Electronic Transactions on Numerical Analysis.
Volume 37, pp. 307-320, 2010.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

AN ANALYSIS OF LOW-RANK MODIFICATIONS OF PRECONDITIONERS FOR SADDLE POINT SYSTEMS*

CHEN GREIF† AND MICHAEL L. OVERTON‡

Dedicated to the memory of our mentor and friend, Gene H. Golub

Abstract. We characterize the spectral behavior of a primal Schur-complement-based block diagonal preconditioner for saddle point systems, subject to low-rank modifications. This is motivated by a desire to reduce as much as possible the computational cost of matrix-vector products with the (1,1) block, while keeping the eigenvalues of the preconditioned matrix reasonably clustered. The formulation leads to a perturbed hyperbolic quadratic eigenvalue problem. We derive interlacing results, highlighting the differences between this problem and perturbed linear eigenvalue problems. As an example, we consider primal-dual interior point methods for semidefinite programs, and express the eigenvalues of the preconditioned matrix in terms of the centering parameter.

Key words. saddle point systems, preconditioners, Schur complement, semidefinite programming

AMS subject classifications. 65F08, 65F10, 90C22

1. Introduction.

Consider the following saddle point system coefficient matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{E}^{-1} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix}. \tag{1.1}$$

We assume that $\mathbf{E}$ and $\mathbf{A}$ have dimensions $n \times n$ and $m \times n$ respectively, with $m < n$, that $\mathbf{E}$ is symmetric positive definite, and that $\mathbf{A}$ has rank m. The use of the inverse in the (1,1) block is purely notational, to highlight the fact that we exclude the semidefinite case. We will, however, assume that $\mathbf{E}$ could be very ill-conditioned. Saddle point systems of the form (1.1) arise in numerous applications, ranging from optimization [18] to solution of PDEs [9] to other areas, and their iterative solution has been subject to extensive study in the last couple of decades; see [4] for a comprehensive survey.

A key for the rapid convergence of an iterative method for a linear system of the form $\mathbf{H}x = b$ is the availability of an effective preconditioner, which we will denote throughout by $\mathbf{K}$. Each step of an *outer* iteration for solving the preconditioned linear system $\mathbf{K}^{-1}\mathbf{H}x = \mathbf{K}^{-1}b$ (using, say, MINRES [19]) requires the solution of an *inner* linear system whose coefficient matrix is $\mathbf{K}$. Therefore, convergence of the outer iteration is fast if the eigenvalues of the preconditioned matrix $\mathbf{K}^{-1}\mathbf{H}$ are clustered, but careful attention must be paid to the conditioning and eigenvalue distribution of the matrix $\mathbf{K}$ itself, which determine the speed of convergence of the inner iteration.

Consider the preconditioner

$$\mathbf{K} = \begin{bmatrix} \mathbf{E}^{-1} + \mathbf{A}^T\mathbf{W}^{-1}\mathbf{A} & 0 \\ 0 & \mathbf{W} \end{bmatrix}, \tag{1.2}$$

with $\mathbf{W}$ an $m \times m$ symmetric positive definite matrix. Here, we have set $\mathbf{K}_{11}$, the $(1,1)$ block of $\mathbf{K}$, to the primal Schur complement of the matrix obtained by replacing $\mathbf{H}_{22} = 0$

*Received January 25, 2010. Accepted for publication May 28, 2010. Published online October 10, 2010. Recommended by M. Benzi.

†Department of Computer Science, The University of British Columbia, Vancouver, BC (greif@cs.ubc.ca). The work of this author was partially supported by the Natural Sciences and Engineering Research Council of Canada.

‡Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (overton@cs.nyu.edu). The work of this author was partially supported by the National Science Foundation under award DMS-0714321, and partially by the Pacific Institute for the Mathematical Sciences.

by the stabilizing negative definite matrix $-\mathbf{W}$. A motivation for this is that even if $\mathbf{E}$ is ill-conditioned, by selecting an appropriate weight matrix $\mathbf{W}$ it is possible to make $\mathbf{K}_{11}$ relatively well conditioned. This will enable us to solve the inner systems efficiently using the conjugate gradient method. For a discussion of preconditioning techniques based on this and related approaches, their analysis, and application to boundary value problems, see [5, 6, 14, 21].

In [14] it is shown that with $\mathbf{K}$ defined as in (1.2), the preconditioned matrix $\mathbf{K}^{-1}\mathbf{H}$ has an eigenvalue 1 of algebraic multiplicity n, that the negative eigenvalues all lie between -1 and 0, and that if $\mathbf{H}_{11}$ is allowed to be singular with nullity p (which is not the case in the current paper) then p negative eigenvalues are exactly -1. This characterization of the clustering of the eigenvalues shows that (1.1) can be solved within a small number of outer iterations. The multiplicities of the eigenvalues of the preconditioned matrix hold regardless of the choice of the weight matrix $\mathbf{W}$, and a good choice may help reduce the overall computational cost, by efficiently solving the inner iterations associated with $\mathbf{K}_{11}$.

This leads to the main question that we investigate in this paper. Suppose we want to consider preconditioners of the type (1.2), ensuring that $\mathbf{K}_{11}$ is well conditioned even when $\mathbf{E}$ is ill-conditioned, but at the same time we aim to reduce the cost of performing matrix-vector products with $\mathbf{K}_{11}$. This may occur when the construction of rows of $\mathbf{A}$ or their multiplication with a vector entails a high computational cost. One way to address this is by replacing $\mathbf{W}^{-1}$ in $\mathbf{K}_{11}$ by a simple, lower rank matrix $\mathbf{V}$. Then, if $\mathbf{V}$ is diagonal and some of its diagonal entries are zero, not all rows of $\mathbf{A}$ are used when forming matrix-vector products with $\mathbf{K}_{11}$. We would like to explore whether this is possible without degrading the condition numbers and the spectral distributions of $\mathbf{K}_{11}$ and $\mathbf{K}^{-1}\mathbf{H}$ too much.

In Section 2, we set the stage for exploring this issue. We set $\mathbf{V} = \mathbf{W}^{-1}$, and provide a few new results on the eigenvalues of the preconditioned matrix, specifically exploring connections to the eigenvalues of the dual Schur complement of (1.1). In Section 3, we present a perturbed hyperbolic quadratic eigenvalue problem and derive new interlacing results. In Section 4, we apply our results to primal-dual interior point methods for semidefinite programming.

For notational convenience, the eigenvalues in the lemmas and theorems below are ordered as follows: eigenvalues of symmetric positive definite matrices are ordered in ascending order; eigenvalues of symmetric indefinite problems are ordered in descending order.

2. Preconditioning with a low-rank weight matrix.

Motivated by the arguments made in the Introduction, consider the following block diagonal matrix, which generalizes (1.2), as a preconditioner for (1.1),

$$\mathbf{K} = \begin{bmatrix} \mathbf{E}^{-1} + \mathbf{A}^T\mathbf{V}\mathbf{A} & 0 \\ 0 & \mathbf{W} \end{bmatrix}, \tag{2.1}$$

where $\mathbf{V}, \mathbf{W}$ are $m \times m$ symmetric matrices. Likewise, the $(1,1)$ block of the preconditioner is now generalized to

$$\mathbf{K}_{11} = \mathbf{E}^{-1} + \mathbf{A}^T\mathbf{V}\mathbf{A}.$$

We will choose $\mathbf{W}$ to be positive definite and $\mathbf{V}$ to be a positive semidefinite rank $m - s$ correction of $\mathbf{W}^{-1}$ as follows,

$$\mathbf{W}^{-1} = \mathbf{V} + \mathbf{Y}^T\mathbf{Y}, \tag{2.2}$$

ETNA
Kent State University
http://etna.math.kent.edu

where $\mathbf{Y} \in \mathbb{R}^{(m-s)\times m}$, $0 \le s \le m$, with full row rank. If $s = m$, $\mathbf{Y}$ is "empty" and $\mathbf{V} = \mathbf{W}^{-1}$, i.e., (2.1) reduces to (1.2). The following lemma includes this case.

LEMMA 2.1. *The preconditioned matrix $\mathbf{K}^{-1}\mathbf{H}$ has an eigenvalue $\varphi = 1$ of algebraic multiplicity $n - m + s$. The corresponding eigenvectors are of the form $(w, \mathbf{W}^{-1}\mathbf{A}w)$. If $s = m$ then* any *set of n linearly independent vectors $w \in \mathbb{R}^n$ qualify. Otherwise, a possible set of eigenvectors $(w, \mathbf{W}^{-1}\mathbf{A}w)$ is defined by $n - m$ vectors w that are linearly independent null vectors of $\mathbf{A}$, and s additional null vectors of $\mathbf{YA}$ that are not null vectors of $\mathbf{A}$, that is, w satisfies $0 \ne \mathbf{A}w \in \mathrm{null}(\mathbf{Y})$.*

Proof. The eigenvalue problem for $\mathbf{K}^{-1}\mathbf{H}$ is

$$\begin{bmatrix} \mathbf{E}^{-1} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix} = \varphi \begin{bmatrix} \mathbf{E}^{-1} + \mathbf{A}^T\mathbf{VA} & 0 \\ 0 & \mathbf{W} \end{bmatrix} \begin{bmatrix} w \\ z \end{bmatrix}. \tag{2.3}$$

From the first block row we have

$$\Big((\varphi - 1)\mathbf{E}^{-1} + \varphi\mathbf{A}^T\mathbf{VA}\Big)w = \mathbf{A}^T z. \tag{2.4}$$

If $\varphi = 1$, (2.4) simplifies to $\mathbf{A}^T\mathbf{VA}w = \mathbf{A}^T z$. In this case, from the second block row of (2.3) we have $z = \mathbf{W}^{-1}\mathbf{A}w$, and hence

$$\mathbf{A}^T\mathbf{VA}w = \mathbf{A}^T\mathbf{W}^{-1}\mathbf{A}w. \tag{2.5}$$

We can readily see that there are vectors $w \ne 0$ that satisfy this equation, and therefore $\varphi = 1$ is indeed an eigenvalue of $\mathbf{K}^{-1}\mathbf{H}$. Notice that if $\mathbf{A}w \ne 0$, then $\mathbf{A}^T\mathbf{W}^{-1}\mathbf{A}w \ne 0$, since $\mathbf{A}$ has full row rank. If $\mathbf{V} = \mathbf{W}^{-1}$ then (2.5) holds for any w. Otherwise, under relation (2.2) between $\mathbf{V}$ and $\mathbf{W}$, (2.5) simplifies to $(\mathbf{YA})^T\mathbf{YA}w = 0$. Since $\mathrm{rank}(\mathbf{Y}) = m - s$ there are s linearly independent vectors other than the null vectors of $\mathbf{A}$ that satisfy this relation, and they are as stated in the lemma. □

In the special case $\mathbf{V} = \mathbf{W}^{-1}$ we can provide further insight. Let us first show that a matrix we will need to invert later is nonsingular.

LEMMA 2.2. *Suppose $\mathbf{V} = \mathbf{W}^{-1}$ and let φ be an eigenvalue of $\mathbf{K}^{-1}\mathbf{H}$. Then the matrix*

$$\mathbf{T}(\varphi) = (\varphi - 1)\mathbf{E}^{-1} + \varphi\mathbf{A}^T\mathbf{VA}$$

is singular if $\varphi = 1$ and nonsingular otherwise.

Proof. First, note that since $\mathbf{H}$ is nonsingular, φ cannot be zero. Since $\mathbf{K}$ is symmetric positive definite, $\mathbf{K}^{1/2}$ exists and the eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ are identical to those of the symmetric matrix $\mathbf{K}^{-1/2}\mathbf{H}\mathbf{K}^{-1/2}$. The inertia of the latter is equal to the inertia of $\mathbf{H}$, and hence we must have that n eigenvalues φ are positive and m are negative. By Lemma 2.1 the multiplicity of the positive eigenvalue $\varphi = 1$ is n, and therefore all the remaining eigenvalues φ must be negative.

If $\varphi = 1$ then $\mathbf{T} = \mathbf{A}^T\mathbf{VA}$, which is singular since it is $n \times n$ but its rank is at most m. If $\varphi \ne 1$ then we must have $\varphi < 0$ by the above inertia considerations, and T in this case is negative definite, hence nonsingular. □

Theorem 2.4 below relates the eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ to the eigenvalues of the dual Schur complement of (1.1),

$$\mathbf{M} = \mathbf{AEA}^T. \tag{2.6}$$

We start with a lemma.

LEMMA 2.3. *Define*

$$\widehat{\mathbf{E}}^{-1} = (\varphi - 1)\mathbf{E}^{-1}, \qquad \widehat{\mathbf{V}} = \varphi\mathbf{V}, \qquad \widehat{\mathbf{M}} = \mathbf{A}\widehat{\mathbf{E}}\mathbf{A}^T.$$

ETNA
Kent State University
http://etna.math.kent.edu

Then, if $\varphi \neq 1$,

$$(\mathbf{A}\mathbf{T}^{-1}\mathbf{A}^T)^{-1} = \widehat{\mathbf{M}}^{-1} + \widehat{\mathbf{V}}.$$

Proof. This follows readily from [10], or can be obtained by using the Sherman-Morrison formula; see also [5, 11]. We have

$$\mathbf{T}^{-1} = (\widehat{\mathbf{E}}^{-1} + \mathbf{A}^T\widehat{\mathbf{V}}\mathbf{A})^{-1} = \widehat{\mathbf{E}} - \widehat{\mathbf{E}}\mathbf{A}^T(\mathbf{I} + \widehat{\mathbf{V}}\mathbf{A}\widehat{\mathbf{E}}\mathbf{A}^T)^{-1}\widehat{\mathbf{V}}\mathbf{A}\widehat{\mathbf{E}}.$$

Thus

$$\begin{aligned}\mathbf{A}\mathbf{T}^{-1}\mathbf{A}^T &= \mathbf{A}\left(\widehat{\mathbf{E}} - \widehat{\mathbf{E}}\mathbf{A}^T(\mathbf{I} + \widehat{\mathbf{V}}\mathbf{A}\widehat{\mathbf{E}}\mathbf{A}^T)^{-1}\widehat{\mathbf{V}}\mathbf{A}\widehat{\mathbf{E}}\right)\mathbf{A}^T \\ &= \widehat{\mathbf{M}} - \widehat{\mathbf{M}}(\mathbf{I} + \widehat{\mathbf{V}}\widehat{\mathbf{M}})^{-1}\widehat{\mathbf{V}}\widehat{\mathbf{M}}.\end{aligned}$$

One can verify that $\mathbf{I} - (\mathbf{I} + \widehat{\mathbf{V}}\widehat{\mathbf{M}})^{-1}\widehat{\mathbf{V}}\widehat{\mathbf{M}} = (\mathbf{I} + \widehat{\mathbf{V}}\widehat{\mathbf{M}})^{-1}$. Finally, it is immediate to see that $\widehat{\mathbf{M}}(\mathbf{I} + \widehat{\mathbf{V}}\widehat{\mathbf{M}})^{-1} = (\widehat{\mathbf{M}}^{-1} + \widehat{\mathbf{V}})^{-1}$, which completes the proof. ∎

Continuing on with considering $\mathbf{V} = \mathbf{W}^{-1}$, the specific choice of a scalar multiple of the identity allows us to relate the eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ to the eigenvalues of $\mathbf{M}$; we denote the latter by

$$0 < \gamma_1 \leq \gamma_2 \leq \cdots \gamma_m. \tag{2.7}$$

We have the following result.

THEOREM 2.4. *Suppose* $\mathbf{V} = \mathbf{W}^{-1} = \beta\mathbf{I}$, *and let* γ_i *be the eigenvalues of* $\mathbf{M}$ *defined in (2.6), ordered as in (2.7). Then, the eigenvalues of the preconditioned matrix* $\mathbf{K}^{-1}\mathbf{H}$ *are given, in descending order, as follows:*

$$\begin{aligned} &\varphi_j = 1, && j = 1, \ldots, n; \\ &-1 < \varphi_{n+j} = \frac{-\beta\gamma_j}{\beta\gamma_j + 1} < 0, && j = 1, \ldots, m. \end{aligned} \tag{2.8}$$

Proof. The multiplicity of $\varphi = 1$ has been established in Lemma 2.1. Consider now $\varphi \neq 1$. The matrix multiplying w on the left hand side of (2.4), namely $\mathbf{T}(\varphi)$, is nonsingular by Lemma 2.2. Multiplying (2.4) by $\mathbf{T}^{-1}$ and using $\mathbf{A}w = \varphi\mathbf{W}z$ from (2.3), we obtain

$$\mathbf{A}\mathbf{T}^{-1}\mathbf{A}^T z = \varphi\mathbf{W}z. \tag{2.9}$$

By Lemma 2.3, (2.9) is equivalent to

$$z = \varphi(\widehat{\mathbf{M}}^{-1} + \widehat{\mathbf{V}})\mathbf{W}z. \tag{2.10}$$

Substituting $\widehat{\mathbf{M}}^{-1} = (\varphi - 1)\mathbf{M}^{-1}$, $\widehat{\mathbf{V}} = \varphi\beta\mathbf{I}$, and $\mathbf{W}^{-1} = \beta\mathbf{I}$, (2.10) is equivalent to

$$\varphi(\varphi - 1)\mathbf{M}^{-1}z = \beta(1 - \varphi^2)z, \tag{2.11}$$

or $\mathbf{M}z = -\frac{1}{\beta}\frac{\varphi}{1+\varphi}z$. Thus $\gamma_j = -\frac{\varphi_{n+j}}{\beta(1+\varphi_{n+j})}$, which gives the second equation of (2.8). ∎

It follows from Theorem 2.4 that the value of β may be used to control the eigenvalues. The larger it is, the closer the negative eigenvalues given in (2.8) are to -1, and hence the smaller the number of expected MINRES iterations. However, there is a tradeoff, because the rate of convergence of the preconditioned *inner* iteration, namely the linear system solve

ETNA
Kent State University
http://etna.math.kent.edu

for $\mathbf{K}_{11}$, depends in a different way on β. Let us make the assumption, valid in the case of semidefinite programming discussed in Section 4, that the costs of multiplication of vectors by $\mathbf{E}$ and $\mathbf{E}^{-1}$ are comparable. Thus $\mathbf{E}$ provides a preconditioner for $\mathbf{K}_{11}$, and the spectrum of $\mathbf{E}\mathbf{K}_{11} = \mathbf{I} + \mathbf{E}\mathbf{A}^T\mathbf{V}\mathbf{A}$ controls the rate of convergence of the conjugate gradient method to solve systems whose coefficient matrix is $\mathbf{K}_{11}$. When $\mathbf{V}$ is a multiple of the identity, the characterization of this spectrum is straightforward.

LEMMA 2.5. *Suppose that $\mathbf{V} = \beta\mathbf{I}$. Then $n - m$ eigenvalues of*

$$\mathbf{E}\mathbf{K}_{11} = \mathbf{I} + \mathbf{E}\mathbf{A}^T\mathbf{V}\mathbf{A}$$

are equal to 1, and the remaining m eigenvalues have the form $1 + \beta\gamma_j$, $j = 1, \ldots, m$.

Proof. This is a consequence of the fact that the nonzero eigenvalues of the matrix product $(\mathbf{E}\mathbf{A}^T\mathbf{V})\mathbf{A}$ equal the nonzero eigenvalues of the product $\mathbf{A}(\mathbf{E}\mathbf{A}^T\mathbf{V})$. ∎

Thus, the wish to make β large to speed up convergence of the outer MINRES iteration conflicts with the desire to make β small to improve the rate of convergence of the inner conjugate gradient iteration.

3. Interlacing for a quadratic eigenvalue problem.

In this section we extend the results of Section 2 to the case where, instead of setting $\mathbf{V}$ to a multiple of the identity matrix, we choose it to have lower rank; the multiplication of vectors by $\mathbf{A}^T\mathbf{V}\mathbf{A}$ in the "inner" iteration is then less costly. We first make an easy generalization of Lemma 2.5 using standard eigenvalue interlacing results, and then we go on to generalize Theorem 2.4 by extending the interlacing results to the quadratic eigenvalue problem that arises.

The discussion that ensues shows that the eigenvalues of the preconditioned matrix can be expressed in terms of a low-rank modification of a hyperbolic quadratic eigenvalue problem (QEP). There is a rich mathematical theory for QEPs; see the excellent review [22] and the recent paper [15]. However, they are not as well understood as their linear eigenvalue problem counterparts. For example, interlacing results for these problems are fairly scarce; see [20].

LEMMA 3.1. *Suppose that $\mathbf{V}$ is diagonal with s diagonal values set to β and the other $m - s$ values equal to zero. Denote the eigenvalues of $\mathbf{E}\mathbf{K}_{11} = \mathbf{I} + \mathbf{E}\mathbf{A}^T\mathbf{V}\mathbf{A}$ by δ_j, $j = 1, \ldots, n$, ordered in ascending order. Then we have*

$$\begin{aligned} &\delta_j = 1, && j = 1, \ldots, n - s;\\ &1 + \beta\gamma_j \le \delta_{n-s+j} \le 1 + \beta\gamma_{j+m-s}, && j = 1, \ldots, s. \end{aligned}$$

Proof. This follows from the interlacing property for symmetric matrices ([12, Theorem 8.1.8], [25, pp. 94-97]), because the matrix $\mathbf{M}\mathbf{V} = \mathbf{A}\mathbf{E}\mathbf{A}^T\mathbf{V}$ is a rank $m - s$ perturbation of $\beta\mathbf{M}$. ∎

This result includes Lemma 2.5 as the special case $s = m$. Likewise, the following result includes Theorem 2.4 as the special case $s = m$.

THEOREM 3.2. *Suppose $\mathbf{W}^{-1} = \beta\mathbf{I}$. Let $\mathbf{V}$ be a diagonal matrix with s of its diagonal values equal to β and the rest zero. Denote the eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ in this case by ν_j, ordered in descending order. Then, for β sufficiently large,*

$$\begin{aligned} &\nu_j > 1, && j = 1, \ldots, m - s;\\ &\nu_j = 1, && j = m - s + 1, \ldots, n;\\ &-1 < \frac{-\beta\gamma_{j+m-s}}{\beta\gamma_{j+m-s} + 1} \le \nu_{n+j} \le \frac{-\beta\gamma_j}{\beta\gamma_j + 1} < 0, && j = 1, \ldots, s;\\ &\nu_j < -1, && j = n + s + 1, \ldots, n + m. \end{aligned}$$

Proof. By Lemma 2.3 and equations (2.9)–(2.10), which hold for any choice of $\mathbf{V}$, we have a quadratic eigenvalue problem in ν,

$$\beta z = \nu\Big((\nu-1)\mathbf{M}^{-1} + \nu\mathbf{V}\Big)z.$$

The case $s=m$ follows from Theorem 2.4. Below we present a proof for the case $s=m-1$ (that is, $\mathbf{V}$ is a rank-1 change of $\mathbf{W}^{-1}$). Consider the spectral decomposition

$$\mathbf{M} = \mathbf{U\Gamma U}^T, \quad \Gamma = \operatorname{diag}(\gamma_1,\dots,\gamma_m).$$

Note that $\mathbf{W}^{-1}-\mathbf{V}$ is diagonal, with $m-s$ nonzero elements, all equal to β. We have

$$\Big(\nu^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}-uu^T) - \nu\mathbf{\Gamma}^{-1} - \beta\mathbf{I}\Big)\tilde{z} = 0,$$

where u is a column vector and $\tilde{z} = \mathbf{U}^T z$. This is a rank-1 change to the diagonal quadratic eigenvalue problem

$$\Big(\underbrace{\varphi^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}) - \varphi\mathbf{\Gamma}^{-1} - \beta\mathbf{I}}_{\Phi(\varphi)}\Big)\hat{z} = 0, \tag{3.1}$$

which corresponds to the case $\mathbf{V}=\mathbf{W}^{-1}$ after performing a step of diagonalization; cf. (2.11). This case is covered by Theorem 2.4. It is straightforward to show that this quadratic eigenvalue problem is hyperbolic [15, Definition 1.1].

By inertia considerations similar to those presented in the proof of Lemma 2.2 we must have n positive and m negative eigenvalues. The existence and multiplicity $n-m+s=n-1$ of the eigenvalue 1 follows from Lemma 2.1. Suppose now that $\nu \neq 1$, and consider the matrix in (3.1), namely $\Phi(\nu)$. It is singular if and only if $\hat{z}$ is an eigenvector of (3.1). But this is covered by Theorem 2.4. If Φ is nonsingular then

$$\begin{aligned}0 &= \det\Big(\nu^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}-uu^T) - \nu\mathbf{\Gamma}^{-1} - \beta\mathbf{I}\Big)\\ &= \det\Big(\nu^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}) - \nu\mathbf{\Gamma}^{-1} - \beta\mathbf{I}\Big)\\ &\quad\times\ \det\Big(\mathbf{I} - \Big(\nu^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}) - \nu\mathbf{\Gamma}^{-1} - \beta\mathbf{I}\Big)^{-1}\nu^2 uu^T\Big).\end{aligned} \tag{3.2}$$

By our assumption, the first determinant on the right hand side of equation (3.2), which is nothing but $\det(\Phi(\nu))$, is nonzero and hence the second determinant must be zero. For any two vectors x and y we have $\det(I+xy^T) = 1 + y^T x$ [7, Lemma 5.1], and hence

$$\begin{aligned}&\det\Big(\mathbf{I} - \Big(\nu^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}) - \nu\mathbf{\Gamma}^{-1} - \beta\mathbf{I}\Big)^{-1}\nu^2 uu^T\Big)\\ &\quad= 1 - \nu^2 u^T\Big(\nu^2(\mathbf{\Gamma}^{-1}+\beta\mathbf{I}) - \nu\mathbf{\Gamma}^{-1} - \beta\mathbf{I}\Big)^{-1} u\\ &\quad= \underbrace{1 - \sum_{j=1}^{m} \frac{\nu^2 u_j^2}{\nu^2(\gamma_j^{-1}+\beta) - \nu\gamma_j^{-1} - \beta}}_{g(\nu)}\ .\end{aligned}$$

The expressions in the denominator can be factored as

$$q_j(\nu) \equiv \nu^2(\gamma_j^{-1}+\beta) - \nu\gamma_j^{-1} - \beta = \big((\gamma_j^{-1}+\beta)\nu + \beta\big)(\nu-1), \quad j=1,\dots,m.$$

ETNA
Kent State University
http://etna.math.kent.edu

Denoting the expression for the determinant by $g(\nu)$, the poles of g are the roots of q_j, namely, 1 and the negative values given in (2.8). We have

$$g'(\nu) = \sum_{j=1}^{m} \frac{u_j^2 \nu(\nu\gamma_j^{-1} + 2\beta)}{\Big(\big((\gamma_j^{-1} + \beta)\nu + \beta\big)\,(\nu - 1)\Big)^2}.$$

When $\nu > 0$ we have $g'(\nu) > 0$, since all quantities in the expression for g' are positive. Therefore, the only eigenvalue that is positive but is not equal to 1 must be larger than 1.

For negative ν, we have $g'(\nu) < 0$ if $\nu\gamma_j^{-1} + 2\beta > 0$. From Theorem 2.4 it follows that the m poles of the function $g(\nu)$ are algebraically larger than -1. For $\nu > -1$ we have $\nu\gamma_j^{-1} + 2\beta > -\gamma_j^{-1} + 2\beta$, and hence for β sufficiently large we have $g'(\nu) < 0$. Thus, the subset of eigenvalues that are negative, $\{\nu_j\}_{j=n+1}^{n+m}$, are equal to or algebraically smaller than their counterparts for $\mathbf{V} = \mathbf{W}^{-1}$, $\{\varphi_j\}_{j=n+1}^{n+m}$, and $m - s = 1$ of them are smaller than -1.

When $m - s$ is larger than 1, the proof is obtained by considering a sequence of rank- 1 changes to a diagonal quadratic eigenvalue problem. The details are omitted for the sake of brevity. $\square$

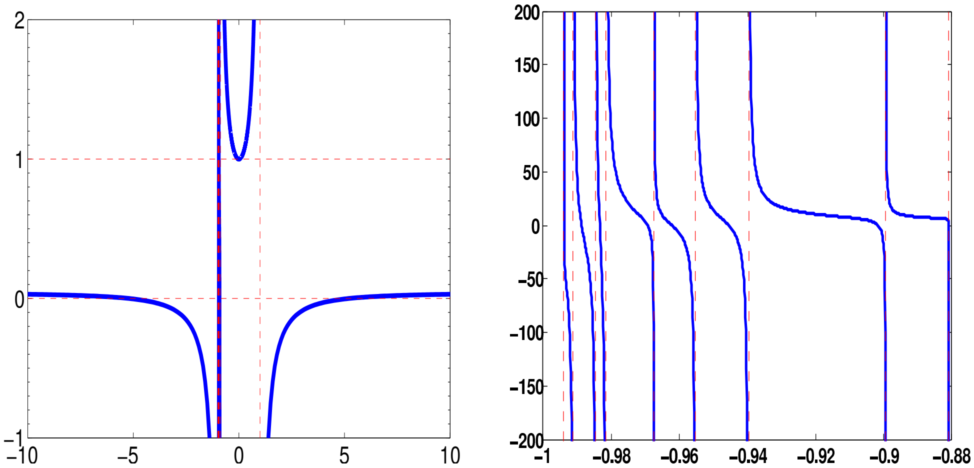

FIG. 3.1. *Left: a schematic illustration of the interlacing phenomenon for the quadratic eigenvalue problem, with a rank-1 change. Most of the variation occurs near* -1*. Right: a close-up view of the interlacing of the negative eigenvalues near* -1*.*

For clarity, we provide a characterization of the function $g(\nu)$, under the conditions stated in Theorem 3.2, for a rank-1 change (that is, $s = m - 1$); see also Fig. 3.1.

- It has poles located at the eigenvalues of the problem described in Theorem 2.4. There are m negative poles, and one positive pole at 1.
- It has a positive derivative for $\nu > 0$ and a negative derivative for $\nu < 0$.
- Between 0 and 1 it has no roots. We have that $g(0) = 1$, and for $0 < \nu < 1$, g is monotonically increasing and approaches ∞ as $\nu \to 1^-$.
- For $\nu > 1$, the function is monotonically increasing. For $\nu \to 1^+$, we have $g(\nu) \to -\infty$. As $\nu \to \infty$, we have

$$g(\nu) \to 1 - \sum_{i=1}^{m} \frac{u_i^2}{\gamma_i^{-1} + \beta}.$$

For β sufficiently large this value is positive and hence one root greater than 1 exists.

ETNA
Kent State University
http://etna.math.kent.edu

- All other roots are negative. Of them, $m-1$ are between -1 and 0. One root is smaller than the smallest pole. The smallest pole is larger than -1, but the root may or may not be smaller than -1.
- As $\nu \to -\infty$, again $g(\nu) \to 1 - \sum_{i=1}^{m} \frac{u_i^2}{\gamma_i^{-1}+\beta}$.

Fig. 3.2 illustrates the effect of a rank-3 modification on the spectrum of the preconditioned matrix. Six eigenvalues of the preconditioned matrix 'escape' from the range where the other eigenvalues are trapped: these are the three leftmost and three rightmost eigenvalues shown in the panel on the right. The tradeoff is interesting: the outer MINRES iteration count will increase by up to six steps, but every inner conjugate gradient step requires three fewer inner products with rows of $\mathbf{A}$ and three fewer inner products with columns of $\mathbf{A}$.

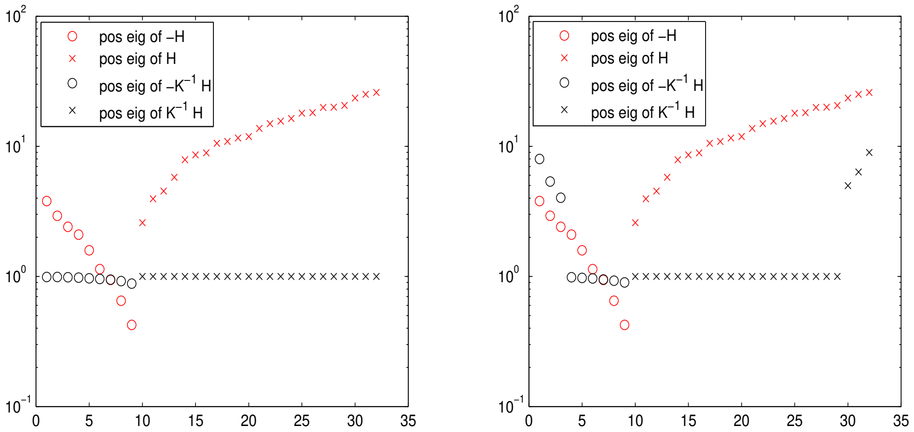

FIG. 3.2. *Eigenvalues of a preconditioned matrix* $\mathbf{K}^{-1}\mathbf{H}$, *with* $n = 23$ *and* $m = 9$, *where* $\mathbf{A}$ *is randomly generated, and* $\mathbf{E}^{-1} = \mathbf{B}^T\mathbf{B}$ *with* $\mathbf{B}$ *randomly generated. Left:* $\mathbf{V} = \mathbf{I}$. *Right:* $\mathbf{V}$ *is diagonal with* 20 *ones and* 3 *zeros.*

4. Application to semidefinite programming. We discuss in this section the applicability of the proposed preconditioning technique to nondegenerate semidefinite programs (SDPs). Preconditioning of SDPs is an important and active research topic [24, 26].

Let $\mathcal{S}^N$ denote the Euclidean space of $N \times N$ real symmetric matrices with inner product $X \bullet Y = \operatorname{tr} XY$, and let $X \succeq 0$ (respectively $X \succ 0$) mean that X is positive semidefinite (positive definite). Consider the primal SDP

$$\begin{array}{ll} \min\limits_{X \in \mathcal{S}^N} & C \bullet X \\ \text{such that} & A_k \bullet X = b_k, \ k = 1, \dots, m, \\ & X \succeq 0, \end{array}$$

where $b \in \mathbb{R}^m$, $C \in \mathcal{S}^N$, and the m data matrices A_k are linearly independent in $\mathcal{S}^N$.

The dual SDP is

$$\begin{array}{ll} \max_{y \in \mathbb{R}^m,\ Z \in \mathcal{S}^N} & b^T y \\ \text{such that} & \sum_{k=1}^{m} y_k A_k + Z = C. \\ & Z \succeq 0. \end{array}$$

In practice, the matrices X and Z almost always have a prescribed block-diagonal structure. The results below all extend to the block diagonal case, but the necessary notation is quite

ETNA
Kent State University
http://etna.math.kent.edu

cumbersome. Note that SDP reduces to linear programming in the case that X and Z are diagonal.

Under the assumption that the SDP has strictly feasible points, that is the primal SDP has a feasible point $X \succ 0$ and the dual SDP has a feasible point (y, Z) with $Z \succ 0$, it is well known that the optimal values of the primal and dual SDP are the same, and that the central path, which consists of triples (X^μ, y^μ, Z^μ) satisfying the primal and dual constraints as well as the centering condition,

$$X^\mu Z^\mu = \mu I, \quad \text{for some } \mu > 0,$$

exists and converges to solutions of the primal and dual SDP as $\mu \downarrow 0$. Primal-dual interior-point path-following methods generate iterates that approximately follow the central path to find solutions to the primal and dual SDP [23]. Widely used publicly available software packages implementing these primal-dual methods include SDPT3, SDPA, and SeDuMi. In contrast, dual-only path-following methods generate only the dual iterates (y, Z), motivated by the fact that Z is generally much more sparse than the primal iterate X [3]. This is because Z is a weighted sum of the data matrices C and the A_k, all of which are generally sparse. On the central path, X is a multiple of the inverse of Z, so it is generally dense. See [8] for a special case where X can be represented efficiently even though it is not sparse.

For both classes of methods, the linear algebra bottleneck that stands in the way of solving large SDPs is as follows. Define

$$n = N^2,$$

and let "vec" map an $N \times N$ matrix to a vector in $\mathbb{R}^n$ by stacking its columns. Let

$$\mathbf{A} = \begin{bmatrix} (\text{vec } A_1)^T \\ \vdots \\ (\text{vec } A_m)^T \end{bmatrix}.$$

The basic linear system that must be solved is

$$\mathbf{H} \begin{bmatrix} \text{vec } \Delta X \\ \Delta y \end{bmatrix} \equiv \begin{bmatrix} \mathbf{E}^{-1} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \text{vec } \Delta X \\ \Delta y \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}. \tag{4.1}$$

For primal-dual methods, the most commonly used formula for $\mathbf{E}$ is the Kronecker product $X \otimes Z^{-1}$, where X and Z are the current primal and dual iterates. The search direction associated with this choice of $\mathbf{E}$ is sometimes called the "HKM" direction and is generally considered more efficient than its primal-dual competitors, the "NT" and "AHO" directions [23]. It follows that the $(1, 1)$ block of $\mathbf{H}$ is $\mathbf{E}^{-1} = X^{-1} \otimes Z$, and hence that the cost of multiplications of vectors by $\mathbf{E}$ and $\mathbf{E}^{-1}$ are similar, assuming the Cholesky factors of X and Z are known (this is always the case, since these matrices cannot be accepted as iterates of the optimization algorithm without checking their positive definiteness). In particular, the preconditioned (1,1) block of (2.1) is $\mathbf{EK}_{11} = \mathbf{I} + (X^{-1} \otimes Z)^{-1}\mathbf{A}^T\mathbf{VA}$ and matrix-vector products with this can be computed efficiently using the identity

$$(X^{-1} \otimes Z)\text{vec}(W) = \text{vec}(ZWX^{-1})$$

(using [16, Lemma 4.3.1], as $X = X^T$).

The discussion above is for primal-dual methods, but for dual-only methods, simply replace X by μZ^{-1}.

ETNA
Kent State University
http://etna.math.kent.edu

One step of block Gauss elimination reduces the system (4.1) to the equivalent normal equations (or Schur complement system)

$$\mathbf{M}\Delta y = f,$$

where $f = \mathbf{AE}f_1 - f_2$ and $\mathbf{M} = \mathbf{AEA}^T$. Thus, using $\mathbf{E} = X \otimes Z^{-1}$, we have

$$\mathbf{M}_{ij} = \mathbf{M}_{ji} = \operatorname{tr} A_i X A_j Z^{-1}, \quad 1 \le i, j \le n.$$

Note that $\mathbf{M} \succ 0$ since $X \succ 0$ and $Z \succ 0$.

In the remainder of this section we develop some analysis that applies on the central path. For any $\mu > 0$, since X^μ and Z^μ commute, there exists an orthogonal matrix Q^μ that simultaneously diagonalizes them, with

$$X^\mu = Q^\mu \ \operatorname{diag}(\lambda_1^\mu, \dots, \lambda_N^\mu)\ (Q^\mu)^T, \quad Z^\mu = Q^\mu \ \operatorname{diag}(\omega_1^\mu, \dots, \omega_N^\mu)\ (Q^\mu)^T, \tag{4.2}$$

and $\lambda_i^\mu \omega_i^\mu = \mu$, $i = 1, \dots, N$. Without loss of generality, assume that

$$\lambda_1^\mu \ge \cdots \ge \lambda_N^\mu \quad \text{and} \quad \omega_1^\mu \le \cdots \le \omega_N^\mu.$$

Letting $\mu \downarrow 0$, we obtain $(X^\mu, y^\mu, Z^\mu) \to (\bar{X}, \bar{y}, \bar{Z})$, which solves the primal and dual SDP [17]. We have the complementarity condition $\bar{X}\bar{Z} = 0$, and there must exist a (not necessarily unique) orthogonal matrix $\bar{Q}$ with

$$\bar{X} = \bar{Q} \ \operatorname{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_N)\ \bar{Q}^T, \quad \bar{Z} = \bar{Q} \ \operatorname{diag}(\bar{\omega}_1, \dots, \bar{\omega}_N)\ \bar{Q}^T,$$

where $\bar{\lambda}_i$ and $\bar{\omega}_i$, the limits of λ_i^μ and ω_i^μ, satisfy $\bar{\lambda}_i \bar{\omega}_i = 0$, $i = 1, \dots, N$. Define

$$r = \operatorname{rank}(\bar{X}),$$

so that

$$\bar{\lambda}_1 \ge \cdots \ge \bar{\lambda}_r > \bar{\lambda}_{r+1} = \cdots = \bar{\lambda}_N = 0.$$

It follows from the complementarity condition that $\bar{Z}$ has rank at most $N - r$. We make the *strict complementarity* assumption that $\bar{Z}$ has rank equal to $N - r$, so that

$$0 = \bar{\omega}_1 = \cdots = \bar{\omega}_r < \bar{\omega}_{r+1} \le \cdots \le \bar{\omega}_N.$$

This holds generically [1], but more importantly, it seems to almost always hold for SDPs that occur in practice. We then have the following trivial lemma.

LEMMA 4.1. *Suppose that the strict complementarity assumption holds. Then the central path eigenvalues satisfy*

$$\lambda_i^\mu = \Theta(1), \quad 1 \le i \le r, \qquad \lambda_i^\mu = \Theta(\mu), \quad r+1 \le i \le N,$$

as $\mu \downarrow 0$.

REMARK 4.2. Recall that the Θ notation signifies a stronger relation than big-O notation: a function $f(n)$ is $\Theta(g(n))$ if f is asymptotically bounded both above and below by g [13, p. 448].

Proof. The first equality holds because for $i \le r$, $\lambda_i^\mu \to \bar{\lambda}_i > 0$ as $\mu \downarrow 0$. The second holds because for $i > r$,

$$\lambda_i^\mu = \frac{\mu}{\omega_i^\mu} \quad \text{and} \quad \omega_i^\mu \to \bar{\omega}_i > 0. \qquad \square$$

Let us partition $\bar{Q} = [\bar{Q}_1\ \bar{Q}_2]$, where $\bar{Q}_1$ has r columns and $\bar{Q}_2$ has $n-r$ columns. We say that the SDP is *primal nondegenerate* if the matrices,

$$\begin{bmatrix} \bar{Q}_1^T A_k \bar{Q}_1 & \bar{Q}_1^T A_k \bar{Q}_2 \\ \bar{Q}_2^T A_k \bar{Q}_1 & 0 \end{bmatrix}, \qquad k = 1, 2, \ldots, m,$$

are linearly independent in $\mathcal{S}^n$, and *dual nondegenerate* if the matrices,

$$\bar{Q}_1^T A_k \bar{Q}_1, \quad k = 1, 2, \ldots, m,$$

span the space $\mathcal{S}^r$. These conditions are well defined even if $\bar{Q}$ is not unique.

In what follows it is convenient to use the notation

$$r^{\overline{2}} = \frac{r(r+1)}{2}.$$

The primal nondegeneracy condition implies that $m \le r^{\overline{2}} + r(n-r)$, and the dual nondegeneracy condition implies that $m \ge r^{\overline{2}}$. Given the strict complementarity assumption, primal nondegeneracy is equivalent to the dual SDP having a unique maximizer, that is, having no other solutions in addition to $(\bar{y}, \bar{Z})$, and dual nondegeneracy is equivalent to the primal SDP having a unique minimizer, that is, no other solutions in addition to $\bar{X}$ [1]. Primal and dual nondegeneracy are generic properties in the sense that randomly generated SDPs with solutions will have both properties, and therefore unique primal and dual solutions, with probability one. However, in practice it is very typical that SDPs are either primal or dual degenerate.

Let $\bar{B}_k = \bar{Q}^T A_k \bar{Q}$. From [16, Lemma 4.3.1], we have

$$\text{vec } \bar{B}_k = (\bar{Q}^T \otimes \bar{Q}^T) \text{ vec } A_k.$$

Thus

$$\bar{\mathbf{B}} \equiv \begin{bmatrix} (\text{vec } B_1)^T \\ \vdots \\ (\text{vec } B_m)^T \end{bmatrix} = \mathbf{A} \left(\bar{Q} \otimes \bar{Q} \right).$$

Each column of $\bar{\mathbf{B}}$ corresponds to an index pair (i, j) identifying two columns of $\bar{Q}$, with $1 \le i, j \le n$. Note that there are $(n-1)^{\overline{2}}$ duplicate columns (one for each pair $i \ne j$). Following [2], we may partition

$$\bar{\mathbf{B}}\mathbf{\Pi} = [\bar{\mathbf{B}}_1\ \bar{\mathbf{B}}_2\ \bar{\mathbf{B}}_3],$$

where $\mathbf{\Pi}$ is a permutation matrix and the columns in $\bar{\mathbf{B}}_1$, $\bar{\mathbf{B}}_2$ and $\bar{\mathbf{B}}_3$ correspond, respectively, to index pairs (i, j) with both i and $j \le r$, exactly one of i, $j \le r$, and neither $\le r$. The dual nondegeneracy condition says that $\bar{\mathbf{B}}_1$ (which has r^2 columns and m rows) has rank $r^{\overline{2}}$ (the other $(r-1)^{\overline{2}}$ columns are redundant), and hence that $m \ge r^{\overline{2}}$.

On the central path, the dual Schur complement matrix is

$$\mathbf{M}^\mu = \mathbf{A}(X^\mu \otimes (Z^\mu)^{-1})\mathbf{A}^T = \mu^{-1}\mathbf{A}(X^\mu \otimes X^\mu)\mathbf{A}^T = \mu^{-1}\mathbf{B}^\mu(\Lambda^\mu \otimes \Lambda^\mu)(\mathbf{B}^\mu)^T,$$

where $\mathbf{B}^\mu = \mathbf{A}(Q^\mu \otimes Q^\mu)$ and $\Lambda^\mu = \text{diag}(\lambda_1^\mu, \ldots, \lambda_N^\mu)$, using Q^μ, λ_i^μ defined in (4.2). It follows that

$$\mu\mathbf{M}^\mu = \mathbf{B}_1^\mu(\Lambda_1^\mu \otimes \Lambda_1^\mu)(\mathbf{B}_1^\mu)^T + \mathbf{B}_2^\mu(\Lambda_1^\mu \otimes \Lambda_2^\mu)(\mathbf{B}_2^\mu)^T + \mathbf{B}_3^\mu(\Lambda_2^\mu \otimes \Lambda_2^\mu)(\mathbf{B}_3^\mu)^T, \tag{4.3}$$

ETNA
Kent State University
http://etna.math.kent.edu

where $\Lambda_1^\mu = \operatorname{diag}(\lambda_1^\mu, \ldots, \lambda_r^\mu)$, $\Lambda_2^\mu = \operatorname{diag}(\lambda_{r+1}^\mu, \ldots, \lambda_N^\mu)$ and, as long as $\mathbf{\Pi}$ is chosen appropriately, $\mathbf{B}_1^\mu$, $\mathbf{B}_2^\mu$ and $\mathbf{B}_3^\mu$ respectively converge to $\bar{\mathbf{B}}_1$, $\bar{\mathbf{B}}_2$ and $\bar{\mathbf{B}}_3$ as $\mu \downarrow 0$. We then have the following lemma.

LEMMA 4.3. *Suppose the strict complementarity and the dual nongeneracy assumptions hold. Then $\mathbf{M}^\mu$ has $\overline{r^2}$ eigenvalues that are $\Theta(\mu^{-1})$ and $m - \overline{r^2}$ eigenvalues that are $O(1)$.*

Proof. Following [2], we consider the scaled dual Schur complement matrix shown in (4.3) as $\mu \downarrow 0$. The second and third terms on the right-hand side converge to zero by Lemma 4.1, while the first converges to $\bar{\mathbf{B}}_1(\bar{\Lambda}_1 \otimes \bar{\Lambda}_1)(\bar{\mathbf{B}}_1)^T$, where $\bar{\Lambda}_1 = \operatorname{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_r)$. The matrix $\bar{\mathbf{B}}_1$ has rank $\overline{r^2}$ by the dual nondegeneracy assumption, while $\bar{\Lambda}_1 \otimes \bar{\Lambda}_1$ is a positive definite diagonal matrix of order r^2. It follows that $\overline{r^2}$ eigenvalues of the $m \times m$ matrix $\mu\mathbf{M}^\mu$ converge to positive numbers. The remaining eigenvalues converge to zero, and since the second and third terms in (4.3) are $O(\mu)$, these eigenvalues are $O(\mu)$ by Lipschitz continuity. Dividing by μ gives the result. $\square$

We are now ready to perform a spectral analysis for the preconditioned system $\mathbf{K}^{-1}\mathbf{H}$ for SDP, assuming that the relevant matrices are evaluated on the central path. The first result is a refinement of Theorem 2.4. Recall that the order of $\mathbf{H}$ in (4.1) is $n + m = N^2 + m$.

THEOREM 4.4. *Suppose that in (4.1),*

$$\mathbf{E}^{-1} = (X^\mu)^{-1} \otimes Z^\mu$$

and that in (2.1),

$$\mathbf{V} = \mathbf{W}^{-1} = \beta\mathbf{I} = \mu^{-\alpha}\mathbf{I}$$

for some $\alpha \geq 0$. Finally, assume that the strict complementarity and dual nondegeneracy conditions hold. Then n eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ are equal to 1, $\overline{r^2}$ eigenvalues φ_i are $-1 + \Theta(\mu^{\alpha+1})$, and $m - \overline{r^2}$ eigenvalues φ_i are $-1 + O(\mu^\alpha)$. These eigenvalues are all algebraically larger than -1.

Proof. The multiplicity of the eigenvalue one follows immediately from Theorem 2.4. From Lemma 4.3, provided strict complementarity and dual nondegeneracy hold, for $\overline{r^2}$ of the eigenvalues of $\mathbf{M}^\mu$ there exists a positive constant c_i independent of μ such that $\gamma_i = \frac{c_i}{\mu}$, and hence by (2.8)

$$\varphi_i = \frac{-c_i\mu^{-(\alpha+1)}}{c_i\mu^{-(\alpha+1)} + 1} = \frac{-c_i}{c_i + \mu^{\alpha+1}}.$$

Therefore for μ sufficiently small we have $\overline{r^2}$ eigenvalues φ_i that are $-1 + \Theta(\mu^{\alpha+1})$. The remaining $m - \overline{r^2}$ eigenvalues correspond to eigenvalues of $\mathbf{M}^\mu$ that are $O(1)$. By a similar argument, these eigenvalues are $-1 + O(\mu^\alpha)$. Furthermore, since $\mu > 0$, these eigenvalues are all larger than -1. $\square$

Next, we present a refinement of Theorem 3.2.

THEOREM 4.5. *Suppose $\mathbf{W}^{-1} = \beta\mathbf{I}$. Let $\mathbf{V}$ be a diagonal matrix with s of its diagonal values equal to β and the rest zero. Denote the eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ in this case by ν_j, ordered in descending order. Suppose, as in the previous theorem, that the iterates follow the central path, that the strict complementarity and dual nondegeneracy conditions hold, and that $\beta = \Theta(\mu^{-\alpha})$. Assume further that $\overline{r^2} \leq s \leq m$. Then, for μ sufficiently small, there are at least $\max(0, \overline{r^2} + s - m)$ eigenvalues of $\mathbf{K}^{-1}\mathbf{H}$ that are $-1 + \Theta(\mu^{\alpha+1})$.*

Proof. By Lemma 4.3, $\overline{r^2}$ eigenvalues γ_i of $\mathbf{M}^\mu$ are $\Theta(\mu^{-1})$ and the remaining $m - \overline{r^2}$ are $O(1)$. Suppose $s = m$. Then by Theorem 4.4 the algebraically smallest negative eigenvalues are $\overline{r^2}$ eigenvalues that are $-1 + \Theta(\mu^{\alpha+1})$. Now suppose $s = m - 1$. By interlacing arguments identical to the ones made in Theorem 3.2, $\overline{r^2} - 1$ of these eigenvalues are trapped between poles of the same magnitude and hence are still of the same order. Since $g'(\varphi) < 0$ for $\varphi < 0$, the remaining negative eigenvalue moves to the left under the perturbation implied by reducing s to $m-1$. However, it is not necessarily of the order $-1+\Theta(\mu^{\alpha+1})$, since it is not trapped by a pole of g on the left. The same arguments can be repeated for $s < m-1$. $\square$

5. Conclusions and future work. We have studied the spectral properties of a primal Schur-complement-based block diagonal preconditioner for saddle point systems, subject to low-rank modifications. A motivation for this approach is the goal of performing matrix-vector products with as few as possible rows of the constraint matrix, while maintaining the effectiveness of the preconditioner. We have taken semidefinite programming as an example. Our focus in this paper is on the analysis, and there is much to do to investigate the practicality of the proposed approach. First, semidefinite programs are typically degenerate, and in such cases, some of our analysis does not hold. Secondly, the strong connection between the spectrum of the preconditioned matrix to that of the dual Schur complement requires a comparison to alternatives that rely on the latter, namely normal equations solvers. Finally, the approach that we have investigated is parameter-dependent, and it would be desirable to explore choices that reduce the overall computational cost.

Acknowledgments. We are grateful to the referees for their constructive comments. In particular, we are indebted to one of them for a thorough and helpful review of our analytical results.

REFERENCES

[1] F. ALIZADEH, J.-P. A. HAEBERLY, AND M. OVERTON, *Complementarity and nondegeneracy in semidefinite programming*, Math. Program., 77 (1997), pp. 111–128.
[2] ———, *Primal-dual interior-point methods for semidefinite programming: convergence rates, stability, and numerical results*, SIAM J. Optim., 8 (1998), pp. 746–768.
[3] S. BENSON, Y. YE, AND X. ZHANG, *Solving large-scale sparse semidefinite programs for combinatorial optimization*, SIAM J. Optim., 10 (2000), pp. 443–461.
[4] M. BENZI, G. GOLUB, AND J. LIESEN, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.
[5] M. BENZI AND M. OLSHANSKII, *An augmented Lagrangian-based approach to the Oseen problem*, SIAM J. Sci. Comput., 28 (2006), pp. 2095–2113.
[6] Z. H. CAO, *Augmentation block preconditioners for saddle point-type matrices with singular* $(1,1)$ *blocks*, Numer. Linear Algebra Appl., 15 (2008), pp. 515–533.
[7] J. DEMMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.
[8] Z. DENG, M. GU, AND M. OVERTON, *Structured primal-dual interior-point methods for banded semidefinite programming*, Oper. Theory Adv. Appl., 202 (2009), pp. 111–141.
[9] H. ELMAN, D. SILVESTER, AND A. WATHEN, *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*, Oxford University Press, Oxford, UK, 2005.
[10] R. FLETCHER, *Practical Methods of Optimization*, Second ed., A Wiley-Interscience Publication, John Wiley & Sons Ltd., Chichester, 1987.
[11] G. GOLUB AND C. GREIF, *On solving block-structured indefinite linear systems*, SIAM J. Sci. Comput., 24 (2003), pp. 2076–2092.
[12] G. GOLUB AND C. VAN LOAN, *Matrix Computations*, Third ed., Johns Hopkins, Baltimore, MD, 1996.
[13] R. GRAHAM, D. KNUTH, AND O. PATASHNIK, *Concrete mathematics*, Second ed., Addison-Wesley Publishing Company, Reading, MA, 1994.
[14] C. GREIF AND D. SCHÖTZAU, *Preconditioners for saddle point linear systems with highly singular* $(1,1)$ *blocks*, Electron. Trans. Numer. Anal., 22 (2006), pp. 114–121. http://etna.math.kent.edu/vol.22.2006/pp114-121.dir.
[15] C. GUO, N. HIGHAM, AND F. TISSEUR, *Detecting and solving hyperbolic quadratic eigenvalue problems*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1593–1613.
[16] R. HORN AND C. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[17] M. KOJIMA, S. SHINDOH, AND S. HARA, *Interior-point methods for the monotone linear complementarity problem in symmetric matrices*, SIAM J. Optim., 7 (1997), pp. 86–125.

[18] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer, New York, 2006.
[19] C. PAIGE AND M. SAUNDERS, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., 12 (1975), pp. 617–629.
[20] B. PLESTENJAK, *Numerical methods for the tridiagonal hyperbolic quadratic eigenvalue problem*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1157–1172.
[21] C. POWELL AND D. SILVESTER, *Optimal preconditioning for Raviart-Thomas mixed formulation of second-order elliptic problems*, SIAM J. Matrix Appl. Anal., 25 (2003), pp. 718–738.
[22] F. TISSEUR AND K. MEERBERGEN, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.
[23] M. TODD, *Semidefinite programming*, Acta Numer., 10 (2001), pp. 515–560.
[24] K. TOH, *Solving large scale semidefinite programs via an iterative solver on the augmented systems*, SIAM J. Optim., 14 (2003), pp. 670–698.
[25] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.
[26] X. ZHAO, D. SUN, AND K. TOH, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim., 20 (2010), pp. 1737–1765.