A BRAMBLE-PASCIAK-LIKE METHOD WITH APPLICATIONS IN OPTIMIZATION

MARTIN STOLL*

Abstract. The Bramble-Pasciak Conjugate Gradient algorithm is a widely used tool in the finite element community. Motivated by a reformulation of the linear system in saddle point form, we introduce Bramble-Pasciak-like methods that can be used to solve problems coming from optimization. We illustrate that the eigenvalues for the preconditioned matrix in this setup have a very similar (sometimes equivalent) structure to the preconditioned matrix of a method which uses a constraint preconditioner. We furthermore give numerical results for optimization examples.

AMS subject classifications. Primary 65F10, 65N22, 65F50 Secondary 76D07

Key words. Saddle point problems, linear systems, Krylov subspaces, preconditioning

1. Introduction. In many applications such as partial differential equations or optimization, systems of the form

$$\underbrace{\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}}_{\mathcal{A}} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ g \end{bmatrix} \tag{1.1}$$

with $A \in \mathbb{R}^{n,n}$ and $C \in \mathbb{R}^{m,m}$ being symmetric matrices arise. We assume that $ker(C) \cap ker(B^T) = \emptyset$ holds for $B \in \mathbb{R}^{m,n}$. 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 assume that A is positive definite and C is positive semi-definite whereas A is typically indefinite for saddle point problems stemming from optimization with C positive definite or positive semi-definite. There are many methods for solving saddle point problems [1] and we give a formulation that represents a framework for many solvers. Furthermore, we introduce a method based on the Bramble-Pasciak CG [2] and a method recently introduced in [6].

2. Reformulation. Using (1.1) we can obtain an equivalent system

$$\begin{aligned} &\left(\sigma \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} + \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} D & F^T \\ F & E \end{bmatrix} \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix}\right) \begin{bmatrix} x \\ y \end{bmatrix} \\ &\qquad = \begin{bmatrix} b \\ g \end{bmatrix} + \begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} D & F^T \\ F & E \end{bmatrix} \begin{bmatrix} b \\ g \end{bmatrix} \end{aligned} \tag{2.1}$$

for all $\sigma \neq 0$, arbitrary symmetric matrices $D \in \mathbb{R}^{n,n}$ and $E \in \mathbb{R}^{m,m}$ and an arbitrary matrix F. We denote the coefficient matrix of Equation 2.1 as $K(\sigma, D, E, F)$. Many well known methods can be represented using (2.1), i.e.

- $K(-1, A^{-1}, 0, 0)$ gives the Schur-complement method for finding y.
- $K(-1, A_0^{-1}, 0, 0)$ gives the classical Bramble-Pasciak configuration for a given A_0, see [2].
- $K(\gamma, I, -I, 0)$ gives Liesen and Parlett's method for a given γ, see [12, 13]
- $K(-\alpha + \beta\gamma, \alpha A_0^{-1} + \beta I, -\beta I, 0)$ gives the combination preconditioning method of Stoll and Wathen, see [17].

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

- $K(1,0,(1+\nu)C^{-1},0)$ for a given ν (in particular $\nu = 1$) this represents the method proposed by Forsgren, Gill, Griffin, see [6].
- $K(0,I,I,0)$ gives the normal equations.
- $K(0,A^{-1},C^{-1},0)$ gives the primal-dual Schur complement method for finding x and y simultaneously.
- $K(1,\hat{A}^{-1}(B^T\hat{S}^{-1}B - \hat{A}^{-1})\hat{A}^{-1},\hat{S}^{-1},-\hat{S}^{-1}B\hat{A}^{-1})$ represents the method presented by Schöberl and Zulehner for the case $C = 0$ [15].

3. Reformulation and non-standard inner products. In this section we illustrate the equivalence of the reformulation and a method proposed by Bramble and Pasciak in [2] which we quickly review. The block-triangular preconditioner

$$\mathcal{P} = \begin{bmatrix} A_0 & 0 \\ B & -I \end{bmatrix} \text{ where } \quad \mathcal{P}^{-1} = \begin{bmatrix} A_0^{-1} & 0 \\ BA_0^{-1} & -I \end{bmatrix} \tag{3.1}$$

is used for the saddle point problem given in (1.1). The resulting preconditioned matrix $\widehat{\mathcal{A}} = \mathcal{P}^{-1}\mathcal{A}$ is not symmetric anymore and therefore no short-term recurrence method can be applied, see [5, 14]. The matrix

$$\mathcal{H} = \begin{bmatrix} A - A_0 & 0 \\ 0 & I \end{bmatrix}$$

defines a bilinear form and under certain conditions on A_0, $\mathcal{H}$ defines the inner product $\langle x,y\rangle_{\mathcal{H}} = x^T\mathcal{H}y$ in which the preconditioned matrix $\widehat{\mathcal{A}}$ is not only self-adjoint but also positive definite; therefore CG can be employed in this inner product, see Algorithm 1.

Algorithm 1 Bramble and Pasciak CG

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

The Bramble-Pasciak CG can also be viewed as the Preconditioned Conjugate Gradient method (PCG) [4, 10] applied to the matrix $\mathcal{H}\mathcal{P}^{-1}\mathcal{A}$. In more detail, solving system (1.1) is equivalent to solving the system

$$\mathcal{H}\mathcal{P}^{-1}\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \mathcal{H}\mathcal{P}^{-1}\begin{bmatrix} b \\ g \end{bmatrix}, \tag{3.2}$$

which can be obtained from (2.1) via

$$\begin{aligned} K(-1, A_0^{-1}, 0, 0) &= \left(-1\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} + \begin{bmatrix} A \\ B \end{bmatrix} A_0^{-1} \begin{bmatrix} A & B^T \end{bmatrix}\right) \\ &= \begin{bmatrix} AA_0^{-1}A - A & AA_0^{-1}B^T - B^T \\ BA_0^{-1}A - B & BA_0^{-1}B^T + C \end{bmatrix} = \widehat{\mathcal{A}}^T\mathcal{H} = \mathcal{H}\widehat{\mathcal{A}}. \end{aligned} \tag{3.3}$$

The sequence of approximations $\left\{x^{(k)}\right\}$ generated by the Bramble-Pasciak CG method satisfies $x^{(k)} \in span\left\{\mathcal{P}^{-1}\mathcal{A}r_0, \ldots, (\mathcal{P}^{-1}\mathcal{A})^{k-1}r_0\right\}$. Applying the (unpreconditioned) conjugate gradient method to solve the linear system with $\mathcal{H}\mathcal{P}^{-1}\mathcal{A}$ will result in $x^{(k)} \in span\left\{\mathcal{H}\mathcal{P}^{-1}\mathcal{A}r_0, \ldots, (\mathcal{H}\mathcal{P}^{-1}\mathcal{A})^{k-1}r_0\right\}$. Thus, the Krylov subspaces will be different and a different sequence of iterates will be formed. Suppose we apply the preconditioned CG method with a symmetric and positive definite preconditioner L to solve (3.2). Using the classical PCG implementation given in [4, 10] we obtain Algorithm 2. It is easily seen that both algorithms are identical when $L = \mathcal{H}$.

Algorithm 2 PCG for solving $\mathcal{H}\mathcal{P}^{-1}\mathcal{A}x = \mathcal{H}\mathcal{P}^{-1}b$ with symmetric and positive definite preconditioner L

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

In order to analyze the relationship between classical CG and the Bramble-Pasciak CG , we illustrate that the positivity condition in the inner product defined by $\mathcal{H}$ relates to condition when CG can be applied to the system matrix given in (3.2).

The preconditioned matrix $\widehat{\mathcal{A}} = \mathcal{P}^{-1}\mathcal{A}$ is self-adjoint in the inner product defined by $\mathcal{H}$ and for the positive definiteness we have to show that $\langle \mathcal{P}^{-1}\mathcal{A}x, x\rangle_{\mathcal{H}} > 0$ which is equivalent to $\langle \mathcal{H}\mathcal{P}^{-1}\mathcal{A}x, x\rangle > 0$. The relation for self-adjointness in a bilinear form is given by $(\mathcal{P}^{-1}\mathcal{A})^T\mathcal{H} = \mathcal{H}\mathcal{P}^{-1}\mathcal{A}$ which means that for $\langle \mathcal{H}\mathcal{P}^{-1}\mathcal{A}x, x\rangle$ to be positive, we only have to look at the field of values of the symmetric matrix

$$\mathcal{H}\mathcal{P}^{-1}\mathcal{A} = \begin{bmatrix} AA_0^{-1}A - A & AA_0^{-1}B^T - B^T \\ BA_0^{-1}A - B & BA_0^{-1}B^T + C \end{bmatrix}.$$

Conditions for this matrix to be positive are given in [11, 16]. Hence, the Bramble-Pasciak CG can be applied whenever the symmetric matrix $\mathcal{H}\mathcal{P}^{-1}\mathcal{A}$ is positive definite. This shows that the alternative formulation given in (2.1) represents a linear system with the matrix $\mathcal{H}\widehat{\mathcal{A}}$, see (3.3) .

We now want to examine whether the alternative formulation should be used for numerical purposes. The preconditioner $\mathcal{P}$ was constructed to alter the spectrum of the preconditioned matrix $\mathcal{A}$ such that good convergence can be expected for $\widehat{\mathcal{A}}$. On the other hand, if we premultiply $\widehat{\mathcal{A}}$ by $\mathcal{H}$ a further convergence enhancement cannot necessarily be expected. Moreover, we expect the convergence with $\mathcal{H}\widehat{\mathcal{A}}$ to be poorer since the premultiplication by $\mathcal{H}$ will destroy the eigenvalue structure achieved by applying the preconditioner $\mathcal{P}$. An alternative would be to use $\mathcal{H}$ as a preconditioner for $\mathcal{H}\widehat{\mathcal{A}}$ as explained above which would result in the eigenstructure of the preconditioned matrix $\widehat{\mathcal{A}}$. In Figure 3.1 we plot the convergence history of the Bramble-Pasciak CG and the classical PCG (with and without preconditioning) when applied to a Stokes problem of dimension 59 that was generated by IFISS [3].

As predicted the unpreconditioned CG method is outperformed by the Bramble-Pasciak CG method. When the preconditioner $L = \mathcal{H}$ is used within PCG the convergence curves are almost identical due to round-off error.

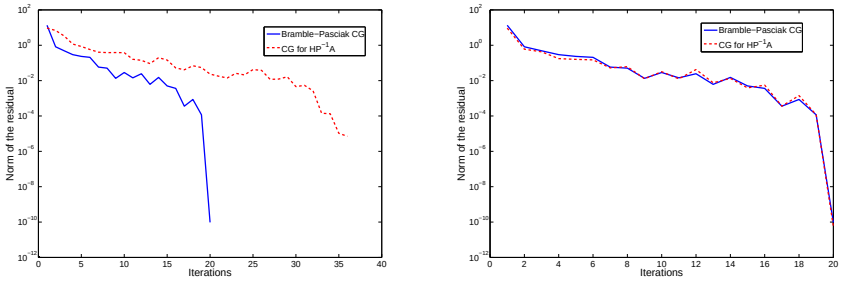

FIG. 3.1. *2-norm of the residual for Bramble-Pasciak* CG *and* CG *for* $\mathcal{H}\widehat{\mathcal{A}}$ *without preconditioning (left) and with preconditioner* $\mathcal{H}$ *(right) for a Stokes problem generated with* IFISS *[3] of dimension* 59.

4. Using the reformulation. In Section 2 we illustrated that different methods for solving saddle point problems can be presented within the same framework, see (2.1). Furthermore, we showed that for a Bramble-Pasciak setup it would not be feasible to use the alternative formulation for numerical experiments due to the fact that we first multiply by $\mathcal{H}$ and then use it as a preconditioner. In this section we want to show how the alternative formulation can be used to generate a Bramble-Pasciak-like method from another method that lies within the same framework. We therefore quickly summarize the method of Forsgren *et al.* introduced in [6].

4.1. The method of Forsgren, Griffin and Gill (FGG). Forsgren, Gill and Griffin work with a saddle point problem of the general form

$$\mathcal{A}(\nu)\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A + (1+\nu)B^TC^{-1}B & \nu B^T \\ \nu B & \nu C \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ g \end{bmatrix} \tag{4.1}$$

where $\nu \in \mathbb{R}$. The case $\nu = -1$ gives the classical 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

$$\mathcal{A}(1) = \begin{bmatrix} A + 2B^TC^{-1}B & B^T \\ B & C \end{bmatrix}. \tag{4.2}$$

Note, that the assumption that C is definite is needed for this formulation. In addition, a general preconditioner

$$\mathcal{P}(\nu) = \begin{bmatrix} G + (1+\nu)B^TC^{-1}B & \nu B^T \\ \nu B & \nu C \end{bmatrix} \tag{4.3}$$

is introduced where G is an approximation to A. Again, $\mathcal{P}(\nu)$ represents different preconditioners for different instances of ν.

The eigenvalues of the preconditioned system $\mathcal{P}(\nu)^{-1}\mathcal{A}(\nu)$ are independent of ν and are given by the eigenvalues of

$$(G + B^TC^{-1}B)^{-1}(A + B^TC^{-1}B) \tag{4.4}$$

and 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. In order to construct a Bramble-Pasciak-like method we consider the preconditioner

$$\mathcal{P}_- = \begin{bmatrix} A_0 & B^T \\ 0 & -C_0 \end{bmatrix} \text{ with } \mathcal{P}_-^{-1} = \begin{bmatrix} A_0^{-1} & A_0^{-1}B^TC_0^{-1} \\ 0 & -C_0^{-1} \end{bmatrix} \tag{4.5}$$

and the bilinear form

$$\mathcal{H}_- = \begin{bmatrix} A_0 & 0 \\ 0 & C - C_0 \end{bmatrix}. \tag{4.6}$$

It is easy to see that the preconditioned matrix $\widehat{\mathcal{A}} = \mathcal{P}_-^{-1}\mathcal{A}$ is self-adjoint in this bilinear form when verifying that $\widehat{\mathcal{A}}^T\mathcal{H}_- = \mathcal{H}_-\widehat{\mathcal{A}}$ holds.

The connection to the method by Forsgren *et al.* can be made by looking at (2.1) in the setup

$$K(1, 0, (1+\nu)C^{-1}, 0) = \begin{bmatrix} A + (1+\nu)B^TC^{-1}B & -\nu B^T \\ -\nu B & \nu C \end{bmatrix}.$$

This matrix can also be expressed as

$$\begin{aligned} \mathcal{H}_-\widehat{\mathcal{A}} &= \begin{bmatrix} I & 0 \\ 0 & C - \frac{1}{(1+\nu)}C \end{bmatrix} \begin{bmatrix} A + B^T(1+\nu)C^{-1}B & B^T - B^T(1+\nu)C^{-1}C \\ -(1+\nu)C^{-1}B & (1+\nu)C^{-1}C \end{bmatrix} \\ &= \begin{bmatrix} A + (1+\nu)B^TC^{-1}B & -\nu B^T \\ -\nu B & \nu C \end{bmatrix} \end{aligned} \tag{4.7}$$

which corresponds to the Bramble-Pasciak-like setting with $C_0 = \frac{1}{(1+\nu)}C$ and $A_0 = I$. We want to stress the fact that for the method of Forsgren *et al.* the matrix C is assumed to be definite whereas the Bramble-Pasciak method only needs the definiteness of the preconditioners C_0 and A_0.

We firstly analyze the Bramble-Pasciak-like method for the case when C is definite and show that it is possible to choose A_0 and C_0 such that $\mathcal{H}_-$ defines an inner product and $\widehat{\mathcal{A}}$ is positive definite within this inner product. This would enable the use of CG for the Bramble-Pasciak-like equivalent of the method introduced [6].

The matrix

$$\mathcal{H}_- = \begin{bmatrix} A_0 & 0 \\ 0 & C - C_0 \end{bmatrix}$$

defines an inner product whenever A_0 is symmetric and positive definite and whenever the symmetric block $C - C_0$ becomes positive definite, i.e. $C - C_0 > 0$ where C is a positive definite matrix. In addition, we need all the eigenvalues of

$$\widehat{\mathcal{A}}^T\mathcal{H}_- = \begin{bmatrix} A + B^TC_0^{-1}B & -B^TC_0^{-1}C + B^T \\ B - CC_0B & CC_0^{-1}C - C \end{bmatrix}$$

to be positive. We use a technique employed in [11, 16] where we split $\widehat{\mathcal{A}}^T\mathcal{H}_-$ as

$$\widehat{\mathcal{A}}^T\mathcal{H}_- = \begin{bmatrix} I & -B^TC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A + B^TC^{-1}B & 0 \\ 0 & CC_0^{-1}C - C \end{bmatrix} \begin{bmatrix} I & 0 \\ -C^{-1}B & I \end{bmatrix}.$$

Since this is an congruence transformation, Sylvester's law of inertia gives that we only have to look at the eigenvalues of

$$\begin{bmatrix} A + B^T C^{-1} B & 0 \\ 0 & C C_0^{-1} C - C \end{bmatrix}.$$

The first block $A + B^T C^{-1} B$ is always positive definite and it can be shown that the block $C C_0^{-1} C - C$ is positive definite whenever $C_0 < C$. Therefore, we are able to reliably apply the CG method to the linear system. The case given in [6] where $C_0 = \frac{1}{1+\nu} C$ fulfills this criterion if the matrix C is definite.

In [17] Stoll and Wathen introduce a preconditioner and bilinear form very similar to the classical Bramble-Pasciak one but with different numerical properties. The main motivation was to have a bilinear form

$$\mathcal{H}_+ = \begin{bmatrix} A_0 & 0 \\ 0 & C + C_0 \end{bmatrix} \tag{4.8}$$

that defines an inner product whenever the preconditioners A_0 and C_0 are positive definite. The preconditioner can also be modified and we get

$$\mathcal{P}_+ = \begin{bmatrix} A_0 & B^T \\ 0 & C_0 \end{bmatrix} \text{ with } \mathcal{P}_+^{-1} = \begin{bmatrix} A_0^{-1} & -A_0^{-1} B^T C_0^{-1} \\ 0 & C_0^{-1} \end{bmatrix}. \tag{4.9}$$

Hence, the preconditioned matrix

$$\widehat{\mathcal{A}} = \mathcal{P}_+^{-1} \mathcal{A} = \begin{bmatrix} A_0^{-1} A - A_0^{-1} B^T C_0^{-1} B & A_0^{-1} B^T + A_0^{-1} B^T C_0^{-1} C \\ C_0^{-1} B & -C_0^{-1} C \end{bmatrix}$$

is self-adjoint in the inner product $\mathcal{H}_+$.

The applicability of CG can be determined by studying the eigenvalues of

$$\widehat{\mathcal{A}}^T \mathcal{H}_+ = \begin{bmatrix} I & -B^T C^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A + B^T C^{-1} B & 0 \\ 0 & -C C_0^{-1} C - C \end{bmatrix} \begin{bmatrix} I & 0 \\ -C^{-1} B & I \end{bmatrix}.$$

Again Sylvester's law of inertia tells us that the eigenvalues of $A + B^T C^{-1} B$ and $-C C_0^{-1} C - C$ will determine the number of positive, negative and zero-eigenvalues of the matrix $\widehat{\mathcal{A}}^T \mathcal{H}$. The block $A + B^T C^{-1} B$ will be positive definite for all C_0 whereas the block $-(C C_0^{-1} C + C)$ will be negative for C_0 being positive definite. Therefore, we cannot reliably apply the CG method. As an alternative, a $\mathcal{H}_+$-MINRES method[1] can always be implemented since an inner product is always at hand due to the definiteness of A_0 and C_0. Another possibility is to use the ITFQMR method of Freund [7] where a simplified version of the nonsymmetric Lanczos process is used based on the identity $\widehat{\mathcal{A}}^T \mathcal{H}_+ = \mathcal{H}_+ \widehat{\mathcal{A}}$. An implementation of ITFQMR is given in [16]. The ITFQMR method is sometimes called simplified QMR or SQMR .

In the case of the block C being positive semi-definite, e.g. $C = 0$, we can use $\mathcal{H}_\pm$-MINRES whenever $\mathcal{H}_\pm$ defines an inner product and ITFQMR whenever $\widehat{\mathcal{A}}^T \mathcal{H}_\pm = \mathcal{H}_\pm \widehat{\mathcal{A}}$ holds.

It should be mentioned here that the preconditioner A_0 in

$$\mathcal{P} = \begin{bmatrix} A_0 & B^T \\ 0 & \pm C_0 \end{bmatrix}$$

can be chosen such that A_0 resembles the structure given by Forsgren *et al.*, i.e. $A_0 = G + B^T C^{-1} B$ which we will call FGG setup.

[1] The $\mathcal{H}_+$-MINRES method is defined in [16] and can be viewed as the standard MINRES procedure with the non-standard inner product defined by $\mathcal{H}_+$.

4.3. Eigenvalue analysis.

4.3.1. Bramble-Pasciak-like setup. In this section we analyze the eigenvalues of the preconditioned matrix $\widehat{\mathcal{A}}$ where the preconditioner is given by

$$\mathcal{P}_+ = \begin{bmatrix} A_0 & B^T \\ 0 & C_0 \end{bmatrix} \text{ with } \mathcal{H}_+ = \begin{bmatrix} A_0 & 0 \\ 0 & C + C_0 \end{bmatrix}$$

as the inner product. Note, the analysis presented here is not based on the assumption of C being positive definite. We restrict ourself to the case where $A_0 = A$ and assume that $\left(\lambda, \begin{bmatrix} x^T & y^T \end{bmatrix}^T\right)$ represents an eigenpair of $\widehat{\mathcal{A}}$. This gives

$$\begin{bmatrix} I - A^{-1}B^TC_0^{-1}B & A^{-1}B^T + A^{-1}B^TC_0^{-1}C \\ C_0^{-1}B & -C_0^{-1}C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}. \tag{4.10}$$

Assuming that $\lambda = 1$ (4.10) reduces to

$$-A^{-1}B^TC_0^{-1}Bx + A^{-1}B^Ty + A^{-1}B^TC_0^{-1}Cy = 0 \tag{4.11}$$

$$C_0^{-1}Bx - C_0^{-1}Cy = y \tag{4.12}$$

and (4.12) gives

$$Bx = (C_0 + C)y. \tag{4.13}$$

Substituting this into (4.11) results in

$$-A^{-1}B^TC_0^{-1}(C_0 + C)y + A^{-1}B^Ty + A^{-1}B^TC_0^{-1}Cy = 0y = 0 \tag{4.14}$$

This shows that no condition on y has to be imposed. In particular, we use $Bx = (C_0 + C)y$ with $y = 0$ and since the kernel of B is $n - m$ dimensional we have at least $n - m$ eigenvalues at 1. In addition, if we choose $(C_0 + C)y \in Im(B)$ we also get that $Bx = (C_0 + C)y$ and since the $Im(B)$ is a m dimensional subspace we have m eigenvalues at 1. This gives n eigenvalues at 1 for the preconditioned matrix $\widehat{\mathcal{A}}$.

The case $\lambda \neq 1$ can be analyzed using

$$x - A^{-1}B^TC_0^{-1}Bx + A^{-1}B^Ty + A^{-1}B^TC_0^{-1}Cy = \lambda x \tag{4.15}$$

$$C_0^{-1}Bx - C_0^{-1}Cy = \lambda y \tag{4.16}$$

Using (4.16) we get $C_0^{-1}Bx = \lambda y + C_0^{-1}Cy$ which we substitute into (4.15) to obtain $x = -A^{-1}B^Ty$. Putting this back into (4.16) gives

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

which gives m negative eigenvalues since both $(BA^{-1}B^T + C)$ and C_0 are assumed to be positive definite.

A similar analysis can be done for the configuration given by

$$\mathcal{P}_- = \begin{bmatrix} A_0 & B^T \\ 0 & -C_0 \end{bmatrix} \text{ and } \mathcal{H}_- = \begin{bmatrix} A_0 & 0 \\ 0 & C - C_0 \end{bmatrix}.$$

With $A_0 = A$ we obtain n unit eigenvalues plus the m positive eigenvalues of

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

As a result the convergence of the Bramble-Pasciak-like setup with $A_0 = A$ is given within at most $m + 1$ steps.

4.3.2. The FGG setup. We now analyze the eigenvalues of the Bramble-Pasciak-like method with the FGG setup of the preconditioner, i.e.

$$\mathcal{P}_+ = \begin{bmatrix} A_0 & B^T \\ 0 & \frac{1}{\alpha}C \end{bmatrix} \Longrightarrow \mathcal{P}_+^{-1} = \begin{bmatrix} A_0^{-1} & -\alpha A_0^{-1}B^T C^{-1} \\ 0 & \alpha C^{-1} \end{bmatrix}$$

where $A_0 = G + B^T C_0^{-1} B$ and $C_0 = \frac{1}{\alpha}C$. To study the eigenvalues of the preconditioned matrix $\widehat{\mathcal{A}} = \mathcal{P}_+^{-1}\mathcal{A}$, we assume again that an eigenpair of $\widehat{\mathcal{A}}$ is given, i.e.

$$A_0^{-1}Ax - \alpha A_0^{-1}B^T C^{-1}Bx + A_0^{-1}B^T y + \alpha A_0^{-1}B^T y = \lambda x \tag{4.17}$$

$$\alpha C^{-1}Bx - \alpha y = \lambda y. \tag{4.18}$$

For $\alpha > 0$ we get from (4.18) that

$$\alpha Bx - \alpha Cy = \lambda Cy$$

and therefore

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

Substituting the last result into (4.17) gives

$$Ax + \frac{\alpha(1-\lambda)}{\alpha+\lambda} B^T C^{-1}Bx = \lambda A_0 x$$

which corresponds to a quadratic eigenvalue problem. Note, that for $\alpha = -1$ the matrix $\widehat{\mathcal{A}}$ reduces to

$$\begin{bmatrix} A_0^{-1}(A + B^T C^{-1}B) & 0 \\ -C^{-1}B & I \end{bmatrix}$$

for which the eigenvalues are given by the n eigenvalues of

$$A_0^{-1}(A + B^T C^{-1}B)$$

and m unit ones. This gives complete coincidence with the setup given in [6].

Similar results can be obtained for

$$\mathcal{P}_- = \begin{bmatrix} A_0 & B^T \\ 0 & -\frac{1}{\alpha}C \end{bmatrix}.$$

Assuming an eigenpair of $\widehat{\mathcal{A}} = \mathcal{P}_-^{-1}\mathcal{A}$ is given, we get

$$\begin{bmatrix} A_0^{-1}A + \alpha A_0^{-1}B^T C^{-1}B & A_0^{-1}B^T - \alpha A_0^{-1}B^T \\ -\alpha C^{-1}B & \alpha I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} \tag{4.19}$$

which for $\alpha = 1$ again reduces to

$$\begin{bmatrix} A_0^{-1}(A + B^T C^{-1}B) & 0 \\ -C^{-1}B & I \end{bmatrix};$$

a matrix that has the eigendistribution of setup presented by Forsgren, Gill and Griffin. Figure 4.1 shows the eigendistribution for the Forsgren *et al.* configuration and the Bramble-Pasciak setup for different values of α for the matrix CVXQP3_S of dimension 175 taken from the CUTEr [8, 9] test set.

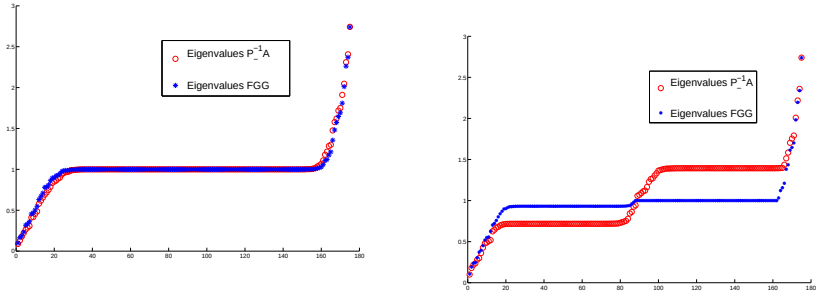

FIG. 4.1. *Eigenvalue distribution for BP-like_ and Forsgren et al. with* $\alpha = 1$ *(left) and* $\alpha = 0.9$ *(right).*

5. Numerical Experiments. The examples in this section are taken from the CUTEr [8, 9] test set. The matrix A is typically indefinite and C will be constructed to be positive definite.

The methods we compare in this section are the CG of Forsgren, Gill and Griffin and the Bramble-Pasciak-like CG with $\mathcal{P}_-$ and $\mathcal{H}_-$.

EXAMPLE 5.1. *In this example, we are looking at the matrix CVXQP1_M which is of size* 1500×1500. *C will be a diagonal matrix with entries of the form* 10^{-k} *on the diagonal where* $2 \leq k \leq 10$. *The preconditioners are defined by* $C_0 = 0.9C$ *and* $A_0 = diag(A) + B^T C_0^{-1} B$. *The results for the Bramble-Pasciak-like_ setup and the Forsgren-Gill-Griffin method are shown in Figure 5.1.*

EXAMPLE 5.2. *Again we are looking at the matrix CVXQP1_M. The matrix C will now be generated using the MATLAB command* `C=1e-1*sprandsym(m,.3)+speye(m);`, *i.e. an identity matrix plus a random sparse perturbation. The preconditioners are again defined by* $C_0 = 0.9C$ *and* $A_0 = diag(A) + B^T C_0^{-1} B$. *The results for the Bramble-Pasciak-like_ setup and the Forsgren-Gill-Griffin method are shown in Figure 5.2.*

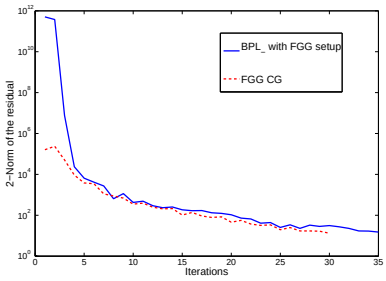

FIG. 5.1. *Diagonal C*

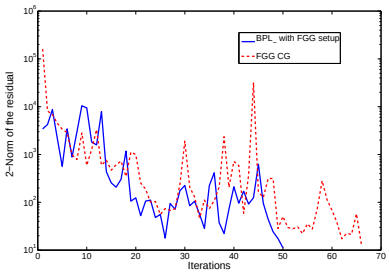

FIG. 5.2. *Randomly perturbed C*

6. Conclusions. We presented a reformulation of the saddle point problem which represents a framework for many well known methods for solving saddle point problems. We employed this structure to introduce a Bramble-Pasciak-like method based on a constrained preconditioning technique. We illustrated that competitive results can be obtained when applying this method to problems coming from optimization.

Acknowledgment. The author would like to thank Sue Dollar, Nick Gould and Andy Wathen for their helpful comments.

REFERENCES

[1] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[2] 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.

[3] 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).

[4] 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, New York, 2005.

[5] V. Faber and T. Manteuffel, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., 21 (1984), pp. 352–362.

[6] A. Forsgren, P. Gill, and J. Griffin, *Iterative solution of augmented systems arising in interior methods*, SIAM Journal on Optimization, to appear.

[7] 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.

[8] N. Gould, D. Orban, and P. Toint, *CUTEr and SifDec: A constrained and unconstrained testing environment, revisited*, ACM Transactions on Mathematical Software (TOMS), 29 (2003), pp. 373–394.

[9] N. I. M. Gould, D. Orban, and P. L. 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.

[10] A. Greenbaum, *Iterative methods for solving linear systems*, vol. 17 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[11] 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).

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

[13] J. Liesen and B. N. Parlett, *On nonsymmetric saddle point matrices that allow conjugate gradient iterations*, submitted to Numerische Mathematik, (2007).

[14] J. Liesen and Z. Strakoš, *On optimal short-term recurrences for generating orthogonal Krylov subspace bases*. submitted to SIAM review, 2006.

[15] 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.

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

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