Technical Report
RAL-TR-2005-026

CCLRC

Using constraint preconditioners with regularized saddle-point problems

H S Dollar N I M Gould W H A Schilders A J Wathen

October 31, 2005

Council for the Central Laboratory of the Research Councils

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

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

ISSN 1358-6254

RAL-TR-2005-026

Using constraint preconditioners with regularized saddle-point problems

H. Sue Dollar[1], Nicholas I. M. Gould[2,3],
Wil H. A. Schilders[4] and Andrew J. Wathen[5]

ABSTRACT

The problem of finding good preconditioners for the numerical solution of a certain important class of indefinite linear systems is considered. These systems are of a 2 by 2 block (KKT) structure in which the (2,2) block (denoted by $-C$) is assumed to be nonzero.

Keller, Gould and Wathen (SIAM J. Matrix Anal. Appl., 21(4):1300-1307, 2000) introduced the idea of using constraint preconditioners that have a specific 2 by 2 block structure for the case of C being zero. We shall give results concerning the spectrum and form of the eigenvectors when a preconditioner of the form considered by Keller, Gould and Wathen is used but the system we wish to solve may have $C \neq 0$. In particular, the results presented here indicate clustering of eigenvalues and, hence, faster convergence of Krylov subspace iterative methods when the entries of C are small; such a situation arises naturally in interior point methods for optimization and we present results for such problems which validate our conclusions.

---

[1] Department of Mathematics, University of Reading, Whiteknights, P.O. Box 220, Reading, Berkshire, RG6 6AX, England, UK. Email: H.S.Dollar@reading.ac.uk .

[2] Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, England, UK. Email: n.i.m.gould@rl.ac.uk .
Current reports available from "www.numerical.rl.ac.uk/reports/reports.shtml".

[3] This work was supported by the EPSRC grants GR/S42170.

[4] Philips Research Laboratories; Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands. Email: wil.schilders@philips.com . Also, Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, PO Box 513, 5600 MB Eindhoven, The Netherlands.

[5] Oxford University Computing Laboratory, Numerical Analysis Group, Wolfson Building, Parks Road, Oxford, OX1 3QD, England, UK. Email: Andy.Wathen@comlab.ox.ac.uk . Current reports available from "web.comlab.ox.ac.uk/oucl/publications/natr/index.html".

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

October 31, 2005

1 Introduction

The solution of systems of the form

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

where $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times m}$ are symmetric and $B \in \mathbb{R}^{m\times n}$, is often required in optimization and other various fields.

Example 1.1 (Nonlinear Programming). Consider the convex nonlinear optimization problem

$$\text{minimize } f(x) \text{ such that } c(x) \geq 0, \tag{1.2}$$

where $x \in \mathbb{R}^n$, and $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $-c : \mathbb{R}^n \mapsto \mathbb{R}^{\widehat{m}}$ are convex and twice differentiable. Primal-dual interior point methods [18] for this problem aim to track solutions to the (perturbed) optimality conditions

$$\nabla f(x) = B^T(x)y \text{ and } Yc(x) = \mu e, \tag{1.3}$$

where y are Lagrange multipliers (dual variables), e is the vector of ones,

$$B(x) = \nabla c(x) \text{ and } Y = \text{diag}\{y_1, y_2, \ldots, y_{\widehat{m}}\},$$

as the positive scalar parameter μ is decreased to zero. The Newton correction $(\Delta x, \Delta y)$ to the solution estimate (x, y) of (1.3) satisfy the equation [3]:

$$\begin{bmatrix} A(x,y) & -B^T(x) \\ YB(x) & C(x) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + B^T(x)y \\ -Yc(x) + \mu e \end{bmatrix},$$

where

$$A(x,y) = \nabla_{xx} f(x) - \sum_{i=1}^{\widehat{m}} y_i \nabla_{xx} c_i(x) \text{ and } C(x) = \text{diag}\{c_1(x), c_2(x), \ldots, c_{\widehat{m}}(x)\}.$$

It is common to eliminate the variables Δy from the Newton system. Since this may introduce unwarranted ill conditioning, it is often better [8] to isolate the effects of poor conditioning by partitioning the constraints so that the values of those indexed by $\mathcal{I}$ are "large" while those indexed by $\mathcal{A}$ are "small", and instead to solve

$$\begin{bmatrix} A + B_{\mathcal{I}}^T C_{\mathcal{I}}^{-1} Y_{\mathcal{I}} B_{\mathcal{I}} & B_{\mathcal{A}}^T \\ B_{\mathcal{A}} & -C_{\mathcal{A}} Y_{\mathcal{A}}^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ -\Delta y_{\mathcal{A}} \end{bmatrix} = \begin{bmatrix} -\nabla f + B_{\mathcal{A}}^T y_{\mathcal{A}} + \mu B_{\mathcal{I}}^T C_{\mathcal{I}}^{-1} e \\ -c_{\mathcal{A}} + \mu Y_{\mathcal{A}}^{-1} e \end{bmatrix}$$

where, for brevity, we have dropped the dependence on x and y. The matrix $C_{\mathcal{A}} Y_{\mathcal{A}}^{-1}$is symmetric and positive definite; as the iterates approach optimality, the entries of this matrix become small. The entries of $B_{\mathcal{I}}^T C_{\mathcal{I}}^{-1} Y_{\mathcal{I}} B_{\mathcal{I}}$ also become small when close to optimality.

Example 1.2 (Stokes). Mixed finite element (and other) discretization of the Stokes equations

$$\begin{aligned} -\nabla^2 \vec{u} + \nabla p &= \vec{f} \quad \text{in } \Omega \\ \nabla \cdot \vec{u} &= 0 \quad \text{in } \Omega, \end{aligned}$$

for the fluid velocity $\vec{u}$ and pressure p in the domain $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$ yields linear systems in the saddle-point form (1.1) (for derivation and the following properties of this example see [6]). The symmetric block A arises from the diffusion terms $-\nabla^2 \vec{u}$ and B^T represents the discrete gradient operator whilst B represents its adjoint, the (negative) divergence. When (inf-sup) stable mixed finite element spaces are employed, $C = 0$, however for equal order and other spaces which are not inherently stable, stabilized formulations yield symmetric and positive semi-definite matrices C which typically have a large-dimensional kernel - for example for the famous $\boldsymbol{Q}_1$–$\boldsymbol{P}_0$ element which has piecewise bilinear velocities and piecewise constant pressures in 2-dimensions, C typically has a kernel of dimension $m/4$.

We shall assume that $0 < m \leq n$ and B is of full rank. Various preconditioners which take the general form

$$\mathcal{P}_{_C} = \begin{bmatrix} G & B^T \\ B & -C \end{bmatrix}, \tag{1.4}$$

where $G \in \mathbb{R}^{n \times n}$ is some symmetric matrix, have been considered (for example, see [4, 13, 17].) When $C = 0$, (1.4) is commonly known as a constraint preconditioner [2, 12]. In practice C is often positive semi-definite (and frequently diagonal).

In interior-point methods a sequence of such problems are solved with the entries in C generally becoming small as the optimization iteration progresses. That is, the regularization is successively reduced as the optimizer gets closer to the minimum. For the Stokes problem, the entries of C are generally small since they scale with the underlying mesh size and so reduce for finer grids. This motivates us to look at the spectral properties of $\mathcal{P}^{-1}\mathcal{A}_{_C}$, where

$$\mathcal{P} = \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix}, \tag{1.5}$$

but $C \neq 0$, Section 2.

The obvious advantage in being able to use such a constraint preconditioner is as follows: if B remains constant in each system of the form (1.1), and we choose G in our preconditioner to remain constant, then the preconditioner $\mathcal{P}$ will be unchanged. Any factorizations required to carry out the preconditioning steps in a Krylov subspace iteration will only need to be done once and then used during each execution of the chosen Krylov subspace iteration, instead of carrying out the factorizations at the beginning of each execution.

For symmetric (and in general normal) matrix systems, the convergence of an applicable iterative method is determined by the distribution of the eigenvalues of the coefficient matrix. It is often desirable for the number of distinct eigenvalues to be small so that the

rate of convergence is rapid. For non-normal systems the convergence is not so readily described, see [11, page 6].

2 Preconditioning $\mathcal{A}_C$ by $\mathcal{P}$

Suppose that we precondition $\mathcal{A}_C$ by $\mathcal{P}$, where $\mathcal{P}$ is defined in (1.5). The decision to investigate this form of preconditioner is motivated in Section 1. We shall use the following assumptions in our theorems:

A1 $B \in \mathbb{R}^{m\times n}$ $(m < n)$ has full rank,

A2 $Z \in \mathbb{R}^{n\times(n-m)}$ is a basis for the nullspace of B,

A3 C has rank $p > 0$ and is factored as EDE^T, where $E \in \mathbb{R}^{m\times p}$ and has orthonormal columns, and $D \in \mathbb{R}^{p\times p}$ is non-singular,

A4 $F \in \mathbb{R}^{m\times(m-p)}$ is such that its columns form a basis for the nullspace of C.

Theorem 2.1. Assume that **A1**–**A4** hold, then the matrix $\mathcal{P}^{-1}\mathcal{A}_C$ has:

- at least $2(m-p)$ eigenvalues at 1,
- its non-unit eigenvalues defined by the finite (and non-unit) eigenvalues of quadratic eigenvalue problem problem

  $$0 = \lambda^2 B^T E D^{-1} E^T B u - \lambda(G + 2B^T E D^{-1} E^T B)u + (A + B^T E D^{-1} E^T B)u,$$

  subject to $u = Zw_1 + B^T(BB^T)^{-1}Ew_2$ for some $w_2 \neq 0$.

Proof. We shall consider the cases of $p = m$ and $0 < p < m$ separately.

Case $p = m$: The generalized eigenvalue problem takes the form

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} G & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{2.6}$$

Expanding this out we obtain

$$Ax + B^T y = \lambda Gx + \lambda B^T y, \tag{2.7}$$
$$Bx - Cy = \lambda Bx. \tag{2.8}$$

From (2.8) we deduce that either $\lambda = 1$ and $y = 0$, or $\lambda \neq 1$. If the former holds, then (2.7) implies that x must satisfy

$$Ax = Gx.$$

Thus, the associated eigenvectors will take the form

$$\begin{bmatrix} x^T & 0^T \end{bmatrix}^T,$$

where $x \neq 0$ satisfies $Ax = Gx$. There is no guarantee that such an eigenvector will exist, and therefore no guarantee that there are any unit eigenvalues.

If $\lambda \neq 1$, then Equation (2.8) and the non-singularity of C gives

$$y = (1-\lambda)\, C^{-1}Bx, \quad x \neq 0.$$

By substituting this into (2.7) and rearranging we obtain the quadratic eigenvalue problem

$$\left(\lambda^2 B^T C^{-1} B - \lambda \left(G + 2B^T C^{-1} B\right) + A + B^T C^{-1} B\right) x = 0. \tag{2.9}$$

The non-unit eigenvalues of (2.6) are therefore defined by the finite (non-unit) eigenvalues of (2.9). Note that since $B^T C^{-1} B$ has rank m, (2.9) has $2n-(n-m) = n+m$ finite eigenvalues, but at most n linearly independent eigenvectors [16, Section 3.1]. Hence, $\mathcal{P}^{-1}\mathcal{A}_C$ has at most n linearly independent eigenvectors associated with the non-unit eigenvalues when $p = m$.

Now, assumption **A2** implies that

$$C^{-1} \;=\; ED^{-1}E^T,$$

and, hence, letting $u = x$ we complete our proof for the case $p = m$.

Case $0 < p < m$: Any $y \in \mathbb{R}^m$ can be written as $y = Ey_e + Fy_f$. Substituting this into (2.6) and premultiplying the resulting generalized eigenvalue problem by

$$\begin{bmatrix} I & 0 \\ 0 & E^T \\ 0 & F^T \end{bmatrix},$$

we obtain

$$\left[\begin{array}{cc|c} A & B^TE & B^TF \\ E^TB & -D & 0 \\ \hline F^TB & 0 & 0 \end{array}\right] \begin{bmatrix} x \\ y_e \\ y_f \end{bmatrix} = \lambda \left[\begin{array}{cc|c} G & B^TE & B^TF \\ E^TB & 0 & 0 \\ \hline F^TB & 0 & 0 \end{array}\right] \begin{bmatrix} x \\ y_e \\ y_f \end{bmatrix}. \tag{2.10}$$

Noting that the (3,3) block has dimension $(m-p) \times (m-p)$ and is a zero matrix in both coefficient matrices, we can apply Theorems 2.1 and 2.3 from [12] to obtain:

- $\mathcal{P}^{-1}\mathcal{A}_C$ has an eigenvalue at 1 with multiplicity $2(m-p)$,
  - there are $m-p$ linearly independent eigenvectors of the form $\begin{bmatrix} 0^T & y_e^T & y_f^T \end{bmatrix}^T$ that correspond to the case $\lambda = 1$,
  - there are at most n eigenvectors of the form $\begin{bmatrix} x^T & 0^T & y_f^T \end{bmatrix}^T$ arising from $Ax = \omega Gx$ with x linearly independent, $\omega = 1$, and $\lambda = 1$,

- the remaining $n - m + 2p$ eigenvalues are defined by the generalized eigenvalue problem

$$N^T \begin{bmatrix} A & B^T E \\ E^T B & -D \end{bmatrix} Nw = \lambda N^T \begin{bmatrix} G & B^T E \\ E^T B & 0 \end{bmatrix} Nw, \tag{2.11}$$

  where N is an $(n+p) \times (n-m+2p)$ basis for the nullspace of $\begin{bmatrix} F^T B & 0 \end{bmatrix}$.

One choice for N is

$$N = \begin{bmatrix} Z & B^T \left(BB^T\right)^{-1} E & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Substituting this into (2.11) we obtain the generalized eigenvalue problem

$$\left[\begin{array}{cc|c} Z^T AZ & Z^T AW & 0 \\ W^T AZ & W^T AW & I \\ \hline 0 & I & -D \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \lambda \left[\begin{array}{cc|c} Z^T GZ & Z^T GW & 0 \\ W^T GZ & W^T GW & I \\ \hline 0 & I & 0 \end{array}\right] \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \tag{2.12}$$

where $W = B^T \left(BB^T\right)^{-1} E$. This generalized eigenvalue problem resembles that of (2.6) in the first case considered in this proof. Therefore, the non-unit eigenvalues of $\mathcal{P}^{-1}\mathcal{A}_C$ are equal to the finite (and non-unit) eigenvalues of the quadratic eigenvalue problem

$$\left(\lambda^2 \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix} - \lambda \begin{bmatrix} Z^T GZ & Z^T GW \\ W^T GZ & W^T GW + 2D^{-1} \end{bmatrix} + \begin{bmatrix} Z^T AZ & Z^T AW \\ W^T AZ & W^T AW + D^{-1} \end{bmatrix}\right) \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 0, \tag{2.13}$$

subject to $w_2 \neq 0$.

Let $S \in \mathbb{R}^{m \times m}$ be the non-singular matrix $\begin{bmatrix} Z & W \end{bmatrix}$, then

$$S^T B^T E D^{-1} B S = \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix}.$$

Hence, the quadratic eigenvalue problem (2.13) can be written as

$$S^T \left(\lambda^2 B^T E D^{-1} E^T B - \lambda(G + 2B^T E D^{-1} E^T B) + A + B^T E D^{-1} E^T B\right) S \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 0,$$

subject to $w_2 = 0$. The non-unit eigenvalues of $\mathcal{P}^{-1}\mathcal{A}_C$ will, therefore, also be defined by the finite (and non-unit) eigenvalues of the equivalent quadratic eigenvalue problem

$$\left(\lambda^2 B^T E D^{-1} E^T B - \lambda(G + 2B^T E D^{-1} E^T B) + A + B^T E D^{-1} E^T B\right) u = 0, \tag{2.14}$$

where u can be expressed as $u = Zw_1 + B^T (BB^T)^{-1} w_2$ for some $w_2 \neq 0$. There are at most $n - m + p$ linearly independent eigenvectors associated with the finite eigenvalues of this quadratic eigenvalue problem, implying that $\mathcal{P}^{-1}\mathcal{A}_C$ has at most $n - m + p$ linearly independent eigenvectors of the form $\begin{bmatrix} x^T & y_e^T & y_f^T \end{bmatrix}^T$. □

The definition of some of the eigenvalues through a quadratic eigenvalue problem is interesting and shall be examined in more detail later on. If the matrix C has full rank (as is generally the case when using interior point methods to solve optimization problems), the requirement that $u = Zw_1 + B^T(BB^T)^{-1}Ew_2$ for some $w_2 \neq 0$ is equivalent to the trivial $u \neq 0$.

Theorem 2.2. Assume that **A1**–**A4** hold. Then the matrix $\mathcal{P}^{-1}\mathcal{A}_C$ has $n+m$ eigenvalues as defined in Theorem 2.1 and $m-p+i+j$ linearly independent eigenvectors. There are

1. $m-p$ eigenvectors of the form $[\ 0^T \quad 0^T \quad y_f^T\ ]^T$ that correspond to the case $\lambda = 1$;
2. i $(0 \leq i \leq n)$ eigenvectors of the form $[\ x^T \quad 0^T \quad y_f^T\ ]^T$ arising from $Ax = \sigma Gx$ with x linearly independent, $\sigma = 1$, and $\lambda = 1$;
3. j $(0 \leq j \leq n-m+p)$ eigenvectors of the form $[\ x^T \quad y_e^T \quad y_f^T\ ]^T$ that correspond to the case $\lambda \neq 1$ and $Cy \neq 0$ with $y = Fy_f + Ey_e$.

Proof. We need only prove that the $m-p+i+j$ eigenvectors of $\mathcal{P}^{-1}\mathcal{A}_C$ defined in the proof of Theorem 2.1 are linearly independent.

We need to show that

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ y_{f1}^{(1)} & \cdots & y_{f(m-p)}^{(1)} \end{bmatrix} \begin{bmatrix} a_1^{(1)} \\ \vdots \\ a_{m-p}^{(1)} \end{bmatrix} + \begin{bmatrix} x_1^{(2)} & \cdots & x_i^{(2)} \\ 0 & \cdots & 0 \\ y_{f1}^{(2)} & \cdots & y_{fi}^{(2)} \end{bmatrix} \begin{bmatrix} a_1^{(2)} \\ \vdots \\ a_i^{(2)} \end{bmatrix} + \begin{bmatrix} x_1^{(3)} & \cdots & x_j^{(3)} \\ y_{e1}^{(3)} & \cdots & y_{ej}^{(3)} \\ y_{f1}^{(3)} & \cdots & y_{fj}^{(3)} \end{bmatrix} \begin{bmatrix} a_1^{(3)} \\ \vdots \\ a_j^{(3)} \end{bmatrix} \tag{2.15}$$

implies that the vectors $a^{(l)}$ $(l = 1, 2, 3)$ are zero vectors. Multiplying (2.15) by $\mathcal{A}_C$ and $\mathcal{P}^{-1}$, and recalling that in the previous equation the first matrix arises from $\lambda_l = 1$ $(l = 1, \ldots, m)$, the second matrix from the case that $\lambda_l = 1$ and $\omega_l = 1$ $(l = 1, \ldots, i)$, and the third matrix from $\lambda_l \neq 1$ $(l = 1, \ldots, j)$, gives

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ y_{f1}^{(1)} & \cdots & y_{f(m-p)}^{(1)} \end{bmatrix} \begin{bmatrix} a_1^{(1)} \\ \vdots \\ a_{m-p}^{(1)} \end{bmatrix} + \begin{bmatrix} x_1^{(2)} & \cdots & x_i^{(2)} \\ 0 & \cdots & 0 \\ y_{f1}^{(2)} & \cdots & y_{fi}^{(2)} \end{bmatrix} \begin{bmatrix} a_1^{(2)} \\ \vdots \\ a_i^{(2)} \end{bmatrix} + \begin{bmatrix} x_1^{(3)} & \cdots & x_j^{(3)} \\ y_{e1}^{(3)} & \cdots & y_{ej}^{(3)} \\ y_{f1}^{(3)} & \cdots & y_{fj}^{(3)} \end{bmatrix} \begin{bmatrix} \lambda_1^{(3)} a_1^{(3)} \\ \vdots \\ \lambda_j^{(3)} a_j^{(3)} \end{bmatrix}. \tag{2.16}$$

Subtracting (2.15) from (2.16) gives

$$\begin{bmatrix} x_1^{(3)} & \cdots & x_j^{(3)} \\ y_{e1}^{(3)} & \cdots & y_{ej}^{(3)} \\ y_{f1}^{(3)} & \cdots & y_{fj}^{(3)} \end{bmatrix} \begin{bmatrix} (\lambda_1^{(3)} - 1)a_1^{(3)} \\ \vdots \\ (\lambda_j^{(3)} - 1)a_j^{(3)} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.17}$$

The linear independence of $\begin{bmatrix} x_l^{(3)T} & y_{el}^{(3)T} \end{bmatrix}^T$ $(l = 1, \ldots, j)$ in (2.14) gives rise to $(\lambda_l^{(3)} - 1)a_l^{(3)} = 0$ $(l = 1, \ldots, j)$. The eigenvalues $\lambda_l^{(3)}$ $(l = 1, \ldots, j)$ are non-unit which implies that $a_l^{(3)} = 0$ $(l = 1, \ldots, j)$.

We also have linear independence of $x_l^{(2)}$ $(l = 1, \ldots, i)$, implies that $a_l^{(2)} = 0$ $(l = 1, \ldots, i)$. Equation 2.15 simplifies to

$$\begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ y_{f1}^{(1)} & \cdots & y_{f(m-p)}^{(1)} \end{bmatrix} \begin{bmatrix} a_1^{(1)} \\ \vdots \\ a_{(m-p)}^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}.$$

However, $y_{fl}^{(1)}$ $(l = 1, \ldots, m - p)$ are linearly independent giving $a_l^{(1)} = 0$.

□

Remark; $\mathcal{P}^{-1}\mathcal{A}_C$ has at least $2(m-p)$ unit eigenvalues, but there is no guarantee that the associated eigenvectors are all linearly independent. However, we can divide these eigenvectors into two groups such that all the eigenvectors in a group are linearly independent and each group has at least $m - p$ members.

2.1 Analysis of the quadratic eigenvalue problem

We note that the quadratic eigenvalue problem (2.14) can have negative and complex eigenvalues, see [16]. The following theorem gives sufficient conditions for general quadratic eigenvalue problems to have real and positive eigenvalues.

Theorem 2.3. Consider the quadratic eigenvalue problem

$$\left(\lambda^2 K - \lambda L + M\right) x = 0, \tag{2.18}$$

where $M, L, \in \mathbb{R}^{n\times n}$ are symmetric positive definite, and $K \in \mathbb{R}^{n\times n}$ is symmetric positive semidefinite. Define $\gamma(M, L, K)$ to be

$$\gamma(M, L, K) = \min\left\{(x^T Lx)^2 - 4(x^T Mx)(x^T Kx) : \|x\|_2 = 1\right\}.$$

If $\gamma(M, L, K) > 0$, then the eigenvalues λ are real and positive.

Proof. From [16, Section 1] we know that under our assumptions the quadratic eigenvalue problem

$$\left(\mu^2 M + \mu L + K\right) x = 0$$

has real and negative eigenvalues. Suppose we divide this equation by μ^2 and set $\lambda = -1/\mu$. The quadratic eigenvalue problem (2.18) is obtained, and since μ is real and negative, λ is real and positive. □

We would like to be able to use the above theorem to show that, under suitable assumptions, all the eigenvalues of $\mathcal{P}^{-1}\mathcal{H}$ are real and positive. Let

$$\widetilde{D} = B^T E D^{-1} E^T B, \tag{2.19}$$

where D and E are as defined in Theorem 2.1. If we assume that $A + \widetilde{D}$ is positive definite, then we may write $A + \widetilde{D} = R^T R$ for some nonsingular matrix R. The quadratic eigenvalue (2.14) is similar to

$$\left(\lambda^2 R^{-T} \widetilde{D} R^{-1} - \lambda R^{-T}(G + 2\widetilde{D})R^{-1} + I\right) z = 0,$$

where $z = Rw$. Thus, if we assume that $A + \widetilde{D}$ and $G + 2\widetilde{D}$ are positive definite, and can show that

$$\gamma(I, R^{-T}(G + 2\widetilde{D})R^{-1}, R^{-T}\widetilde{D}R^{-1}) > 0,$$

where $\gamma(\cdot,\cdot,\cdot)$ is as defined in Theorem 2.3, then we can apply the above theorem to show that (2.14) has real and positive eigenvalues.

Let us assume that $\|z\|_2 = 1$, then

$$\begin{aligned}
& \left(z^T R^{-T}\left(G + 2\widetilde{D}\right) R^{-1} z\right)^2 - 4 z^T z z^T R^{-T} \widetilde{D} R^{-1} z \\
= \; & \left(z^T R^{-T} G R^{-1} z + 2 z^T R^{-T} \widetilde{D} R^{-1} z\right)^2 - 4 z^T R^{-T} \widetilde{D} R^{-1} z \\
= \; & \left(z^T R^{-T} G R^{-1} z\right)^2 + 4 z^T R^{-T} \widetilde{D} R^{-1} z \left(z^T R^{-T} G R^{-1} z + z^T R^{-T} \widetilde{D} R^{-1} z - 1\right) \\
= \; & \left(w^T G w\right)^2 + 4 w^T \widetilde{D} w \left(w^T G w + w^T \widetilde{D} w - 1\right),
\end{aligned} \tag{2.20}$$

where $1 = \|z\|_2 = \|Rw\|_2 = \|w\|_{A+\widetilde{D}}$. Clearly, we can guarantee that (2.20) is positive if

$$w^T G w + w^T \widetilde{D} w > 1 \quad \text{for all } w \text{ such that } \|w\|_{A+\widetilde{D}} = 1,$$

that is

$$\frac{w^T G w + w^T \widetilde{D} w}{w^T\left(A + \widetilde{D}\right) w} > \frac{w^T\left(A + \widetilde{D}\right) w}{w^T\left(A + \widetilde{D}\right) w} \quad \text{for all } w \neq 0.$$

Rearranging we find that we require

$$w^T G w > w^T A w$$

for all $w \neq 0$. Thus we need only scale any positive definite G such that $\frac{w^T G w}{w^T w} > \|A\|_2^2$ for all $w \neq 0$ to guarantee that (2.20) is positive for all w such that $\|w\|_{A+\widetilde{D}} = 1$. For example, we could choose $G = \alpha I$, where $\alpha > \|A\|_2^2$.

Using the above in conjunction with Theorem 2.1 we obtain the following technical result:

Theorem 2.4. Suppose that **A1**–**A4** hold and $\widetilde{D}$ is as defined in (2.19). Further, assume that $A + \widetilde{D}$ and $G + 2\widetilde{D}$ are symmetric positive definite, $\widetilde{D}$ is symmetric positive semidefinite and

$$\min \left\{ (z^T G z)^2 + 4(z^T \widetilde{D} z)(z^T G z + z^T \widetilde{D} z - 1) : \|z\|_{A+\widetilde{D}} = 1 \right\} > 0,$$

then all the eigenvalues of $\mathcal{P}^{-1}\mathcal{A}_C$ are real and positive.

Observing that the coefficient matrices in (2.10) are of the form of those considered by Gould, Hribar and Nocedal [9], we could apply a projected preconditioned conjugate gradient method to solve (1.1) if all the eigenvalues of $\mathcal{P}^{-1}\mathcal{A}_C$ are real and positive and we have a decomposition of C as in **A3**. Theorem 2.4 therefore gives conditions which allow us to use such a method. Dollar gives a variant of this method in which no decomposition of C is required, see [5, Section 5.5].

3 Convergence

In the context of this paper, the convergence of an iterative method under preconditioning is not only influenced by the spectral properties of the coefficient matrix, but also by the relationship between m, n and p. We can determine an upper bound on the number of iterations of an appropriate Krylov subspace method by considering minimum polynomials of the coefficient matrix.

Definition 3.1. *Let $\mathcal{A} \in \mathbb{R}^{(n+m)\times(n+m)}$. The monic polynomial f of minimum degree such that $f(\mathcal{A}) = 0$ is called the minimum polynomial of $\mathcal{A}$.*

Krylov subspace theory states that iteration with any method with an optimality property, e.g. GMRES, will terminate when the degree of the minimum polynomial is attained, [15]. In particular, the degree of the minimum polynomial is equal to the dimension of the corresponding Krylov subspace (for general b), [14, Proposition 6.1].

Theorem 3.1. Suppose that the assumptions of Theorem 2.4 hold. The dimension of the Krylov subspace $\mathcal{K}(\mathcal{P}^{-1}\mathcal{A}_C, b)$ is at most $\min\{n - m + 2p + 2, n + m\}$.

Proof. Suppose that $0 < p < m$. As in the proof to Theorem 2.1, the generalized eigenvalue problem can be written as

$$\left[\begin{array}{cc|c} A & B^TE & B^TF \\ E^TB & -D & 0 \\ \hline F^TB & 0 & 0 \end{array}\right] \left[\begin{array}{c} x \\ y_e \\ y_f \end{array}\right] = \lambda \left[\begin{array}{cc|c} G & B^TE & B^TF \\ E^TB & 0 & 0 \\ \hline F^TB & 0 & 0 \end{array}\right] \left[\begin{array}{c} x \\ y_e \\ y_f \end{array}\right]. \qquad (3.21)$$

Thus, applying Theorem 3.5 from [12], the dimension of the Krylov subspace $\mathcal{K}(\mathcal{P}^{-1}\mathcal{A}_C, b)$ is at most $\min\{n+p-(m-p)+2, n+m\} = \min\{n-m+2p+2, n+m\}$.

If $p = m$, then trivially $\mathcal{K}(\mathcal{P}^{-1}\mathcal{A}_C, b)$ has dimension at most $\min\{n-m+2p+2, n+m\}$.

□

3.1 Clustering of eigenvalues when $\|C\|$ is small

When using interior-point methods to solve optimization problems, the matrix C is generally diagonal and of full rank. In this case, Theorem 3.1 would suggest that there is little advantage of using a constraint preconditioner of the form $\mathcal{P}$ over any other preconditioner. However, in interior-point methods the entries of C also become small as we get close to optimality and, hence, $\|C\|$ is small. In the following we shall assume that the norm considered is the ℓ_2 norm, but the results can be generalized to other norms.

Theorem 3.2. Let $\zeta > 0$, $\delta \geq 0$, $\varepsilon \geq 0$ and $\delta^2 + 4\zeta(\delta - \varepsilon) \geq 0$ then the roots of the quadratic function

$$\lambda^2\zeta - \lambda(\delta + 2\zeta) + \varepsilon + \zeta = 0$$

satisfy

$$\lambda = 1 + \frac{\delta}{2\zeta} \pm \mu, \quad \mu \leq \sqrt{2}\max\left\{\frac{\delta}{2\zeta}, \sqrt{\frac{|\delta-\varepsilon|}{\zeta}}\right\}$$

Proof. The roots of the quadratic equation satisfies

$$\begin{aligned} \lambda &= \frac{\delta + 2\zeta \pm \sqrt{(\delta+2\varepsilon)^2 - 4(\varepsilon+\zeta)}}{2\zeta} = 1 + \frac{\delta}{2\zeta} \pm \frac{\sqrt{\delta^2 + 4\zeta(\delta-\varepsilon)}}{2\zeta} \\ &= 1 + \frac{\delta}{2\zeta} \pm \sqrt{\left(\frac{\delta}{2\zeta}\right)^2 + \frac{\delta-\varepsilon}{\zeta}} \end{aligned}$$

If $\dfrac{\delta-\varepsilon}{\zeta} \geq 0$, then

$$\begin{aligned}\sqrt{\left(\frac{\delta}{2\zeta}\right)^2+\frac{\delta-\varepsilon}{\zeta}} &\leq \sqrt{2\max\left\{\left(\frac{\delta}{2\zeta}\right)^2,\frac{\delta-\varepsilon}{\zeta}\right\}}\\ &= \sqrt{2}\max\left\{\frac{\delta}{2\zeta},\sqrt{\frac{\delta-\varepsilon}{\zeta}}\right\}.\end{aligned}$$

If $\dfrac{\delta-\varepsilon}{\zeta} \leq 0$, then the assumption $\delta^2+4\zeta(\delta-\varepsilon)\geq 0$ implies that

$$\left(\frac{\delta}{2\zeta}\right)^2 \geq \frac{\varepsilon-\delta}{\zeta} \geq 0.$$

Hence,

$$\sqrt{\left(\frac{\delta}{2\zeta}\right)^2+\frac{\delta-\varepsilon}{\zeta}} \leq \frac{\delta}{2\zeta} < \sqrt{2}\max\left\{\frac{\delta}{2\zeta},\sqrt{\frac{\varepsilon-\delta}{\zeta}}\right\}.$$

□

Remark: the important point to notice is that if $\zeta \gg \delta$ and $\zeta \gg \varepsilon$, then $\lambda \approx 1$ in Theorem 3.2.

Theorem 3.3. Assume that **A1**–**A4** hold, $A+\widetilde{D}$ and $G+2\widetilde{D}$ are symmetric positive definite, $\widetilde{D}$ is symmetric positive semidefinite and

$$\min\left\{(z^TGz)^2+4(z^T\widetilde{D}z)(z^TGz+z^T\widetilde{D}z-1) : \|z\|_{A+\widetilde{D}}=1\right\} > 0,$$

then all the eigenvalues of $\widetilde{P}^{-1}\mathcal{A}_C$ are real and positive. In addition, the eigenvalues λ of (2.14) subject to $E^TBu \neq 0$, will also satisfy

$$|\lambda-1| \leq \mathcal{O}(\|C\|).$$

Proof. That the eigenvalues of $\widetilde{P}^{-1}\mathcal{A}_C$ are real and positive follows directly from Theorem 2.4.

Suppose that $C = EDE^T$ is a reduced singular value decomposition of C, where the columns of $E \in \mathbb{R}^{m\times p}$ are orthogonal and $D \in \mathbb{R}^{p\times p}$ is diagonal with entries d_j that are non-negative and in non-increasing order.

In the following, $\|.\| = \|.\|_2$, so that

$$\|C\| = \|D\| = d_1.$$

Premultiplying the quadratic eigenvalue problem (2.14) by u^T gives

$$\begin{aligned} 0 &= \lambda^2 u^T \widetilde{D} u - \lambda(u^T G u + 2u^T \widetilde{D} u) \\ &\quad +(u^T A u + u^T \widetilde{D} u). \end{aligned} \tag{3.22}$$

Assume that $v = E^T B u$ and $\|v\| = 1$, where u is an eigenvector of the above quadratic eigenvalue problem, then

$$u^T \widetilde{D} u = v^T D^{-1} v = \frac{v_1^2}{d_1} + \frac{v_2^2}{d_2} + \ldots + \frac{v_m^2}{d_m} \geq \frac{v^T v}{d_1} = \frac{1}{\|C\|}.$$

Hence,

$$\frac{1}{u^T \widetilde{D} u} \leq \|C\|.$$

Let $\zeta = u^T \widetilde{D} u$, $\delta = u^T G u$ and $\varepsilon = u^T A u$, then (3.22) becomes

$$\lambda^2 \zeta - \lambda(\delta + 2\zeta) + \varepsilon + \zeta = 0.$$

From Lemma 3.2, λ must satisfy

$$\lambda = 1 + \frac{\delta}{2\zeta} \pm \mu, \quad \mu \leq \sqrt{2} \max\left\{ \frac{\delta}{2\zeta}, \sqrt{\frac{|\delta - \varepsilon|}{\zeta}} \right\}.$$

Now $\delta \leq c\|G\|$, $\varepsilon \leq c\|G\|$, where c is an upper bound on $\|u\|$ and u are eigenvectors of (2.14) subject to $\left\|E^T B u\right\| = 1$. Hence, the eigenvalues of (2.14) subject to $E^T B u \neq 0$ satisfy

$$|\lambda - 1| = \mathcal{O}(\|C\|).$$

□

This clustering of part of the spectrum of $\mathcal{P}^{-1}\mathcal{A}_C$ will often translate into a speeding up of the convergence of a selected Krylov subspace method, [1, Section 1.3].

3.2 Numerical Examples

We shall verify our theoretical results by considering some simple saddle point systems.

Example 3.4 (C nonsingular). Consider the matrices

$$\mathcal{A}_C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

so that $m = p = 1$ and $n = 2$. The preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}_C$ has eigenvalues at $\frac{1}{2}$, $2 - \sqrt{2}$ and $2 + \sqrt{2}$. The corresponding eigenvectors are $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$, $\begin{bmatrix} 1 & 0 & (\sqrt{2} - 1) \end{bmatrix}^T$

and $\begin{bmatrix} 1 & 0 & -(\sqrt{2}+1) \end{bmatrix}^T$ respectively. The preconditioned system $\widetilde{P}^{-1}\mathcal{A}_C$ has all non-unit eigenvalues, but this does not go against Theorem 2.1 because $m-p=0$. With our choices of $\mathcal{A}_C$ and $\widetilde{P}$, and setting $D=I$ and $E=I$ ($C=EDE^T$), the quadratic eigenvalue problem (2.14) is

$$\left(\lambda^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\right) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0.$$

This quadratic eigenvalue problem has three finite eigenvalues which are $\lambda = \frac{1}{2}$, $\lambda = 2-\sqrt{2}$ and $\lambda = 2+\sqrt{2}$.

Example 3.5 (C semidefinite). Consider the matrices

$$\mathcal{A}_C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

so that $m=2$, $n=2$ and $p=1$. The preconditioned matrix $\mathcal{P}^{-1}\mathcal{A}_C$ has two unit eigenvalues and a further two at $\lambda = 2-\sqrt{2}$ and $\lambda = 2+\sqrt{2}$. There is just one linearly independent eigenvector associated with the unit eigenvector; specifically this is $\begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T$. For the non-unit eigenvalues, the eigenvectors are $\begin{bmatrix} 0 & 1 & 0 & (\sqrt{2}-1) \end{bmatrix}^T$ and $\begin{bmatrix} 0 & 1 & 0 & -(\sqrt{2}+1) \end{bmatrix}^T$ respectively.

Since $2(m-p)=2$, we correctly expected there to be at least two unit eigenvalues, Theorem 2.1. The remaining eigenvalues will be defined by the quadratic eigenvalue problem (2.14):

$$\left(\lambda^2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0, \quad u_2 \neq 0,$$

where $D=[1]$ and $E = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ are used as factors of C. This quadratic eigenvalue problem has three finite eigenvalues of which two correspond to the case $u = Z^T w_1 + B^T(BB^T)^{-1}Ew_2$ for some $w_2 \neq 0$, i.e. $u_2 \neq 0$. These are $\lambda = 2-\sqrt{2}$ and $\lambda = 2+\sqrt{2}$; the corresponding eigenvectors have $u_1 = 0$.

Example 3.6 (C with small entries). Suppose that $\mathcal{A}_C$ and $\mathcal{P}$ are as in Example 3.4, but $C=[10^{-a}]$ for some positive real number a. Setting $D = 10^{-a}I$ and $E=I$ ($C = EDE^T$), the quadratic eigenvalue problem (2.14) is

$$\left(\lambda^2 \begin{bmatrix} 10^a & 0 \\ 0 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 2+2\times 10^a & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 1+10^a & 0 \\ 0 & 1 \end{bmatrix}\right) \begin{bmatrix} x_y \\ x_z \end{bmatrix} = 0.$$

This quadratic eigenvalue problem has three finite eigenvalues: $\lambda = \frac{1}{2}$,

$$\lambda = 1 + 10^{-a} \pm 10^{-a}\sqrt{1+10^a}.$$

For large values of a, $\lambda \approx 1 + 10^{-a} \pm 10^{-\frac{a}{2}}$; the eigenvalues will be close to 1.

The CUTEr test set [10] provides a set of quadratic programming problems. We shall use the problem CVXQP2_S in the following examples. This problem is very small with $n = 100$ and $m = 25$. "Barrier" penalty terms (in this case 1.1) are added to the diagonal of A to simulate systems that might arise during and iteration of an interior-point method for such problems. We shall set $G = \text{diag}(A)$, and $C = \alpha \times \text{diag}(0, \ldots, 0, 1, \ldots, 1)$, where α is a positive, real parameter that we will change.

All tests were performed on a dual Intel Xeon 3.20GHz machine with hyperthreading and 2GiB of RAM. It was running Fedora Core 2 (Linux kernel 2.6.8) with MATLAB® 7.0. The linear systems were solved using the Simplified Quasi-Minimal Residual Algorithm (SQMR) [7] – MATLAB® code for SQMR can be obtained from the MATLAB® Central File Exchange at `http://www.mathworks.fr/matlabcentral/`. We terminate the iteration when the value of residual is reduced by at least a factor of 10^{-8}.

In Figure 3.1 we compare the performance (in terms of iteration count) between using a preconditioner of the form $\mathcal{P}$ and one of the form $\mathcal{P}_C$, equations (1.5) and (1.4) respectively. The matrix C used in this set of results takes the form αI. Although the SQMR method doesn't have an optimality property as was assumed in Section 3, as α becomes smaller, we hope that the difference between the number of iterations required by the two preconditioners decreases. We observe that, in this example, once $\alpha \leq 10^{-3}$ there is little benefit in reproducing C in the preconditioner.

In Figure 3.2 we also compare the performance (in terms of iteration count) between using a preconditioner of the form $\mathcal{P}$ and one of the form $\mathcal{P}_C$, Equations (1.5) and (1.4) respectively. However, we have now set $C = \alpha \times \text{diag}(0, \ldots, 0, 1, \ldots, 1)$, where $\text{rank}C = \lfloor m/2 \rfloor$. We observe that the convergence is faster in the second figure - this is as we would expect because of there now being a guarantee of at least 24 unit eigenvalues in the preconditioned system compared to the possibility of none. Similar results can be found in [5] where a projected preconditioned conjugate gradient method has been used to solve the linear systems instead of SQMR.

4 Conclusion

In this paper, we have investigated a new class of preconditioner for indefinite linear systems that incorporate the (1,2) and (2,1) blocks of the original matrix. These blocks are often associated with constraints. We have shown that if C has rank $p > 0$, then the preconditioned system has at least $2(m - p)$ unit eigenvalues, regardless of the structure of G. In addition, we have shown that if the entries of C are very small, then we will expect an additional $2p$ eigenvalues to be clustered around 1 and, hence, for the number of iterations required by our chosen Krylov subspace method to be dramatically reduced. These later results are of particular relevance to interior point methods for optimization.

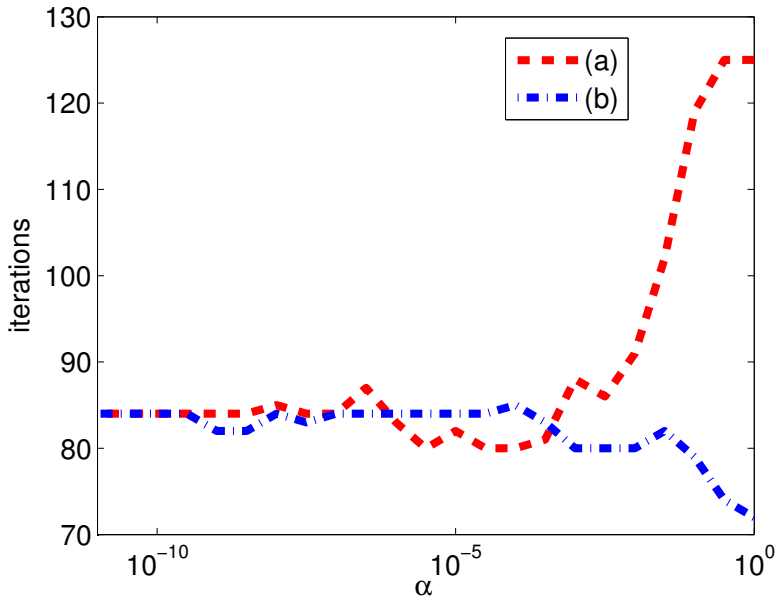

Figure 3.1: Number of SQMR iterations when either (a) $\mathcal{P}$ or (b) $\mathcal{P}_C$ are used as preconditioners for $C = \alpha I$.

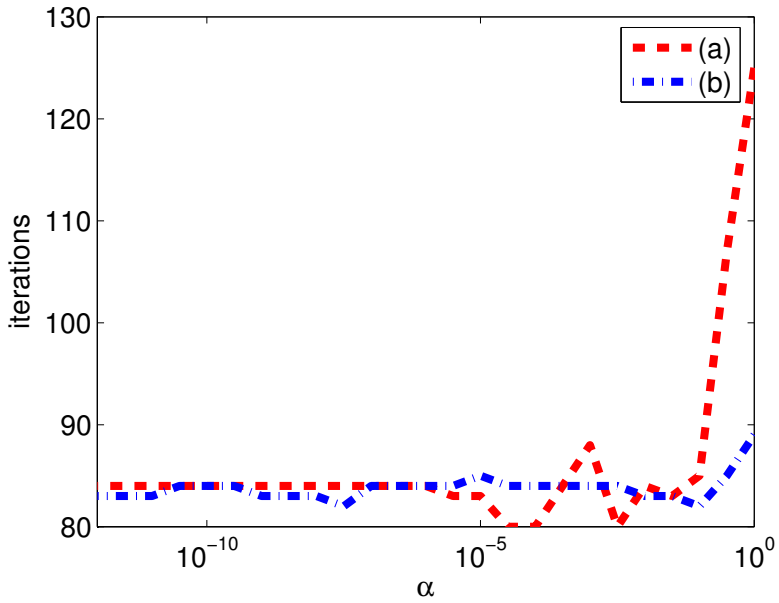

Figure 3.2: Number of SQMR iterations when either (a) $\mathcal{P}$ or (b) $\mathcal{P}_C$ are used as preconditioners for $C = \alpha \times \text{diag}(0, \ldots, 0, 1, \ldots, 1)$, where $\text{rank} C = \lfloor m/2 \rfloor$.

References

[1] O. Axelsson and V. A. Barker. *Finite element solution of boundary value problems. Theory and computation*, volume 35 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Reprint of the 1984 original.

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

[3] L. Bergamaschi, J. Gondzio, and G. Zilli. Preconditioning indefinite systems in interior point methods for optimization. *Comput. Optim. Appl.*, 28(2):149–171, 2004.

[4] H. S. Dollar. Extending constraint preconditioners for saddle point problems. Technical Report NA-05/02, Oxford University Computing Laboratory, 2005. (Submitted to SIAM J. Matrix Anal. Appl.).

[5] H. S. Dollar. *Iterative Linear Algebra for Constrained Optimization*. PhD thesis, Oxford University, 2005.

[6] H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite Elements and Fast Iterative Solvers: with applications in Incompressible Fluid Dynamics*. Oxford University Press, Oxford, 2005.

[7] R. W. Freund and N. M. Nachtigal. A new Krylov-subspace method for symmetric indefinite linear systems. In W. F. Ames, editor, *Proceedings of the 14th IMACS World Congress on Computational and Applied Mathematics*, pages 1253–1256. IMACS, 1994.

[8] N. I. M. Gould. On the accurate determination of search directions for simple differentiable penalty functions. *IMA Journal of Numerical Analysis*, 6:357–372, 1986.

[9] N. I. M. Gould, M. E. Hribar, and J. Nocedal. On the solution of equality constrained quadratic programming problems arising in optimization. *SIAM J. Sci. Comput.*, 23(4):1376–1395, 2001.

[10] N. I. M. Gould, D. Orban, and Ph. L. Toint. CUTEr (and SifDec), a constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, 2003.

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

[12] C. Keller, N. I. M. Gould, and A. J. Wathen. Constraint preconditioning for indefinite linear systems. *SIAM J. Matrix Anal. Appl.*, 21(4):1300–1317, 2000.

[13] I. Perugia and V. Simoncini. Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations. *Numer. Linear Algebra Appl.*, 7(7-8):585–616, 2000. Preconditioning techniques for large sparse matrix problems in industrial applications (Minneapolis, MN, 1999).

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

[15] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869, 1986.

[16] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Rev.*, 43(2):235–286, 2001.

[17] K.-C. Toh, K.-K. Phoon, and S.-H. Chan. Block preconditioners for symmetric indefinite linear systems. *Internat. J. Numer. Methods Engrg.*, 60(8):1361–1381, 2004.

[18] S. J. Wright. *Primal-dual interior-point methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.