MAX-MIN EIGENVALUE PROBLEMS, PRIMAL-DUAL INTERIOR POINT ALGORITHMS, and TRUST REGION SUBPROBLEMS

Franz Rendl *   Robert J. Vanderbei †   Henry Wolkowicz ‡

University of Waterloo
Department of Combinatorics and Optimization
Faculty of Mathematics
Waterloo, Ontario N2L 3G1, Canada
e-mail henry@orion.uwaterloo.ca

November 13, 1993

Technical Report CORR 93-30

Abstract

Two Primal-dual interior point algorithms are presented for the problem of maximizing the smallest eigenvalue of a symmetric matrix over diagonal perturbations. These algorithms prove to be simple, robust, and efficient. Both algorithms are based on transforming the

problem to one over the cone of positive semidefinite matrices. One of the algorithms does this transformation through an intermediate transformation to a trust region subproblem. This allows the removal of a dense row.

1 INTRODUCTION

Consider the *max-min eigenvalue problem*

$$(MMP) \quad \omega^* := \max_{v^t e = 0} \lambda_{\min}(C^b - \text{Diag}\,(v)), \qquad (1.1)$$

where C^b is an $(n+1) \times (n+1)$ real symmetric matrix, $v \in \Re^{n+1}$, and $\text{Diag}\,(v)$ denotes the diagonal matrix formed from the vector v. Applications for this problem are many and varied, see e.g. [10, 11]. Essential in many of these applications is a fast algorithm for (MMP), since it has to be solved many times within the application. For example, this problem provides bounds in branch and bound codes for graph bisection, see e.g. [3, 7]. In this paper we study two primal-dual interior point algorithms which show that these problems can be solved robustly and quickly for very large dimensions.

Typically, there is a loss of differentiability due to multiplicity of the smallest eigenvalue at the optimum. In fact, a singleton eigenvalue characterizes differentiability. Since the smallest eigenvalue is a concave function, subgradient approaches can be used to solve (1.1), see e.g. [2]. More recently, it has been shown that Newton-based algorithms with local quadratic convergence are still possible, see e.g. [12]; though the local convergence depends on correctly identifying the multiplicity of the smallest eigenvalue. Since the problem can be rephrased as minimizing a scalar ω subject to all the eigenvalues $\lambda_i(C^b - \text{Diag}\,(v)) \leq t$, degeneracy corresponds to a high multiplicity of the optimal smallest eigenvalue. Both approaches mentioned above have difficulty with high degeneracy. However, just as in ordinary linear programming, the interior point methods we present do not seem to bothered by degeneracy.

Problem (1.1) can be rephrased using positive semidefinite constraints which avoid the nondifferentiability, see e.g. [12, 1]. Interior point methods for problems involving matrix inequalities are studied in [1, 8, 14, 19, 21, 17]. (See the latter for a historical overview.) In this paper we study two equivalent differentiable formulations to (MMP). The first is derived using duality theory for an equivalent max-min trust region subproblem in conjunction with an interior point approach. We then compare this with the second approach which is studied in [14].

---

[0]This paper was presented at the NATO Advanced Study Institute on Algorithms for Continuous Optimization, Il Ciocco, Italy, September, 1993.

The paper is organized as follows. In Section 2 we present several equivalent problems to (MMP). This includes a parametric trust region subproblem as well as a parametric quadratic programming problem. These lead to the first algorithm which we use to solve (MMP). This reformulation of (MMP) is dependent on a given row and column of C^b. The flexibility in choosing the column can have advantages for large sparse problems. We also discuss an alternative reformulation of (MMP) using the cone of positive semidefinite matrices. The motivation for the two formulations is in the applications area as well as in testing out two forms of the algorithm. The equivalent formulations allow applications to 0,1 quadratic programming. In Section 3 we present the first and second derivatives of the matrix functions. In Section 4 we present the primal-dual interior point methods. This includes the matrix derivatives and duality theory. These algorithms performed extremely well and exhibited all the favourable properties of interior point methods for linear programming, e.g.: the iteration count was not affected by the dimension of the problem; degeneracy which here refers to a high multiplicity of the smallest eigenvalue at the optimum, did not affect performance. The algorithms are very robust and did not fail on any problems. The numerical results are reported in detail in [14].

1.1 Preliminaries

We briefly summarize the notation and definitions used in the paper. See e.g. [5] for more details on the various matrix results.

We work in the space of *real symmetric matrices*, denoted $\mathcal{S}$, with the *trace inner product* $\langle A, B \rangle := \operatorname{trace} AB$. The set of *positive semidefinite matrices*, denoted $\mathcal{P}$, forms a closed convex cone, which is *self-polar*, i.e. the *polar cone*

$$\mathcal{P}^+ := \left\{ K = K^t : \operatorname{trace} KP \geq 0, \ \forall P \in \mathcal{P} \right\} = \mathcal{P}.$$

The space $\mathcal{S}$ is endowed with the Loëwner partial order, i.e. $A \succ$ (resp. $\succeq$) B denotes $A - B$ is positive definite (resp. semidefinite).

For $v \in \Re^n$, $\operatorname{Diag}(v)$ denotes the diagonal matrix formed from the vector v. Conversely, for a matrix M, $\operatorname{diag}(M)$, with lower case d, denotes the column vector formed from the diagonal of M. The vector $e \in \Re^n$ denotes the *vector of ones*; while e_i is the *i-th unit vector*; and E_i is the zero matrix with 1 in the i,i position. For a rectangular matrix M, $M^\dagger$ denotes the *Moore-Penrose generalized inverse*. $\mathcal{R}(M)$, $\mathcal{N}(M)$ denote *range space* and *null*

3

space, respectively. For a square matrix M, $\det(M)$ denotes the *determinant*. For two $m \times n$ matrices M, N, the *Hadamard product*, or entrywise product, is denoted $M \circ N$.

2 EQUIVALENT FORMULATIONS

Without loss of generality, we can subtract elements summing to 0 from the diagonal and assume that the element $C_{1,1}^b = 0$, i.e.

$$C^b = \left[\begin{array}{cc} 0 & -b^t \\ -b & C \end{array} \right].$$

(In general, it may be more advantageous to choose a particular dense row and column for the vector b.) Therefore, from Theorem 4.1 in [13], an equivalent max-min problem is the *parametric trust region subproblem*

$$(TRP) \quad (n+1)\omega^* = \max_{u^t e = 0} \ \min_{x^t x = n} x^t(C - \mathrm{Diag}\,(u))x - 2b^t x, \qquad (2.1)$$

where $x, u \in \Re^n$. Then, from the duality theory in Theorem 5.1 in [16], we get the equivalent problem

$$(n+1)\omega^* = \max_{u^t e = 0} \ \max_{C - \mathrm{Diag}\,(u) - \lambda I \succeq 0} n\lambda - b^t(C - \mathrm{Diag}\,(u) - \lambda I)^\dagger b, \qquad (2.2)$$

where $\cdot^\dagger$ denotes the Moore-Penrose generalized inverse. With $y = u + \lambda e$, this is further equivalent to

$$(D) \quad \begin{array}{ll} (n+1)\omega^* = \max & f(y) := y^t e - b^t(C - \mathrm{Diag}\,(y))^\dagger b \\ \text{subject to} & C - \mathrm{Diag}\,(y) \succeq 0 \\ & y \in \Re^n, \end{array} \qquad (2.3)$$

where e is the vector of ones. This is in the form of a dual linear program where our linear operator $\mathrm{Diag}\,(y)$ corresponds to the matrix A^t in linear programming. Therefore, we label it with (D). We solve (MMP) by solving (D). Note that the Moore-Penrose inverse is never actually used in the numerical algorithms, since we will restrict $C - \mathrm{Diag}\,(y) \succ 0$. We summarize the relationships between the various problems in the following.

Theorem 2.1 *Suppose that y solves (2.3) with value $f(y) = (n+1)\omega^*$. Let:*

$$\lambda := \frac{y^t e}{n}, \quad u := y - \lambda e;$$

$$\bar{x} = C_y^\dagger b, \quad \alpha := \bar{x}^t \bar{x};$$

$$x = \bar{x} + w, \quad for\ some\ w \in \mathcal{N}(C_u)\ such\ that\ x^t x = n; \qquad (2.4)$$

and

$$\bar{z} := \begin{pmatrix} 1 \\ \bar{x} \end{pmatrix}, \quad z := \begin{pmatrix} 1 \\ x \end{pmatrix}, \quad t := \lambda + b^t \bar{x}, \quad v := \begin{pmatrix} \frac{n}{n+1} t \\ u - \frac{t}{n+1} e \end{pmatrix}.$$

Then:

1. *u, λ solves (2.2) with optimal value $(n+1)\omega^*$;*

2. *$\alpha \leq n$, w exists for (2.4) and u, x solves the TRP (2.1) with optimal value $(n+1)\omega^*$,*

3. *$z, \bar{z}$ are eigenvectors for the optimal eigenvalue ω^* with optimal perturbation v for (MMP).*

Proof: Statement 1. follows from the translation $y = u + \lambda e$. The duality theory in [16] yields 2. Also, it is shown in [13] that the so-called *hard case* holds at the optimum perturbation. Therefore, $\alpha \leq n$. Now we have that u solves (TRP) with $u^t e = 0$ and x solves the inner minimization problem for this u. Therefore, the following positive semidefiniteness conditions and stationarity conditions hold with Lagrange multiplier λ, see e.g. [4, 15],

$$C - \operatorname{diag}(u) - \lambda I \succeq 0, \quad (C - \operatorname{diag}(u) - \lambda I)\bar{x} = (C - \operatorname{diag}(u) - \lambda)x = b. \quad (2.5)$$

Let $C_u := C - \operatorname{diag}(u)$ and $D(t) := \begin{bmatrix} t & -b^t \\ -b & C_u \end{bmatrix}$. Then λ is an eigenvalue of $D(t)$ with eigenvector $\bar{z}$, since $(C_u - \lambda I)\bar{x} = b$. By (2.5), we could also choose x in the definition of z and t, rather than $\bar{x}$. Moreover, the optimality conditions $C_u - \lambda I \geq 0$ implies $\lambda \leq \lambda_{\min}(C_u)$. Therefore, by the interlacing

theorem for eigenvalues, e.g. [5], $\lambda = \lambda_{\min}(D(t))$. To find the optimal shift v, we see that this implies

$$\lambda - \frac{t}{n+1} = \frac{n\lambda - b^t x}{n+1}$$

is the smallest eigenvalue of the shifted matrix

$$D(t) - \frac{t}{n+1}I = C^b - \text{diag}\,(v),$$

thereby defining the $n+1$ dimensional vector

$$v := \begin{pmatrix} \frac{n}{n+1}t \\ u - \frac{t}{n+1}e \end{pmatrix}, \quad t = \lambda + b^t x. \tag{2.6}$$

Therefore, v is the optimal perturbation of C^b and

$$(n+1)\omega^* = n\lambda - b^t x, \tag{2.7}$$

since the latter is the optimal value of (TRP), see (2.2).
$\square$

The dual problem to problem (D) is the min-max of the Lagrangian, i.e.

$$\min_{X \succeq 0} \max_y f(y) + \text{trace}\, X(C - \text{Diag}\,(y)).$$

The inner problem is the unconstrained maximum of a concave function, and so we can delete the maximization by adding the stationary point condition as a constraint, i.e. we get that the primal problem, or the dual of (D), is

$$\begin{array}{rll} \omega^* = \min & f(y) + \text{trace}\, X(C - \text{Diag}\,(y)) \\ (P) \qquad \text{subject to} & \nabla f(y) - \text{diag}\,(X) = 0 \\ & X \succeq 0. \end{array} \tag{2.8}$$

Moreover, there is no duality gap between primal and dual problems, since Slater's constraint qualification holds for (D). In addition, if y, X is a feasible pair for the primal and dual problems, then the duality gap is just

$$\text{trace}\, X(C - \text{Diag}\,(y)). \tag{2.9}$$

If f is a linear function, then we see that the objective function of the primal becomes trace CX, as expected. For general duality results over general cones, including the cone of positive semidefinite matrices, see ([18]).

We can now apply interior point methods to problem (P) or (D), e.g. we can consider the following dual log-barrier and shifted log-barrier problems with log-barrier parameter $\mu \downarrow 0$:

1.
$$
\begin{aligned}
\max \quad & f(y) + \mu \log \det(C - \text{Diag}\,(y)) \\
\text{subject to} \quad & y \in \Re^n;
\end{aligned}
\tag{2.10}
$$

2.
$$
\begin{aligned}
\max \quad & f(u) \\
\text{subject to} \quad & \mu \log(1 + \tfrac{1}{\mu} \det(C - \text{Diag}\,(y))) \geq 0 \\
& u \in \Re^n.
\end{aligned}
\tag{2.11}
$$

Here det denotes *determinant*. Note that the functions in the two problems are concave. (See the comment at the end of Section 3.)

2.1 ALTERNATIVE FORMULATION

A different formulation for (MMP) is given in e.g. [12, 1, 17, 14]

$$
\omega^* = \max\{\omega : v^t e = 0,\ C^b - \text{Diag}\,(v) \succeq \omega I\}.
\tag{2.12}
$$

Using the relation $y^s = v + \omega e$, we get the equivalent problem

$$
\omega^* = \max\{e^t y^s :\ C^b - \text{Diag}\,(y^s) \succeq 0\}.
\tag{2.13}
$$

The dual is the min-max of the Lagrangian

$$
\min_{X^a \succeq 0} \max_{y^s}\ e^t y^s + \text{trace}\, X^a(C^b - \text{Diag}\,(y^s)).
$$

By setting the first order derivatives of the inner unconstrained problem to 0, we conclude that $e - \text{diag}\,(X^a) = 0$, and $e^t y^s - \text{trace}\, X^a \text{Diag}\,(y^s) = 0$. Therefore, the dual program to (2.13) is

$$
\begin{aligned}
\min \quad & \text{trace}\, C^b X^a \\
\text{subject to} \quad & \text{diag}\,(X^a) = e \\
& X^a \succeq 0.
\end{aligned}
\tag{2.14}
$$

7

This dual can be derived directly from the paired duality results in Theorem 4.1 in [18] upon noting that Diag and diag are adjoint linear operators. Such dual pairs of linear programs over cones behave much like ordinary linear programs over the nonnegative orthant, e.g. the duality gap for a given feasible pair X^a, y^s is $\left\langle X^a, (C^b - \text{Diag}\,(y^s)) \right\rangle$, while complementary slackness corresponds to a zero duality gap.

We can relate this dual pair with our previous equivalent formulation by using the Schur complement.

Theorem 2.2 *Let X^a, y^s be an optimal pair for the dual programs (2.13) and (2.14), with optimal value ω^*. Partition*

$$X^a := \begin{bmatrix} 1 & a^t \\ a & W \end{bmatrix}, \; y^s := \begin{bmatrix} s \\ y \end{bmatrix}.$$

Let:

$$v := y^s - \omega^* e, \; t := \frac{n+1}{n}(s - \omega^*), \; \bar{x} := a;$$

$$\lambda := t - b^t \bar{x} \quad u := y - \lambda e;$$

and

$$x := \bar{x} + w, \quad \text{for some } w \in \mathcal{N}(C - \text{Diag}\,(u)) \text{ such that } x^t x = n;$$

Then 1.,2.,3. of Theorem 2.1 hold. Moreover, $X := W - aa^t$ solves (2.8), the dual of (D).

Proof: First note that the Schur Complement Theorem, see e.g. [9], implies that X^a is positive definite (resp. semidefinite) if and only if the Schur complement $W - aa^t$ is positive definite (resp. semidefinite). Moreover, complementary slackness $X^a(C^b - \text{diag}\,(y^s)) = 0$ implies that $(C - \text{diag}\,(y))a = b$, i.e. a satisfies the optimality conditions for the trust region subproblem. The results now follow from Theorem 2.1.

$\square$

3 DERIVATIVES

In this section we provide the derivatives for Newton's method. Recall that the objective function is

$$f(y) := y^t e - b^t (C - \text{Diag}\,(y))^{-1} b.$$

8

Denote the log-barrier terms by

$$g(y) := \log \det(C - \operatorname{Diag}(y)),$$

and

$$h(y) := \mu \log(1 + \frac{1}{\mu} \det(C - \operatorname{Diag}(y))).$$

In addition, we let e_i denote the i-th unit vector, E_i denote the 0 matrix with 1 in the (i, i) position, and

$$C_y := C - \operatorname{Diag}(y), \ x_y := C_y^{-1} b.$$

From

$$C_y x_y = b, \tag{3.1}$$

we differentiate with respect to y to get

$$\frac{\partial C_y}{\partial y} x_y + C_y \frac{\partial x_y}{\partial y} = 0,$$

i.e. the $n \times n$ Jacobian

$$\frac{\partial x_y}{\partial y} = C_y^{-1} \operatorname{Diag}(x_y), \tag{3.2}$$

while

$$\frac{\partial x_y}{\partial y_i} = C_y^{-1}(E_i x_y). \tag{3.3}$$

Therefore

$$\begin{aligned} \frac{\partial f(u)}{\partial y_i} &= 1 - b^t C_y^{-1}(E_i x_y) \\ &= 1 - \|E_i x_y\|^2. \end{aligned} \tag{3.4}$$

This yields the gradient

$$\frac{\partial f(u)}{\partial y} = e - \operatorname{diag}(x_y x_y^t). \tag{3.5}$$

Given a symmetric matrix B,

$$\frac{\partial \det B}{\partial B} = \operatorname{adj}(B),$$

where adj denotes the *adjoint matrix*. Therefore, by Cramer's rule,

$$
\begin{aligned}
\frac{\partial g(y)}{\partial y_i} &= \frac{-1}{\det C_y}\operatorname{trace} E_i \operatorname{adj}(C_y) \\
&= -\operatorname{trace} E_i C_y^{-1},
\end{aligned}
\tag{3.6}
$$

and

$$
\frac{\partial g(y)}{\partial y} = -\operatorname{diag}(C_y^{-1}).
\tag{3.7}
$$

If we differentiate (3.1) twice we get

$$
-E_i \frac{\partial x_y}{\partial y_j} - E_j \frac{\partial x_y}{\partial y_i} + C_y \frac{\partial^2 x_y}{\partial y_i \partial y_j} = 0,
$$

which yields

$$
\frac{\partial^2 x_u}{\partial y_i \partial y_j} = C_y^{-1}\left(E_i \frac{\partial x_y}{\partial y_j} + E_j \frac{\partial x_y}{\partial y_i}\right).
\tag{3.8}
$$

Also

$$
\begin{aligned}
\frac{\partial^2 f(u)}{\partial y_i \partial y_j} &= -2(E_i x_y)^t \frac{\partial (E_i x_y)}{\partial y_j} \\
&= -2(E_i x_y)^t E_i \frac{\partial x_y}{\partial y_j} \\
&= -2(E_i x_y)^t C_y^{-1}(E_j x_y).
\end{aligned}
\tag{3.9}
$$

Therefore,

$$
\frac{\partial^2 f(u)}{\partial y^2} = -2 x_y x_y^t \circ C_y^{-1},
\tag{3.10}
$$

where $\circ$ denotes the *Hadamard product*.

For a nonsingular matrix function $B(t)$, we can differentiate both sides of $B(t)B(t)^{-1} = I$ and obtain the derivative

$$
\frac{\partial B(t)^{-1}}{\partial t} = -B(t)^{-1}\frac{\partial B(t)}{\partial t} B(t)^{-1}.
$$

Therefore, differentiating g twice yields:

$$
\begin{aligned}
\frac{\partial g(y)}{\partial y_i \partial y_j} &= -\frac{\partial (\operatorname{trace} E_i C_y^{-1})}{\partial y_j} \\
&= -\operatorname{trace} E_i C_y^{-1} E_j C_y^{-1},
\end{aligned}
\tag{3.11}
$$

and

$$
\frac{\partial g(y)}{\partial y^2} = -C_y^{-1} \circ C_y^{-1}.
\tag{3.12}
$$

Similarly,

$$
\begin{aligned}
\frac{\partial h(y)}{\partial y_i} &= \frac{-\mu}{1+\det(C_y)/\mu}\frac{1}{\mu}\operatorname{trace} E_i \operatorname{adj}(C_y) \\
&= \frac{-\det(C_y)}{1+\det(C_y)/\mu}\operatorname{trace} E_i C_y^{-1},
\end{aligned}
\tag{3.13}
$$

and

$$\frac{\partial h(y)}{\partial y} \quad = \quad \frac{-\det(C_y)}{1+\det(C_y)/\mu}\mathrm{diag}\left(C_y^{-1}\right). \tag{3.14}$$

In summary, the gradients and Hessians are:

$$\nabla f = e - \mathrm{diag}\left(x_y x_y^t\right); \tag{3.15}$$

$$\nabla g = -\mathrm{diag}\left(C_y^{-1}\right); \tag{3.16}$$

$$\nabla h = -\frac{-\det(C_y)}{1+\det(C_y)/\mu}\mathrm{diag}\left(C_y^{-1}\right); \tag{3.17}$$

$$\nabla^2 f = -2x_y x_y^t \circ C_y^{-1}; \tag{3.18}$$

$$\nabla^2 g = -C_y^{-1} \circ C_y^{-1}. \tag{3.19}$$

Note that the above implies that the functions f, g are concave when C_y is positive definite, since Schur's Theorem on Hadamard products, see e.g. [6], states that the cone of positive semidefinite matrices is closed under the Hadamard product.

4 PRIMAL-DUAL METHODS

The optimality condition, or stationarity condition, for the unconstrained dual log-barrier problem (2.10) is

$$e - \mathrm{diag}\left(x_y x_y^t\right) - \mu\mathrm{diag}\left(C_y^{-1}\right) = 0. \tag{4.1}$$

Comparing this with the optimality conditions for (2.3) we get the Lagrange multiplier estimate, or primal variable estimate,

$$X = \mu C_y^{-1}. \tag{4.2}$$

The Newton direction for solving (2.10) is found by solving the system

$$(2x_y x_y^t \circ C_y^{-1} + \mu C_y^{-1} \circ C_y^{-1})\delta y = e - \mathrm{diag}\left(x_y x_y^t\right) - \mu\mathrm{diag}\left(C_y^{-1}\right). \tag{4.3}$$

11

The first two terms after the equality correspond to a step to optimality while the last term is a centering step in the dual space. This provides a dual log-barrier method. Alternatively, by adding a slack variable matrix, we can make (2.3) look more like a linear programming constrained problem, i.e. we get

$$(D) \quad \begin{aligned} \omega^* = \max \quad & f(y) := y^t e - b^t (C - \mathrm{Diag}\,(y))^{\dagger} b \\ \text{subject to} \quad & C - \mathrm{Diag}\,(y) - Z = 0 \\ & u \in \Re^n, \ Z \succeq 0. \end{aligned} \qquad (4.4)$$

The primal-dual log-barrier problem is then

$$(B) \quad \begin{aligned} \omega_\mu^* = \max \quad & f(y) + \mu \log \det(Z) \\ \text{subject to} \quad & C - \mathrm{Diag}\,(y) - Z = 0. \end{aligned} \qquad (4.5)$$

The Lagrangian for (4.5) is

$$L(y, Z, X) = f(y) + \mu \log \det(Z) + \mathrm{trace}\, X(C - \mathrm{Diag}\,(y) - Z),$$

while the optimality conditions $F(y, Z, X) = 0$ are:

$$\begin{aligned} e - \mathrm{diag}\,(x_y x_y^t) - \mathrm{diag}\,(X) &= 0 \quad \text{(primal feasibility)} \\ -X + \mu Z^{-1} &= 0 \quad \text{(complementary slackness)} \\ C - \mathrm{Diag}\,(y) - Z &= 0 \quad \text{(dual feasibility)} \\ Z \succeq 0, \ X \succeq 0. & \end{aligned} \qquad (4.6)$$

Here, X, Z are the primal and dual matrices, respectively. Note that the first two optimality conditions correspond to stationarity of the Lagrangian for (4.5), e.g. the second condition is obtained from differentiating the Lagrangian with respect to Z. Recall that the derivative of $\log \det(Z)$ is Z^{-1}. The comments in brackets correspond to the corresponding equations for a linear programming problem, i.e. the corresponding linear programming equations would be: $b - Ax = 0$, $XZe = \mu e$, $c - A^t y - z = 0$. Moreover, note that the transpose, or adjoint operator, of $\mathrm{diag}\,(\cdot)$ is $\mathrm{Diag}\,(\cdot)$.

In our framework, the primal feasibility corresponds to feasibility of x_y in (P), see (2.8). If $\mathrm{diag}\,(X) = 0$, then this corresponds to x_y being feasible for the trust region subproblem. This emphasizes the fact that we do not get feasibility for this subproblem due to the hard case occurring at the optimum. However, at the optimum we also have $ZX = 0$, which means

that the columns of X and also of $X^{\frac{1}{2}}$ form a basis for the null space of $Z = C - \text{Diag}\,(y)$. Therefore. we can use a linear combination of the columns of $X^{\frac{1}{2}}$ to add to x_y to get a feasible optimal solution to the subproblem.

Note that if μ is fixed and we solve the optimality conditions (4.6), then the duality gap (2.9) becomes

$$n\mu.$$

We now apply Newton's method to solve the optimality conditions, $F(y, Z, X) = 0$, while maintaining the positivity of X, Z, i.e. the system $-F'\delta = F$ is:

$$
\begin{bmatrix}
2x_y x_y^t \circ C_y^{-1} & 0 & \text{diag}\,(\cdot) \\
0 & \mu Z^{-1} \cdot Z^{-1} & I \\
\text{Diag}\,(\cdot) & I & 0
\end{bmatrix}
\begin{pmatrix}
\delta y \\
\delta Z \\
\delta X
\end{pmatrix}
=
\begin{pmatrix}
(e - \text{diag}\,(x_y x_y^t) - \text{diag}\,(X)) \\
(-X + \mu Z^{-1}) \\
(C - \text{Diag}\,(y) - Z)
\end{pmatrix}
$$
$$(4.7)$$

The middle equation implies that

$$\delta Z = Z - \frac{1}{\mu} Z X Z - \frac{1}{\mu} Z(\delta X) Z.$$

Moreover, since the final equation is linear, if a stepsize of 1 is taken or if this equation is satisfied by the initial solution estimates, then it is satisfied by each iteration. We assume that this is indeed the case. Therefore, we have

$$\delta Z = -\text{Diag}\,(\delta y), \quad C_y^{-1} = Z^{-1}.$$

We therefore can eliminate δZ and obtain

$$
\begin{bmatrix}
-\frac{1}{\mu} Z \cdot Z & \text{Diag}\,(\cdot) \\
\text{diag}\,(\cdot) & -\nabla^2 f(y)
\end{bmatrix}
\begin{pmatrix}
\delta X \\
\delta y
\end{pmatrix}
=
\begin{pmatrix}
-Z + \frac{1}{\mu} Z X Z \\
\nabla f(y) - \text{diag}\,(X)
\end{pmatrix}
$$

After multiplying through by μ and Z^{-1} and eliminating δX, we get

$$
\begin{bmatrix}
I & -\mu Z^{-1} \text{Diag}\,(\cdot) Z^{-1} \\
0 & \text{diag}\,(\mu Z^{-1} \text{Diag}\,(\cdot) Z^{-1}) - \nabla^2 f(y)
\end{bmatrix}
\begin{pmatrix}
\delta X \\
\delta y
\end{pmatrix}
=
$$
$$
\begin{pmatrix}
\mu Z^{-1} - X \\
\text{diag}\,(-\mu Z^{-1} + X) + \nabla f(y) - \text{diag}\,(X)
\end{pmatrix},
$$

so that

$$\delta X = \mu Z^{-1} - X + \mu Z^{-1} \text{Diag}\,(\delta y) Z^{-1}. \qquad (4.8)$$

In summary, we have the simplified Newton directions for our algorithm 1:

$$
\begin{aligned}
\delta y &= ((2x_y x_y^t + \mu Z^{-1}) \circ Z^{-1})^{-1}(e - \operatorname{diag}(x_y x_y^t) - \mu \operatorname{diag}(Z^{-1})); \\
\delta Z &= -\operatorname{Diag}(\delta y); \\
X + \delta X &= \mu Z^{-1}(Z + \operatorname{Diag}(\delta y))Z^{-1}.
\end{aligned}
\tag{4.9}
$$

Therefore, for the Newton steplength of 1, we can check that both X, Z are positive definite by checking $Z \pm \operatorname{Diag}(\delta y)$, respectively. We discuss the line search in more detail in Section 4.2, below.

4.1 ALTERNATIVE ALGORITHM

The alternative formulation in (2.13) and its dual in (2.14) yield the following primal-dual optimality conditions:

$$
\begin{aligned}
e - \operatorname{diag}(X^a) &= 0 \quad (\text{primal feasibility}) \\
-X^a + \mu Z^{-1} &= 0 \quad (\text{complementary slackness}) \\
C^b - \operatorname{Diag}(y^s) - Z &= 0 \quad (\text{dual feasibility}) \\
Z \succeq 0, \; X^a &\succeq 0.
\end{aligned}
\tag{4.10}
$$

Here X^a, Z are the primal and dual matrices again, but they are not the same as in the above formulation. In fact, they are one dimension larger. This system is more like a linear programming primal dual algorithm in that there are two linear equations and one nonlinear one formed from the complementary slackness conditions.

We now apply Newton's method to solve the optimality conditions, while maintaining the positivity of X^a, Z, i.e. we solve

$$
\begin{bmatrix}
0 & 0 & \operatorname{diag}(\cdot) \\
0 & \mu Z^{-1} \cdot Z^{-1} & I \\
\operatorname{Diag}(\cdot) & I & 0
\end{bmatrix}
\begin{pmatrix}
\delta y^s \\
\delta Z \\
\delta X^a
\end{pmatrix}
=
\begin{pmatrix}
(e - \operatorname{diag}(X^a)) \\
(-X^a + \mu Z^{-1}) \\
(C^b - \operatorname{Diag}(y^s) - Z)
\end{pmatrix}.
\tag{4.11}
$$

This leads to similar equations for the steps $\delta y^s, \delta Z, \delta X^a$, except that the nonlinear term is discarded. The equations in simplified form, i.e. when we assume that the slack variable $Z = C^b - \operatorname{Diag}(y^s)$, are:

$$
\begin{aligned}
\delta y^s &= (\mu Z^{-1} \circ Z^{-1})^{-1}(e - \mu \operatorname{diag}(Z^{-1})); \\
\delta Z &= -\operatorname{Diag}(\delta y^s); \\
X^a + \delta X^a &= \mu Z^{-1}(Z + \operatorname{Diag}(\delta y^s))Z^{-1}.
\end{aligned}
\tag{4.12}
$$

14

Again, we can check that both X^a, Z are positive definite by checking $Z \pm$ Diag (δy^s). The above equations yield our second algorithm.

4.2 LINE SEARCHES

For algorithm 1, the line search consists in verifying that the new X and Z are positive definite. At the current iterates $X, Z \succ 0$, we need to check that

$$Z - t\text{Diag}\,(\delta y) \succ 0, \text{ for } Z,$$

where $t > 0$ is the step length. For X, when the error $\|\mu Z^{-1} - X\|$ is small, then a reasonable heuristic is to check

$$Z + t\text{Diag}\,(\delta y) \succ 0, \text{ for } X.$$

Let $Z = R^t R$ be the Cholesky factorization of Z. Then the above is equivalent to finding $t > 0$ such that

$$I \pm tR^{-t}(\text{Diag}\,(\delta y))R^{-1} \succ 0.$$

With this change, the problem is equivalent to the step length problem in ordinary linear programming, i.e. suppose that the matrix Diag (δy) is not negative semidefinite (in which case the step length can be infinite and so is chosen to be 1). Then we get that the step for Z is

$$t_Z := \min\{1, \frac{1}{\lambda_{\max}(R^{-t}(\text{Diag}\,(\delta y))R^{-1})}\};$$

while, if the matrix Diag (δy) is not positive semidefinite (in which case the step length can be infinite and so is chosen to be 1) the step for X is

$$t_X := \min\{1, \frac{1}{-\lambda_{\min}(R^{-t}(\text{Diag}\,(\delta y))R^{-1})}\}.$$

By the above we see that we can use any upper bound for the maximum eigenvalue and any lower bound for the minimum eigenvalue. Note that for an $n \times n$ symmetric matrix K, see [20],

$$\lambda_{\max}(K) \leq \frac{\text{trace}\,K}{n} + \sqrt{n-1}\sqrt{\frac{\text{trace}\,K^2}{n} - \left(\frac{\text{trace}\,K}{n}\right)^2};$$

15

$$\lambda_{\min}(K) \geq \frac{\operatorname{trace} K}{n} - \sqrt{n-1} \sqrt{\frac{\operatorname{trace} K^2}{n} - \left(\frac{\operatorname{trace} K}{n}\right)^2}.$$

We can find the traces using Z^{-1}, e.g.

$$\operatorname{trace}\left(R^{-t}(\operatorname{Diag}(\delta y))R^{-1}\right) = \operatorname{trace}\operatorname{Diag}(\delta y)Z^{-1} = \delta y^t \operatorname{diag}(Z^{-1}).$$

Alternatively, to find the bounds, we could first shift $R^{-t}(\operatorname{Diag}(\delta y))R^{-1}$ by a multiple of the identity to guarantee that it is positive definite and then calculate its norm. We then shift by a multiple of the identity to guarantee that it is negative definite and find the norm of the negative of the matrix. By shifting back, we get the values for the largest and smallest eigenvalues. Alternatively, we could calculate the largest and smallest eigenvalues of $\operatorname{Diag}(\delta y)Z^{-1}$ directly.

Note that if $t_X = 1$, then δX is a good step for X. However, if $t_X < 1$, then $X + t_X \delta X$ is not necessarily equal to $\mu Z^{-1}(Z + \operatorname{Diag}(t_X \delta y))Z^{-1}$. This situation almost never arose in practice. But an extra safeguard was added to ensure that X remain positive definite.

After finding these maximum steplengths, we then multiply the step length by .90 to guarantee that we do not get too close to the boundary. Because of the nonlinearity in algorithm 1, we then take one steplength for both variables; while algorithm 2 uses different steplengths in the primal and dual variables.

From numerical tests, it appears that the step length 1 for X is usually not too large, i.e. it does not usually lose positive definiteness. But, the step length for Z is much too large immediately after μ is decreased. When μ is not decreased, the step length 1 seems good for both variables.

An efficient line search could be done using a Lanczos type algorithm to calculate the smallest eigenvalues of $C_y - \alpha \delta y$, $C_y + \beta \delta y$. Or inverse iteration can be used so the matrices from the Cholesky factorization do not get inverted.

4.3 THE ALGORITHM

The algorithms presented above differ only in the optimality conditions that Newton's method solves. We present the following outline for the algorithm 1:

Algorithm 4.1 *Suppose that C^b is given. Find initial estimate y so that $Z = C - \text{Diag}(y)$ is positive definite and well conditioned. Find an initial μ and set $X = \mu Z^{-1}$ so that X is well conditioned, e.g. set $\mu = \frac{\|Z\|}{\|Z^{-1}\|}$. Repeat the following steps.*

1. *If the convergence criteria is satisfied, e.g. if $\mu <$ given tolerance, then:*

    (a) *If X is not positive definite, then recalculate it using (4.12) with $t_X \delta y$.*

    (b) *Recalculate $\mu = \max\{\frac{\text{trace}(XZ)}{10n}, \frac{\|e - \text{diag}(x_y x_y^t) - \text{diag}(X)\|}{10n}\}$, to ensure primal feasibility.*

    (c) *STOP if the convergence criteria, μ ¡ given tolerance, is still satisfied.*

2. *Solve the Newton system for the directions $\delta y, \delta Z, \delta X$.*

3. *Find a step length that preserves positive definiteness for the primal and dual variables X, Z, and update the variables.*

4. *Check the error in the current Newton system with the current μ. If this error has decreased from the last iteration, then update $\mu = \frac{\text{trace}(XZ)}{10n}$.*

5   CONCLUSION

We have derived two primal-dual algorithms for maximizing the minimum eigenvalue of a diagonally perturbed symmetric matrix. Both algorithms have been tested extensively. Test results for the second algorithm are given in [14]. The results for the first algorithm, which is based on the trust region subproblem, were similar. Therefore, if the matrix has a dense row, the firt algorithm should be used since the dense row can be eliminated.

References

[1] F. ALIZADEH. Combinatorial optimization with semidefinite matrices. In *Proceedings of the Second Annual Integer Programming and Combinatorial Optimization Conference*, Carnegie-Mellon University, 1992.

[2] J. CULLUM, W.E. DONATH, and P.WOLFE. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. *Mathematical Programming Study*, 3:35–55, 1975.

[3] J. FALKNER, F. RENDL, and H. WOLKOWICZ. A computational study of graph partitioning. Technical Report CORR, Department of Combinatorics and Optimization, Waterloo, Ont, 1992. Submitted to Math. Progr.

[4] D.M. GAY. Computing optimal locally constrained steps. *SIAM J. Sci. Statist. Comput.*, 2:186–197, 1981.

[5] R. HORN and C. JOHNSON. *Matrix Analysis*. Cambridge University Press, New York, 1985.

[6] R. HORN and C. JOHNSON. *Topics in Matrix Analysis*. Cambridge University Press, New York, 1991.

[7] A. P. KAMATH and N. K. KARMARKAR. A continuous method for computing bounds in integer quadratic optimization problems. *Journal of Global Optimization*, 2(3):229–241, 1992.

[8] Y. E. NESTEROV and A. S. NEMIROVSKY. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1993.

[9] D. OUELLETTE. Schur complements and statistics. *Linear Algebra and Its Applications*, 36:187–295, 1981.

[10] M.L. OVERTON. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM J. Matrix Analysis and Applications*, 9:256–268, 1988.

[11] M.L. OVERTON. Large-scale optimization of eigenvalues. *SIAM J. Optimization*, 2:88–120, 1992.

[12] M.L. OVERTON and R.S. WOMERSLEY. Second derivatives for optimizing eigenvalues of symmetric matrices. Technical Report 627, Computer Science Department, NYU, 1993.

[13] S. POLJAK and H. WOLKOWICZ. Convex relaxations of 0-1 quadratic programming. Technical Report 93-18, DIMACS, 1993.

[14] F. RENDL, R. J. VANDERBEI, and H. WOLKOWICZ. A primal-dual interior point method for the max-min eigenvalue problem. Technical Report CORR 93-20, SOR 93-15, Department of Combinatorics and Optimization, Waterloo, Ont, 1993.

[15] D.C. SORENSEN. Trust region methods for unconstrained minimization. In M.J.D. Powell, editor, *Nonlinear Optimization 1981*. Academic Press, London, 1982.

[16] R.J. STERN and H. WOLKOWICZ. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. Technical Report CORR 92-38, University of Waterloo, Waterloo, Canada, 1992.

[17] L. VANDENBERGHE and S. BOYD. Primal-dual potential reduction method for problems involving matrix inequalities. Technical report, Electrical Engineering Department, Stanford University, Stanford, CA 94305, 1993.

[18] H. WOLKOWICZ. Some applications of optimization in matrix theory. *Linear Algebra and its Applications*, 40:101–118, 1981.

[19] H. WOLKOWICZ. Explicit solutions for interval semidefinite linear programming. Research Report CORR 93-29, University of Waterloo, 1993.

[20] H. WOLKOWICZ and G.P.H. STYAN. Bounds for eigenvalues using traces. *Linear Algebra and its Applications*, 29:471–506, 1980.

[21] B. YANG and R. J. VANDERBEI. The simplest semidefinite programs are trivial. Technical Report SOR-93-??, Program in Statistics & Operations Research, Princeton University, Princeton, NJ, 1993.