Interior-Point Algorithms for Semidefinite Programming Based on a Nonlinear Formulation*

Samuel Burer† Renato D.C. Monteiro‡ Yin Zhang§

December 1999
Revised May 2001

Abstract

Recently in [5], the authors of this paper introduced a nonlinear transformation to convert the positive definiteness constraint on an $n \times n$ matrix-valued function of a certain form into the positivity constraint on n scalar variables while keeping the number of variables unchanged. Based on this transformation, they proposed a first-order interior-point algorithm for solving a special class of linear semidefinite programs. In this paper, we extend this approach and apply the transformation to general linear semidefinite programs, producing nonlinear programs that have not only the n positivity constraints, but also n additional nonlinear inequality constraints. Despite this complication, the transformed problems still retain most of the desirable properties. We propose first-order and second-order interior-point algorithms for this type of nonlinear program and establish their global convergence. Computational results demonstrating the effectiveness of the first-order method are also presented.

Keywords: semidefinite program, semidefinite relaxation, nonlinear programming, interior-point methods.

AMS 1991 subject classification: 90C06, 90C27, 90C30.

1 Introduction

Semidefinite programming (SDP) is a generalization of linear programming in which a linear function of a matrix variable X is maximized or minimized over an affine subspace

*Computational results reported in this paper were obtained on an SGI Origin2000 computer at Rice University acquired in part with support from NSF Grant DMS-9872009.

†School of Mathematics, Georgia Tech., Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grants INT-9910084 and CCR-9902010. (E-mail: burer@math.gatech.edu).

‡School of ISyE, Georgia Tech., Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grants INT-9600343, INT-9910084 and CCR-9902010. (Email: monteiro@isye.gatech.edu).

§Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA. This author was supported in part by DOE Grant DE-FG03-97ER25331, DOE/LANL Contract 03891-99-23 and NSF Grant DMS-9973339. (Email: zhang@caam.rice.edu).

of symmetric matrices subject to the constraint that X be positive semidefinite. Due to its nice theoretical properties and numerous applications, SDP has received considerable attention in recent years. Among its theoretical properties is that semidefinite programs can be solved to a prescribed accuracy in polynomial time. In fact, polynomial-time interior-point algorithms for SDP have been extensively studied, and these algorithms, especially the primal-dual path-following algorithms, have proven to be efficient and robust in practice on small- to medium-sized problems.

Even though primal-dual path-following algorithms can in theory solve semidefinite programs very efficiently, they are unsuitable for solving large-scale problems in practice because of their high demand for storage and computation. In [1], Benson *et al* have proposed another type of interior-point algorithm — a polynomial-time potential-reduction dual-scaling method — that can better take advantage of the special structure of the SDP relaxations of certain combinatorial optimization problems. Moreover, the efficiency of the algorithm has been demonstrated on several specific large-scale applications (see [2, 6]). A drawback to this method, however, is that its operation count per iteration can still be quite prohibitive due to its reliance on Newton's method as its computational engine. This drawback is especially critical when the number of constraints of the SDP is significantly greater than the size of the matrix variable.

In addition to — but in contrast with — Benson *et al*'s algorithm, several methods have recently been proposed to solve specially structured large-scale SDP problems, and common to each of these algorithms is the use of gradient-based, nonlinear programming techniques. In [8], Helmberg and Rendl have introduced a first-order bundle method to solve a special class of SDP problems in which the trace of the primal matrix X is fixed. This subclass includes the SDP relaxations of many combinatorial optimization problems, e.g., Goemans and Williamson's maxcut SDP relaxation and Lovász's theta number SDP. The primary tool for their spectral bundle method is the replacement of the positive-semidefiniteness constraint on the dual slack matrix S with the equivalent requirement that the minimum eigenvalue of S be nonnegative. The maxcut SDP relaxation has received further attention from Homer and Peinado in [9]. By using the original form of the Goemans-Williamson relaxation, i.e., by not employing the change of variables $X = VV^T$, where $V \in \Re^{n\times n}$ is the original variable, they show how the maxcut SDP can be reformulated as an unconstrained maximization problem for which a standard steepest ascent method can be used. Burer and Monteiro [3] improved upon the idea of Homer and Peinado by simply noting that, without loss of generality, V can be required to be lower triangular. More recently, Vavasis [14] has shown that the gradient of the classical log-barrier function of the dual maxcut SDP can be computed in time and space proportional to the time and space needed for computing the Cholesky factor of the dual slack matrix S. Since S is sparse whenever the underlying graph is sparse, Vavasis's observation may potentially lead to efficient, gradient-based implementations of the classical log-barrier method that can exploit sparsity.

In our recent paper [5], we showed how a class of linear and nonlinear SDPs can be reformulated into nonlinear optimization problems over very simple feasible sets of the form $\Re^n_{++} \times \Re^m$, where n is the size of matrix variable X, m is a problem-dependent, nonnegative integer, and $\Re^n_{++}$ is the positive orthant of $\Re^n$. The reformulation is based on the idea of eliminating the positive definiteness constraint on X by first applying the substitution $X = LL^T$ as was done in [3], where L is a lower triangular matrix, and then using a novel elimination scheme to reduce the number of variables and constraints. We also showed how to compute the gradient of the resulting nonlinear objective function efficiently, hence enabling the application of existing nonlinear programming techniques to many SDP problems.

In [5], we also specialized the above approach to the subclass of linear SDPs in which the diagonal of the primal matrix X is fixed. By reformulating the dual SDP and working directly in the space of the transformed problem, we devised a globally convergent, gradient-based nonlinear interior-point algorithm that simultaneously solve the original primal and dual SDPs. We remark that the class of fixed-diagonal SDPs includes most of the known SDP relaxations of combinatorial optimization problems.

More recently, Vanderbei and Benson [13] have shown how the positive semidefinite constraint on X can be replaced by the n nonlinear, concave constraints $[D(X)]_{ii} \geq 0$, for $i = 1, \ldots, n$, where $D(X)$ is the unique diagonal matrix D appearing in the standard factorization $X = LDL^T$ of a positive semidefinite matrix X. Moreover, they show how these concave constraints can be utilized in the solution of any linear SDP using an interior-point algorithm for general convex, nonlinear programs. Since the discussion of Vanderbei and Benson's method is mainly of theoretical nature in [13], the question of whether or not this method offers practical advantages on large-scale SDPs is yet to be determined.

In this paper, we extend the ideas of [5] to solve general linear SDPs. More specifically, in [5] we showed that if the diagonal of the primal matrix variable X was constrained to equal a vector d, then the dual SDP could be transformed to a nonlinear programming problem over the simple feasible set $\Re^n_{++} \times \Re^m$, where n is the size of the matrix variable and m is the number of additional primal constraints. The general case described here (that is, the case in which the diagonal of X is not necessarily constrained as above) is based on similar ideas but requires that the feasible points of the new nonlinear problem satisfy n nonlinear inequality constraints in addition to lying in the set $\Re^n_{++} \times \Re^m$. These new inequality constraints, however, can be handled effectively from an algorithmic standpoint using ideas from interior-point methods.

We propose two interior-point algorithms for solving general linear SDPs based on the above ideas. The first is a generalization of the first-order (or gradient-based) log-barrier algorithm presented in [5], whereas the second is a potential reduction algorithm that employs the use of second-derivative information via Newton's method. We believe that the first algorithm is a strong candidate for solving large, sparse SDPs in general form and that the second algorithm will also have relevance for solving small- to medium-sized SDPs, even though our current perspective is mainly theoretical.

This paper is organized as follows. In Section 2, we introduce the SDP problem studied in this paper along with our corresponding assumptions, and we briefly summarize the main results of our previous paper. We then reformulate the SDP into the nonlinear programming problem mentioned in the previous subsection and introduce and analyze a certain Lagrangian function which will play an important role in the algorithms developed in this paper. In Sections 3 and 4, respectively, we develop and prove the convergence of the two aforementioned algorithms — one being a first-order log-barrier algorithm, another a second-order potential reduction algorithm — for solving the SDP. In Section 5, we present computational results that show the performance of the first-order log-barrier algorithm on a set of SDPs that compute the so-called Lovász theta numbers of some graphs that are randomly generated and some that are from the literature. Also in Section 5, we discuss some of the advantages and disadvantages of the two algorithms presented in the paper. In the last section, we conclude the paper with a few final comments.

1.1 Notation and terminology

We use $\Re$, $\Re^n$, and $\Re^{n\times n}$ to denote the space of real numbers, real n-dimensional column vectors, and real $n\times n$ matrices, respectively, and $\Re^n_+$ and $\Re^n_{++}$ to denote those subsets of $\Re^n$ consisting of the entry-wise nonnegative and positive vectors, respectively. By $\mathcal{S}^n$ we denote the space of real $n\times n$ symmetric matrices, and we define $\mathcal{S}^n_+$ and $\mathcal{S}^n_{++}$ to be the subsets of $\mathcal{S}^n$ consisting of the positive semidefinite and positive definite matrices, respectively. We write $A \succeq 0$ and $A \succ 0$ to indicate that $A \in \mathcal{S}^n_+$ and $A \in \mathcal{S}^n_{++}$, respectively. We will also make use of the fact that each $A \in \mathcal{S}^n_+$ has a unique matrix square root $A^{1/2}$ that satisfies $A = A^{1/2}A^{1/2}$. $\mathcal{L}^n$ denotes the space of real $n\times n$ lower triangular matrices, and $\mathcal{L}^n_+$, and $\mathcal{L}^n_{++}$ are the subsets of $\mathcal{L}^n$ consisting of those matrices with nonnegative and positive diagonal entries, respectively. In addition, we define $\mathcal{L}^n_0 \subset \mathcal{L}^n$ to be the set of all $n\times n$ strictly lower triangular matrices.

We let $\mathrm{tr}(A)$ denote the trace of a matrix $A \in \Re^{n\times n}$, namely the sum of the diagonal elements of A. Moreover, for $A, B \in \Re^{n\times n}$, we define $A \bullet B \equiv \mathrm{tr}(A^T B)$. In addition, for $u, v \in \Re^n$, $u * v \in \Re^n$ denotes the Hadamard product of u and v, i.e., the entry-wise multiplication of u and v, and if $u \in \Re^n_{++}$, we define u^{-1} to be the unique vector satisfying $u * u^{-1} = e$, where e is the vector of all ones. We also define $\mathrm{Diag} : \Re^n \to \Re^{n\times n}$ by $\mathrm{Diag}(u) = U$, where U is the diagonal matrix having $U_{ii} = u_i$ for all $i = 1,\ldots,n$, and $\mathrm{diag} : \Re^{n\times n} \to \Re^n$ is defined to be the adjoint of Diag, i.e., $\mathrm{diag}(U) = u$, where $u_i = U_{ii}$ for all $i = 1,\ldots,n$. We use the notation $e_i \in \Re^n$ to denote the i-th coordinate vector that has a 1 in position i and zeros elsewhere.

We will use $\|\cdot\|$ to denote both the Euclidean norm for vectors and its induced operator norm, unless otherwise specified. The Frobenius norm of a matrix A is defined as $\|A\|_F = (A \bullet A)^{1/2}$.

2 The SDP Problem and Preliminary Results

In this section, we introduce a general-form SDP problem, state two standard assumptions on the problem, and discuss the optimality conditions and the central path for the SDP. We then describe the transformation that converts the dual SDP into a constrained nonlinear optimization problem. The consideration of optimality conditions for the new problem leads us to introduce a certain Lagrangian function, for which we then develop derivative formulas and several important results. We end the section with a detailed description of the properties of a certain "primal estimate" associated with the Lagrangian function; these properties will prove crucial for the development of the algorithms in Sections 3 and 4.

2.1 The SDP problem and corresponding assumptions

In this paper, we study the following slight variation of the standard form SDP problem:

$$\begin{array}{rl} \max & C \bullet X \\ \text{s.t.} & \operatorname{diag}(X) \geq d \\ & \mathcal{A}(X) = b, \ X \succeq 0 \end{array} \qquad (P)$$

where $X \in \mathcal{S}^n$ is the matrix variable, and the data of the problem is given by the matrix $C \in \mathcal{S}^n$, the vectors $d \in \Re^n$ and $b \in \Re^m$, and the linear function $\mathcal{A} : \mathcal{S}^n \to \Re^m$, which is defined by $[\mathcal{A}(X)]_k = A_k \bullet X$ for a given set of matrices $\{A_k\}_{k=1}^m \subset \mathcal{S}^n$. We remark that (P) differs from the usual standard form SDP $\max\{C \bullet X : \mathcal{A}(X) = b, X \succeq 0\}$ only by the additional inequality $\operatorname{diag}(X) \geq d$, but we also note that every standard form SDP can be written in the form of (P) by simply adding the redundant constraint $\operatorname{diag}(X) \geq d$ for any nonpositive vector $d \in \Re^n$. So, in fact, the form (P) is as general as the usual standard form SDP.

The need for considering the general form (P) rather than the usual standard form SDP arises from the requirement that, in order to apply the transformation alluded to in the introduction, the dual SDP must possess a special structure. The dual SDP of (P) is

$$\begin{array}{rl} \min & d^T z + b^T y \\ \text{s.t.} & \operatorname{Diag}(z) + \mathcal{A}^*(y) - C = S \\ & z \leq 0, \ S \succeq 0 \end{array} \qquad (D)$$

where $S \in \mathcal{S}^n$ is the matrix variable, $z \in \Re^n$ and $y \in \Re^m$ are the vector variables and $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ is the adjoint of the operator $\mathcal{A}$ defined by $\mathcal{A}^*(y) = \sum_{j=1}^m y_j A_j$. The term $\operatorname{Diag}(z)$ found in the equality constraint of (D) is precisely the "special structure" which our transformation will exploit. We will describe the transformation in more detail in the following subsection.

We denote by $\mathcal{F}^0(P)$ and $\mathcal{F}^0(D)$ the sets of strictly feasible solutions for problems (P)

and (D), respectively, i.e.,

$$\mathcal{F}^0(P) \equiv \{X \in \mathcal{S}^n : \mathrm{diag}(X) > d,\ \mathcal{A}(X) = b,\ X \succ 0\},$$
$$\mathcal{F}^0(D) \equiv \{(z, y, S) \in \Re^n \times \Re^m \times \mathcal{S}^n : \mathrm{Diag}(z) + \mathcal{A}^*(y) - C = S,\ z < 0,\ S \succ 0\},$$

and we make the following assumptions throughout our presentation.

Assumption 1: The set $\mathcal{F}^0(P) \times \mathcal{F}^0(D)$ is nonempty.

Assumption 2: The matrices $\{A_k\}_{k=1}^m$ are linearly independent.

Note that, when a standard form SDP is converted to the form (P) by the addition of the redundant constraint $\mathrm{diag}(X) \geq d$, for any nonpositive $d \in \Re^n$, the strict inequality $\mathrm{diag}(X) > d$ in the definition of $\mathcal{F}^0(P)$ is redundant, i.e., for each feasible $X \succ 0$, the inequality $\mathrm{diag}(X) > d$ is automatically satisfied. Hence, $\mathcal{F}^0(P)$ equals the usual set of strictly feasible solutions defined by $\{X \in \mathcal{S}^n : \mathcal{A}(X) = b, X \succ 0\}$. In particular, if we assume that the usual set of interior solutions is nonempty, then $\mathcal{F}^0(P)$ is also nonempty. In addition, it is not difficult to see that the set $\{(y, S) \in \Re^m \times \mathcal{S}^n : \mathcal{A}^*(y) - C = S, S \succ 0\}$ of dual strictly feasible solutions for the usual standard form SDP is nonempty if and only if the set $\mathcal{F}^0(D)$ is nonempty. In total, we conclude that, when a standard form SDP is put in the form (P), Assumption 1 is equivalent to the usual assumption that the original primal and dual SDPs both have strictly feasible solutions.

Under Assumption 1, it is well-known that problems (P) and (D) both have optimal solutions X^* and (z^*, y^*, S^*), respectively, such that $C \bullet X^* = d^T z^* + b^T y^*$. This last condition, called strong duality, can be alternatively expressed as the requirement that $X^* \bullet S^* = 0$ and $[\mathrm{diag}(X^*) - d]^T(-z^*) = 0$, or equivalently that $X^* S^* = 0$ and $[\mathrm{diag}(X^*) - d] * (-z^*) = 0$. In addition, under Assumptions 1 and 2, it is well-known that, for each $\nu > 0$, the problems

$$(P_\nu) \qquad \max \left\{ C \bullet X + \nu \log(\det X) + \nu \sum_{i=1}^n \log(X_{ii} - d_i) : X \in \mathcal{F}^0(P) \right\},$$
$$(D_\nu) \qquad \min \left\{ d^T z + b^T y - \nu \log(\det S) - \nu \sum_{i=1}^n \log(-z_i) : (z, y, S) \in \mathcal{F}^0(D) \right\}$$

have unique solutions X_ν and (z_ν, y_ν, S_ν), respectively, such that

$$X_\nu S_\nu = \nu I, \qquad [\mathrm{diag}(X_\nu) - d] * (-z_\nu) = \nu e, \tag{1}$$

where $I \in \Re^{n \times n}$ is the identity matrix and $e \in \Re^n$ is the vector of all ones. The set of solutions $\{(X_\nu, z_\nu, y_\nu, S_\nu) : \nu > 0\}$ is known as the primal-dual central path for problems (P) and (D). In the upcoming sections, this central path will play an important role in the development of algorithms for solving problems (P) and (D).

2.2 The transformation

In this subsection, we present the primary result of [5] which allows us to transform problem (D) into an equivalent nonlinear program with n nonlinear inequality constraints. We then introduce a certain Lagrangian function associated with the new nonlinear program and prove some key results regarding this function.

Recall that $\mathcal{L}_0^n \subset \mathcal{L}^n$ denotes the set of all $n \times n$ strictly lower triangular matrices. The following result is stated and proved in [5] (see theorem 4 of section 6 therein).

Theorem 2.1 *The following statements hold:*

(a) *for each* $(w, y) \in \Re^n_{++} \times \Re^m$, *there exists a unique* $(\tilde{L}, z) \in \mathcal{L}_0^n \times \Re^n$ *such that*

$$\operatorname{Diag}(z) + \mathcal{A}^*(y) - C = \left(\operatorname{Diag}(w) + \tilde{L}\right)\left(\operatorname{Diag}(w) + \tilde{L}\right)^T; \tag{2}$$

(b) *the functions* $\tilde{L}(w, y)$ *and* $z(w, y)$ *defined according to (2) are each infinitely differentiable and analytic on their domain* $\Re^n_{++} \times \Re^m$;

(c) *the spaces* $\Re^n_{++} \times \Re^m$ *and*

$$\{(z, y, S) \in \Re^n \times \Re^m \times \mathcal{S}^n : \operatorname{Diag}(z) + \mathcal{A}^*(y) - C = S,\ S \succ 0\} \tag{3}$$

are in bijective correspondence according to the assignment $(w, y) \mapsto (z, y, S)$, *where* $z = z(w, y)$, $S = LL^T$ *and* $L \equiv \operatorname{Diag}(w) + \tilde{L}(w, y)$.

It is important to note that the set in (3) differs from the strictly feasible set $\mathcal{F}^0(D)$ in that the inequality $z < 0$ is not enforced.

An immediate consequence of Theorem 2.1 is that the dual SDP (D) can be recast as the nonlinear program

$$\begin{array}{ll} \inf & d^T z(w, y) + b^T y \\ \text{s.t.} & z(w, y) < 0,\ w > 0 \end{array} \qquad (NLD)$$

where $w \in \Re^n$ and $y \in \Re^m$ are the vector variables. A few remarks are in order concerning (NLD) and its relationship to (D). Firstly, the functions $\tilde{L}(w, y)$ and $z(w, y)$ introduced in the above theorem cannot be uniquely extended to the boundary of $\Re^n_{++} \times \Re^m$, and so it is necessary that w be strictly positive in (NLD). Secondly, the vector constraint $z(w, y) < 0$, which arises directly from the corresponding constraint of (D), could equivalently be replaced by $z(w, y) \le 0$. We have chosen the strict inequality because, with $z(w, y) < 0$, there is a bijective correspondence between $\mathcal{F}^0(D)$ and the feasible set of (NLD). Finally, because the elements of w are not allowed to take the value zero, (NLD) does not in general have an optimal solution. In fact, only when $(d, b) = (0, 0)$ does (NLD) have an optimal solution set.

Even though the feasible set of (NLD) does not include its boundary points, we may still consider the hypothetical situation in which the inequalities $w > 0$ and $z(w,y) < 0$ are relaxed to $w \geq 0$ and $z(w,y) \leq 0$. In particular, we can investigate the first-order optimality conditions of the resulting hypothetical, constrained nonlinear program using the Lagrangian function $\ell : \Re^n_+ \times \Re^m \times \Re^n_+ \to \Re$ defined by

$$\ell(w,y,\lambda) = d^T z(w,y) + b^T y + \lambda^T z(w,y) \equiv (d+\lambda)^T z(w,y) + b^T y. \tag{4}$$

If additional issues such as regularity are ignored in this hypothetical discussion, then the first-order necessary conditions for optimality could be stated as follows: if $(w,y) \in \Re^n_+ \times \Re^m$ is a local minimum of the function $d^T z(w,y) + b^T y$ subject to the constraint that $z(w,y) \leq 0$, then there exists $\lambda \in \Re^n_+$ such that

$$\nabla_w \ell(w,y,\lambda) \geq 0, \tag{5a}$$
$$\nabla_y \ell(w,y,\lambda) = 0, \tag{5b}$$
$$w * \nabla_w \ell(w,y,\lambda) = 0, \tag{5c}$$
$$z(w,y) * \lambda = 0. \tag{5d}$$

One may suspect that these optimality conditions are of little use since they are based on the hypothetical assumption that $z(w,y)$ and $\tilde{L}(w,y)$ are defined on the boundary of $\Re^n_{++} \times \Re^m$. In the following sections, however, we show that these are precisely the conditions which guarantee optimality when satisfied "in the limit" by a sequence of points $\{(w^k, y^k, \lambda^k)\} \subset \Re^n_{++} \times \Re^m \times \Re^n_{++}$.

2.3 The first and second derivatives of the Lagrangian

In this subsection, we establish some key derivative results for the Lagrangian function ℓ introduced in the last subsection. In addition to the definition (4) of the Lagrangian, we define the functions L and S, each respectively mapping the set $\Re^n_{++} \times \Re^m$ into $\mathcal{L}^n_{++}$ and $\mathcal{S}^n_{++}$, by the formulas

$$L(w,y) = \mathrm{Diag}(w) + \tilde{L}(w,y), \tag{6}$$
$$S(w,y) = L(w,y)L(w,y)^T. \tag{7}$$

We note that $S(w,y)$ and $L(w,y)$ are the positive definite slack matrix and its Cholesky factor, respectively, which are associated with (w,y) via the bijective correspondence of Theorem 2.1.

By (4), it is evident that the derivatives of the Lagrangian are closely related to the derivatives of the function $h_v : \Re^n_{++} \times \Re^m \to \Re$ defined as

$$h_v(w,y) = v^T z(w,y), \tag{8}$$

for all $(w,y) \in \Re^n_{++} \times \Re^m$, where $v \in \Re^n$ is an arbitrary, fixed vector. Theorems 2.3 and 2.4 below establish the derivatives of h_v based on an auxiliary matrix X that is defined in the following proposition. Since Proposition 2.2 is an an immediate consequence of lemma 3 of [5], we omit its proof. (See also proposition 7 in [5].)

Proposition 2.2 *Let $L \in \mathcal{L}^n_{++}$ and $v \in \Re^n$ be given. Then the system of linear equations*

$$\mathrm{diag}(X) = v, \quad (XL)_{ij} = 0, \textit{ for all } i > j, \tag{9}$$

has a unique solution X in $\mathcal{S}^n$.

We remark that the proof of the following theorem is basically identical to the one of theorem 2 in [5]. It is included here for the sake of completeness and for paving the way towards the derivation of the second derivatives of h_v in Theorem 2.4.

Theorem 2.3 *Let $(w, y) \in \Re^n_{++} \times \Re^m$ and $v \in \Re^n$ be given. Let $X \in \mathcal{S}^n$ denote the unique solution of (9), where $L \equiv L(w, y)$. Then:*

(a) $\nabla_w h_v(w, y) = 2\,\mathrm{diag}(XL)$;

(b) $\nabla_y h_v(w, y) = -\mathcal{A}(X)$.

Proof. To prove (a), it suffices to show that $(\partial h_v/\partial w_i)(w, y) = 2(XL)_{ii}$ for all $i = 1, \ldots, n$. Differentiating (8) with respect to w_i, we obtain

$$\begin{aligned} \frac{\partial h_v}{\partial w_i}(w, y) &= v^T \left(\frac{\partial z}{\partial w_i}(w, y)\right) = \mathrm{Diag}(v) \bullet \mathrm{Diag}\left(\frac{\partial z}{\partial w_i}(w, y)\right) \\ &= X \bullet \mathrm{Diag}\left(\frac{\partial z}{\partial w_i}(w, y)\right), \end{aligned} \tag{10}$$

where the last equality follows from the fact that $\mathrm{diag}(X) = v$. Now differentiating (2) with respect to w_i, we obtain

$$\mathrm{Diag}\left(\frac{\partial z}{\partial w_i}(w, y)\right) = L\left(\mathrm{Diag}(e_i) + \frac{\partial \tilde{L}}{\partial w_i}(w, y)\right)^T + \left(\mathrm{Diag}(e_i) + \frac{\partial \tilde{L}}{\partial w_i}(w, y)\right) L^T. \tag{11}$$

Taking the inner product of both sides of this equation with X and using the fact that X is symmetric, we obtain

$$\begin{aligned} X \bullet \mathrm{Diag}\left(\frac{\partial z}{\partial w_i}(w, y)\right) &= 2\,(XL) \bullet \left(\frac{\partial \tilde{L}}{\partial w_i}(w, y) + \mathrm{Diag}(e_i)\right) \\ &= 2(XL) \bullet \mathrm{Diag}(e_i) = 2\,(XL)_{ii}, \end{aligned} \tag{12}$$

where the second equality follows from the fact that $(\partial \tilde{L}/\partial w_i)(w, y)$ is strictly lower triangular and XL is upper triangular in view of (9). Combining (10) and (12), we conclude that (a) holds.

Differentiating (8) with respect to y_k for a fixed $k \in \{1, \ldots, m\}$, we obtain

$$\frac{\partial h_v}{\partial y_k}(w, y) \;=\; \mathrm{Diag}(v) \bullet \mathrm{Diag}\left(\frac{\partial z}{\partial y_k}(w, y)\right) \;=\; X \bullet \mathrm{Diag}\left(\frac{\partial z}{\partial y_k}(w, y)\right), \tag{13}$$

where the second equality is due to the fact that $\mathrm{diag}(X) = v$. Differentiating (2) with respect to y_k, we obtain

$$\mathrm{Diag}\left(\frac{\partial z}{\partial y_k}(w, y)\right) = L\left(\frac{\partial \tilde{L}}{\partial y_k}(w, y)\right)^T + \left(\frac{\partial \tilde{L}}{\partial y_k}(w, y)\right) L^T - A_k. \tag{14}$$

Taking the inner product of both sides of this equation with X and using arguments similar to the ones above, we conclude that

$$X \bullet \mathrm{Diag}\left(\frac{\partial z}{\partial y_k}(w, y)\right) = 2(XL) \bullet \left(\frac{\partial \tilde{L}}{\partial y_k}(w, y)\right) - A_k \bullet X = -A_k \bullet X.$$

Statement (b) now follows from (13), the last identity, and the definition of $\mathcal{A}$. ■

Theorem 2.4 *Let $(w, y) \in \Re^n_{++} \times \Re^m$ and $v \in \Re^n$ be given. Let X denote the unique solution of (9) in $\mathcal{S}^n$, where $L \equiv L(w, y)$. Then, for every $i, j \in \{1, \ldots, n\}$ and $k, l \in \{1, \ldots, m\}$, there hold:*

$$\frac{\partial^2 h_v}{\partial w_i \partial w_j}(w, y) \;=\; 2\,\mathrm{tr}(W_i^T X W_j), \tag{15}$$

$$\frac{\partial^2 h_v}{\partial w_i \partial y_k}(w, y) \;=\; 2\,\mathrm{tr}(W_i^T X Y_k), \tag{16}$$

$$\frac{\partial^2 h_v}{\partial y_k \partial y_l}(w, y) \;=\; 2\,\mathrm{tr}(Y_k^T X Y_l), \tag{17}$$

where

$$W_i \equiv \mathrm{Diag}(e_i) + \frac{\partial \tilde{L}}{\partial w_i}(w, y), \qquad Y_k \equiv \frac{\partial \tilde{L}}{\partial y_k}(w, y). \tag{18}$$

Moreover, if $X \succeq 0$ then $\nabla^2 h_v(w, y, \lambda) \succeq 0$.

Proof. We will only prove (15) since the proofs of equations (16) and (17) follow by similar arguments. Note also that the proof of (15) is similar to the proofs of Theorems 2.3(a) and 2.3(b), and so the proof has been somewhat abridged. Indeed, differentiating (8) with respect to w_i and then with respect to w_j, we obtain

$$\frac{\partial^2 h_v}{\partial w_i \partial w_j}(w, y) \;=\; X \bullet \mathrm{Diag}\left(\frac{\partial^2 z}{\partial w_i \partial w_j}(w, y)\right). \tag{19}$$

Now differentiating (11) with respect to w_j, we obtain

$$\mathrm{Diag}\left(\frac{\partial^2 z}{\partial w_i \partial w_j}(w, y)\right) \;=\; L\left(\frac{\partial^2 \tilde{L}}{\partial w_i \partial w_j}(w, y)\right)^T + W_i W_j^T + W_j W_i^T + \left(\frac{\partial^2 \tilde{L}}{\partial w_i \partial w_j}(w, y)\right) L^T,$$

which immediately implies

$$X \bullet \operatorname{Diag}\left(\frac{\partial z}{\partial w_i \partial w_j}(w,y)\right) = 2\operatorname{tr}(W_i^T X W_j). \tag{20}$$

Combining (19) and (20), we conclude that (15) holds.

Now assume that $X \succeq 0$. Using (15), (16) and (17), it is straightforward to see that $q^T \nabla^2 h_v(w,y,\lambda) q = 2\,\|X^{1/2}R\|_F^2 \geq 0$ for every vector $q \in \Re^{n+m}$, where

$$R \equiv q_1 W_1 + \ldots + q_n W_n + q_{n+1} Y_1 + \ldots + q_{n+m} Y_m. \tag{21}$$

This proves the final statement of the theorem. ■

Before giving the first and second derivatives of the Lagrangian as corollaries to the Theorems 2.3 and 2.4, we establish another technical result that will be used later in Section 4.

Lemma 2.5 *Let $(w,y) \in \Re^n_{++} \times \Re^m$ and $v \in \Re^n$ be given, and let $X \in \mathcal{S}^n$ denote the unique solution of (9), where $L \equiv L(w,y)$. Suppose also that $q \in \Re^{n+m}$. Then, if $X \succ 0$ and $q^T \nabla^2 h_v(w,y) q = 0$, then $\mathcal{A}^*(q_y)$ is a diagonal matrix, where $q_y \in \Re^m$ is the vector of the last m components of q.*

Proof. Define $\nabla^2 h_v \equiv \nabla^2 h_v(w,y)$, and recall from the proof of Theorem 2.4 that the positive semidefiniteness of X implies

$$q^T \nabla^2 h_v q = 2\,\|X^{1/2}R\|_F^2 \geq 0 \tag{22}$$

for all $q \in \Re^{n+m}$, where R is given by (21). Now, using the hypotheses of the lemma, it is straightforward to see from (22) that $R = 0$, which implies $LR^T + RL^T = 0$. Using the definition of R, (11), (14), and (18), we have

$$\begin{aligned} 0 = LR^T + RL^T &= \sum_{i=1}^n q_i (LW_i^T + W_i L^T) + \sum_{k=1}^m q_{n+k}(LY_k^T + Y_k L^T) \\ &= \sum_{i=1}^n q_i D_i + \sum_{k=1}^m q_{n+k}(D_k + A_k), \end{aligned}$$

where D_i is defined as $\operatorname{Diag}((\partial z/\partial w_i)(w,y))$ and D_k is defined similarly. From the above equation, it is thus evident that $\sum_{k=1}^m q_{n+k} A_k \equiv \mathcal{A}^*(q_y)$ is a diagonal matrix. ■

We remark that the final statement of Theorem 2.4 can be strengthened using Lemma 2.5 if the linear independence of the matrices $\{e_i e_i^T\}_{i=1}^n \cup \{A_k\}_{k=1}^m$ is assumed. In particular, it can be shown that $\nabla^2 h_v(w,y) \succ 0$ whenever $X \succ 0$ and the above collection of $n+m$ data matrices is linearly independent. Such an assumption, however, is stronger than our Assumption 2, and since we intend that the results in this paper be directly applicable to SDPs that satisfy the usual assumptions, we only assume the linear independence of $\{A_k\}_{k=1}^m$.

Theorems 2.3 and 2.4 and Lemma 2.5 have immediate consequences for the derivatives of the Lagrangian ℓ, detailed in the following definition and corollary. Note that, in the result below, we define

$$\hat{\nabla}^2 \ell(w, y, \lambda) \equiv \begin{bmatrix} \nabla^2_{ww}\ell & \nabla^2_{wy}\ell \\ \nabla^2_{yw}\ell & \nabla^2_{yy}\ell \end{bmatrix} \tag{23}$$

to be the $(n+m) \times (n+m)$ leading principal block of the Hessian $\nabla^2 \ell(w, y, \lambda)$ of the Lagrangian function.

Definition 2.6 *For any $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n$, let $X(w, y, \lambda)$ denote the unique solution of (9) in $\mathcal{S}^n$ with $v \equiv \lambda + d$ and $L \equiv L(w, y)$. We refer to $X(w, y, \lambda)$ as the* primal estimate *for (P) associated with (w, y, λ).*

Corollary 2.7 *Let $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n$ be given and define $L \equiv L(w, y)$ and $X \equiv X(w, y, \lambda)$. Then:*

(a) $\nabla_w \ell(w, y, \lambda) = 2\,\mathrm{diag}(XL)$;

(b) $\nabla_y \ell(w, y, \lambda) = b - \mathcal{A}(X)$.

(c) $\nabla_\lambda \ell(w, y, \lambda) = z(w, y)$.

(d) if $X \succeq 0$, then $\hat{\nabla}^2 \ell(w, y, \lambda) \succeq 0$;

(e) if $X \succ 0$ and $q \in \Re^{n+m}$ satisfies $q^T \hat{\nabla}^2 \ell(w, y) q = 0$, then $\mathcal{A}^*(q_y)$ is a diagonal matrix, where $q_y \in \Re^m$ is the vector of the last m components of q.

We again mention that had we assumed linear independence of the matrices $\{e_i e_i^T\}_{i=1}^n \cup \{A_k\}_{k=1}^m$, we would also be able to claim that $\hat{\nabla}^2 \ell(w, y, \lambda) \succ 0$ if $X \succ 0$. However, with the weaker Assumption 2, the claim does not necessarily hold.

2.4 Properties of the primal estimate

This subsection establishes several important properties of the primal estimate $X(w, y, \lambda)$ given by Definition 2.6. The following proposition is the analogue of lemma 5 of [5].

Lemma 2.8 *Let $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n$ be given and define $L \equiv L(w, y)$, $S \equiv S(w, y)$, $X \equiv X(w, y, \lambda)$, and $\nabla_w \ell \equiv \nabla_w \ell(w, y, \lambda)$. Then:*

(a) XL is upper triangular, or equivalently, $L^T X L$ is diagonal;

(b) $X \succeq 0$ if and only if $\nabla_w \ell \geq 0$; in addition, $X \succ 0$ if and only if $\nabla_w \ell > 0$;

(c) $w * \nabla_w \ell = 2\,\mathrm{diag}(L^T X L)$, hence $w^T \nabla_w \ell = 2\,\mathrm{tr}(L^T X L) = 2\, X \bullet S$.

Proof. The upper triangularity of XL follows directly from (9). Since L^T and XL are both upper triangular, so is the product L^TXL which is also symmetric. Hence L^TXL must be diagonal. On the other hand, if L^TXL is diagonal, say it equals D, then $XL = L^{-T}D$ is upper triangular. So, (a) follows.

To prove the first part of (b), we note that the nonsingularity of L implies that $X \succeq 0$ if and only if $L^TXL \succeq 0$, but since L^TXL is diagonal by (a), $L^TXL \succeq 0$ if and only if $\mathrm{diag}(L^TXL) \geq 0$. Given that both L^T and XL are upper triangular matrices, it is easy to see that $\mathrm{diag}(L^TXL)$ is the Hadamard product of $\mathrm{diag}(L^T)$ and $\mathrm{diag}(XL)$. Since $\mathrm{diag}(L^T) > 0$, it follows that $\mathrm{diag}(L^TXL) \geq 0$ if and only if $\mathrm{diag}(XL) \geq 0$. The first statement of (b) now follows from the sequence of implications just derived and the fact that $\nabla_w \ell = 2\,\mathrm{diag}(XL)$ by Corollary 2.7(a).

The second part of (b) can be proved by an argument similar to the one given in the previous paragraph; we need only replace the inequalities by strict inequalities.

Statement (c) follows from (7), Proposition 2.7(a), and the simple observation that the diagonal of L^TXL is the Hadamard product of $\mathrm{diag}(L^T) = w$ and $\mathrm{diag}(XL)$ since both L^T and XL are upper triangular. ■

The following proposition establishes that the matrix $X(w, y, \lambda)$ plays the role of a (possibly infeasible) primal estimate for any $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n_+$, hence giving justification to its name in Definition 2.6. In particular, the proposition gives necessary and sufficient conditions for $X(w, y, \lambda)$ to be a feasible or strictly feasible solution of (P). It is interesting to note that these conditions are based entirely on the gradient of the Lagrangian function $\ell(w, y, \lambda)$.

Proposition 2.9 *Let $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n_+$, and define $L \equiv L(w, y)$, $S \equiv S(w, y)$, $X \equiv X(w, y, \lambda)$, $\nabla_w \ell \equiv \nabla_w \ell(w, y, \lambda)$, and $\nabla_y \ell \equiv \nabla_y \ell(w, y, \lambda)$. Then:*

(a) X is feasible for (P) if and only if $\nabla_w \ell \geq 0$ and $\nabla_y \ell = 0$;

(b) X is strictly feasible for (P) if and only if $\nabla_w \ell > 0$ and $\nabla_y \ell = 0$.

Proof. By the definition of X, we have $X \in \mathcal{S}^n$ and $\mathrm{diag}(X) = d + \lambda \geq d$. The theorem is an easy consequence of Corollary 2.7(b) and Lemma 2.8(b). ■

The following proposition provides a measure of the duality gap, or closeness to optimality, of points $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n_+$ and $X(w, y, \lambda)$, when both (w, y) and $X(w, y, \lambda)$ are feasible for (NLP) and (P), respectively.

Proposition 2.10 *Let $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n_+$, and define $z \equiv z(w, y)$, $S \equiv S(w, y)$, $X \equiv X(w, y, \lambda)$, $\nabla_w \ell \equiv \nabla_w \ell(w, y, \lambda)$, and $\nabla_y \ell \equiv \nabla_y \ell(w, y, \lambda)$. If $z \leq 0$, $\nabla_w \ell \geq 0$ and $\nabla_y \ell = 0$, then X is feasible for (P) and (z, y, S) is feasible for (D), and*

$$d^T z + b^T y - C \bullet X = \lambda^T(-z) + \frac{1}{2} w^T \nabla_w \ell.$$

Proof. The feasibility of X follows from Proposition 2.9, and that of (z, y, S) from the definitions of z and S, and the assumption $z \leq 0$. The above equality follows from the substitutions $C = \text{Diag}(z) + \mathcal{A}^*(y) - S$ and $\text{diag}(X) = d + \lambda$ as well as from Lemma 2.8(c). ∎

3 A Log-Barrier Algorithm

It is well known that under a homeomorphic transformation ξ, any path in the domain of ξ is mapped into a path in the range of ξ, and vice versa. Furthermore, given any continuous function f from the range of ξ to $\Re$, the extremers of f in the range of ξ are mapped into corresponding extremers of the composite function $f(\xi(\cdot))$ in the domain of ξ. In particular, if f has a unique minimizer in the range of ξ, then this minimizer is mapped into the unique minimizer of $f(\xi(\cdot))$ in the domain of ξ.

In view of these observations, it is easy to see that, under the transformation introduced in Section 2, the central path of the SDP problem (D) becomes a new "central path" in the transformed space. Furthermore, since the points on the original central path are the unique minimizers of a defining log-barrier function corresponding to different parameter values, the points on the transformed central path are therefore unique minimizers of the transformed log-barrier function corresponding to different parameter values. In general, however, it is possible that extraneous, non-extreme stationary points could be introduced to the transformed log-barrier function by the nonlinear transformations applied. In this section, we show that the transformed log-barrier functions in fact have no such non-extreme stationary points, and we use this fact to establish a globally convergent log-barrier algorithm for solving the primal and dual SDP.

In the first subsection, we describe the central path in the transformed space, and then some technical results that ensure the convergence of a sequence of primal-dual points are given in the second subsection. Finally, the precise statement of the log-barrier algorithm as well as its convergence are presented in the last subsection.

3.1 The central path in the transformed space

Given the strict inequality constraints of (NLD), a natural problem to consider is the following log-barrier problem associated with (NLD), which depends on the choice of a constant $\nu > 0$:

$$(NLD_\nu) \qquad \min \left\{ f(w,y) - 2\nu \sum_{i=1}^{n} \log w_i - \nu \sum_{i=1}^{n} \log(-z_i(w,y)) : z(w,y) < 0,\ w > 0 \right\},$$

where $f(w,y) \equiv d^T z(w,y) + b^T y$ and, for all $i = 1, \ldots, n$, $z_i(w,y)$ is the i-th coordinate function of $z(w,y)$. (The reason for the factor 2 will become apparent shortly.) Not surprisingly, (NLD_ν) is nothing but the standard dual log-barrier problem (D_ν) introduced in

Section 2.1 under the transformation given by Theorem 2.1, i.e., (D_ν) is equivalent to the nonlinear program

$$\min\left\{ f(w,y) - \nu\,\log(\det(S(w,y))) - \nu\sum_{i=1}^{n}\log(-z_i(w,y)) : z(w,y) < 0,\ w > 0 \right\},$$

which is exactly (NLD_ν) after the simplification

$$\log(\det S) = \log(\det(LL^T)) = 2\ \log(\det L) = 2\ \log(w_1 \cdots w_n) = 2\sum_{i=1}^{n}\log w_i,$$

where $S \equiv S(w,y)$ and $L \equiv L(w,y)$ and the next-to-last equality follows from the fact that the determinant of a triangular matrix is the product of its diagonal entries.

Recall from the discussion in Section 2.1 that the primal and dual log-barrier problems (D_ν) and (P_ν) each have unique solutions (z_ν, y_ν, S_ν) and X_ν, respectively, such that (1) holds. One can ask whether (NLD_ν) also has a unique solution, and if so, how this unique solution relates to (z_ν, y_ν, S_ν) and X_ν. The following theorem establishes that (NLD_ν) does in fact have a unique stationary point (w_ν, y_ν) which is simply the inverse image of the point (z_ν, y_ν, S_ν) under the bijective correspondence given in Theorem 2.1.

Theorem 3.1 *For each $\nu > 0$, the problem (NLD_ν) has a unique minimum point, which is also its unique stationary point. This minimum (w_ν, y_ν) is equal to the inverse image of the point (z_ν, y_ν, S_ν) under the bijective correspondence of Theorem 2.1. In particular, $z(w_\nu, y_\nu) = z_\nu$ and $S(w_\nu, y_\nu) = S_\nu$. Moreover, $X(w_\nu, y_\nu, -\nu\, z(w_\nu, y_\nu)^{-1}) = X_\nu$.*

Proof. Let (w, y) be a stationary point of (NLD_ν), and define $\lambda \equiv -\nu\, z(w,y)^{-1}$, $z \equiv z(w,y)$, $L \equiv L(w,y)$, $S \equiv S(w,y)$, $X \equiv X(w,y,\lambda)$, $\nabla\ell \equiv \nabla\ell(w,y,\lambda)$, $\nabla f \equiv \nabla f(w,y,\lambda)$, $\nabla_w z \equiv \nabla_z z(w,y)$ and $\nabla_y z \equiv \nabla_y z(w,y)$. Since (w,y) is a stationary point, it satisfies the first-order optimality conditions of (NLD_ν)

$$\nabla_w f - 2\,\nu\, w^{-1} - \nu\,[\nabla_w z] z^{-1} = 0, \qquad \nabla_y f - \nu\,[\nabla_y z] z^{-1} = 0.$$

(Recall that $\nabla_w z$ is an $n \times n$ matrix and that $\nabla_y z$ is an $m \times n$ matrix.) Using the definitions of f and ℓ, we easily see that $\nabla\ell = \nabla f + [\nabla z]\lambda$. Using this relation, the definition of λ, we easily see that the above optimality conditions are equivalent to

$$\nabla_w \ell * w = 2\,\nu\, e, \qquad \nabla_y \ell = 0, \tag{24}$$

where $e \in \Re^n$ is the vector of all ones. It is now clear from (24) and Proposition 2.9 that X is a strictly feasible solution of (P).

Corollary 2.7(a) implies that the first equation of (24) is equivalent to $\mathrm{diag}(XL) * w = \nu e$, and so the equality $w = \mathrm{diag}(L^T)$ implies that $\mathrm{diag}(XL) * \mathrm{diag}(L^T) = \nu e$, which in turn implies that $\mathrm{diag}(L^T X L) = \nu e$ since XL is upper triangular by the definition of X. Since $L^T X L$ is diagonal, it follows that $L^T X L = \nu I$, and hence that $XS = \nu I$. Note also that, by the definitions of λ and $X(w,y,\lambda)$, $[\text{diag}(X) - d] * (-z) = \nu\, e$. Hence, X and (z,y,S) satisfy the conditions of (1), and this clearly implies $X = X_\nu$ and $(z,y,S) = (z_\nu, y_\nu, S_\nu)$. We conclude that (NLD_ν) has a unique stationary point satisfying all the conditions stated in the theorem. That this stationary point is also a global minimum follows from the fact that (z_ν, y_ν, S_ν) is the global minimum of (D_ν). ■

3.2 Sufficient conditions for convergence

In accordance with the discussion in the last paragraph of Section 2.2, we now consider the Lagrangian function $\ell(w,y,\lambda)$ only on the open set

$$\Omega \equiv \Re^n_{++} \times \Re^m \times \Re^n_{++}.$$

Given a sequence of points $\{(w^k, y^k, \lambda^k)\}_{k\geq 0} \subset \Omega$, we define $L^k \equiv L(w^k, y^k)$, $S^k \equiv S(w^k, y^k)$, $z^k \equiv z(w^k, y^k)$, $X^k \equiv X(w^k, y^k, \lambda^k)$ and $\nabla \ell^k \equiv \nabla \ell(w^k, y^k, \lambda^k)$ for all $k \geq 0$. The following result gives sufficient conditions for the sequences $\{(z^k, y^k, S^k)\}$, $\{X^k\}$ and $\{(w^k, \lambda^k, L^k)\}$ to be bounded.

Lemma 3.2 *Let $\{(w^k, y^k, \lambda^k)\}_{k\geq 0} \subset \Omega$ be a sequence of points such that*

- $z^k < 0$ *and* $\nabla_w \ell^k \geq 0$ *for all* $k \geq 0$,
- $\lim_{k\to\infty} \nabla_y \ell^k = 0$, *and*
- *the sequences* $\{(w^k)^T \nabla_w \ell^k\}$ *and* $\{(\lambda^k)^T z^k\}$ *are both bounded.*

Then the sequences $\{(z^k, y^k, S^k)\}$, $\{X^k\}$ and $\{(w^k, \lambda^k, L^k)\}$ are bounded.

Proof. By Assumption 1, there exists a point $\bar{X}^0 \in \mathcal{F}^0(P)$. Define $\bar{\lambda}^0 \equiv \text{diag}(\bar{X}^0) - d$. By the definition of $\mathcal{F}^0(P)$, we have $\bar{X}^0 \succ 0$ and $\bar{\lambda}^0 > 0$. Since $\bar{X}^0 \succ 0$, there exists $\eta > 0$ such that $\eta^{-1} I \succ \bar{X}^0 \succ \eta I$. Define

$$\mathcal{N}^0 \equiv \{X \in \mathcal{S}^n : \eta^{-1} I \succ X \succ \eta I \text{ and } \text{diag}(X) > d\}.$$

Clearly, $\mathcal{N}^0$ is a bounded open set containing $\bar{X}^0$. Hence, by the linearity of $\mathcal{A}$, $\mathcal{A}(\mathcal{N}^0)$ is an open set containing $b = \mathcal{A}(\bar{X}^0)$. By Corollary 2.7(b) and the assumption that $\lim_{k\to\infty} \nabla_y \ell^k = 0$, we conclude that $\lim_{k\to\infty} \mathcal{A}(X^k) = b$, and hence that $\mathcal{A}(X^k) \in \mathcal{A}(\mathcal{N}^0)$ for all k sufficiently large, say $k \geq k_0$. Hence, there exists $\tilde{X}^k \in \mathcal{N}^0$ such that $\mathcal{A}(X^k - \tilde{X}^k) = 0$ for all $k \geq k_0$. We define $\tilde{\lambda}^k \equiv \text{diag}(\tilde{X}^k) - d$ for each $k \geq k_0$, and since $\tilde{X}^k \in \mathcal{N}^0$, we have $\tilde{\lambda}^k > 0$ for all such k and moreover that $\{\tilde{\lambda}^k\}_{k\geq k_0}$ is a bounded sequence.

Now let $(\bar{z}^0, \bar{y}^0, \bar{S}^0)$ be a point in $\mathcal{F}^0(D)$, that is, a feasible solution of (D) such that $\bar{S}^0 \succ 0$ and $\bar{z}^0 < 0$. For each $k \geq k_0$, we combine the information from the previous paragraph, the fact that diag and $\mathcal{A}$ are the adjoints of Diag and $\mathcal{A}^*$, respectively, and the inequalities $(\lambda^k)^T \bar{z}^0 < 0$ and $(\tilde{\lambda}^k)^T z^k < 0$ to obtain the following inequality:

$$\begin{aligned} & (X^k - \tilde{X}^k) \bullet (S^k - \bar{S}^0) \\ = \ & (X^k - \tilde{X}^k) \bullet \left[\mathrm{Diag}(z^k - \bar{z}^0) + \mathcal{A}^*(y^k - \bar{y}^0) \right] \\ = \ & \mathrm{diag}(X^k - \tilde{X}^k)^T (z^k - \bar{z}^0) + \mathcal{A}(X^k - \tilde{X}^k)^T (y^k - \bar{y}^0) \\ = \ & (\lambda^k - \tilde{\lambda}^k)^T (z^k - \bar{z}^0) = (\lambda^k)^T z^k - (\lambda^k)^T \bar{z}^0 - (\tilde{\lambda}^k)^T z^k + (\tilde{\lambda}^k)^T \bar{z}^0 \\ \geq \ & (\lambda^k)^T z^k + (\tilde{\lambda}^k)^T \bar{z}^0. \end{aligned}$$

Using this inequality and the fact that $\tilde{X}^k \in \mathcal{N}^0$ for all $k \geq k_0$, we have

$$\begin{aligned} & X^k \bullet S^k + \eta^{-1} I \bullet \bar{S}^0 - (\lambda^k)^T z^k - (\tilde{\lambda}^k)^T \bar{z}^0 \\ \geq \ & X^k \bullet S^k + \tilde{X}^k \bullet \bar{S}^0 - (\lambda^k)^T z^k - (\tilde{\lambda}^k)^T \bar{z}^0 \\ \geq \ & X^k \bullet \bar{S}^0 + \tilde{X}^k \bullet S^k \geq X^k \bullet \bar{S}^0 + \eta I \bullet S^k \geq 0 \end{aligned}$$

for all $k \geq k_0$, where the last inequality follows from the fact that $X^k \succeq 0$, which itself follows from Proposition 2.9(b) and the assumption that $\nabla_w \ell \geq 0$. By Proposition 2.8, the assumption that $\{(w^k)^T \nabla_w \ell^k\}$ is bounded implies that $\{X^k \bullet S^k\}$ is bounded which, together with the fact that $\{(\lambda^k)^T z^k\}$ and $\{\tilde{\lambda}^k\}_{k \geq k_o}$ are bounded, implies that the left-hand side of the above inequality is bounded for all $k \geq k_0$. It thus follows from the positive definiteness of $\bar{S}^0$ and ηI that both $\{X^k\}$ and $\{S^k\}$ are bounded.

The boundedness of $\{S^k\}$ clearly implies the boundedness of $\{L^k\}$ and hence the boundedness of $\{w^k\} = \{\mathrm{diag}(L^k)\}$. In addition, since $\lambda^k = \mathrm{diag}(X^k) - d$ for all $k \geq 0$, the boundedness of $\{X^k\}$ implies that $\{\lambda^k\}$ is bounded which, together with the boundedness of $\{(\lambda^k)^T z^k\}$, implies that $\{z^k\}$ is bounded. Now, using the boundedness of $\{S^k\}$ and $\{z^k\}$ along with Assumption 2, we easily see that $\{y^k\}$ is bounded. ∎

In Section 2.2, we used a hypothetical discussion of the optimality conditions of (NLD) to motivate the use of the Lagrangian function ℓ. In the following theorem, we see that the hypothetical optimality conditions (5) do in fact have relevance to the solution of (NLD). In particular, the theorem shows that if $\{(w^k, y^k, \lambda^k)\}$ is a sequence of points satisfying (5a) for each $k \geq 0$ and if (5b), (5c) and (5d) are satisfied in the limit, then any accumulation points of the corresponding sequences $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$ are optimal solutions of (P) and (D), respectively.

Theorem 3.3 *Let $\{(w^k, y^k, \lambda^k)\}_{k \geq 0} \subset \Omega$ be a sequence of points such that $z^k < 0$ and $\nabla_w \ell^k \geq 0$ for all $k \geq 0$, and such that*

$$\lim_{k \to \infty} \nabla_y \ell^k = 0 \quad \text{and} \quad \lim_{k \to \infty} w^k * \nabla_w \ell^k = \lim_{k \to \infty} z^k * \lambda^k = 0.$$

Then:

(a) the sequences $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$ are bounded, and;

(b) any accumulation points of $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$ are optimal solutions of (P) and (D), respectively.

Proof. The proof of statement (a) follows immediately from Lemma 3.2. To prove (b), let X^∞, $(z^\infty, y^\infty, S^\infty)$ and λ^∞ be accumulation points of the sequences $\{X^k\}$, $\{(z^k, y^k, S^k)\}$ and $\{\lambda^k\}$, respectively, where the boundedness of $\{\lambda^k\}$ also follows from Lemma 3.2. The assumptions and Proposition 2.9 imply that

$$\lim_{k\to\infty} \mathcal{A}(X^k) - b = \lim_{k\to\infty} \nabla_y \ell^k = 0,$$

$X^k \succeq 0$ and $\mathrm{diag}(X^k) = d + \lambda^k$ for all $k \geq 0$. This clearly implies that $X^\infty \succeq 0$, $\mathcal{A}(X^\infty) = b$ and

$$\mathrm{diag}(X^\infty) = d + \lambda^\infty \geq d,$$

that is, X^∞ is a feasible solution of (P). Since each (z^k, y^k, S^k) is a feasible solution of (D), it follows that $(z^\infty, y^\infty, S^\infty)$ is also a feasible solution of (D). Moreover, by Proposition 2.8, we have that $(w^k)^T \nabla_w \ell^k = 2\, X^k \bullet S^k$ for all $k \geq 0$, from which it follows that $X^\infty \bullet S^\infty = 0$, and also that $[\mathrm{diag}(X^k) - d] * z^k = \lambda^k * z^k$ for all $k \geq 0$, from which it follows that $[\mathrm{diag}(X^\infty) - d] * z^\infty = 0$. We have thus shown that X^∞ and $(z^\infty, y^\infty, S^\infty)$ are optimal solutions of (P) and (D). ∎

3.3 A globally convergent log-barrier algorithm

In this short subsection, we introduce a straightforward log-barrier algorithm for solving (NLD). The convergence of the algorithm is a simple consequence of Theorem 3.3.

Let constants $\gamma_1 \in [0, 1]$, $\gamma_2 \geq 1$, and $\Gamma > 0$ be given, and for each $\nu > 0$, define $\mathcal{N}(\nu) \subset \Re^n_{++} \times \Re^m$ to be the set of all points (w, y) satisfying

- $z(w, y) < 0$,
- $2\gamma_1 \nu\, e \leq w * \nabla_w \ell \leq 2\gamma_2 \nu\, e$,
- $\|\nabla_y \ell\| \leq \Gamma \mu$,

where $\nabla \ell \equiv \nabla \ell(w, y, -\nu z(w, y)^{-1})$ and $e \in \Re^n$ is the vector of all ones. Note that each $(w, y) \in \mathcal{N}(\nu)$ satisfies $\nabla_w \ell(w, y, -\nu z(w, y)^{-1}) > 0$ and that the unique minimizer (w_ν, y_ν) of (NLD_ν) is in $\mathcal{N}(\nu)$. (See the proof of Theorem 3.1 and equation (24) in particular.) We propose the following algorithm:

Log Barrier Algorithm:

Let $\sigma \in (0, 1)$ and $\mu_0 > 0$ be given, and set $k = 0$.

For $k = 0, 1, 2, \ldots$

1. Use an unconstrained minimization method to solve (NLD_{μ_k}) approximately, obtaining a point $(w^k, y^k) \in \mathcal{N}(\mu_k)$.
2. Set $\mu_{k+1} = \sigma \mu_k$, increment k by 1, and return to step 1.

End

We stress that since (NLD_{μ_k}) has a unique stationary point for all $\mu_k > 0$ which is also the unique minimum, step 1 of the algorithm will succeed using any reasonable unconstrained minimization method. Specifically, any convergent, gradient-based method will eventually produce a point in the set $\mathcal{N}(\mu_k)$.

If we define $\lambda_k \equiv -\nu z(w^k, y^k)^{-1}$ for all $k \geq 0$, then based on the definition of $\mathcal{N}(\nu)$ and Proposition 2.8(b), the algorithm clearly produces a sequence of points $\{(w^k, y^k, \lambda^k)\}_{k\geq 0}$ that satisfies the hypotheses of Theorem 3.3. Hence, the log-barrier algorithm converges in the sense of the theorem.

4 A Potential Reduction Algorithm

In this section, we describe and prove the convergence of a potential reduction interior-point algorithm for solving (NLD). The basic idea is to produce a sequence of points $\{(w^k, y^k, \lambda^k)\}_{k\geq 0}$ satisfying the hypotheses of Theorem 3.3 via the minimization of a special merit function defined inside Ω. This minimization is performed using an Armijo line search along the Newton direction of a related equality system.

Throughout this section, we assume that a point $(w^0, y^0, \lambda^0) \in \Omega$ is given that satisfies $z(w, y) < 0$ and $\nabla_w \ell(w^0, y^0, \lambda^0) > 0$.

4.1 Definitions, technical results and the algorithm

We define $f^+ \equiv f(w^0, y^0) + 1$ and

$$\Xi \equiv \left\{(w, y, \lambda) \in \Omega : \nabla_w \ell(w, y, \lambda) > 0,\ z(w, y) < 0,\ f(w, y) < f^+\right\}. \tag{25}$$

Note that $(w^0, y^0, \lambda^0) \in \Xi$. The potential reduction algorithm, which we state explicitly at the end of this subsection, will be initialized with the point (w^0, y^0, λ^0) and will subsequently produce a sequence of points $\{(w^k, y^k, \lambda^k)\}_{k\geq 0}$ such that $(w^k, y^k, \lambda^k) \in \Xi$ for all $k \geq 0$. The requirement that $\nabla_w \ell(w^k, y^k, \lambda^k)$ be nonnegative for all k is reasonable in light of our goal of producing a sequence satisfying the hypotheses of Theorem 3.3. The third requirement that $f(w^k, y^k)$ be less than f^+ for all $k \geq 0$ is technical and will be used to prove special properties of the sequence produced by the algorithm.

We also define $F : \Xi \to \Omega$ by

$$F(w, y, \lambda) \equiv \begin{bmatrix} w * \nabla_w \ell(w, y, \lambda) \\ \nabla_y \ell(w, y, \lambda) \\ -\lambda * z(w, y) \end{bmatrix} \equiv \begin{bmatrix} w \\ e \\ -\lambda \end{bmatrix} * \nabla \ell(w, y, \lambda), \tag{26}$$

where $e \in \Re^m$ is the vector of all ones. For $i = 1, \ldots, 2n + m$, we let F_i denote the i-th coordinate function of F. Note that $\{F_1, \ldots, F_n\}$ are the n scalar functions corresponding to the elements of $w * \nabla_w \ell(w, y, \lambda)$. Similar relationships hold between $\{F_{n+1}, \ldots, F_{n+m}\}$ and $\nabla_y \ell(w, y, \lambda)$, as well as between $\{F_{n+m+1}, \ldots, F_{2n+m}\}$ and $-\lambda * z(w, y)$. In addition, we define $N \equiv \{1, \ldots, n\} \cup \{n + m + 1, \ldots, 2n + m\}$.

With the definition of F, our goal of producing a sequence of points satisfying the hypotheses of Theorem 3.3 can be stated more simply as the goal of producing a sequence $\{(w^k, y^k, \lambda^k)\}_{k\geq 0} \subset \Xi$ such that

$$\lim_{k\to\infty} F(w^k, y^k, \lambda^k) = 0. \tag{27}$$

The primary tool which allows us to accomplish (27) is the merit function $\mathcal{M} : \Xi \to \Re$ given by

$$\mathcal{M}(w, y, \lambda) \equiv \zeta \log \|F\|^2 - \sum_{i\in N} \log F_i - \log(f^+ - f), \tag{28}$$

where $F \equiv F(w, y, \lambda)$, $f \equiv f(w, y)$ and ζ is an arbitrary constant satisfying $\zeta > n$. The usefulness of this merit function comes from the fact that (27) can be accomplished via the iterative minimization of $\mathcal{M}$ by taking a step along the Newton direction of the system $F(w, y, \lambda) = 0$ at the current point. In what follows, we investigate this minimization scheme.

Since we will apply Newton's method to the nonlinear system $F(w, y, \lambda) = 0$, we need to study the nonsingularity of the Jacobian $F'(w, y, \lambda)$. To simplify our notation, for all $(w, y, \lambda) \in \Omega$ we define

$$\nabla\ell \equiv \nabla\ell(w, y, \lambda), \quad \nabla^2\ell \equiv \nabla^2\ell(w, y, \lambda), \quad \hat{\nabla}^2\ell \equiv \hat{\nabla}^2\ell(w, y, \lambda), \quad \nabla z \equiv \nabla z(w, y).$$

Recall that $\hat{\nabla}^2\ell$ is the leading principal block of $\nabla^2\ell$ as defined in (23). Straightforward differentiation of $F(w, y, \lambda)$ yields

$$\begin{aligned} F'(w, y, \lambda) &= \text{Diag}\left(\begin{bmatrix} e \\ 0 \\ -e \end{bmatrix} * \nabla\ell\right) + \text{Diag}\left(\begin{bmatrix} w \\ e \\ -\lambda \end{bmatrix}\right) \nabla^2\ell \\ &= \text{Diag}\left(\begin{bmatrix} \nabla_w\ell \\ 0 \\ -z \end{bmatrix}\right) + \text{Diag}\left(\begin{bmatrix} w \\ e \\ -\lambda \end{bmatrix}\right) \begin{bmatrix} \hat{\nabla}^2\ell & \nabla z \\ \nabla z^T & 0 \end{bmatrix}. \end{aligned}$$

Now multiplying $F'(w, y, \lambda)$ by the diagonal matrix $\text{Diag}([w^{-T},\ e^T,\ \lambda^{-T}]^T)$, we obtain

$$P(w, y, \lambda) \equiv \text{Diag}\left(\begin{bmatrix} w^{-1} * \nabla_w\ell \\ 0 \\ -\lambda^{-1} * z \end{bmatrix}\right) + \begin{bmatrix} \hat{\nabla}^2\ell & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \nabla z \\ -\nabla z^T & 0 \end{bmatrix}. \tag{29}$$

For $(w, y, \lambda) \in \Omega$, the Newton equation $F'(w, z, \lambda)[\Delta w, \Delta y, \Delta\lambda] = -F(w, y, \lambda)$ is clearly equivalent to

$$P(w, y, \lambda) \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta\lambda \end{bmatrix} = - \begin{bmatrix} w^{-1} \\ e \\ \lambda^{-1} \end{bmatrix} * F(w, y, \lambda). \tag{30}$$

Note that $P(w, y, \lambda)$ is a $(2n+m) \times (2n+m)$ matrix which is in general asymmetric. We will use the matrix $P(w, y, \lambda)$ to help establish the nonsingularity of $F'(w, y, \lambda)$ in the following lemma.

Lemma 4.1 *For $(w, y, \lambda) \in \Xi$, the matrix $P(w, y, \lambda)$ is positive definite and consequently, the Jacobian $F'(w, y, \lambda)$ is nonsingular.*

Proof. Since $F'(w, y, \lambda)$ is the product of $P(w, y, \lambda)$ with a positive diagonal matrix, it suffices to prove the first part of the lemma. Combining the fact that $(w, y, \lambda) \in \Xi$ with Lemma 2.8(b) and Corollary 2.7(d), we see that $\hat{\nabla}^2 \ell \succeq 0$. (However, it is not necessarily positive definite even though $X \succ 0$; see the discussion after Corollary 2.7). Moreover, we have

$$w^{-1} * \nabla_w \ell > 0 \quad \text{and} \quad -\lambda^{-1} * z > 0. \tag{31}$$

Hence, we conclude from (29) that $P \equiv P(w, y, \lambda)$ is the sum of two positive semidefinite matrices and one skew-symmetric matrix. It follows that P is positive semidefinite.

It remains to show that P is invertible, or equivalently that $(\Delta w, \Delta y, \Delta \lambda) = (0, 0, 0)$ is the unique solution to the system

$$P(w, y, \lambda) \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in \Re^{n+m+n}, \tag{32}$$

where the sizes of the zero-vectors on the right-hand side should be clear from the context. So let $(\Delta w, \Delta y, \Delta \lambda)$ be a solution to (32). Pre-multiplying both sides of (32) by the row vector $[\Delta w^T, \Delta y^T, \Delta \lambda^T]$ and using (29), we obtain a sum of three terms corresponding to the three matrices in (29) that add to zero. By skew-symmetry of the third matrix in (29), the corresponding term is zero; and by positive semidefiniteness of the first two matrices, the first two terms are both nonnegative and thus each of them must be zero. The term corresponding to the first matrix in (29) leads to

$$\Delta w^T [\mathrm{diag}(w^{-1} * \nabla_w \ell)] \Delta w + \Delta \lambda^T [\mathrm{Diag}(-\lambda^{-1} * z)] \Delta \lambda = 0,$$

which, together with (31), implies that $(\Delta w, \Delta \lambda) = (0, 0)$. Rewriting (32) to reflect this information, we obtain the equations

$$\hat{\nabla}^2 \ell \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \Re^{n+m}, \qquad \nabla z^T \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = 0 \in \Re^n, \tag{33}$$

where again the sizes of the zero-vectors should be clear from the context.

Since $(w, y, \lambda) \in \Xi$, we see from Lemma 2.8(b) that $X(w, y, \lambda) \succ 0$. This fact together with the first equation of (33) implies that the hypotheses of Corollary 2.7(e) hold with $q = (0, \Delta y)$. It follows that $\mathcal{A}^*(\Delta y)$ is a diagonal matrix. Now, let $v \in \Re^n$ satisfy $\mathrm{Diag}(v) =$

$\mathcal{A}^*(\Delta y)$, and let $X \in \mathcal{S}^n$ denote the unique solution of (9) with respect to v and $L \equiv L(w, y)$. Using the second equation of (33), the fact that $\nabla h_v = \nabla z\, v$, and Theorem 2.3(b), we obtain

$$\begin{aligned} 0 &= v^T [\nabla_y z]^T \Delta y = \Delta y^T [\nabla_y z] v = \Delta y^T \nabla_y h_v = -\Delta y^T \mathcal{A}(X) = -\mathcal{A}^*(\Delta y) \bullet X \\ &= -\mathrm{Diag}(v) \bullet \mathrm{Diag}(X) = -\mathrm{Diag}(v) \bullet \mathrm{Diag}(v) = -\|v\|^2, \end{aligned}$$

where the fifth equality is due to the identity $\mathrm{Diag}(v) = \mathcal{A}^*(\Delta y)$ and the sixth is due to the fact that $\mathrm{diag}(X) = v$. Hence, we conclude that $\mathrm{Diag}(v) = \mathcal{A}^*(\Delta y) = 0$. By Assumption 2, this implies that $\Delta y = 0$, thus completing the proof that P is positive definite. ∎

We remark that, if we had assumed linear independence of the entire collection $\{e_i e_i\}_{i=1}^n \cup \{A_k\}_{k=1}^m$, then the proof that $P(w, y, \lambda)$ is positive definite in the above proof would have been trivial due to the fact that $X(w, y, \lambda) \succ 0$ and $\hat{\nabla}^2 \ell(w, y, \lambda) \succ 0$. (see the discussion after Corollary 2.7). In any case, even though the proof was more difficult, our weaker Assumption 2 still suffices to establish the nonsingularity of $F'(w, y, \lambda)$.

A direct consequence of Lemma 4.1 is that, for each $(w, y, \lambda) \in \Xi$, the Newton direction $(\Delta w, \Delta y, \Delta \lambda)$ for the system $F(w, y, \lambda) = 0$ exists at (w, y, λ). Stated differently, Lemma 4.1 shows that the system

$$F'(w, y, \lambda)[\Delta w, \Delta y, \Delta \lambda] = -F(w, y, \lambda) \tag{34}$$

has a unique solution for all $(w, y, \lambda) \in \Xi$. The following lemmas show that this Newton direction is a descent direction for f (when $(\Delta w, \Delta y)$ is used as the direction) and also for $\mathcal{M}$.

Lemma 4.2 *Let $(w, y, \lambda) \in \Xi$, and let $(\Delta w, \Delta y, \Delta \lambda)$ be the Newton direction at (w, y, λ) given by (34). Then $(\Delta w, \Delta y)$ is a descent direction for f at (w, y).*

Proof. Let $\hat{P}$ be the $(n+m) \times (n+m)$ leading minor of $P(w, y, \lambda)$, that is,

$$\hat{P} \equiv \mathrm{Diag}\left(\begin{bmatrix} w^{-1} \\ 0 \end{bmatrix} * \hat{\nabla} \ell \right) + \hat{\nabla}^2 \ell, \tag{35}$$

where $\hat{\nabla} \ell$ consists of the first $n+m$ components of $\nabla \ell$. Note that $\hat{P}$ is positive definite since $P(w, y, \lambda)$ is positive definite by Lemma 4.1.

Equation (34) implies that (30) holds. Using (26), (29), and (35), it is easy to see that (30) can be rewritten as the following two equations:

$$\hat{P} \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} = -\hat{\nabla} \ell - \nabla z (\Delta \lambda) \qquad \text{and} \qquad -\lambda^{-1} * z * \Delta \lambda - \nabla z^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} = z.$$

Solving for $\Delta \lambda$ in the second equation and substituting the result in the first equation, we obtain

$$\hat{P} \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} = -\hat{\nabla} \ell - \nabla z \left(-z^{-1} * \lambda * \nabla z^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} - \lambda \right).$$

Multiplying the above equation on the left by the row vector $[\Delta w^T, \Delta y^T]$, noting that $\hat{\nabla}\ell = \nabla f + \nabla z \lambda$, where $\nabla f \equiv \nabla f(w,y)$, and using the positive definiteness of $\hat{P}$, we have

$$\begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix}^T \nabla f + \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix}^T \nabla z \, \mathrm{Diag}(-z^{-1} * \lambda) \nabla z^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} < 0.$$

The fact that $-z^{-1} * \lambda > 0$ clearly implies that the matrix $\nabla z \, \mathrm{Diag}(-z^{-1} * \lambda) \nabla z^T$ is positive semidefinite. This combined with the above inequality proves that $(\Delta w, \Delta y)$ is a descent direction for f at (w, y). ■

Lemma 4.3 *Let $(w, y, \lambda) \in \Xi$, and let $(\Delta w, \Delta y, \Delta \lambda)$ be the Newton direction at (w, y, λ) given by (34). Then $(\Delta w, \Delta y, \Delta \lambda)$ is a descent direction for $\mathcal{M}$ at (w, y, λ).*

Proof. We first state a few simple results which we then combine to prove the lemma. Let $f \equiv f(w,y)$, $\nabla f \equiv \nabla f(w,y)$, $F \equiv F(w,y,\lambda)$, $\nabla F \equiv \nabla F(w,y,\lambda)$, and $\nabla M \equiv \nabla M(w,y,\lambda)$. Then equation (34) implies that

$$F^T \nabla F^T \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta \lambda \end{bmatrix} = -F^T F = -\|F\|^2 \qquad \text{and} \qquad \nabla F_i^T \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta \lambda \end{bmatrix} = -F_i \quad \forall \; i \in N. \tag{36}$$

We have from Lemma 4.2 that

$$\nabla f^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} < 0. \tag{37}$$

In addition, using (28), we have that

$$\nabla \mathcal{M} = 2\zeta \frac{\nabla F F}{\|F\|^2} - \sum_{i \in N} \frac{\nabla F_i}{F_i} + \frac{1}{f^+ - f} \begin{bmatrix} \nabla f \\ 0 \end{bmatrix}, \tag{38}$$

where we note that $\nabla_\lambda f(w,y) = 0 \in \Re^n$.

Now, using (36), (37), (38) and the inequalities $\zeta > n$ and $f(w,y) < f^+$, we have

$$\nabla \mathcal{M}^T \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta \lambda \end{bmatrix} < 2\zeta \frac{(-\|F\|^2)}{\|F\|^2} - \sum_{i \in N} \frac{(-F_i)}{F_i} = -2\zeta + 2n < 0, \tag{39}$$

which proves that $(\Delta w, \Delta y, \Delta \lambda)$ is a descent direction for $\mathcal{M}$ at (w, y, λ). ■

Given $(w, y, \lambda) \in \Xi$ and $\alpha \in \Re$, we define

$$(w(\alpha), y(\alpha), \lambda(\alpha)) \equiv (w, y, \lambda) + \alpha(\Delta w, \Delta y, \Delta \lambda),$$

where $(\Delta w, \Delta y, \Delta\lambda)$ is the Newton direction given by (34). An important step in the potential reduction algorithm is the Armijo line search: given $(w, y, \lambda) \in \Xi$ and $\sigma \in (0,1)$, the line search selects a step-size $\alpha > 0$ such that $(w(\alpha), y(\alpha), \lambda(\alpha)) \in \Xi$ and

$$\mathcal{M}_\alpha - \mathcal{M} \leq \sigma\alpha \nabla\mathcal{M}^T \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta\lambda \end{bmatrix}, \tag{40}$$

where $\mathcal{M}_\alpha \equiv \mathcal{M}(w(\alpha), y(\alpha), \lambda(\alpha))$, $\mathcal{M} \equiv \mathcal{M}(w, y, \lambda)$ and $\nabla\mathcal{M} \equiv \nabla\mathcal{M}(w, y, \lambda)$. Due to the fact that Ξ is an open set and also due to Lemma 4.3, such an α can be found in a finite number of steps.

We are now ready to state the potential reduction algorithm.

Potential Reduction (PR) Algorithm:

Let $\zeta > n$ and $\sigma, \rho \in (0,1)$ be given, and set $k = 0$.

For $k = 0, 1, 2, \ldots$

1. Solve system (34) for $(w, y, \lambda) = (w^k, y^k, \lambda^k)$ to obtain the Newton direction $(\Delta w^k, \Delta y^k, \Delta\lambda^k)$.
2. Let j_k be the smallest nonnegative integer such that $(w^k(\rho^{j_k}), y^k(\rho^{j_k}), \lambda^k(\rho^{j_k})) \in \Xi$ and such that (40) holds with $\alpha = \rho^{j_k}$.
3. Set $(w^{k+1}, y^{k+1}, \lambda^{k+1}) = (w^k(\rho^{j_k}), y^k(\rho^{j_k}), \lambda^k(\rho^{j_k}))$, increment k by 1, and return to step 1.

End

We remark that, due to Lemma 4.3, Algorithm PR monotonically decreases the merit function $\mathcal{M}$.

4.2 Convergence of Algorithm PR

In this subsection, we prove the convergence of the potential reduction algorithm given in the previous subsection. A key component of the analysis is the boundedness of the sequence produced by the algorithm, which is established in Lemmas 4.5 and 4.7.

Let $\{(w^k, y^k, \lambda^k)\}_{k \geq 0}$ be the sequence produced by the potential reduction algorithm, and define $f^k \equiv f(w^k, y^k)$, $F^k \equiv F(w^k, y^k, \lambda^k)$, $L^k \equiv L(w^k, y^k)$, $X^k \equiv X(w^k, y^k, \lambda^k)$, $S^k \equiv S(w^k, y^k)$, and $z^k \equiv z(w^k, y^k)$ for all $k \geq 0$.

Lemma 4.4 *The sequence $\{F^k\}$ is bounded. As a result, the sequences $\{X^k \bullet S^k\}$ and $\{\mathcal{A}(X^k) - b\}$ are also bounded.*

Proof. Consider the function $p : \Omega \to \Re$ defined by

$$p(r, s, t) = \zeta \, \log \|(r, s, t)\|^2 - \sum_{i=1}^n \log r_i - \sum_{i=1}^n \log t_i,$$

where ζ is the same constant appearing in (28). It is not difficult to verify (see Monteiro and Pang [12], for example) that p is coercive, i.e., for every sequence $\{(r^k, s^k, t^k)\} \subset \Omega$, if $\lim_{k\to\infty} \|(r^k, s^k, t^k)\| = \infty$, then $\lim_{k\to\infty} p(r^k, s^k, t^k) = \infty$. This property implies that the level set

$$\Lambda(\gamma) \equiv \{(r, s, t) \in \Re^n_{++} \times \Re^m \times \Re^n_{++} : p(r, s, t) \leq \gamma\}$$

is compact for all $\gamma \in \Re$.

The definition of $\mathcal{M}$ implies that, for all $(w, y, \lambda) \in \Xi$,

$$\mathcal{M}(w, y, \lambda) = p(F(w, y, \lambda)) - \log(f^+ - f(w, y)). \tag{41}$$

By the assumption that both the primal SDP (P) and the dual SDP (D) are feasible, weak duality implies that there exists a constant f^- such that the dual objective value $f(w, y) \geq f^-$ for all $(w, y) \in \Re^n_{++} \times \Re^m$. Hence, (41) implies that

$$p(F(w, y, \lambda)) \leq \mathcal{M}(w, y, \lambda) + \log(f^+ - f^-) \tag{42}$$

for all $(w, y, \lambda) \in \Xi$. Now combining (42) with the fact that the potential reduction algorithm decreases the merit function $\mathcal{M}$ in each iteration, we see that

$$p(F^k) \leq \bar{\gamma} \equiv \mathcal{M}(w^0, y^0, \lambda^0) + \log(f^+ - f^-) \tag{43}$$

for all $k \geq 0$. This in turn shows that

$$\{F^k\} \subset \Lambda(\bar{\gamma}).$$

We conclude that $\{F^k\}$ is contained in a compact set and hence is bounded.

The boundedness of $\{X^k \bullet S^k\}$ and $\{\mathcal{A}(X^k) - b\}$ now follows immediately from (26), Lemma 2.8(c) and Corollary 2.7(b). ∎

Lemma 4.5 *The sequences $\{(z^k, y^k, S^k)\}$, $\{L^k\}$, and $\{w^k\}$ are bounded.*

Proof. It suffices to show that the sequences $\{S^k\}$ and $\{z^k\}$ are bounded since, as in the proof of Lemma 3.2, the boundedness of $\{S^k\}$ and $\{z^k\}$ immediately implies the boundedness of $\{y^k\}$, $\{L^k\}$ and $\{w^k\}$. Let $\bar{X}^0 \in \mathcal{F}^0(P)$, define $\bar{\lambda}^0 \equiv \mathrm{diag}(\bar{X}^0) - d > 0$ and observe that, for each $k \geq 0$,

$$\begin{aligned} 0 &\leq \bar{X}^0 \bullet S^k = \bar{X}^0 \bullet \left[\mathrm{Diag}(z^k) + \mathcal{A}^*(y^k) - C\right] \\ &= \mathrm{diag}(\bar{X}^0)^T z^k + \mathcal{A}(\bar{X}^0)^T y^k - C \bullet \bar{X}^0 \\ &= (\mathrm{diag}(\bar{X}^0) - d)^T z^k + d^T z^k + b^T y^k - C \bullet \bar{X}^0 \\ &= (\bar{\lambda}^0)^T z^k + f^k - C \bullet \bar{X}^0 \\ &< (\bar{\lambda}^0)^T z^k + f^+ - C \bullet \bar{X}^0. \end{aligned}$$

It follows from the inequalities $\bar{\lambda}^0 > 0$ and $z^k < 0$ that $\{z^k\}$ is bounded. In addition, since $\bar{X}^0 \succ 0$ and $S^k \succeq 0$ for all $k \geq 0$, the above relation and the boundedness of $\{z^k\}$ imply that $\{S^k\}$ is bounded. ∎

Lemma 4.6 *The sequence $\{C \bullet X^k\}$ is bounded.*

Proof. By Lemma 4.4, there exists $\eta > 0$ such that $X^k \bullet S^k - (z^k)^T \lambda^k < \eta$ for all $k \geq 0$. This implies the following relation, which holds for all $k \geq 0$:

$$\begin{aligned} 0 \leq X^k \bullet S^k - (z^k)^T \lambda^k &= X^k \bullet \Big[\mathrm{Diag}(z^k) + \mathcal{A}^*(y^k) - C\Big] + (-z^k)^T \lambda^k \\ &= \mathrm{diag}(X^k)^T z^k + \mathcal{A}(X^k)^T y^k - C \bullet X^k - (z^k)^T \lambda^k \\ &= (\lambda^k + d)^T z^k + b^T y^k + (\mathcal{A}(X^k) - b)^T y^k - C \bullet X^k - (z^k)^T \lambda^k \\ &= f^k - (\mathcal{A}(X^k) - b)^T y^k - C \bullet X^k < \eta. \end{aligned}$$

Since $\{f^k\}$ is bounded and since $\{\mathcal{A}(X^k) - b\}$ and $\{y^k\}$ are bounded by Lemmas 4.4 and 4.5, we conclude from the above relation that $\{C \bullet X^k\}$ is bounded. ∎

Lemma 4.7 *The sequences $\{X^k\}$ and $\{\lambda^k\}$ are bounded.*

Proof. Let $(\bar{z}^0, \bar{y}^0, \bar{S}^0) \in \mathcal{F}^0(D)$ and note that $\bar{z}^0 < 0$. Consider the following relation, which holds for all $k \geq 0$:

$$\begin{aligned} 0 \leq X^k \bullet \bar{S}^0 &= X^k \bullet \big[\mathrm{Diag}(\bar{z}^0) + \mathcal{A}^*(\bar{y}^0) - C\big] \\ &= \mathrm{diag}(X^k)^T \bar{z}^0 + \mathcal{A}(X^k)^T \bar{y}^0 - C \bullet X^k \\ &\leq b^T \bar{y}^0 + (\mathcal{A}(X^k) - b)^T \bar{y}^0 - C \bullet X^k. \end{aligned}$$

From this relation, Lemmas 4.4 and 4.6, and the fact that $\bar{S}^0 \succ 0$, we conclude that $\{X^k\}$ is bounded. In addition, since $\mathrm{diag}(X^k) = d + \lambda^k$ for all $k \geq 0$, we conclude that $\{\lambda^k\}$ is bounded. ∎

The following theorem proves the convergence of the potential reduction algorithm. We remark that the key result is the convergence of $\{F^k\}$ to zero, which is stated in part (a) of the theorem. Part (b) is already implied by Lemmas 4.5 and 4.7, and once (a) has been established, part (c) follows immediately from Theorem 3.3.

Theorem 4.8 *Let $\{(w^k, y^k, \lambda^k)\}$ be the sequence produced by algorithm PR. Then:*

(a) $\lim_{k \to \infty} F^k = 0$;

(b) the sequences $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$ are bounded;

(c) any accumulation points of $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$ are optimal solutions of (P) and (D), respectively.

Proof. To prove (a), assume for contradiction that (a) does not hold. Then Lemma 4.4 implies that there exists a convergent subsequence $\{F^k\}_{k \in K}$ such that $F^\infty \equiv \lim_{k \in K} F^k \neq 0$. By Lemmas 4.5 and 4.7, we may also assume that the sequence $\{(w^k, y^k, \lambda^k)\}_{k \in K}$ converges to a point $(w^\infty, y^\infty, \lambda^\infty)$.

Since $F^\infty \neq 0$, since $\{F^k\}$ is bounded, and due to the weak duality between (P) and (D), there exist constants $\eta_1, \eta_2, \eta_3 \in \Re$ such that

$$\eta_1 < \lim_{k\in K}\left(\zeta \log \|F^k\|^2\right), \quad \eta_2 < \lim_{k\in K}\left(-\sum_{i\in N}\log F_i^k\right), \quad \eta_3 < \lim_{k\in K}\left(-\log(f^+ - f^k)\right). \tag{44}$$

These three inequalities together imply that

$$\lim_{k\in K} f^k \neq f^+ \qquad \text{and} \qquad F_i^\infty \neq 0 \;\; \forall \;\; i\in N$$

since otherwise, $\mathcal{M}(w^k, y^k, \lambda^k)$ would tend towards infinity, an impossibility since the algorithm has produced a sequence which monotonically decreases the merit function $\mathcal{M}$. Hence, we conclude that $F^\infty \in \Omega$, and so we clearly have that $(w^\infty, y^\infty, \lambda^\infty) \in \Xi$. It follows that the Newton direction $(\Delta w^\infty, \Delta y^\infty, \Delta\lambda^\infty)$ of system (34) exists at $(w^\infty, y^\infty, \lambda^\infty)$, and in addition, the sequence $\{(\Delta w^k, \Delta y^k, \Delta\lambda^k)\}_{k\in K}$ of Newton directions converges to $(\Delta w^\infty, \Delta y^\infty, \Delta\lambda^\infty)$. Moreover, by the inequality (39) found in the proof of Lemma 4.3, we have that

$$\nabla\mathcal{M}(w^\infty, y^\infty, \lambda^\infty)^T \begin{bmatrix}\Delta w^\infty\\ \Delta y^\infty\\ \Delta\lambda^\infty\end{bmatrix} < 2n - 2\zeta < 0. \tag{45}$$

Since $\{(w^k, y^k, \lambda^k)\}_{k\in K}$ converges in Ξ and since $\mathcal{M}$ is continuous on Ξ, it follows that $\{\mathcal{M}(w^k, y^k, \lambda^k)\}_{k\in K}$ converges. Using the relation

$$\mathcal{M}^{k+1} - \mathcal{M}^k \leq \sigma\rho^{j_k}(\nabla\mathcal{M}^k)^T \begin{bmatrix}\Delta w^k\\ \Delta y^k\\ \Delta\lambda^k\end{bmatrix} < \sigma\rho^{j_k}(2n - 2\zeta) < 0$$

for all $k \geq 0$, where $\mathcal{M}^{k+1} \equiv \mathcal{M}(w^{k+1}, y^{k+1}, \lambda^{k+1})$, $\mathcal{M}^k \equiv \mathcal{M}(w^k, y^k, \lambda^k)$ and $\nabla\mathcal{M}^k \equiv \nabla M(w^k, y^k, \lambda^k)$ and where the first and second inequalities follow from (40) and (39), respectively, we clearly see that $\lim_{k\in K}\rho^{j_k} = 0$ since the left-hand side tends to zero as $k \in K$ tends to infinity. This implies $\lim_{k\in K} j_k = \infty$.

Since j_k tends to infinity as $k \in K$ tends to infinity, we conclude that the Armijo line search requires more and more trial step-sizes as $k \in K$ increases. Recall that the line search has two simultaneous objectives: given $(w, y, \lambda) \in \Xi$, the line search finds a step-size $\alpha > 0$ such that $(w(\alpha), y(\alpha), \lambda(\alpha)) \in \Xi$ and such that relation (40) is satisfied. Since $\{(w^k, y^k, \lambda^k)\}_{k\in K}$ and $\{(\Delta w^k, \Delta y^k, \Delta\lambda^k)\}_{k\in K}$ converge to $(w^\infty, y^\infty, \lambda^\infty) \in \Xi$ and $(\Delta w^\infty, \Delta y^\infty, \Delta\lambda^\infty)$, respectively, where Ξ is an open set, it is straightforward to see that there exist $\tilde{j} \geq 0$ and $\tilde{k} \in K$ such that

$$(w^k(\rho^j), y^k(\rho^j), \lambda^k(\rho^j)) \in \Xi \tag{46}$$

for all $j \geq \tilde{j}$ and all $k \in K$ such that $k \geq \tilde{k}$. Hence, due to the fact that $\lim_{k \in K} j_k = \infty$, there exists $\hat{k} > \tilde{k}$ such that, for all $k \geq \hat{k}$, we have $j_k > \tilde{j}$, which implies (46) holds with $j = j_k - 1$ but (40) is not satisfied for the step-size $\rho^{j_k - 1}$, i.e.,

$$\frac{\mathcal{M}(w^k(\rho^{j_k-1}), y^k(\rho^{j_k-1}), \lambda^k(\rho^{j_k-1})) - \mathcal{M}^k}{\rho^{j_k-1}} > \sigma(\nabla \mathcal{M}^k)^T \begin{bmatrix} \Delta w^k \\ \Delta y^k \\ \Delta \lambda^k \end{bmatrix}.$$

Letting $k \in K$ tend to infinity in the above expression, we obtain

$$\nabla \mathcal{M}(w^\infty, y^\infty, \lambda^\infty)^T \begin{bmatrix} \Delta w^\infty \\ \Delta y^\infty \\ \Delta \lambda^\infty \end{bmatrix} \geq \sigma \nabla \mathcal{M}(w^\infty, y^\infty, \lambda^\infty)^T \begin{bmatrix} \Delta w^\infty \\ \Delta y^\infty \\ \Delta \lambda^\infty \end{bmatrix},$$

which contradicts (45) and the fact that $\sigma \in (0, 1)$. Hence, we conclude that statement (a) does in fact hold.

Statements (b) and (c) hold as discussed prior to the statement of the theorem. ∎

5 Computational Results and Discussion

In this section, we discuss some of the advantages and disadvantages of the two algorithms presented in Sections 3 and 4, and we also present some computational results for the first-order log-barrier algorithm.

5.1 First-order versus second-order

It is a well-known phenomenon in nonlinear programming that first-order (f-o) methods, i.e., those methods that use only gradient information to calculate their search directions, typically require a large number of iterations for convergence to a high accuracy, while second-order (s-o) methods, i.e., those that also use Hessian information, converge to the same accuracy in far fewer iterations. The benefit of f-o methods over s-o methods, on the other hand, is that gradient information is typically much less expensive to obtain than Hessian information, and so f-o iterations are typically much faster than s-o iterations.

For many problems, s-o approaches are favored over f-o approaches since the small number of expensive iterations produces an overall solution time that is better than the f-o method's large number of inexpensive iterations. For other problems, the reverse is true. Clearly, the relative advantages and disadvantages must be decided on a case-by-case analysis.

For semidefinite programming, the current s-o interior-point methods (either primal-dual or dual-scaling) have proven to be very robust for solving small- to medium-sized problems to high accuracy, but their performance on large-sized problems (with large n and/or m) has been mostly discouraging because the cost per iteration increases dramatically with the problem size. In fact, these methods are often inappropriate for obtaining solutions of even low accuracy. This void has been filled by f-o methods, which have proven capable of obtaining moderate accuracy in a reasonable amount of time (see the discussion in the introduction).

It is useful to consider the two algorithms presented in this paper in light of the above comments. We feel that the f-o log-barrier algorithm will have its greatest use for the solution of large SDPs. In fact, in the next section we give some computational results indicating this is the case when n is of moderate size and m is large. The s-o potential reduction method, however, will most likely not have an immediate impact except possibly on small- to medium-sized problems. In addition, there may be advantages of the potential-reduction algorithm over the conventional s-o interior-point methods. For example, the search direction computation may be less expensive in the (w, y)-space, either if one solves the Newton system directly or approximates its solution using the conjugate gradient method. (This is a current topic of investigation.) Overall, the value of the potential reduction method is two-fold: (i) it demonstrates that the convexity of the Lagrangian in the neighborhood Ξ allows one to develop s-o methods for the transformed problem; and (ii) such s-o methods may have practical advantages for solving small- to medium-sized SDPs.

5.2 Log-barrier computational results

Given a graph G with vertex set $V = \{1, \ldots, n\}$ and edge set E, the Lovász theta number ϑ of G (see [11]) can be computed as the optimal value of the following primal-dual SDP pair:

$$\max\{(ee^T) \bullet X : I \bullet X = 1,\ (e_i e_j^T + e_j e_i^T) \bullet X = 0\ \forall\ (i,j) \in E,\ X \succeq 0\}$$
$$\min\left\{ \lambda : \lambda I + \sum_{(i,j)\in E} y_{ij}(e_i e_j^T + e_j e_i^T) - ee^T = S,\ S \succeq 0 \right\},$$

where $X, S \in \mathcal{S}^n$, $\lambda \in \Re$, and $y \in \Re^{|E|}$ are the variables, I is the $n \times n$ identity matrix, $e \in \Re^n$ is the vector of all ones, and $e_k \in \Re^n$ is the k-th coordinate vector. Note that both the primal and dual problems have strictly feasible solutions.

We ran the log-barrier algorithm on nineteen graphs for which n was of small to moderate size but m was large. In particular, the size of m makes the solution of most of these Lovász theta problems difficult for second-order interior-point methods. The first nine graphs were randomly generated graphs on 100 vertices varying in edge density from 10% to 90%, while the last ten graphs are the complements of test graphs used in the Second DIMACS Challenge on the Maximum Clique Problem [10]. (Note that, for these graphs, the Lovász theta number gives an upper bound on the size of a maximum clique.)

We initialized the log-barrier algorithm with $\lambda = n + 2$, $y_{ij} = 0$ for all $(i,j) \in E$, and the specific $w > 0$ corresponding to $z = -e$. (Such a w was found by a direct Cholesky factorization.) In this way, we were able to begin the algorithm with a feasible point. The initial value of μ was set to 1, and after each log-barrier subproblem was solved, μ was decreased by a factor of 10. Moreover, the criterion for considering a subproblem to be solved was slightly altered from the theoretical condition described in Section 3.3. For the computational results, we found it more efficient to consider a subproblem solved once the norm of the gradient of the barrier function became less than 10^{-3}. The overall algorithm was terminated once μ reached the value 10^{-6}. Finally, we solved each subproblem using a limited-memory BFGS approach with a strong Wolfe-Powell line search.

Our computer code was programmed in ANSI C and run on an SGI Origin2000 with 16 300MHz R12000 processors and 10 Gigabytes of RAM at Rice University, although we stress that our code is not parallel. In Table 5.2, we give the results of the log-barrier algorithm on the nineteen test problems. In the first four columns, information regarding the problems are listed, including the problem name, the sizes of n and m, and a lower bound on the optimal value for the SDP. We remark that $n = |V|$ and $m = |E| + 1$ and that the lower bounds were computed with the primal SDP code described in [4]. In the next four columns, we give the objective value obtained by our code, the relative accuracy of the final solution with respect to the given lower bound, the time in seconds taken by the method, and the number of iterations performed by the method. From the table, we see that the method can obtain a nearly optimal solution (as evidenced by the good relative accuracies) in a small amount of time even though m can be quite large. We also see that the number of iterations is relatively large, which is not surprising since the method is a first-order algorithm.

6 Concluding Remarks

Conventional interior-point algorithms based on Newton's method are generally too costly for solving large-scale semidefinite programs. In search of alternatives, some recent papers have focused on formulations that facilitate in one way or another the application of gradient-based algorithms. The present paper is one of the efforts in this direction.

In this paper, we apply the nonlinear transformation derived in [5] to a general linear SDP problem to obtain a nonlinear program with both positivity constraints on variables and additional inequality constraints as well. Under the standard assumptions of the primal-dual strict feasibility and the linear independence of constraint matrices, we establish global convergence for a log-barrier algorithmic framework and a potential-reduction algorithm.

Our initial computational experiments indicate that the log-barrier approach based on our transformation is promising for solving at least some classes of large-scale SDP problems including, in particular, problems where the number of constraints is far greater than the size of the matrix variables. The potential reduction algorithm is also interesting from a theoretical standpoint and for the advantages it may provide for solving small- to medium-scale problems. We believe both methods are worth more investigation and improvement.

Table 1: Performance of the Log-Barrier Algorithm on Lovász Theta Graphs

name	n	m	low bd	obj val	acc	time	iter
rand1	100	497	32.1165	32.1181	4.8e−05	850	29741
rand2	100	992	22.1225	22.1234	4.2e−05	528	18877
rand3	100	1487	17.0210	17.0221	6.4e−05	629	21264
rand4	100	1982	13.1337	13.1355	1.4e−04	682	22560
rand5	100	2477	10.4669	10.4678	8.6e−05	696	22537
rand6	100	2972	8.3801	8.3814	1.5e−04	829	24539
rand7	100	3467	7.0000	7.0001	2.1e−05	137	4106
rand8	100	3962	5.0000	5.0000	9.5e−06	176	5218
rand9	100	4457	4.0000	4.0000	4.5e−06	118	3612
MANN-a9.co	45	74	17.4750	17.4752	9.7e−06	19	6150
brock200-1.co	200	5068	27.4540	27.4585	1.6e−04	3605	16083
brock200-4.co	200	6813	21.2902	21.2946	2.1e−04	4544	20092
c-fat200-1.co	200	18368	12.0000	12.0029	2.5e−04	2560	9337
c-fat200-5.co	200	11429	60.3453	60.3475	3.7e−05	1121	3924
johnson08-4-4.co	70	562	14.0000	14.0004	3.1e−05	28	2519
johnson16-2-4.co	120	1682	8.0000	8.0002	3.0e−05	161	2386
keller4.co	171	5102	14.0047	14.0136	6.4e−04	3112	20723
san200-0.7-1.co	200	5972	30.0000	30.0002	5.5e−06	273	973
sanr200-0.7.co	200	6034	23.8102	23.8374	1.1e−03	6596	29074

References

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

[2] S. Benson and Y. Ye. Approximating Maximum Stable Set and Minimum Graph Coloring Problems with the Positive Semidefinite Relaxation. Working paper, Department of Management Science, University of Iowa, Iowa, November 1999.

[3] S. Burer and R. D. C. Monteiro. A Projected Gradient Algorithm for Solving the Maxcut SDP Relaxation. Working paper, School of ISyE, Georgia Tech, USA, December 1998. (To appear in *Optimization Methods and Software*.)

[4] S. Burer and R. D. C. Monteiro. A Nonlinear Programming Algorithm for Solving Semidefinite Programs via Low-rank Factorization. Working paper, School of ISyE, Georgia Tech, USA, March 2001. (Submitted to *Mathematical Programming*.)

[5] S. Burer, R. D. C. Monteiro, and Y. Zhang. Solving a Class of Semidefinite Programs via Nonlinear Programming. Submitted to *Mathematical Programming A*. (See also Technical Report TR99-17, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA, September 1999.)

[6] C. Choi and Y. Ye. Application of Semidefinite Programming to Circuit Partitioning. In *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*, P. M. Pardalos, editor, Kluwer, 2000.

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

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

[9] M. Peinado and S. Homer. Design and performance of parallel and distributed approximation algorithms for maxcut. *Journal of Parallel and Distributed Computing*, 46:48–61, 1997.

[10] D. Johnson and M. Trick. *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge*. AMS, 1996.

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

[12] R. D. C. Monteiro and J. S. Pang. A potential reduction Newton method for constrained equations. *SIAM Journal on Optimization*, 9:729–754, 1999.

[13] R. J. Vanderbei and H. Yurttan Benson. On Formulating Semidefinite Programming Problems as Smooth Convex Nonlinear Optimization Problems. Technical Report ORFE 99-01, Dept. of Operations Research and Financial Engineering, Princeton University, Princeton NJ, November 1999.

[14] S. Vavasis. A Note on Efficient Computation of the Gradient in Semidefinite Programming. Working paper, Department of Computer Science, Cornell University, September 1999.