A Bramble-Pasciak-like method with applications in optimization

H. Sue Dollar[†], Nicholas I. M. Gould[†][∗], Martin Stoll[∗], Andrew J. Wathen[∗]

June 13, 2008

Abstract

Saddle-point systems arise in many applications areas, in fact in any situation where an extremum principle arises with constraints. The Stokes problem describing slow viscous flow of an incompressible fluid is a classic example coming from partial differential equations and in the area of Optimization such problems are ubiquitous.

In this manuscript we show how new approaches for the solution of saddle-point systems arising in Optimization can be derived from the Bramble-Pasciak Conjugate Gradient approach widely used in PDEs and more recent generalizations thereof. In particular we derive a class of new solution methods based on the use of Preconditioned Conjugate Gradients in non-standard inner products and demonstrate how these can be understood through more standard machinery. We show connections to Constraint Preconditioning and give the results of numerical computations on a number of standard Optimization test examples.

1   Introduction

Saddle-point systems arise in many applications areas, in fact in any situation where an extremum principle arises with constraints. The Stokes problem describing slow viscous flow of an incompressible fluid is a classic example coming from partial differential equations; here minimisation of viscous energy (or flow induced by a body force or boundary forcing) is constrained by conservation of mass (see for example [6]). In the area of Optimization, such problems of finding optima in the prescence of constraints are ubiquitous.

We consider saddle-point problems of the general form

$$Kz \equiv \left[\begin{array}{cc} A & B^T \\ B & -C \end{array} \right] \left[\begin{array}{c} x \\ y \end{array} \right] = \left[\begin{array}{c} b \\ g \end{array} \right] \equiv d \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{m \times m}$ are symmetric matrices and $B \in \mathbb{R}^{m \times n}$. We assume that K is non-singular, sufficient conditions for which are that B is of full rank, C is positive semidefinite and A is positive definite on the kernel of B (see [1]). In practice, the properties for the blocks A and C usually vary with the underlying application. In the context of partial differential equations and mixed finite element methods we can usually assume that A is positive definite and C is positive semi-definite whereas in many problems arising in Optimization, A can be indefinite (for example when directions of negative curvature arise in sequential quadratic programming). The notation used in this paper is reasonably standard, but note that in line with the possibility of indefinite blocks, $\lambda_{\min}(M)$ denotes leftmost eigenvalue and $\lambda_{\max}(M)$ the rightmost eigenvalue of a symmetric matrix M.

There are many methods for solving saddle-point problems (see [1] for a survey). In this paper we give a formulation that represents a framework for many solvers, some already known and some new. In particular, we introduce a new method based on the Bramble-Pasciak CG method [3] and a variant

---

[†]Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxon OX11 0QX, UK (S.Dollar@rl.ac.uk, N.Gould@rl.ac.uk).

[∗]Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford, OX1 3QD, United Kingdom, (nick.gould@comlab.ox.ac.uk, martin.stoll@comlab.ox.ac.uk, andy.wathen@comlab.ox.ac.uk)

of a method recently introduced by Forsgren, Gill and Griffin [9] which extends the idea of constraint preconditioning ([16]). An important feature of our reformulation is that the various methods can be interpreted as Preconditioned Conjugate Gradient methods in non-standard inner products (see [30]). However we also show how such methods can be thought of as acting in the standard ℓ_2 inner product with different preconditioning.

We derive a method similar to the Bramble-Pasciak method which is very similar to the method of Forsgren, Gill and Griffin in the case that C is positive definite but which is well defined and an effective method even if C is semidefinite (including the extreme case that it is zero). We present the results of numerical computations on Optimization examples from the CUTEr test set ([14]).

2 Reformulation

It follows directly that any solution z to (1) also satisfies

$$
\left(\sigma \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} + \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} D & F^T \\ F & E \end{bmatrix} \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix}
$$
$$
= \sigma \begin{bmatrix} b \\ g \end{bmatrix} + \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} D & F^T \\ F & E \end{bmatrix} \begin{bmatrix} b \\ g \end{bmatrix}
$$

(2)

for arbitrary σ, symmetric matrices $D \in \mathbb{R}^{n \times n}$ and $E \in \mathbb{R}^{m \times m}$ and a matrix $F \in \mathbb{R}^{m \times n}$. We denote the coefficient matrix and right-hand side of (2) as $K(\sigma, D, E, F)$ and $d(\sigma, D, E, F)$, respectively, and note that $K = K(1, 0, 0, 0)$ and $d = d(1, 0, 0, 0)$. Many well-known methods can be represented using this reformulation. For example,

- $K(0, I, I, 0)$ gives the normal equations for (1);

- $K(-1, A^{-1}, 0, 0)$ gives the Schur-complement method for finding y when A is nonsingular;

- $K(0, A^{-1}, C^{-1}, 0)$ gives the primal-dual Schur complement method for finding x and y simultaneously when both A and C are nonsingular; and

- $K(1, 0, (1 + \nu)C^{-1}, 0)$ for a given ν (in particular $\nu = 1$) gives the system to which Forsgren, Gill, Griffin apply the preconditioned conjugate gradient (PCG) method, see [9]. The matrices C and $A + B^T C^{-1} B$ are assumed to be positive definite.

There are also a variety of methods that solve (1) by applying the conjugate gradient (CG) method within a non-standard inner-product. The general framework (2) may also be used to represent such methods:

- PCG applied to $K(-1, A_0^{-1}, 0, 0)$ (with appropriate preconditioner) gives the well-respected Bramble-Pasciak configuration for a given A_0, see [3]. The matrices A and $BA^{-1}B^T + C$ are assumed to be positive definite, and A_0 is such that $A - A_0$ is also symmetric and positive definite.

- PCG applied to $K(-\gamma, I, -I, 0)$ (with appropriate preconditioner) gives Liesen and Parlett's method for a given γ, see [18, 19]. The matrix A is assumed to be positive definite and γ lies in the interval $[\lambda_{\max}(C), \lambda_{\min}(A)]$. This method extends that of Benzi and Simoncini [2] to the case where $C \neq 0$.

- PCG applied to $K(-(\alpha + \beta\gamma), \alpha A_0^{-1} + \beta I, -\beta I, 0)$ (with appropriate preconditioner) gives an example of the combination preconditioning method of Stoll and Wathen, see [29]. The assumptions of both Bramble-Pasciak and Liesen and Parlett must hold, and α, β and γ must be chosen such that $K(-(\alpha + \beta\gamma), \alpha A_0^{-1} + \beta I, -\beta I, 0) = \alpha K(-1, A_0^{-1}, 0, 0) + \beta K(-\gamma, I, -I, 0)$ is positive definite.

- PCG applied to $K(1, A_0^{-1}(B^T C_0^{-1} B - A_0^{-1})A_0^{-1}, C_0^{-1}, -C_0^{-1} B A_0^{-1})$ (with appropriate preconditioner) represents the method presented by Schöberl and Zulehner for the case $C = 0$ [26] for given A_0 and

C_0. The matrix A is assumed to be positive definite on the kernel of B, A_0 is such that $A_0 - A$ is symmetric and positive definite, and C_0 is such that $BA_0^{-1}B^T - C_0$ is symmetric and positive definite.

It may not be obvious why the above formulations produce algorithms that (in exact arithmetic) produce iterates which are equivalent to those produced by the CG methods within a non-standard inner-product. In Section 3, we will reveal why we can reformulate the methods as above. We intend that these reformulations will provide the reader with an insight into the properties of the non-standard inner-product CG methods without having to use the non-standard inner-product. Of course, simply reformulating (1) as (2) in itself offers no immediate advantage. However, if $K(\sigma, D, E, F)$ possesses one of more desirable properties, (2) may be preferable to (1).

Before considering the non-standard inner-product conjugate gradient methods, we will consider what properties need to hold to guarantee that $K(\sigma, D, E, F)$ is symmetric and positive definite (and thus one may use methods such as CG rather than MINRES). Clearly, D and E need both be symmetric. Furthermore, we may factorize $K(\sigma, D, E, F)$ as

$$
\begin{aligned}
K(\sigma, D, E, F) &= \begin{bmatrix} \Theta_1 & \Theta_2^T \\ \Theta_2 & \Theta_3 \end{bmatrix} \\
&= \begin{bmatrix} I & \Theta_2^T \Theta_3^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Theta_1 - \Theta_2^T \Theta_3^{-1} \Theta_2 & 0 \\ 0 & \Theta_3 \end{bmatrix} \begin{bmatrix} I & 0 \\ \Theta_3^{-1} \Theta_2 & I \end{bmatrix}
\end{aligned}
\tag{3}
$$

or

$$
K(\sigma, D, E, F) = \begin{bmatrix} I & 0 \\ \Theta_2 \Theta_1^{-1} & I \end{bmatrix} \begin{bmatrix} \Theta_1 & 0 \\ 0 & \Theta_3 - \Theta_2 \Theta_1^{-1} \Theta_2^T \end{bmatrix} \begin{bmatrix} I & \Theta_1^{-1} \Theta_2^T \\ 0 & I \end{bmatrix},
\tag{4}
$$

where

$$
\begin{aligned}
\Theta_1 &= \sigma A + ADA + B^T FA + AF^T B + B^T EB, \tag{5} \\
\Theta_2 &= \sigma B + BDA - CFA + BF^T B - CEB, \tag{6} \\
\Theta_3 &= BDB^T - CFB^T - BF^T C + CEC - \sigma C. \tag{7}
\end{aligned}
$$

Using Sylvester's law of inertia [13], we obtain the following theorem:

Theorem 2.1. *Let Θ_1, Θ_2 and Θ_3 be as defined in (5)–(7). $K(\sigma, D, E, F)$ is symmetric and positive definite if and only if*

- *D and E are symmetric,*
- *Θ_3 is positive definite, and*
- *$\Theta_1 - \Theta_2^T \Theta_3^{-1} \Theta_2$ is positive definite.*

Equivalently, $K(\sigma, D, E, F)$ is symmetric and positive definite if and only if

- *D and E are symmetric,*
- *Θ_1 is positive definite, and*
- *$\Theta_3 - \Theta_2 \Theta_1^{-1} \Theta_2^T$ is positive definite.*

Clearly, $K(\sigma, D, E, F)$ is symmetric and positive definite if and only if $K^{-1} K(\sigma, D, E, F) K^{-1}$ is symmetric and positive definite. This is equivalent to requiring that

$$
\sigma K^{-1} + \begin{bmatrix} D & F^T \\ F & E \end{bmatrix}
$$

be symmetric and positive definite. We will consider different cases for A and C separately. Proofs for the following corollaries may be found in Appendix A.

Corollary 2.2. *If A is symmetric and nonsingular, and*

$$
\begin{aligned}
S_A &= C + BA^{-1}B^T, \\
\Upsilon_1 &= D + \sigma A^{-1} - \sigma A^{-1} B^T S_A^{-1} B A^{-1}, \\
\Upsilon_2 &= \sigma F + S_A^{-1} B A^{-1}, \\
\Upsilon_3 &= E - \sigma S_A^{-1},
\end{aligned}
$$

then $K(\sigma, D, E, F)$ is symmetric and positive definite if and only if

- D *and* E *are symmetric,*

- Υ_3 *is positive definite, and*

- $\Upsilon_1 - \Upsilon_2^T \Upsilon_3^{-1} \Upsilon_2$ *is positive definite.*

Corollary 2.3. *If C is symmetric and nonsingular, and*

$$
\begin{aligned}
S_C &= A + B^T C^{-1} B, \\
\Delta_1 &= D + \sigma S_C^{-1}, \\
\Delta_2 &= F + \sigma C^{-1} B S_C^{-1}, \\
\Delta_3 &= E + \sigma C^{-1} B S_C^{-1} B^T C^{-1} - C^{-1},
\end{aligned}
$$

then $K(\sigma, D, E, F)$ is symmetric and positive definite if and only if

- D *and* E *are symmetric,*

- Δ_1 *is positive definite, and*

- $\Delta_3 - \Delta_2 \Delta_1^{-1} \Delta_2^T$ *is positive definite.*

Corollary 2.4. *If $C = 0$, the columns of $Z \in \mathbb{R}^{n \times (n-m)}$ span the nullspace of B, $B^\dagger = \left(B^T B\right)^{-1} B$ is the Moore-Penrose inverse of B [13] and*

$$
\begin{aligned}
S_Z &= Z^T A Z, \\
\Gamma_1 &= D + \sigma Z S_Z^{-1} Z^T, \\
\Gamma_2 &= F + \sigma B^{\dagger T} \left(I - A Z S_Z^{-1} Z^T\right), \\
\Gamma_3 &= E + \sigma B^{\dagger T} \left(A Z S_Z Z^T A - A\right) B^\dagger.
\end{aligned}
$$

$K(\sigma, D, E, F)$ is symmetric and positive definite if and only if

- D *and* E *are symmetric,*

- Γ_1 *is positive definite,*

- $\Gamma_3 - \Gamma_2 \Gamma_1^{-1} \Gamma_2^T$ *is positive definite.*

Conditions for the case where C is rank-deficient but nonzero may be derived by factoring C as

$$
C = U^T \begin{bmatrix} \widehat{C} & 0 \\ 0 & 0 \end{bmatrix} U,
$$

where $\widehat{C}$ is nonsingular and U is unitary. Premultiplying K by $\begin{bmatrix} I & 0 \\ 0 & U \end{bmatrix}$ and post multiplying by the inverse of this matrix reveals a saddle-point system to which either Corollary 2.3 or Corollary 2.4 could be applied.

3 Equivalence of non-standard inner-product CG methods and standard PCG methods

In this section, we illustrate the equivalence of the reformulation and a class of methods that apply CG with a non-standard inner-product. Such examples can be found in [2, 3, 8, 19, 26, 30]. These methods all have a common framework. That is, matrices P and H are formed such that P is nonsingular, H is symmetric and positive definite, and $P^{-1}K$ is self-adjoint in the inner-product $\langle \cdot, \cdot \rangle$ defined by $\langle x, y \rangle_H = x^T H y$, i.e., $HP^{-1}K$ is symmetric. Moreover, P and H are chosen such that $P^{-1}K$ is positive definite in the inner-product $\langle \cdot, \cdot \rangle_H$. The methods then use this non-standard inner-product within the CG method, as illustrated in Algorithm 1. At iteration k of Algorithm 1, $\text{span}\{p^{(0)}, p^{(1)}, \ldots, p^{(k-1)}\} = \text{span}\{r^{(0)}, r^{(1)}, \ldots, r^{(k-1)}\}$, $r^{(k)T} H r^{(j)} = 0$ and $p^{(k)T} H P^{-1} K p^{(j)} = 0$ for all $j < k$, see [19, Theorem 3.2]. Hence, Algorithm 1 may be reformulated as Algorithm 2.

Algorithm 1 Non-standard inner-product CG (variant 1)

Given $z^{(0)} = 0$, set $r^{(0)} = P^{-1}\left(d - Kz^{(0)}\right)$ and $p^{(0)} = r^{(0)}$
for $k = 0, 1, \ldots$ **do**
$\quad \alpha = \frac{\langle r^{(k)}, p^{(k)} \rangle_H}{\langle P^{-1}K p^{(k)}, p^{(k)} \rangle_H}$
$\quad z^{(k+1)} = z^{(k)} + \alpha p^{(k)}$
$\quad r^{(k+1)} = r^{(k)} - \alpha P^{-1} K p^{(k)}$
$\quad \beta = \frac{\langle P^{-1}K r^{(k+1)}, p^{(k)} \rangle_H}{\langle P^{-1}K p^{(k)}, p^{(k)} \rangle_H}$
$\quad p^{(k+1)} = r^{(k+1)} - \beta p^{(k)}$
end for

Algorithm 2 Non-standard inner-product CG (variant 2)

Given $z^{(0)} = 0$, set $r^{(0)} = P^{-1}\left(d - Kz^{(0)}\right)$ and $p^{(0)} = r^{(0)}$
for $k = 0, 1, \ldots$ **do**
$\quad \alpha = \frac{r^{(k)T} H r^{(k)}}{p^{(k)T} H P^{-1} K p^{(k)}}$
$\quad z^{(k+1)} = z^{(k)} + \alpha p^{(k)}$
$\quad r^{(k+1)} = r^{(k)} - \alpha P^{-1} K p^{(k)}$
$\quad \beta = \frac{r^{(k+1)T} H r^{(k+1)}}{r^{(k)T} H r^{(k)}}$
$\quad p^{(k+1)} = r^{(k+1)} + \beta p^{(k)}$
end for

We observe that $P^{-1}K$ is self-adjoint and positive definite in the inner-product $\langle \cdot, \cdot \rangle_H$ if and only if $HP^{-1}K$ is symmetric and positive definite. Thus, an alternative method for solving (1) is to apply the PCG method (with a preconditioner L) to the equivalent symmetric and positive definite system

$$HP^{-1}Kz = HP^{-1}d. \tag{8}$$

Such a method is given by Algorithm 3. Eliminating $s^{(k)}$ from Algorithm 3 we obtain Algorithm 4.

Observe that when $L = H$, Algorithms 2 and 4 are equivalent (in exact arithmetic). Thus, application of a non-standard inner-product CG method with matrices H and P produces iterates that are equivalent to those formed by applying the standard PCG method to the symmetric and positive definite problem (8) with symmetric and positive definite preconditioner H. The convergence of Algorithms 1 and 2 can, hence, be described by the eigenvalues of $P^{-1}K$, and that of Algorithms 3 and 4 are described by the eigenvalues of $L^{-1}HP^{-1}K$.

Algorithm 3 Preconditioned conjugate gradient method for solving $HP^{-1}Kx = HP^{-1}b$ with symmetric and positive definite preconditioner L

Given $z^{(0)} = 0$, set $s^{(0)} = HP^{-1}\left(d - Kz^{(0)}\right)$
Solve $Lq^{(0)} = s^{(0)}$ and set $p^{(0)} = q^{(0)}$
for $k = 0, 1, \ldots$ **do**
  $\alpha = \frac{s^{(k)T}q^{(k)}}{p^{(k)T}HP^{-1}Kp^{(k)}}$
  $z^{(k+1)} = z^{(k)} + \alpha p^{(k)}$
  $s^{(k+1)} = s^{(k)} - \alpha HP^{-1}Kp^{(k)}$
  Solve $Lq^{(k+1)} = s^{(k+1)}$
  $\beta = \frac{s^{(k+1)T}q^{(k+1)}}{s^{(k)T}q^{(k)}}$
  $p^{(k+1)} = z^{(k+1)} + \beta p^{(k)}$
end for

Algorithm 4 Simplified version of preconditioned conjugate gradient method for solving $HP^{-1}Kx = HP^{-1}b$ with symmetric and positive definite preconditioner L

Given $z^{(0)} = 0$, set $q^{(0)} = L^{-1}HP^{-1}\left(d - Kz^{(0)}\right)$ and $p^{(0)} = q^{(0)}$
for $k = 0, 1, \ldots$ **do**
  $\alpha = \frac{q^{(k)T}Lq^{(k)}}{p^{(k)T}HP^{-1}Kp^{(k)}}$
  $z^{(k+1)} = z^{(k)} + \alpha p^{(k)}$
  $q^{(k+1)} = q^{(k)} - \alpha L^{-1}HP^{-1}Kp^{(k)}$
  $\beta = \frac{q^{(k+1)T}Lq^{(k+1)}}{q^{(k)T}Lq^{(k)}}$
  $p^{(k+1)} = q^{(k+1)} + \beta p^{(k)}$
end for

We note that, in our framework (2), the matrix HP^{-1} corresponds to

$$\sigma I + \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} D & F^T \\ F & E \end{bmatrix}.$$

We will illustrate the above observations by considering the Bramble-Pasciak CG method. This method assumes that A is symmetric and positive definite and sets

$$H = \begin{bmatrix} A - A_0 & 0 \\ 0 & S_0 \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} A_0 & 0 \\ B & -S_0 \end{bmatrix},$$

where A_0 is an approximation to A, $A - A_0$ is symmetric and positive definite, and S_0 is a symmetric and positive definite approximation to $C + BA^{-1}B^T$. We note that Bramble and Pasciak [3] only considered $S_0 = I$ but this was extended to more general S_0 in [17, 20, 27]. Setting $\sigma = -1$, $D = A_0^{-1}$, $E = 0$, and $F = 0$, we can confirm that

$$HP^{-1} = \sigma I + \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} D & F^T \\ F & E \end{bmatrix}.$$

The entries in the matrix $HP^{-1}K$ are independent of the choice of S_0.

Corollary 2.2 implies that $HP^{-1}K$ will be symmetric and positive definite if and only if both $BA^{-1}B^T + C$ and $A_0^{-1} - A^{-1}$ are positive definite. We note that these are exactly the same conditions as those derived by Klawonn [17].

Finally, in Figure 1, we plot the convergence history of the Bramble-Pasciak CG method and the standard PCG method without preconditioning (CG) and with preconditioner $L = H$ (PCG) when applied to a Stokes problem of dimension 59 that was generated by IFISS [5]. We set $A_0 = 0.5A$ and $S_0 = I$. The matrix P was constructed such that $P^{-1}K$ has good convergence properties but we have no reason to expect that $HP^{-1}K$ will also have good convergence properties. We would therefore expect that the Bramble-Pasciak and PCG methods will outperform the unpreconditioned PCG method: Figure 1 confirms our prediction. As expected, when the preconditioner $L = H$ is used within PCG the convergence curve is almost identical to that of the Bramble-Pasciak CG method (the slight deviation is due to round-off error).

4   Using the reformulation

In Sections 2 and 3, we illustrated that different methods for solving saddle-point problems can be presented within the same framework, see (2). Furthermore, we showed that non-standard inner-product CG methods for solving saddle-point systems can be reformulated as standard PCG methods. In this section, we will review the properties of the Forsgren, Gill and Griffin method and derive a Bramble-Pasciak-style method that may have similar convergence properties. Our new method allows us to relax the assumption that C is symmetric and positive definite which is required in the Forsgren, Gill and Griffin approach.

4.1   The method of Forsgren, Griffin and Gill (FGG)

Forsgren, Gill and Griffin [9] work with a saddle-point problem of the general form

$$K(\nu) \begin{bmatrix} x \\ y \end{bmatrix} \equiv \begin{bmatrix} A + (1+\nu)B^T C^{-1} B & -\nu B^T \\ -\nu B & \nu C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b + (1+\nu)B^T C^{-1} g \\ -\nu g \end{bmatrix} \tag{9}$$

where $\nu \in \mathbb{R}$ and, as we have already observed, $K(\nu) \equiv K(1, 0, (1+\nu)C^{-1}, 0)$ in our general setting. We emphasize the fact that C must be definite in this formulation. The case $\nu = -1$ gives the standard saddle-point formulation, $\nu = 0$ a condensed system which is equivalent to the Schur-complement method for finding the solution, and $\nu = 1$ the doubly augmented system with matrix

$$K(1) = \begin{bmatrix} A + 2B^T C^{-1} B & -B^T \\ -B & C \end{bmatrix}. \tag{10}$$
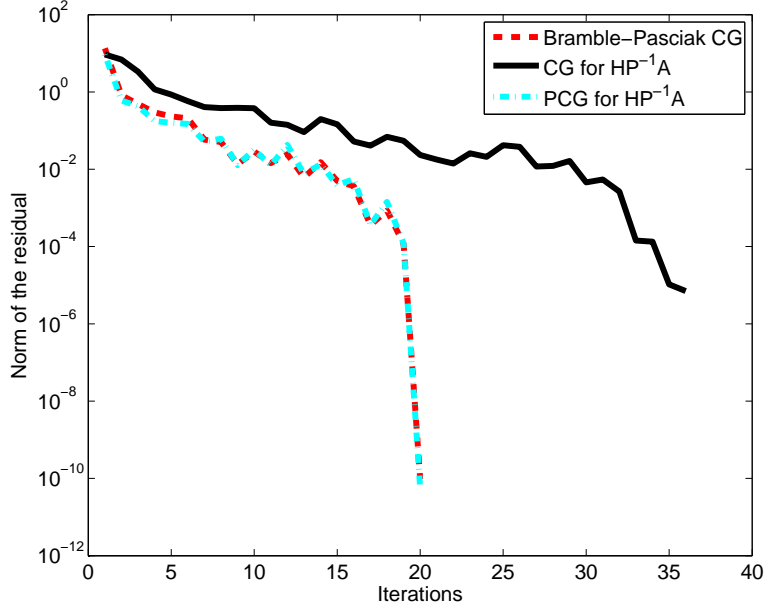
Figure 1: Bramble-Pasciak CG unpreconditioned CG for $HP^{-1}Kz = HP^{-1}d$, and PCG for $HP^{-1}Kz = HP^{-1}d$ with preconditioner H for a Stokes problem of dimension 59.

Note that $K(\nu)$ is positive definite if $A + B^T C^{-1} B > 0$, $C > 0$ and $\nu > 0$. In addition, a general preconditioner

$$M(\nu) = \begin{bmatrix} G + (1+\nu)B^T C^{-1} B & -\nu B^T \\ -\nu B & \nu C \end{bmatrix} \tag{11}$$

is introduced, where G is an approximation to A and $G + B^T C^{-1} B > 0$. Again, $M(\nu)$ represents different preconditioners for different instances of ν. In practice it is often useful to use the decomposition

$$\begin{bmatrix} G + (1+\nu)B^T C^{-1} B & -\nu B^T \\ -\nu B & \nu C \end{bmatrix} = \begin{bmatrix} I & (1+\nu)B^T C^{-1} \\ 0 & -\nu I \end{bmatrix} \begin{bmatrix} G & B^T \\ B & -C \end{bmatrix}$$

to solve a system with the preconditioner $M(\nu)$. Note that the first factor here is easily inverted and the second factor is in the form of a constraint preconditioner ([16]). The eigenvalues of the preconditioned system $M(\nu)^{-1}K(\nu)$ are independent of ν and are given by the eigenvalues of

$$(G + B^T C^{-1} B)^{-1}(A + B^T C^{-1} B) \tag{12}$$

together with m unit eigenvalues. Therefore, in exact arithmetic, convergence is given in at most $n + 1$ steps.

4.2 A Bramble-Pasciak-like approach

In this section, we show the equivalence of the method proposed by Forsgren *et al.* and a Bramble-Pasciak-like method. Multiplying K by -1 and block-symmetrically permuting its rows (and columns) we obtain a matrix of the form

$$\overline{K} = \begin{bmatrix} C & -B \\ -B^T & -A \end{bmatrix}.$$

Applying Bramble-Pasciak to this matrix we obtain

$$\overline{H} = \begin{bmatrix} C - C_0 & 0 \\ 0 & A_0 \end{bmatrix} \quad \text{and} \quad \overline{P} = \begin{bmatrix} C_0 & 0 \\ -B^T & -A_0 \end{bmatrix},$$

8

where A_0 and C_0 are symmetric and nonsingular. Removing the permutation and multiplication we obtain the preconditioner

$$P = \begin{bmatrix} A_0 & B^T \\ 0 & -C_0 \end{bmatrix} \text{ with } P^{-1} = \begin{bmatrix} A_0^{-1} & A_0^{-1}B^TC_0^{-1} \\ 0 & -C_0^{-1} \end{bmatrix} \tag{13}$$

and the bilinear form matrix

$$H = \begin{bmatrix} A_0 & 0 \\ 0 & C - C_0 \end{bmatrix}. \tag{14}$$

It is easy to show that $HP^{-1}K$ is symmetric (i.e., $P^{-1}K$ self-adjoint in the bilinear form given by H) and that

$$HP^{-1}K = K(1, 0, C_0^{-1}, 0).$$

Thus, if C is positive definite and $C_0 = \frac{1}{1+\nu}C$, we obtain the FGG reformulation (see Section 2).

We stress the fact that the FGG method assumes that the matrix C is definite whereas the derivation of Bramble-Pasciak-like method does not require this assumption.

Typically, the properties of A and C depend on the underlying application. We will now discuss some of the common cases. Since, as we noted in Section 3, the iteration is applied implicitly to a system with matrix $P^{-1}K$, and as the eigenvalues of such a matrix influence convergence, we analyze the eigenvalues of the matrix $P^{-1}K$ for various choices of A_0 and C_0. If A_0 and C_0 can be chosen such that H and $HP^{-1}K$ are positive definite, then this would enable us to solve (1) by applying PCG (with preconditioner H) to the equivalent system (8). Using H as the preconditioner enables us rewrite the method in the style of a non-standard inner-product CG method (Algorithms 1 and 2 in Section 3).

If H is symmetric and positive definite, but $HP^{-1}K$ is symmetric but not positive definite, then we cannot reliably apply the CG method. However, we may solve (8) by applying MINRES [21] and using H as the preconditioner. In a similar manner to that employed in Section 3, we can show that the iterates generated are equivalent (in exact arithmetic) to those of the $H-$MINRES method defined in [28]. Another possibility is to use the ITFQMR method of Freund [10] where a simplified version of the nonsymmetric Lanczos process is used based on the identity $HP^{-1}K = KP^{-T}H$. An implementation of ITFQMR is given in [28] - the ITFQMR method is also known as simplified QMR or sQMR.

In the case of the block C being positive semi-definite, e.g. $C = 0$, we can use H-MINRES whenever H is positive definite and ITFQMR whenever $HP^{-1}K = KP^{-T}H$ holds.

4.2.1 C positive definite

The case where C is positive definite can sometimes be found in optimization [9] (as well as other areas [1]) and usually occurs because of some explicit regularization [23]. Optimality conditions imply that $A + B^TC^{-1}B$ should be positive definite. Suppose that we set $C_0 = C$, then the eigenvalues of $P^{-1}K$ are given by the following theorem:

Theorem 4.1. *Let*

$$K = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \quad and \quad P = \begin{bmatrix} A_0 & B^T \\ 0 & -C \end{bmatrix}$$

with nonsingular A_0 and C. Then $P^{-1}K$ has

- *m eigenvalues at 1,*

- *the remaining n eigenvalues are defined by the generalized eigenvalue problem*

$$\left(A + B^TC^{-1}B\right)x = \lambda A_0 x.$$

Proof. It is straightforward to show that

$$P^{-1}K = \begin{bmatrix} A_0^{-1}\left(A + B^TC^{-1}B\right) & 0 \\ -C^{-1}B & I \end{bmatrix}.$$
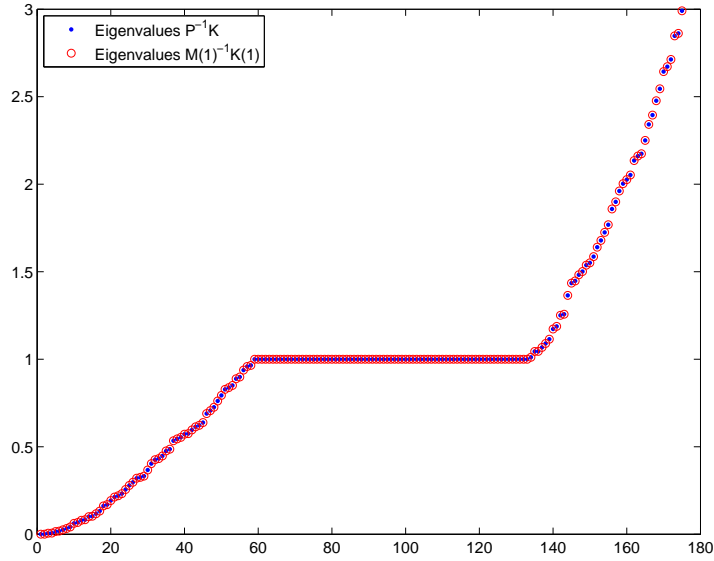
Figure 2: Eigenvalue distribution for $P(\nu)^{-1}K(\nu)$ and $P^{-1}K$, where $C_0 = C$ and $A_0 = G + B^T C^{-1} B$.

Hence, there are m eigenvalues equal to 1 and the remaining eigenvalues satisfy

$$\left(A + B^T C^{-1} B \right) x = \lambda A_0 x.$$

$\square$

Note that if $A_0 = G + B^T C^{-1} B$, then $P^{-1}K$ will have the same eigenvalues as $M(\nu)^{-1}K(\nu)$, where $K(\nu)$ and $M(\nu)$ are defined by equations (9) and (10), respectively. We confirm these results by considering the matrix CVXQP3_S of dimension 175 taken from the CUTEr [15] test set and comparing the eigenvalues of $M(\nu)^{-1}K(\nu)$ and $P^{-1}K$, see Figure 2. In this example, $C = I$ and $G = diag(A)$.

However, if $C_0 = C$, then H in (14) will be singular and, therefore, Algorithms 1 and 2 may breakdown. Suppose that we instead choose $C_0 = (1+\nu)^{-1}C$, where $\nu \neq -1$. If A_0 is chosen to be a symmetric and positive definite matrix, H will be symmetric and positive definite if and only if $\nu > 0$ or $\nu < -1$. Applying Corollary 2.3 with $\sigma = 1$, $D = 0$, $E = C_0^{-1}$, and $F = 0$, we find that $HP^{-1}K$ is positive definite if and only if $A + B^T C^{-1} B$ and $C_0^{-1} - C^{-1}$ are both positive definite. If $C_0 = (1+\nu)^{-1}C$, then $C_0^{-1} - C^{-1}$ is positive definite if and only if $\nu > 0$. This confirms the result by Forsgren *et al.* that $K(\nu)$ is positive definite if $A + B^T C^{-1} B > 0$ and $\nu > 0$, see Section 4.1. Theorem 4.2 provides results on the eigenvalues of the resulting matrix $P^{-1}K$:

Theorem 4.2. *Let B have rank $r > 0$ and $Z \in \mathbb{R}^{n \times (n-r)}$ be such that its columns span the nullspace of B. Additionally, let*

$$K = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \quad and \quad P = \begin{bmatrix} A_0 & B^T \\ 0 & -(1+\nu)^{-1}C \end{bmatrix}$$

with nonsingular A_0 and C, where $\nu \neq 0$ and $\nu \neq -1$. Suppose that the generalized eigenvalue problem $Z^T A Z x_z = \lambda Z^T A_0 Z x_z$ has j $(0 \leq j \leq n - r)$ eigenvalues equal to $1 + \nu$. Then $P^{-1}K$ has

- *at least j eigenvalues at $1 + \nu$,*

- *the remaining eigenvalues satisfy the quadratic eigenvalue problem*

$$\lambda^2 A_0 x - \lambda \left(A + (1+\nu) \left(A_0 + B^T C^{-1} B \right) \right) x + (1+\nu) \left(A + B^T C^{-1} B \right) x = 0$$

10

subject to $\lambda \neq 0$ *and* $\lambda \neq 1 + \nu$.

Proof. Assume that $\left(\lambda, \begin{bmatrix} x^T & y^T \end{bmatrix}^T\right)$ represents an eigenpair of $P^{-1}K$. Then

$$Ax + B^T y = \lambda\left(A_0 x + B^T y\right), \tag{15}$$

$$Bx - Cy = -\frac{\lambda}{1+\nu}Cy. \tag{16}$$

$$\tag{17}$$

Let $\lambda = 1 + \nu$. Equation (16) implies that $Bx = 0$. Let $Z \in \mathbb{R}^{n \times (n-r)}$ be such that its columns span the nullspace of B and $Y \in \mathbb{R}^{n \times r}$ be such that its columns span the range of the columns of B^T. If $x = Y x_y + Z x_z$, then $Bx = 0$ implies that $x_y = 0$. Premultiplying (15) by $\begin{bmatrix} Y & Z \end{bmatrix}^T$ and substituting in $x = Z x_z$ we obtain

$$Y^T A Z x_z + (BY)^T y = (1+\nu)\left(Y^T A_0 Z x_z + (BY)^T y\right), \tag{18}$$

$$Z^T A Z x_z = (1+\nu) Z^T A Z x_z.$$

Hence, $x_z \neq 0$ if and only if $1 + \nu$ is an eigenvalue of the generalized eigenvalue problem $Z^T A Z x_z = \lambda Z^T A_0 Z x_z$. Given such an x_z, y can be defined using (18).

Let $\lambda \neq 1 + \nu$. Equation (16) implies that

$$y = \frac{1+\nu}{1+\nu-\lambda} C^{-1} Bx.$$

Substituting this into (15) an rearranging we obtain the quadratic eigenvalue problem

$$\lambda^2 A_0 x - \lambda\left(A + (1+\nu)\left(A_0 + B^T C^{-1} B\right)\right)x + (1+\nu)\left(A + B^T C^{-1} B\right)x = 0.$$

This completes the proof.

$\square$

Figure 3 shows the eigendistribution for $M(\nu)^{-1} K(\nu)$ and $P^{-1}K$, where $\nu = 0.1$, $C_0 = (1+\nu)C$ and $A_0 = diag(A) + B^T C^{-1} B$. The eigenvalues predicted by Theorem 4.2 are also plotted. As before, we consider the matrix CVXQP3_S from the CUTEr test set with $C = I$.

In the case where a factorization of $A_0 = G + B^T C_0^{-1} B$ should be avoided, it may be helpful to decompose the matrix P as

$$P = \begin{bmatrix} G + B^T C_0^{-1} B & B^T \\ 0 & -C_0 \end{bmatrix} = \begin{bmatrix} G & B^T \\ B & -C_0 \end{bmatrix}\begin{bmatrix} I & 0 \\ C_0^{-1} B & I \end{bmatrix}.$$

Forsgren *et al.* recommend a similar trick for their method, see Section 4.1.

4.2.2 A positive definite and C positive semi-definite

If A is positive definite, then we may let $A_0 = A$; the analysis presented here is not based on the assumption that C is positive definite. The eigenvalues of $P^{-1}K$ are defined by Theorem 4.3.

Theorem 4.3. *Let*

$$K = \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \quad and \quad P = \begin{bmatrix} A_0 & B^T \\ 0 & -(1+\nu)^{-1}C \end{bmatrix}$$

where A is positive definite, C_0 is symmetric and (positive or negative) definite, and $C - C_0$ is nonsingular. Then $P^{-1}K$ has
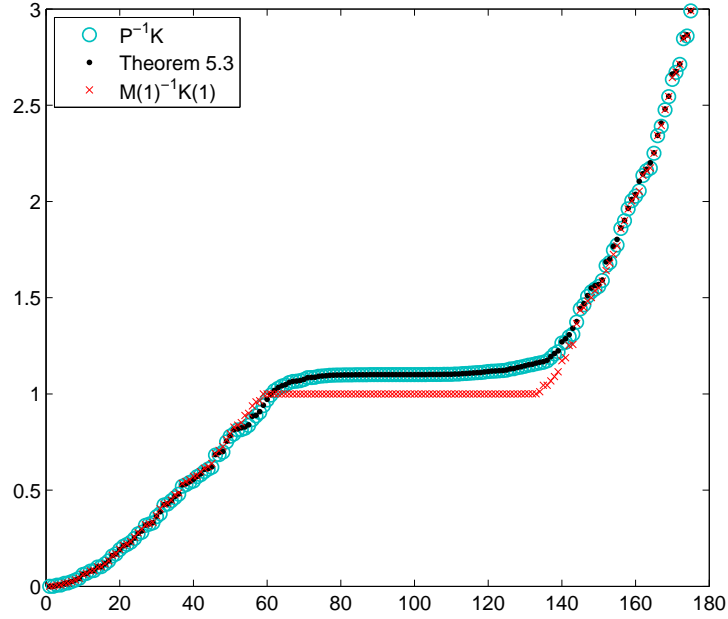
- *n eigenvalues at 1,*

Figure 3: Eigenvalue distribution for $M(\nu)^{-1}K(\nu)$, $P^{-1}K$, and the distribution predicted by Theorem 4.2, where $\nu = 0.1$, $C_0 = (1 + \nu)\,C$ and $A_0 = G + B^T C^{-1} B$.

- *the remaining m eigenvalues are defined by the generalized eigenvalue problem*

$$\left(C + BA^{-1}B^T\right)y = \lambda C_0 y.$$

Proof. Assume that $\left(\lambda, \begin{bmatrix} x^T & y^T \end{bmatrix}^T\right)$ represents an eigenpair of $P^{-1}K$. Then

$$Ax + B^T y = \lambda\left(Ax + B^T y\right) \tag{19}$$
$$Bx - Cy = -\lambda C_0 y. \tag{20}$$

Let $\lambda = 1$. Equation (19) trivially holds. Equation (20) implies that

$$Bx = (C - C_0)\,y.$$

By assumption, $C - C_0$ is nonsingular and, hence, there are n linearly independent eigenvectors of the form

$$\begin{bmatrix} x \\ (C - C_0)^{-1} Bx \end{bmatrix}$$

associated with $\lambda = 1$.

Let $\lambda \neq 1$. Equation (19) implies that $Ax + B^T y = 0$. Therefore, $x = -A^{-1}B^T y$. Substituting this into (20) gives the generalized eigenvalue problem

$$\left(C + BA^{-1}B^T\right)y = \lambda C_0 y.$$

This completes the proof.

$\square$

12

As a result the convergence of the Bramble-Pasciak-like setup with $A_0 = A$ is given in at most $m + 1$ steps. If $C + BA^{-1}B^T$ and C_0 are both positive definite, then all of the eigenvalues of $P^{-1}K$ will be positive, however, C_0 must be chosen such that $C - C_0$ is positive definite in order to guarantee that H is positive definite. If $C + BA^{-1}B^T$ is positive definite and C_0 is negative definite, then $P^{-1}K$ will have m negative eigenvalues.

The case of A definite and C semi-definite typically occurs when working with the mixed finite element formulation of the Stokes problem, see [6]. Such examples can be easily generated using the IFISS package (cf. [5]). Instead of setting $A_0 = A$, A_0 is generally chosen to be a symmetric and positive definite approximation to A, e.g. an Incomplete Cholesky decomposition and C_0 an approximation to the positive or negative Schur-complement. A more general analysis for arbitrary choices of A_0 and C_0 could be performed in the same manner as that in [30]. However, this does not appear to produce useful bounds for the eigenvalues.

4.2.3 Neither A nor C are positive definite

The case where neither A nor C are positive definite is a more severe case since we cannot set $A_0 = A$ or $C_0 = \frac{1}{1+\nu}C$ and expect to obtain a positive definite matrix H. One remedy is to modify A so that the result is positive definite. This may either be achieved during an attempted sparse factorization of A by suitable modifications to its diagonal entries [7, 11] or by modifying the 1×1 and 2×2 diagonal blocks of a computed sparse indefinite factorization, see [4, 24, 25] and [12, Section 4.4.2.2].

5 Numerical Experiments

In this section, we provide examples to show how the Bramble-Pasciak like method can be applied to different problems. The examples in this section are taken either from the CUTEr [15] test set or are generated using the IFISS software package [5]. We will again use the structure presented in Section 4.2 where different setups of the original matrix are analyzed. The methods we compare in this section are the CG of Forsgren, Gill and Griffin (when applicable, Section 4.1) and the Bramble-Pasciak-like CG. We will also compare MINRES [6, 21] where the preconditioner is defined as the block-diagonal matrix

$$P = \left[\begin{array}{cc} A_0 & 0 \\ 0 & M_0 \end{array} \right],$$

where M_0 is a given matrix, and $H-$MINRES (Section 4.2).

C positive definite

In this example, we consider the matrix $CVXQP1\_M$ from CUTEr which is of size 1500×1500. C will either be the identity matrix or a diagonal matrix with entries of the form 10^{-k} on the diagonal where $2 \leq k \leq 10$. We set $M_0 = 0.9C$, $C_0 = M_0$, and $A_0 = diag(A) + B^T C_0^{-1} B$. The right-hand side is such that z is the vector of all ones. The results for the Bramble-Pasciak-like setup and the Forsgren-Gill-Griffin method are shown in Figure 4 and Figure 5. Throughout this section, we compare the relative residuals, where the size of the residual is measured by with the Euclidean norm.

It can be seen that, for these two choices of C, the performance of the Bramble-Pasciak-like method is very similar to that of by the FGG method. It should be noted that if the decomposed form of the Bramble-Pasciak-like preconditioner (see Section 4.2) and the FGG preconditioner (see Section 4.1) are used, we would expect similar timings for each iteration.

A positive definite

The examples we consider in this section are generated using IFISS to discretize the Stokes problem with mixed finite elements: this provides a configuration where A is positive definite and C is positive semi-
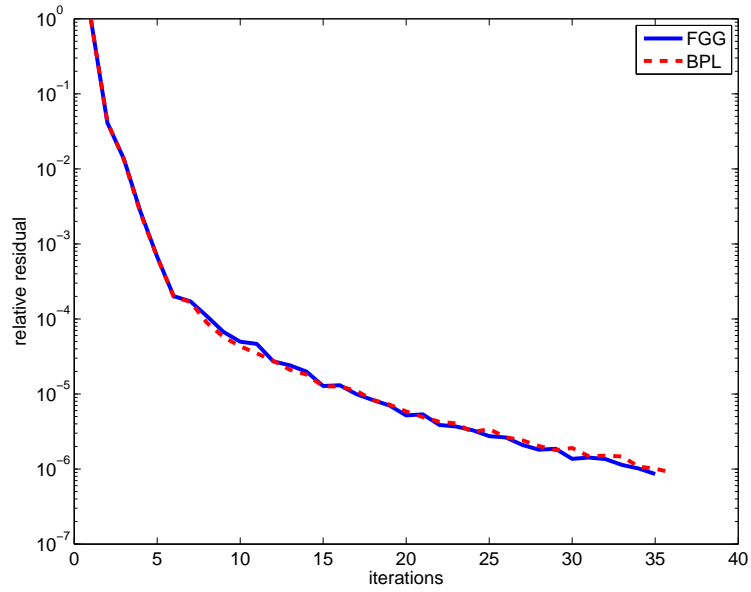
Figure 4: Comparison of the FGG method and the Bramble-Pasciak-like CG method (BPL) method for the matrix *CVXQP1_M* with $C = I$.
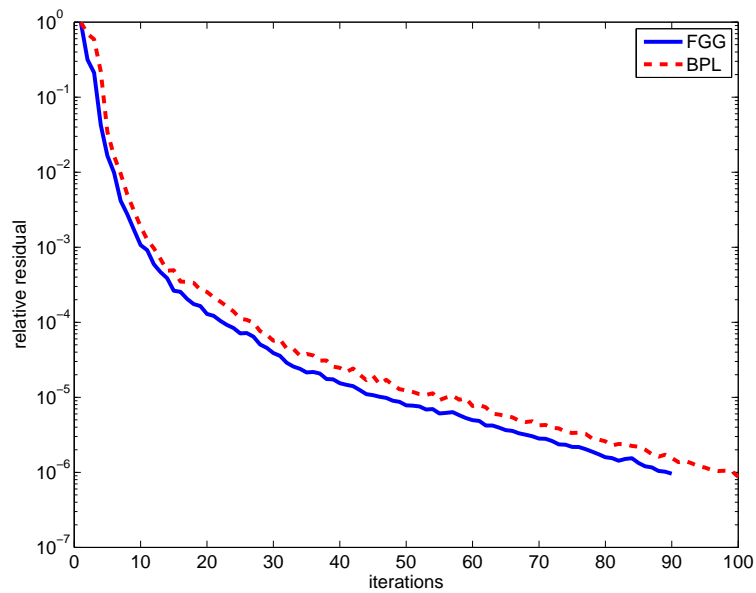

Figure 5: Comparison of the FGG method and the Bramble-Pasciak-like CG method (BPL) method for the matrix *CVXQP1_M* with a random diagonal and positive definite matrix C.
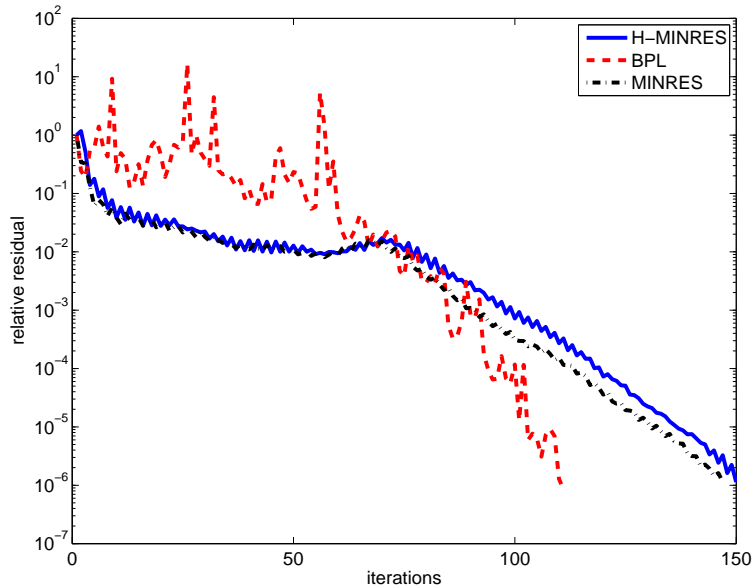
14

Figure 6: Comparison of the H-MINRES method, the Bramble-Pasciak-like CG method (BPL) and the preconditioned MINRES method for the backward-facing step.

definite, see [6]. The first test matrix is of size 6659×6659 and describes the flow over a backward-facing step. The matrix A_0 is taken to be the incomplete Cholesky factorization with zero fill-in [22]. A matrix M_0 is generated by IFISS to be the positive-definite pressure mass matrix. It can be seen from the results in Figure 6 that the Bramble-Pasciak-like CG method (with $C_0 = M_0$) is initially outperformed by the H-MINRES method (with $C_0 = -M_0$) and the standard preconditioned MINRES method. However, the latter two methods then almost stagnate for a large number of iterations (in terms of the Euclidean norm applied to the residual) and, hence, the Bramble-Pasciak-like CG method reaches the desired relative tolerace of 10^{-6} in significantly fewer iterations. We note that the Bramble-Pasciak-like method might breakdown but we observe good behaviour for this problem.

The second test matrix is of size 9539×9539 and describes the flow over a channel domain [6]. The matrix A_0 is chosen such that $A_0 = .9A$ and M_0 is again generated by IFISS as the positive-definite pressure mass matrix. The results given in Figure 7 show that the Bramble-Pasciak-like CG method (with $C_0 = M_0$) out performs the H-MINRES method (with $C_0 = -M_0$) and preconditioned MINRES. We note that the Bramble-Pasciak-like method is not guaranteed to work for this example but, again, we observe good results.

A indefinite and C semi-definite

In this example, we again consider examples from the CUTEr testset where the block A is typically indefinite with zero eigenvalues and the matrix C is positive semi-definite. In [9] it is assumed that the matrix C if semi-definite has a zero block in the lower corner. In order guarantee this structure for real world examples some preprocessing might be necessary. We present an example where we again consider the CUTEr matrix $CVXQP1\_M$ with the block

$$C = \left[\begin{array}{cc} \widehat{C} & 0 \\ 0 & \tilde{C} \end{array} \right] \in \mathbb{R}^{m \times m}$$

where $\widehat{C}$ is a matrix with eigenvalues at zero and $\tilde{C}$ is generated using the MATLAB command
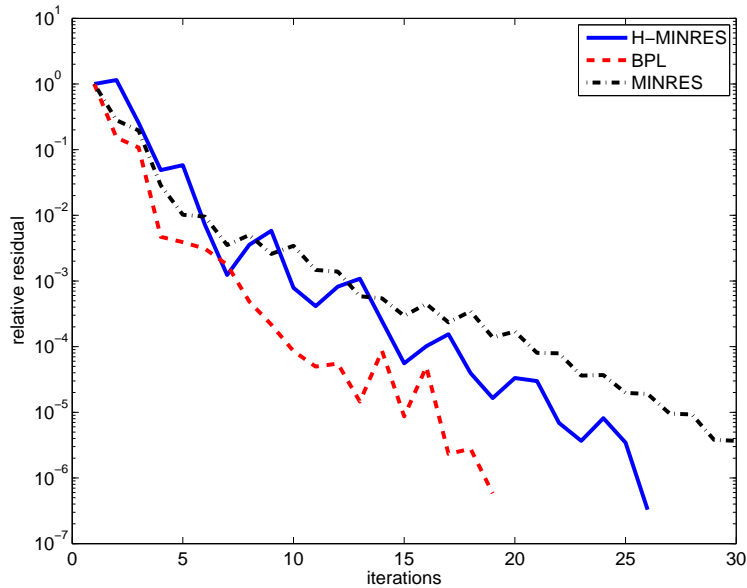
15

Figure 7: Comparison of the H-MINRES method, the Bramble-Pasciak-like CG method (BPL) and the preconditioned MINRES method for the flow over the channel domain.

```
1e-1*sprandsym(p,.1)+1e1*speye(p);
```
with $p = m - 3$.

We set A_0 to be the modified Cholesky preconditioner of A, as presented in Section 4.2.3, and then create a Schur-complement type matrix $M_0 = C + BA_0^{-1}B^T$. Note that we can always reliably apply $H-$MINRES when $C_0 = -M_0$. We will also provide results for the Bramble-Pasciak-like method with $C_0 = M_0$ (which is not guaranteed to work in the case of semi-definite C) and results using ITFQMR for the choice $C_0 = M_0$. From the results given in Figure 8, it can be observed that preconditioned MINRES needs more iterations than the other methods to achieve the given relative tolerance of 10^{-6}. The other methods all perform similarly and converge in a couple of iterations.

The second example in this section is again taken from CUTEr. In particular, we use the matrix *CONT050* which is of size 4998×4998. The setup for C is the same as for *CVXQP1_M* and we again compute a modified Cholesky matrix A_0 for A which we then use to generate a Schur-complement-type matrix $C_0 = C + BA_0^{-1}B^T$. The results are shown in Figure 9.

6 Conclusions

We have presented a reformulation of the saddle-point problem which represents a framework for many well known solution methods for such problems. We have employed this structure to introduce a Bramble-Pasciak-like method related to a constraint preconditioning technique. We have illustrated that competitive results are obtained when applying this method to problems arising in Optimization.

References

[1] M. BENZI, G. H. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.
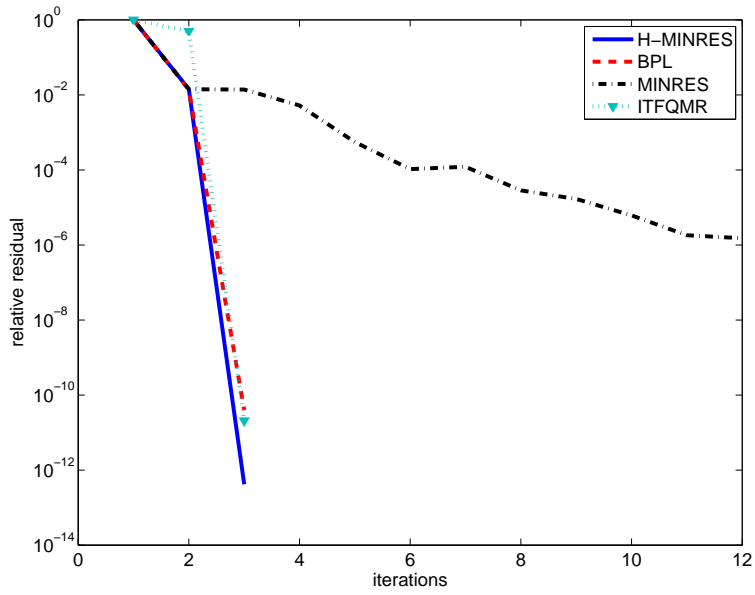
Figure 8: Comparison of the H-MINRES method, the Bramble-Pasciak-like CG method (BPL), the IT-FQMR method and the preconditioned MINRES method for the matrix $CVXQP1\_M$ with indefinite A and semi-definite C.
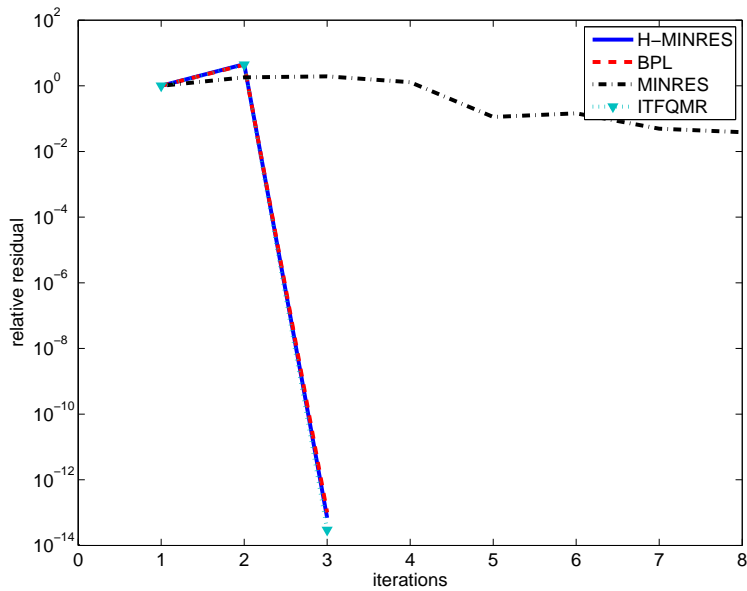

Figure 9: Comparison of the H-MINRES method, the Bramble-Pasciak-like CG method (BPL), the IT-FQMR method and the preconditioned MINRES method for the matrix $CONT050$ indefinite A and semi-definite C.

[2] M. BENZI AND V. SIMONCINI, *On the eigenvalues of a class of saddle point matrices*, Numer. Math., 103 (2006), pp. 173–196.

[3] J. H. BRAMBLE AND J. E. PASCIAK, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp., 50 (1988), pp. 1–17.

[4] S. H. CHENG AND N. J. HIGHAM, *A modified Cholesky algorithm based on a symmetric indefinite factorization*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 1097–1110 (electronic).

[5] H. C. ELMAN, A. RAMAGE, AND D. J. SILVESTER, *Algorithm 866: IFISS, a Matlab toolbox for modelling incompressible flow*, ACM Transactions on Mathematical Software, 33 (2007).

[6] H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford and New York, 2005.

[7] H.-R. FANG AND D. P. O'LEARY, *Modified cholesky algorithms: A catalog with new approaches*, Mathematical Programming, to appear.

[8] B. FISCHER, A. RAMAGE, D. J. SILVESTER, AND A. J. WATHEN, *Minimum residual methods for augmented systems*, BIT, 38 (1998), pp. 527–543.

[9] A. FORSGREN, P. GILL, AND J. GRIFFIN, *Iterative solution of augmented systems arising in interior methods*, SIAM Journal on Optimization, to appear, 8 (2007), pp. 666–690.

[10] R. W. FREUND, *Transpose-free quasi-minimal residual methods for non-Hermitian linear systems*, in Recent advances in iterative methods, vol. 60 of IMA Vol. Math. Appl., Springer, New York, 1994, pp. 69–94.

[11] P. E. GILL AND W. MURRAY, *Newton-type methods for unconstrained and linearly constrained optimization*, Mathematical Programming, 7 (1974), pp. 311–350.

[12] P. E. GILL, W. MURRAY, AND M. H. WRIGHT, *Practical Optimization*, Academic Press, London, 1981.

[13] G. H. GOLUB AND C. F. VAN LOAN, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, third ed., 1996.

[14] N. I. M. GOULD, D. ORBAN, AND P. TOINT, CUTEr*, a constrained and unconstrained testing environment, revisited*, Technical Report TR/PA/01/04, CERFACS, Toulouse, France, 2001. Preliminary version of the paper published in ACM Transactions on Mathematical Software, vol. 29, nber 4, december 2003.

[15] N. I. M. GOULD, D. ORBAN, AND P. L. TOINT, *CUTEr and SifDec: A constrained and unconstrained testing environment, revisited*, ACM Transactions on Mathematical Software (TOMS), 29 (2003), pp. 373–394.

[16] C. KELLER, N. I. M. GOULD, AND A. J. WATHEN, *Constraint preconditioning for indefinite linear systems*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1300–1317.

[17] A. KLAWONN, *Block-triangular preconditioners for saddle point problems with a penalty term*, SIAM J. Sci. Comput., 19 (1998), pp. 172–184. Special issue on iterative methods (Copper Mountain, CO, 1996).

[18] J. LIESEN, *A note on the eigenvalues of saddle point matrices*, Tech. Rep. 10-2006, TU Berlin, 2006.

[19] J. LIESEN AND B. N. PARLETT, *On nonsymmetric saddle point matrices that allow conjugate gradient iterations*, Numerische Mathematik to appear, (2007).

[20] A. Meyer and T. Steidten, *Improvements and experiments on the Bramble-Pasciak Type CG for mixed problems in elasticity*, tech. rep., TU Chemnitz, 2001.

[21] C. C. Paige and M. A. Saunders, *Solutions of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.

[22] Y. Saad, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.

[23] M. A. Saunders, *Cholesky-based methods for sparse least squares: the benefits of regularization*, in Linear and nonlinear conjugate gradient-related methods (Seattle, WA, 1995), SIAM, Philadelphia, PA, 1996, pp. 92–100.

[24] R. B. Schnabel and E. Eskow, *A new modified Cholesky factorization*, SIAM Journal on Scientific Computing, 11 (1991), pp. 1136–1158.

[25] ——, *A revised modified Cholesky factorization algorithm*, SIAM Journal on Optimization, 9 (1999), pp. 1135–1148.

[26] J. Schöberl and W. Zulehner, *Symmetric Indefinite Preconditioners for Saddle Point Problems with Applications to PDE-Constrained Optimization Problems*, SIAM Journal on Matrix Analysis and Applications, 29 (2007), p. 752.

[27] V. Simoncini, *Block triangular preconditioners for symmetric saddle-point problems*, Appl. Numer. Math., 49 (2004), pp. 63–80.

[28] M. Stoll and A. J. Wathen, *The Bramble-Pasciak-plus preconditioner for saddle point problems*, Technical Report 07/13, Oxford University Computing Laboratory, 2007.

[29] ——, *Combination preconditioning and self-adjointness for saddle point problems*, Technical Report 07/11, Oxford University Computing Laboratory, 2007.

[30] ——, *Combination preconditioning and the Bramble-Pasciak+ preconditioner*, SIAM J. Matrix Anal. Appl, accepted, (2008).

Appendix A   Proofs

Let

$$\sigma K^{-1} + \begin{bmatrix} D & F^T \\ F & E \end{bmatrix} = \begin{bmatrix} \Omega_1 & \Omega_2^T \\ \Omega_2 & \Omega_3 \end{bmatrix},$$

for given values of Ω_1, Ω_2 and Ω_3. We may factorize this as

$$\begin{bmatrix} \Omega_1 & \Omega_2^T \\ \Omega_2 & \Omega_3 \end{bmatrix} = \begin{bmatrix} I & \Omega_2^T\Omega_3^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Omega_1 - \Omega_2^T\Omega_3^{-1}\Omega_2 & 0 \\ 0 & \Omega_3 \end{bmatrix} \begin{bmatrix} I & 0 \\ \Omega_3^{-1}\Omega_2 & I \end{bmatrix}. \tag{21}$$

Using Sylvester's law of inertia, $\sigma K^{-1} + \begin{bmatrix} D & F^T \\ F & E \end{bmatrix}$ is positive definite if and if Ω_3 and $\Omega_1 - \Omega_2^T\Omega_3^{-1}\Omega_2$ are both positive definite.

Equivalently, we may use the factorization

$$\begin{bmatrix} \Omega_1 & \Omega_2^T \\ \Omega_2 & \Omega_3 \end{bmatrix} = \begin{bmatrix} I & 0 \\ \Omega_2\Omega_1^{-1} & I \end{bmatrix} \begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_3 - \Omega_2\Omega_1^{-1}\Omega_2^T \end{bmatrix} \begin{bmatrix} I & \Omega_1^{-1}\Omega_2^T \\ 0 & I \end{bmatrix}. \tag{22}$$

Using Sylvester's law of inertia, $\sigma K^{-1} + \begin{bmatrix} D & F^T \\ F & E \end{bmatrix}$ is positive definite if and if Ω_1 and $\Omega_3 - \Omega_2\Omega_1^{-1}\Omega_2^T$ are both positive definite.

If A is nonsingular, then

$$K^{-1} = \begin{bmatrix} A^{-1} - A^{-1}B^T S^{-1} B A^{-1} & A^{-1}B^T S^{-1} \\ S^{-1} B A^{-1} & -S^{-1} \end{bmatrix},$$

where $S = C + BA^{-1}B^T$. Use of factorization (21) completes the proof of Corollary 2.2.

If C is nonsingular, then

$$K^{-1} = \begin{bmatrix} S^{-1} & S^{-1}B^T C^{-1} \\ C^{-1}BS^{-1} & C^{-1}BS^{-1}B^T C^{-1} - C^{-1} \end{bmatrix},$$

where $S = A + B^T C^{-1} B$. Use of factorization (22) completes the proof of Corollary 2.3.

If $C = 0$, the columns of $Z \in \mathbb{R}^{n \times (n-m)}$ span the nullspace of B, and $B^\dagger$ be the Moore-Penrose inverse of B, then

$$K^{-1} = \begin{bmatrix} ZS^{-1}Z^T & \left(I - ZS^{-1}Z^T A\right) B^\dagger \\ B^{\dagger T}\left(I - AZS^{-1}Z^T\right) & B^{\dagger T}\left(AZS^{-1}Z^T A - A\right)B^\dagger \end{bmatrix},$$

where $S = Z^T A Z$. Use of factorization (22) completes the proof of Corollary 2.4.