Electronic Transactions on Numerical Analysis.
Volume xx, pp. 1–17, 2024.

ISSN 1068–9613.

ETNA
Kent State University **and**
Johann Radon Institute (RICAM)

AUGMENTATION-BASED PRECONDITIONERS FOR SADDLE-POINT SYSTEMS WITH SINGULAR LEADING BLOCKS*

SUSANNE BRADLEY† AND CHEN GREIF‡

Abstract. We consider the iterative solution of symmetric saddle-point matrices with a singular leading block. We develop a new ideal positive definite block-diagonal preconditioner that yields a preconditioned operator with four distinct eigenvalues. We offer a few techniques for making the preconditioner practical, and illustrate the effectiveness of our approach with numerical experiments. The novelty of the paper lies in the generality of the assumptions made: as long as the saddle-point matrix is nonsingular, there is no assumption on the specific rank of the leading block. Current ideal preconditioners typically rely either on invertibility or a high nullity of the leading block, and the new technique aims to bridge over this gap. A spectral analysis is offered, accompanied by numerical experiments.

Key words. saddle-point systems, preconditioning, augmentation, Schur complement

AMS subject classifications. 65F08, 65F10, 65F15

1. Introduction.

Consider the saddle-point system

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}, \tag{1.1}$$

where $A \in \mathbb{R}^{n \times n}$ is symmetric positive semidefinite and $B \in \mathbb{R}^{m \times n}$ has full row rank, with $m < n$. We denote the coefficient matrix by

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

We assume throughout that $\mathcal{K}$ is invertible. A necessary and sufficient condition for this is that $\ker(A) \cap \ker(B) = \{0\}$; see [1, Theorem 3.2]. Thus, the nullity of A must be no greater than m, or $\mathcal{K}$ will necessarily be singular. We therefore say that a leading block A with nullity m is *lowest-rank* or *maximally rank-deficient*. Under the assumptions above, the matrix $\mathcal{K}$ is symmetric and indefinite, and the solution of the linear system (1.1) poses several numerical challenges; we refer to the survey of [1] for an overview of solution methods.

Our focus is on positive definite preconditioners, which maintain symmetry of the preconditioned operator and can therefore be used with a symmetric iterative solver such as MINRES [16]. When A is positive definite, the preconditioner of Murphy, Golub, and Wathen [14]

$$\mathcal{M}_1 = \begin{bmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{bmatrix}$$

has the property that the preconditioned operator $\mathcal{M}_1^{-1}\mathcal{K}$ has three distinct eigenvalues, meaning that a preconditioned iterative solver (such as MINRES) will converge within three iterations in exact arithmetic. In practice, the matrices A and $BA^{-1}B^T$ are too expensive to form and solve for exactly, so approximations must be sought.

The case in which A is singular has been less studied; see [6, 10, 11] for preconditioning approaches in this setting. Golub, Greif, and Varah [10] have analyzed the positive definite block-diagonal preconditioner

$$\mathcal{M}_2 = \begin{bmatrix} A + B^TWB & 0 \\ 0 & B(A + B^TWB)^{-1}B^T \end{bmatrix},$$

*Work supported by Natural Sciences and Engineering Research Council of Canada

†Department of Computer Science, University of British Columbia, V6T 1Z4. Email: smbrad@cs.ubc.ca
‡Department of Computer Science, University of British Columbia, V6T 1Z4. Email: greif@cs.ubc.ca

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where $W \in \mathbb{R}^{m\times m}$ is a positive semidefinite matrix such that $A + B^TWB$ is positive definite. This can be considered a generalization of $\mathcal{M}_1$, in which a semidefinite term is first added to the leading block to make it positive definite. Because of the requirement that $\ker(A) \cap \ker(B) = \{0\}$, the matrix $A + B^TWB$ is necessarily positive definite if W is positive definite (though this is not a necessary condition unless A is lowest-rank).

While the preconditioned operator $\mathcal{M}_2^{-1}\mathcal{K}$ is not guaranteed to have a fixed, small number of distinct eigenvalues, it is shown in [10, Theorem 2.5] that the eigenvalues are bounded within the intervals $\left[-1, \frac{1-\sqrt{5}}{2}\right] \cup \left[1, \frac{1+\sqrt{5}}{2}\right]$. However, from [6, Theorem 3.5] and [11, Theorem 4.1], we can observe that $\mathcal{M}_2^{-1}\mathcal{K}$ does have exactly two distinct eigenvalues when A has maximal nullity.

Contribution of this paper. At present, to the best of our knowledge the literature provides ideal positive definite block-diagonal preconditioners that yield preconditioned operators with a small number of distinct eigenvalues (and, therefore, will lead to convergence of a preconditioned iterative solver in a small number of iterations in the absence of round-off error) in the cases where A has full rank and where A has maximal nullity. In this work, we bridge the gap between the full-rank and minimal-rank (or maximal-nullity) cases by providing such a preconditioner for cases in which $(n - m) < \operatorname{rank}(A) < n$. This is potentially meaningful because on the one hand we cannot invert A and given its assumed rank deficiency, the Schur complement $BA^{-1}B^T$ does not exist either, making it difficult to develop standard preconditioners. And on the other hand unique algebraic properties that have been studied in [6, 10, 11] for the maximal-nullity case cannot be applied either.

Outline. We provide relevant mathematical background in Section 2 and describe our preconditioning approach in Section 3. We then provide numerical experiments in Section 4 and concluding remarks in Section 5.

2. Mathematical background.

In this section, we provide some existing results that will aid us in developing and analyzing our preconditioner. Section 2.1 describes previous strategies in the literature for augmenting a rank-deficient leading block A, and Section 2.2 describes some special properties of matrices with maximally rank-deficient leading blocks. We then use these techniques to provide an alternative proof of a result in [11] for matrices with a maximally rank-deficient A, and we use the insights of this alternative proof to adapt this approach to matrices with non-maximally rank-deficient A in Section 3.

2.1. Leading block augmentation.

Our strategy for preconditioning involves augmenting the leading block A so that it becomes positive definite, rather than semidefinite. We observe that (1.1) can be reformulated as (see, for example, [8, 9]):

$$\begin{bmatrix} A + B^TWB & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f + B^TWg \\ g \end{bmatrix},$$

where W is an $m \times m$ matrix. We will assume W is positive semidefinite and the leading block

$$A_W = A + B^TWB \tag{2.1}$$

is positive definite. An advantage of this approach is that a positive definite leading block will provide flexibility in both forming and analyzing our preconditioners later in this paper. This approach proved effective in [2] for fluid flow problems. We also recall the following result [8, 9]:

LEMMA 2.1. *Let*

$$\mathcal{K}(W) = \begin{bmatrix} A_W & B^T \\ B & 0 \end{bmatrix},$$

where $W \in \mathbb{R}^{m\times m}$*. If* $\mathcal{K}$ *and* $\mathcal{K}(W)$ *are both nonsingular, then*

$$\mathcal{K}^{-1} = (\mathcal{K}(W))^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & W \end{bmatrix}.$$

2.2. Matrix properties when $\operatorname{nullity}(A) = m$**.** When A has maximal nullity – that is, when $\operatorname{nullity}(A) = m$ – the blocks of $\mathcal{K}$ and those of the augmented matrix $\mathcal{K}(W)$ interact in unique ways, which provide useful tools in the design and analysis of preconditioners.

Estrin and Greif [6, Theorem 3.5] provide the following result on the Schur complement of $\mathcal{K}(W)$:

PROPOSITION 2.2. *Suppose* $\operatorname{nullity}(A) = m$ *and let* $W \in \mathbb{R}^{m\times m}$ *be an invertible matrix. Then*

$$B(A + B^T W B)^{-1} B^T = W^{-1}.$$

We also recall the following result [7, Corollary 2.1] applying to more general matrices, which we will use repeatedly in our analyses:

LEMMA 2.3. *For matrices* $M, N \in \mathbb{R}^{n\times n}$ *with* $\operatorname{rank}(M) = r, \operatorname{rank}(N) = n - r$ *and* $M + N$ *nonsingular, the matrix* $(M + N)^{-1}M$ *is a projector with rank* r*. Moreover,*

$$M(M + N)^{-1}N = 0.$$

A recent article by the authors [3] provides eigenvalue bounds for saddle-point systems with a rank-deficient leading block. We will use the following result [3, Theorem 7] in our analyses:

THEOREM 2.4. *When* $\operatorname{rank}(A) = n - m$*, the positive eigenvalues of* $\mathcal{K}$ *are greater than or equal to*

$$\min\left\{\mu^+_{\min}(1 - \cos(\theta_{\min})), \sigma_{\min}\sqrt{1 - \cos(\theta_{\min})}\right\}$$

where: $\mu^+_{\min}$ *denotes the smallest positive eigenvalue of* A*;* $\sigma_{\min}$ *the smallest singular value of* B*; and* $\theta_{\min}$ *the minimum principal angle between* $\operatorname{range}(A)$ *and* $\operatorname{range}(B^T)$*.*

2.3. Preconditioning when $\operatorname{nullity}(A) = m$**.** We consider the block-diagonal preconditioner [11]

$$\mathcal{M}_W = \begin{bmatrix} A_W & 0 \\ 0 & W^{-1} \end{bmatrix}, \tag{2.2}$$

where W is positive definite and A_W is as defined in (2.1). Let us denote the blocks of the split preconditioned operator $\mathcal{M}_W^{-1/2}\mathcal{K}\mathcal{M}_W^{-1/2}$ as follows:

$$\mathcal{M}_W^{-1/2}\mathcal{K}\mathcal{M}_W^{-1/2} = \begin{bmatrix} A_W^{-1/2} A A_W^{-1/2} & A_W^{-1/2} B^T W^{1/2} \\ W^{1/2} B A_W^{1/2} & 0 \end{bmatrix} =: \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & 0 \end{bmatrix}.$$

LEMMA 2.5. *When* $\operatorname{rank}(A) = n - m$*, the blocks of* $\mathcal{M}_W^{-1/2}\mathcal{K}\mathcal{M}_W^{-1/2}$ *satisfy the following:*

(i) All nonzero eigenvalues of $\tilde{A}$ *are equal to 1;*
(ii) All singular values of $\tilde{B}$ *are equal to 1;*
(iii) The subspaces $\operatorname{range}(\tilde{A})$ *and* $\operatorname{range}(\tilde{B}^T)$ *are orthogonal.*

Proof. To prove (i), we note that $\tilde{A}$ is similar to $A_W^{-1}A$, which is a projector by Lemma 2.3. Lemma 2.2 gives us that $BA_W^{-1}B^T = W^{-1}$, and therefore

$$\tilde{B}\tilde{B}^T = W^{1/2}BA_W^{-1}B^TW^{1/2} = I,$$

which proves (ii). We prove (iii) by showing that $\mathrm{range}(\tilde{B}^T) \subseteq \ker(\tilde{A})$. We write

$$\tilde{A}\tilde{B}^T = A_W^{-1/2}AA_W^{-1}B^TW^{-1/2} = 0,$$

where the second equality follows from the result of [6, Proposition 2.6], which shows that $A_W^{-1}B^T$ is a null-space matrix of A. ◻

We now consider what the results of Lemma 2.5 tell us about the eigenvalues of $\mathcal{M}_W^{-1}\mathcal{K}$ when $\mathrm{rank}(A) = n - m$. The orthogonality of $\mathrm{range}(\tilde{A})$ and $\mathrm{range}(\tilde{B}^T)$ means that the value of $\cos(\theta_{\mathrm{min}})$ in Theorem 2.4 is 1, and thus that the positive eigenvalues are greater than or equal to the minimum of the smallest positive eigenvalue of $\tilde{A}$ and the smallest singular value of $\tilde{B}$. These are both equal to 1, by parts (i)-(ii) of Lemma 2.5. Because the maximal eigenvalues of $\tilde{A}$ and singular values of $\tilde{B}$ are also equal to 1, all negative eigenvalues are equal to -1 and all positive eigenvalues are less than or equal to 1 (as a consequence of [17, Lemma 2.1]). This yields the following result, which is also shown via a different proof method in [11, Theorem 4.1]; we refer to their proof for derivation of the multiplicities of the eigenvalues.

PROPOSITION 2.6. *When* $\mathrm{rank}(A) = n - m$*, the matrix* $\mathcal{M}_W^{-1}\mathcal{K}$ *has two distinct eigenvalues given by* 1 *and* -1 *with algebraic multiplicities* n *and* m*, respectively.*

Proposition 2.6 tells us that when A has maximal nullity there is a block-diagonal preconditioner that yields a preconditioned operator with two distinct eigenvalues. This is similar to the block-diagonal preconditioner of [14], which yields a preconditioner with three distinct eigenvalues in the case that A is positive definite. What has not yet been developed is a preconditioner that gives a small fixed number of distinct eigenvalues for the "in-between" case where A is rank-deficient, but not lowest-rank. This is the focus of the next section.

3. Block diagonal preconditioning for non-maximal nullity.

3.1. Preconditioner derivation. Let us now consider the case in which A has nullity k, with $k < m$. We will now consider how we can devise a preconditioner to preserve (perhaps approximately) the properties listed in Lemma 2.5 in the case where we no longer have maximal nullity.

Let us consider a general block-diagonal preconditioner of the form

$$\mathcal{M} = \begin{bmatrix} A+G & 0 \\ 0 & C \end{bmatrix},$$

where C is positive definite and G is a semidefinite matrix such that $A + G$ is positive definite. As before, let us define the split preconditioned system:

$$\mathcal{M}^{-1/2}\mathcal{K}\mathcal{M}^{-1/2} = \begin{bmatrix} (A+G)^{-1/2}A(A+G)^{-1/2} & (A+G)^{-1/2}B^TC^{-1/2} \\ C^{-1/2}B(A+G)^{-1/2} & 0 \end{bmatrix}$$
$$=: \begin{bmatrix} \tilde{A} & \tilde{B}^T \\ \tilde{B} & 0 \end{bmatrix}.$$

Property (i) of Lemma 2.5 holds whenever $\mathrm{rank}(G) = k$; see Lemma 2.3. It is also straightforward to verify, using a similar process as in the proof of Lemma 2.5, that Property (ii) holds if and only if

$$C = B(A+G)^{-1}B^T.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Property (iii) of Lemma 2.5 holds because, in that Lemma's setting,

$$A(A+G)^{-1}B^T = 0.$$

We can write this as

$$\begin{aligned} A(A+G)^{-1}B^T &= (A+G-G)(A+G)^{-1}B^T \\ &= B - G(A+G)^{-1}B^T. \end{aligned} \tag{3.1}$$

Suppose that G has rank k, as we have already established will ensure Property (i). Then, as a consequence of Lemma 2.3, $G(A+G)^{-1}$ is a projector onto the range of G. From (3.1) we see that Property (iii) will hold if $G(A+G)^{-1}$ is a projector onto the range of B^T; however, this is clearly not possible if $\mathrm{rank}(G) = k < m$. But we note that if we set

$$G = B^T W_k B,$$

where W_k is a symmetric positive semidefinite matrix of rank k, this matrix will be a projector onto a rank-k subspace of $\mathrm{range}(B^T)$. While Property (iii) will not hold in this case because we will not have $\tilde{A}\tilde{B}^T = 0$, we instead have that $\mathrm{nullity}(\tilde{A}\tilde{B}^T) = k$ (which is the highest nullity we can achieve, as from (3.1) we have a rank-k term being subtracted from B).

Thus, we consider the preconditioner:

$$\mathcal{M}_k = \begin{bmatrix} A_k & 0 \\ 0 & S_k \end{bmatrix}, \tag{3.2}$$

where $A_k = A + B^T W_k B$ and $S_k = BA_k^{-1}B^T$, with $\mathrm{rank}(W_k) = \mathrm{nullity}(A) = k$. This is the same preconditioner analyzed in [10], but with the additional assumption that $\mathrm{rank}(W_k) = k$.

Remark 1. We note that, when A has maximal nullity, the preconditioner $\mathcal{M}_k$ reduces to that of Greif and Schötzau defined in eq. (2.2). When A is positive definite, then $\mathcal{M}_k$ is equivalent to the preconditioner $\mathcal{M}_1$.

3.2. Analysis of $\mathcal{M}_k$.

We present some lemmas that will be necessary for our analysis.

LEMMA 3.1. *When* $\mathrm{rank}(W_k) = \mathrm{nullity}(A) = k$,

$$(BA_k^{-1}B^T)^{-1} = W_k + (BB^T)^{-1}B(A - AVA)B^T(BB^T)^{-1},$$

where $V = Z(Z^TAZ)^{-1}Z^T$ *with* $Z \in \mathbb{R}^{n\times(n-m)}$ *being a null-space matrix of* B.

Proof. The proof follows by considering the block inverses of $\mathcal{K}$ and

$$\mathcal{K}(W_k) := \begin{bmatrix} A_k & B^T \\ B & 0 \end{bmatrix}.$$

Let $Z \in \mathbb{R}^{n\times(n-m)}$ denote a matrix whose columns form a basis for $\ker(B)$. The inverse of $\mathcal{K}$ is (see [1, Eq. (3.8)]):

$$\mathcal{K}^{-1} = \begin{bmatrix} V & (I - VA)B^T(BB^T)^{-1} \\ (BB^T)^{-1}B(I - AV) & -(BB^T)^{-1}B(A - AVA)B^T(BB^T)^{-1} \end{bmatrix},$$

where $V = Z(Z^TAZ)^{-1}Z^T$; we note that Z^TAZ must be nonsingular for any nonsingular $\mathcal{K}$ (see [1]). The result then follows from Lemma 2.1 and the fact that the (2,2)-block of $(\mathcal{K}(W_k))^{-1}$ is equal to $-(BA_k^{-1}B^T)^{-1}$ (see [1, Eq. (3.4)]). ∎

LEMMA 3.2. *The matrix* VA *is a projector. Moreover, when* $\mathrm{rank}(W_k) = \mathrm{nullity}(A) = k$*, the following results hold:*

(i) The matrix $A_k^{-1}A$ is a projector;
(ii) The matrices VA and $A_k^{-1}A$ commute.

Proof. By writing $VA = Z(Z^TAZ)^{-1}Z^TA$, it is clear that VA is a projector onto $\ker(B)$. Item (i) holds because of Lemma 2.3.

To verify (ii), we first note that

$$VAA_k^{-1}A = VA,$$

because AA_k^{-1} is a projector (this follows from the fact that $A_k^{-1}A = (AA_k^{-1})^T$ is a projector) onto the range of A. Because $A_k^{-1}A = I - A_k^{-1}B^TW_kB$, we can write

$$A_k^{-1}AZ = Z - A_k^{-1}B^TWBZ = Z.$$

Therefore,

$$\begin{aligned} A_k^{-1}AVA &= A_k^{-1}AZ(Z^TAZ)^{-1}Z^TA \\ &= Z(Z^TAZ)^{-1}Z^TA \\ &= VA \\ &= VAA_k^{-1}A. \end{aligned}$$

□

THEOREM 3.3. *Let $\mathcal{K}$ be nonsingular with A having nullity k, and let $W_k \in \mathbb{R}^{m\times m}$ be a rank-k matrix such that $A + B^TW_kB$ is positive definite. The preconditioned operator $\mathcal{M}_k^{-1}\mathcal{K}$ has four distinct eigenvalues:*

- $\lambda = -1$ *with multiplicity k;*
- $\lambda = 1$ *with multiplicity $n - m + k$;*
- $\lambda = \frac{1\pm\sqrt{5}}{2}$, *each with multiplicity $m - k$.*

Proof. We consider the eigenvalue equations for the preconditioned system:

$$Ax + B^Ty = \lambda A_kx; \tag{3.3a}$$

$$Bx = \lambda S_ky. \tag{3.3b}$$

From (3.3b) we obtain $y = \frac{1}{\lambda}S_k^{-1}Bx$. Substituting this into (3.3a) and re-arranging yields

$$A_k^{-1}Ax + \frac{1}{\lambda}A_k^{-1}B^TS_k^{-1}Bx - \lambda x = 0. \tag{3.4}$$

By Lemma 3.1, we can write

$$\begin{aligned} A_k^{-1}B^TS_k^{-1}B = A_k^{-1}&B^TW_kB \\ &+ A_k^{-1}B^T(BB^T)^{-1}B(A - AVA)B^T(BB^T)^{-1}B. \end{aligned} \tag{3.5}$$

As was discussed in the proof of Lemma 3.2, VA is a projector onto $\ker(B)$, meaning that $I - VA$ is a projector onto $\mathrm{range}(B)$. Because $B^T(BB^T)^{-1}B$ is an orthogonal projector onto this subspace, we have

$$(I - VA)B^T(BB^T)^{-1}B = I - VA.$$

Similarly, $B^T(BB^T)^{-1}B(I - AV) = I - AV$. Thus, we can further simplify (3.5), using relations we developed in Lemma 3.2:

$$\begin{aligned} A_k^{-1}B^TS_k^{-1}B &= A_k^{-1}B^TW_kB + A_k^{-1}(A - AVA) \\ &= I - A_k^{-1}AVA \\ &= I - VA. \end{aligned}$$

We can thus rewrite (3.4) as

$$A_k^{-1}Ax - \frac{1}{\lambda}VAx + \left(\frac{1}{\lambda} - \lambda\right)x = 0. \tag{3.6}$$

By Lemma 3.2, $A_k^{-1}A$ and VA are commuting projectors; thus, they have the same eigenvectors. Because VA has rank $n-m$ and $A_k^{-1}A$ has rank $n-k$, we have

$$\mathrm{range}(VA) \subseteq \mathrm{range}(A_k^{-1}A) \text{ and } \ker(A_k^{-1}A) \subseteq \ker(VA).$$

We now consider x in the ranges/kernels of these projectors.

Case I: When $x \in \ker(A)$, (3.6) becomes

$$\left(\frac{1}{\lambda} - \lambda\right)x = 0. \tag{3.7}$$

We note that x cannot be zero, as (3.3a) would necessarily imply $y = 0$. Thus, (3.7) gives k eigenvectors corresponding to each of the eigenvalues $\lambda = \pm 1$.

Case II: When $x \in \mathrm{range}(VA)$ (and therefore also in $\mathrm{range}(A_k^{-1}A)$), (3.6) becomes

$$(1-\lambda)\,x = 0,$$

which gives $n-m$ additional eigenvectors corresponding to the eigenvalue $\lambda = 1$.

Case III: if $x \in \ker(VA)$ and $\mathrm{range}(A_k^{-1}A)$ (we know there are $m-k$ such vectors because the projectors commute), (3.6) becomes

$$\left(1 + \frac{1}{\lambda} - \lambda\right)x = 0,$$

which gives the eigenvalues $\lambda = \frac{1\pm\sqrt{5}}{2}$, each with geometric multiplicity $m-k$.

Cases I-III account for all $n+m$ eigenvectors of $\mathcal{M}_k^{-1}\mathcal{K}$. □

3.3. Schur complement approximations. In practice, the blocks A_k and S_k of the ideal preconditioner $\mathcal{M}_k$ defined in (3.2) are too expensive to invert exactly. While developing suitable approximation strategies for these terms often requires some knowledge of the problem at hand, we provide here two strategies for approximately inverting the Schur complement S_k.

First, recall from Lemma 2.2 that when A has maximal nullity we have $S_k^{-1} = W_k$. Thus, when A has high but not maximal nullity, it is reasonable to use an approximation of the form

$$S_k^{-1} \approx W_k + \beta I, \tag{3.8}$$

where β is a small positive value. We add the βI term because if A is not maximally rank-deficient then W_k will be singular. We refer to this strategy as the "WkI Schur complement approximation."

For our second strategy, recall that Lemma 3.1 tells us that

$$\begin{aligned} S_k^{-1} &= W_k + (BB^T)^{-1}B(A - AVA)B^T(BB^T)^{-1} \\ &= W_k + (BB^T)^{-1}BA\underbrace{(I - VA)}_{=:P}B^T(BB^T)^{-1}. \end{aligned}$$

Since VA is a projector whose range is $\ker(B)$ and whose kernel is $\ker(Z^TA)$, the matrix $P = (I - VA)$ has range given by $\ker(Z^TA)$ and kernel given by $\ker(B)$. Thus, we consider replacing the projector $(I - VA)$ by the orthogonal projector onto $\mathrm{range}(B)$, defined by $P_B = B^T(BB^T)^{-1}B$. This matrix has the same kernel as P but a different range, and has the advantage of yielding a considerably simpler second term, as we can write:

$$\begin{aligned}(BB^T)^{-1}BAP_BB^T(BB^T)^{-1} &= (BB^T)^{-1}BAB^T(BB^T)^{-1}BB^T(BB^T)^{-1}\\ &= (BB^T)^{-1}BAB^T(BB^T)^{-1}.\end{aligned}$$

Thus, we can also consider the Schur complement approximation:

$$S_k^{-1} \approx W_k + (BB^T)^{-1}BAB^T(BB^T)^{-1}. \tag{3.9}$$

We note that this modified second term is similar to the BFBt preconditioner proposed by Elman [5] for the Navier-Stokes equations; thus, we refer to this as the "BFBt Schur complement approximation."

4. Numerical experiments.

In this section we consider implementations of the block-diagonal preconditioner described in Section 3. All experiments are run in MATLAB R2021a on a commodity desktop PC. We report computation times for all experiments. The code is not optimized for efficiency and the measurements do not represent what would be possible with an optimized, state-of-the-art code base; they are included as a way to compare the computational costs of different approaches and validate our analytical observations.

4.1. Selection of weight matrix.

Here we detail our general approach for choosing W_k. For simplicity, all our matrices W_k are diagonal matrices with either 1 or 0 on the diagonal; thus, the augmented matrix A_k is equal to A in addition to k terms of the form b^Tb, where b is a single row of B. Hence, our task of selecting W_k becomes the task of selecting which rows of B to use in to augment A.

We begin by forming a matrix A_{drop} formed by eliminating very small elements of A (for our purposes, we eliminate those matrix entries whose absolute values are less than machine epsilon times the largest magnitude entry in A). We then select rows of B that increase the structural rank of A_{drop} until the matrix $A_{drop} + \sum_i b_i^T b_i$ has full structural rank. These selected rows of b do not guarantee that the augmented matrix $A + \sum_i b_i^T b_i$ has full numerical rank or is sufficiently well-conditioned to avoid convergence problems, so in some cases we add additional rows of B; in this case, we greedily select the sparsest rows of B to reduce fill-in of A_k.

We note that, in general, this approach of selecting W_k does *not* guarantee a "minimal-rank" augmentation; that is, the rank of W_k may be greater than the nullity of A. Finding a W_k with rank exactly equal to the nullity of A such that the augmented matrix A_k is sufficiently well-conditioned to avoid numerical difficulty requires knowledge of the null-space of A and of which vectors in B will span that null space. That said, in many practical applications, for example in problems arising from discretizations of PDEs, some information on the discretized differential operators and their null space is often available and comes handy.

4.2. Constrained optimization problems.

Problem statement. Given a positive semidefinite Hessian matrix $H \in \mathbb{R}^{n\times n}$, vectors $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ and a Jacobian matrix $J \in \mathbb{R}^{m\times n}$, consider the primal-dual pair of quadratic programs (QP) in standard form:

$$\min_x c^Tx + \frac{1}{2}x^THx \quad \text{s.t. } Jx = b, \quad x \geq 0; \tag{4.1a}$$

$$\min_{x,y,z} b^Ty - \frac{1}{2}x^THx \quad \text{s.t. } J^Ty + z - Hx = c, \quad z \geq 0, \tag{4.1b}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where y and z are vectors of Lagrange multipliers. In linear programming problems, we have $H = 0$.

Each step of a primal-dual interior-point method (IPM) to solve (4.1) requires solving a linear system of the form [15]:

$$\begin{bmatrix} H + X^{-1}Z & J^T \\ J & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -c - HxJ^T y + \tau X^{-1} e \\ b - Jx \end{bmatrix}.$$

Here X and Z are diagonal matrices whose diagonal entries are the components of x and z, respectively, and $\tau > 0$ is the barrier parameter, which governs the progress of the interior-point iterations; see [15] for full details. Some entries of the diagonal matrices X and Z approach zero as the IPM iterations proceed, so the leading block of the saddle-point matrix becomes increasingly ill-conditioned, with the largest magnitude entries occurring along the diagonal. Thus the leading block may become nearly singular or numerically singular, particularly if H is singular.

Description of test problems. We use an implementation of the predictor-corrector algorithm of Mehrotra [13]. The matrices for linear programming problems were obtained from the Sparse Suite matrix collection [4], and the quadratic programming problems are from TOMLAB*. A summary of the test suite of LP problems used in our experiments is given in Table 4.1.

Problem ID	m	n	nnz($\mathcal{K}$)
lp_80bau3b	2,262	12,061	35,325
lp_bandm	305	472	2,966
lp_capri	271	482	2,378
lp_finnis	497	1,064	3,824
lp_fit1p	627	1,677	11,545
lp_ganges	1,309	1,706	8,643
lp_lofti	153	366	1,502
lp_maros_r7	3,136	9,408	154,256
lp_osa_14	2,337	5,497	371,894
lp_osa_30	4,350	104,374	708,862
lp_pilot87	2,030	6,680	81,629
lp_scfxm1	330	600	3,332
lp_scsd8	397	2,750	11,334
lp_stair	356	614	4,617
lp_standmps	467	1,274	5,152
lp_stocfor2	2,157	3,045	12,402
lp_truss	1,000	8,806	36,642
lp_vtp_base	198	346	1,397

Table 4.1: Summary of linear programming (LP) problems used in numerical experiments. The value $nnz(\mathcal{K})$ gives the number of nonzeros arising in the saddle-point system in each interior-point method iteration.

*Test matrices available at https://tomopt.com/tomlab/.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Comparison of different augmentation and approximation strategies. In this experiment we consider preconditioners of the form

$$\mathcal{M} = \begin{bmatrix} \tilde{A}_{aug} & 0 \\ 0 & B\hat{A}_{aug}^{-1}B^T \end{bmatrix}, \tag{4.2}$$

where $\tilde{A}_{aug}$ and $\hat{A}_{aug}$ are approximations (potentially the same approximation) of an augmented leading block A. Our experiments are on matrices that arise while applying an interior-point method on an LPs, so the leading block A is diagonal. We consider three augmentation strategies:

1. Partial augmentation: we take $A_{aug} = A + B^T W_k B$, where we form W_k by selecting just enough rows of B such that $A_{drop} + B^T W_k B$ has full structural rank, where A_{drop} is the matrix obtained by setting to zero all elements of A with absolute value less than or equal to machine-epsilon times the largest absolute magnitude value of A.
2. Full augmentation: we take $A_{aug} = A + B^T B$.
3. Identity augmentation: we take $A_{aug} = A + \rho I$, for some positive ρ.

For A_{aug} arising from partial and full augmentation, we consider three approximations for $\tilde{A}_{aug}$ and $\hat{A}_{aug}$ in (4.2):

1. Ideal approximation (ID): $\tilde{A}_{aug} = \hat{A}_{aug} = A_{aug}$. (This is too expensive to use in practice but we include it here for comparison purposes.)
2. Diagonal approximation (D): $\tilde{A}_{aug} = \hat{A}_{aug} = \mathrm{diag}(A_{aug})$.
3. Incomplete Cholesky approximation (IC) : $\tilde{A}_{aug} = \mathrm{IC}(A_{aug})$ and $\hat{A}_{aug} = \mathrm{diag}(A_{aug})$. We use ICT with drop tolerance of 0.01.

For the identity-based augmentation, the matrix A_{aug} is diagonal, so we invert it exactly (that is, $\tilde{A}_{aug} = \hat{A}_{aug} = A_{aug}$).

Problem ID	Partial			Full			Identity
	ID	D	IC	ID	D	IC	ID
`80bau3b`	5 (0.03)	22 (0.03)	230 (0.02)	18 (2.0)	122 (0.02)	254 (0.01)	43 (0.02)
`maros_r7`	22 (3.7)	22 (0.2)	56 (0.1)	2 (2.2)	19 (0.1)	26 (0.1)	11 (0.1)

Table 4.2: MINRES iteration counts for partial, full and identity-augmentation preconditioners for the `lp_80bau3b` and `lp_maros_r7` problems, using various block approximation strategies (ID=ideal, D=diagonal, IC=incomplete Cholesky). Time per iteration (in seconds) is given in parentheses.

Problem ID	Partial augmentation			Full augmentation		
	Rank(W)	nnz(A_W)	nnz(IC(A_W))	Rank(W)	nnz(A_W)	nnz(IC(A_W))
`80bau3b`	2	12,249	12,101	2,262	456,943	14,183
`maros_r7`	2,511	1,101,752	31,343	3,136	1,230,928	10,761

Table 4.3: Comparison of memory usage for partial and full augmentation for the `lp_80bau3b` and `lp_maros_r7` problem.

We use matrices that arise from IPMs on the test problems `lp_80bau3b` and `lp_maros_r7`. Iteration counts and time per iteration are given in Tables 4.2 and 4.3.

We observe that for `lp_80bau_3b`, the partial augmentation preconditioner outperforms the full augmentation preconditioner in terms of both iteration count and memory usage. This is because the leading block of this matrix is only mildly rank-deficient, so we only need a low-rank augmentation to make it nonsingular (which leads to a much sparser augmented matrix than the full augmentation); additionally, when we fully augment this matrix we are far away from the "ideal" amount of augmentation (i.e., the rank of augmentation that would yield a small fixed number of distinct eigenvalues in an ideally-preconditioned iterative solver) because the leading block is nowhere near lowest-rank.

In contrast, the leading block for `lp_maros_r7` is highly rank-deficient, as even the minimal amount of augmentation to obtain a structurally nonsingular leading block requires using most of the rows of B (2,511, when m for this problem is 3,136). And we observe that, in cases like these where the nullity of the leading block is high, we are close enough to the lowest-rank case that full augmentation performs well. In this case, it actually performs better than the partial augmentation in terms of iteration counts and computation time because the fully augmented leading block is more well-conditioned than the partially augmented leading block. Recall that our procedure for choosing W_k only looks at structural rank, and does not guarantee that the augmented matrix is actually nonsingular (so we may still encounter numerical difficulties without further augmentation).

Finally, we note that the incomplete Cholesky approximation strategy is less effective than the diagonal approximation strategy. One reason for this is that by the time IPM matrices are singular, the largest magnitude entries tend to occur along the diagonal; thus, a diagonal leading block approximation is generally effective (as we will see in the next set of experiments). The other reason is that, as previously mentioned, when we used the incomplete Cholesky in the leading block we avoided using the inverse of the incomplete Cholesky factors in the Schur complement approximation to avoid introducing too much computational expense. Thus, the Schur complement approximation is not equal to $B\tilde{A}_{aug}^{-1}B^T$ (where $\tilde{A}_{aug}$ is the selected leading block approximation); and as we saw in Section 3, this has an impact on the theoretical properties of the preconditioned operator.

Running partial augmentation preconditioners on LP test suites. Here we consider preconditioning the complete set of problems described in Table 4.4. The matrices reported below are the first matrices for which the IPM generates a matrix with a numerically singular leading block. We consider the partial augmentation preconditioner of the form (4.2) with the diagonal leading block approximation strategy: that is, we define P_D using $\tilde{A}_{aug} = \hat{A}_{aug} = \mathrm{diag}(A_{aug})$. In all cases, we select W_k by augmenting A until the matrix $A_{drop} + B^T W_k B$ is structurally nonsingular. MINRES solver tolerance is set to a relative residual norm of 10^{-8}.

Eigenvalues of the preconditioned operator $P_D^{-1}\mathcal{K}$ are shown in Figure 4.1 for `lp_fit1p` problem. There is strong clustering of eigenvalues near $1, \frac{1\pm\sqrt{5}}{2}$.

Using preconditioned MINRES iterations in an IPM. Here we consider using preconditioned inner solves in an IPM solver. For our test problems, we use the LP `lp_stocfor2` and the TOMLAB QP problem 37 (which has $m = 490$; $n = 1275$; 3,288 nonzeros in the Jacobian matrix; and 290 in the Hessian). Our preconditioning approach at each iteration is as follows:

- If the leading block A is nonsingular, we use the preconditioner

$$\mathcal{M}_{LP} = \begin{bmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{bmatrix}$$

Problem ID	rank(W_k)	nnz(A_k)	P_D	
			Iters	Time per iter
80bau3b	1	12,117	20	0.02
bandm	5	1,444	40	0.003
capri	13	2,230	67	0.003
finnis	29	11,184	77	0.006
fit1p	5	2,545	28	0.06
ganges	88	2,690	41	0.01
lofti	13	966	194	0.001
maros_r7	64	73,102	26	0.2
osa_14	34	98,459,317	171	0.06
osa_30	4	354,880,632	80	0.1
pilot87	5	133,798	37	0.2
scfxm1	1	840	32	0.003
scsd8	36	16,826	6	0.003
stair	33	9,994	11	0.006
standmps	2	557,906	65	0.004
stocfor2	61	3,411	9	0.1
truss	15	18,468	34	0.005
vtp_base	10	3,126	125	0.002

Table 4.4: MINRES iteration counts and time per iteration (in seconds) of the partial augmentation preconditioners with diagonal approximations of A_k.

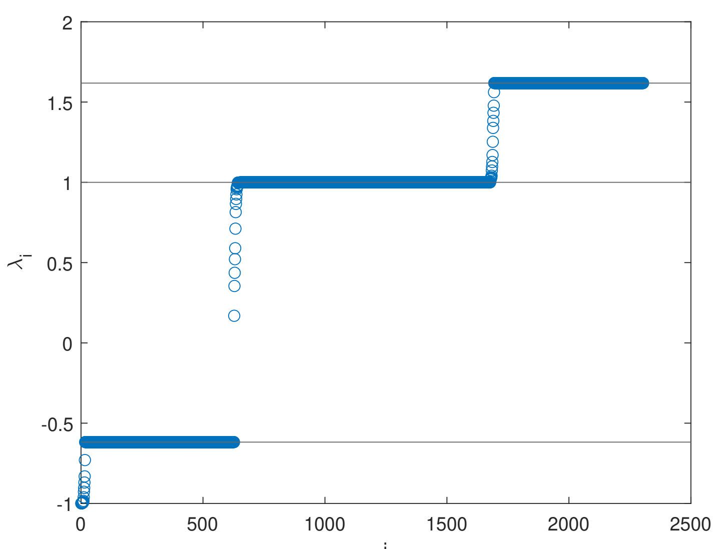

Fig. 4.1: Eigenvalues of preconditioned operator $P_D^{-1}\mathcal{K}$ for matrix arising in the IPM solution of the lp_fit1p problem. Horizontal lines are shown at $y = \pm 1, \frac{1\pm\sqrt{5}}{2}$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

for the LP (recall that in this context A is diagonal), and

$$\mathcal{M}_{QP} = \begin{bmatrix} \mathrm{IC}(A) & 0 \\ 0 & B(\mathrm{diag}(A))^{-1}B^T \end{bmatrix}$$

for the QP, with an ICT drop tolerance of 0.01.

- If the leading block A is singular, we select the lowest-rank W_k to make $A_{drop} + B^T W_k B$ nonsingular, and use the preconditioner

$$\mathcal{M} = \begin{bmatrix} \mathrm{diag}(A_k) & 0 \\ 0 & B(\mathrm{diag}(A_k))^{-1}B^T \end{bmatrix}.$$

We solve the IPM to a duality gap tolerance of 10^{-6} and use an inner tolerance of 10^{-7} for the MINRES solves.

We see that for both problems, using inexact solves results in modestly more IPM iterations, as we would expect. For the LP, the leading block was nonsingular for the first 21 iterations and numerically singular for the final 10. For the QP, the leading block was nonsingular for the first 22 iterations and singular for the last 16. Notice that the average MINRES iteration counts are correspondingly higher for the QP. This is because, at the LP steps with a nonsingular leading block, we were able to use an ideal preconditioner because the leading block is diagonal, and convergence was always achieved in roughly three iterations. Additionally, the nonzero Hessian in the QP has some additional terms in the leading block that are dropped in the diagonal leading block approximation once the leading block becomes singular.

Problem		Direct inner solve	MINRES inner solve		
ID	Type	IPM iterations	IPM iterations	Inner iters (average)	
				Predictor	Corrector
`stocfor2`	LP	27	31	4.1	4.1
`TOMLAB37`	QP	31	38	35.1	36.6

Table 4.5: Comparison of IPM iterations using a direct vs. preconditioned MINRES solver for the inner linear system solves. Average number of inner MINRES iterations are reported for both the predictor and corrector steps.

Testing different block approximation strategies. Here we test the WkI Schur complement approximation strategy (see Eq. (3.8)). We use a matrix that arises at the 20th iteration of the IPM solution for the LP `maros_r7` and use $\beta = 0.5$. As we have seen in our earlier LP experiments, by the time the IPM iterations have advanced enough to create a numerically singular leading block, the diagonal has enough large entries that the augmented matrix A_k is mostly diagonally dominant. Thus, using $\mathrm{diag}(A_k)$ is often effective in approximating A_k. We include comparisons between the preconditioners in which:

- A_k approximated by $\mathrm{diag}(A_k)$ and S_k^{-1} is approximated by $B\,\mathrm{diag}(A_k)^{-1}B^T$ (the preconditioner P_D explored in the previous set of experiments);
- A_k is approximated by $\mathrm{diag}(A_k)$ and S_k^{-1} is approximated by $W_k + \beta I$ ("Diagonal+WkI" or "D+WkI").

For this experiment, our weight matrix W_k has rank 2,911 (the minimum required to achieve structural nonsingularity of $A_{drop} + B^T W_k B$).

A convergence plot is shown in Figure 4.2. The P_D preconditioner converges in 11 iterations and 1.4 seconds (0.1 seconds per iteration), and the Diagonal+WkI preconditioner in 102 iterations and 0.18 seconds (0.0018 seconds per iteration). While this is a significantly higher iteration count, we notice that this preconditioner is extremely cheap (in that it is fully diagonal) and thus results in faster computational time overall. We note that a basic Jacobi iteration on the original system (or Jacobi on the leading block combined with the WkI approximation of the Schur complement) does not lead to convergence. Thus, the leading block augmentation has utility in arriving at this surprisingly simple-looking preconditioner.

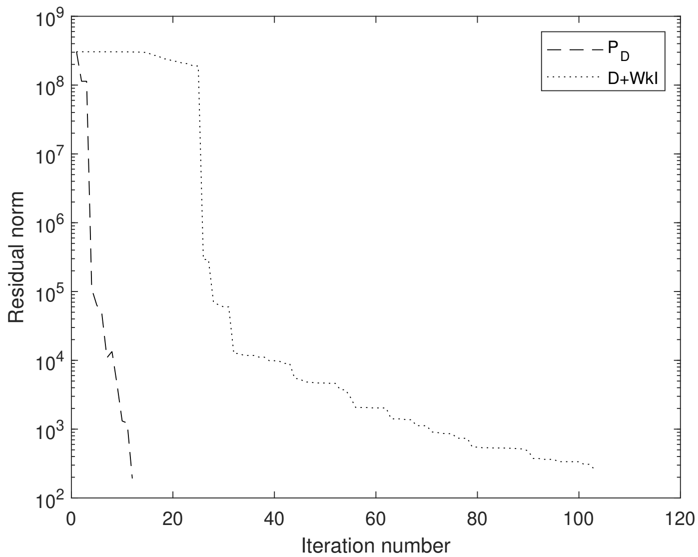

Fig. 4.2: Comparison of block approximation strategies (diagonal leading block + $B(\mathrm{diag}(A_k))^1B^T$ Schur complement; Diagonal leading block+WkI Schur complement) for a matrix arising from an IPM on the `lp_maros_r7` problem.

4.3. A geophysical inverse problem.

Problem statement. Here we consider the example of a geophysical inverse problem described in [12], which involves recovering a model based on observations of a field. The regularized problem is defined by

$$\begin{aligned}\min_{m,u}\ & \frac{1}{2}||Qu-b||^2+\frac{\beta}{2}||W(m-m_{ref})||^2\\ \text{s.t.}\ & A(m)u=q,\end{aligned}$$

where β is a regularization parameter, m is a model, m_{ref} is a reference model, W is a weight matrix, and $A(m)$ is a nonlinear map that encodes the model conditions of the field being considered. If Gauss-Newton iterations are used, the linear system to be solved at each step takes the form

$$\begin{bmatrix} Q^TQ & 0 & F^T \\ 0 & \beta W^TW & G^T \\ F & G & 0 \end{bmatrix}\begin{bmatrix}\delta u\\ \delta m\\ \delta\lambda\end{bmatrix} = -\begin{bmatrix} r_u\\ r_m\\ r_\lambda\end{bmatrix},$$

where F is a sparse, large, nonsingular matrix that stands for the value of the nonlinear map A at the current iterate, m_k, and G is the Jacobian of A evaluated at the current iterate, m_k. In the typical case of sparse observations, G is sparse and Q^TQ has high nullity.

Testing different block approximation strategies. In this experiment we test the BFBt Schur complement approximation strategy (Eq. (3.9)). We set the regularization parameter $\beta = 10^{-3}$. The leading block is highly singular, so we augment A by all of B to avoid numerical difficulties (as simply augmenting by enough rows of B to make the augmented matrix structurally nonsingular still leads to a matrix that is highly ill-conditioned).

Recall that the BFBt Schur complement approximation requires two solves for BB^T. Fortunately, for the geophysics problem, this term is sparse and banded. Thus, in computing this approximation, we will solve exactly for the BB^T terms.

We note that the augmented matrix $A + B^TB$ has an interesting structure, as we can see in Figure 4.3: if we partition the matrix into four blocks with the (1,1)-block of size m and the (2,2)-block of size $n - m$, we observed that the (1,1)- and (2,2)-blocks are banded (e.g., for a problem with $m = 9,261$ and $n = 17,261$, the bandwidths are 848 and 421, respectively), and can therefore be solved less expensively than the entire matrix $A + B^TB$. Thus, we use block Jacobi as a preconditioner for an inner preconditioned conjugate gradient (PCG) solver for A_k.

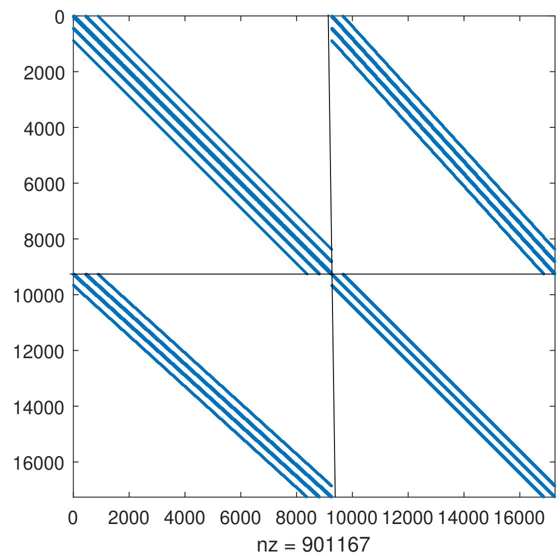

Fig. 4.3: Sparsity pattern of $A_k = A + B^TB$ for a geophysics problem with $m = 9,261$ and $n = 17,261$.

Thus, in these experiments, we compare the preconditioners in which:

- A_k is inverted exactly (which is generally not practical for large problems but is included here for validation and comparison), and S_k^{-1} is approximated with the BFBt approximation. We denote this by "Akinv+BFBt."
- A_k is inverted approximately using CG to an inner tolerance of 0.1, with block Jacobi as a preconditioner, and S_k^{-1} is approximated by the BFBt approximation. We denote

this by "CG+BFBt."

We use MINRES for the Akinv+BFBt preconditioner and FGMRES(30) for the CG+BFBt.

m	n	Akinv+BFBt		CG+BFBt	
		Iters	Time per iter	Iters	Time per iter
2,197	3,195	6	0.21	9	0.20
4,913	9,009	6	1.07	10	0.76
9,261	17,261	8	2.87	10	2.26

Table 4.6: Results (solver iteration counts and time per iteration) geophysics problems of varying size. Akinv+BFBt = exact solve for A_k, BFBt approximation for S_k; CG+BFBt = block Jacobi preconditioned CG for A_k, BFBt for S_k.

Results are shown in Table 4.6. The Akinv+BFBt preconditioner performs well in terms of iteration count, but includes a very expensive term in the A_k solve. We note, however, that the number of preconditioned iterations is very close to what we would expect of the ideal preconditioner (with exact solves for both A_k and S_k), which highlights the effectiveness of the BFBt Schur complement approximation for this problem. The CG+BFBt preconditioner achieves similar convergence to the Akinv+BFBt – in particular, the number of iterations appears to be independent of problem size – and is modestly less expensive per iteration in terms of compute time (we avoid the direct solve for A_k, but have some added expense from the inner CG solves and additional orthogonalization for FGMRES). On average, the inner PCG solves required 28.7 iterations for the first test problem (with $m = 2,197$ and $n = 3,195$), 35.1 iterations for the second problem (with $m = 4,913$ and $n = 9,009$), and 35.8 iterations for the third (with $m = 9,261$ and $n = 17,261$). For larger problems, we speculate that CG+BFBt will outperform Akinv+BFBt by larger margins.

5. Concluding remarks. We have developed a block-diagonal preconditioner for saddle-point systems with a singular leading block. We showed how, by augmenting A with a weight matrix of just high enough rank to overcome its nullity, we yield a preconditioned operator with a small fixed number of distinct eigenvalues. In doing so, we have closed a gap in the existing literature, in analyzing a preconditioning approach for a scenario where the leading block of the saddle-point matrix is neither full rank nor does it have nullity equal to the number of rows of B.

Specifically, we have considered block preconditioners that are based on approximating the augmented leading block of the saddle-point matrix and the augmented Schur complement. Typically, the construction of the weight matrix W_k and the selection of effective approximations may be guided by the problem at hand (for example, in cases where the matrix blocks and Schur complement arise from well-studied discretized differential operators). We have provided some general approaches that may work for different problems. For A_k, we have included diagonal (for LPs), incomplete Cholesky (for QPs), block Jacobi and inner PCG iterations (for geophysics); and for S_k, the $B(\mathrm{diag}(A_k))^{-1}B^T$ and WkI approximations (for the optimization problems), and the BFBt approximation (for the geophysics problem).

We have restricted our attention to diagonal weight matrices with all ones and zeros along the diagonal and have described a method that looks only at the structural rank of a modified augmented matrix. Future work may include more sophisticated choices of the weight matrix, which may in turn yield faster convergence.

REFERENCES

[1] M. BENZI, G. H. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numerica, 14 (2005), pp. 1–137.

[2] M. BENZI AND M. A. OLSHANSKII, *An augmented Lagrangian-based approach to the Oseen problem*, SIAM J. Sci. Comput., 28 (2006), pp. 2095–2113.

[3] S. BRADLEY AND C. GREIF, *Eigenvalue bounds for double saddle-point systems*, Journal of Computational and Applied Mathematics, 424 (2023).

[4] T. A. DAVIS AND Y. HU, *The university of florida sparse matrix collection*, ACM Trans. Math. Softw., 38 (2011).

[5] H. C. ELMAN, *Preconditioning for the steady-state Navier–Stokes equations with low viscosity*, SIAM Journal on Scientific Computing, 20 (1999), pp. 1299–1316.

[6] R. ESTRIN AND C. GREIF, *On nonsingular saddle-point systems with a maximally rank deficient leading block*, SIAM Journal on Matrix Analysis and Applications, 36 (2015), pp. 367–384.

[7] ———, *Towards an optimal condition number of certain augmented Lagrangian-type saddle-point matrices*, Numerical Linear Algebra with Applications, 23 (2016), pp. 693–705.

[8] R. FLETCHER, *An Ideal Penalty Function for Constrained Optimization*, IMA Journal of Applied Mathematics, 15 (1975), pp. 319–342.

[9] G. H. GOLUB AND C. GREIF, *On solving block-structured indefinite linear systems*, SIAM J. Sci. Comput., 24 (2003), pp. 2076–2092.

[10] G. H. GOLUB, C. GREIF, AND J. M. VARAH, *An algebraic analysis of a block diagonal preconditioner for saddle point systems*, SIAM Journal on Matrix Analysis and Applications, 27 (2005), pp. 779–792.

[11] C. GREIF AND D. SCHÖTZAU, *Preconditioners for the discretized time-harmonic Maxwell equations in mixed form*, Numer. Linear Algebra Appl., 14 (2007), pp. 281–297.

[12] E. HABER, U. M. ASCHER, AND D. OLDENBURG, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems, 16 (2000), p. 1263.

[13] S. MEHROTRA, *On the implementation of a primal-dual interior point method*, SIAM Journal on Optimization, 2 (1992), pp. 575–601.

[14] M. F. MURPHY, G. H. GOLUB, AND A. J. WATHEN, *A note on preconditioning for indefinite linear systems*, SIAM Journal on Scientific Computing, 21 (2000), pp. 1969–1972.

[15] J. NOCEDAL AND S. J. WRIGHT, *Numerical optimization*, Springer Series in Operations Research and Financial Engineering, Springer, New York, second ed., 2006.

[16] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM Journal on Numerical Analysis, 12 (1975), pp. 617–629.

[17] T. RUSTEN AND R. WINTHER, *A preconditioned iterative method for saddlepoint problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 887–904.