Semidefinite Programming in the Space of Partial Positive Semidefinite Matrices

Samuel Burer*

May 2, 2002
Revised: May 29, 2002
Revised: October 24, 2002
Revised: January 16, 2003

Abstract

We build upon the work of Fukuda et al. [9] and Nakata et al. [26], in which the theory of partial positive semidefinite matrices has been applied to the semidefinite programming (SDP) problem as a technique for exploiting sparsity in the data. In contrast to their work, which improves an existing algorithm that is based on the HRVW/KSH/M search direction, we present a primal-dual path-following algorithm that is based on a new search direction, which, roughly speaking, is defined completely within the space of partial symmetric matrices. We show that the proposed algorithm computes a primal-dual solution to the SDP problem having duality gap less than a fraction $\varepsilon > 0$ of the initial duality gap in $\mathcal{O}(n \log(\varepsilon^{-1}))$ iterations, where n is the size of the matrices involved. Moreover, we present computational results showing that the algorithm possesses several advantages over other existing implementations.

Keywords: Semidefinite programming, sparsity, matrix completion, numerical experiments.

1 Introduction

The semidefinite programming (SDP) problem has been studied extensively in recent years, and many different types of algorithms for solving SDPs have been proposed. Various primal-dual interior-point methods for linear programming can be extended to SDP with equivalent iteration complexities, typically $\mathcal{O}(\sqrt{n}\log(\varepsilon^{-1}))$, where n is the size of matrices in the SDP problem and $\varepsilon > 0$ is the desired fractional reduction in the duality gap. For example, see [1, 15, 17, 22, 23, 25, 28, 30, 32]. In practice, these methods have many advantages, including applicability to any standard form SDP, accurate primal-dual optimal solutions in a small number of iterations, and exploitation of sparsity in certain key stages of the algorithm. On the other hand, they also exhibit some notable disadvantages, such as the need to compute, store, and work with dense matrices — in particular, handling the $n \times n$ primal iterate X and the $m \times m$ Schur complement matrix M, where m is the number of linear constraints in the primal SDP, as well as solving the Schur complement equation involving M.

Techniques for dealing with the disadvantages of primal-dual methods have also been developed. For example, to avoid working with the dense matrix X (while maintaining the use of M), Benson et al. [2] have developed a polynomial-time interior-point method that involves only the dual variables

*Department of Management Sciences, University of Iowa, Iowa City, IA 52242-1000, USA. (Email: samuel-burer@uiowa.edu). Based on research supported by the National Science Foundation under grant CCR-0203426.

(S, y) and the lower Cholesky factor L of S, since S and L are generally sparse when the SDP data is sparse. In contrast, others have eliminated the need to compute and store M (while maintaining the use of primal-dual iterates (X, S, y)) by using iterative methods such as the preconditioned conjugate gradient method to solve the Schur complement equation (see [20, 27, 29]). When solving the Schur complement equation using an iterative method, however, an inevitable side effect is the increased difficulty of obtaining accurate primal-dual optimal solutions due to the ill-conditioning of the matrix near optimality.

Other methods, the so-called first-order nonlinear programming algorithms for SDP, depart even more significantly from the standard primal-dual interior-point methods. Generally speaking, these methods solve special classes of SDPs, work in either the primal or dual space, operate on sparse matrices (or compact representations of dense matrices), and sacrifice the underlying theoretical guarantee of polynomial convergence for better opportunities to exploit sparsity and structure. As a result of these algorithmic choices as well as the ill-conditioning that is inherent near optimality, these methods typically can compute optimal solutions of low to medium accuracy in a reasonable balance of iterations and time. See [4, 6, 5, 8, 13, 14] for more information on this class of algorithms.

So far, no one has proposed a method that possesses theoretical polynomial convergence, can solve any standard form SDP, works in both the primal and dual spaces, and can aggressively exploit sparsity in all stages of computation, including the complete avoidance of dense matrices. In this paper, we propose such a method and explore its theoretical and practical characteristics.

The basic idea of the method presented in this paper is drawn from the recent work of Fukuda et al. [9], in which they show that the theory of partial positive semidefinite matrices can be applied to SDPs to help better take advantage of sparsity. In particular, their "completion method" demonstrates that primal-dual interior-point algorithms can be implemented using a certain "partial" representation of the dense matrix variable X. Computational results given in Nakata et al. [26], which employ the sparse representation of X together with the computation and storage of M in each iteration, indicate the efficiency of the completion method on several classes of problems.

The completion method can be viewed as a computational enhancement of an existing primal-dual path-following implementation that is based on the HRVW/KSH/M search direction (which was first defined in Helmberg et al. [15]). From a theoretical point of view, however, the completion method is not known to converge in polynomial time, with the main obstacle being how to measure the proximity of a partial primal-dual solution to the central path. (See the concluding comments of section 5 in [9], where a polynomial potential-reduction algorithm is discussed but the problem of a path-following algorithm is considered open.) In addition, since the completion method employs the Schur complement matrix M directly, there is a practical limitation to the size of SDP that can be solved by this method. Of course, a simple idea to eliminate the direct use of M would be to use an iterative method to solve the Schur complement equation.

The method of this paper improves upon the completion method of Fukuda et al. in two primary ways. The first is theoretical: the method is a polynomial-time path-following algorithm based entirely on partial positive semidefinite matrices, where the main idea is a reformulation of the central path that yields search directions in the space of partial matrices and that also motivates a new neighborhood of the central path, which has some critical properties when viewed in the context of matrix completion. The second is practical: when the Schur complement equation in our method is solved using an iterative method, our approach provides even more opportunity to take advantage of the sparsity of the SDP data. In Section 5, computational results are given to demonstrate this.

Computational results are also given comparing our method with two other successful methods: a primal-dual interior-point method that possesses polynomial convergence but computes and stores

both X and M; and a dual-only first-order algorithm that works exclusively with sparse matrices but does not possess polynomial convergence. The overall conclusion of this paper is that our method achieves several advantages that were previously found only in separate algorithms: theoretically strong convergence, applicability to any SDP, a primal-dual framework, and the opportunity to exploit sparsity in all stages of computation.

1.1 Basic notation and terminology

In this paper, $\Re$, $\Re^p$, and $\Re^{p\times q}$ denote the typical Euclidean spaces, and $\mathcal{S}^p$, $\mathcal{S}^p_+$ and $\mathcal{S}^p_{++}$ denote symmetric, symmetric positive semidefinite, and symmetric positive definite matrices, respectively. Lower triangular matrices are denoted by $\mathcal{L}^p$, and those with positive diagonal entries are signified by $\mathcal{L}^p_{++}$. Similarly, we define $\mathcal{U}^p$ and $\mathcal{U}^p_{++}$ for upper triangular matrices.

For any $v \in \Re^p$, v_i is the i-th component of v; for any $A \in \Re^{p\times q}$, A_{ij}, $A_{i\cdot}$, $A_{\cdot j}$ denote the standard subparts of A. For vectors or matrices, the notation $\cdot^T$ denotes transpose, and for any $A \in \Re^{p\times p}$, $\mathrm{tr}(A) = \sum_{i=1}^p A_{ii}$ denotes the trace function. The standard inner product on $\Re^{p\times q}$ is denoted as $A \bullet B = \mathrm{tr}(A^T B) = \sum_{i=1}^p \sum_{j=1}^q A_{ij}B_{ij}$, and the Frobenius norm, which is induced by the inner product $\bullet$, is denoted $\|A\|_F = (A \bullet A)^{1/2}$. For any $A \in \mathcal{S}^p$ (which necessarily has real eigenvalues), we note that $\mathrm{tr}(A)$ also equals the sum of the eigenvalues of A. The maximum and minimum eigenvalues of A are denoted by $\lambda_{\max}[A]$ and $\lambda_{\min}[A]$. For any $A \in \Re^{p\times q}$, we define the 2-norm of A to be $\|A\| = \sqrt{\lambda_{\max}[A^T A]}$. Some important inequalities are as follows: for all $A, B \in \Re^{p\times q}$, $A \bullet B \le \|A\|_F \|B\|_F$, and for all $A \in \Re^{p\times q}$, $\|A\| \le \|A\|_F$.

If a function f defined between $\Re^p$ and $\Re^q$ is twice differentiable at $v \in \Re^p$, then the first derivative of f at v is denoted as the linear operator $f'(v) : \Re^p \to \Re^q$, which operates on $a \in \Re^p$ as $f'(v)[a]$. In addition, if $q = 1$, then the gradient $\nabla f(v) \in \Re^p$ uniquely satisfies $\nabla f(v)^T a = f'(v)[a]$ for all $a \in \Re^p$. The second derivative of f at v is denoted by the bilinear map $f''(v) : \Re^p \times \Re^p \to \Re^q$, which is written as $f''(v)[a, b]$ for all $(a, b) \in \Re^p \times \Re^p$. If $b = a$, then we write $f''(v)[a]^{(2)}$. If f is also invertible, then f^{-1} denotes its inverse, and if f is linear, then its adjoint is denoted by f^*. In addition, if $p = q$ and f is linear, then f is called positive definite if $v^T f(v) > 0$ for all $v \ne 0$.

The function $\mathrm{chol} : \mathcal{S}^p_{++} \to \mathcal{L}^p_{++}$ computes the lower Cholesky factor; that is, $\mathrm{chol}(S) = L$ where $S = LL^T$. Letting M^{-1} denote the matrix inverse of M and inv the matrix inverse function, we have that $\mathrm{inv}'(M)[A] = -M^{-1}AM^{-1}$. We also let argmax denote the unique optimal solution of a given maximization problem, and we define argmin similarly. The matrix I denotes the identity matrix (of appropriate size), and for any $A \in \Re^{p\times p}$, $\mathrm{diag}(A)$ extracts the diagonal of A. Also, for any $M \in \mathcal{S}^p_+$, $M^{1/2}$ denotes the matrix square root of M.

2 The Partial SDP Problem

We consider the standard-form primal SDP problem

$$(\hat{P}) \qquad \min \left\{ C \bullet \hat{X} : \mathcal{A}(\hat{X}) = b,\ \hat{X} \in \mathcal{S}^n_+ \right\}$$

and its dual

$$(\hat{D}) \qquad \max \left\{ b^T y : \mathcal{A}^*(y) + \hat{S} = C,\ \hat{S} \in \mathcal{S}^n_+ \right\},$$

where the variables are $(\hat{X}, \hat{S}, y) \in \mathcal{S}^n \times \mathcal{S}^n \times \Re^m$ and the data are $C \in \mathcal{S}^n$, $b \in \Re^m$, and $\{A_k\}_{k=1}^m \subset \mathcal{S}^n$. The symbol $\mathcal{A}$ denotes the linear map $\mathcal{A} : \mathcal{S}^n \to \Re^m$ defined by $\mathcal{A}(\hat{X})_k = A_k \bullet \hat{X}$, and its adjoint $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ is defined by $\mathcal{A}^*(y) = \sum_{k=1}^m y_k A_k$. We use similar notation for the sets of

primal-dual feasible solutions and primal-dual interior feasible solutions as in [22] — $\mathcal{F}(\hat{P}) \times \mathcal{F}(\hat{D})$ and $\mathcal{F}^0(\hat{P}) \times \mathcal{F}^0(\hat{D})$, respectively — and we also make the following standard assumptions:

$\hat{\mathbf{A}}$1. the matrices $\{A_k\}_{k=1}^m$ are linearly independent;

$\hat{\mathbf{A}}$2. the set of primal-dual interior feasible solutions $\mathcal{F}^0(\hat{P}) \times \mathcal{F}^0(\hat{D})$ is nonempty.

It is well-known that, under assumptions Â1 and Â2, both $(\hat{P})$ and $(\hat{D})$ have optimal solutions $\hat{X}^*$ and $(\hat{S}^*, y^*)$, which are characterized by the equivalent conditions that the duality gap $\hat{X}^* \bullet \hat{S}^*$ is zero and that the matrix product $\hat{X}^*\hat{S}^*$ is zero. Moreover, for every $\nu > 0$, there exists a unique primal-dual feasible solution $(\hat{X}_\nu, \hat{S}_\nu, y_\nu)$, which satisfies the perturbed optimality equation $\hat{X}\hat{S} = \nu I$. The set of all solutions $\hat{\mathcal{C}} \equiv \{(\hat{X}_\nu, \hat{S}_\nu, y_\nu) : \nu > 0\}$ is known as the central path, and $\hat{\mathcal{C}}$ serves as the basis for path-following algorithms that solve $(\hat{P})$ and $(\hat{D})$. The basic idea is to construct a sequence $\{(\hat{X}^k, \hat{S}^k, y^k)\}_{k\geq 0} \subset \mathcal{F}^0(\hat{P}) \times \mathcal{F}^0(\hat{D})$ that stays in a neighborhood of the central path such that the duality gap $\hat{X}^k \bullet \hat{S}^k$ goes to zero.

A scaled measure of the duality gap that proves useful in the presentation and analysis of path-following algorithms is

$$\mu(\hat{X}, \hat{S}) \equiv \frac{\hat{X} \bullet \hat{S}}{n}, \qquad \forall \ (\hat{X}, \hat{S}) \in \mathcal{S}^n \times \mathcal{S}^n. \tag{1}$$

Note that, for all $(\hat{X}, \hat{S}) \in \mathcal{S}^n_+ \times \mathcal{S}^n_+$, we have $\mu(\hat{X}, \hat{S}) > 0$ unless $\hat{X}\hat{S} = 0$. Moreover, $\mu(\hat{X}_\nu, \hat{S}_\nu) = \nu$ for all points $(\hat{X}_\nu, \hat{S}_\nu, y_\nu)$ on the central path.

2.1 The positive semidefinite matrix completion

Recently, Fukuda et al. [9] introduced techniques for exploiting sparsity using ideas from the theory of matrix completions. In this section, we recapitulate their main results and introduce corresponding notation that will be used throughout the paper.

Let $V = \{1, \ldots, n\}$ denote the row and column indices of an $n \times n$ matrix. Also define the *aggregate density pattern* E of the data $\{C\} \cup \{A_k\}_{k=1}^m$ as follows:

$$E \equiv \{(i,j) \in V \times V : \ \exists \ \ y \in \Re^m \ \text{ such that } \ [C - \mathcal{A}^*(y)]_{ij} \neq 0 \ \}.$$

We assume throughout that $\{(i,i) : i \in V\} \subseteq E$, that is, E contains all of the diagonal entries. Notice also that E is symmetric in the sense that $(i,j) \in E$ if and only if $(j,i) \in E$ because, by definition, $C - \mathcal{A}^*(y) \in \mathcal{S}^n$. (We also remark that the alternative terminology "aggregate sparsity pattern" has been used in [9] to describe E.)

Given any $(\hat{S}, y) \in \mathcal{F}(\hat{D})$, it is clear from the definition of E that those elements of $V \times V$, which correspond to the nonzeros of $\hat{S}$, are contained in E. Hence, $\bar{E} \equiv V \times V \setminus E$ represents the generic sparsity pattern of the variable $\hat{S}$ of $(\hat{D})$. Unlike $\hat{S}$, the variable $\hat{X}$ of $(\hat{P})$ has no sparsity in general, but the sparsity represented by $\bar{E}$ does affect the primal problem in terms of evaluation of the objective function $C \bullet \hat{X}$ and the constraints $\mathcal{A}(\hat{X})$. In particular, it is not difficult to see that the quantities $C \bullet \hat{X}$ and $A_k \bullet \hat{X}$ are dependent upon only those entries $\hat{X}_{ij}$ of $\hat{X}$ where $(i,j) \in E$. In other words, the entries $\hat{X}_{ij}$ for $(i,j) \in \bar{E}$ are irrelevant for the objective function and constraints, but still, they do impact the positive semidefiniteness constraint $\hat{X} \in \mathcal{S}^n_+$. These were precisely the observations that were exploited in [9], as we detail next.

Given a symmetric $G \subseteq V \times V$, we define the following subset of $\mathcal{S}^n$, which has the density pattern G:

$$\mathcal{S}^G \equiv \{\hat{M} \in \mathcal{S}^n : \hat{M}_{ij} = 0 \text{ for all } (i,j) \notin G\}.$$

We also define the corresponding operator $\pi^G : \mathcal{S}^n \to \mathcal{S}^G$ which performs orthogonal projection onto $\mathcal{S}^G$:

$$[\pi^G(\hat{M})]_{ij} = \begin{cases} \hat{M}_{ij} & (i,j) \in G, \\ 0 & (i,j) \notin G. \end{cases}$$

We then define the following subsets of $\mathcal{S}^G$:

$$\begin{aligned} \mathcal{S}^G_+ &= \mathcal{S}^G \cap \mathcal{S}^n_+, \\ \mathcal{S}^G_{++} &= \mathcal{S}^G \cap \mathcal{S}^n_{++}, \\ \mathcal{S}^{G?}_+ &= \{M \in \mathcal{S}^G : \ \exists\ \hat{M} \in \mathcal{S}^n_+ \text{ such that } \pi^G(\hat{M}) = M\}, \\ \mathcal{S}^{G?}_{++} &= \{M \in \mathcal{S}^G : \ \exists\ \hat{M} \in \mathcal{S}^n_{++} \text{ such that } \pi^G(\hat{M}) = M\}. \end{aligned}$$

In words, we describe the last two sets defined above as follows: $\mathcal{S}^{G?}_+$ and $\mathcal{S}^{G?}_{++}$ consist of those matrices in $\mathcal{S}^G$ that can be completed to matrices in $\mathcal{S}^n_+$ and $\mathcal{S}^n_{++}$, respectively. We use the question mark (?) notation to illustrate the informal idea that, for example, $M \in \mathcal{S}^{G?}_+$ is a positive semidefinite matrix except that the entries M_{ij} for $(i,j) \notin G$ have yet to be specified. In addition, an important observation is that $\mathcal{S}^{G?}_{++}$ is an open subset of $\mathcal{S}^G$, which will play an important role when we investigate the derivatives of functions defined on $\mathcal{S}^{G?}_{++}$.

Using these ideas from matrix completion along with the discussion of E above, it is not difficult to see that problems $(\hat{P})$ and $(\hat{D})$ are equivalent to the following two problems, respectively:

$$\min\left\{C \bullet X : \mathcal{A}(X) = b,\ X \in \mathcal{S}^{E?}_+\right\}, \qquad \max\left\{b^T y : \mathcal{A}^*(y) + S = C,\ S \in \mathcal{S}^E_+\right\}.$$

It is interesting to note that the above equivalence holds even when E is replaced by any symmetric $F \supseteq E$. In fact, for technical reasons that will become clear later, it is desirable to apply this idea with an F that satisfies specific structural properties, as discussed next.

It is straightforward to identify a symmetric $G \subseteq V \times V$ with a simple graph $\tilde{G}$ on V, and we make the following graph theoretic definitions. G is said to be *chordal* if $\tilde{G}$ is chordal, that is, every cycle in $\tilde{G}$ having length greater than three has a chord. A *perfect elimination ordering* for G is an ordering $(v_1, \ldots, v_n)$ of the vertices $V = \{1, \ldots, n\}$ of $\tilde{G}$ such that, for each $1 \le i \le n-1$, the $\tilde{G}$-neighbors of v_i in $\{v_{i+1}, \ldots, v_n\}$ form a clique in $\tilde{G}$. A fundamental fact (see Fulkerson and Gross [10]) is that G is chordal if and only if it has a perfect elimination ordering.

Now let F be a symmetric extension of E, i.e., $F \supseteq E$, that satisfies two properties: (i) F is chordal; and (ii) the standard ordering $(1, \ldots, n)$ is a perfect elimination ordering for F. We then define the pair of SDP problems

$$(P) \qquad \min\left\{C \bullet X : \mathcal{A}(X) = b,\ X \in \mathcal{S}^{F?}_+\right\}$$

and

$$(D) \qquad \max\left\{b^T y : \mathcal{A}^*(y) + S = C,\ S \in \mathcal{S}^F_+\right\},$$

which, from the discussion above, are equivalent to $(\hat{P})$ and $(\hat{D})$, respectively. It is worthwhile to note that, under the assumption that no numerical cancellations occur during the calculation of the lower Cholesky factor $L \in \mathcal{L}^n_{++}$ of $S \in \mathcal{S}^E_{++}$, the symmetrized density pattern of L yields a chordal extension F of E such that $(1, \ldots, n)$ is a perfect elimination ordering.

How do the standard assumptions Â1 and Â2 translate to the problems (P) and (D)? It is not difficult to see that, with the analogous definitions $\mathcal{F}(P) \times \mathcal{F}(D)$ and $\mathcal{F}^0(P) \times \mathcal{F}^0(D)$, both of the following assumptions hold easily:

A1. the matrices $\{A_k\}_{k=1}^m$ are linearly independent;

A2. the set of interior feasible solutions $\mathcal{F}^0(P) \times \mathcal{F}^0(D)$ is nonempty.

2.2 The partial central path

A critical observation is that the central path equation $\hat{X}\hat{S} = \nu I$ implies $\hat{X}^{-1}$ has the same density pattern as $\hat{S}$. That is, $\hat{X}^{-1} \in \mathcal{S}^F$, as was proven by Grone et al. in [12] (and used extensively by Fukuda et al. in [9]). This observation represents an important connection between the spaces $\mathcal{S}^{F?}_{++}$ and $\mathcal{S}^n_{++}$:

Theorem 2.1 *Let $X \in \mathcal{S}^{F?}_{++}$. Then there exists a unique $\hat{X} \in \mathcal{S}^n_{++}$ satisfying $\pi^F(\hat{X}) = X$ and $\hat{X}^{-1} \in \mathcal{S}^F$. Moreover,*

$$\hat{X} = \mathit{argmax}\left\{\det(\hat{Y}) : \ \pi^F(\hat{Y}) = X, \ \hat{Y} \in \mathcal{S}^n_{++}\right\}.$$

As in [9], we call $\hat{X}$ the *maximum-determinant positive definite completion* of X, and we also let $\hat{X} : \mathcal{S}^{F?}_{++} \to \mathcal{S}^n_{++}$ denote the function that yields $\hat{X}$ from X, that is, $\hat{X} \equiv \hat{X}(X)$. Using the function $\hat{X}$ and the direct correspondence $\hat{S} = S$ between the spaces of problems $(\hat{D})$ and (D), the central path equation $\hat{X}\hat{S} = \nu I$ can be described in terms of X and S as $\hat{X}(X)S = \nu I$.

For the algorithm presented in this paper, we wish to reformulate the central path equation $\hat{X}(X)S = \nu I$ once more, and so we now introduce some notation and a few small results. We define the following sets of lower triangular matrices, each of which have a density pattern equal to the lower triangular part of F:

$$\begin{aligned}\mathcal{L}^F &\equiv \{L \in \mathcal{L}^n : \ L_{ij} = 0 \text{ for all } i \geq j \text{ such that } (i,j) \notin F\}, \\ \mathcal{L}^F_{++} &\equiv \mathcal{L}^F \cap \mathcal{L}^n_{++}.\end{aligned}$$

Noting the standard fact that the Cholesky factorization has no fill-in when the associated density pattern is chordal and $(1, \ldots, n)$ is a perfect elimination ordering, we see that, for all $(X, S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++}$, $\mathrm{chol}(\hat{X}(X)^{-1}) \in \mathcal{L}^F_{++}$ and $\mathrm{chol}(S) \in \mathcal{L}^F_{++}$. We thus define

$$V : \mathcal{S}^{F?}_{++} \to \mathcal{L}^F_{++}, \qquad V(X) \equiv \mathrm{chol}(\hat{X}(X)^{-1}), \tag{2}$$

$$L : \mathcal{S}^F_{++} \to \mathcal{L}^F_{++}, \qquad L(S) \equiv \mathrm{chol}(S). \tag{3}$$

Using these definitions, it is now possible to further reformulate the central path equation $\hat{X}(X)S = \nu I$.

Proposition 2.2 *Let $(X, S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++}$ and $\nu > 0$. Then $\hat{X}(X)S = \nu I$ if and only if $V(X)^{-1}L(S) = \sqrt{\nu}\, I$.*

Proof. Let $\hat{X} \equiv \hat{X}(X)$, $V \equiv V(X)$, and $L \equiv L(S)$. From (2) and (3), we see that $\hat{X}S = \nu I$ is equivalent to $V^{-1}LL^TV^{-T} = \nu I$, which itself shows that $V^{-1}L$ is the lower Cholesky factor of νI, that is, $V^{-1}L = \sqrt{\nu}\, I$. ∎

Proposition 2.2 now allows us to characterize the point (X_ν, S_ν, y_ν) on the central path $\mathcal{C}$ corresponding to $\nu > 0$ as the unique solution of the system

$$\mathcal{A}(X) = b, \tag{4a}$$

$$\mathcal{A}^*(y) + S = C, \tag{4b}$$

$$V(X)^{-1}L(S) = \sqrt{\nu}\, I. \tag{4c}$$

Having expressed the central path in terms of the variables (X, S), we now wish to express the duality gap in terms of X and S as well. Given $(\hat{X}, \hat{S})$ and defining $(X, S) = (\pi^F(\hat{X}), \hat{S})$, we have

$$\hat{X} \bullet \hat{S} = \hat{X} \bullet S = \pi^F(\hat{X}) \bullet S = X \bullet S.$$

Alternatively, given (X, S) and letting $\hat{X}$ be any completion of X, we see that the equality also holds. Hence, $X \bullet S$ is the appropriate measure of the duality gap in the space $\mathcal{F}(P) \times \mathcal{F}(D)$. Furthermore, from (1) and the above equality, we have $\mu(\hat{X}, \hat{S}) = \mu(X, S)$.

The equation of the previous paragraph introduces a simple but important idea that will be used several times throughout the paper, and so we now give it a verbal description in order to make it simpler to refer to. Given $A, B \in \mathcal{S}^F$ and $\hat{A} \in \mathcal{S}^n$ such that $\pi^F(\hat{A}) = A$, we see that $\hat{A} \bullet B = \pi^F(\hat{A}) \bullet B = A \bullet B$, and we say that $(\hat{A}, A, B)$ are *trace-compatible*.

Given our usage of the equation (4c) in this paper, we also wish to define the square root of the scaled duality gap measure $\mu(\cdot, \cdot)$:

$$\rho(X, S) \equiv \mu(X, S)^{1/2}, \qquad \forall \ \ (X, S) \in \mathcal{S}_+^{F?} \times \mathcal{S}_+^F. \tag{5}$$

Note that, using the fact that $(\hat{X}(X), X, S)$ are trace-compatible, (2), (3), and (1), along with standard properties of the trace function and the Frobenius norm, we easily have that

$$\rho(X, S) = \left\| V(X)^{-1} L(S) \right\|_F / \sqrt{n}. \tag{6}$$

Equation (6) will come in handy throughout the presentation of this paper.

2.3 Nonsingularity of the partial central path

In Section 4, we will develop a primal-dual path-following algorithm based on the central path equations (4), and so in this subsection, we consider the nonsingularity of the Jacobian of the equations defining the central path, which will be necessary for the existence of the SDP direction proposed in Section 4.

Noting that $V(X)^{-1}$ is generically dense, it is not difficult to see that the left-hand sides of the equations (4) are not "square" since they map a point $\mathcal{S}_{++}^{F?} \times \mathcal{S}_{++}^F \times \Re^m$ to $\Re^m \times \mathcal{S}^F \times \mathcal{L}^n$. As has become standard in the SDP literature, however, we can reconfigure the central path equations to obtain a square system. In this case, we replace (4c) with $L(S) - \sqrt{\nu}\, V(X) = 0$, which yields a system of equations $H(X, S, y) = (0, 0, 0)$, where $H : \mathcal{S}_{++}^{F?} \times \mathcal{S}_{++}^F \times \Re^m \to \Re^m \times \mathcal{S}^F \times \mathcal{L}^F$ is given by

$$H(X, S, y) \equiv \begin{pmatrix} \mathcal{A}(X) \\ \mathcal{A}^*(y) + S \\ L(S) - \sqrt{\nu}\, V(X) \end{pmatrix}. \tag{7}$$

Note that the definition of H is dependent on $\nu > 0$, which we consider fixed in the subsequent discussion. We remark that (7) is reminiscent of the central path equation $\hat{S} - \nu \hat{X}^{-1} = 0$, where the usual complementarity equation $\hat{X}\hat{S} = \nu I$ has been reconfigured. This formulation is not appropriate for primal-dual path-following algorithms because, at any $(\hat{X}, \hat{S}, y) \in \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n \times \Re^m$, the resulting Newton direction $(\Delta\hat{X}, \Delta\hat{S}, \Delta y)$ has the property that $\Delta\hat{X}$ depends only on $\hat{X}$ and not on $(\hat{S}, y)$, i.e., the Newton direction is not "primal-dual." In fact, an analogous system for partial matrices uses $S - \nu \hat{X}(X)^{-1} = 0$ instead of $\hat{X}(X)S = \nu I$, but this system also suffers a similar drawback. Hence, we have chosen to model the central path as in (7), in part because, in Section 4, (7) will yield a primal-dual Newton direction due to the special structure of the functions L and V.

We now wish to investigate the Jacobian of H and to determine whether it is nonsingular (perhaps under suitable conditions). Since the derivative of H clearly depends on the derivatives of $V(\cdot)$ and $L(\cdot)$, we first describe these in the set of propositions and corollaries below (whose proofs are not difficult and are hence left to the reader).

Proposition 2.3 *Let $M \in \mathcal{S}^n_{++}$. Then the first derivative of chol$(\cdot)$ at M is given by the invertible, linear map chol$'(M) : \mathcal{S}^n \to \mathcal{L}^n$, which is defined by the following: for all $N \in \mathcal{S}^n$, $K' \equiv$ chol$'(M)[N] \in \mathcal{L}^n$ is the unique solution of the equation $N = K'K^T + K(K')^T$, where $K \equiv$ chol(M).*

Corollary 2.4 *Let $S \in \mathcal{S}^F_{++}$. Then the first derivative of $L(\cdot)$ at S is the invertible, linear map $L'(S) : \mathcal{S}^F \to \mathcal{L}^F$, which is defined by the following: for all $B \in \mathcal{S}^F$, $L' \equiv L'(S)[B] \in \mathcal{L}^F$ is the unique solution of the equation*

$$B = L'L^T + L(L')^T, \tag{8}$$

where $L \equiv L(S)$.

Proposition 2.5 *Let $X \in \mathcal{S}^{F?}_{++}$. Then the linear map $\hat{X}'(X) : \mathcal{S}^F \to \mathcal{S}^n$ is defined by the following: for all $A \in \mathcal{S}^F$, $\hat{X}' \equiv \hat{X}'(X)[A] \in \mathcal{S}^n$ uniquely satisfies the requirements*

$$\pi^F(\hat{X}') = A, \qquad \hat{X}^{-1}\hat{X}'\hat{X}^{-1} \in \mathcal{S}^F, \tag{9}$$

where $\hat{X} \equiv \hat{X}(X)$.

Corollary 2.6 *Let $X \in \mathcal{S}^{F?}_{++}$. Then the linear map $V'(X) : \mathcal{S}^F \to \mathcal{L}^F$ is defined by the following: for all $A \in \mathcal{S}^F$, $V' \equiv V'(X)[A] \in \mathcal{L}^F$ is the unique solution of the equation*

$$-\hat{X}^{-1}\hat{X}'\hat{X}^{-1} = V'V^T + V(V')^T, \tag{10}$$

where $V \equiv V(X)$, $\hat{X} = \hat{X}(X)$, and $\hat{X}' \equiv \hat{X}'(X)[A]$. In addition, $V'(X)$ is invertible.

Having described the derivatives of $V(\cdot)$ and $L(\cdot)$, we now turn to the derivative of H. From (7), we see that the linear map $H' : \mathcal{S}^F \times \mathcal{S}^F \times \Re^m \to \Re^m \times \mathcal{S}^F \times \mathcal{L}^F$ is defined by

$$H'(X,S,y)[A,B,c] = \begin{pmatrix} \mathcal{A}(A) \\ \mathcal{A}^*(c) + B \\ L'(S)[B] - \sqrt{\nu}\, V'(X)[A] \end{pmatrix}. \tag{11}$$

In Lemma 2.8, Corollary 2.9, and Theorem 2.10, we show that $H'(X,S,y)$ is invertible as long as the product $V(X)^{-1}L(S)$ is sufficiently close to some positive multiple of the identity matrix, but first, we need a technical lemma whose proof is straightforward that will prove useful below and also throughout the paper.

Lemma 2.7 *Let $J \in \mathcal{L}^n$. Then $\|J\|_F \leq \|J + J^T\|_F/\sqrt{2}$, with equality holding if and only if J is strictly lower triangular.*

Lemma 2.8 *Let $(X,S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++}$. Define $V \equiv V(X)$ and $L \equiv L(S)$, and let $V'(X)$ and $L'(S)$ be as in Corollaries 2.6 and 2.4, respectively. Then, for all $A \in \mathcal{S}^F$ and for all $\beta > 0$,*

$$-A \bullet \left(V'(X)^{-1} \circ L'(S)\right)[A] \geq \left\|L^{-1}AL^{-T}\right\|_F^2 \left((1+\sqrt{2})\beta^3 - \sqrt{2}(\beta + \|Q\|_F)^3\right), \tag{12}$$

where $Q \equiv V^{-1}L - \beta I$.

Proof. With $L' \equiv L'(S)[A]$, we see from (8) that

$$A = L'L^T + L(L')^T. \tag{13}$$

Also defining $\hat{X} \equiv \hat{X}(X)$ and $\hat{X}' \equiv \hat{X}'(X)[V'(X)^{-1}[L']]$, we see from (9) and (10) that

$$\pi^F(\hat{X}') = V'(X)^{-1}[L'], \qquad -\hat{X}^{-1}\hat{X}'\hat{X}^{-1} = L'V^T + V(L')^T. \tag{14}$$

Now using (14), (2), and the trace-compatibility of $(\hat{X}', V'(X)^{-1}[L'], A)$, we see that the left-hand side of (12) equals

$$\begin{aligned} -A \bullet V'(X)^{-1}[L'] = -A \bullet \hat{X}' &= A \bullet \hat{X}\left(L'V^T + V(L')^T\right)\hat{X} \\ &= 2\, A \bullet V^{-T}V^{-1}L'V^{-1}. \end{aligned} \tag{15}$$

Introducing the notation $\tilde{A} \equiv L^{-1}AL^{-T}$ and $\tilde{L} \equiv L^{-1}L'$ so that $\tilde{A} = \tilde{L} + \tilde{L}^T$ by (13) and using the definition of Q, we observe that

$$\begin{aligned} A \bullet V^{-T}V^{-1}L'V^{-1} &= \tilde{A} \bullet \left(V^{-1}L\right)^T \left(V^{-1}L\right) \tilde{L} \left(V^{-1}L\right) \\ &= \tilde{A} \bullet (Q + \beta I)^T (Q + \beta I)\, \tilde{L}\, (Q + \beta I)\,. \end{aligned} \tag{16}$$

Expanding the right-hand argument of the inner-product just obtained, we see that

$$\begin{aligned} (Q + \beta I)^T (Q + \beta I)\, \tilde{L}\, (Q + \beta I) = Q^TQ\tilde{L}Q + \beta Q^TQ\tilde{L} + \beta Q^T\tilde{L}Q + \beta^2 Q^T\tilde{L} + \\ \beta Q\tilde{L}Q + \beta^2 Q\tilde{L} + \beta^2\tilde{L}Q + \beta^3\tilde{L} \end{aligned} \tag{17}$$

Now combining (15), (16), and (17), applying Lemma 2.7 with $J = \tilde{L}$, and using standard properties of the trace function and the Frobenius norm, we have

$$\begin{aligned} -A \bullet V'(X)^{-1}[L'] &= 2\,\tilde{A} \bullet Q^TQ\tilde{L}Q + 2\,\beta\,\tilde{A} \bullet Q^TQ\tilde{L} + 2\,\beta\,\tilde{A} \bullet Q^T\tilde{L}Q + 2\,\beta^2\,\tilde{A} \bullet Q^T\tilde{L} + \\ &\qquad 2\,\beta\;\tilde{A} \bullet Q\tilde{L}Q + 2\,\beta^2\,\tilde{A} \bullet Q\tilde{L} + 2\;\tilde{A} \bullet \beta^2\tilde{L}Q + 2\,\beta^3\,\tilde{A} \bullet \tilde{L} \\ &\geq \beta^3\|\tilde{A}\|_F^2 - \sqrt{2}\,\|\tilde{A}\|_F^2\|Q\|_F^3 - 3\sqrt{2}\,\beta\,\|\tilde{A}\|_F^2\|Q\|_F^2 - 3\sqrt{2}\,\beta^2\,\|\tilde{A}\|_F^2\|Q\|_F \\ &= \|\tilde{A}\|_F^2\left((1+\sqrt{2})\beta^3 - \sqrt{2}\left(\|Q\|_F^3 + 3\,\beta\,\|Q\|_F^2 + 3\,\beta^2\,\|Q\|_F + \beta^3\right)\right) \\ &= \|\tilde{A}\|_F^2\left((1+\sqrt{2})\beta^3 - \sqrt{2}\left(\|Q\|_F + \beta\right)^3\right), \end{aligned}$$

which proves the lemma. (Note that, within the inequality, we have also used the equality $2\,\beta^3\,\tilde{A} \bullet \tilde{L} = \beta^3\,\|\tilde{A}\|_F^2$.) ∎

Corollary 2.9 *Let $(X, S) \in \mathcal{S}_{++}^{F?} \times \mathcal{S}_{++}^F$. Define $V \equiv V(X)$ and $L \equiv L(S)$, and let $V'(X)$ and $L'(S)$ be as in Corollaries 2.6 and 2.4, respectively. Then, if $\|V^{-1}L - \beta\, I\|_F \leq \beta/6$ for some $\beta > 0$, the operator $-V'(X)^{-1} \circ L'(S)$ is positive definite.*

Proof. By (12), $-V'(X)^{-1} \circ L'(S)$ is positive definite as long there exists some β such that

$$(1+\sqrt{2})\beta^3 - \sqrt{2}\left(\|Q\|_F + \beta\right)^3 > 0,$$

where $Q \equiv V^{-1}L - \beta I$, which is equivalent to

$$\|Q\|_F < \left[2^{-1/6}(1+\sqrt{2})^{1/3} - 1\right]\beta.$$

Since the coefficient in front of β is approximately 0.1951, the result follows. ∎

Corollary 2.9 now allows us to prove that H' is nonsingular under certain conditions.

Theorem 2.10 *Let $(X, S, y) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++} \times \Re^m$, and suppose there exists some $\beta > 0$ such that $\|V^{-1}L - \beta I\|_F \leq \beta/6$, where $V \equiv V(X)$ and $L \equiv L(S)$. Then the linear map $H' : \mathcal{S}^F \times \mathcal{S}^F \times \Re^m \to \Re^m \times \mathcal{S}^F \times \mathcal{L}^F$ defined by (11) is invertible.*

Proof. To show that H' is invertible, we show that $(0, 0, 0)$ is the only solution of the equation $H'(X, S, y)[A, B, c] = (0, 0, 0)$, where $(A, B, c) \in \mathcal{S}^F \times \mathcal{S}^F \times \Re^m$. As is standard in the SDP literature, it is not difficult to see that this equation can be reduced to the $m \times m$ system

$$\left(\mathcal{A} \circ V'(X)^{-1} \circ L'(S) \circ \mathcal{A}^*\right)(c) = 0.$$

Since $V'(X)^{-1} \circ L'(S)$ is negative definite by Corollary 2.9 and since $\mathcal{A}^*$ is injective by assumption A1, we conclude that $c = 0$, which immediately implies that $(A, B) = (0, 0)$, as desired. ∎

The above theorem will help us establish the existence of the Newton direction that will be used as the basis for our algorithm to solve problems (P) and (D) in Section 4. Moreover, the theorem motivates the need for the neighborhood condition $\|V^{-1}L - \beta I\|_F \leq \beta/6$ that we will formally introduce next in Section 3.

3 Technical Results

In this section, we prove several results that will be used for establishing the polynomial convergence of the algorithm that we propose in Section 4.

3.1 Properties of the partial central path map

Given $\gamma \in [0, 1/6]$, we define a feasible neighborhood of the central path as follows:

$$\mathcal{N}(\gamma) \equiv \left\{(X, S, y) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D) : \ \|V(X)^{-1}L(S) - \rho(X, S)\, I\|_F \leq \gamma\, \rho(X, S)\right\}. \tag{18}$$

Clearly $\mathcal{N}(\gamma)$ is nonempty as $(X_\nu, S_\nu, y_\nu) \in \mathcal{N}(\gamma)$ for all $\nu > 0$. Note also that, by Theorem 2.10 with $\beta = \rho(X, S)$ as well as the fact that $\gamma \leq 1/6$, $H'(X, S, y)$ is invertible for all $(X, S, y) \in \mathcal{N}(\gamma)$.

We now wish to establish several fundamental results concerning both the central path function $V(X)^{-1}L(S)$ and the neighborhood $\mathcal{N}(\gamma)$. The first result establishes how the neighborhood condition can be restated simply as an inequality on $\mathrm{tr}(V(X)^{-1}L(S))$.

Proposition 3.1 *$(X, S, y) \in \mathcal{N}(\gamma)$ if and only if (X, S, y) is primal-dual feasible and*

$$\mathrm{tr}\left(V(X)^{-1}L(S)\right) \geq \Gamma\, \rho(X, S), \tag{19}$$

where $\Gamma \equiv n - \gamma^2/2$. Moreover, for any $(X, S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++}$, it holds that $\mathrm{tr}(V(X)^{-1}L(S)) \leq n\, \rho(X, S)$.

Proof. Letting $V \equiv V(X)$, $L \equiv L(S)$, and $\rho \equiv \rho(X, S)$ and using (6), we have

$$\begin{aligned}\left\|V^{-1}L - \rho\, I\right\|_F^2 &= \left(V^{-1}L - \rho\, I\right) \bullet \left(V^{-1}L - \rho\, I\right) = V^{-1}L \bullet V^{-1}L - 2\,\rho\, V^{-1}L \bullet I + \rho^2\, I \bullet I \\ &= \left\|V^{-1}L\right\|_F^2 - 2\,\rho\,\mathrm{tr}(V^{-1}L) + n\,\rho^2 = 2\,n\,\rho^2 - 2\,\rho\,\mathrm{tr}(V^{-1}L),\end{aligned}$$

from which the first statement of the proposition follows using (18). The second statement of the proposition also follows from the above equations which imply $\|V^{-1}L - \rho\, I\|_F^2/(2\,\rho) = n\,\rho - \mathrm{tr}(V^{-1}L)$. ∎

The next proposition establishes some results concerning the second derivative of the function $\hat{V}(\hat{X})^{-1}\hat{L}(\hat{S})$, which, as described in (20), is analogous to $V(X)^{-1}L(S)$ but is defined on all of $\mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$.

Proposition 3.2 *Let $\hat{V} : \mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$ and $\hat{L} : \mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$ be defined by*

$$\hat{V}(\hat{X}) = \mathit{chol}(\hat{X}^{-1}), \qquad \hat{L}(\hat{S}) = \mathit{chol}(\hat{S}), \qquad \forall\ \hat{X}, \hat{S} \in \mathcal{S}^n_{++}. \tag{20}$$

Then the function $\Phi : \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$ defined by $\Phi(\hat{X}, \hat{S}) = \hat{V}(\hat{X})^{-1}\hat{L}(\hat{S})$ satisfies the following:

(a) for any fixed $\hat{S} \in \mathcal{S}^n_{++}$, $\mathrm{tr}(\Phi(\cdot, \hat{S}))$ is a strictly concave function;

(b) for all $\hat{A}, \hat{B} \in \mathcal{S}^n$,

$$\left\| \Phi''(\hat{X}, \hat{S})[\hat{A}, \hat{B}]^{(2)} \right\|_F \leq \frac{1}{\sqrt{2}} \|\hat{V}^{-1}\hat{L}\| \left(\|\hat{V}^T \hat{A} \hat{V}\|_F + \|\hat{L}^{-1}\hat{B}\hat{L}^{-T}\|_F \right)^2, \tag{21}$$

where $\hat{V} \equiv \hat{V}(\hat{X})$ and $\hat{L} \equiv \hat{L}(\hat{S})$.

Proof. In certain places throughout this proof, we will avoid the use of the hat ($\hat{\cdot}$) notation, which indicates fully dense matrices, in order to simplify the notation, though the meanings of the expressions will be clear from the context. Also to simplify notation, we define $\hat{U} : \mathcal{S}^n_{++} \to \mathcal{U}^n_{++}$ to be defined by $\hat{U}(\hat{X}) = \hat{V}(\hat{X})^{-T}$. With this definition, we see that $\Phi(\hat{X}, \hat{S}) = \hat{U}(\hat{X})^T \hat{L}(\hat{S})$ and that $\hat{X} = \hat{U}(\hat{X})\hat{U}(\hat{X})^T$.

To prove both (a) and (b), we consider the second derivative of Φ. Using (20) along with arguments similar to those found in the derivation of $V'(X)$ and $L'(S)$ in Section 2.3, we see that, for all $A, B \in \mathcal{S}^n$,

$$\Phi' \equiv \Phi'(\hat{X}, \hat{S})[A, B] = (U')^T L + U^T L', \tag{22}$$

where $U \equiv \hat{U}(\hat{X})$ and $L \equiv \hat{L}(\hat{S})$ and $U' \equiv \hat{U}'(\hat{X})[A] \in \mathcal{U}^n$ and $L' \equiv \hat{L}'(\hat{S})[B] \in \mathcal{L}^n$ are, respectively, the unique solutions of the equations

$$A = U'U^T + U(U')^T, \qquad B = L'L^T + L(L')^T. \tag{23}$$

Differentiating once again, we see that,

$$\Phi'' \equiv \Phi''(\hat{X}, \hat{S})[A, B]^{(2)} = (U'')^T L + 2\,(U')^T L' + U^T L'', \tag{24}$$

where $U'' \equiv \hat{U}''(\hat{X})[A]^{(2)} \in \mathcal{U}^n$ and $L'' \equiv \hat{L}''(\hat{S})[B]^{(2)} \in \mathcal{L}^n$ are, respectively, the unique solutions of the equations

$$0 = U''U^T + 2\,U'(U')^T + U(U'')^T, \qquad 0 = L''L^T + 2\,L'(L')^T + L(L'')^T. \tag{25}$$

We now prove (a). Letting h denote the function $\mathrm{tr}(\Phi(\cdot, \hat{S}))$, where $\hat{S}$ is fixed, it is straightforward to verify that $h''(\hat{X})[A]^{(2)} = \mathrm{tr}((U'')^T L)$, where U'' and L are defined as above. From (25), we have

$$U^{-1}U'' + (U'')^T U^{-T} = -2\left(U^{-1}U'\right)\left(U^{-1}U'\right)^T, \tag{26}$$

which implies that $\mathrm{diag}(U^{-1}U'') \leq 0$ since the right-hand side of (26) is negative semidefinite, which in turn implies that $\mathrm{diag}(U'') \leq 0$ since $U^{-1} \in \mathcal{U}^n_{++}$. It follows that $h''(\hat{X})[A]^{(2)} < 0$ unless $\mathrm{diag}(U'') = 0$. So suppose $\mathrm{diag}(U'') = 0$. Then, by (26), we see

$$\mathrm{diag}\left(\left(U^{-1}U'\right)\left(U^{-1}U'\right)^T\right) = 0 \quad \iff \quad U^{-1}U' = 0 \quad \iff \quad U' = 0 \quad \iff \quad A = 0.$$

Thus, we conclude that for all $A \neq 0$, $h''(\hat{X})[A]^{(2)} < 0$. This proves that h is strictly concave.

We now prove (b). Using (22) through (25), Lemma 2.7, and standard properties of the 2-norm and the Frobenius norm, we have

$$\begin{aligned}
\left\|\Phi''\right\|_F &\leq \|(U'')^T L\|_F + 2\|(U')^T L'\|_F + \|U^T L''\|_F \\
&\leq \|U^T L\| \left(\|U^{-1}U''\|_F + 2\|U^{-1}U'\|_F\|L^{-1}L'\|_F + \|L^{-1}L''\|_F\right) \\
&\leq \|U^T L\| \left(\sqrt{2}\,\|(U^{-1}U')(U^{-1}U')^T\|_F + 2\|U^{-1}U'\|_F\|L^{-1}L'\|_F + \sqrt{2}\|(L^{-1}L')(L^{-1}L')^T\|_F\right) \\
&\leq \|U^T L\| \left(\sqrt{2}\,\|U^{-1}U'\|_F^2 + 2\|U^{-1}U'\|_F\|L^{-1}L'\|_F + \sqrt{2}\|L^{-1}L'\|_F^2\right) \\
&\leq \|U^T L\| \left(\frac{1}{\sqrt{2}}\,\|U^{-1}AU^{-T}\|_F^2 + \|U^{-1}AU^{-T}\|_F\|L^{-1}BL^{-T}\|_F + \frac{1}{\sqrt{2}}\|L^{-1}BL^{-T}\|_F^2\right) \\
&\leq \frac{1}{\sqrt{2}}\|U^T L\| \left(\|U^{-1}AU^{-T}\|_F + \|L^{-1}BL^{-T}\|_F\right)^2.
\end{aligned}$$

The result now follows from the definition of $\hat{U}(\cdot)$. ∎

The next result plays a crucial role in the analysis of Section 4. In words, the theorem says that, given a fixed pair $(X, S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++}$, the maximum-determinant completion $\hat{X}(X)$ also maximizes the function $\mathrm{tr}(\hat{V}(\cdot)^{-1}L(S))$ among all positive definite completions of X.

Theorem 3.3 *Let $(X, S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++}$, and let $\hat{V} : \mathcal{S}^n_{++} \to \mathcal{L}^n_{++}$ be defined as in Proposition 3.2. Then*

$$\hat{X}(X) = \mathit{argmax}\left\{\mathrm{tr}\left(\hat{V}(\hat{Y})^{-1}L(S)\right) :\ \pi^F(\hat{Y}) = X,\ \hat{Y} \in \mathcal{S}^n_{++}\right\}.$$

Proof. Noting that $\hat{L}(S) = L(S)$ and letting $L \equiv L(S)$, we have from Proposition 3.2 that $h(\hat{Y}) \equiv \mathrm{tr}(\hat{V}(\hat{Y})^{-1}L)$ is a strictly concave function of $\hat{Y}$. Hence, since the constraints of the optimization problem under consideration are convex, any stationary point of this problem is a unique global maximum, and so we prove the theorem by showing that $\hat{X} \equiv \hat{X}(X)$ is a stationary point.

The derivative $h'(\hat{Y}) : \mathcal{S}^n \to \Re$ of h at $\hat{Y}$ is given by

$$h' \equiv h'(\hat{Y})[\hat{A}] = -\mathrm{tr}\left(\hat{V}^{-1}\hat{V}'\hat{V}^{-1}L\right), \tag{27}$$

for all $\hat{A} \in \mathcal{S}^n$, where $\hat{V} \equiv \hat{V}(\hat{Y})$ and $\hat{V}' \equiv \hat{V}'(\hat{Y})[\hat{A}]$ is the unique solution of the system

$$-\hat{Y}^{-1}\hat{A}\hat{Y}^{-1} = \hat{V}'\hat{V}^T + \hat{V}(\hat{V}')^T. \tag{28}$$

Pre-multiplying (28) by $\hat{V}^{-1}$ and post-multiplying by $\hat{V}^{-T}$ and using the fact that $\hat{Y}^{-1} = \hat{V}\hat{V}^T$, we see that $-\hat{V}^T\hat{A}\hat{V} = \hat{V}^{-1}\hat{V}' + (\hat{V}')^T\hat{V}^{-T}$, which shows that $-\mathrm{diag}(\hat{V}^T\hat{A}\hat{V}) = 2\,\mathrm{diag}(\hat{V}^{-1}\hat{V}')$. Applying this equality to (27) and letting $W \in \mathcal{S}^n_{++}$ be the diagonal matrix defined by $W_{jj} = \hat{V}_{jj}^{-1}L_{jj}/2$, we deduce that

$$h' = \frac{1}{2}\mathrm{diag}(\hat{V}^T\hat{A}\hat{V})^T\mathrm{diag}(\hat{V}^{-1}L) = W \bullet \hat{V}^T\hat{A}\hat{V} = \hat{V}W\hat{V}^T \bullet \hat{A},$$

which implies that $\nabla h(\hat{Y}) = \hat{V}W\hat{V}^T$.

Let $\bar{F} \equiv V \times V \setminus F$. Considering that the variables $\hat{Y}_{ij}$ for $(i,j) \in F$ can be eliminated by the equation $\pi^F(\hat{Y}) = X$, we see that a stationary point $\hat{Y}$ of the optimization problem satisfies $\pi^{\bar{F}}(\nabla h(\hat{Y})) = 0$, that is, $\nabla h(\hat{Y}) = \hat{V}W\hat{V}^T \in \mathcal{S}^F$. Since $W \in \mathcal{S}^n_{++}$ is diagonal, $\hat{V}W\hat{V}^T \in \mathcal{S}^F$ is equivalent to $\hat{V}\hat{V}^T = \hat{Y}^{-1} \in \mathcal{S}^F$, which is precisely the condition that $\hat{X}$ satisfies uniquely by Theorem 2.1. So $\hat{X}$ is a stationary point of the optimization problem, which completes the proof. ∎

3.2 Miscellaneous results

In this subsection, we catalog a few results that will prove useful in Section 4. The first two results give details about the system $H'(X,S,y)[A,B,c] = (0,0,R)$, where H' is given as in (11).

Lemma 3.4 *Let $(X,S,y) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++} \times \Re^m$ and $R \in \mathcal{L}^F$ be given, and suppose that $(A,B,c) \in \mathcal{S}^F \times \mathcal{S}^F \times \Re^m$ satisfies $H'(X,S,y)[A,B,c] = (0,0,R)$. Then*

$$A \bullet B = \left(V^{-1}V' + (V')^T V^{-T}\right) \bullet V^{-1}BV^{-T} = 0, \tag{29}$$

where $V \equiv V(X)$ and $V' \equiv V'(X)[A]$.

Proof. From (11), we see that $\mathcal{A}(A) = 0$ and $\mathcal{A}^*(c) + B = 0$. Hence, $A \bullet B = -A \bullet \mathcal{A}^*(c) = -c^T\mathcal{A}(A) = 0$. Also, letting $\hat{X} \equiv \hat{X}(X)$, $\hat{X}' \equiv \hat{X}'(X)[A]$ and using (9), (10), (2), and the trace-compatibility of $(\hat{X}', A, B)$, we see that

$$A \bullet B = \hat{X}' \bullet B = -\hat{X}\left(V'V^T + V(V')^T\right)\hat{X} \bullet B = \left(V^{-1}V' + (V')^T V^{-T}\right) \bullet V^{-1}BV^{-T},$$

which completes the proof. ∎

Proposition 3.5 *Let the conditions of Lemma 3.4 hold, and define $Q \equiv V^{-1}L - \sqrt{\nu}\, I$, where $L \equiv L(S)$ and $\nu > 0$ is as in (7). Suppose that $\|Q\|_F < \sqrt{\nu}/\sqrt{2}$. Then*

$$\|V^{-1}V'\|_F \le \frac{1}{\sqrt{2\nu}}\left(\sqrt{\nu} - \sqrt{2}\,\|Q\|_F\right)^{-1} \|V^{-1}(RL^T + LR^T)V^{-T}\|_F, \tag{30}$$

$$\|V^{-1}BV^{-T}\|_F \le \sqrt{\nu}\left(\sqrt{\nu} - \sqrt{2}\,\|Q\|_F\right)^{-1} \|V^{-1}(RL^T + LR^T)V^{-T}\|_F. \tag{31}$$

Proof. From (11), we have $B = L'(S)^{-1}[R + \sqrt{\nu}\,V'] = (R + \sqrt{\nu}\,V')L^T + L(R + \sqrt{\nu}\,V')^T$. Hence, letting $\tilde{R} \equiv RL^T + LR^T$,

$$V^{-1}BV^{-T} = V^{-1}\tilde{R}V^{-T} + \sqrt{\nu}\,V^{-1}V'(V^{-1}L)^T + \sqrt{\nu}\,V^{-1}L(V^{-1}V')^T.$$

It follows that

$$\frac{1}{\sqrt{\nu}}V^{-1}BV^{-T} - \sqrt{\nu}\,\left(V^{-1}V' + (V')^TV^{-T}\right) = \frac{1}{\sqrt{\nu}}\,V^{-1}\tilde{R}V^{-T} + V^{-1}V'Q^T + Q(V^{-1}V')^T,$$

which, from (29), implies

$$\begin{aligned}
&\max\left\{\frac{1}{\sqrt{\nu}}\left\|V^{-1}BV^{-T}\right\|_F, \sqrt{\nu}\left\|V^{-1}V' + (V')^TV^{-T}\right\|_F\right\} \le \\
&\left\|\frac{1}{\sqrt{\nu}}\,V^{-1}BV^{-T} - \sqrt{\nu}\,\left(V^{-1}V' + (V')^TV^{-T}\right)\right\|_F \le \frac{1}{\sqrt{\nu}}\,\|V^{-1}\tilde{R}V^{-T}\|_F + 2\,\|Q\|_F\|V^{-1}V'\|_F
\end{aligned} \tag{32}$$

Applying Lemma 2.7 with $J = V^{-1}V'$ to (32), we see that

$$\sqrt{2}\left(\sqrt{\nu} - \sqrt{2}\,\|Q\|_F\right)\|V^{-1}V'\|_F \le \frac{1}{\sqrt{\nu}}\,\|V^{-1}\tilde{R}V^{-T}\|_F,$$

which proves (30). To prove (31), we combine (30) and (32) to obtain

$$\begin{aligned}
\frac{1}{\sqrt{\nu}}\|V^{-1}BV^{-T}\|_F &\le \frac{1}{\sqrt{\nu}}\,\|V^{-1}\tilde{R}V^{-T}\|_F\left(1 + \frac{\sqrt{2}\,\|Q\|_F}{\sqrt{\nu} - \sqrt{2}\|Q\|_F}\right) \\
&= \left(\sqrt{\nu} - \sqrt{2}\|Q\|_F\right)^{-1}\|V^{-1}\tilde{R}V^{-T}\|_F.
\end{aligned}$$

■

We next establish several inequalities that relate to bounds on the maximum and minimum eigenvalues of $\hat{X}(X)S$ for $(X,S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++}$.

Proposition 3.6 *Let $(X,S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++}$, and define $\hat{X} \equiv \hat{X}(X)$, $V \equiv V(X)$, $L \equiv L(S)$, and $\rho \equiv \rho(X,S)$. Suppose $\|V^{-1}L - \rho\, I\| \leq \gamma\, \rho$ for some $\gamma \geq 0$ satisfying $\gamma^2 + 2\gamma < 1$. Then:*

(a) $\|V^{-1}L\| \leq (\gamma+1)\rho$;

(b) $\|L^{-1}V\| \leq (1-(\gamma^2+2\gamma))^{-1/2}\rho^{-1}$.

Proof. Let $Q \equiv V^{-1}L - \rho\, I$ and note that

$$\begin{aligned}\|Q\|^2 = \left\|Q^TQ\right\| &= \left\|L^TV^{-T}V^{-1}L - \rho V^{-1}L - \rho L^TV^{-T} + \rho^2 I\right\| \\ &= \left\|L^TV^{-T}V^{-1}L - \rho^2 I - \rho(Q+Q^T)\right\|.\end{aligned}$$

Hence, from (2), (3), standard properties of $\|\cdot\|$, and the assumptions of the proposition,

$$\begin{aligned}\|\hat{X}S - \rho^2 I\| &= \|L^TV^{-T}V^{-1}L - \rho^2 I\| \\ &= \|L^TV^{-T}V^{-1}L - \rho^2 I - \rho(Q+Q^T) + \rho(Q+Q^T)\| \\ &\leq \|Q\|^2 + 2\,\rho\,\|Q\| \leq (\gamma^2+2\gamma)\rho^2. \end{aligned} \tag{33}$$

Recall from the definition of ρ that $\hat{X} \bullet S = X \bullet S = n\rho^2$, that is, $n\rho^2$ equals the sum of the eigenvalues of $\hat{X}S$. It follows that $\lambda_{\max}[\hat{X}S] \geq \rho^2$ and that $\lambda_{\min}[\hat{X}S] \leq \rho^2$. Hence,

$$\|\hat{X}S - \rho^2 I\| = \max\{\lambda_{\max}[\hat{X}S] - \rho^2, \rho^2 - \lambda_{\min}[\hat{X}S]\}.$$

Thus, using (33), (2), and (3), we have

$$\|V^{-1}L\|^2 = \|L^TV^{-T}V^{-1}L\| = \|\hat{X}S\| = \lambda_{\max}[\hat{X}S] \leq \|\hat{X}S - \rho^2 I\| + \rho^2 \leq (\gamma^2+2\gamma+1)\rho^2,$$

which proves (a). Similarly,

$$\begin{aligned}\|L^{-1}V\|^2 = \|V^TL^{-T}L^{-1}V\| &= \|S^{-1}\hat{X}^{-1}\| = \lambda_{\min}[\hat{X}S]^{-1} \\ &\leq \left(\rho^2 - \|\hat{X}S - \rho^2 I\|\right)^{-1} \leq \left(\rho^2 - (\gamma^2+2\gamma)\rho^2\right)^{-1},\end{aligned}$$

which proves (b). ■

We remark that the condition $(X,S,y) \in \mathcal{N}(\gamma)$ implies that the hypotheses of Proposition 3.6 holds since $\|V^{-1}L - \rho\, I\| \leq \|V^{-1}L - \rho\, I\|_F$.

Finally, we state the following proposition, which follows as a direct extension of lemmas 3.4 and 3.5 of Monteiro and Tsuchiya [24].

Proposition 3.7 *Let $(X,S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++}$ and $\bar{X}, \bar{S} \in \mathcal{S}^n$. Define $\hat{X} \equiv \hat{X}(X)$, $V \equiv V(X)$ and $L \equiv L(S)$. Suppose there exists some $\tau \in (0,1)$ such that*

$$\max\left\{\|V^T(\bar{X} - \hat{X})V\|, \|L^{-1}(\bar{S} - S)L^{-T}\|\right\} \leq \tau.$$

Then $\bar{X}, \bar{S} \in \mathcal{S}^n_{++}$,

$$\max\left\{\|V^{-1}\bar{V}\|, \|\bar{V}^{-1}V\|, \|L^{-1}\bar{L}\|, \|\bar{L}^{-1}L\|\right\} \leq \frac{1}{\sqrt{1-\tau}} \qquad \textit{and} \qquad \|\bar{V}^{-1}\bar{L}\| \leq \frac{\|V^{-1}L\|}{1-\tau},$$

where $\bar{V} = chol(\bar{X}^{-1})$ and $\bar{L} = chol(\bar{S})$.

4 The Partial Primal-Dual Algorithm

The algorithm described in this section is based on the same ideas that typical path-following algorithms are based on — namely, the use of a Newton direction to decrease the duality gap and a bound on the step-size to ensure proximity to the central path. Using these ideas, we establish the polynomiality of the algorithm in Theorem 4.9.

Suppose that $(X, S, y) \in \mathcal{N}(\gamma)$ where $\gamma \in [0, 1/6]$. Then, for a fixed constant $0 \leq \sigma < 1$, we define the Newton direction $(\Delta X, \Delta S, \Delta y)$ at (X, S, y) as the solution of the system

$$H'(X, S, y)[\Delta X, \Delta S, \Delta y] = (0, 0, \sigma\, \rho\, V - L)\,, \tag{34}$$

where $\nu \equiv \rho^2$ is implicit in the definition of H, $\rho \equiv \rho(X, S)$, $V \equiv V(X)$, and $L \equiv L(S)$. Note that $(\Delta X, \Delta S, \Delta y)$ is well-defined by Theorem 2.10. We also make the following definitions for all $\alpha \in \Re$ such that $(X_\alpha, S_\alpha) \in \mathcal{S}_+^{F?} \times \mathcal{S}_+^F$:

$$\begin{aligned}(X_\alpha, S_\alpha, y_\alpha) &\equiv (X, S, y) + \alpha(\Delta X, \Delta S, \Delta y),\\ \mu_\alpha &\equiv \mu(X_\alpha, S_\alpha),\\ \rho_\alpha &\equiv \rho(X_\alpha, S_\alpha).\end{aligned}$$

We have the following proposition.

Proposition 4.1 *Let $(X, S, y) \in N(\gamma)$, and define $\mu \equiv \mu(X, S)$, $\rho \equiv \rho(X, S)$ and $\zeta = (\Delta X \bullet S + X \bullet \Delta S)(n\mu)^{-1}$. Then for all $\alpha \in \Re$ such that $(X_\alpha, S_\alpha, y_\alpha) \in \mathcal{S}_+^{F?} \times \mathcal{S}_+^F$,*

$$\mu_\alpha = \mu(1 + \alpha\,\zeta), \qquad\qquad \rho_\alpha \leq \rho(1 + \alpha\,\zeta/2). \tag{35}$$

Proof. Note that $\Delta X \bullet \Delta S = 0$ from (29). Using (1), we thus have

$$\mu_\alpha = \frac{(X + \alpha\Delta X) \bullet (S + \alpha\Delta S)}{n} = \frac{X \bullet S + \alpha\zeta n\mu}{n} = \mu + \alpha\zeta\mu,$$

which proves the equality. Similarly using (5), we see that

$$\rho_\alpha = \rho(1 + \alpha\,\zeta)^{1/2} \leq \rho(1 + \alpha\,\zeta/2),$$

where the inequality follows from the real-number relation $1 + 2x \leq 1 + 2x + x^2$. ∎

With regard to the above proposition, it is important to mention that we anticipate that ζ is negative due to the fact that $\sigma < 1$, which would imply that $\mu_\alpha < \mu$ and $\rho_\alpha < \rho$, that is, the duality gap decreases along the direction $(\Delta X, \Delta S, \Delta y)$. This, however, must be proven under certain assumptions as will be shown below. For the discussion of the generic algorithm next, it would be useful for the reader to assume that $\zeta < 0$.

We now state the generic, primal-dual path-following algorithm that we study in this section.

Algorithm SDP:

Let $\varepsilon > 0$, $\gamma \in [0, 1/6]$, $\sigma \in [0, 1)$, and $(X^0, S^0, y^0) \in \mathcal{N}(\gamma)$ be given. Set $k = 0$.

Repeat until $X^k \bullet S^k \leq \varepsilon$ **do**

1. Set $(X, S, y) \equiv (X^k, S^k, y^k)$ and $\rho \equiv \rho(X, S)$.
2. Compute the Newton direction $(\Delta X, \Delta S, \Delta y)$ at (X, S, y).
3. Choose $\alpha \geq 0$ such that $(X_\alpha, S_\alpha, y_\alpha) \in \mathcal{N}(\gamma)$.
4. Set $(X^{k+1}, S^{k+1}, y^{k+1}) = (X_\alpha, S_\alpha, y_\alpha)$ and increment k by 1.

End

The remainder of this section is devoted to determining constants γ and σ and a constant step-size α such that Algorithm SDP terminates within a polynomial number of loops, where the polynomial depends on n, ε, and $X^0 \bullet S^0$.

To this end, we introduce constants $\gamma \geq 0$, $\delta \in [0, \sqrt{n})$, and $\tau \in (0,1)$ satisfying

$$\gamma \leq 1/6, \qquad 0 < 1 - (\gamma^2 + 2\gamma) \leq 1/\sqrt{2} \tag{36}$$

and

$$2\,(\gamma+1)\left(\delta^2+\gamma^2\right)^{1/2} \leq \left(1-\sqrt{2}\,\gamma\right)\left(1-(\gamma^2+2\,\gamma)\right)\tau. \tag{37}$$

Note that, for example, the triple $(\gamma, \delta, \tau) = (0.138, 0.138, 0.79)$ satisfies (36) and (37) irrespective of the value of n. We also define

$$\sigma \equiv 1 - \delta/\sqrt{n}. \tag{38}$$

In addition, we make the mild assumption that n is large enough so that

$$\delta\sqrt{n} \geq \tau\gamma(\gamma+1), \tag{39}$$

$$\sigma \geq \tau\theta, \tag{40}$$

$$\sqrt{n} \geq \frac{\gamma^2(\sigma-\tau\theta)(1-\tau)}{2\sqrt{2}\,\tau^2(\gamma+1)}, \tag{41}$$

where $\theta \equiv 1+\gamma(\gamma+1)/(2n)$. In fact, taking $(\gamma, \delta, \tau) = (0.138, 0.138, 0.79)$ as above shows that (39) through (41) are also satisfied for all values of n.

Lemma 4.2 *Define $V \equiv V(X)$, $L \equiv L(S)$, and $\rho \equiv \rho(X,S)$, and suppose that $(X,S) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^{F}_{++}$ satisfies $\|V^{-1}L - \rho I\|_F \leq \gamma\rho$. Then*

$$\|\sigma\rho I - V^{-1}L\|_F \leq \left(\delta^2+\gamma^2\right)^{1/2}\rho.$$

Proof. Using the definition of the Frobenius norm, we have

$$\begin{aligned}
\|\sigma\rho I - V^{-1}L\|_F^2 &= \|\rho I - V^{-1}L + (\sigma-1)\rho I\|_F^2 \\
&= \|\rho I - V^{-1}L\|_F^2 + 2(\sigma-1)\rho(\rho I - V^{-1}L) \bullet I + (\sigma-1)^2 n\rho^2, \\
&= \|\rho I - V^{-1}L\|_F^2 + 2(\sigma-1)\rho\left(n\rho - \mathrm{tr}(V^{-1}L)\right) + (\sigma-1)^2 n\rho^2 \\
&\leq \gamma^2\rho^2 + (\sigma-1)^2 n\rho^2,
\end{aligned}$$

where the inequality follows by assumption and by the second statement of Proposition 3.1. Since $(\sigma-1)^2 n = \delta^2$ by (38), the result follows. ∎

Proposition 4.3 *Let $(X,S,y) \in \mathcal{N}(\gamma)$. Then*

$$\max\left\{\|V^T\hat{X}'V\|_F, \|L^{-1}\Delta S L^{-T}\|_F\right\} \leq \tau,$$

where $\hat{X}' \equiv \hat{X}'(X)[\Delta X]$. As a result, $\hat{X} + \hat{X}' \in \mathcal{S}^n_{++}$ and $S + \Delta S \in \mathcal{S}^F_{++}$, where $\hat{X} \equiv \hat{X}(X)$.

Proof. Let $V \equiv V(X)$, $L \equiv L(S)$, and $\rho \equiv \rho(X,S)$. Also letting $V' \equiv V'(X)[\Delta X]$ and using (10) and (2), we see that

$$-V^T\hat{X}'V = V^{-1}V' + (V')^T V^{-T} \qquad \Longrightarrow \qquad \|V^T\hat{X}'V\|_F \leq 2\,\|V^{-1}V'\|_F.$$

Note that the hypotheses of Proposition 3.5 hold with $R \equiv \sigma\rho V - L$, $\nu \equiv \rho^2$, and $Q \equiv V^{-1}L - \rho I$. Hence, using (30), standard properties of norms, (18), Lemma 4.2, Proposition 3.6(a), (36), and (37), we have

$$\begin{aligned}
2\,\|V^{-1}V'\|_F &\leq \frac{\sqrt{2}}{\rho}\left(\rho - \sqrt{2}\,\|Q\|_F\right)^{-1}\|V^{-1}((\sigma\rho V - L)L^T + L(\sigma\rho V - L)^T)V^{-T}\|_F \\
&\leq \frac{2\sqrt{2}}{\rho}\left(\rho - \sqrt{2}\,\gamma\,\rho\right)^{-1}\|V^{-1}(\sigma\rho V - L)L^T V^{-T}\|_F \\
&\leq \frac{2\sqrt{2}}{\rho^2}\left(1 - \sqrt{2}\,\gamma\right)^{-1}\|\sigma\rho I - V^{-1}L\|_F\|V^{-1}L\| \\
&\leq 2\sqrt{2}\left(1 - \sqrt{2}\,\gamma\right)^{-1}\left(\delta^2 + \gamma^2\right)^{1/2}(\gamma + 1) \\
&\leq 2\left(1 - (\gamma^2 + 2\,\gamma)\right)^{-1}\left(1 - \sqrt{2}\,\gamma\right)^{-1}\left(\delta^2 + \gamma^2\right)^{1/2}(\gamma + 1) \leq \tau. \qquad (42)
\end{aligned}$$

Now using (31) and Proposition 3.6(b) along with similar arguments, we have

$$\begin{aligned}
\|L^{-1}\Delta S L^{-T}\|_F &\leq \|L^{-1}V\|^2\,\|V^{-1}\Delta S V^{-T}\|_F \\
&\leq \|L^{-1}V\|^2\,\rho\left(\rho - \sqrt{2}\,\|Q\|_F\right)^{-1}\|V^{-1}((\sigma\rho V - L)L^T + L(\sigma\rho V - L)^T)V^{-T}\|_F \\
&\leq 2\left(1 - (\gamma^2 + 2\gamma)\right)^{-1}\left(1 - \sqrt{2}\,\gamma\right)^{-1}\left(\delta^2 + \gamma^2\right)^{1/2}(\gamma + 1) \leq \tau,
\end{aligned}$$

which concludes the proof of the first statement of the proposition. The second statement follows from Proposition 3.7 with $\bar{X} = \hat{X} + \hat{X}'$ and $\bar{S} = S + \Delta S$ (and the fact that $\Delta S \in \mathcal{S}^F$). ∎

Corollary 4.4 *Let $(X, S, y) \in \mathcal{N}(\gamma)$. Then for all $0 \leq \alpha \leq 1$, $(X_\alpha, S_\alpha) \in \mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++}$.*

Proof. By Proposition 4.3, we see that $S_1 = S + \Delta S \in \mathcal{S}^F_{++}$. Since S_α is a convex combination of S and S_1 for any $0 \leq \alpha \leq 1$, it follows that $S_\alpha \in \mathcal{S}^F_{++}$. Likewise, $\hat{X} + \alpha\hat{X}' \in \mathcal{S}^n_{++}$ for all $0 \leq \alpha \leq 1$. Noting that, by Theorem 2.1 and (9),

$$\pi^F(\hat{X} + \alpha\hat{X}') = X + \alpha\Delta X = X_\alpha$$

we see that $\hat{X} + \alpha\hat{X}'$ is a positive definite completion of X_α, which implies $X_\alpha \in \mathcal{S}^{F?}_{++}$. ∎

Lemma 4.5 *Let $(X, S, y) \in \mathcal{N}(\gamma)$, and define $V \equiv V(X)$, $L \equiv L(S)$, $\rho \equiv \rho(X, S)$, $V' \equiv V'(X)[\Delta X]$, and $L' \equiv L'(S)[\Delta S]$. Also let Γ be defined as in Proposition 3.1. Then*

$$V^{-1}L \bullet \left(V^{-1}L' - V^{-1}V'V^{-1}L\right) \leq \left((\sigma - 1)n + \frac{1}{2}\tau(\gamma + 1)\gamma\right)\rho^2 \qquad (43)$$

$$\begin{aligned}
\frac{\Gamma}{n}V^{-1}L \bullet \left(V^{-1}L' - V^{-1}V'V^{-1}L\right) - \rho\,\mathrm{tr}\left(V^{-1}V'(\rho I - V^{-1}L)\right) \leq \\
\left((\sigma - 1)\Gamma + \frac{1}{2}\,\tau\gamma^2\left(1 + \frac{1}{2n}\,\gamma(\gamma + 1)\right)\right)\rho^2. \qquad (44)
\end{aligned}$$

Proof. We first prove some simple bounds that will allow us to prove (43) and (44) more easily. Defining $P \equiv V^{-1}V'\left(\rho I - V^{-1}L\right)$ and using standard properties of $\mathrm{tr}(\cdot)$, $\|\cdot\|$, and $\|\cdot\|_F$ along with (18), Proposition 3.6(a), and (42) (which appears inside the proof of Proposition 4.3), we see that

$$\begin{aligned}
\left|V^{-1}L \bullet P\right| &= \left|(V^{-1}V')^T(V^{-1}L) \bullet (\rho I - V^{-1}L)\right| \leq \|V^{-1}V'\|_F\|V^{-1}L\|\,\|\rho I - V^{-1}L\|_F \\
&\leq \frac{1}{2}\tau(\gamma + 1)\gamma\rho^2 \qquad (45)
\end{aligned}$$

Similarly,

$$\left|\left(V^{-1}L - \rho I\right) \bullet P\right| \leq \|V^{-1}V'\|_F \|V^{-1}L - \rho I\|_F^2 \leq \frac{1}{2}\tau\gamma^2\rho^2 \tag{46}$$

Now, the equation (34) for the Newton direction $(\Delta X, \Delta S, \Delta y)$ shows that $L' - \rho V' = \sigma\rho V - L$, which implies $V^{-1}L' = \sigma\rho I - V^{-1}L + \rho V^{-1}V'$. Substituting for $V^{-1}L'$ in the left-hand side of (43) and using (6), the second statement of Proposition 3.1, and (45), we see that

$$\begin{aligned} V^{-1}L \bullet \left(V^{-1}L' - V^{-1}V'V^{-1}L\right) &= V^{-1}L \bullet \left(\sigma\rho I - V^{-1}L + \rho V^{-1}V' - V^{-1}V'V^{-1}L\right) \\ &= \sigma\rho\,\mathrm{tr}(V^{-1}L) - \|V^{-1}L\|_F^2 + V^{-1}L \bullet P \\ &\leq (\sigma - 1)n\rho^2 + \frac{1}{2}\tau(\gamma+1)\gamma\rho^2, \end{aligned}$$

as desired. Using similar arguments along with (46) and the fact that $\Gamma/n = 1 - \gamma^2/(2n)$, (44) is proven as follows:

$$\begin{aligned} &\frac{\Gamma}{n} V^{-1}L \bullet \left(V^{-1}L' - V^{-1}V'V^{-1}L\right) - \rho\,\mathrm{tr}\left(V^{-1}V'(\rho I - V^{-1}L)\right) \\ &\quad = \frac{\Gamma}{n}\sigma\rho\,\mathrm{tr}(V^{-1}L) - \frac{\Gamma}{n}\|V^{-1}L\|_F^2 + \left(\frac{\Gamma}{n}V^{-1}L - \rho I\right) \bullet P \\ &\quad = \frac{\Gamma}{n}\sigma\rho\,\mathrm{tr}(V^{-1}L) - \frac{\Gamma}{n}\|V^{-1}L\|_F^2 + \left(V^{-1}L - \rho I\right) \bullet P - \frac{\gamma^2}{2n}V^{-1}L \bullet P \\ &\quad \leq \Gamma\sigma\rho^2 - \Gamma\rho^2 + \frac{1}{2}\tau\gamma^2\rho^2 + \frac{1}{4n}\tau(\gamma+1)\gamma^3\rho^2. \end{aligned}$$

■

Proposition 4.6 *Let $(X, S, y) \in \mathcal{N}(\gamma)$, and define $\mu \equiv \mu(X, S)$. Then*

$$\zeta \equiv \frac{\Delta X \bullet S + X \bullet \Delta S}{n\mu} \leq -\frac{\delta}{\sqrt{n}}.$$

Hence, $\mu(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$ decrease from (X, S, y) along the direction $(\Delta X, \Delta S, \Delta y)$.

Proof. Let $\hat{X} \equiv \hat{X}(X)$, $\hat{X}' \equiv \hat{X}'(X)[\Delta X]$, $V \equiv V(X)$, $V' \equiv V'(X)[\Delta X]$, $L \equiv L(S)$, $L' \equiv L'(S)[\Delta S]$ and $\rho \equiv \rho(X, S)$. Then using Theorem 2.1, (9) with $A = \Delta X$, the fact that $S \in \mathcal{S}^F$ and $\Delta S \in \mathcal{S}^F$, (10) with $A = \Delta X$, (8) with $B = \Delta S$, (2), and (3), we see that

$$\begin{aligned} \Delta X \bullet S + X \bullet \Delta S &= \hat{X}' \bullet S + \hat{X} \bullet \Delta S \\ &= \left(-\hat{X}(V'V^T + V(V')^T)\hat{X}\right) \bullet LL^T + V^{-T}V^{-1} \bullet (L'L^T + L(L')^T) \\ &= 2\,V^{-T}V^{-1} \bullet L'L^T - 2\,\hat{X}V'V^T\hat{X} \bullet LL^T \\ &= 2\,V^{-T}V^{-1} \bullet L'L^T - 2\,V^{-T}V^{-1}V'V^{-1} \bullet LL^T \\ &= 2\,V^{-1}L \bullet \left(V^{-1}L' - V^{-1}V'V^{-1}L\right) \\ &\leq 2\left((\sigma - 1)n + \frac{1}{2}\tau(\gamma+1)\gamma\right)\rho^2, \end{aligned} \tag{47}$$

where the inequality follows from (43). By the definition of ζ, the inequality just proven, (38), and (39), we have

$$\zeta \leq 2(\sigma - 1) + \tau(\gamma+1)\gamma/n \leq -2\delta/\sqrt{n} + \delta/\sqrt{n} = -\delta/\sqrt{n}.$$

The conclusion that the duality gap measures $\mu(\cdot,\cdot)$ and $\rho(\cdot,\cdot)$ decrease along $(\Delta X, \Delta S, \Delta y)$ can now be seen from Proposition 4.1. ■

Proposition 4.7 *Let $(X, S, y) \in \mathcal{N}(\gamma)$, and define the constant Γ as in Proposition 3.1 and the functions Φ, $\hat{V}$, and $\hat{L}$ as in Proposition 3.2. Suppose that $\alpha \geq 0$ satisfies the inequality*

$$\alpha \leq \frac{\gamma^2(\sigma - \tau\theta)(1-\tau)}{2\sqrt{2}\,\tau^2(\gamma+1)} \frac{1}{\sqrt{n}}, \tag{48}$$

where $\theta \equiv 1 + \gamma(\gamma+1)/(2n)$. Then

$$\mathrm{tr}(\Phi(\hat{X} + \alpha\hat{X}', S_\alpha)) = \mathrm{tr}(\hat{V}(\hat{X} + \alpha\hat{X}')^{-1}\hat{L}(S_\alpha)) \geq \Gamma\rho_\alpha,$$

where $\hat{X} \equiv \hat{X}(X)$ and $\hat{X}' \equiv \hat{X}'(X)[\Delta X]$.

Proof. We first remark that the right-hand side of (48) is nonnegative by (40) and is less than or equal to 1 by (41), and so $0 \leq \alpha \leq 1$, which clearly shows that $\hat{X} + \alpha\hat{X}' \in \mathcal{S}^n_{++}$ and $S_\alpha \in \mathcal{S}^n_{++}$ by Proposition 4.3. Hence, Φ is defined at $(\hat{X} + \alpha\hat{X}', S_\alpha)$.

Define $\hat{V} \equiv \hat{V}(\hat{X})$, $\hat{V}' \equiv \hat{V}'(\hat{X})[\Delta X]$, $\hat{L} \equiv \hat{L}(S)$, and $\hat{L} \equiv \hat{L}'(S)[\Delta S]$. In addition, define $V \equiv V(X)$, $V' \equiv V'(X)[\Delta X]$, $L \equiv L(S)$, and $L' \equiv L'(S)[\Delta S]$. Noting that $V(\cdot) = \hat{V}(\hat{X}(\cdot))$ and that $\hat{L}(\cdot)$ is identical to $L(\cdot)$ on the domain $\mathcal{S}^F_{++}$, we see that

$$V = \hat{V}, \quad V' = \hat{V}', \quad L = \hat{L}, \quad L' = \hat{L}'. \tag{49}$$

The Taylor integral formula implies that

$$\Phi(\hat{X} + \alpha\hat{X}', S_\alpha) = \Phi(\hat{X}, S) + \alpha\,\Phi'(\hat{X}, S)[\hat{X}', \Delta S] + \alpha^2\,T_\alpha, \tag{50}$$

where

$$T_\alpha \equiv \int_0^1 (1-t)\Phi''(\hat{X} + t\alpha\hat{X}', S_{t\alpha})[\hat{X}', \Delta S]^{(2)} dt. \tag{51}$$

Analyzing the first two components of (50), we first see by (49) that $\Phi(\hat{X}, S) = \hat{V}^{-1}\hat{L} = V^{-1}L$. Secondly, letting $\rho \equiv \rho(X, S)$, we have

$$\begin{aligned}
\Phi'(\hat{X}, S)[\hat{X}', \Delta S] &= -\hat{V}^{-1}\hat{V}'\hat{V}^{-1}\hat{L} + \hat{V}^{-1}\hat{L}' \\
&= V^{-1}L' - V^{-1}V'V^{-1}L \\
&= \sigma\rho I - V^{-1}L + V^{-1}V'(\rho I - V^{-1}L).
\end{aligned}$$

where the third equality comes from substituting for $V^{-1}L'$ as was done in the proof of Lemma 4.5. Hence, we can rewrite (50) as

$$\begin{aligned}
\Phi(\hat{X} + \alpha\hat{X}', S_\alpha) &= V^{-1}L + \alpha\left(\sigma\rho I - V^{-1}L + V^{-1}V'(\rho I - V^{-1}L)\right) + \alpha^2\,T_\alpha \\
&= (1-\alpha)V^{-1}L + \alpha\sigma\rho I + \alpha P + \alpha^2\,T_\alpha,
\end{aligned} \tag{52}$$

where $P \equiv V^{-1}V'(\rho I - V^{-1}L)$. Taking the trace of (52) and using Proposition 3.1 and Proposition 4.1 where $\zeta \equiv (\Delta X \bullet S + X \bullet \Delta S)/(n\rho^2)$, we see

$$\begin{aligned}
\mathrm{tr}\left(\Phi(\hat{X} + \alpha\hat{X}', S_\alpha)\right) &\geq (1-\alpha)\Gamma\rho + \alpha\sigma n\rho + \alpha\mathrm{tr}(P) + \alpha^2\,\mathrm{tr}(T_\alpha) \\
&\geq \Gamma\left(\rho_\alpha - \alpha\rho(1 + \zeta/2)\right) + \alpha\sigma n\rho + \alpha\mathrm{tr}(P) + \alpha^2\,\mathrm{tr}(T_\alpha) \\
&= \Gamma\rho_\alpha - \alpha\Gamma\rho\left(1 + \frac{1}{2n\rho^2}\left(\Delta X \bullet S + X \bullet \Delta S\right)\right) + \alpha\sigma n\rho + \alpha\mathrm{tr}(P) + \alpha^2\,\mathrm{tr}(T_\alpha) \\
&= \Gamma\rho_\alpha - \alpha\Gamma\rho + \alpha\sigma n\rho + \alpha\rho^{-1}\left(\rho\mathrm{tr}(P) - \frac{\Gamma}{2n}\left(\Delta X \bullet S + X \bullet \Delta S\right)\right) + \alpha^2\,\mathrm{tr}(T_\alpha).
\end{aligned}$$

From (47) (which is inside the proof of Proposition 4.6), we have $\Delta X \bullet S + X \bullet \Delta S = 2\,V^{-1}L \bullet \left(V^{-1}L' - V^{-1}V'V^{-1}L\right)$. Thus, letting $\theta \equiv 1 + \gamma(\gamma+1)/(2n)$, we can apply (44) to the above inequality to get

$$\begin{aligned}
\operatorname{tr}\left(\Phi(\hat{X} + \alpha\hat{X}', S_\alpha)\right) &\geq \Gamma\rho_\alpha - \alpha\Gamma\rho + \alpha\sigma n\rho + \alpha\rho\left(-(\sigma-1)\Gamma - \tau\gamma^2\theta/2\right) + \alpha^2\operatorname{tr}(T_\alpha) \\
&= \Gamma\rho_\alpha + \alpha\sigma n\rho + \alpha\rho\left(-\sigma\Gamma - \tau\gamma^2\theta/2\right) + \alpha^2\operatorname{tr}(T_\alpha) \\
&= \Gamma\rho_\alpha + \alpha\rho\left(\sigma\gamma^2/2 - \tau\gamma^2\theta/2\right) + \alpha^2\operatorname{tr}(T_\alpha) \\
&= \Gamma\rho_\alpha + \alpha\rho\gamma^2\left(\sigma - \tau\theta\right)/2 + \alpha^2\operatorname{tr}(T_\alpha),
\end{aligned}$$

where the second equality comes from the definition $\Gamma \equiv n - \gamma^2/2$.

From the above inequality, the statement of the proposition will hold if

$$\rho\gamma^2\left(\sigma - \tau\theta\right)/2 + \alpha\operatorname{tr}(T_\alpha) \geq 0, \tag{53}$$

and so we now devote our efforts to establishing (53). We start by providing a bound on $\|T_\alpha\|_F$, which can be obtained as follows from (51), the standard properties of integration, and Proposition 3.2(b):

$$\begin{aligned}
\|T_\alpha\|_F &\leq \int_0^1 (1-t)\left\|\Phi''(\hat{X} + t\alpha\hat{X}', S_{t\alpha})[\hat{X}', \Delta S]^{(2)}\right\|_F dt \\
&\leq \frac{1}{\sqrt{2}}\int_0^1 (1-t)\|\hat{V}_{t\alpha}^{-1}\hat{L}_{t\alpha}\|\left(\|\hat{V}_{t\alpha}^T\hat{X}'\hat{V}_{t\alpha}\|_F + \|\hat{L}_{t\alpha}^{-1}\Delta S\hat{L}_{t\alpha}^{-T}\|_F\right)^2 dt \\
&\leq \frac{1}{\sqrt{2}}\int_0^1 (1-t)\|\hat{V}_{t\alpha}^{-1}\hat{L}_{t\alpha}\|\left(\|V^{-1}\hat{V}_{t\alpha}\|^2\|V^T\hat{X}'V\|_F + \|\hat{L}_{t\alpha}^{-1}L\|^2\|L^{-1}\Delta SL^{-T}\|_F\right)^2 dt,
\end{aligned}$$

where $\hat{V}_{t\alpha} \equiv \hat{V}(\hat{X} + t\alpha\hat{X}')$ and $\hat{L}_{t\alpha} \equiv \hat{L}(S_{t\alpha})$. We note that the hypotheses of Proposition 3.7 hold with $(\hat{X}, S)$, $(\hat{X} + t\alpha\hat{X}', S_{t\alpha})$, and the scalar $t\alpha\tau$ due to Proposition 4.3. Now using Propositions 4.3, 3.7, and 3.6(a) as well as (49) and simple integration with respect to t, the above inequality shows that

$$\begin{aligned}
\|T_\alpha\|_F &\leq \frac{1}{\sqrt{2}}\int_0^1 (1-t)\frac{\|\hat{V}^{-1}\hat{L}\|}{1-t\alpha\tau}\left((1-t\alpha\tau)^{-1}\tau + (1-t\alpha\tau)^{-1}\tau\right)^2 dt \\
&= 2\sqrt{2}\,\tau^2\|V^{-1}L\|\int_0^1 \frac{1-t}{(1-t\alpha\tau)^3}\,dt \\
&\leq 2\sqrt{2}\,\tau^2\left((\gamma+1)\rho\right)\left(2(1-\alpha\tau)\right)^{-1} \\
&\leq \sqrt{2}\,\tau^2(1-\tau)^{-1}(\gamma+1)\rho
\end{aligned}$$

Hence,

$$\begin{aligned}
\rho\gamma^2(\sigma-\tau\theta)/2 + \alpha\operatorname{tr}(T_\alpha) &\geq \rho\gamma^2(\sigma-\tau\theta)/2 - \alpha\left|\operatorname{tr}(T_\alpha)\right| \geq \rho\gamma^2(\sigma-\tau\theta)/2 - \alpha\sqrt{n}\,\|T_\alpha\|_F \\
&\geq \rho\gamma^2(\sigma-\tau\theta)/2 - \alpha\sqrt{2}\tau^2(1-\tau)^{-1}(\gamma+1)\sqrt{n}\rho \\
&= \rho\left(\gamma^2(\sigma-\tau\theta)/2 - \alpha\sqrt{2}\tau^2(1-\tau)^{-1}(\gamma+1)\sqrt{n}\right) \geq 0,
\end{aligned}$$

where the last inequality follows from (48). This completes the proof of the proposition. ∎

Corollary 4.8 *Let $(X, S, y) \in \mathcal{N}(\gamma)$, and suppose $\alpha \geq 0$ satisfies (48). Then $(X_\alpha, S_\alpha, y_\alpha) \in \mathcal{N}(\gamma)$.*

Proof. As discussed in the proof of Proposition 4.7, we have $\alpha \leq 1$, and so $(X_\alpha, S_\alpha) \in \mathcal{S}_{++}^{F?} \times \mathcal{S}_{++}^{F}$ by Corollary 4.4. In addition, the Newton system defining $(\Delta X, \Delta S, \Delta y)$ clearly shows that $(X_\alpha, S_\alpha, y_\alpha) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$.

As was shown in the proof of Corollary 4.4, $\hat{X} + \alpha \hat{X}'$ is a positive definite completion of X_α, where $\hat{X} \equiv \hat{X}(X)$ and $\hat{X}' \equiv \hat{X}'(X)[\Delta X]$. By Theorem 3.3, we have that, among all positive definite completions of X_α, the maximum-determinant completion $\hat{X}(X_\alpha)$ of X_α maximizes the function $\mathrm{tr}(\hat{V}(\cdot)^{-1}\hat{L}(S_\alpha)$, where $\hat{V}(\cdot)$ and $\hat{L}(\cdot)$ are as in Proposition 3.2. Thus, using the fact that $V(\cdot) = \hat{V}(\hat{X}(\cdot))$ and combining Theorem 3.3 and Proposition 4.7, we see

$$\mathrm{tr}(V(X_\alpha)^{-1}L(S_\alpha)) = \mathrm{tr}(\hat{V}(\hat{X}(X_\alpha))^{-1}\hat{L}(S_\alpha)) \geq \mathrm{tr}(\hat{V}(\hat{X} + \alpha\hat{X}')^{-1}\hat{L}(S_\alpha)) \geq \Gamma\rho_\alpha.$$

Combining this inequality with the fact that $(X_\alpha, S_\alpha, y_\alpha) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D)$ and applying Proposition 3.1, we conclude that $(X_\alpha, S_\alpha, y_\alpha) \in \mathcal{N}(\gamma)$. ∎

Theorem 4.9 *Let $n \geq 1$ and constants $\gamma \geq 0$, $\delta \in [0, \sqrt{n})$, and $\tau \in (0, 1)$ be given such that (36) through (41) are satisfied. Suppose that Algorithm SDP is initialized with $(X^0, S^0, y^0) \in \mathcal{N}(\gamma)$ and a tolerance $\varepsilon \geq 0$, and suppose that the constant step-size $\alpha \geq 0$ is used in each iteration of the algorithm, where α is given by the right-hand side of (48). Then, for each $k \geq 0$, the sequence $\{(X^k, S^k, y^k)\}_{k\geq 0}$ produced by the algorithm satisfies the following:*

(a) $(X^k, S^k, y^k) \in \mathcal{N}(\gamma)$;

(b) $\mu(X^k, S^k) \leq (1 - \alpha\,\delta/\sqrt{n})^k \mu(X^0, S^0)$.

As a consequence, Algorithm SDP terminates with a point (X^k, S^k, y^k) satisfying $X^k \bullet S^k \leq \varepsilon$ in at most $\mathcal{O}(n \log(X^0 \bullet S^0/\varepsilon))$ iterations.

Proof. Item (a) follows from Corollary 4.8 and induction on k. Likewise, item (b) follows from the combination of Proposition 4.1 and Proposition 4.6 along with induction on k. The bound on the number of iterations now follows from (b) and the standard argument which shows that the duality gap is reduced in each iteration by a factor which is on the order of $\mathcal{O}(1 - \alpha\,\delta/\sqrt{n})$, which is $\mathcal{O}(1 - 1/n)$ due to (48). ∎

5 Computational Issues and Results

In this section, we discuss several computational issues related to Algorithm-SDP, and then present computational results comparing our method with three other SDP implementations.

5.1 Implementation features of Algorithm SDP

We now demonstrate how the theoretical presentation of Algorithm SDP in Section 4 can be specified into a practical implementation.

First, as is typical with practical primal-dual interior-point algorithms, we implement Algorithm SDP as an infeasible method. Thus, we do not require full primal-dual feasibility of the iterates (X, S, y) but rather only require $X \in \mathcal{S}_{++}^{F?}$ and $S \in \mathcal{S}_{++}^{F}$ and then define the Newton direction $(\Delta X, \Delta S, \Delta y)$ by

$$H'(X, S, y)[\Delta X, \Delta S, \Delta y] = (b - \mathcal{A}(X), C - \mathcal{A}^*(y) - S, \sigma\,\rho\,V - L), \tag{54}$$

where $\rho \equiv \rho(X, S)$ and $0 \leq \sigma < 1$, instead of by (34). In particular, this makes the choice of initial iterate trivial (which in fact is chosen as described in the SDPT3 user's guide [31] in all of the computational results of presented in the following subsections).

Second, we choose a different stopping criterion than originally given for Algorithm SDP. A more practical stopping criterion is based on the relative duality gap and relative feasibility of the current iterate (X, S, y), which we define respectively as

$$\frac{X \bullet S}{1 + (|C \bullet X| + |b^T y|)/2} \qquad \text{and} \qquad \max\left\{ \frac{\|b - \mathcal{A}(X)\|}{1 + \|b\|}, \frac{\|C - \mathcal{A}^*(y) - S\|_F}{1 + \|C\|_F} \right\}.$$

Target values for the gap and feasibility can then be specified at run-time.

Third, in our implementation of Algorithm SDP, we do not bother to stay in the neighborhood $\mathcal{N}(\gamma)$, since our computational experience indicates that this does not yield a substantial practical improvement. Instead, we essentially take α as large as possible while keeping $(X_\alpha, S_\alpha, y_\alpha)$ in $\mathcal{S}^{F?}_{++} \times \mathcal{S}^F_{++} \times \Re^m$. In fact, as is common in SDP implementations, we differentiate two step-sizes, α_p for the primal and α_d for the dual. Then, for the primal and dual separately, the actual step-size is calculated by estimating the infeasible boundary step-size $\bar{\alpha}$ to within an absolute accuracy of 1.0e$-$2 using a simple bisection method and then choosing the step-size slightly less than $\min(\bar{\alpha}, 1)$.

Fourth, we have found it advantageous for reducing the duality gap in the early stages of Algorithm SDP to update X and S in each iteration by performing an alternative linesearch in the spaces of V and L. More specifically, we choose $\alpha_p \leq 1$ and $\alpha_d \leq 1$ so that $L_{\alpha_d} \equiv L(S) + \alpha_d L'(S)[\Delta S]$ and $V_{\alpha_p} \equiv V(X) + \alpha_p V'(X)[\Delta X]$ are close to the boundary of $\mathcal{L}^F_{++}$ and then define $X_{\alpha_p} = \pi^F(V^{-T}_{\alpha_p} V^{-1}_{\alpha_p})$ and $S_{\alpha_d} = L_{\alpha_d} L^T_{\alpha_d}$. (We note that the calculation of X_{α_p} and S_{α_d} can be done efficiently; see below). This update method, however, does not effectively achieve feasibility, and so it is always necessary to revert back to the typical linesearch in the space of X and S.

Fifth, our choice of σ in each iteration is adaptive rather than constant as in the statement of Algorithm SDP. Roughly speaking, we choose σ conservatively whenever we are experiencing small step-sizes but then more aggressively when our step-sizes are larger. In particular, we set $\sigma = 1 - \gamma \min(\alpha_p, \alpha_d)$, where α_p and α_d are the successful step-sizes of the preceding iteration and $\gamma \in [0.1, 0.3]$ is an adaptive parameter that is incrementally increased when step-sizes satisfying $\alpha_p = \alpha_d = 1$ are encountered and incrementally reduced otherwise. As such, our typical values for σ are smaller than those for other SDP implementations. (For example, SDPT3 employs a constant $\gamma = 0.9$.) In addition, it is not immediately clear whether a predictor-corrector approach will improve the choice of σ or even whether such an approach would be computationally efficient (see following subsections), and so our current implementation of Algorithm SDP does not use a predictor-corrector strategy.

Finally, we mention some details regarding the calculation of the Newton direction $(\Delta X, \Delta S, \Delta y)$. As with other primal-dual path-following algorithms, the calculation can be reduced to the solution of the system $M \Delta y = h$, where M is the so-called Schur complement matrix and h is in accordance with the system (54). Here, M is the $m \times m$ matrix representation of the linear operator $-\mathcal{A} \circ V'(X)^{-1} \circ L'(S) \circ \mathcal{A}^*$, so that M is positive definite by Corollary 2.9. Two fundamental techniques for calculating Δy can then be considered: either (1) solution of the system via forward and backward substitution after the direct formation and factorization of M; or (2) solution of the equation via an iterative method. We present numerical results for both methods in later subsections. It is important to note, however, that M has no inherent sparsity (as is typical with other methods) and that M is nonsymmetric (which is atypical). Hence, the first method for calculating Δy requires Gaussian elimination with pivoting, and the second requires an efficient iterative method for nonsymmetric systems (like BiCGSTAB as we have chosen in the computational results). Thus, the ill-conditioning of M near optimality can have a negative impact upon both methods. A natural way to reduce this impact in the case of BiCGSTAB is to perform some pre-conditioning of the linear system, but this has not been implemented in the current version of Algorithm SDP since more investigation is necessary to develop reasonable preconditioning strategies.

Having described the key implementation choices of Algorithm SDP, we now consider the basic operations of the algorithm and in particular discuss their computational complexities. From the statement of Algorithm SDP and the definition of the Newton direction, we see that the main operations are checking $X \in \mathcal{S}^{F?}_{++}$ and $S \in \mathcal{S}^F_{++}$ and evaluating $V(X)$, $V'(X)^{-1}[N]$, $L(S)$, and $L'(S)[B]$, for any $N \in \mathcal{L}^F$ and $B \in \mathcal{S}^F$. To describe the complexity of these operations, we introduce a few definitions. For each $j = 1, \ldots, n$, we define

$$K_j \equiv \{i \in V : (i,j) \in F,\ i \geq j\}.$$

That is, for each $j = 1, \ldots, n$, K_j is the set of row indices of the j-th column of the lower part of F. We have the following fact (detailed in Fukuda et al. [9]), which expresses the chordal structure F as a union of dense blocks, or cliques:

$$F = \cup_{j=1}^n K_j \times K_j. \tag{55}$$

We also define

$$f_2 \equiv \sum_{j=1}^n |K_j|^2.$$

A common way to check if S is in $\mathcal{S}^F_{++}$ is to simply attempt the calculation of $L(S)$ (then $S \in \mathcal{S}^F_{++}$ if and only if the calculation is successful), and standard methods for calculating $L(S)$ show that the time required is $\mathcal{O}(f_2)$. Moreover, in a similar manner, the defining equation (8) shows that $L'(S)[B]$ can also be calculated in time $\mathcal{O}(f_2)$. Hence, each of the key operations involving S is $\mathcal{O}(f_2)$.

To calculate the times required for the key operations involving X, we introduce some additional notation and ideas. First, for any $P, Q \subseteq V$ and any $W \in \mathcal{S}^n$, we let $W_{PQ} \in \Re^{|P| \times |Q|}$ denote the matrix obtained from W by deleting all rows $p \notin P$ and all columns $q \notin Q$. Second, it is clear that (55) can be simplified to $F = \cup_{r=1}^\ell C_r \times C_r$, where $\{C_r\}_{r=1}^\ell$ are the maximal members of $\{K_j\}_{j=1}^n$, i.e., those members of $\{K_j\}_{j=1}^n$ that are not properly contained in any other members. We then define

$$f_3 \equiv \sum_{r=1}^\ell |C_r|^3,$$

and have the following critical theorem, proved in Grone et al. in [12]:

Theorem 5.1 *Let $X \in \mathcal{S}^F$. Then:*

(a) $X \in \mathcal{S}^{F?}_+$ if and only if $X_{C_r C_r} \in \mathcal{S}^{|C_r|}_+$ for all $r = 1, \ldots, \ell$;

(b) $X \in \mathcal{S}^{F?}_{++}$ if and only if $X_{C_r C_r} \in \mathcal{S}^{|C_r|}_{++}$ for all $r = 1, \ldots, \ell$.

This theorem shows immediately that testing if $X \in \mathcal{S}^{F?}_{++}$ can be done in time $\mathcal{O}(f_3)$ by simply attempting the factorizations of the submatrices $X_{C_r C_r}$ of X.

Next, we determine the time required for $V(X)$ and $V'(X)^{-1}[A]$ by considering the following proposition, which gives a formula for $V(X)$ that will be convenient for computation. In the proposition, π^{Fl} is the operator that is defined similarly to π^F except that it projects onto $\mathcal{L}^F$ instead of $\mathcal{S}^F$.

Proposition 5.2 *Let $X \in \mathcal{S}^{F?}_{++}$. Then $V \equiv V(X) \in \mathcal{L}^F_{++}$ uniquely satisfies the equation*

$$\pi^{Fl}(XV) = \pi^{Fl}(V^{-T}). \tag{56}$$

Proof. We first consider solving the simpler equation $\pi^{Fl}(XQ) = 0$ for $Q \in \mathcal{L}^F$. Because the j-th column of this equation can be expressed compactly as $X_{K_jK_j}q_j = 0$, where q_j is the nonzero part of $Q_{\cdot j}$, we conclude from Theorem 5.1(b) that $q_j = 0$, which implies that $Q = 0$.

Now suppose $V_1, V_2 \in \mathcal{L}^F_{++}$ each satisfy (56), and note that the right-hand side of (56) is a diagonal matrix. For $i = 1, 2$, let D_i be the diagonal matrix defined by the diagonal entries of V_i. Then (56) implies that $\pi^{Fl}(XV_1D_1) = \pi^{Fl}(XV_2D_2)$, or equivalently that $\pi^{Fl}(X(V_1D_1 - V_2D_2)) = 0$. Using the result of the previous paragraph, this shows $V_1D_1 = V_2D_2$. Examining the jj-th position of this equation and employing the definition of D_i, we see that $[V_1]_{jj} = [V_2]_{jj}$, which in turn implies that $D_1 = D_2$ and hence $V_1 = V_2$. In other words, at most one $V \in \mathcal{L}^F_{++}$ satisfies (56).

We now show that $V \equiv V(X)$ satisfies (56), which will complete the proof of the proposition. For $i \geq j$ such that $(i, j) \in F$, consider the ij-th entry of the matrix $(\hat{X} - X)V$, where $\hat{X} \equiv \hat{X}(X)$:

$$[(\hat{X} - X)V]_{ij} = \sum_{k=j}^{n} [\hat{X} - X]_{ik} V_{kj}.$$

We claim that the above expression equals zero. So suppose for contradiction that it is nonzero. Then there exists $k \geq j$ such that $[\hat{X} - X]_{ik} \neq 0$ and $V_{kj} \neq 0$. Because $\pi^F(\hat{X}) = X$ and $V \in \mathcal{L}^F_{++}$, this implies that $(i, k) \notin F$ and $(k, j) \in F$. However, due to the chordal structure of F and the fact that the ordering $(1, \ldots, n)$ is a perfect elimination ordering for F, we have that $(i, j), (k, j) \in F$ imply $(i, k) \in F$, which is a contradiction. Hence, the above expression equals 0. Said differently, we have $\pi^{Fl}((\hat{X} - X)V) = 0$, which implies $\pi^{Fl}(XV) = \pi^{Fl}(\hat{X}V) = \pi^{Fl}(V^{-T})$, as desired. ∎

As the proposition and its proof indicate, the nonzero part $v_j \in \Re^{|K_j|}$ of the j-th column of V is simply the solution of the system $X_{K_jK_j}v_j = V_{jj}^{-1}e_1$, where $e_1 \in \Re^{|K_j|}$ has a one in its first position and zeros elsewhere. Hence, as long as the factorizations of $X_{K_jK_j}$ for $j = 1, \ldots, n$ are readily available, the calculation of V can be done in time $\mathcal{O}(f_2)$.

Are these factorizations readily available, however? The operation to verify $X \in \mathcal{S}^{F?}_{++}$ only yields the factorizations of $X_{C_rC_r}$ for $r = 1, \ldots, \ell$, and so the factorizations of $X_{K_jK_j}$ are not explicitly available. This is not a significant obstacle, however, since it is possible to reorder the vertices V (in a preprocessing phase, for example) so that factorizations of the matrices $X_{K_jK_j}$ are embedded in a natural manner in the upper Cholesky factorizations of the matrices $X_{C_rC_r}$. Moreover, this reordering can be done without altering the chordal structure of F or the property that $(1, \ldots, n)$ is a perfect elimination ordering. This property is discussed in detail in section 2.1 of [9] and section 2.2 of [26], where it is described as a *perfect elimination ordering induced from an ordering of maximal cliques satisfying the running intersection property*, and this feature has been implemented into Algorithm SDP.

Differentiating equation (56) with respect to X in the direction $A \in \mathcal{S}^F$ and defining $N \equiv V'(X)[A]$, we see that

$$\pi^{Fl}(AV) = -\pi^{Fl}(V^{-T}N^TV^{-T}) - \pi^{Fl}(XN).$$

Note that the first term on the right-hand side does not require the full matrix $V^{-T}N^TV^{-T}$ but rather just its diagonal. The above equation also provides a convenient form for calculating $A = V'(X)^{-1}[N]$ for an arbitrary $N \in \mathcal{L}^F$, once X and V are available. In fact, it is not difficult to see that A can be computed from N in time $\mathcal{O}(f_2)$.

We summarize the complexities obtained from the preceding discussion in the following proposition.

Proposition 5.3 *Let $X, S \in \mathcal{S}^F$. Determining if X is in $\mathcal{S}^{F?}_{++}$ requires time $\mathcal{O}(f_3)$, and under the mild assumption that certain calculations are stored upon the successful determination of $X \in \mathcal{S}^{F?}_{++}$,*

calculation of $V(X)$ and $V'(X)^{-1}[N]$, for any $N \in \mathcal{L}^F$, require time $\mathcal{O}(f_2)$. Determining if S is in $\mathcal{S}^F_{++}$ is performed by attempting the computation of $L(S)$, which requires time $\mathcal{O}(f_2)$. Upon the successful computation of $L(S)$, the calculation of $L'(S)[B]$, for any $B \in \mathcal{S}^F$, requires time $\mathcal{O}(f_2)$.

5.2 Comparison with a Standard Primal-Dual Method

In order to see how Algorithm SDP compares with other primal-dual path-following implementations, in this subsection we compare Algorithm SDP with SDPT3 version 3.0, a successful implementation by Todd, Toh, and Tütüncü (see [31]). We have chosen to run SDPT3 with the HRVW/KSH/M direction using the Mehrotra predictor-corrector strategy. Both algorithms use the same starting point and terminate when the relative duality gap is less than 1.0e−4 and the relative feasibility is less than 1.0e−5.

We remark that these moderate values for the target gap and feasibility have been chosen for two reasons. First, it makes our presentation consistent with later subsections where larger SDPs are considered and solved to the same accuracies. Second, we have chosen moderate values in keeping with our discussion of the previous subsection concerning how the ill-conditioning of M near optimality affects the calculation of Δy in Algorithm SDP. In fact, for all but a few of the problems presented below (most notably the *control* and *qap* instances), Algorithm SDP can easily obtain even higher accuracy.

For these computational results, we solve the system $M\Delta y = h$ in Algorithm SDP using either Gaussian elimination or BiCGSTAB, depending on the stage of the algorithm. In the early stages of the algorithm when the conditioning of M is good, we employ BiCGSTAB. As soon as the number of multiplications by M (or equivalently, the number of evaluations of $-\mathcal{A} \circ V'(X)^{-1} \circ L'(S) \circ \mathcal{A}^*$) exceeds m during one call to the BiCGSTAB subroutine, however, we switch to Gaussian elimination. Assuming that evaluations of the functions are $\mathcal{A}$ and $\mathcal{A}^*$ require time $\mathcal{O}(f_2)$ (as is the case for most of the test problems below), the direct method requires time $\mathcal{O}(mf_2)$ to form the matrix M and then an additional $\mathcal{O}(m^3)$ time to factor M and solve for Δy. We have thus chosen problems having $mf_2 + m^3 \leq$ 1.0e+9 as a heuristic guide for selecting problems for which the direct method is not too computationally intensive. In particular, no problems having $m > 1000$ have been selected.

The test problems come from the SDPLIB collection of problems maintained by Borchers [3], and their statistics are listed in Table 1. The first three columns are self-explanatory, and the last two give the percentage of nonzeros in the iterates S and X represented by the density patterns E and F. Complete computational results on a 2.4 GHz Pentium 4 computer are given in Table 2 and are self-explanatory, with times given in seconds.

We remark that, although Algorithm SDP is designed primarily for sparse problems, i.e., when F is a relatively small subset of $V \times V$, it can of course be applied to dense problems with $F = V \times V$, as with a few of the test problems in Table 1. We have included these problems because we feel it is instructive to compare the performance of Algorithm SDP and SDPT3 on such instances.

The results indicate several interesting points. First and foremost, both methods were able to solve all problems to the desired accuracy in a reasonable amount of time. Second, Algorithm SDP outperformed SDPT3 on several sets problems (the *arch*, *max*, *qp*, and *ss* problems, as well as a subset of the *mcp* problems) for which the chordal pattern F was very small, indicating Algorithm SDP's capability for exploiting this structure. On the other hand, SDPT3 significantly outperformed Algorithm SDP on the *control* and *qap* problems, as Algorithm SDP was challenged by the conditioning and density of these problems. In addition, SDPT3 was faster on the remaining *mcp* problems, most likely due to the related fact that Algorithm SDP consistently required more iterations than SDPT3, which itself is indicative of the strong convergence properties of the

problem	n	m	dens E (%)	dens F (%)
arch2	335	174	3.29	6.87
arch4	335	174	3.29	6.87
arch8	335	174	3.29	6.87
control5	75	351	45.61	46.88
control6	90	496	45.42	46.76
control7	105	666	45.28	46.68
control8	120	861	45.18	46.43
gpp100	100	101	100.00	100.00
gpp124-2	124	125	100.00	100.00
gpp124-3	124	125	100.00	100.00
gpp124-4	124	125	100.00	100.00
maxG11	800	800	0.75	2.61
mcp250-2	250	250	2.75	15.27
mcp250-3	250	250	4.89	36.76
mcp250-4	250	250	8.51	58.63
mcp500-1	500	500	0.90	2.58
mcp500-2	500	500	1.38	11.99
qap7	50	358	100.00	100.00
qap8	65	529	100.00	100.00
qap9	82	748	100.00	100.00
qpG11	1600	800	0.25	0.73
ss30	426	132	4.43	9.85
theta1	50	104	100.00	100.00
theta2	100	498	100.00	100.00
truss5	331	208	3.31	3.31
truss6	451	172	0.88	0.88
truss7	301	86	0.99	0.99
truss8	628	496	3.18	3.18

Table 1: Statistics for SDPLIB Test Problems

PROBLEM	OBJECTIVE VALUE				R-GAP		R-FEAS		ITER		TIME (s)	
	AS		T3		AS	T3	AS	T3	AS	T3	AS	T3
	PRIMAL	DUAL	PRIMAL	DUAL								
arch2	−6.71397e−01	−6.71546e−01	−6.71485e−01	−6.71523e−01	8.9e−05	3.8e−05	7.6e−13	8.4e−12	50	17	8	20
arch4	−9.72501e−01	−9.72657e−01	−9.72561e−01	−9.72644e−01	7.9e−05	8.3e−05	1.4e−11	2.9e−11	52	18	10	20
arch8	−7.05642e+00	−7.05716e+00	−7.05689e+00	−7.05702e+00	9.2e−05	1.7e−05	1.4e−10	3.1e−10	42	17	8	19
control5	−1.68824e+01	−1.68838e+01	−1.68835e+01	−1.68836e+01	7.8e−05	1.1e−05	1.4e−08	1.3e−07	52	18	37	11
control6	−3.73018e+01	−3.73048e+01	−3.73035e+01	−3.73047e+01	7.8e−05	3.0e−05	1.7e−08	4.4e−08	52	19	95	21
control7	−2.06237e+01	−2.06253e+01	−2.06249e+01	−2.06251e+01	7.6e−05	1.3e−05	2.7e−08	1.2e−07	50	21	207	43
control8	−2.02856e+01	−2.02865e+01	−2.02861e+01	−2.02865e+01	4.6e−05	2.0e−05	1.5e−08	5.5e−08	51	20	429	71
gpp100	4.49437e+01	4.49435e+01	4.49445e+01	4.49435e+01	2.2e−06	2.3e−05	8.2e−06	3.1e−10	42	12	7	1
gpp124-2	4.68623e+01	4.68623e+01	4.68639e+01	4.68623e+01	8.1e−07	3.5e−05	7.0e−06	4.2e−10	44	12	15	2
gpp124-3	1.53014e+02	1.53014e+02	1.53017e+02	1.53014e+02	2.0e−06	2.1e−05	7.8e−06	4.1e−10	41	11	15	2
gpp124-4	4.18990e+02	4.18987e+02	4.19008e+02	4.18987e+02	5.7e−06	5.1e−05	9.1e−06	1.5e−09	38	12	16	2
maxG11	−6.29127e+02	−6.29165e+02	−6.29127e+02	−6.29165e+02	6.1e−05	6.0e−05	5.6e−06	9.8e−16	39	11	20	65
mcp250-2	−5.31888e+02	−5.31932e+02	−5.31926e+02	−5.31930e+02	8.2e−05	8.3e−06	2.3e−06	7.1e−16	34	10	5	4
mcp250-3	−9.81095e+02	−9.81175e+02	−9.81168e+02	−9.81173e+02	8.1e−05	5.0e−06	3.2e−06	1.3e−15	34	10	26	4
mcp250-4	−1.68183e+03	−1.68196e+03	−1.68188e+03	−1.68196e+03	8.0e−05	5.0e−05	8.5e−06	1.8e−15	34	9	52	5
mcp500-1	−5.98103e+02	−5.98154e+02	−5.98122e+02	−5.98154e+02	8.5e−05	5.4e−05	2.9e−06	7.9e−16	35	11	3	19
mcp500-2	−1.06996e+03	−1.07006e+03	−1.07003e+03	−1.07006e+03	9.0e−05	2.3e−05	2.0e−06	4.0e−15	35	11	57	23
qap7	4.24833e+02	4.24811e+02	4.24763e+02	4.24787e+02	7.1e−05	2.1e−05	4.3e−06	2.7e−10	40	12	29	2
qap8	7.56972e+02	7.56941e+02	7.56778e+02	7.56863e+02	6.6e−05	1.3e−05	9.3e−06	3.1e−10	39	12	33	4
qap9	1.40999e+03	1.40991e+03	1.40988e+03	1.40986e+03	7.1e−05	7.6e−05	8.0e−06	4.3e−10	39	11	100	7
qpG11	−2.44858e+03	−2.44866e+03	−2.44864e+03	−2.44866e+03	3.1e−05	8.6e−06	5.4e−06	2.5e−15	39	12	27	281
ss30	−2.02378e+01	−2.02396e+01	−2.02389e+01	−2.02396e+01	8.7e−05	3.2e−05	3.0e−06	5.2e−11	46	17	7	71
theta1	−2.29992e+01	−2.30002e+01	−2.29998e+01	−2.30000e+01	4.5e−05	1.1e−05	4.9e−06	2.6e−13	39	9	1	0
theta2	−3.28782e+01	−3.28793e+01	−3.28776e+01	−3.28795e+01	3.3e−05	5.8e−05	1.2e−15	1.9e−12	42	9	48	3
truss5	1.32638e+02	1.32630e+02	1.32636e+02	1.32635e+02	6.4e−05	1.3e−05	6.5e−10	5.7e−08	45	15	5	5
truss6	9.01022e+02	9.00985e+02	9.00999e+02	9.00999e+02	4.1e−05	6.8e−06	2.6e−08	6.0e−07	43	21	2	3
truss7	9.00038e+02	8.99999e+02	9.00012e+02	9.00000e+02	4.4e−05	1.5e−05	1.3e−11	2.2e−07	42	22	0	2
truss8	1.33117e+02	1.33105e+02	1.33114e+02	1.33113e+02	8.8e−05	1.1e−05	5.1e−09	3.7e−07	48	14	66	27

Table 2: Results of Algorithm SDP (AS) and SDPT3 (T3) on the SDPLIB Test Problems

HRVW/KSH/M direction when combined with the Merhrotra predictor-corrector strategy.

5.3 Comparison with the Completion Method

In this subsection, we compare Algorithm SDP with the completion method (CM) of Fukuda et al. on problems for which the large size of m requires the solution of the Schur complement equation by an iterative method. As in the previous subsection, Algorithm SDP is run with BiCGSTAB as the iterative solver, and the SDPs are solved to an accuracy of 1.0e−4 for the relative duality gap and 1.0e−5 for relative feasibility.

As described in [9, 26], CM stores X and S in the same sparse format as Algorithm SDP, and the search direction in each iteration is the sparse projection of the HRVW/KSH/M direction. Moreover, the sparsity of X and S is exploited in the formation of the Schur complement matrix M, which is then factored directly to solve for Δy. Here, however, we have implemented our own version of CM which computes Δy using an iterative method, namely the conjugate gradient method, which is appropriate since, in this case, M is symmetric positive definite. Other algorithmic choices for our implementation of CM mimic those of SDPT3 of the previous subsection, except that the predictor-corrector method has not been implemented due to its need to solve an extra $m \times m$ system in each iteration.

In CM, multiplication by M is equivalent to evaluation of the operator $\mathcal{A}(\hat{X}\mathcal{A}^*(\cdot)S^{-1})$, where $\hat{X} \equiv \hat{X}(X)$. It is not difficult to see that, using the sparse matrices V and L and assuming that evaluations of $\mathcal{A}$ and $\mathcal{A}^*$ require time $\mathcal{O}(f_2)$ (as assumed in the previous subsection), this operator can be evaluated in time $\mathcal{O}(nf)$, where $f = \sum_{j=1}^{n} |K_j|$, by using sparse triangular solves. We thus have that

$$f_2 = \sum_{j=1}^{n} |K_j|^2 < \sum_{j=1}^{n} n|K_j| = nf.$$

Since f_2 is the time required to multiply by M in Algorithm SDP, this demonstrates an advantage of Algorithm SDP over CM.

For comparing Algorithm SDP with CM, we have chosen fifteen test problems from several sources; the problems are listed in Table 3. The first ten are Lovász theta SDPs based on graphs used in the Second DIMACS Challenge [16]. The first nine use the original SDP formulation of Lovász (see [21]) in which both E and F are completely dense due to the fact that C is the matrix of all ones, while the tenth employs a different formulation (see [19]) which better respects the sparsity of the underlying graph. Experimentally, we have found that the general conditioning of the first formulation is better than that of the second, and so the first formulation should be preferred whenever the underlying graph is relatively dense (which is the case for the first nine but not the tenth). In addition, in the case of the first nine problems, we have that f_2 and nf are both $\mathcal{O}(n^3)$, which allows us to compare Algorithm SDP and CM on a more equal footing for these problems. The next three, the so-called *vibra* problems, were studied in [18], and the final two were test problems in the Seventh DIMACS Challenge [7]. Both methods were given an upper bound of 5 hours computation time and so were terminated after the first iteration past this time limit, if necessary. The computational results are listed in Table 4.

The table shows that Algorithm SDP converged on all problems except the *vibra* problems, for which achieving both small gap and small feasibility was difficult. (We remark that Algorithm SDP was terminated when progress became slow.) Interestingly, however, on all three *vibra* problems, the objective values achieved by Algorithm SDP were close to optimal (see [18]). CM also failed to converge on the *vibra* problems as well as *copo23*. In terms of running times, Algorithm SDP outperformed CM in nearly all instances — approximately twice as fast for the first nine and an

problem	n	m	dens E (%)	dens F (%)
brock200-1.co	200	5067	100.00	100.00
brock200-4.co	200	6812	100.00	100.00
c-fat200-1.co	200	18367	100.00	100.00
hamming6-4.co	64	1313	100.00	100.00
hamming8-4.co	256	11777	100.00	100.00
johnson16-2-4.co	120	1681	100.00	100.00
keller4.co	171	5101	100.00	100.00
san200-0.7-1.co	200	5971	100.00	100.00
sanr200-0.7.co	200	6033	100.00	100.00
MANN-a27.co	379	1081	2.03	3.12
vibra3	1185	544	0.62	1.53
vibra4	2545	1200	0.30	0.90
vibra5	6801	3280	0.11	0.45
copo14	560	1275	1.17	1.17
copo23	2300	5820	0.31	0.31

Table 3: Statistics for Test Problems with Large m

order of magnitude faster or more for the last six problems.

5.4 Comparison with a First-Order Method

Finally, we compare Algorithm SDP with the first-order method (BMZ) of Burer, Monteiro, and Zhang [6]. BMZ is a dual-only method that solves a special class of so-called "fixed-diagonal" SDPs by optimizing the log-barrier function for a decreasing sequence of barrier parameters $\{\nu_k\}_{k\geq 0}$ using a first-order, gradient-based nonlinear programming approach. Two of the key features of BMZ are that it works only with S and L and that its function and gradient evaluations each take time $\mathcal{O}(f_2)$, which matches the complexity of the fundamental operations of Algorithm SDP.

We compare Algorithm SDP and BMZ on the fourteen problems shown in Table 5. The first two problems (which come from the Gset test problem suite [14]) and last six problems (which come from the Seventh DIMACS Challenge) are Lovász theta SDPs and employ the sparse formulation (as mentioned in the previous subsection). The remaining problems are maximum cut SDP relaxations (see [11]) and also come from the Seventh DIMACS Challenge. Each method was given an upper bound of 5 hours running time on each problem and was thus terminated upon completion of the first iteration after 5 hours, if necessary. BMZ was stopped once the log-barrier subproblem corresponding to $\nu = 1.0\text{e}{-4}$ was solved, which yielded a comparable accuracy to Algorithm SDP's stopping criterion. The computational results are shown in Table 6.

The results show that both methods had difficulty solving such large problems in the time allotted. Even still, when comparing objective values, on twelve of the fourteen problems, Algorithm SDP made more progress towards optimality than BMZ. An advantage of BMZ, of course, is that each iterate is dual feasible, while an advantage of Algorithm SDP is that primal information is produced in addition to dual information.

6 Concluding Remarks

The results of this paper involve both theoretical and practical aspects of solving SDPs. Theoretically, we have shown that it is possible to express the central path using sparse equations rather than the usual dense ones, and we have moreover shown how to measure the proximity of a partial

PROBLEM	OBJECTIVE VALUE				R-GAP		R-FEAS		ITER		TIME (s)	
	AS		CM		AS	CM	AS	CM	AS	CM	AS	CM
	PRIMAL	DUAL	PRIMAL	DUAL								
brock200-1.co	−2.74560e+01	−2.74570e+01	−2.74554e+01	−2.74571e+01	3.7e−05	6.0e−05	9.4e−06	4.8e−07	44	20	1170	2647
brock200-4.co	−2.12929e+01	−2.12938e+01	−2.12926e+01	−2.12940e+01	4.1e−05	6.5e−05	8.8e−06	4.6e−07	44	18	1165	2466
c-fat200-1.co	−1.19995e+01	−1.20000e+01	−1.19996e+01	−1.20000e+01	4.4e−05	3.7e−05	5.0e−06	4.8e−07	44	17	717	6097
hamming6-4.co	−5.33333e+00	−5.33334e+00	−5.33320e+00	−5.33340e+00	1.7e−06	3.3e−05	9.6e−06	9.9e−08	41	12	1	0
hamming8-4.co	−1.60000e+01	−1.60000e+01	−1.59999e+01	−1.60001e+01	5.0e−06	1.1e−05	7.9e−06	3.2e−07	46	18	510	89
johnson16-2-4.co	−7.99993e+00	−8.00035e+00	−7.99999e+00	−8.00010e+00	4.5e−05	1.3e−05	3.7e−06	3.0e−14	41	12	21	3
keller4.co	−1.40119e+01	−1.40125e+01	−1.40121e+01	−1.40123e+01	4.5e−05	1.4e−05	7.1e−06	4.7e−07	44	19	861	1663
san200-0.7-1.co	−3.00001e+01	−3.00000e+01	−2.99991e+01	−3.00000e+01	1.6e−05	3.1e−05	7.5e−06	4.3e−07	45	22	311	433
sanr200-0.7.co	−2.38355e+01	−2.38365e+01	−2.38357e+01	−2.38364e+01	3.9e−05	3.2e−05	8.5e−06	4.2e−07	44	19	1191	3696
MANN-a27.co	−1.32761e+02	−1.32763e+02	−1.32751e+02	−1.32764e+02	1.8e−05	9.7e−05	1.8e−06	2.3e−06	42	31	3	270
vibra3	−1.72333e+02	−1.72684e+02	−1.72458e+02	−1.72648e+02	2.0e−03	1.1e−03	4.9e−06	8.3e−07	58	98	94	19071
vibra4	−1.65349e+02	−1.65724e+02	−1.24071e+02	−1.89703e+02	2.2e−03	4.2e−01	4.3e−09	4.4e−06	64	57	1044	18188
vibra5	−1.65570e+02	−1.66063e+02	−9.34045e+00	−7.96259e+02	4.0e−03	8.2e+02	1.5e+00	8.1e+02	56	8	5500	18598
copo14	2.65890e−05	−7.14668e−05	1.11364e−05	−2.92301e−05	9.9e−05	4.0e−05	1.3e−06	4.2e−07	50	21	36	603
copo23	−5.27584e−05	−6.03132e−05	2.52874e−02	−4.42279e−02	8.5e−05	6.7e−02	6.8e−06	3.8e−07	63	23	822	22684

Table 4: Results of Algorithm SDP (AS) and the Completion Method (CM) on Test Problems with Large m

problem	n	m	dens E (%)	dens F (%)
G43	1001	10991	2.39	55.46
G51	1001	6910	1.58	16.12
brock400-1.co	401	20478	25.90	90.92
p-hat300-1.co	301	34218	75.95	98.28
toruspm3-8-50	512	512	1.56	15.60
torusg3-8	512	512	1.56	15.60
toruspm3-15-50	3375	3375	0.24	6.08
torusg3-15	3375	3375	0.24	6.08
hamming-07-5-6	129	1921	24.44	73.71
hamming-08-3-4	257	16385	50.19	93.27
hamming-09-5-6	513	54273	41.55	92.82
hamming-09-8	513	2817	2.53	15.98
hamming-10-2	1025	24065	4.77	38.40
hamming-11-2	2049	58369	2.88	36.57

Table 5: Statistics for Maxcut and Theta Test Problems

primal-dual solution to the central path, which was a question left open in [9]. Combining these ideas, we have also shown how to solve the SDP in polynomial-time using a "partial" Newton direction. Even still, there seem to be many interesting theoretical questions left open by the ideas presented in this paper. For example, can the nonsingularity of H' be established under weaker conditions than presented in Theorem 2.10? Or can a wider neighborhood of the central path be used to improve the iteration complexity of the method? (The relatively small step-size established in Section 4 was forced by the neighborhood, not by positive semidefiniteness of the new iterates.) Or can other directions with better properties be defined?

Of course, one of the most appealing aspects of applying the idea of matrix completions to SDP is the prospect of actually solving sparse SDPs more efficiently, and the results of Section 5 indicate that the algorithm proposed in this paper is highly effective on varying classes of problems — especially for those having a small density pattern F. An area of further investigation for Algorithm SDP is the conditioning of the Schur complement matrix M near optimality, particularly as it affects the convergence of the BiCGSTAB subroutine. Currently, it is unclear how preconditioning techniques can be best employed to mitigate the inevitable ill-conditioning.

Overall, we feel that Algorithm SDP makes a significant contribution to the existing algorithms for SDP, allowing one to solve any SDP in a primal-dual framework while taking advantage of sparsity in all stages of computation.

Acknowledgements

The author would like to thank Stephen Vavasis and Kim Chuan Toh for helpful comments that have improved the presentation of Section 5. In addition, the author is in debt to two anonymous referees, who provided extremely detailed and insightful comments on the first and second drafts of this paper.

References

[1] F. Alizadeh, J.-P.A. Haeberly, and M.L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM Journal*

PROBLEM	OBJECTIVE VALUE			GAP MEAS		R-FEAS		ITER		TIME (s)	
	AS		BMZ	AS (r-gap)	BMZ (ν)	AS	BMZ	AS	BMZ	AS	BMZ
	PRIMAL	DUAL									
G43	−2.82795e+02	−2.81556e+02	−4.83252e+02	5.3e−02	1.0e−02	1.6e−01	0.0e+00	29	737	18221	18029
G51	−3.62057e+02	−3.50632e+02	−3.49226e+02	2.4e−02	1.0e−04	1.6e−01	0.0e+00	30	4783	20317	18003
brock400-1.co	−3.96791e+01	−3.97140e+01	−4.05332e+01	8.6e−04	1.0e−04	3.9e−06	0.0e+00	37	7401	20819	18001
p-hat300-1.co	−9.66567e+00	−1.01670e+01	−1.93734e+01	1.1e−03	1.0e−03	8.7e−05	0.0e+00	37	15455	19039	18003
toruspm3-8-50	−5.27764e+02	−5.27809e+02	−5.27810e+02	8.7e−05	1.0e−05	5.1e−06	0.0e+00	36	7814	88	510
torusg3-8	−4.57317e+07	−4.57359e+07	−4.57360e+07	9.2e−05	1.0e−05	5.0e−06	0.0e+00	36	5692	96	521
toruspm3-15-50	−3.47481e+03	−3.47513e+03	−3.47516e+03	9.2e−05	1.0e−04	5.4e−06	0.0e+00	39	2559	21460	18008
torusg3-15	−3.13381e+08	−3.13457e+08	−3.13458e+08	2.4e−04	1.0e−04	9.7e−06	0.0e+00	38	2829	20996	18005
hamming-07-5-6	−4.26654e+01	−4.26667e+01	−4.26677e+01	1.3e−05	1.0e−05	6.0e−06	0.0e+00	37	7405	38	117
hamming-08-3-4	−2.55381e+01	−2.56004e+01	−2.56700e+01	4.7e−05	1.0e−04	6.2e−06	0.0e+00	38	27447	2794	18001
hamming-09-5-6	−8.47006e+01	−8.54005e+01	−1.11574e+02	3.3e−02	1.0e−02	4.1e−02	0.0e+00	30	4091	18112	18011
hamming-09-8	−2.23997e+02	−2.24000e+02	−2.24003e+02	1.3e−05	1.0e−05	8.4e−06	0.0e+00	39	5674	78	334
hamming-10-2	−1.02183e+02	−1.02402e+02	−1.02427e+02	4.3e−05	1.0e−05	3.1e−06	0.0e+00	40	4431	14463	18002
hamming-11-2	−1.22074e+02	−1.65422e+03	−2.22378e+02	2.1e+00	1.0e−03	1.5e+03	0.0e+00	24	404	19540	18029

Table 6: Results of Algorithm SDP (AS) and the Burer-Monteiro-Zhang First-Order Method (BMZ) on Maxcut and Theta Test Problems

on Optimization, 8:746–768, 1998.

[2] S. Benson, Y. Ye, and X. Zhang. Solving large-scale sparse semidefinite programs for combinatorial optimization. *SIAM Journal on Optimization*, 10:443–461, 2000.

[3] B. Borchers. SDPLIB 1.2, a library of semidefinite programming test problems. *Optimization Methods and Software*, 11(1):683–690, 1999.

[4] S. Burer and R.D.C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Manuscript, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA, March 2001. Revised August 2001. To appear in *Mathematical Programming*.

[5] S. Burer, R.D.C. Monteiro, and Y. Zhang. A computational study of a gradient-based log-barrier algorithm for a class of large-scale SDPs. Manuscript, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlatna, GA 30332, USA, June 2001. Revised January 2002. To appear in *Mathematical Programming*.

[6] S. Burer, R.D.C. Monteiro, and Y. Zhang. Solving a class of semidefinite programs via nonlinear programming. *Mathematical Programming*, 93:97–122, 2002.

[7] Seventh DIMACS implementation challenge on semidefinite and related optimization problems, November 2000. See the website: `http://dimacs.rutgers.edu/Challenges/`.

[8] M. Fukuda and M. Kojima. Interior-point methods for lagrangian duals of semidefinite programs. Manuscript, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguru-ku, Tokyo 152, Japan, December 2000.

[9] M. Fukuda, M. Kojima, K. Murota, and K. Nakata. Exploiting sparsity in semidefinite programming via matrix completion I: General framework. *SIAM Journal on Optimization*, 11(3):647–674, 2001.

[10] D.R. Fulkerson and O.A. Gross. Incidence matrices and interval graphs. *Pacific Journal of Mathematics*, 15:835–855, 1965.

[11] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of ACM*, pages 1115–1145, 42 (1995).

[12] R. Grone, C.R. Johnson, E.M. Sá, and H. Wolkowicz. Positive definite completions of partial hermitian matrices. *Linear Algebra Applications*, 58:109–124, 1984.

[13] C. Helmberg and K.C. Kiwiel. A spectral bundle method with bounds. ZIB Preprint SC-99-37, Konrad-Zuse-Zentrum, Berlin, Germany, December 1999. To appear in *Mathematical Programming*.

[14] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10:673–696, 2000.

[15] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 6:342–361, 1996.

[16] D. Johnson and M. Trick. *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge*. American Mathematical Society, 1996.

[17] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7:86–125, 1997.

[18] M. Kočvara and M. Stingl. Pennon: A code for convex nonlinear and semidefinite programming. Research Report 290, Institute of Applied Mathematics, University of Erlangen, 2002.

[19] M. Laurent, S. Poljak, and F. Rendl. Connections between semidefinite relaxations of the max-cut and stable set problems. *Mathematical Programming*, 77:225–246, 1997.

[20] C-J. Lin and R. Saigal. On solving large scale semidefinite programming problems: a case study of quadratic assigment problem. Technical Report, Dept. of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2177, 1997.

[21] L. Lovász. On the Shannon Capacity of a graph. *IEEE Transactions of Information Theory*, IT-25(1):1–7, January 1979.

[22] R.D.C. Monteiro. Primal-dual path following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7:663–678, 1997.

[23] R.D.C. Monteiro. Polynomial convergence of primal-dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions. *SIAM Journal on Optimization*, 8:797–812, 1998.

[24] R.D.C. Monteiro and T. Tsuchiya. Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming. *SIAM Journal on Optimization*, 9:551–577, 1999.

[25] R.D.C. Monteiro and Y. Zhang. A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. *Mathematical Programming*, 81:281–299, 1998.

[26] K. Nakata, K. Fujisawa, M. Fukuda, M. Kojima, and K. Murota. Exploiting sparsity in semidefinite programming via matrix completion II: Implementation and numerical results. Manuscript, Department of Applied Physics, University of Tokyo, Tokyo, Japan, February 2001. Revised July 2001. To appear in *Mathematical Programming, Series B*.

[27] K. Nakata, K. Fujisawa, and M. Kojima. Using the conjugate gradient method in interior-points for semidefinite programs. *Proceedings of the Institute of Statistical Mathematics*, 46:297–316, 1998. In Japanese.

[28] Y. E. Nesterov and M. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8:324–364, 1998.

[29] K.C. Toh and M. Kojima. Solving some large scale semidefinite programs via the conjugate residual method. *SIAM Journal on Optimization*, 12:669–691, 2002.

[30] L. Tuncel. Potential reduction and primal-dual methods. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, pages 235–265. Kluwer Academic Publishers, 2000.

[31] R. H. Tütüncü, K. C. Toh, and M. J. Todd. *SDPT3: A Matlab software package for semidefinite-quadratic-linear programming, version 3.0*, August 2001. Available from `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`.

[32] Y. Zhang. On extending some primal–dual interior–point algorithms from linear programming to semidefinite programming. *SIAM Journal on Optimization*, 8:365–386, 1998.