matrix and a scaled mass matrix. An efficient preconditioner for these systems is given when a fixed number of steps of the Chebyshev-semi iteration (described below) can be applied, as shown in [31].

In [32, 33] Wathen provided eigenvalue bounds for mass matrices coming from different two or three dimensional finite elements. We are going to use these bounds to derive, first, bounds for the convergence of relaxed Jacobi (and the optimal relaxation parameter) and, second, bounds for the eigenvalues of $A_0^{-1}M$, where A_0 now represents the matrix that is used to get the n^{th} Chebyshev semi-iterate approximation. First, suppose we want to solve the linear system

$$M\hat{x} = \hat{b} \tag{16}$$

where M is the mass matrix introduced in Section 2 and $\hat{b}$ is an arbitrary right hand side. The iterates of the relaxed Jacobi method [28] are given by

$$x^{(m+1)} = (I - \omega D^{-1}M)x^{(m)} + \omega D^{-1}\hat{b}$$

where $D = \mathrm{diag}(A)$. In [32] Wathen gives bounds for the matrix $D^{-1}M$ for different finite element discretizations. These bounds are easily obtained, even for irregular meshes – if D_n is the diagonal of the nth element matrix E_n, then the smallest and largest eigenvalues are given by the smallest and largest eigenvalues of $D_n^{-1}E_n$ over all n. This is just an $\mathcal{O}(n)$ calculation, and only needs to be done once for each mesh. We consider square Q1 elements here, and in this case

$$\lambda(D^{-1}M) \in [1/4, 9/4].$$

By using this it is easy to see that the eigenvalues of the iteration matrix, $I - \omega D^{-1}M$ satisfy

$$\lambda(I - \omega D^{-1}M) \in [1 - \frac{9\omega}{4}, 1 - \frac{\omega}{4}].$$

The optimal relaxation parameter ω is computed such that the interval is symmetric, i.e., for $\omega = 4/5$ we get $\lambda(I - \omega D^{-1}M) \in [-4/5, 4/5]$ (see [33]). In summary, $\omega = 4/5$ is the optimal relaxation parameter for square Q1 elements in 2D. In 3D for the equivalent elements we have $\lambda(D^{-1}M) \in [1/8, 27/8]$ and hence the optimal relaxation parameter is $\omega = 4/7$ with $\lambda(I - \omega D^{-1}M) \in [-13/14, 13/14]$.

If we now label the iteration matrix as

$$S = I - \omega D^{-1}M,$$

then in a polynomial iterative method (see [34]) we take a linear combination of iterates $x^{(k)}$ to produce $z^{(k)} = \sum_{i=0}^{k} \alpha_i x^{(i)}$ which is, ideally, a better approximation to the solution x. We can think of the terms α_i as being coefficients of a polynomial of degree k, $p_k(z)$. Note we require $p_k(1) = 1$ for the iterates to satisfy $z^{(k)} = x$ for all k if $x^{(0)} = x$. We can write the error as

$$x - z^{(k)} = p_k(S)(x - x^{(0)}).$$

If we start with an initial guess of zero, then this reduces to

$$x - z^{(k)} = p_k(S)x,$$

and, rearranging, we can write the k^{th} iterate as

$$z^{(k)} = (I - p_k(S))x.$$

Now, from (16), we can write

$$z^{(k)} = (I - p_k(S))M^{-1}\hat{b}.$$

If we precondition M with the k^{th} polynomial iterate, then we are actually preconditioning with the matrix A_0 which is the matrix with inverse

$$A_0^{-1} = (I - p_k(S))M^{-1}.$$

For the application of the Bramble-Pasciak CG we need good eigenvalue bounds for the matrix $A_0^{-1}M$. In the case of the Chebyshev semi-iteration we have $A_0^{-1}M$ given by

$$A_0^{-1}M = I - p_k(S).$$

We choose the underlying polynomial $p_k(z)$ so that it is small on the eigenvalues and $p_k(1) = 1$. If we only know the extremal eigenvalues but nothing about the distribution of the remaining eigenvalues, then the best choice of polynomial is the scaled and shifted Chebyshev polynomials. If $\lambda(S) \in [-\rho, \rho]$,

$$p_k(S) = \widehat{T_k}(S) = \frac{T_k(S/\rho)}{T_k(1/\rho)},$$

where T_k are the Chebyshev polynomials, i.e. $T_0(t) = 1$, $T_1(t) = t$, $T_2(t) = t^2 - 1/2$, $T_3(t) = t^3 - 3/4t$, etc. This then gives the Chebyshev semi-iterative method [35, 36]. An efficient way to compute the iterates $z^{(k+1)}$ of the Chebyshev semi-iterative method is given by the three term recurrence relation

$$z^{(k+1)} = \vartheta_{k+1}(Sz^{(k)} + g - z^{(k-1)}) + z^{(k-1)}, \tag{17}$$

where $\vartheta_{k+1} = \frac{T_k(1/\rho)}{\rho T_{k+1}(1/\rho)}$ and $g = \omega D^{-1}\hat{b}$. (see Varga [34, Chapter 5]). The values ϑ_{k+1} can be computed using the bounds for eigenvalues of the iteration matrix and the recursive definition of the Chebyshev polynomials. We know that $\lambda(S) \in [-\rho, \rho]$ and so the range of $\lambda(\widehat{T_k}(S))$ can be computed as well as $\lambda(A_0^{-1}M)$. These values are tabulated for square Q1 elements in 2D (Table I) and cubic Q1 elements in 3D (Table II) below.

We now go back to the construction of preconditioners A_0 and A_1 that we want to use in the Bramble-Pasciak method based on the presented analysis of the Chebyshev semi-iteration. Recall that in order for $\mathcal{H}$ to define an inner product we need the blocks $M - A_0$ and $\beta M - A_1$ to be positive definite. We only discuss the case $M - A_0 > 0$ since $\beta M - A_1 > 0$ can be seen as a scaled version of this condition. The condition $M - A_0 > 0$ is equivalent to $A_0^{-1}M - I$ being positive definite. In fact, if all the eigenvalues of $A_0^{-1}A$ are larger then one we have positive definiteness of $M - A_0$. Using the bounds presented in Tables (I) and (II) we have a good knowledge of the minimal eigenvalue λ_{min} of $A_0^{-1}M$ and choosing any parameter γ_0 with $0 < \gamma_0 < \lambda_{min}$ guarantees the positivity of $M - \gamma_0 A_0$. Analogously, A_1 can be scaled to give the positive definiteness of $\beta M - \gamma_1 A_1 > 0$. This is a key advantage of our approach: the required eigenvalue bounds to ensure positive definiteness of $A_0^{-1}M - I$ are not commonly available when the Bramble-Pasciak method is used, whereas here we have simple a priori calculations that give precisely the tight bounds required. For notational convenience we will only use A_0 and A_1 for the rest of the paper assuming that they are scaled appropriately.

3.2. Eigenvalue analysis

Consider a general saddle point matrix (1),

$$\mathcal{A} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}.$$

Based on the results of [15], [9], we would like to precondition this using a block triangular preconditioner

$$\mathcal{P}_E = \begin{bmatrix} A & 0 \\ B & -S \end{bmatrix},$$

Table I. Upper and lower bounds for $\lambda(A_0^{-1}M)$, and ϵ_k s.t. $|1-\lambda_k| \leq \epsilon_k$, where A_0 is given by k steps of the Chebyshev semi-iteration (without scaling) for square Q1 elements (in 2D)

k	Lower bound (λ_{min})	Upper bound (λ_{max})	ϵ_k	$\frac{\epsilon_k}{\epsilon_{k-1}}$
1	0.200000000000000	1.800000000000000	0.800000000000	—
2	0.529411764705882	1.470588235294118	0.470588235294	0.58823529
3	0.753846153846154	1.246153846153846	0.246153846153	0.52307692
4	0.875486381322957	1.124513618677043	0.124513618677	0.50583657
5	0.937560975609756	1.062439024390244	0.062439024390	0.50146341
6	0.968757627532341	1.031242372467659	0.031242372467	0.50036612
7	0.984375953616112	1.015624046383888	0.015624046383	0.50009154
8	0.992187619207471	1.007812380792529	0.007812380792	0.50002288
9	0.996093764901104	1.003906235098896	0.003906235098	0.50000572
10	0.998046876862643	1.001953123137357	0.001953123137	0.50000143
11	0.999023437732830	1.000976562267170	0.000976562267	0.50000035
12	0.999511718779104	1.000488281220896	0.000488281220	0.50000008
13	0.999755859378638	1.000244140621363	0.000244140621	0.50000002
14	0.999877929687954	1.000122070312045	0.000122070312	0.50000000
15	0.999938964843807	1.000061035156194	0.000061035156	0.50000000
16	0.999969482421881	1.000030517578119	0.000030517578	0.50000000
17	0.999984741210937	1.000015258789063	0.000015258789	0.50000000
18	0.999992370605466	1.000007629394531	0.000007629394	0.50000000
19	0.999996185302733	1.000003814697268	0.000003814697	0.50000000
20	0.999998092651366	1.000001907348635	0.000001907348	0.50000000

where S is some approximation to the Schur complement $BA^{-1}B^T$. It is well known that at least in the symmetric – or in general the self-adjoint – case, the effectiveness of a preconditioner is dependent on the clustering of the eigenvalues of the preconditioned system, so we want to find the eigenvalues of the matrix $\mathcal{P}_E^{-1}\mathcal{A}$. Note that we can write $\mathcal{P}_E^{-1}$ explicitly as

$$\mathcal{P}_E^{-1} = \begin{bmatrix} A^{-1} & 0 \\ S^{-1}BA^{-1} & -S^{-1} \end{bmatrix},$$

and so

$$\mathcal{P}_E^{-1}\mathcal{A} = \begin{bmatrix} I & A^{-1}B^T \\ 0 & S^{-1}BA^{-1}B^T \end{bmatrix}.$$

Suppose that $w = [u\ v]^T \neq 0$ and λ are an eigenvector and eigenvalue respectively for the generalized eigenvalue problem $\mathcal{A}w = \lambda\mathcal{P}_E w$. Then we have that

$$\begin{bmatrix} I & A^{-1}B^T \\ 0 & S^{-1}BA^{-1}B^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix},$$

and so, if $v \neq 0$, $S^{-1}BA^{-1}B^Tv = \lambda v$, or, if $v = 0$, then we have an eigenvalue at 1. Let μ_{min} and μ_{max} denote the smallest and largest eigenvalues of $\mathcal{P}_E^{-1}\mathcal{A}$ respectively. The magnitude of these will clearly depend on our approximation to S.

Table II. Upper and lower bounds for $\lambda(A_0^{-1}M)$, and ϵ_k s.t. $|1-\lambda_k| \leq \epsilon_k$, where A_0 is given by k steps of the Chebyshev semi-iteration (without scaling) for square Q1 elements (in 3D)

k	Lower bound	Upper bound	ϵ_k	$\frac{\epsilon_k}{\epsilon_{k-1}}$
1	0.071428571428571	1.928571428571429	0.928571428571	—
2	0.242152466367712	1.757847533632288	0.757847533632	0.81614349
3	0.433470861268694	1.566529138731306	0.566529138731	0.74755028
4	0.597147975231673	1.402852024768327	0.402852024768	0.71108791
5	0.720776486124292	1.279223513875708	0.279223513875	0.69311681
6	0.808846507572246	1.191153492427754	0.191153492427	0.68458952
7	0.869897818711463	1.130102181288537	0.130102181288	0.68061629
8	0.911689150016541	1.088310849983459	0.088310849983	0.67878070
9	0.940130893927575	1.059869106072427	0.059869106072	0.67793601
10	0.959435805298037	1.040564194701956	0.040564194701	0.67754802
11	0.972523033141943	1.027476966858060	0.027476966858	0.67736995
12	0.981390172808929	1.018609827191073	0.018609827191	0.67728826
13	0.987396479797611	1.012603520202398	0.012603520202	0.67725079
14	0.991464472578415	1.008535527421550	0.008535527421	0.67723360
15	0.994219521276648	1.005780478723391	0.005780478723	0.67722572
16	0.996085332018970	1.003914667981061	0.003914667981	0.67722210
17	0.997348906791884	1.002651093208068	0.002651093208	0.67722044
18	0.998204627482791	1.001795372516598	0.001795372517	0.67721968
19	0.998784139008908	1.001215860991333	0.001215860991	0.67721933
20	0.999176595616630	1.000823404383702	0.000823404383	0.67721917

Now consider our application, where we have the system (9), coupled with the 'ideal' preconditioner

$$\mathcal{P} = \begin{bmatrix} M & 0 & 0 \\ 0 & \beta M & 0 \\ -K & M & -KM^{-1}K \end{bmatrix}.$$

Then, in this idealized case, applying the general results above we see that if λ is an eigenvalue of the preconditioned system either $\lambda = 1$ or

$$(KM^{-1}K)^{-1}(KM^{-1}K + \frac{1}{\beta}M)v = \lambda v.$$

Corollary 3.3 in [10] gives us that, in the second case,

$$\frac{1}{\beta}\alpha_1 h^4 + 1 \leq \lambda \leq \frac{1}{\beta}\alpha_2 + 1, \tag{18}$$

where α_1 and α_2 are constants independent of h and independent of β. These constants depend only on the discretization used, and are not large in general. For example, in 2D for **Q1** square elements the constants are $\alpha_1 = 1/1296$ and $\alpha_2 = 1/4\pi^2$.

We have given the eigenvalue bounds for the ideal case, but in practice we will approximate the Schur complement as $\widetilde{K}M^{-1}\widetilde{K}$, where $\widetilde{K}$ is an approximation to the stiffness matrix. To

consider this effect, suppose that there exists a constant η such that $\|\widetilde{K}^{-1}K - I\| \leq \eta$. Then, using a duality argument from Braess and Peisker in the case of the biharmonic equation [37, Section 4], we obtain that the generalized Rayleigh quotients satisfy

$$(1-\eta)^2 \leq \frac{w^T K^T M^{-1} K w}{w^T \widetilde{K}^T M^{-1} \widetilde{K} w} \leq (1+\eta)^2 \quad \forall\ w. \tag{19}$$

Thus, we can write

$$\frac{w^T K M^{-1} K + \frac{1}{\beta} M w}{w^T \widetilde{K} M^{-1} \widetilde{K} w} = \frac{w^T K M^{-1} K + \frac{1}{\beta} M w}{w^T K M^{-1} K w} \cdot \frac{w^T K M^{-1} K w}{w^T \widetilde{K} M^{-1} \widetilde{K} w}, \tag{20}$$

and so eigenvalues of

$$(\widetilde{K} M^{-1} \widetilde{K})^{-1} (K M^{-1} K + \frac{1}{\beta} M) \tag{21}$$

are bounded above and below by the product of the maximal and minimal values of the terms on the right hand side of (20). Therefore, using (18) and (19), we have that these eigenvalues lie in the interval $[(1-\eta)^2, (1+\eta)^2(\frac{1}{\beta}\alpha_2 + 1)]$.

We now leave the idealized case and return to the 2×2 block structure to consider the more general case of (1) preconditioned by a lower block triangular preconditioner (10), where the (1,1) block of the preconditioner is now some approximation to A. To obtain eigenvalue bounds for the preconditioned system in this case we first consider the symmetric and positive-definite block-diagonal preconditioner

$$\mathcal{P}_D = \begin{bmatrix} A_0 & 0 \\ 0 & S_0 \end{bmatrix}$$

and the generalized eigenvalue problem $\mathcal{A}u = \lambda \mathcal{P} u$. Using $v = \mathcal{P}^{1/2} u$ we get $\mathcal{P}_D^{-1/2} \mathcal{A} \mathcal{P}_D^{-1/2} v = \lambda \mathcal{P}_D^{-1/2} \mathcal{P} \mathcal{P}_D^{-1/2} v$ where v is scaled such that $\|v\| = 1$. This gives rise to a new generalized eigenvalue problem $\tilde{\mathcal{A}} v = \lambda \tilde{P} v$ with

$$\tilde{\mathcal{A}} \equiv \begin{bmatrix} A_0^{-1/2} A A_0^{-1/2} & A_0^{-1/2} B^T S_0^{-1/2} \\ S_0^{-1/2} B A_0^{-1/2} & -S_0^{-1/2} C S_0^{-1/2} \end{bmatrix} \equiv \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & -\tilde{C} \end{bmatrix}$$

and

$$\tilde{\mathcal{P}} \equiv \begin{bmatrix} I & 0 \\ S_0^{-1/2} B A_0^{-1/2} & -I \end{bmatrix} \equiv \begin{bmatrix} I & 0 \\ \tilde{B} & -I \end{bmatrix}.$$

This eigenvalue problem can hence be reformulated as

$$\tilde{A} v_1 + \tilde{B}^T v_2 = \lambda v_1 \tag{22}$$

$$\tilde{B} v_1 - \tilde{C} v_2 = \lambda \tilde{B} v_1 - \lambda v_2. \tag{23}$$

Recall that in our case $C = 0$ and hence the term $\tilde{C} v_2$ will be equal to zero. Assuming now that $v_2 = 0$ yields $\tilde{A} v_1 = \lambda v_1$ with λ an eigenvalue of the symmetric positive definite matrix $\tilde{A}$ if only $(1-\lambda)\tilde{B} v_1 = 0$. The case $v_1 = 0$ implies that $\tilde{B}^T v_2 = 0$, but since $\tilde{B}^T$ is of full rank $v_2 = 0$. Thus, we assume that $v_1 \neq 0$ and $v_2 \neq 0$. If we multiply (22) on the left by the conjugate transpose v_1^*, we obtain

$$v_1^* \tilde{A} v_1 + v_1^* \tilde{B}^T v_2 = \lambda v_1^* v_1 \Longrightarrow v_1^* \tilde{B}^T v_2 = \lambda v_1^* v_1 - v_1^* \tilde{A} v_1. \tag{24}$$

The conjugate transpose of (23) multiplied on the right by v_2 gives

$$v_1^* \tilde{B}^T v_2 = \bar{\lambda} v_1^* \tilde{B}^T v_2 - \bar{\lambda} v_2^* v_2. \tag{25}$$

Using (24) gives for (25)

$$(1-\bar{\lambda})(\lambda v_1^* v_1 - v_1^* \tilde{A} v_1), = -\bar{\lambda} v_2^* v_2 \tag{26}$$

which can be further rewritten

$$(\lambda - |\lambda|^2) \|v_1\|^2 + (\bar{\lambda} - 1) v_1^* \tilde{A} v_1 + \bar{\lambda} \|v_2\|^2 = 0. \tag{27}$$

We analyze (27) further knowing that λ is real, see (12), and under the assumption that $\|v\| = 1$ with $\|v_2\|^2 = 1 - \|v_1\|^2$ and get

$$(\lambda - \lambda^2) \|v_1\|^2 + \lambda v_1^* \tilde{A} v_1 + \lambda - \lambda \|v_1\|^2 - v_1^* \tilde{A} v_1 = 0. \tag{28}$$

We then get for λ

$$\lambda_\pm = \frac{1 + v_1^* \tilde{A} v_1 \pm \sqrt{(1 + v_1^* \tilde{A} v_1)^2 - 4 \|v_1\|^2 v_1^* \tilde{A} v_1}}{2 \|v_1\|^2}. \tag{29}$$

We now want to find bounds for λ based on the eigenvalues of $\tilde{A}$. Note that $v_1^* \tilde{A} v_1 > 0$ for all v_1. Since $\tilde{A}$ is a symmetric and positive definite matrix and $\|v_1\|, \|v_2\| \leq 1$, we have the following bounds:

$$\hat{\mu}_{min}^{\tilde{A}} := v_1^* v_1 \mu_{min}^{\tilde{A}} \leq v_1^* \tilde{A} v_1 \leq \mu_{max}^{\tilde{A}} v_1^* v_1 =: \hat{\mu}_{max}^{\tilde{A}},$$

where $\mu_{min}^{\tilde{A}}$ and $\mu_{max}^{\tilde{A}}$ define the minimal and maximal eigenvalues of $\tilde{A}$ respectively. Based on the results presented in Section 3.1 we can find bounds for $\hat{\mu}_{min}^{\tilde{A}}$: since $\tilde{A} = A_0^{-1/2} A A_0^{-1/2}$ which has the same eigenvalues as $A_0^{-1}A$. Assuming that we use 10 steps of the Chebyshev semi-iterative method for a $Q1$ element in 2D , we get that $\mu_{min}^{\tilde{A}} = \mu_{min}^{A_0^{-1}A} = 0.9980/\gamma_0$ and $\mu_{max}^{\tilde{A}} = \mu_{max}^{A_0^{-1}A} = 1.0020/\gamma_0$ for A_0 being scaled by γ_0 so that $\mathcal{H}$ defines an inner product. In order to be able to provide upper and lower bounds for the eigenvalues given by (29), $\|v_1\|$ has to be known or at least bounds for it. For a small example of dimension 147×147 the eigenvalues of $\mathcal{P}^{-1}\mathcal{A}$ and $\|v_1\|$ are shown in Figure 1 where we observe that $0.9 < \|v_1\| \leq 1$. In practice, $\|v_1\|$ is typically not known and so these bounds are not of much practical use.

4. Block-diagonal preconditioners

The saddle point matrix (1) is typically symmetric and indefinite and one candidate to solve such problems is MINRES introduced by Paige and Saunders in [2]. In order to be able to apply MINRES to $\mathcal{A}$ we need a symmetric and positive definite preconditioner such as the block diagonal preconditioner

$$\mathcal{P}_1 = \begin{bmatrix} A_0 & 0 \\ 0 & S_0 \end{bmatrix} \tag{30}$$

where A_0 is a symmetric positive definite preconditioner for the left upper block of $\mathcal{A}$ and S_0 is a preconditioner based on the Schur complement $BA^{-1}B^T$ of $\mathcal{A}$. Preconditioners of such a

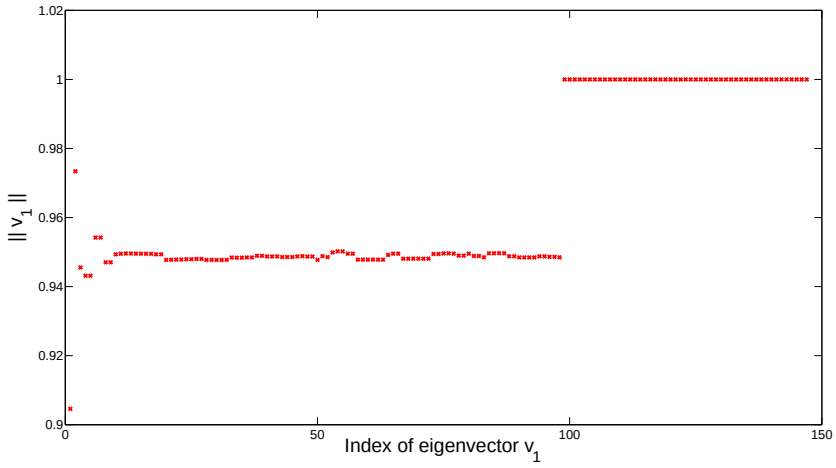


Figure 1. Norm of eigenvector block v_1.

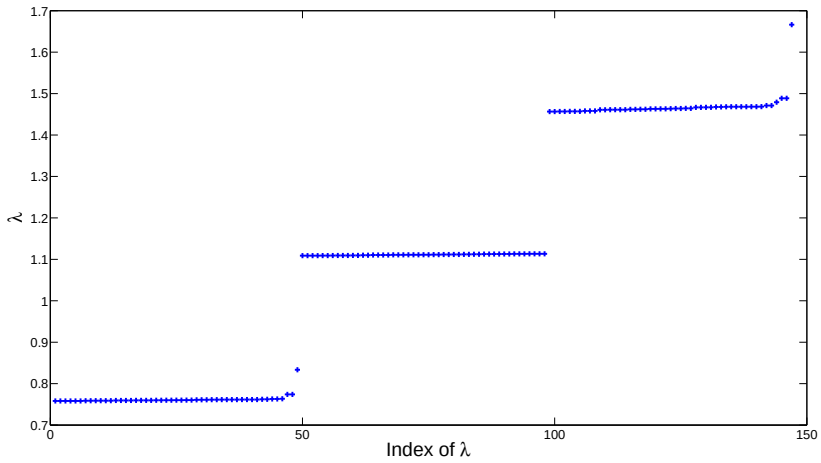


Figure 2. Eigenvalues of $\mathcal{P}^{-1}\mathcal{A}$.

form are well studied when the underlying matrix comes from the finite element treatment of a partial differential equation (see [9, 12, 38] for further details). These preconditioners were also recently used by Rees *et al.* to solve systems arising in the context of PDE constrained optimization (see [10]).

If we have a system of the form (9), then such a block diagonal preconditioner is the one

considered in [10]:

$$\mathcal{P}_{BD} = \begin{bmatrix} M & 0 & 0 \\ 0 & \beta M & 0 \\ 0 & 0 & KM^{-1}K^T \end{bmatrix}. \tag{31}$$

Here $KM^{-1}K^T$ is taken to be the approximation to the Schur complement $\frac{1}{\beta}M + KM^{-1}K^T$. If the system is discretized using Q1 finite elements, as is the case here, then Rees *et al.* show that if λ is an eigenvalue of the preconditioned system $\mathcal{P}_{BD}^{-1}\mathcal{A}$ then λ satisfies one of

$$\lambda = 1, \tag{32}$$

$$\frac{1}{2}\left(1 + \sqrt{5 + \frac{2\alpha_1 h^4}{\beta}}\right) \leq \lambda \leq \frac{1}{2}\left(1 + \sqrt{5 + \frac{2\alpha_2}{\beta}}\right) \tag{33}$$

$$\text{or } \frac{1}{2}\left(1 - \sqrt{5 + \frac{2\alpha_2}{\beta}}\right) \leq \lambda \leq \frac{1}{2}\left(1 - \sqrt{5 + \frac{2\alpha_1 h^4}{\beta}}\right), \tag{34}$$

where α_1, α_2 are positive constants independent of the mesh size h and β, but which will change depending on the discretization used – as in the previous section, $\alpha_1 = 1/1296$ and $\alpha_2 = 1/4\pi^2$ for the discretization considered here. More details can be found in [10].

To make this an effective preconditioner, further approximations to the blocks are made so that a solve with $\mathcal{P}_{BD}$ is cheaper. Therefore, the mass matrix is approximated by a certain number of iterations of the Chebyshev semi-iterative method (see Section 3.1), and the Schur complement by $\tilde{K}M^{-1}\tilde{K}^T$, where $\tilde{K}$ represents a fixed number of algebraic multigrid V-cycles from the `HSL` package `HSL_MI20` [29]. This approximation will have a similar effect on the bounds (32 - 34) as the corresponding approximation had in the previous section.

5. Numerical results

The problem

We illustrate our method using the following example, which is Example 5.1 in [10]. Let $\Omega = [0,1]^m$, where $m = 2, 3$, and consider the problem

$$\min_{y,u} \frac{1}{2}||y - \bar{y}||^2_{L_2(\Omega)} + \frac{\beta}{2}||u||^2_{L_2(\Omega)}$$

$$\text{s.t.} \qquad -\nabla^2 y = u \text{ in } \Omega \tag{35}$$

$$y = \bar{y}|_{\partial\Omega} \text{ on } \partial\Omega \tag{36}$$

where, in 2D,

$$\bar{y} = \begin{cases} (2x_1 - 1)^2(2x_2 - 1)^2 & \text{if } (x_1, x_2) \in [0, \frac{1}{2}]^2 \\ 0 & \text{otherwise} \end{cases}$$

and, in 3D,

$$\bar{y} = \begin{cases} (2x_1 - 1)^2(2x_2 - 1)^2(2x_3 - 1)^2 & \text{if } (x_1, x_2, x_3) \in [0, \frac{1}{2}]^3 \\ 0 & \text{otherwise} \end{cases}$$

i.e. $\bar{y}$ is bi- or tri-quadratic (depending on whether $m = 2$ or 3) with a peak of unit height at the origin and is zero outside $[0, \frac{1}{2}]^m$. Note that we set $f \equiv 0$ here.

We discretize the problem using square Q1 finite elements, and in our preconditioner we approximate M by 10 steps of the Chebyshev semi-iteration and K by one or two V-cycles of the `HSL` package `HSL_MI20` [29] (via a MATLAB interface). We choose $\gamma = 0.9$ as the scaling parameter – this guarantees positive definiteness of $\mathcal{H}$ in (11). To get a fair comparison we use the stopping criterion $\|r_k\|_2/\|r_0\|_2 \leq$ `tol` for both methods, which is different to the norms in which both methods naturally converge.

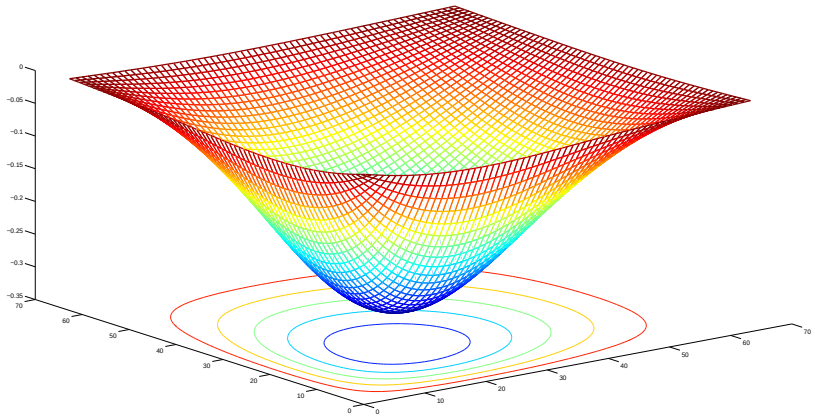


Figure 3. Control for $\beta = 1e-2$

Numerical Experiments

The numerical results shown in Tables III to X indicate that the Bramble-Pasciak CG shows mesh-independent convergence when only one V-cycle of the AMG preconditioner is applied. This is in contrast to MINRES for this problem which needs 2 V-cycles to show mesh independence convergence behavior, even when using the correct norm as a convergence test (see [10]). We note that in every case shown the Bramble-Pasciak CG had fewer iterations than MINRES applied to the same problem. Recall, the cost of one step of the Bramble-Pasciak CG and one step of the block diagonally preconditioned MINRES are essentially the same concerning the number of applications of the preconditioners A_0, A_1 and S_0. Additionally, the Bramble-Pasciak CG requires the evaluation of the inner products with $\mathcal{H}$ which makes an iteration step more expensive. This is one reason why the timings for MINRES and the Bramble-Pasciak CG do not differ as much as the iteration numbers indicate. On the other hand, we expect the timings to better reflect the difference between MINRES and the Bramble-Pasciak CG when both methods are implemented in a low level language. At the moment the AMG preconditioner relies on a fast FORTRAN implementation with the whole algorithm including the matrix multiplication being implemented in MATLAB. A low level implementation of

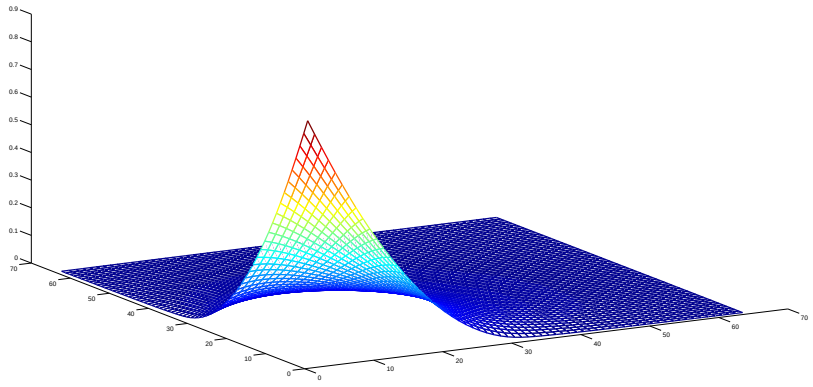


Figure 4. State for $\beta = 1e-2$

both methods allows for problems of much larger dimensions to be solved in which case the multiplication with the system matrix becomes increasingly expensive. This means the extra iterations MINRES needs for convergence will effect the timings more significantly.

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	7	0.019	11	0.030	7	0.021	12	0.033
3	8	0.033	12	0.050	7	0.027	15	0.054
4	8	0.086	13	0.142	10	0.077	16	0.120
5	8	0.317	13	0.517	9	0.230	18	0.454
6	8	1.474	13	2.405	13	1.418	21	2.215
7	10	7.701	15	11.625	15	6.850	24	10.647
8	11	35.838	17	55.866	17	32.617	28	52.366
9	7	103.490	11	164.655	33	290.333	47	433.174

Table III. 2D, $\beta = 1e-2$, 1 V-cycle (AMG),10 steps Chebyshev semi-iteration

6. Conclusions

In this paper we have shown that the Bramble-Pasciak CG is a method well suited for problems from PDE constrained optimization. The drawback of parameter estimation to ensure a positive definite non-standard inner product, typically identified with this method, can be

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	8	0.022	11	0.031	7	0.020	12	0.033
3	8	0.034	11	0.048	7	0.027	13	0.048
4	7	0.077	11	0.122	7	0.057	13	0.100
5	7	0.283	11	0.463	7	0.187	13	0.328
6	8	1.503	11	2.076	7	0.833	14	1.562
7	8	6.304	11	8.701	9	4.440	15	7.064
8	8	27.045	11	37.042	9	19.079	16	32.393
9	7	103.490	11	164.655	16	152.681	23	214.146

Table IV. 2D, $\beta = 1e-2$, 2 V-cycles (AMG),10 steps Chebyshev semi-iteration

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	14	0.039	20	0.056	16	0.043	22	0.058
3	15	0.064	22	0.094	20	0.071	30	0.104
4	15	0.164	22	0.241	20	0.145	31	0.222
5	15	0.608	22	0.935	20	0.473	32	0.776
6	15	2.766	22	4.067	22	2.270	32	3.290
7	15	11.634	23	18.007	24	10.525	34	14.771
8	16	52.771	24	79.295	26	48.917	36	67.096
9	17	252.251	26	389.925	45	396.867	61	565.242

Table V. 2D, $\beta = 1e-4$, 1 V-cycle (AMG),10 steps Chebyshev semi-iteration

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	14	0.039	19	0.053	14	0.038	20	0.053
3	15	0.065	21	0.091	20	0.072	29	0.102
4	15	0.167	22	0.246	20	0.150	30	0.222
5	15	0.610	22	0.930	17	0.445	31	0.754
6	15	2.839	22	4.201	17	1.876	31	3.324
7	15	11.951	23	18.452	17	7.962	32	14.658
8	15	50.890	22	74.776	17	34.284	30	59.129
9	15	225.498	22	331.960	32	295.189	43	392.516

Table VI. 2D, $\beta = 1e-4$, 2 V-cycles (AMG),10 steps Chebyshev semi-iteration

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	7	0.028	12	0.049	11	0.039	17	0.059
3	8	0.288	13	0.467	12	0.269	20	0.442
4	8	3.707	13	6.034	14	4.018	22	6.189
5	9	41.697	14	65.251	18	52.926	26	75.285

Table VII. 3D, $\beta = 1e-2$, 1 V-cycle (AMG),10 steps Chebyshev semi-iteration

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	7	0.029	11	0.046	11	0.041	14	0.050
3	8	0.281	13	0.459	12	0.275	18	0.403
4	8	3.908	13	6.407	12	3.851	18	5.628
5	8	39.808	13	65.122	14	47.045	20	66.080

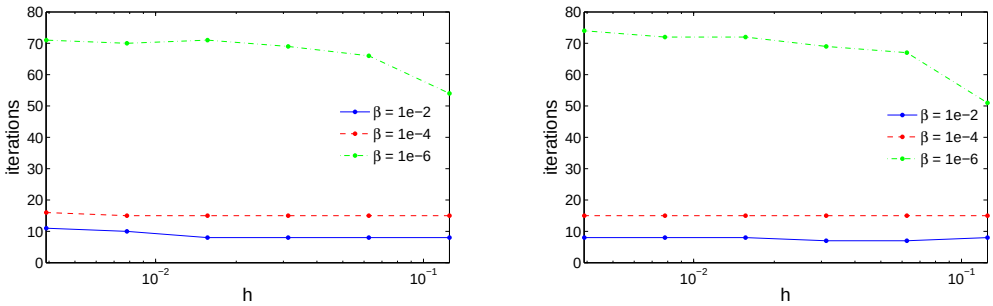
Table VIII. 3D, $\beta = 1e-2$, 2 V-cycles (AMG),10 steps Chebyshev semi-iteration

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	13	0.053	18	0.074	17	0.059	27	0.092
3	14	0.485	23	0.801	24	0.503	39	0.811
4	14	6.508	22	10.298	29	8.013	41	11.241
5	14	65.374	23	107.799	32	91.923	46	130.982

Table IX. 3D, $\beta = 1e-4$, 1 V-cycle (AMG),10 steps Chebyshev semi-iteration

	Bramble Pasciak				MINRES			
	tol=1e-4		tol=1e-8		tol=1e-4		tol=1e-8	
N	its	time	its	time	its	time	its	time
2	11	0.047	15	0.063	13	0.047	21	0.073
3	13	0.462	21	0.754	23	0.501	35	0.753
4	14	6.904	21	10.400	25	7.700	39	11.879
5	14	70.171	22	110.956	27	87.958	44	141.832

Table X. 3D, $\beta = 1e-4$, 2 V-cycles (AMG),10 steps Chebyshev semi-iteration



(a) 1 AMG V-cycle (b) 2 AMG V-cycles

Figure 5. Iterations for convergence for different β

easily overcome when the Chebyshev semi-iteration is used as the preconditioners can be scaled very cheaply based on simple a priori calculations. We have shown an eigenvalue analysis for an idealized case as well as for a general setup. The competitiveness of the Bramble-Pasciak CG is illustrated by the numerical results where we show that this method also gives mesh independence for only one V-cycle of the AMG Schur complement preconditioner.

REFERENCES

1. Benzi M, Golub GH, Liesen J. Numerical solution of saddle point problems. *Acta Numer* 2005; **14**:1–137.
2. Paige CC, Saunders MA. Solutions of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal* 1975; **12**(4):617–629.
3. Hestenes MR, Stiefel E. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand* 1952; **49**:409–436 (1953).
4. Ashby S, Manteuffel T, Saylor P. A taxonomy for conjugate gradient methods. *SIAM J. Numer. Anal* 1990; **27**(6):1542–1568.
5. Dollar S. Iterative linear algebra for constrained optimization. PhD Thesis, University of Oxford 2005. Http://web.comlab.ox.ac.uk/oucl/research/na/thesis/thesisdollar.pdf.
6. Tröltzsch F. *Optimale Steuerung partieller Differentialgleichungen: Theorie, Verfahren und Anwendungen*. Vieweg Verlag, Wiesbaden, 2005.
7. Hinze M, Pinnau R, Ulbrich M, Ulbrich S. *Optimization with PDE Constraints*. Mathematical Modelling: Theory and Applications, Springer-Verlag, New York, 2009.
8. Stoll M, Wathen A. Preconditioning for PDE constrained optimization with control constraints. *In preparation* 2009; .
9. Elman HC, Silvester DJ, Wathen AJ. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Numerical Mathematics and Scientific Computation, Oxford University Press: New York, 2005.
10. Rees T, Dollar HS, Wathen A. Optimal solvers for PDE-Constrained Optimization. *Technical Report NA-08/10*, Oxford University Computing Laboratory 2008.
11. Bramble JH, Pasciak JE. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comp* 1988; **50**(181):1–17.
12. Wathen A, Silvester D. Fast iterative solution of stabilised Stokes systems. I. Using simple diagonal preconditioners. *SIAM J. Numer. Anal* 1993; **30**(3):630–649.
13. Elman HC. Multigrid and Krylov subspace methods for the discrete Stokes equations. *Seventh Copper Mountain Conference on Multigrid Methods*, vol. CP 3339, Melson ND, Manteuffel TA, McCormick SF, Douglas CC (eds.), NASA: Hampton, VA, 1996; 283–299. URL `citeseer.ist.psu.edu/106673.html`.
14. Langer U, Kuhn M. Parallel solvers for large-scale, coupled finite and boundary element equations. Unpublished.

15. Klawonn A. Block-triangular preconditioners for saddle point problems with a penalty term. *SIAM J. Sci. Comput* 1998; **19**(1):172–184. Special issue on iterative methods (Copper Mountain, CO, 1996).
16. Carstensen C, Kuhn M, Langer U. Fast parallel solvers for symmetric boundary element domain decomposition equations. *Numer. Math* 1998; **79**(3):321–347.
17. Ainsworth M, Sherwin S. Domain decomposition preconditioners for p and hp finite element approximation of Stokes equations. *Comput. Methods Appl. Mech. Engrg.* 1999; **175**(3-4):243–266.
18. Zulehner W. Analysis of iterative methods for saddle point problems: a unified approach. *Math. Comp* 2002; **71**(238):479–505.
19. Gatica GN, Heuer N. Conjugate gradient method for dual-dual mixed formulations. *Math. Comp* 2002; **71**(240):1455–1472.
20. Axelsson O, Neytcheva M. Preconditioning methods for linear systems arising in constrained optimization problems. *Numer. Linear Algebra Appl* 2003; **10**(1-2):3–31. Dedicated to the 60th birthday of Raytcho Lazarov.
21. Axelsson O, Neytcheva M. Robust preconditioners for saddle point problems. *Numerical methods and applications*, *Lecture Notes in Comput. Sci*, vol. 2542. Springer: Berlin, 2003; 158–166.
22. Peters J, Reichelt V, Reusken A. Fast iterative solvers for discrete Stokes equations. *SIAM J. Sci. Comput* 2005; **27**(2):646–666.
23. Stoll M. Solving Linear Systems using the Adjoint. PhD Thesis, University of Oxford 2009.
24. Elman HC. Preconditioners for saddle point problems arising in computational fluid dynamics. *Appl. Numer. Math.* 2002; **43**(1-2):75–89. 19th Dundee Biennial Conference on Numerical Analysis (2001).
25. Stoll M, Wathen A. Combination preconditioning and the Bramble–Pasciak$^+$ preconditioner. *SIAM J. Matrix Anal. Appl* 2008; **30**(2):582–608, doi:10.1137/070688961. URL `http://link.aip.org/link/?SML/30/582/1`.
26. Simoncini V. Block triangular preconditioners for symmetric saddle-point problems. *Appl. Numer. Math* 2004; **49**(1):63–80.
27. Rozložník M, Simoncini V. Krylov subspace methods for saddle point problems with indefinite preconditioning. *SIAM J. Matrix Anal.Appl* 2002; **24**(2):368–391, doi:10.1137/S0895479800375540. URL `http://link.aip.org/link/?SML/24/368/1`.
28. Saad Y. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics: Philadelphia, PA, 2003.
29. Boyle J, Mihajlovic MD, Scott JA. HSL_MI20: an efficient AMG preconditioner. *Technical Report RAL-TR-2007-021*, Rutherford Appleton Laboratory (CCLRC) 2007.
30. Ashby SF, Holst MJ, Manteuffel TA, Saylor PE. The role of the inner product in stopping criteria for conjugate gradient iterations. *BIT* 2001; **41**(1):26–52.
31. Rees T, Wathen A. Chebyshev Semi-iteration in Preconditioning. *Technical Report NA-08/14*, Oxford University Computing Laboratory 2008.
32. Wathen AJ. Realistic eigenvalue bounds for the Galerkin mass matrix. *IMA J. Numer. Anal.* 1987; **7**(4):449–457.
33. Wathen AJ. On relaxation of Jacobi iteration for consistent and generalized mass matrices. *Comm. Appl. Numer. Methods* 1991; **7**(2):93–102.
34. Varga RS. *Matrix iterative analysis*. Prentice-Hall Inc.: Englewood Cliffs, N.J., 1962.
35. Golub GH, Varga RS. Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods. I. *Numer. Math.* 1961; **3**:147–156.
36. Golub GH, Varga RS. Chebyshev semi-iterative methods, successive over-relaxation iterative methods, and second order Richardson iterative methods. II. *Numer. Math.* 1961; **3**:157–168.
37. Braess D, Peisker P. On the numerical solution of the biharmonic equation and the role of squaring matrices for preconditioning. *IMA J. Numer. Anal.* 1986; **6**(4):393–404.
38. Silvester D, Wathen A. Fast iterative solution of stabilised Stokes systems. II. Using general block preconditioners. *SIAM J. Numer. Anal* 1994; **31**(5):1352–1367.

RECENT REPORTS

2009

01/09	A Mass and Solute Balance Model for Tear Volume and Osmolarity in The Normal And The Dry Eye	Gaffney Tiffany Yokoi Bron
02/09	Diffusion and permeation in binary solutions	Peppin
03/09	On the modelling of biological patterns with mechanochemical models: insights from analysis and computation	Moreo Gaffney Garcia-Aznar Doblare
04/09	Stability analysis of reaction-diffusion systems with timem-dependent coefficients on growing domains	Madzvamuse Gaffney Maini
05/09	Onsager reciprocity in premelting solids	Peppin Spannuth Wettlaufer
06/09	Inherent noise can facilitate coherence in collective swarm motion	Yates *et al.*
07/09	Solving the Coupled System Improves Computational Efficiency of the Bidomain Equations	Southern Plank Vigmond Whiteley
08/09	Model reduction using a posteriori analysis	Whiteley
09/09	Equilibrium Order Parameters of Liquid Crystals in the Laudau-De Gennes Theory	Majumdar
10/09	Landau-De Gennes theory of nematic liquid crystals: the Oseen-Frank limit and beyond	Majumdar Zarnescu
11/09	A Comparison of Numerical Methods used for Finite Element Modelling of Soft Tissue Deformation	Pathmanathan Gavaghan Whiteley
12/09	From Individual to Collective Behaviour of Unicellular Organisms: Recent Results and Open Problems	Xue Othmer Erban
13/09	Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions	Erban Chapman
14/09	Chaste: a test-driven approach to software development for physiological modelling	Pitt-Francis *et al.*

15/09	Block triangular preconditioners for PDE constrained optimization	Rees Stoll
16/09	From microscopic to macroscopic descriptions of cell migration on growing domains	Baker Yates Erban

Copies of these, and any other OCCAM reports can be obtained from:

Oxford Centre for Collaborative Applied Mathematics
Mathematical Institute
24 - 29 St Giles'
Oxford
OX1 3LB
England
www.maths.ox.ac.uk/occam