Complexity of an Infeasible Interior-Point Algorithm for Finding an Approximate Solution of a Semi-Definite Program with no Regularity Assumption

Robert M. Freund

Revised, December, 1995

Abstract. A semi-definite program (SDP) is an optimization problem of the form minimize $C \bullet X$ subject to $A_i \bullet X = b_i$, $i = 1, \ldots, m$, where the variable X is a matrix in R^{nxn} and X is restricted to be a symmetric and positive semi-definite matrix ($X \succeq 0$), and " $\bullet$ " denotes the inner product on matrices. When an SDP problem satisfies a regularity condition (akin to the Slater condition) on the primal and the dual, it is well-known that interior-point methodologies from linear programming can be extended to yield theoretically and practically efficient algorithms for SDP. However, a non-regular SDP problem can be poorly behaved, even having a finite duality gap whether or not the primal and/or the dual program attain their optima. We examine the complexity of finding approximate solutions to instances of SDP, even in the possible absence of any regularity condition, by an algorithm that is an extension and generalization of infeasible-interior-point path-following algorithms. Our main results, Theorem 6.1 and Theorem 6.2, give complexity bounds on the number of Newton steps needed by the algorithm to find an approximate solution to an instance of SDP, even with no regularity assumption. These complexity bounds are subject to the interpretation of the sense of approximation, when the problem and/or its dual attain or do not attain their optimal values, and when there is a zero or a positive duality gap. The bounds in these theorems depend on the desired feasibility and optimality tolerances, the initial feasibility and optimality gaps, the size n of the variable matrix X, and two (relative) condition numbers δ_1 and δ_2. The condition numbers δ_1 and δ_2 depend on the problem instance and the desired optimality tolerance, and are measured using a norm induced by the starting point of the algorithm, and so are relative to the initial starting point. When a regularity condition is satisfied, then the algorithm and the bounds obtained specialize to the best complexity bounds known for (well-behaved) instances of SDP.

Key words: semi-definite program, positive semi-definite matrix, interior-point method, condition number, Newton method, linear program.

1. Introduction

This study is concerned with the efficiency of computing an approximate solution of a semi-definite program (SDP):

$$\text{P:} \qquad \underset{X}{\text{minimize}} \quad C \bullet X$$

$$\text{s.t.} \quad A_i \bullet X = b_i \,, \quad i = 1, \ldots, m$$

$$X \succeq 0 \,,$$

where C, A_i ($i = 1, \ldots, m$), and X are nxn matrices, X is symmetric, the " $\bullet$ " operation is the inner product on matrices (also, $C \bullet X = \text{trace}(CX)$), and " $\succeq$ " denotes the Löwner partial order, i.e., $X \succeq Y$ ($X \succ Y$) denotes that $X - Y$ is symmetric and positive semi-definite (positive definite), and where (P) is not necessarily assumed to satisfy any regularity condition. Let z^*_P denote the optimal value of (P), and note that without any assumptions, it is possible that $z^*_P = +\infty$, $z^*_P = -\infty$, and that z^*_P might not be attained even if z^*_P is finite. The problem (P) is seen as a generalization of the linear programming problem and has important applications in (smooth) constrained optimization (see Alizadeh [1], [2], [3], Nesterov and Nemirovskii [21], Jarre [15]), combinatorial optimization (see Alizadeh [3], Delorme and Poljak [12], Goemans and Williamson [14], and Lovasz and Schrijver [17], for example), and in systems and control theory (see [11], for example).

Recently, much attention has been focused on the use of interior-point methods for solving the SDP problem, spurred by the work of Alizadeh [1], [2], [3], and by Nesterov and Nemirovskii [21], who independently demonstrated the theoretical effectiveness of interior-point methods for SDP, and the practical use of interior-point methods has been explored in Alizadeh et. al. [4], [5], Boyd and El Ghaoui [10], Rendl et. al. [23], [24], and Vandenberghe and Boyd [27], [28], for example. Indeed, if (P) and its dual satisfy a regularity condition akin to a Slater condition, then most of the methodology and complexity results stemming from the application of interior-point methods to linear programming carry over and are applicable with suitable minor modification to the SDP problem, see for example, [1], [2], [3], [21] cited above, plus Kojima et. al. [19], Jarre [15], Boyd and El Ghaoui [10], and Vandenberghe and Boyd [27], [28]. However, all of these applications of interior-point methods rely on a suitable regularity condition on (P), namely that both (P) and its dual (see below) each have strictly positive definite feasible solutions, or a stronger assumption. The contribution of this study is the development of an algorithm and associated complexity analysis for the efficient approximate solution of (P) in the possible absence of any regularity condition. The algorithm is given in Section 4, and the complexity results are presented in Theorem 6.1 and Theorem 6.2 of Section 6. The interpretation of these complexity results in the absence of a regularity condition on (P) is also presented in Section 6 immediately following the statements of the two theorems.

By constructing a suitable Lagrangian function, one obtains the following dual program (D) of (P):

$$\text{D:} \qquad \underset{y,\,S}{\text{maximize}} \quad b^T y$$

$$\text{s.t.} \quad \sum_{i=1}^{m} y_i A_i + S = C$$

$$y \in R^m, \; S \succeq 0,$$

see, for example, Alizadeh [3]. If X and (y, S) are feasible for the primal (P) and the dual (D), respectively, then the duality gap is seen to be $C \bullet X - b^t y = X \bullet S \geq 0$, see A.6 in the Appendix. Let z^*_D denote the optimal value of (D). Under the regularity condition that the primal has a positive definite feasible solution $X \succ 0$ and the dual has a feasible solution (y, S) with $S \succ 0$, then it is well-known that both the primal and dual problem attain their optima and that there is no duality gap, i.e., $z^*_D = z^*_P$, see Alizadeh [3] for example. In the absence of any regularity condition, the optimal duality gap can be positive and the primal and/or the dual might not attain their optima. For example, consider the following SDP instance:

$$\text{P1:} \qquad \underset{X}{\text{minimize}} \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \bullet X$$

$$\text{s.t.} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \bullet X = 2$$

$$X \succeq 0,$$

The dual of P1 is:

D1: maximize $2y$

s.t. $\begin{bmatrix} 1 & -y \\ -y & 0 \end{bmatrix} \succeq 0 .$

Note that the optimal value of P1 is $z^*_{P1} = 0$, but that this value is not attained ($x_{11} = \varepsilon,\ x_{12} = x_{21} = 1,\ x_{22} = \frac{1}{\varepsilon}$ is feasible for any $\varepsilon > 0$). However, the optimal value of D1 is $z^*_{D1} = 0$ and is attained (just set $y = 0$).

A second example is:

P2: $\underset{X}{\text{minimize}} \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet X$

s.t. $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \bullet X = 0$

$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \bullet X = 0$

$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \bullet X = 0$

$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix} \bullet X = 10$

$X \succeq 0 .$

The dual of P2 then is

$$\underset{y_1, y_2, y_3, y_4}{\text{maximize}} \quad 10\, y_4$$

$$\text{s.t.} \quad \begin{bmatrix} 0 & 1 + y_4 & -y_2 \\ 1 + y_4 & -y_1 & -y_3 \\ -y_2 & -y_3 & -2y_4 \end{bmatrix} \succeq 0 \ .$$

It is straightforward to verify that P2 has an optimal value of $z^*_{P2} = 0$, attained at

$$X^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

and that D2 has an optimal value of $z^*_{D2} = -10$, attained at

$$y_1^* = 0, \ y_2^* = 0, \ y_3^* = 0, \text{ and } y_4^* = -1 \ .$$

Therefore, P2 and D2 have a finite optimal duality gap of $z^*_{P2} - z^*_{D2} = 10$, even though they both attain their optima.

Note that problem D1 has no positive definite solution, nor do problems P2 or D2 . Borwein and Wolkowicz ([6], [7], [8], [9]) develop a dual problem for problems like (P) that yields no optimal duality gap, by constructing the dual based on the smallest face of the primal cone that contains the primal feasible region. However, this construction is not computationally tractable without prior knowledge of the structure of the relative interior of the primal feasible region, and so it is not pursued herein.

Also, we point out that it is possible to have an SDP instance such that the primal feasible region contains only irrational points, even if all of the data for the SDP is rational. Delorme and Poljak [12] construct an SDP that is a relaxation of the problem MAXSAT, and they show that for the 5-cycle, that the optimal value of the SDP is $\frac{32}{25 + 5\sqrt{5}}$. By combining their primal and dual problems into one problem with a constraint enforcing a zero duality gap, one easily constructs a feasible region based on rational data that contains no rational points.

Given the possibility that an instance of SDP may not attain its optimum, may exhibit a finite duality gap, and/or may have no rational feasible points, we are motivated to seek an approximate solution to (P) and (D), i.e., a solution X that is almost feasible for (P), a solution (y, S) that is almost feasible for (D), and that has a duality gap $X \bullet S$ that is no greater than a given optimality tolerance ε^*. This paper develops an algorithm, based on Newton's method, for computing such an approximately feasible and approximately optimal solution to (P) and (D), and also develops a complexity analysis of the algorithm. The main complexity results, Theorem 6.1 and Theorem 6.2, give upper bounds on the number of Newton steps needed by the algorithm to find approximate feasible and approximately optimal solutions to an instance of SDP, even with no regularity assumption. However, because the algorithm computes only approximate solutions, these complexity bounds are subject to interpretation. Remarks concerning the interpretation of the bounds to cases when (P) and/or (D) attain or do not attain their optimal values, and when (P) and (D) have a zero or a positive duality gap, are presented immediately following Theorem 6.1 and Theorem 6.2. The bounds in these theorems depend on the desired feasibility and optimality tolerances, the initial feasibility and optimality gaps, the size n of the variable matrix X, and two (relative) condition numbers δ_1 and δ_2. The first condition number, δ_1, measures the distance from the starting point to the set of feasible solutions, roughly speaking, using norms induced by the starting point. The second condition number, δ_2, measures the minimum norm of approximately-optimal solutions, roughly speaking, again using norms induced by the starting point. When a regularity condition is satisfied, then the algorithm and the bounds obtained specialize to the best complexity bounds known for (well-behaved) instances of SDP.

The notation used in this paper is presented below. In Section 2, we define a parametric family of SDP problems and a parametric family of SDP barrier problems, that will be (approximately) solved parametrically to obtain an approximate solution to (P) and (D). Basic properties of this parametric family are presented as well. Section 3 presents a Newton method for the barrier problem, and derives properties of the Newton direction. Section 4 contains the algorithm, based on Newton's method, for finding an approximate solution to the problem (P) and its dual (D). This algorithm can be viewed as an extension (with certain modifications) of the algorithm developed in [13]. In Section 5, we define two relative condition numbers, δ_1 and δ_2, that are important in the complexity analysis. The condition numbers δ_1 and δ_2 depend on the problem instance and the optimality tolerance, and are measured using a norm induced by the starting point of the algorithm, and so are relative to the initial starting point. Interpretations of these two condition numbers are also given in Section 5. Section 6 contains the main results of the paper, in Theorem 6.1 and Theorem 6.2. Theorem 6.1 presents an upper bound on the number of iterations (Newton steps) of the algorithm needed to obtain an approximately feasible solution to (P) and (D). Theorem 6.2 presents an upper bound on the number of iterations of the algorithm needed to obtain an approximately optimal solution to (P) and (D). The interpretation of these bounds, in light of the sense of approximation of solutions and in cases where (P) and/or (D) are not well-behaved, is also presented in Section 6. These bounds depend on the desired feasibility and optimality tolerances, the initial feasibility and optimality gaps, the size n of the variable matrix X, and the two (relative) condition numbers δ_1 and δ_2. Section 7 contains proofs of some of the results, and Section 8 contains concluding remarks. Finally, the Appendix reviews some basic properties of traces of matrices and norms of matrices.

Acknowledgment. I would like Masakazu Kojima, Dimitris Bertsimas, Florian Jarre, and James Renegar, for discussions that have contributed to the results contained herein. I would like to thank an anonymous referee for comments that have considerably improved the paper, and for suggesting the use of Proposition 2.1.

Notation: Let $R^{n \times n}$ ($S^{n \times n}$) denote the set of real (real and symmetric) $n \times n$ matrices, and let I denote the identity matrix. If M is a square ($n \times n$) matrix, the trace of M is denoted as $\operatorname{tr}(M) = \sum_{i=1}^{n} M_{ii}$, and $\operatorname{tr}(AB)$ is often denoted as $A \bullet B$. The Frobenius norm of M is defined as $\|M\| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij}^2} = \sqrt{\operatorname{tr}(M^T M)}$, and the p-norm of the matrix M (considered as an n^2-vector) is $\|M\|_p = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |M_{ij}|^p \right)^{1/p}$, for $p \in [1, \infty]$. Thus $\|M\| = \|M\|_2$. The spectral norm of M is denoted $|M| = \max \{ \|Mx\| \mid \|x\| \le 1 \}$ (where $\|x\|$ is the Euclidean norm of x). When M is a symmetric matrix, then $|M| = \max_i \{ |\lambda_i(M)| \}$, where $\lambda_1(M), \ldots, \lambda_n(M)$ are the n ordered eigenvalues of M, i.e., $\lambda_1(M) \ge \ldots \ge \lambda_n(M)$. Let $\lambda(M)$ denote the vector of ordered eigenvalues of M, i.e., $\lambda(M) = (\lambda_1(M), \ldots, \lambda_n(M))^T$. Let $S_{+}^{n \times n}$ ($S_{++}^{n \times n}$) denote the set of real symmetric positive semi-definite (positive-definite) matrices in $R^{n \times n}$. We denote the Löwner partial ordering on ($S^{n \times n}$) by "$\succeq$", and we write $A \succeq B$ if and only if $A - B \in S_{+}^{n \times n}$, $A \succ B$ if and only if $A - B \in S_{++}^{n \times n}$. In particular, if $M \in S_{+}^{n \times n}$, then $\lambda(M) \geq 0$ and $|M| = \lambda_1(M)$.

If $v \in R^n$ is a vector, then $||v||_p$ is the usual p-norm, i.e., $||v||_p = \left(\sum_{j=1}^{n} |v_j|^p \right)^{1/p}$, for $p \in [1, \infty]$. The n-vector of ones is denoted by e.

If $X \succ 0$ is given, X can be used to induce a norm over $S^{n \times n}$ as follows. Let V be obtained as a factor of X, i.e., $X = VV^T$, then define $||M||_X = ||V^T M V||$. In Section 5 it is shown that this norm is well-defined, i.e., if $X = VV^T = NN^T$, then $||V^T M V|| = ||N^T M N||$, and so $||M||_X$ does not depend on the choice of factorization of X.

2. Parametric Families of SDP Barrier Problems

Consider a semi-definite program (SDP) of the form

$$
\begin{aligned}
P: \quad & \underset{X}{\text{minimize}} \quad C \bullet X \\
& \text{s.t.} \quad A_i \bullet X = b_i, \quad i = 1, \ldots, m \qquad (2.1a) \\
& \qquad\quad X \succeq 0,
\end{aligned}
$$

where the data for P is the array $(A_1, \ldots, A_m, b, C)$ which satisfies $A_1, \ldots, A_m, C \in S^{n \times n}$, $b \in R^m$. (Note that there is no loss of generality in assuming that $A_1, \ldots, A_m, C$ are symmetric matrices.) An easily derived dual of P is given by:

$$\text{D:} \quad \underset{y,\,S}{\text{maximize}} \quad b^T y$$

$$\text{s.t.} \quad \sum_{i=1}^{m} y_i A_i + S = C \tag{2.1b}$$

$$y \in R^m, \ S \succeq 0,$$

see Alizadeh [3], for example. Let z^*_P and z^*_D denote the optimal values of P and D, respectively.

For ease of notation, we denote by $A \bullet X$ the m-vector $(A_1 \bullet X, \ldots, A_m \bullet X)^T$. We also denote by $y A$ the matrix $\sum_{i=1}^{m} y_i A_i$. One easily derives that if X and (y, S) are feasible solutions of P and D, then the duality gap $C \bullet X - b^T y$ satisfies $C \bullet X - b^T y = X \bullet S \geq 0$, by A.6. In the absence of a regularity condition on P and/or D, there is no guarantee that P and/or D will be feasible, will attain its optimum if it is feasible, or that the duality gap $z^*_P - z^*_D$ will be zero.

Given a feasibility tolerance $\varepsilon^f > 0$ and an optimality tolerance $\varepsilon^* > 0$, a triplet (X, y, S) is an $(\varepsilon^f, \varepsilon^*)$-solution of P and D if (X, y, S) satisfies:

$$\|A \bullet X - b\|_2 \leq \varepsilon^f \tag{2.2a}$$

$$\|y A + S - C\| \leq \varepsilon^f \tag{2.2b}$$

$$C \bullet X - b^T y \leq \varepsilon^*. \tag{2.2c}$$

The triplet (X, y, S) is an ε^f-feasible solution of P and D if (X, y, S) satisfies (2.2a) and (2.2b), and is called an ε^*-optimal solution if (X, y, S) satisfies (2.2c).

Let $\left(\hat{X}, \hat{y}, \hat{S}\right)$ be a starting point that will be used to initiate the algorithm, that satisfies:

$$\hat{X} \succ 0 \tag{2.3a}$$

$$\hat{S} \succ 0 \tag{2.3b}$$

$$\hat{X}\hat{S} = \hat{\theta} I \text{ for some } \hat{\theta} > 0 . \tag{2.3c}$$

Conditions (2.3a) and (2.3b) simply state that $\hat{X}$ and $\hat{S}$ are positive definite (symmetric) matrices, and condition (2.3c) is an analog of the "centering condition" typically used in interior-point methods for linear programming. Note that one obvious way to generate $\left(\hat{X}, \hat{y}, \hat{S}\right)$ is to set $\hat{X} = \hat{S} = I, \ \hat{y} = 0$.

The starting point $\left(\hat{X}, \hat{y}, \hat{S}\right)$ can be used to create parametric primal and dual family of SDPs:

$$P(\varepsilon): \quad \underset{X}{\text{minimize}} \quad \left(C + \varepsilon\left[\hat{y}A + \hat{S} - C\right]\right) \bullet X$$

$$\text{s.t.} \quad A \bullet X = b + \varepsilon[A \bullet \hat{X} - b]$$

$$X \succeq 0 ,$$

and

$$D(\varepsilon): \quad \underset{y,\,S}{\text{maximize}} \quad \left(b + \varepsilon\left[A \bullet \hat{X} - b\right]\right)^T y$$

$$\text{s.t.} \quad \sum_{i=1}^{m} y_i A_i + S = C + \varepsilon\left[\hat{y}A + \hat{S} - C\right]$$

$$S \succeq 0 .$$

At $\varepsilon = 1$, $(X, y, S) = \left(\hat{X}, \hat{y}, \hat{S}\right)$ is feasible for $P(\varepsilon)$, $D(\varepsilon)$. At $\varepsilon = 0$, $P(\varepsilon)$ is P and $D(\varepsilon)$ is D.

Let $z^*_P(\varepsilon)$ and $z^*_D(\varepsilon)$ denote the optimal values of $P(\varepsilon)$ and $D(\varepsilon)$, respectively.

Proposition 2.1. If P and D both have feasible solutions, then $P(\varepsilon)$ and $D(\varepsilon)$ both have feasible solutions, and $z_P^*(\varepsilon) = z_D^*(\varepsilon)$, for all $\varepsilon > 0$. Furthermore, $z_D^* \le \lim_{\varepsilon \to 0^+} \inf z_D^*(\varepsilon) \le \lim_{\varepsilon \to 0^+} \sup z_P^*(\varepsilon) \le z_P^*$.

Proof: Because both P and D have feasible solutions, z_P^* and z_D^* are finite. For any $\delta > 0$, there exists $\bar{X}$ and $(\bar{y}, \bar{S})$ feasible for P and D, respectively, such that $C \bullet \bar{X} \le z_P^* + \delta$ and $b^T\bar{y} \ge z_D^* - \delta$. Then for any $\varepsilon > 0$, $X_\varepsilon = \varepsilon\hat{X} + (1-\varepsilon)\bar{X}$ is feasible for $P(\varepsilon)$ and $(y_\varepsilon, S_\varepsilon) = (\varepsilon\hat{y} + (1-\varepsilon)\bar{y}, \varepsilon\hat{S} + (1-\varepsilon)\bar{S})$ is feasible for $D(\varepsilon)$. Since $\hat{X} \succ 0$ and $\hat{S} \succ 0$, $X_\varepsilon \succ 0$ and $S_\varepsilon \succ 0$, whereby $z_P^*(\varepsilon) = z_D^*(\varepsilon)$, see Alizadeh [3] for example. Also note that

$$z_P^*(\varepsilon) \le \left[C + \varepsilon\left(\hat{y}A + \hat{S} - C\right)\right] \bullet X_\varepsilon$$

$$= (1-\varepsilon)C \bullet \left[\varepsilon\hat{X} + (1-\varepsilon)\bar{X}\right] + \varepsilon\left[\hat{y}A + \hat{S}\right] \bullet \left(\varepsilon\hat{X} + (1-\varepsilon)\bar{X}\right)$$

$$= (1-\varepsilon)^2 C \bullet \bar{X} + \varepsilon(1-\varepsilon)C \bullet \hat{X} + \varepsilon\left[\hat{y}A + \hat{S}\right] \bullet \left(\varepsilon\hat{X} + (1-\varepsilon)\bar{X}\right)$$

$$\le (1-\varepsilon)^2\left(z_P^* + \delta\right) + \varepsilon(1-\varepsilon)C \bullet \hat{X} + \varepsilon\left[\hat{y}A + \hat{S}\right] \bullet \left(\varepsilon\hat{X} + (1-\varepsilon)\bar{X}\right).$$

Therefore $\lim_{\varepsilon \to 0^+} \sup z_P^*(\varepsilon) \le z_P^* + \delta$. As this is true for any $\delta > 0$, it is also true for $\delta = 0$. A similar argument for the dual problem completes the proof. ■

Given $(\hat{X}, \hat{y}, \hat{S})$ satisfying (2.3), the parametric logarithmic barrier problem induced by $(\hat{X}, \hat{y}, \hat{S})$ is defined as:

$$\begin{aligned} BP(\omega, \varepsilon): \quad & \underset{X}{\text{minimize}} \quad \left(C + \varepsilon\left[\hat{y}A + \hat{S} - C\right]\right) \bullet X - \omega\, ln\,(\det(X)) \\ & \text{s.t.} \quad A \bullet X = b + \varepsilon[A \bullet \hat{X} - b] \qquad (2.4) \\ & \qquad\quad X \succ 0, \end{aligned}$$

where $\omega > 0$ is a barrier parameter and ε controls the extent of infeasibility of X in the original problem P. The matrix X solves $BP(\omega, \varepsilon)$ if and only if there exists (y, S) that satisfy:

$$A \bullet X = b + \varepsilon[A \bullet \hat{X} - b], \; X \succeq 0 \tag{2.5a}$$

$$yA + S = C + \varepsilon[\hat{y}A + \hat{S} - C], \; S \succeq 0 \tag{2.5b}$$

$$XS = \omega I. \tag{2.5c}$$

Note that for $\varepsilon = 0$ and $\omega = 0$, then a solution (X, y, S) of (2.5) solves P and D, since (2.5a) and (2.5b) ensure primal and dual feasibility and (2.5c) ensures a zero duality gap. We also remark that from (2.3) we have that $(X, y, S, \omega, \varepsilon) = (\hat{X}, \hat{y}, \hat{S}, \hat{\theta}, 1)$ satisfies (2.5). Given $(X, y, S, \omega, \varepsilon) = (\hat{X}, \hat{y}, \hat{S}, \hat{\theta}, 1)$ as a starting point, we would like to solve (2.5) parametrically in ε and ω as $\varepsilon \to 0$ and $\omega \to 0$. However, since (2.5c) is a nonlinear equation in X and S, we cannot in general solve (2.5) exactly. Instead we will consider the following conditions, which define a γ–approximate solution to $BP(\omega, \varepsilon)$ for a given constant γ (we will typically use $\gamma = \frac{1}{4}$):

$$A \bullet X = b + \varepsilon[A \bullet \hat{X} - b], \; X \succ 0 \tag{2.6a}$$

$$yA + S = C + \varepsilon[\hat{y}A + \hat{S} - C] \tag{2.6b}$$

$$\left\| I - \left(\tfrac{1}{\omega}\right) V^T S V \right\| \le \gamma, \text{ where } X = VV^T. \tag{2.6c}$$

Note that (2.6a) and (2.6b) correspond to (2.5a) and (2.5b). However, (2.6c) is different. If X is factorized into $X = VV^T$, then note that (2.6c) is a weaker condition than (2.5c), since if $XS = \omega I$, then $VV^T S = \omega I$, and rearranging yields $I - \left(\frac{1}{\omega}\right) V^T SV = 0$. Also note from A.13 that (2.6c) is satisfied or not independent of the factorization, i.e., $\left\| I - \left(\frac{1}{\omega}\right) V^T S V \right\| = \left\| I - \left(\frac{1}{\omega}\right) N^T S N \right\|$ wherever $VV^T = NN^T$. This being the case we observe:

<u>Proposition 2.2</u>. For $\gamma \ge 0$, $(X, y, S, \theta, \varepsilon) = (\hat{X}, \hat{y}, \hat{S}, \hat{\theta}, 1)$ is a γ–approximate solution of $BP(\omega, \varepsilon) = BP(\hat{\theta}, 1)$. ■

We also have:

Proposition 2.3. If (X, y, S) is a γ-approximate solution of $BP(\omega, \varepsilon)$ and $\varepsilon \geq 0$, $\omega > 0$, $\gamma < 1$, then

(i) X is a feasible solution of $P(\varepsilon)$

(ii) (y, S) is a feasible solution of $D(\varepsilon)$

(iii) the duality gap satisfies $\omega(n - \sqrt{n}\gamma) \leq X \bullet S \leq \omega(n + \sqrt{n}\gamma)$.

Proof: (i) and (ii) are apparent from (2.6a) and (2.6b) since (2.6c) implies that $S \succ 0$. To prove (iii), let $R = I - \frac{1}{\omega} V^T S V$ in (2.6c). Then $\|R\| \leq \gamma$. We have $X \bullet S = \mathrm{tr}(XS)$ $= \mathrm{tr}(V V^T S) = \mathrm{tr}(V^T S V) = \omega\, \mathrm{tr}(I - R) = n\omega - \omega\, \mathrm{tr}(R)$. However, $|\mathrm{tr}(R)| = \left| \sum_{i=1}^{n} R_{ii} \right| \leq \sqrt{n} \sqrt{\sum_{i=1}^{n} R_{ii}^2} \leq \sqrt{n}\, \|R\| \leq \sqrt{n}\gamma$, completing the proof. ■

Motivated by Proposition 2.1 and Proposition 2.3, we would like to solve for a γ-approximate solution of $BP(\omega, \varepsilon)$ for a sequence of values of $\varepsilon \to 0$ and $\omega \to 0$.

3. Newton's Method

Suppose that $(\bar{X}, \bar{y}, \bar{S})$ is a given γ-approximate solution to $BP(\bar{\omega}, \bar{\varepsilon})$ for given $\bar{\varepsilon} \geq 0$, $\bar{\omega} \geq 0$, and we want to obtain a γ-approximate solution (X, y, S) to $BP(\beta\bar{\omega}, \alpha\bar{\varepsilon})$ for some given values of α and β that are used to update $\varepsilon = \alpha\bar{\varepsilon}$ and $\omega = \beta\bar{\omega}$. Here we describe a Newton method for this problem.

We start with a "re-scaling" of the data that generalizes the re-scaling used commonly in interior-point methods. Using the ideas in Alizadeh [3], we first factorize:

$$\bar{X} = \bar{V}\bar{V}^T, \tag{3.1a}$$

and define:

$$\bar{A}_i = \bar{V}^T A_i \bar{V}, \quad i = 1, \ldots, m, \tag{3.1b}$$

$$\overline{C} \quad = \quad \overline{V}^T C \overline{V}. \tag{3.1c}$$

Then, keeping in mind that $\varepsilon = \alpha \bar{\varepsilon}$ and $\omega = \beta \overline{\omega}$ for given values of α and β, we define the following Newton equations to obtain a Newton direction $D \in S^{n \times n}$ and dual multipliers $y \in R^m$.

$$\overline{A}_i \bullet D = (\alpha - 1)\, \bar{\varepsilon} \left(A_i \bullet \widehat{X} - b_i \right), \; i = 1, \ldots, m \;, \tag{3.2a}$$

$$\sum_{i=1}^{m} y_i \overline{A}_i + \beta \overline{\omega} I - \beta \overline{\omega} D = \overline{C} + \alpha\, \bar{\varepsilon} \left(\hat{y} \overline{A} + \overline{V}^T \widehat{S} \overline{V} - \overline{C} \right). \tag{3.2b}$$

(The equations (3.2) are derived by re-scaling $BP(\omega, \varepsilon)$ using (3.1) and writing down the optimality conditions for the quadratic approximate of $BP(\omega, \varepsilon)$ at $X = \overline{X}$ and $(\omega, \varepsilon) = (\beta \overline{\omega}, \alpha \bar{\varepsilon})$. Therefore, equations (3.2) will always have a unique solution in D, and will have a unique solution in y if the matrices $A_1, \ldots, A_m$ are independent, i.e., if there is no nontrivial solution $v_1, \ldots, v_m$ of $\sum_{i=1}^{m} v_i A_i = 0$.)

Note from (3.2b) that D will be a symmetric matrix because $\overline{A}_1, \ldots, \overline{A}_m, \widehat{S}, \overline{C}, I$ are symmetric matrices. Note also that the Newton equations (3.2) use only primal re-scaling. Primal-dual re-scaling (see Alizadeh et. al. [4], [5], Kojima et. al. [19]) may be preferable in practice, and this issue is discussed in the concluding remarks in Section 8.

Because D is a Newton direction for the "re-scaled" problem, we will define the point induced by the Newton direction D to be:

$$X \quad = \quad \overline{V}[I + D]\overline{V}^T \tag{3.3a}$$

for the primal and

$$S \quad = \quad \beta \overline{\omega}\, \overline{V}^{-T} [I - D] \overline{V}^{-1}, \tag{3.3b}$$

for the dual, with y given from the solution to (3.2).

We have the following Lemma concerning this method.

Lemma 3.1. Suppose that $(\overline{X}, \overline{y}, \overline{S})$ is a γ–approximate solution of BP$(\overline{\omega}, \overline{\varepsilon})$ for $\gamma < 1$ and that α and β are given scalars. Let (D, y) be the solution of the Newton equations (3.2), and let (X, S) be defined from (3.3.) Then, if $\|D\| \le \sqrt{\gamma}$, (X, y, S) is a γ–approximate solution of BP$(\beta\overline{\omega}, \alpha\overline{\varepsilon})$.

Proof: We must show that (X, y, S) satisfies (2.6) for $\varepsilon = \alpha\overline{\varepsilon}$, $\omega = \beta\overline{\omega}$. Direct substitution can be used to verify (2.6a) and (2.6b) as follows. For $i = 1, \ldots, m$,

$$
\begin{aligned}
A_i \bullet X &= A_i \bullet \left(\overline{V}\,\overline{V}^T + \overline{V}D\overline{V}^T\right) = A_i \bullet \overline{X} + A_i \bullet \overline{V}D\overline{V}^T = b_i + \overline{\varepsilon}\left(A_i \bullet \hat{X} - b_i\right) + \overline{V}^T A_i \overline{V} \bullet D \\
&= b_i + \overline{\varepsilon}\left(A_i \bullet \hat{X} - b_i\right) + \overline{A}_i \bullet D = b_i + \alpha\overline{\varepsilon}\left(A_i \bullet \hat{X} - b_i\right),
\end{aligned}
$$

using (3.3a), (3.2a),. and A.1 and A.2, showing (2.6a).

From (3.1), (3.2b), and (3.3b) we have

$$
\overline{V}^T C \overline{V} + \alpha\,\overline{\varepsilon}\left(\hat{y}\left(\overline{V}^T A \overline{V}\right) + \overline{V}^T \hat{S} \overline{V} - \overline{V}^T C \overline{V}\right) = \sum_{i=1}^{m} y_i \overline{V}^T A_i \overline{V} + \overline{V}^T S \overline{V}.
$$

Pre-multiplying and post-multiplying by $\overline{V}^{-T}$ and $\overline{V}^{-1}$, we obtain

$$
\sum_{i=1}^{m} y_i A_i + S = C + \alpha\;\overline{\varepsilon}\left(\hat{y} A + \hat{S} - C\right),
$$

showing (2.6b). Also, $|D| \le \|D\| \le \sqrt{\gamma} < 1$ (see A.7), and so $-1 < \lambda_i(D) < 1$, $i = 1, \ldots, n$, which implies that $I - D \succ 0$ and $I + D \succ 0$, which proves that $X \succ 0$ and $S \succ 0$ from (3.3).

Finally, we must show that (2.6c) holds. Because $I + D \succ 0$, we can factorize $I + D = UU^T = UU$ for some symmetric U that is a square root of $I + D$. From (3.3a) we have that $X = \overline{V}UU^T\overline{V}^T$ and so $W = \overline{V}U$ is a factor of X, i.e., $X = WW^T$. Therefore, to prove that (2.6c) holds, we will show that $\left\| I - \frac{1}{\beta\overline{\omega}} W^T S W \right\| \le \gamma$. Note from (3.3b), that $I - \frac{1}{\beta\overline{\omega}} W^T S W =$
$I - \frac{1}{\beta\overline{\omega}} U^T \overline{V}^T (\beta\overline{\omega}) \overline{V}^{-T} (I - D) \overline{V}^{-1} \overline{V} U = I - U^T (I - D) U =$

$]I - U(2I - U^2)U = I - 2U^2 + U^4 = (U^2 - I)(U^2 - I) = D^2$, because U is symmetric and so $U^T = U$. Therefore, via A.9,

$\left\| I - \left(\frac{1}{\beta \overline{\omega}}\right) W^T S W \right\| = \|D^2\| \leq \|D\|^2 \leq \gamma$, which completes the proof. ■

Lemma 3.1 states that if the current point $(\overline{X}, \overline{y}, \overline{S})$ is a γ–approximate solution of $BP(\overline{\omega}, \overline{\varepsilon})$, and that if the norm of the Newton direction D is not larger than $\sqrt{\gamma}$, then the new point (X, y, S) will be a γ–approximate solution of $BP(\beta \overline{\omega}, \alpha \overline{\varepsilon})$.

4. The Algorithm for Finding an Approximate Solution of SDP

In this section, we present an algorithm for finding an approximate solution of the SDP problem P. The data for the algorithm is the data for P $(A_1, \ldots, A_m, b, C)$, the data for the starting point $(\widehat{X}, \widehat{y}, \widehat{S})$ (assumed to satisfy 2.3), the feasibility and optimality tolerances $(\varepsilon^f, \varepsilon^*)$, and the constant γ used to define a γ–approximate solution of $BP(\omega, \varepsilon)$. We will now make the further assumption that the initial duality gap exceeds the optimality tolerance, i.e., $\widehat{X} \bullet \widehat{S} = n\widehat{\theta}$ satisfies $\widehat{X} \bullet \widehat{S} \geq \varepsilon^*$. (This assumption can be enforced simply by rescaling $\widehat{X}$ and/or $\widehat{S}$.) The algorithm appears below, with an explanation of the steps given following.

Algorithm $(A_1, \ldots, A_m, b, C, \widehat{X}, \widehat{y}, \widehat{S}, \varepsilon^f, \varepsilon^*, \gamma)$

Step 0 (Initialize) $(X^0, y^0, S^0) = (\widehat{X}, \widehat{y}, \widehat{S})$, $\varepsilon^0 = 1$, $\omega^0 = \widehat{\theta}$, $k = 0$

$$\widehat{\beta} = 1 + \frac{\sqrt{\gamma} - \gamma}{\sqrt{n} - \sqrt{\gamma}}$$

$$\varepsilon' = \min\left\{ \varepsilon^f / \|A \bullet \widehat{X} - b\|_2 \,,\, \varepsilon^f / \|\widehat{y}A + \widehat{S} - C\| \right\}$$

Step 1 (Evaluate Current Point) $(\overline{X}, \overline{y}, \overline{S}) = (X^k, y^k, S^k)$, $\overline{\varepsilon} = \varepsilon^k$, $\overline{\omega} = \omega^k$

Factor $\overline{X} = \overline{V}\,\overline{V}^T$, and compute

$$\overline{A}_i = \overline{V}^T A_i \overline{V}\ ,\ i=1,\ldots,m,\quad \overline{C} = \overline{V}^T C \overline{V}\ .$$

If $\overline{\varepsilon} \le \varepsilon'$ and $\overline{X} \bullet \overline{S} \le \varepsilon^*$, halt. (4.1)

If $\overline{\varepsilon} \le \varepsilon'$ and $\overline{X} \bullet \overline{S} > \varepsilon^*$, go to Step 2. (4.2)

If $\overline{\varepsilon} > \varepsilon'$ and $\left(||\overline{X}||_{\hat{S}} + ||\overline{S}||_{\hat{X}}\right)\left(\frac{\overline{\varepsilon}}{\overline{\omega}}\right) \le 2n\left(1 + \gamma + \hat{\beta}\right)$, go to Step 3. (4.3)

If $\overline{\varepsilon} > \varepsilon'$ and $\left(||\overline{X}||_{\hat{S}} + ||\overline{S}||_{\hat{X}}\right)\left(\frac{\overline{\varepsilon}}{\overline{\omega}}\right) > 2n\left(1 + \gamma + \hat{\beta}\right)$, go to Step 4. (4.4)

<u>Step 2 (Shrink $\overline{\omega}$)</u>. Set $\alpha = 1$. Solve (4.5) for (D, y) and the smallest value of β:

$$\overline{A}_i \bullet D = 0,\ i=1,\ldots,m\ . \tag{4.5a}$$

$$\sum_{i=1}^{m} y_i\overline{A}_i + \beta\overline{\omega} I - \beta\overline{\omega} D = \overline{C} + \overline{\varepsilon}\left(\hat{y}\overline{A} + \overline{V}^T\hat{S}\overline{V} - \overline{C}\right) \tag{4.5b}$$

$$\mathrm{tr}\left(D^2\right) \le \gamma \tag{4.5c}$$

$$\beta\overline{\omega} \ge \frac{\varepsilon^*}{n + \sqrt{n}\gamma} \tag{4.5d}$$

Go to Step 5.

<u>Step 3 (Shrink $\overline{\omega}$ and $\overline{\varepsilon}$)</u>. Solve (4.6) for (D, y) and the smallest value of δ:

$$\overline{A}_i \bullet D = (\alpha - 1)\overline{\varepsilon}\left(A_i \bullet \hat{X} - b_i\right),\ i=1,\ldots,m\ . \tag{4.6a}$$

$$\sum_{i=1}^{m} y_i\overline{A}_i + \beta\overline{\omega} I - \beta\overline{\omega} D = \overline{C} + \alpha\,\overline{\varepsilon}\left(\hat{y}\overline{A} + \overline{V}^T\hat{S}\overline{V} - \overline{C}\right). \tag{4.6b}$$

$$\mathrm{tr}\left(D^2\right) \le \gamma \tag{4.6c}$$

$$\beta = \max\left\{\delta,\ \frac{\varepsilon^*}{\overline{\omega}(n + \sqrt{n}\gamma)}\right\} \tag{4.6d}$$

$$\alpha = \max\left\{\delta,\ \varepsilon'/\overline{\varepsilon}\right\}. \tag{4.6e}$$

Go to Step 5.

Step 4 (Expand $\bar{\omega}$). Set $\alpha = 1$. Set $\beta = \hat{\beta}$. Solve (4.7) for (D, y):

$$\bar{A}_i \bullet D = 0, \; i = 1, \ldots, m \quad . \tag{4.7a}$$

$$\sum_{i=1}^{m} y_i \bar{A}_i + \beta \bar{\omega} I - \beta \bar{\omega} D = \bar{C} + \bar{\varepsilon}\left(\hat{y}\bar{A} + \bar{V}^T \hat{S} \bar{V} - \bar{C}\right) \tag{4.7b}$$

Go to Step 5.

Step 5 (Update Iterates)

$$X = \bar{V}(I + D)\bar{V}^T$$

$$S = (\beta\bar{\omega})\bar{V}^{-T}(I - D)\bar{V}^{-1}$$

$$\varepsilon = \alpha\bar{\varepsilon}$$

$$\omega = \beta\bar{\omega}$$

$$\left(X^{k+1}, y^{k+1}, S^{k+1}\right) = (X, y, S), \; \varepsilon^{k+1} = \varepsilon, \; \omega^{k+1} = \omega \; .$$

$k \leftarrow k+1$. Go to Step 1.

We now given an explanation and partial rationalization of the steps in the algorithm. In Step 0, the initial values of the algorithm are set. The constant $\hat{\beta}$ is computed, and will be explained later. The constant ε' is used to check for approximate feasibility: if $\left(\bar{X}, \bar{y}, \bar{S}\right)$ is a γ-approximate solution of BP$(\bar{\omega}, \bar{\varepsilon})$ and $\bar{\varepsilon} \leq \varepsilon'$, then from (2.6) it follows that (2.2a) and (2.2b) are satisfied, i.e., $\bar{X}$ and $\left(\bar{y}, \bar{S}\right)$ are approximately feasible for P and D, respectively.

In Step 1, the data are first "re-scaled" as in (3.1). Then conditions (4.1) through (4.4) are checked, and the algorithm proceeds differently according to which of the four conditions are satisfied. If (4.1) is satisfied, it follows that $\left(\bar{X}, \bar{y}, \bar{S}\right)$ satisfies (2.2) and so $(\overline{X}, \overline{y}, \overline{S})$ is an $(\varepsilon^f, \varepsilon^*)$-solution of P and D. In this case, the algorithm halts.

If condition (4.2) is satisfied, the algorithm proceeds to Step 2. In this case, (4.2) indicates that $(\overline{X}, \overline{y}, \overline{S})$ satisfies (2.2a) and (2.2b), and so $\overline{X}$ and $(\overline{y}, \overline{S})$ are approximately feasible for P and D, respectively. Therefore, there is no need to modify $\overline{\varepsilon}$, and so α is set to $\alpha = 1$ at Step 2. However, the duality gap $\overline{X} \bullet \overline{S}$ is still larger than ε^*, and so we would like to shrink $\overline{\omega}$ to $\beta \overline{\omega}$ for some value of $\beta < 1$ and as small as possible, and then solve the Newton equations (3.2) (which are given in Step 2 as (4.5a) and (4.5b)) to obtain the Newton direction D and dual multipliers y. From Lemma 3.1, the new iterate values will be a γ–approximate solution of $BP(\overline{\omega}, \beta \overline{\varepsilon})$ so long as $||D|| \le \sqrt{\gamma}$, or equivalency if $tr(D^2) \le \gamma$ (see A.3), and this is stipulated in (4.5c). Finally, since we only need a duality gap of ε^*, from Proposition 2.3 (iii) we can restrict β to satisfy (4.5d). We will show at the end of this section that Step 2 can be carried out efficiently, i.e., Step 2 requires hardly any more work than solving the Newton equations (4.5a) and (4.5b) and using the quadratic formula.

We now focus on (4.3) and (4.4). If neither (4.1) nor (4.2) are satisfied, then $\overline{\varepsilon} > \varepsilon'$ and we would like to shrink both $\overline{\varepsilon}$ and $\overline{\omega}$ to new values $\varepsilon = \alpha \overline{\varepsilon}$ and $\omega = \beta \overline{\omega}$ for $\alpha = \beta = \delta$ for a small fraction $\delta < 1$ that is nicely bounded away from 1. However, we are not quite able to accomplish this. Instead, it may be necessary at some iterations to leave $\overline{\varepsilon}$ as it is (i.e., set $\alpha = 1$), and instead increase $\overline{\omega}$ by the scalar quality $\hat{\beta} > 1$ (defined in Step 0). Nevertheless, we will be able to bound the number of iterations at which $\overline{\omega}$ is increased. The logic that controls whether or not the algorithm shrinks $\overline{\varepsilon}$ and $\overline{\omega}$, or instead increases $\overline{\omega}$ and leaves $\overline{\varepsilon}$ as is, is determined in (4.3) and (4.4), and will be discussed in Section 6.

If (4.3) is satisfied, the algorithm proceeds to Step 3. In this case, we would like to choose a small value of δ and set $\alpha = \delta$ and $\beta = \delta$, and then solve the Newton equation (3.2) (stated here as (4.6a) and (4.6b)). However, as there is no need to decrease β lower than $\frac{\varepsilon^*}{\overline{\omega}\,(n + \sqrt{n}\,\gamma)}$ (for then the new iterate would have duality gap less than ε^*), condition (4.6d) is introduced. Similarly, as there is no need to decrease α lower than $\varepsilon'/\overline{\varepsilon}$ (for the new iterates will satisfy (2.2a) and (2.2b)), condition (4.6e) is introduced. Also, from Lemma 3.1, we will need (4.6c) to be satisfied to ensure that the new iterate will be a γ–approximate solution of $BP(\beta \overline{\omega}, \alpha \overline{\varepsilon})$. We will show at the end of this section that Step 3 can be carried out efficiently, i.e., that Step 3 requires hardly any more work than solving the Newton equations (4.6a) and (4.6b), and using the quadratic formula.

Finally, suppose that (4.4) is satisfied. Then the algorithm will go to Step 4, where $\bar{\varepsilon}$ is left unchanged ($\alpha = 1$) and $\bar{\omega}$ is increased to $\hat{\beta}\,\bar{\omega}$, where $\hat{\beta}$ is defined in Step 0. Equations (4.7a) and (4.7b) are simply the Newton equations (3.1) for these values. We will show that if $\beta = \hat{\beta}$ then $||D|| \leq \gamma$ in (4.7), and so the new iterate values will be a γ–approximate solution of $BP\left(\bar{\varepsilon},\ \hat{\beta}\,\bar{\omega}\right)$. We will also demonstrate an upper bound on the number of iterations that visit Step 4.

Last of all, in Step 5, the iterate values are updated as in (3.3), and the algorithm continues.

We now discuss the efficiency of computations in Steps 2 and 3 of the algorithm. Let us first examine Step 2. For β fixed, the Newton equations (4.5a, b) will yield a unique value of D. Furthermore, it is easy to see from (4.5b) that D will be linear in $1/\beta$, i.e., the solution D can be written as:

$$D = \bar{D} + \bar{G}\left(\frac{1}{\beta}\right), \tag{4.8}$$

where it is straightforward from (4.5a, b) to compute $\bar{D}$ and $\bar{G}$. We then test the lower bound for β given in (4.5d) and check if $\frac{1}{\beta} = \frac{\bar{\omega}\left(n + \sqrt{n}\,\gamma\right)}{\varepsilon^*}$ yields a value of D in (4.8) for which (4.5c) is satisfied, by performing the necessary arithmetic operations. If (4.5c) is satisfied, Step 2 is complete. If not, notice from (4.8) that $\mathrm{tr}\left(D^2\right)$ is quadratic in $1/\beta$, and so we can compute the lowest value of β such that (4.5c) is satisfied by applying the quadratic formula to the quadratic equation $\mathrm{tr}\left(D^2\right) = \gamma$, which is quadratic in $1/\beta$ from (4.8). Thus Step 2 can be computed efficiently.

The same arguments used in the analysis of Step 2 also apply to Step 3, with suitable modification. For fixed α and β, the Newton equations (4.6a, b) will yield a unique value of D. Furthermore, it is easy to see from (4.6b) that D will be linear in $\left(\frac{1}{\beta}\right)$ and in $\left(\frac{\alpha}{\beta}\right)$, i.e., the solution D can be written as:

$$D = \bar{D} + \bar{G}\left(\frac{1}{\beta}\right) + \bar{H}\left(\frac{\alpha}{\beta}\right), \tag{4.9}$$

where it is straightforward to compute $\overline{D}$, $\overline{G}$, and $\overline{H}$, from (4.6a, b). Then the inequality (4.6c) can be written as a quadratic inequality in the two variables $\left(\frac{1}{\beta}\right)$ and $\left(\frac{\alpha}{\beta}\right)$, namely, (4.6c) becomes, via (4.9), the inequality:

$$\operatorname{tr}\left(\overline{D}^2\right) + \operatorname{tr}\left(\overline{G}^2\right)\left(\frac{1}{\beta}\right)^2 + \operatorname{tr}\left(\overline{H}^2\right)\left(\frac{\alpha}{\beta}\right)^2$$

$$+ 2\operatorname{tr}\left(\overline{D}\,\overline{G}\right)\left(\frac{1}{\beta}\right) + 2\operatorname{tr}\left(\overline{D}\,\overline{H}\right)\left(\frac{\alpha}{\beta}\right) + 2\operatorname{tr}\left(\overline{G}\,\overline{H}\right)\left(\frac{1}{\beta}\right)\left(\frac{\alpha}{\beta}\right) \le \gamma. \quad (4.10)$$

We first set $\alpha = \beta = \delta$, and then (4.10) is quadratic in $\left(\frac{1}{\delta}\right)$, and so we can use the quadratic formula to find the smallest value of δ satisfying (4.10). If $\delta \ge \frac{\varepsilon^*}{\overline{\omega}(n + \sqrt{n}\,\gamma)}$ and $\delta \ge \frac{\varepsilon'}{\overline{\varepsilon}}$, then $\alpha = \beta = \delta$ is minimal in Step 3 and we are done. If not, we must also check the cases $\beta = \frac{\varepsilon^*}{\overline{\omega}(n + \sqrt{n}\,\gamma)}$ and $\alpha = \delta$, (whereby (4.10) is quadratic in α) and/or $\alpha = \frac{\varepsilon'}{\overline{\varepsilon}}$ and $\beta = \delta$ (whereby (4.10) is quadratic in $\frac{1}{\beta}$), and so in all cases, Step 3 can be computed efficiently using the quadratic formula.

5. Condition Numbers Relative to the Starting Point $\left(\hat{X}, \hat{y}, \hat{S}\right)$

In this section, we define two condition numbers for the problem instance $(A_1, \ldots, A_m, b, C)$ of SDP, which will be denoted by δ_1 and δ_2. These condition numbers are defined relative to the starting point $\left(\hat{X}, \hat{y}, \hat{S}\right)$ that is used to start the algorithm. The first condition number, δ_1, is a function of the problem instance $(A_1, \ldots, A_m, b, C)$ and the starting point $\left(\hat{X}, \hat{y}, \hat{S}\right)$. The second condition number, δ_2, is a function of the problem instance $(A_1, \ldots, A_m, b, C)$, the starting point $\left(\hat{X}, \hat{y}, \hat{S}\right)$, and the optimality tolerance ε^*. In Section 6, we will give a complexity bound on the number of iterations of the algorithm that is needed to solve SDP for an $(\varepsilon^f, \varepsilon^*)$-solution of SDP, that depends only on the initial infeasibilities of $\hat{X}$ and $\left(\hat{y}, \hat{S}\right)$ for P and D, the initial duality gap $\hat{X} \bullet \hat{S}$, as well as n, ε^f, ε^*, and the condition numbers δ_1 and δ_2.

Before presenting the definitions of δ_1 and δ_2, we start with a property of factorizations of matrices.

Proposition 5.1. Let $X = VV^T = NN^T \succ 0$. Then for any matrix M, $\|V^TMV\| = \|N^TMN\|$, and $|V^TMV| = |N^TMN|$.

Proof: $\|V^TMV\|^2 = \mathrm{tr}(V^TM^TVV^TMV) = \mathrm{tr}(M^TVV^TMVV^T) = \mathrm{tr}(M^TNN^TMNN^T) = \mathrm{tr}(N^TM^TNN^TMN) = \|N^TMN\|^2$, which establishes the first equality. To prove the second equality, note that $|V^TMV| = \max\{\|V^TMVx\| \mid \|x\| \le 1\}$. Let $\bar{x}$ be a vector where this maximum is achieved, and let $\bar{y} = N^{-1}V\bar{x}$. Then $\|\bar{y}\|^2 = \bar{x}^TV^TN^{-T}N^{-1}V\bar{x} = \bar{x}^TV^TV^{-T}V^{-1}V\bar{x} = \|\bar{x}\|^2 \le 1$, and so $|N^TMN|^2 \ge \|N^TMN\bar{y}\|^2 = \bar{y}^TN^TM^TNN^TMN\bar{y} = \bar{x}^TV^TN^{-T}N^TM^TVV^TMV\bar{x} = \bar{x}^TV^TM^TVV^TMV\bar{x} = \|V^TMV\bar{x}\|^2 = |V^TMV|^2$. A parallel argument shows that $|N^TMN| \le |V^TMV|$, completing the proof. ■

Definition 5.1. Suppose $X \succ 0$. Define the (relative) Frobenius norm of a matrix $M \in S^{n\times n}$ to be $\|M\|_X = \|V^TMV\|$ where X is factorized into $X = VV^T$, and define the (relative) spectral norm to be $|M|_X = |V^TMV|$ where X is factorized into $X = VV^T$.

From Proposition 5.1, $\|M\|_X$ and $|M|_X$ are well defined.

We now define the first condition number, δ_1, as follows:

$$\delta_1 = \max\left\{ \begin{array}{ll} \min_{X} |X - \hat{X}|_{\hat{X}^{-1}}, & \min_{y,\,S} |S - \hat{S}|_{\hat{S}^{-1}} \\ \text{s.t. } A \bullet X = b & \text{s.t. } yA + S = C \\ X \in S^{n\times n} & S \in S^{n\times n} \end{array} \right\}. \tag{5.1}$$

This condition number is a generalization of a similar idea developed in [13] and in Kojima [18]. Note that δ_1 is nonnegative and is positive unless both $\hat{X}$ and $(\hat{y}, \hat{S})$ are feasible for P and D, respectively. The primal part of (5.1) measures how close $\hat{X}$ is to the affine manifold $\{X \mid A \bullet X = b\}$, measured in the spectral norm induced by $\hat{X}^{-1}$. The dual part of (5.1) has an analogous interpretation.

Remark 5.1. It is possible to ensure that $\delta_1 \le 2$ by an appropriate construction of the starting point $(\hat{X}, \hat{y}, \hat{S})$.

To see why this remark is true, consider the primal problem first, and compute any symmetric matrix X that satisfies $A \bullet X = b$, and set $\hat{X} = (\|X\|)I$. Then $|X - \hat{X}|_{\hat{X}^{-1}} = (\|X\|)^{-1}|X - (\|X\|)I| \le 1 + (\|X\|)^{-1}|X| \le 2$, since $\|X\| \ge |X|$, see A.7. Therefore the primal part of (5.1) will have objective value at most 2, and a similar construction with the dual will produce $\delta_1 \le 2$.

Let $\tilde{X}$ and $(\tilde{y}, \tilde{S})$ be optimal solutions to the problems defined in (5.1). Then

$$A \bullet \tilde{X} = b, \qquad |\tilde{X} - \hat{X}|_{\hat{X}^{-1}} \le \delta_1 \tag{5.2a}$$

$$\tilde{y}A + \tilde{S} = C, \qquad |\tilde{S} - \hat{S}|_{\hat{S}^{-1}} \le \delta_1 . \tag{5.2b}$$

A useful property of the condition number δ_1 is given below.

Proposition 5.2. Suppose $\bar{X} \succ 0$ and $\bar{S} \succ 0$. Then

(i) $$\|\hat{S} - \tilde{S}\|_{\bar{X}} \le \delta_1 \|\bar{X}\|_{\hat{S}},$$

and

(ii) $$\|\hat{X} - \tilde{X}\|_{\bar{S}} \le \delta_1 \|\bar{S}\|_{\hat{X}}.$$

Proof: Consider the factorizations $\bar{X} = \bar{V}\bar{V}^T$, $\bar{S} = \bar{U}\bar{U}^T$, $\hat{X} = \hat{V}\hat{V}^T$, $\hat{S} = \hat{U}\hat{U}^T$. Then

$\widehat{X}^{-1} = \widehat{V}^{-T}\widehat{V}^{-1}$, $\widehat{S}^{-1} = \widehat{U}^{-T}\widehat{U}^{-1}$, and

$$\|\widehat{S} - \widetilde{S}\|_{\overline{X}} = \|\overline{V}^T(\widehat{S} - \widetilde{S})\overline{V}\|$$

$$= \|\overline{V}^T\widehat{U}\widehat{U}^{-1}(\widehat{S} - \widetilde{S})\widehat{U}^{-T}\widehat{U}^T\overline{V}\|$$

$$\leq \|\overline{V}^T\widehat{U}\widehat{U}^T\overline{V}\|\,|\widehat{U}^{-1}(\widehat{S} - \widetilde{S})\widehat{U}^{-T}| \qquad \text{(from A.10)}$$

$$= \|\widehat{U}^T\overline{V}\,\overline{V}^T\widehat{U}\|\,|\widehat{S} - \widetilde{S}|_{\widehat{S}^{-1}} \qquad \text{(from A.14)}$$

$$\leq \delta_1\|\widehat{U}^T\overline{X}\widehat{U}\| \qquad \text{(from 5.2b)}$$

$$= \delta_1\|\overline{X}\|_{\widehat{S}},$$

which shows (i). A parallel argument for (ii) completes the proof. ■

The second condition number, δ_2, is defined as follows, and depends on $(A_1, \ldots, A_m, b, C)$, the starting point $(\widehat{X}, \widehat{y}, \widehat{S})$, and the optimality tolerance ε^*:

$$\delta_2 = \delta_2(\varepsilon^*) = \min_{X,\,y,\,S} \left(\frac{1}{2n}\right)\left(\widehat{X}^{-1} \bullet X + \widehat{S}^{-1} \bullet S\right) \tag{5.3}$$

$$\text{s.t.} \quad A \bullet X = b, \quad X \succeq 0$$

$$yA + S = C, \quad S \succeq 0$$

$$C \bullet X - b^T y \leq \varepsilon^*.$$

The constraints of (5.3) state that (X, y, S) must range over the values that are primal and dual feasible and have a duality gap of at most ε^*. The objective function (save the $1/2n$ factor) can be written as

$$\widehat{X}^{-1} \bullet X + \widehat{S}^{-1} \bullet S = \operatorname{tr}\left(\widehat{V}^{-1}X\widehat{V}^{-T} + \widehat{U}^{-1}S\widehat{U}^{-T}\right) = \sum_{i=1}^{n} \lambda_i\left(\widehat{V}^{-1}X\widehat{V}^{-T} + \widehat{U}^{-1}S\widehat{U}^{-T}\right)$$

(where $\hat{X} = \hat{V}\hat{V}^T$ and $\hat{S} = \hat{U}\hat{U}^T$) and so measures the sum of the (nonnegative) eigenvalues of X and S, appropriately weighted and summed. In this way, $\delta_2(\varepsilon^*)$ measures the 1-norm of the eigenvalue vector of $\hat{V}^{-1} X \hat{V}^{-T} + \hat{U}^{-1} S \hat{U}^{-T}$ and is a measure of the minimum norm of feasible solutions with duality gaps less than or equal to ε^*. A related interpretation of $\delta_2(\varepsilon^*)$ is seen from the following proposition:

<u>Proposition 5.3</u>. $\delta_2 = \delta_2(\varepsilon^*) \le \min_{X,\, y,\, S} \left\{ \max\left\{ |X|_{\hat{X}^{-1}},\ |S|_{\hat{S}^{-1}} \right\} \right\}$ (5.4)

$$\text{s.t.} \quad A \bullet X = b, \quad X \succeq 0$$

$$yA + S = C, \quad S \succeq 0$$

$$C \bullet X - b^T y \le \varepsilon^* .$$

Here we see that $\delta_2(\varepsilon^*)$ is bounded by the spectral norms of primal-dual feasible solutions with duality gap at most ε^*, where the norms are induced by the starting point $(\hat{X}, \hat{y}, \hat{S})$.

<u>Proof of Proposition 5.3</u>: Let (X, y, S) solve (5.4). Then (X, y, S) is feasible for (5.3), and

$$\hat{X}^{-1} \bullet X = \hat{V}^{-T}\hat{V}^{-1} \bullet X = \operatorname{tr}\left(\hat{V}^{-1} X \hat{V}^{-T}\right) = \sum_{i=1}^{n} \lambda_i \left(\hat{V}^{-1} X \hat{V}^{-T}\right)$$

$$\le n\left|\hat{V}^{-1} X \hat{V}^{-T}\right| = n\,|X|_{\hat{X}^{-1}} ,$$

where $\hat{X} = \hat{V}\hat{V}^T$.

Similarly, one easily obtains $\hat{S}^{-1} \bullet S \le n\,|S|_{\hat{S}^{-1}}$.

Therefore $\quad \delta_2 \le \frac{1}{2n}\left(\hat{X}^{-1} \bullet X + \hat{S}^{-1} \bullet S\right) \le \frac{1}{2}|X|_{\hat{X}^{-1}} + \frac{1}{2}|S|_{\hat{S}^{-1}}$

$$\le \max\left\{ |X|_{\hat{X}^{-1}}, |S|_{\hat{S}^{-1}} \right\} . \quad \blacksquare$$

Remark 5.2 (Properties of $\delta_2(\varepsilon^*)$)

(i) $\delta_2(\varepsilon^*)$ is decreasing in ε^*.

(ii) If $z_P^* = z_D^*$ and both P and D attain their optimal values, then $\delta_2(\varepsilon^*)$ is finite for all $\varepsilon^* \geq 0$.

(iii) If $z_P^* = z_D^*$ and P and/or D does not attain its respective optimal value, then $\delta_2(\varepsilon^*) \to +\infty$ as $\varepsilon^* \to 0$.

(iv) If $z_P^* > z_D^*$, then $\delta_2(\varepsilon^*)$ is finite for all $\varepsilon^* > z_P^* - z_D^*$, and $\delta_2(\varepsilon^*) = +\infty$ for all $\varepsilon^* < z_P^* - z_D^*$.

The properties follow directly from the definition of δ_2 in (5.3). To illustrate property (iii), consider example P1 of the Introduction. In order for X and (y, S) to be feasible for P1 and D1, respectively, and have a duality gap of ε^*, the value of X must be

$$X = \begin{bmatrix} \varepsilon^* & 1 \\ 1 & \frac{1}{\varepsilon^*} \end{bmatrix}.$$

For any starting point $\hat{X} \succ 0$, $\hat{X}^{-1} \bullet X \to +\infty$ as $\varepsilon^* \to 0$, and so $\delta_2 = \delta_2(\varepsilon^*) \to +\infty$ as $\varepsilon^* \to 0$.

The condition number $\delta_2(\varepsilon^*)$ is a generalization of similar ideas developed in [13] as well as Kojima [18]. Note from Proposition 5.3 that as $\lambda_n(\hat{X})$ and $\lambda_n(\hat{S})$ increases, then the upper bound on $\delta_2(\varepsilon^*)$ decreases. In the special case when SDP is simply a linear program of bit-size L, we can set $\hat{S} = \hat{X} = O(2^L)I$, and then $\delta_2(\varepsilon^*)$ will be less than or equal to 1, and so $(\ln(\delta_2))^+$ will be zero.

The interpretation of δ_1 and δ_2 as condition numbers relative to the starting point $(\hat{X}, \hat{y}, \hat{S})$ is due to Renegar [25].

6. Complexity of the Algorithm for SDP

In this section, we analyze the complexity of the algorithm. We will prove two main complexity results:

Theorem 6.1. Suppose $\gamma = 1/4$. The algorithm will find an ε^f-feasible solution of P and D in at most T_f iterations, where T_f satisfies

$$T_f \leq \left\lceil (2 + 4\sqrt{n} + 30n\,\delta_1)\, ln\left(\frac{\max\left\{\|A \bullet \hat{X} - b\|_2, \|\hat{y}A + \hat{S} - C\|\right\}}{\varepsilon^f}\right)\right\rceil + \lfloor 4\sqrt{n}\,(ln\,(\delta_2))^+ \rfloor,$$

where $\delta_2 = \delta_2(\varepsilon^*)$.

Theorem 6.2. Suppose $\gamma = 1/4$. The algorithm will find an ε^*-optimal solution of P and D in at most T_o iterations, where T_o satisfies

$$T_o \leq \left\lceil (2 + 4\sqrt{n} + 30n\,\delta_1)\left[ln\left(\frac{(5/4)\hat{X} \bullet \hat{S}}{\varepsilon^*}\right) + 2\,(ln\,(\delta_2))^+\right]\right\rceil + \lfloor 4\sqrt{n}\,(ln\,(\delta_2))^+ \rfloor,$$

where $\delta_2 = \delta_2(\varepsilon^*)$.

Before proceeding with the proofs of these two theorems, we first interpret the two complexity bounds for several different cases of the behavior of the semidefinite programs P and D.

Case 1: $z_P^* = z_D^*$ and both P and D attain their optimal values. First note in Theorem 6.1 that the iteration bound on T_f depends on the logarithm of the ratio $\max\left\{\|A \bullet \hat{X} - b\|_2, \|\hat{y}A + S - C\|\right\}/\varepsilon^f$. The smaller the initial infeasibility of the starting point and the larger the feasibility tolerance ε^f, the smaller will be this ratio. Also, the bound depends on n and on δ_1 and δ_2. The bound is linear in δ_1. However, as pointed out in Remark 5.1, one can always ensure that $\delta_1 \leq 2$ by an appropriate construction of the starting point $(\hat{X}, \hat{y}, \hat{S})$. Also, in the very special case when the starting point $(\hat{X}, \hat{y}, \hat{S})$ is feasible for both P and D, then $\delta_1 = 0$, and the bound T_f depends on $\sqrt{n}$, similar to the best complexity bounds for linear programming. The bound also depends on the logarithm of $\delta_2 = \delta_2(\varepsilon^*)$. Recall that δ_2 is a measure of the minimum norm of feasible solutions of P and D with duality gap less than or equal to ε^*. From Remark 5.2 (i) and (ii), $\delta_2(\varepsilon^*) \leq \delta_2(0)$ and $\delta_2(0)$

is a measure of the minimum norm of optimal solutions of P and D . Therefore, the dependence of the bound on T_f on the logarithm of δ_2 is reasonable, as the bound depends on the logarithm of the norm of optimal solutions in a manner similar to linear programming. However, whereas in the case of linear programming this bound can in turn be bounded by the size of the data (i.e., O(L)), there is no corresponding data-dependent bound for SDP. As discussed in Alizadeh [3], for example, there are instances of P where the optimal solution grows double-exponentially in the dimension n of the problem, and for those instances $ln(\delta_2)$ is exponential in n . Nevertheless, if one accepts the notion that the logarithm of the norm of the optimal solution is a reasonable measure of the difficulty of the problem, then both δ_1 and δ_2 enter the bound on T_f in a reasonable way. The iteration bound on T_o in Theorem 6.2 is similar, and depends on δ_1, $ln(\delta_2)$, and the ratio of the initial to the desired duality gap. The smaller the initial duality gap and the larger the optimality tolerance ε^*, the smaller will be the ratio. Finally, from Remark 2.1, the primal and dual objective function values of the iterates (X^k, y^k, S^k) will each approach the interval $[z_P^* - \varepsilon^*, z_P^* + \varepsilon^*]$ as ε^f goes to zero. In this way, as both ε^f an ε^* go to zero, the final iterative values of the algorithm will approach feasibility and optimal objective function values in the limit. However, a bound on the rate of convergence of this limit has not been determined.

Case 2: $z_P^* = z_D^*$, but P and/or D does not attain its respective optimal value. (This case arises, for example, in Karisch et. al. [16].) All of the above remarks from Case 1 still pertain to this case, except those concerning the role of δ_2. Recall that δ_2 is a measure of the minimum norm of feasible solutions of P and D with duality gap less than or equal to ε^*. For any given positive value of ε^*, $\delta_2 = \delta_2(\varepsilon^*)$ will be finite. However, from Remark 5.2 (iii), $\delta_2 = \delta_2(\varepsilon^*) \to +\infty$ as $\varepsilon^* \to 0$. Therefore, the bounds on T_f and T_o will go to $+\infty$ as $\varepsilon^* \to 0$. (One natural question concerns how quickly $ln(\delta_2(\varepsilon^*))$ grows as ε^* approaches zero. The author has not been able to create an example where $ln(\delta_2(\varepsilon^*))$ grows exponentially in either n or in the size of the data (L), and it would be interesting to find such an example.) Again, if one accepts the notion that the logarithm of the minimum norm of ε^*-optimal solutions of P and D is a reasonable measure of the difficulty of finding an ε^*-optimal solution, then as in Case 1, δ_2 enters the bound on T_f and T_o in a reasonable way.

Case 3: P and D exhibit a nonzero duality gap, i.e., $z_P^* > z_D^*$. From Remark 5.2 (iv), $\delta_2(\varepsilon^*) = +\infty$ whenever $\varepsilon^* < z_P^* - z_D^*$. Therefore, when the optimality tolerance is less than the duality gap, the bounds in Theorem 6.1 and Theorem 6.2 are both $+\infty$ and so have no value whatsoever. It is curious to note, however, that in the case when the optimality tolerance is greater than the duality gap, i.e., $\varepsilon^* > z_P^* - z_D^*$, then the bounds in these theorems are finite and all of the relevant remarks in Case 1 and in Case 2 remain valid. (However, it is hard to conceive of an instance of P where one might know the value of the finite duality gap a priori.) From Remark 2.1, as ε^f goes to zero the primal and dual objective function values of the final iterates (X^k, y^k, S^k) will approach the interval $[z_D^* - \varepsilon^*, z_P^* + \varepsilon^*]$ in the limit. However, once again a bound on the rate of convergence of this limit has not been determined.

The proof of Theorem 6.1 and Theorem 6.2 will follow as a consequence of the following five Lemmas. The first Lemma gives a bound on the size of the Newton step.

Lemma 6.1. Suppose $(\bar{X}, \bar{y}, \bar{S})$ is a γ-approximate solution of $BP(\bar{\omega}, \bar{\varepsilon})$ and $\gamma < 1$. Let $\alpha \in [0, 1]$ and $\beta > 0$ be given and let (D, y) be the solution of the Newton equations (3.2) via (3.1). Then

$$\|D\| \le \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} + \frac{(1-\alpha)\bar{\varepsilon}}{\beta\bar{\omega}}\delta_1\|\bar{X}\|_{\hat{S}} + \frac{(1-\alpha)\bar{\varepsilon}}{(1-\gamma)\bar{\omega}}\delta_1\|\bar{S}\|_{\hat{X}}.$$

Proof: Let $\bar{X} = \bar{V}\bar{V}^T$. We first have

$$\begin{aligned}\left\|I - \frac{1}{\beta\bar{\omega}}\bar{V}^T\bar{S}\bar{V}\right\| &= \left\|\frac{1}{\beta}\left(I - \frac{1}{\bar{\omega}}\bar{V}^T\bar{S}\bar{V}\right) + \frac{\beta-1}{\beta}I\right\| \\ &\le \frac{1}{\beta}\left\|\left(I - \frac{1}{\bar{\omega}}\bar{V}^T\bar{S}\bar{V}\right)\right\| + \frac{|1-\beta|}{\beta}\|I\| \\ &\le \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} \quad \text{(from (2.6c))}. \end{aligned} \tag{6.1}$$

Next, let $(\tilde{X}, \tilde{y}, \tilde{S})$ be the optimal solutions to the problems defined in (5.1), and so $(\tilde{X}, \tilde{y}, \tilde{S})$ satisfy (5.2). Then for $i = 1, \ldots, m$,

$$\begin{aligned}&\bar{A}_i \bullet \left(D - \bar{V}^{-1}(\hat{X} - \tilde{X})\bar{V}^{-T}(\alpha-1)\bar{\varepsilon}\right) = (\alpha-1)\bar{\varepsilon}\left(A_i \bullet \hat{X} - b_i\right) \\ &\quad -(\alpha-1)\bar{\varepsilon}\bar{V}^T A_i \bar{V} \bullet \bar{V}^{-1}(\hat{X} - \tilde{X})\bar{V}^{-T} = (\alpha-1)\bar{\varepsilon}\left(A_i \bullet \hat{X} - b_i - A_i \bullet \hat{X} + b_i\right) = 0, \end{aligned} \tag{6.2}$$

from (3.2a) and (3.1b) and (5.2a).

From (3.2b), (3.1b), (2.6) and (5.2b) we have

$$\left[D + \left(-I + \frac{1}{\beta\,\overline{\omega}}\overline{V}^T \overline{S}\,\overline{V} - \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\overline{V}^T\left(\hat{S}-\tilde{S}\right)\overline{V}\right)\right]$$

$$= \frac{1}{\beta\,\overline{\omega}}\sum_{i=1}^{m}\left(y_i - \overline{y}_i + (1-\alpha)\,\overline{\varepsilon}\,\hat{y}_i - (1-\alpha)\,\overline{\varepsilon}\,\tilde{y}_i\right)\overline{A}_i. \tag{6.3}$$

Combining (6.2) and (6.3) we obtain

$$\left[D + \left(-I + \frac{1}{\beta\,\overline{\omega}}\overline{V}^T \overline{S}\,\overline{V} - \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\overline{V}^T\left(\hat{S}-\tilde{S}\right)\overline{V}\right)\right] \bullet \left[D - (\alpha-1)\,\overline{\varepsilon}\,\overline{V}^{-1}\left(\hat{X}-\tilde{X}\right)\overline{V}^{-T}\right] = 0. \tag{6.4}$$

From A.12, we have

$$\begin{aligned}
\|D\| &\le \left\| I - \frac{1}{\beta\,\overline{\omega}}\overline{V}^T \overline{S}\,\overline{V} + \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\overline{V}^T\left(\hat{S}-\tilde{S}\right)\overline{V}\right\| + \overline{\varepsilon}\,(1-\alpha)\left\|\overline{V}^{-1}\left(\hat{X}-\tilde{X}\right)\overline{V}^{-T}\right\| \\
&\le \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} + \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\left\|\overline{V}^T\left(\hat{S}-\tilde{S}\right)\overline{V}\right\| + \overline{\varepsilon}\,(1-\alpha)\left\|\overline{V}^{-1}\left(\hat{X}-\tilde{X}\right)\overline{V}^{-T}\right\|
\end{aligned}$$

(from (6.1))

$$\begin{aligned}
&= \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} + \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\left\|\hat{S}-\tilde{S}\right\|_{\overline{X}} + \overline{\varepsilon}\,(1-\alpha)\left\|\hat{X}-\tilde{X}\right\|_{\overline{X}^{-1}} \\
&\le \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} + \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\left\|\hat{S}-\tilde{S}\right\|_{\overline{X}} + \frac{\overline{\varepsilon}\,(1-\alpha)}{(1-\gamma)\,\overline{\omega}}\left\|\hat{X}-\tilde{X}\right\|_{\overline{S}} \qquad \text{(from A.11)} \\
&\le \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} + \frac{(1-\alpha)\,\overline{\varepsilon}}{\beta\,\overline{\omega}}\,\delta_1\left\|\overline{X}\right\|_{\hat{S}} + \frac{\overline{\varepsilon}\,(1-\alpha)}{(1-\gamma)\,\overline{\omega}}\,\delta_1\left\|\overline{S}\right\|_{\hat{X}}
\end{aligned}$$

where the last inequality follows from Proposition 5.2. ■

The next Lemma shows that every iteration of the algorithm is a γ-approximate solution of BP (ω, ε).

Lemma 6.2. For $k = 0, 1, \ldots, (X^k, y^k, S^k)$ is a γ-approximate solution of $BP(\omega^k, \varepsilon^k)$.

Proof: The result is obviously true for $k = 0$, from Proposition 2.2. Now supposing that (X^k, y^k, S^k) is a γ-approximate solution of $BP(\omega^k, \varepsilon^k)$, we must show that $(X^{k+1}, y^{k+1}, S^{k+1})$ is a γ-approximate solution of $BP(\omega^{k+1}, \varepsilon^{k+1})$. From Lemma 3.1, this will be true if the Newton direction D satisfies $\|D\| \leq \sqrt{\gamma}$ at each iteration. There are three cases to consider, depending on whether the algorithm visits Step 2, 3, or 4, at iteration k. If the algorithm visits Steps 2 or 3, then (4.5c) or (4.6c) ensures that

$$\|D\|^2 = \mathrm{tr}(D^T D) = \mathrm{tr}(D^2) \leq \gamma,$$

so $\|D\| \leq \sqrt{\gamma}$ as needed. If the algorithm visits Step 4, then from Lemma 6.1, with $\alpha = 1$ and $\beta = \hat{\beta}$,

$$\|D\| \leq \frac{\gamma}{\hat{\beta}} + \frac{\hat{\beta} - 1}{\hat{\beta}} \sqrt{n} = \sqrt{\gamma} \quad \text{from the construction of } \hat{\beta} \text{ at Step 0.}$$

Therefore in any of the cases, $\|D\| \leq \sqrt{\gamma}$, and so the new iterate will be a γ-approximate solution. ■

The next two Lemmas bound the values of α and β at Steps 2 and 3.

Lemma 6.3. Suppose that $\gamma = 1/4$. If the algorithm is at Step 2, then unless

$\beta \bar{\omega} = \dfrac{\varepsilon^*}{n + \sqrt{n}\,\gamma}$, β must satisfy $\beta \leq 1 - \dfrac{1}{2 + 4\sqrt{n}}$.

Proof: We first show that for $\alpha = 1$ and $1 \geq \beta \geq 1 - \dfrac{\sqrt{\gamma} - \gamma}{\sqrt{n} + \sqrt{\gamma}}$, that $\mathrm{tr}(D^2) \leq \gamma$.

To see this, note from Lemma 6.1 that

$$\begin{aligned} \|D\| &\leq \frac{\gamma}{\beta} + \frac{|1-\beta|}{\beta}\sqrt{n} = \frac{\gamma}{\beta} + \frac{1-\beta}{\beta}\sqrt{n} = \frac{\gamma + \sqrt{n}}{\beta} - \sqrt{n} \\ &\leq \frac{\sqrt{n} + \sqrt{\gamma}}{\sqrt{n} + \gamma}(\gamma + \sqrt{n}) - \sqrt{n} = \sqrt{\gamma}. \end{aligned}$$

Therefore $\operatorname{tr}(D^2) = ||D||^2 \le \gamma$, and so the smallest value of β satisfying (4.5) must be less than or equal to $1 - \frac{\sqrt{\gamma}-\gamma}{\sqrt{n}+\sqrt{\gamma}} = 1 - \frac{1}{2+4\sqrt{n}}$. ■

Lemma 6.4. Suppose that $\gamma = 1/4$. If the algorithm is at Step 3, then unless

$\beta = \frac{\varepsilon^*}{\bar{\omega}(n+\sqrt{n}\gamma)}$, β must satisfy $\beta \le 1 - \frac{1}{2+4\sqrt{n}+30n\delta_1}$.

Also, unless $\alpha = \varepsilon'/\bar{\varepsilon}$, then α must satisfy

$$\alpha \le 1 - \frac{1}{2+4\sqrt{n}+30n\delta_1}.$$

Proof: It is sufficient to show that if $\alpha = \beta = \delta$, and if

$$1 \ge \delta \ge 1 - \frac{1}{2+4\sqrt{n}+30n\delta_1},$$

then $\operatorname{tr}(D^2) \le \gamma$. To see this, note from Lemma 6.1 that

$$||D|| \le \frac{\gamma}{\beta} + \frac{1-\beta}{\beta}\sqrt{n} + \frac{(1-\alpha)}{\beta}\left(\frac{\bar{\varepsilon}}{\bar{\omega}}\right)\delta_1||\bar{X}||_{\hat{S}} + \frac{(1-\alpha)}{(1-\gamma)}\frac{\bar{\varepsilon}}{\bar{\omega}}\delta_1||\bar{S}||_{\hat{X}}$$

$$\le \frac{\gamma+\sqrt{n}}{\delta} - \sqrt{n} + \frac{(1-\delta)\bar{\varepsilon}}{\delta(1-\gamma)\bar{\omega}}\delta_1\left(||\bar{X}||_{\hat{S}} + ||\bar{S}||_{\hat{X}}\right)$$

$$\le \frac{\gamma+\sqrt{n}}{\delta} - \sqrt{n} + \frac{(1-\delta)}{\delta(1-\gamma)}\delta_1(2n)(1+\gamma+\hat{\beta}) \quad \text{(from (4.3))}$$

$$\le \frac{1/4+\sqrt{n}}{\delta} - \sqrt{n} + \left(\frac{1-\delta}{\delta}\right)(2n\delta_1)\left(\frac{2.75}{0.75}\right) \quad \text{(since } \gamma = 1/4 \text{ and } \hat{\beta} \le 1.5\text{)}$$

$$= \frac{1/4+\sqrt{n}}{\delta} - \sqrt{n} + \left(\frac{1-\delta}{\delta}\right)\left(\frac{22}{3}\right)n\delta_1$$

$$\le \frac{1/4+\sqrt{n}}{\delta} - \sqrt{n} + \frac{22.5}{3\delta}n\delta_1 - \frac{22.5n\delta_1}{3}$$

$$= \frac{1/4 + \sqrt{n} + \frac{22.5}{3} n \delta_1}{\delta} - \left(\sqrt{n} + \frac{22.5}{3} n \delta_1 \right)$$

$$= \frac{1/4 (1 + 4 \sqrt{n} + 30 n \delta_1)}{\delta} - \frac{1}{4} (4 \sqrt{n} + 30 n \delta_1)$$

$$\leq \frac{1}{4} (2 + 4 \sqrt{n} + 30 n \delta_1) - \frac{1}{4} (4 \sqrt{n} + 30 n \delta_1) = \frac{1}{2} = \sqrt{\gamma} .$$

Therefore, $tr (D^2) = || D ||^2 \leq \gamma$, and so the smallest value of δ satisfying (4.6) will be less than or equal to $1 - \frac{1}{2 + 4 \sqrt{n} + 30 n \delta_1}$. ■

The last Lemma gives a bound on the number of iterations that visit Step 4.

Lemma 6.5. Suppose that $\gamma = 1/4$. Let T_4 be the number of iterations of the algorithm that visit Step 4. Then

$$T_4 \leq \lfloor 4 \sqrt{n} (ln (\delta_2))^+ \rfloor . \quad ■$$

The proof of this Lemma is deferred to the next Section.

Proof of Theorem 6.1:

At the start of the algorithm, $\varepsilon^0 = 1$. In order for (X^k , y^k , S^k) to be an ε^f-feasible solution of P and D, ε^k must be less than or equal to ε', where ε' is defined in Step 0 of the algorithm. Examining the rules for the choice of α in Steps 2, 3, and 4 of the algorithm, notice the algorithm only visits Step 2 after $\varepsilon^k \leq \varepsilon'$. Therefore, in order to bound the number of iterations before $\varepsilon^k \leq \varepsilon'$, it is sufficient to bound the number of iterations k that the algorithm visits Step 3 and Step 4 and $\varepsilon^k > \varepsilon'$. Let T_3 and T_4 denote the number of iterations k that the algorithm visits Step 3 and Step 4, and $\varepsilon^k > \varepsilon'$, respectively. From Lemma 6.5, we know that $T_4 \leq \lfloor 4 \sqrt{n} (ln (\delta_2))^+ \rfloor$. In order to bound T_3, note from Lemma 6.4 that unless $\alpha = \varepsilon' / \bar{\varepsilon}$ at Step 3 (implying $\varepsilon^{k+1} = \varepsilon = \alpha \bar{\varepsilon} = \varepsilon'$), that at Step 3 we must have

$$\alpha \leq 1 - \frac{1}{2 + 4 \sqrt{n} + 30 n \delta_1} .$$

$$\text{Let } T = \left\lceil (2 + 4\sqrt{n} + 30 n \delta_1) \, ln \left(\frac{\max\{ \| A \bullet \hat{X} - b \|_2 , \| \hat{y} A + \hat{S} - C \| \}}{\varepsilon f} \right) \right\rceil$$

$$= -(2 + 4\sqrt{n} + 30 n \delta_1) \, ln \, (\varepsilon').$$

Then after the algorithm visits Step 3 T times, the value of ε will satisfy

$$\varepsilon \le \left(1 - \frac{1}{2 + 4\sqrt{n} + 30 n \delta_1}\right)^T$$

whereby

$$ln \, \varepsilon \le T \, ln \left(1 - \frac{1}{2 + 4\sqrt{n} + 30 n \delta_1}\right) \le \frac{-T}{2 + 4\sqrt{n} + 30 n \delta_1} \le ln \, (\varepsilon')$$

and so $\varepsilon \le \varepsilon'$. Therefore $T_3 \le T$, completing the proof. ■

Proof of Theorem 6.2:

From Proposition 2.3, a bound on the duality gap at iteration k is $X^k \bullet S^k \le \omega^k (n + \sqrt{n}\gamma)$. Therefore, in order for $X^k \bullet S^k \le \varepsilon^*$, it is sufficient that $\omega^k \le \frac{\varepsilon^*}{n + \sqrt{n}\gamma}$. The starting value of ω is $\omega^0 = \hat{X} \bullet \hat{S}$, and $\omega^{k+1} \le \omega^k$ if the algorithm visits Steps 2 and 3, whereas $\omega^{k+1} > \omega^k$ if the algorithm visits Step 4. Let T_2, T_3, and T_4, denote the number of iterations that the algorithm visits Steps 2, 3, and 4, respectively, while the duality gap exceeds ε^*. After each visit to Step 4, ω^k increases by the quantity $\hat{\beta}$, and note that

$$\hat{\beta} = 1 + \frac{\sqrt{\gamma} - \gamma}{\sqrt{n} - \sqrt{\gamma}} = 1 + \frac{1}{4\sqrt{n} - 2} \le 1 + \frac{1}{2\sqrt{n}}, \tag{6.5}$$

at $\gamma = 1/4$. After each visit to Step 3, ω^k decreases by the quantity β where

$$\beta \le 1 - \frac{1}{2 + 4\sqrt{n} + 30 n \delta_1},$$

according to Lemma 6.4 (unless $\omega^{k+1} = \beta\,\overline{\omega} = \dfrac{\varepsilon^*}{n + \sqrt{n}\,\gamma}$, and so $\left(X^{k+1}\right) \bullet \left(S^{k+1}\right) \le \varepsilon^*$, and ε^*-optimality is achieved). Similarly, after each visit to Step 2, ω^k decreases by the quantity β where

$$\beta \le 1 - \frac{1}{2 + 4\sqrt{n}} \le 1 - \frac{1}{2 + 4\sqrt{n} + 30\,n\,\delta_1}$$

according to Lemma 6.3 (unless $\omega^{k+1} = \beta\,\overline{\omega} = \dfrac{\varepsilon^*}{n + \sqrt{n}\,\gamma}$, and so $\left(X^{k+1}\right) \bullet \left(S^{k+1}\right) \le \varepsilon^*$, and ε^*-optimality is achieved).

$$\text{Let } T = \left\lceil \left(2 + 4\sqrt{n} + 30\,n\,\delta_1\right)\left[\ln\left(\frac{5\,\hat{X} \bullet \hat{S}}{4\,\varepsilon^*}\right) + 2\left(\ln\left(\delta_2\right)\right)^+ \right] \right\rceil. \qquad (6.6)$$

Then after T_4 visits to Step 4 and a total of T visits to Steps 2 and 3, the value of ω will satisfy

$$\omega \le \left(\omega^0\right)\left(1 - \frac{1}{2 + 4\sqrt{n} + 30\,n\,\delta_1}\right)^T \left(\hat{\beta}\right)^{T_4}, \text{ and so}$$

$$\ln\omega \le \ln\left(\omega^0\right) + T\ln\left(1 - \frac{1}{2 + 4\sqrt{n} + 30\,n\,\delta_1}\right) + T_4 \ln\left(\hat{\beta}\right)$$

$$\le \ln\left(\hat{X} \bullet \hat{S}/n\right) - \frac{T}{2 + 4\sqrt{n} + 30\,n\,\delta_1} + \frac{T_4}{2\sqrt{n}} \qquad \text{(from (6.5))}$$

$$\le \ln\left(\hat{X} \bullet \hat{S}/n\right) - \ln\left(\frac{5\,\hat{X} \bullet \hat{S}}{4\,\varepsilon^*}\right) - 2\left(\ln\left(\delta_2\right)\right)^+ + 2\left(\ln\left(\delta_2\right)\right)^+$$

(from Lemma (6.5) and (6.6))

$$= \ln\left(\frac{4\,\varepsilon^*}{5\,n}\right) = \ln\left(\frac{\varepsilon^*}{n(1+\gamma)}\right) \le \ln\left(\frac{\varepsilon^*}{n + \gamma\sqrt{n}}\right).$$

Therefore $\omega \le \varepsilon^*/(n + \gamma\sqrt{n})$, and so the duality gap will be at most ε^* (from Proposition 2.3).

Therefore $T_2 + T_3 \le T$, and this inequality plus Lemma 6.5 shows that

$$T_2+T_3+T_4 \le \left[(2+4\sqrt{n}+30\,n\,\delta_1)\left[ln\left(\frac{(5/4)\,\hat{X}\bullet\hat{S}}{\varepsilon^*}+2\,(ln\,(\delta_2))^+\right)\right]\right]+\lfloor 4\sqrt{n}\,(ln\,(\delta_2))^+\rfloor. \quad \blacksquare$$

7. Proof of Lemma 6.5

Note that if the optimization problem (5.3) is not feasible, then $\delta_2 = +\infty$ and Lemma 6.5 is trivially true. Therefore, we suppose in this section that (5.3) is feasible. It is easy to see that (5.3) attains its optimum at some point (X^*, y^*, S^*). We first present two propositions, that are extensions of results in Mizuno [20] for the case of linear programming:

Proposition 7.1. If (X^*, y^*, S^*) is an optimal solution of (5.3) and $(\bar{X}, \bar{y}, \bar{S})$ is a feasible solution of BP $(\bar{\omega}, \bar{\varepsilon})$ for $\bar{\varepsilon} \in [0, 1]$, then

$$\left(\bar{\varepsilon}\hat{X} + (1-\bar{\varepsilon})X^*\right)\bullet\bar{S} + \left(\bar{\varepsilon}\hat{S} + (1-\bar{\varepsilon})S^*\right)\bullet\bar{X} \tag{7.1}$$

$$= \bar{X}\bullet\bar{S} + \left(\bar{\varepsilon}\hat{X} + (1-\bar{\varepsilon})X^*\right)\bullet\left(\bar{\varepsilon}\hat{S} + (1-\bar{\varepsilon})S^*\right).$$

Proof: We have

$$A\bullet\left(\bar{\varepsilon}\hat{X} + (1-\bar{\varepsilon})X^* - \bar{X}\right)$$

$$= \bar{\varepsilon}A\bullet\hat{X} + (1-\bar{\varepsilon})b - \left(b + \bar{\varepsilon}\left(A\bullet\hat{X} - b\right)\right) = 0$$

from (5.3) and (2.4). Similarly, we have from (5.3) and (2.4) that

$$\left(\bar{\varepsilon}\hat{y} + (1-\bar{\varepsilon})y^* - \bar{y}\right)A + \left(\bar{\varepsilon}\hat{S} + (1-\bar{\varepsilon})S^* - \bar{S}\right)$$

$$= \bar{\varepsilon}\left(\hat{y}A + \hat{S}\right) + (1-\bar{\varepsilon})C - \left(C + \bar{\varepsilon}\left(\hat{y}A + \hat{S} - C\right)\right) = 0 .$$

Combining these two expressions gives

$\left(\bar{\varepsilon}\hat{X} + (1-\bar{\varepsilon})X^* - \bar{X}\right)\bullet\left(\bar{\varepsilon}\hat{S} + (1-\bar{\varepsilon})S^* - \bar{S}\right) = 0$, and rearranging yields the desired result. $\blacksquare$

Proposition 7.2. If $(\bar{X}, \bar{y}, \bar{S})$ is a γ-approximate solution of BP $(\bar{\omega}, \bar{\varepsilon})$ for some $\gamma \in (0, 1)$, and $\bar{\omega} \geq \dfrac{\varepsilon^*}{n(1+\gamma)}$, then

$$\left(\|\bar{S}\|_{\hat{X}} + \|\bar{X}\|_{\hat{S}}\right)\left(\frac{\bar{\varepsilon}}{\bar{\omega}}\right) \leq 2n(1+\gamma) + 2n\,\hat{\theta}\left(\frac{\bar{\varepsilon}}{\bar{\omega}}\right)\max\{\delta_2, 1\}.$$

Proof: Let $\hat{X} = \hat{V}\hat{V}^T$, $\hat{S} = \hat{U}\hat{U}^T$, be factorizations of $\hat{X}$ and $\hat{S}$. Then

$$\bar{\varepsilon}\left(\|\bar{S}\|_{\hat{X}} + \|\bar{X}\|_{\hat{S}}\right) = \bar{\varepsilon}\left(\left\|\hat{V}^T\bar{S}\hat{V}\right\| + \left\|\hat{U}^T\bar{X}\hat{U}\right\|\right)$$

$$\leq \bar{\varepsilon}\left(\mathrm{tr}\left(\hat{V}^T\bar{S}\hat{V}\right) + \mathrm{tr}\left(\hat{U}^T\bar{X}\hat{U}\right)\right) \qquad \text{(from (A.8))}$$

$$\leq \bar{\varepsilon}\bar{S} \bullet \hat{X} + \bar{X} \bullet \hat{S} \qquad \text{(from (A.1 and A.2))}$$

$$\leq \bar{S} \bullet \left(\bar{\varepsilon}\hat{X} + (1-\bar{\varepsilon})X^*\right) + \bar{X} \bullet \left(\bar{\varepsilon}\hat{S} + (1-\bar{\varepsilon})S^*\right)$$

(from (A.6))

$$= \bar{X} \bullet \bar{S} + \left(\bar{\varepsilon}\hat{X} + (1-\bar{\varepsilon})X^*\right) \bullet \left(\bar{\varepsilon}\hat{S} + (1-\bar{\varepsilon})S^*\right)$$

(from Proposition 7.1)

$$\leq \bar{\omega}n(1+\gamma) + \bar{\varepsilon}^2\hat{\theta}n + \bar{\varepsilon}(1-\bar{\varepsilon})\left(\hat{X} \bullet S^* + \hat{S} \bullet X^*\right) + (1-\bar{\varepsilon})^2 X^* \bullet S^*$$

(from (2.3) and Proposition 2.3)

$$\leq \bar{\omega}n(1+\gamma) + \bar{\varepsilon}^2\hat{\theta}n + \bar{\varepsilon}(1-\bar{\varepsilon})\hat{\theta}\left(\hat{S}^{-1} \bullet S^* + \hat{X}^{-1} \bullet X^*\right) + (1-\bar{\varepsilon})^2\varepsilon^*$$

(from (2.3) and (5.3))

$$\leq \bar{\omega}n(1+\gamma) + \bar{\varepsilon}^2\hat{\theta}n + \bar{\varepsilon}(1-\bar{\varepsilon})\hat{\theta}(2n\delta_2) + (1-\bar{\varepsilon})^2\bar{\omega}n(1+\gamma)$$

(from (5.3) and the statement of the proposition)

$$\leq 2\bar{\omega} n(1+\gamma) + 2\,\hat{\theta}\,\bar{\varepsilon} n(\bar{\varepsilon} + (1 - \bar{\varepsilon})\delta_2)$$

$$\leq 2\bar{\omega} n(1+\gamma) + 2\,\hat{\theta}\,\bar{\varepsilon} n(\max\{\delta_2, 1\})$$

Dividing both sides by $\bar{\omega}$ proves the result. ■

Proposition 7.3. All iterations of the algorithm satisfy

$$\omega^k \geq \frac{\varepsilon^*}{n(1+\gamma)}.$$

Proof: For $k = 0$, $\omega^0 = \hat{\theta} = (n\hat{\theta})(1/n) \geq \frac{\varepsilon^*}{n} \geq \frac{\varepsilon^*}{n(1+\gamma)}$, by the assumption that $\hat{X} \bullet \hat{S} \geq \varepsilon^*$. Suppose $\omega^k \geq \frac{\varepsilon^*}{n(1+\gamma)}$, then we must prove that $\omega^{k+1} \geq \frac{\varepsilon^*}{n(1+\gamma)}$. At Step 1, with $(\bar{X}, \bar{y}, \bar{S}, \bar{\varepsilon}, \bar{\omega}) = (X^k, y^k, S^k, \varepsilon^k, \omega^k)$, either (4.1), (4.2), (4.3), or (4.4) will be satisfied. If (4.1) is satisfied, the algorithm halts and there is nothing to prove. If (4.2) or (4.3) is satisfied, then (4.5d) or (4.6d) ensures that $\omega^{k+1} = \beta\bar{\omega} \geq \frac{\varepsilon^*}{n + \sqrt{n}\gamma} \geq \frac{\varepsilon^*}{n(1+\gamma)}$. If (4.4) is satisfied, $\omega^{k+1} = \hat{\beta}\omega^k > \omega^k \geq \frac{\varepsilon^*}{n(1+\gamma)}$, proving the result. ■

Proposition 7.4. If the algorithm visits Step 4 at iteration k, then

$$\frac{\omega^k}{\varepsilon^k} \leq \hat{\theta}\max\{\delta_2, 1\} \text{ and } \frac{\omega^{k+1}}{\varepsilon^{k+1}} \leq \hat{\theta}\max\{\delta_2, 1\}.$$

Proof: If the algorithm visits Step 4, then from (4.4), Proposition 7.2, and Proposition 7.3,

$$2n(1+\gamma+\hat{\beta}) < \left(\|\bar{X}\|_{\hat{S}} + \|\bar{S}\|_{\hat{X}}\right)\left(\frac{\bar{\varepsilon}}{\bar{\omega}}\right) \leq 2n(1+\gamma) + 2n\,\hat{\theta}\left(\frac{\bar{\varepsilon}}{\bar{\omega}}\right)\max\{\delta_2, 1\}.$$

Therefore

$$\hat{\beta} < \hat{\theta}\left(\frac{\bar{\varepsilon}}{\bar{\omega}}\right)\max\{\delta_2, 1\},$$

whereby

$$\frac{\omega^k}{\varepsilon^k} = \frac{\overline{\omega}}{\overline{\varepsilon}} < \left(\frac{1}{\hat{\beta}}\right)\hat{\theta}\max\{\delta_2, 1\} < \hat{\theta}\max\{\delta_2, 1\}. \tag{7.2}$$

Also, from (7.2), we have

$$\frac{\omega^{k+1}}{\varepsilon^{k+1}} = \frac{\hat{\beta}\,\omega^k}{\varepsilon^k} < \hat{\theta}\max\{\delta_2, 1\}. \quad \blacksquare$$

Finally, we have:

Proof of Lemma 6.5: Note that ε^k is nonincreasing in k, that is, ε^k never is increased at any iteration because $\alpha \leq 1$ at every iteration (see Lemma 6.4). Therefore, if $\varepsilon^k = \overline{\varepsilon} \leq \varepsilon'$ (condition (4.1) or (4.2)), then condition (4.3) or (4.4) cannot hold at any subsequent iteration, so once the algorithm visits Step 1 or Step 2, the algorithm will never visit Step 3 or Step 4 at any subsequent iterations. We will prove the result by examining changes in the ratio ω^k/ε^k when the algorithm visits Steps 3 or 4. When the algorithm is at Step 4,

$$\left(\omega^{k+1}/\varepsilon^{k+1}\right) = \hat{\beta}\left(\omega^k/\varepsilon^k\right). \tag{7.3}$$

When the algorithm is at Step 3,

$$\frac{\left(\omega^{k+1}/\varepsilon^{k+1}\right)}{\left(\omega^k/\varepsilon^k\right)} = \frac{\max\left\{\delta, \varepsilon^*/\overline{\omega}(n + \gamma\sqrt{n})\right\}}{\max\{\delta, \varepsilon'/\overline{\varepsilon}\}} \geq \frac{\delta}{\max\{\delta, \varepsilon'/\overline{\varepsilon}\}}.$$

This last quantity will be equal to one, unless $\alpha = \varepsilon'/\overline{\varepsilon}$ at Step 3 and so $\varepsilon^{k+1} = \varepsilon'$ so the algorithm will next visit Step 1 or 2, never visiting Step 4 again. Therefore, for all iterations prior to first visiting Step 1 or Step 2, $\frac{\omega^k}{\varepsilon^k}$ is nondecreasing. Furthermore, whenever the algorithm visits Step 4, (7.3) states that $\frac{\omega^k}{\varepsilon^k}$ increases by the factor $\hat{\beta}$.

Therefore, from Proposition 7.4,

$$\left(\frac{\omega^0}{\varepsilon^0}\right)\hat{\beta}^{T_4} \leq \hat{\theta}\max\{\delta_2, 1\}.$$

However, $\frac{\omega^0}{\varepsilon^0} = \hat{\theta}$, so that $\hat{\beta}^{T_4} \leq \max\{\delta_2, 1\}$, (7.4) whereby $\left(\frac{1}{\hat{\beta}}\right)^{T_4} \max\{\delta_2, 1\} \geq 1$,

and so $T_4 \, ln\left(\frac{1}{\hat{\beta}}\right) + ln\,(\max\{\delta_2, 1\}) \geq 0$.

However, $ln\,(\max\{\delta_2, 1\}) = (ln\,(\delta_2))^+$

and $ln\left(\frac{1}{\hat{\beta}}\right) \leq \frac{1}{\hat{\beta}} - 1 = \frac{-1}{4\sqrt{n} - 1}$ for $\gamma = \frac{1}{4}$.

Therefore $T_4 \leq (4\sqrt{n} - 1)(ln\,(\delta_2))^+ \leq 4\sqrt{n}\,(ln\,(\delta_2))^+$

which implies

$$T_4 \leq \lfloor 4\sqrt{n}(ln\,(\delta_2))^+ \rfloor . \quad \blacksquare$$

8. Discussion and Concluding Remarks

We conclude this study with a discussion of several points concerning the interpretation of the results contained herein and the relation to other avenues of research investigation.

Bounding the size of near-optimal solutions As discussed in Section 5, the (relative) condition number δ_2 is a measure of the size of near-optimal solutions of (P) and (D), i.e., δ_2 is a lower bound on the size of primal and dual solutions with a duality gap no greater than the desired value ε^* (see (5.3) and Proposition 5.3, for example). In a manner similar to many interior-point algorithms for linear programming (see Todd and Ye [26] and [13]), the algorithm of Section 4 provides a lower bound on δ_2 whenever the algorithm visits Step 4, as the following Lemma demonstrates:

Lemma 8.1: Suppose $\gamma = \frac{1}{4}$. Let T_4 denote the number of iterations of the algorithm that visit Step 4. If $T_4 > 0$, then

$$\delta_2 \geq \left(1 + \frac{1}{4\sqrt{n} - 2}\right)^{T_4}.$$

This Lemma provides a lower bound on δ_2, which from (5.3) or Proposition 5.3 provides a lower bound on the size of ε^*-optimal solution of (P) and (D).

<u>Proof of Lemma 8.1</u> If $\delta_2 > 1$, then from (7.4) it follows that

$$\delta_2 \geq (\hat{\beta})^{T_4},$$

and the result follows by substituting $\gamma = \frac{1}{4}$ in the formula for $\hat{\beta}$ in Step 0 of the algorithm. It thus remains to show that $\delta_2 > 1$ if $T_4 > 0$.

Suppose $T_4 > 0$. Then there is some iteration k where the algorithm visits Step 4. From (7.2) it follows that

$$\frac{\omega^k}{\varepsilon^k} < \hat{\theta} \max\{\delta_2, 1\}. \tag{8.1}$$

However, from the proof of Lemma 6.5, it follows that

$$\frac{\omega^0}{\varepsilon^0} \leq \frac{\omega^k}{\varepsilon^k}, \tag{8.2}$$

because $\frac{\omega^i}{\varepsilon^i}$ is nondecreasing in i for $0 \leq i \leq k$. Combining (8.1) and (8.2) we obtain

$$\hat{\theta} = \frac{\hat{\theta}}{1} = \frac{\omega^0}{\varepsilon^0} < \hat{\theta} \max\{\delta_2, 1\},$$

whereby $\delta_2 > 1$. ■

<u>Rescaling and symmetric Newton directions</u>. The Newton equations (3.2) were derived by taking the quadratic approximation to the parametric logarithmic barrier problem $BP(\omega, \varepsilon)$ of (2.4) at the primal point $\overline{X}$ factorized and rescaled via (3.1). One pleasant feature of this derivation is that the Newton direction D is automatically a symmetric matrix, since all of the other relevant matrices in (3.2b) are symmetric as well. This is evidently due to the fact that only primal information ($\overline{X}$ and its factorization $\overline{V}\,\overline{V}^T$) is used in the rescaling, and the Newton direction is derived for the primal problem via a quadratic approximation. This is in contrast to other Newton methods for semi-definite programming, which compute the Newton direction for the Karush-Kuhn-Tucker

optimality conditions, and so use rescaling based on both primal and dual information. In these methods, there is a need to symmetrize the Newton direction via techniques that are explored in Kojima et. al. [19], Alizadeh et. al. [4], [5], Vandenberghe and Boyd [27], and Nesterov and Todd [22]. However, from both a computational as well as an aesthetic point of view, the use of primal-dual rescaling has advantages over the primal rescaling used herein. Computational experience has demonstrated the typically superior performance of primal-dual rescaling for linear programming, and so it is reasonable to extrapolate that primal-dual rescaling would be superior as well for semi-definite programming. Also, from an aesthetic point of view, primal-dual rescaling is superior because it better reflects the natural duality and symmetry of the primal and the dual problem, and uses more information than does primal-only rescaling. The main reason that primal-only scaling has been used in this paper is its simplicity and the relative ease of manipulating the arithmetic to derive the desired results. A practical implementation of the algorithm in this paper should surely consider primal-dual scaling. Furthermore, although it is reasonable to believe that the theoretical results in this paper could be proved with a version of the algorithm that uses primal-dual rescaling, it would be presumptuous to assume that such proofs would be "easily" obtained.

General behavior of semi-definite programming problems. One of the fascinating features about semi-definite programming is that when an instance of SDP has a positive definite feasible solution in the primal and in the dual (i.e., it satisfies a primal-and-dual Slater condition), then the SDP instance is extremely well-behaved (optima exist for the primal and the dual, with no duality gap, and virtually all of the linear programming interior-point tools and methodologies can be extended to SDP). In this case SDP is as well-behaved in theory and in practice as linear programming. In this study, we have extended interior-point methods to derive certain results regarding the complexity of obtaining approximate solutions in the case when an SDP instance does not satisfy a primal-and-dual Slater condition. An interesting question is whether there are any conditions weaker than the primal-and-dual Slater condition that will allow for an instance of SDP to be well-behaved, i.e., to exhibit primal and dual optimal solutions with no duality gap.

Appendix

This appendix reviews some basic properties of the trace of a matrix and norms defined on square matrices. Recall that $\mathrm{tr}(M) = \sum_{i=1}^{n} M_{ii}$, $A \bullet B = \mathrm{tr}(AB)$, $||M|| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij}^2}$, and $|M| = \max_i \{|\lambda_i(M)|\}$, where $\lambda_1(M), \ldots, \lambda_n(M)$ are the ordered eigenvalues of M. The following five properties follow directly from these definitions:

A.1 $A \bullet (BC) = AB \bullet C$.

A.2 $A \bullet B = B \bullet A$.

A.3 $||M|| = \sqrt{\mathrm{tr}(M^T M)}$.

A.4 $||M|| = ||M^T||$.

A.5 If $R^T = R^{-1}$, then $||R^T M R|| = ||M||$ and $|R^T M R| = |M|$.

A.6 If $X \succeq 0$ and $S \succeq 0$, then $X \bullet S \geq 0$, and $X \bullet S = 0$ if and only if $XS = 0$.

Proof: Let $X = VV^T$. Then $X \bullet S = VV^T \bullet S = \mathrm{tr}(VV^T S) = \mathrm{tr}(V^T S V)$ (from A.1 and A.2), and finally $\mathrm{tr}(V^T S V) \geq 0$ since $V^T S V \succeq 0$. If $X \bullet S = 0$, then $\mathrm{tr}(V^T S V) = 0$, which implies $V^T S V = 0$. As S can be factorized into $S = UU^T$, we have $V^T U U^T V = 0$, so that $V^T U = 0$, and $XS = VV^T UU^T = V(V^T U)U^T = 0$. ■

A.7 If M is symmetric, $|M| \leq ||M||$.

Proof: $M = QDQ^T$, where $Q^{-1} = Q^T$ and D is a diagonal matrix, and D_{ii} is an eigenvalue M, $i = 1, \ldots, n$. Then $||M|| = \sqrt{\mathrm{tr}(M^T M)} = \sqrt{\mathrm{tr}(QDQ^T QDQ^T)} = \sqrt{\mathrm{tr}(D^2)} = \sqrt{\sum_{i=1}^{n} D_{ii}^2} \geq \max_i |D_i| = |M|$. ■

A.8 If $M \succeq 0$, $||M|| \le \mathrm{tr}(M)$.

Proof: From the proof of A.7,

$$||M|| = \sqrt{\sum_{i=1}^{n} D_{ii}^2} \le \sum_{j=1}^{n} |D_{ii}| = \sum_{j=1}^{n} D_{ii} = \mathrm{tr}(D)$$

$$= \mathrm{tr}\left(DQ^TQ\right) = \mathrm{tr}\left(QDQ^T\right) = \mathrm{tr}(M). \quad \blacksquare$$

A.9 $||AB|| \le ||A||\,||B||$.

Proof: $$||AB|| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{k=1}^{n} A_{ik}B_{kj}\right)^2}$$

$$\le \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left(\sum_{k=1}^{n} A_{ik}^2\right)\left(\sum_{k=1}^{n} B_{kj}^2\right)\right)}$$

$$= \sqrt{\left(\sum_{i=1}^{n}\sum_{k=1}^{n} A_{ik}{}^2\right)\left(\sum_{j=1}^{n}\sum_{k=1}^{n} B_{kj}^2\right)}$$

$$= ||A||\,||B||. \quad \blacksquare$$

A.10 If N is a symmetric matrix, then $||B^TNB|| \le ||B^TB||\,|N|$

and $||B^TNB|| \le |B^TB|\,||N||$.

Proof: Let $N = R^TDR$, $R^T = R^{-1}$, D diagonal, and $B^TB = QEQ^T$, $Q^T = Q^{-1}$, E diagonal. Then $||E|| = ||B^TB||$, $|E| = |B^TB|$, $||N|| = ||D||$, $|N| = |D|$. Because $B^TR^TRB = B^TB = QE^{1/2}E^{1/2}Q^T$, it follows from Proposition 5.1 that

$$||B^TNB|| = ||B^TR^TDRB|| = ||QE^{1/2}DE^{1/2}Q^T|| = ||E^{1/2}DE^{1/2}||$$

$$= ||DE||.$$

However, as D and E are diagonal, $||DE|| \le ||D||\,|E| = ||N||\,|B^TB|$ and $||DE|| \le |D|\,||E|| = |N|\,||B^TB||$, proving the result. $\blacksquare$

A.11 If $\left\| I - \frac{1}{\bar{\omega}} \bar{V}^T \bar{S} \bar{V} \right\| \leq \gamma < 1$, where $\bar{X} = \bar{V}\bar{V}^T$, then for any matrix M,

$$||M||_{\bar{X}^{-1}} \leq \frac{1}{\bar{\omega}(1-\gamma)} ||M||_{\bar{S}} .$$

Proof: Let $R = I - \frac{1}{\bar{\omega}} \bar{V}^T \bar{S} \bar{V}$,. Then $||R|| \leq 1$, and so by A.7, $|R| \leq ||R|| \leq \gamma$, whereby $\lambda_i(I - R) \geq 1 - \gamma$, $i = 1, \ldots, n$.

Therefore $|(I - R)^{-1}| \leq \frac{1}{1-\gamma}$.

Let $\bar{S} = \bar{U}\bar{U}^T$. Then $\frac{1}{\bar{\omega}}(I - R)^{-1} = \bar{V}^{-1}\bar{U}^{-T}\bar{U}^{-1}\bar{V}^{-T}$. We have:

$$||M||_{\bar{X}^{-1}} = \left\|\bar{V}^{-1} M \bar{V}^{-T}\right\| = \left\|\bar{V}^{-1}\bar{U}^{-T}\bar{U}^T M \bar{U}\bar{U}^{-1}\bar{V}^{-T}\right\|$$

$$\leq \left\|\bar{U}^T M \bar{U}\right\| \left|\bar{V}^{-1}\bar{U}^{-T}\bar{U}^{-1}\bar{V}^{-T}\right| \qquad \text{(from A.10)}$$

$$= ||M||_{\bar{S}} \left|\frac{1}{\bar{\omega}}(I - R)^{-1}\right|$$

$$\leq \frac{1}{\bar{\omega}(1-\gamma)} ||M||_{\bar{S}} \quad \blacksquare$$

A.12 If $D, G, H \in S^{n \times n}$, and $(D + G) \bullet (D - H) = 0$, then $||D|| \leq ||G|| + ||H||$.

Proof: $||D||^2 - ||D||\,||G - H|| - G \bullet H \leq ||D||^2 + D \bullet (G - H) - G \bullet H$

$$= (D + G) \bullet (D - H) = 0$$

Therefore, from the quadratic formula,

$$||D|| \leq \frac{||G-H|| + \sqrt{||G-H||^2 + 4G \bullet H}}{2} = \frac{||G-H|| + ||G+H||}{2} \leq ||G|| + ||H||. \quad \blacksquare$$

A.13 If $0 \prec \bar{V}\bar{V}^T = \bar{N}\bar{N}^T$ and $\bar{S} \succeq 0$, then

$$\left\|I - \left(\frac{1}{\varpi}\right)\bar{V}^T\bar{S}\bar{V}\right\| = \left\|I - \left(\frac{1}{\varpi}\right)\bar{N}^T\bar{S}\bar{N}\right\|.$$

Proof: $$\left\|I - \left(\frac{1}{\varpi}\right)\bar{V}^T\bar{S}\bar{V}\right\|^2 = \mathrm{tr}\left(\left(I - \frac{1}{\varpi}\bar{V}^T\bar{S}\bar{V}\right)\left(I - \frac{1}{\varpi}\bar{V}^T\bar{S}\bar{V}\right)\right)$$

$$= \mathrm{tr}\left(I - \frac{2}{\varpi}\bar{V}^T\bar{S}\bar{V} + \left(\frac{1}{\varpi}\right)^2\bar{V}^T\bar{S}\bar{V}\bar{V}^T\bar{S}\bar{V}\right)$$

$$= \mathrm{tr}\left(I - \frac{2}{\varpi}\bar{S}\bar{V}\bar{V}^T + \left(\frac{1}{\varpi}\right)^2\bar{S}\bar{V}\bar{V}^T\bar{S}\bar{V}\bar{V}^T\right)$$

$$= \mathrm{tr}\left(I - \frac{2}{\varpi}\bar{S}\bar{N}\bar{N}^T + \left(\frac{1}{\varpi}\right)^2\bar{S}\bar{N}\bar{N}^T\bar{S}\bar{N}\bar{N}^T\right)$$

$$= \mathrm{tr}\left(I - \frac{2}{\varpi}\bar{N}^T\bar{S}\bar{N} + \left(\frac{1}{\varpi}\right)^2\bar{N}^T\bar{S}\bar{N}\bar{N}^T\bar{S}\bar{N}\right) = \left\|I - \frac{1}{\varpi}\bar{N}^T\bar{S}\bar{N}\right\|^2 \quad \blacksquare$$

A.14 $\|A^TA\| = \|AA^T\|$ and $|A^TA| = |AA^T|$.

Proof: $\|A^TA\|^2 = \mathrm{tr}(A^TAA^TA) = \mathrm{tr}(AA^TAA^T)$ (from A.1 and (A.2)

$$= \mathrm{tr}(AA^T)^T(AA^T) = \|AA^T\|^2.$$

To show the second equality, it suffices to show that A^TA and AA^T have the same nonzero eigenvalues. To see this, let $\theta \neq 0$ be an eigenvalue of A^TA. Then there exists $\upsilon \neq 0$ such that $A^TA\upsilon = \theta\upsilon$. Thus $AA^T(A\upsilon) = \theta(A\upsilon)$, and so θ is an eigenvalue of AA^T. ■

A.15 If $A \succeq 0$ and B is symmetric, $A \bullet B \leq \mathrm{tr}(A)\,|B|$.

Proof: Let $A = QDQ^T$, where $Q^{-1} = Q^T$ and D is a nonnegative diagonal matrix.

Then $$A \bullet B = \mathrm{tr}(QDQ^TB) = \mathrm{tr}(DQ^TBQ)$$

$$= \sum_i D_{ii}(Q^TBQ)_{ii}.$$

However, $\left(Q^T B Q\right)_{ii} \leq \left|Q^T B Q\right| = |B|$, from A.5.

Therefore $A \bullet B = \sum_i D_{ii}\left(Q^T B Q\right)_{ii} \leq \sum_i D_{ii}|B| = \operatorname{tr}(D)|B| = \operatorname{tr}(A)|B|$. ■

References

[1] F. Alizadeh, "Combinatorial optimization with interior point methods and semi-definite matrices", Ph.D. thesis, Computer Science Department, University of Minnesota, Minneapolis, Minnesota, 1991.

[2] F. Alizadeh, "Optimization over the positive-definite cone: interior point methods and combinatorial applications," in *Advances in Optimization and Parallel Computing,* P. Pardalos, ed., North-Holland, 1992.

[3] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization," to appear in *SIAM Journal on Optimization,* 1994.

[4] F. Alizadeh, J. P. Haeberly, and M. Overton, "A new primal-dual interior-point method for semidefinite programming," Technical Report 659, March, 1994.

[5] F. Alizadeh, J. P. Haeberly, and M. Overton, "Primal-dual interior-point method for semidefinite programming," August, 1994.

[6] J. Borwein and H. Wolkowicz, "Facial reduction for a cone-convex programming problem," *J. Austral. Math. Soc.* (Series A) 30 (1981) 369-380.

[7] J. Borwein and H. Wolkowicz, "Characterization of optimality for the abstract convex program with finite dimensional range," *J. Austral. Math. Soc.* (Series A) 30 (1981) 390-411.

[8] J. Borwein and H. Wolkowicz, "Regularizing the abstract convex program," *Journal of Mathematical Analysis and Applications* 83 2 (1981) 495-530.

[9] J. Borwein and H. Wolkowicz, "Some applications of optimization in matrix theory," *Linear Algebra and its Applications* 40:101-118 (1981).

[10] S. Boyd and L. El Ghaoui, "Method of centers for minimizing generalized eigenvalues", *Linear Algebra and its Applications*, 188, 189: 63-111 (1993).

[11] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory,* SIAM, 1994.

[12] C. Delorme and S. Poljak, "Laplacian eigenvalues and the maximum cut problem," *Mathematical Programming* 62 (1993) 557-574.

[13] R. M. Freund, 1993, "An infeasible-start algorithm for linear programming whose complexity depends on the distance from the starting point to the optimal solution," to appear in *Annals of Operations Research.*

[14] M. Goemans and D. Williamson, ".878-Approximation algorithms for MAX CUT and MAX 2SAT," (1994), Department of Mathematics, MIT, Cambridge, Mass.

[15] F. Jarre, "An interior point method for minimizing the maximum eigenvalue of a linear combination of matrices," *SIAM J. on Control and Optimization* 31 5 (1993) 1360-1377.

[16] C. Karisch, F. Rendl, and H. Wolkowicz, "Quadratic Lagrangean relaxation for the quadratic assignment problem," University of Waterloo Research Report, Waterloo, Ontario, in progress.

[17] L. Lovasz and A. Schrijver, "Cones of matrices and set-functions, and 0-1 optimization," *SIAM J. Optimization*, 1 (1991) 166-190.

[18] M. Kojima, 1993, "Basic lemmas in polynomial-time infeasible-interior-point methods for linear programs," to appear in *Annals of Operations Research.*

[19] M. Kojima, S. Shindoh, and S. Hara, "Interior-point methods for the monotone linear complementarity problem in symmetric matrices," Research Report, Department of Information Sciences Tokyo Institute of Technology, Tokyo, Japan, April, 1994.

[20] S. Mizuno, "Polynomiality of the Kojima-Megiddo-Mizuno infeasible interior point algorithm for linear programming," Technical Report no. 1006, School of Industrial Engineering and Operations Research, Cornell University, Ithaca, NY, 1992.

[21] Y. Nesterov and A. Nemirovskii, *Interior point polynomial methods in convex programming: Theory and Applications*, SIAM, 1993.

[22] Y. Nesterov and M. Todd, "Self-scaled cones and interior-point methods in nonlinear programming," working paper, CORE, Catholic University of Louvain, Louvain-la-Neuve, Belgium, April, 1994.

[23] F. Rendl, R. Vanderbei, and H. Wolkowicz, "Interior-point methods for max-min eigenvalue problems," Technical Report SOR-93-15, Princeton University, July, 1993.

[24] F. Rendl, R. Vanderbei, and H. Wolkowicz, "Max-min eigenvalue problems, primal-dual interior point algorithms, and trust region subproblems," CORR 93-30, University of Waterloo, November, 1993.

[25] J. Renegar, private communication, 1994.

[26] M. Todd and Y. Ye, "Approximate Farkas Lemmas and Stopping Rules for Iterative Infeasible-Point Algorithms for Linear Programming," manuscript, November, 1994.

[27] L. Vandenberghe and S. Boyd, "A primal-dual potential reduction method for problems involving matrix inequalities," 1993, to appear in *Mathematical Programming Series B.*

[28] L. Vandenberghe and S. Boyd, "A polynomial-time algorithms for determining quadratic Lyapunov functions for nonlinear systems," in Circuit Theory and Design, H. Dedieu, editor, ECCTD '93, Elseveier Science Publishers, 1993, 1065-1068.