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

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

December, 1999

Abstract

Recently, the authors of this paper introduced a nonlinear transformation to convert the positive definiteness constraint on an $n \times n$ matrix function of a certain form into the positivity constraint on n scalar variables while keeping the number of variables unchanged [4]. Based on this transformation, they proposed interior point algorithms for solving a special class of linear semidefinite programs [5]. 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 interior-point algorithms for this type of nonlinear program and establish their global convergence.

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 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

*This is Technical Report TR99-27, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005.

†School of Mathematics, Georgia Tech., Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grant 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 Grant INT-9600343 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).

in recent years. Among its theoretical properties is that semidefinite programs can be solved 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, especially when the number of constraints of the SDP is significantly greater than the size of the matrix variable.

In addition to Benson *et al*'s algorithm, several algorithms have recently been proposed to solve the SDP problem, 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]. Using the change of variables $X = VV^T$, where $V \in \Re^{n\times n}$, they eliminate the positive-semidefiniteness constraint on X and 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 [12] 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 which can exploit sparsity.

In our recent paper [4], 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 $\Re^n_{++} \times \Re^N$, where n is the size of matrix variable X, N 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 definite constraint on the dual slack matrix S by first applying the substitution $S = LL^T$ (as was done in [3] to the primal matrix X) 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 then specialized the approach of [4] 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 pair of globally convergent nonlinear interior-point algorithms which simultaneously solve the original primal and dual SDPs. We remark that the class of SDPs studied in [5] includes the SDP relaxations of most, if not all, combinatorial optimization problems.

Most recently, Vanderbei and Benson [11] 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 [11], the question of whether or not this method offers practical advantages on large-scale SDPs is still to be determined.

In this paper, we continue the work of [4, 5], and show how to solve general linear SDPs using the ideas of our first two papers. In [5], we showed that if the diagonal of the primal matrix variable X was fixed, then the SDP could be recast as 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. In contrast, the general case described here requires that the feasible points of the new nonlinear problem not only lie in $\Re^n_{++} \times \Re^m$ but also satisfy n additional nonlinear inequality constraints. These new inequality constraints, however, can be effectively handled from an algorithmic standpoint using ideas from interior point methods. We believe that the algorithms presented in this paper, especially the log-barrier algorithm, are good candidates for solving large, sparse SDPs in general form.

Since this paper is a continuation of our previous work, there are a number of results in this paper that either have appeared in [4] or [5], or are extensions or specializations of results in those two papers. For the sake of self-containedness, however, we will include proofs for most of these results except for a few that require excessively long proofs.

1.1 Organization of the paper

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 papers. 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 two algorithms — one being a log-barrier algorithm, another a potential reduction algorithm — for solving the SDP. The key difference between the two algorithms is that, while the log-barrier algorithm can employ any reasonable minimization algorithm to produce a sequence of points which loosely follow the central path, the potential reduction algorithm utilizes search directions which are Newton directions for a certain optimality system of nonlinear equations and does not explicitly stay close to the central path. In the last section, we conclude the paper with a few final comments.

1.2 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. $\mathcal{L}^n$ 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.

We let $\operatorname{tr}(A)$ 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 $A \bullet B \equiv \operatorname{tr}(A^T B)$. 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 $\operatorname{Diag} : \Re^n \to \Re^{n\times n}$ by $\operatorname{Diag}(u) = U$, where U is the diagonal matrix having $U_{ii} = u_i$ for all $i = 1, \ldots, n$, and $\operatorname{diag} : \Re^{n\times n} \to \Re^n$ is defined to be the adjoint of Diag, i.e., $\operatorname{diag}(U) = u$, where $u_i = U_{ii}$ for all $i = 1, \ldots, n$.

We will use $\|\cdot\|$ to denote both the Euclidean norm for vectors and its induced operator norm, unless otherwise specified.

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 which 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.

2.1 The SDP problem and corresponding assumptions

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

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

where $X \in \mathcal{S}^n$ is the matrix variable, and the data of the problem are given by the matrix $C \in \mathcal{S}^n$, the vectors $d \in \Re^n$ and $b \in \Re^m$, and the linear function $\mathbf{A} : \mathcal{S}^n \to \Re^m$, which is defined by $[\mathbf{A}(X)]_j = A_j \bullet X$ for a given set of matrices $\{A_j\}_{j=1}^m \subset \mathcal{S}^n$. We remark that (P) differs from the usual standard form SDP $\max\{C \bullet X : \mathbf{A}(X) = b, X \succeq 0\}$ 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

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

where $S \in \mathcal{S}^n$ is the matrix variable, $z \in \Re^n$ and $y \in \Re^m$ are the vector variables and $\mathbf{A}^* : \Re^m \to \mathcal{S}^n$ is the adjoint of the operator $\mathbf{A}$ defined by $\mathbf{A}^*(y) = \sum_{j=1}^m y_j A_j$. 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.,

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

and we make the following assumptions throughout our presentation.

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

Assumption A.2: The matrices $\{A_j\}_{j=1}^m$ are linearly independent.

Note that, when a standard form SDP is converted to the form (P) by the addition of the redundant constraint $\operatorname{diag}(X) \geq d$, for any nonpositive $d \in \Re^n$, the strict inequality $\operatorname{diag}(X) > d$ 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 $\{X \in \mathcal{S}^n : \mathbf{A}(X) = b, X \succ 0\}$. 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 $\{(y, S) \in \Re^m \times \mathcal{S}^n : \mathbf{A}^*(y) - C = S, S \succ 0\}$ 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 A.1 is equivalent to the usual assumption that the original primal and dual SDPs both have strictly feasible solutions.

Under Assumption A.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 A.1 and A.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 [4] 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.

Let $\mathcal{L}_0^n \subset \mathcal{L}^n$ denote the set of all $n \times n$ strictly lower triangular matrices. We have the following theorem, which is a specialization of the results established in [4] (see Section 3.3 therein).

Theorem 2.1 *Under Assumption A.2, there 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

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

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

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

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

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

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 $w \ge 0$ and $z(w,y) \le 0$. 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

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

The first-order necessary conditions for optimality could then 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 $z(w,y) \le 0$, then there exists $\lambda \in \Re^n_+$ such that

$$\nabla_w \ell(w,y,\lambda) \ge 0, \tag{5a}$$

$$\nabla_y \ell(w,y,\lambda) = 0, \tag{5b}$$

$$w * \nabla_w \ell(w,y,\lambda) = 0, \tag{5c}$$

$$z(w,y) * \lambda = 0. \tag{5d}$$

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 derivatives and the primal estimate

We define the functions f, L and S, each respectively mapping the set $\Re^n_{++} \times \Re^m$ into $\Re$, $\mathcal{L}^n_{++}$ and $\mathcal{S}^n_{++}$, by the formulas

$$f(w,y) = d^T z(w,y) + b^T y, \tag{6}$$

$$L(w,y) = \mathrm{Diag}(w) + \tilde{L}(w,y), \tag{7}$$

$$S(w,y) = L(w,y)L(w,y)^T. \tag{8}$$

We note that $f(w,y)$ is simply the objective value of (NLD) at (w,y) and that $S(w,y)$ and $L(w,y)$ are the positive definite slack matrix and its Cholesky factor which are associated with (w,y) via the bijective correspondence of Theorem 2.1.

The following result appeared in [4] in a more general form.

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

$$diag(X) = v, \quad (XL)_{ij} = 0, \quad i > j, \quad X \in \mathcal{S}^n, \tag{9}$$

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

Proof. We give a constructive proof for the proposition. We note that: (i) (9) forms a square linear system of equations for the lower entries of $X \in \mathcal{S}^n$; and (ii) $L \in \mathcal{L}^n_{++}$, hence $L_{jj} > 0$ for $j = 1, \cdots, n$ and $L_{kj} = 0$ for all $k < j$. It is already known from the first equation in (9) that $X_{ii} = v_i$ for all $i = 1, \cdots, n$. We can uniquely determine $X_{ij} = X_{ji}$ for all $i > j$ from the second equation in (9) by the recursive formula

$$X_{ij} = -\frac{1}{L_{jj}} \sum_{k>j}^{n} X_{ik} L_{kj}, \quad \forall i > j, \tag{10}$$

in a "bottom-to-top and right-to-left" order (or some other workable order) such that all the $X_{ik} = X_{ki}$ terms in the right-hand side have already been determined before their appearance in the sum. This completes the proof. ∎

In the sequel, we will use the following definition.

Definition 2.3 *For any given $L \in \mathcal{L}^n_{++}$ and $v \in \Re^n$, let $X[L](v)$ be the unique solution of (9) in $\mathcal{S}^n$. Moreover, for any given $(w,y,\lambda) \in \Re^n_{++} \times \Re^m \times \Re^n$, let*

$$X(w,y,\lambda) \equiv X[L(w,y)](d+\lambda),$$

where $L(w,y)$ is defined in (7).

It has been shown in [4, 5] that the first and second derivatives of $f(w, y)$ have close ties with the matrix $X \equiv X[L(w, y)](d) \in \mathcal{S}^n$, which satisfies (9) with $L = L(w, y)$ and the right-hand side $v = d$. Comparing (4) and (6), we see that the only difference between the Lagrangian function $\ell(w, y, \lambda)$ and the objective function $f(w, y)$ is that the vector sum $d + \lambda$ appears in $\ell(w, y, \lambda)$ while the vector d alone appears in $f(w, y)$. By adding λ to d, we obtain the following result for the Lagrangian function $\ell(w, y, \lambda)$, which is an analogy of Proposition 2.2 in [5] for the objective function $f(w, y)$. Before we state the result, we define

$$\hat{\nabla}^2 \ell(w, y, \lambda) \equiv \begin{bmatrix} \nabla^2_{ww} \ell & \nabla^2_{wy} \ell \\ \nabla^2_{yw} \ell & \nabla^2_{yy} \ell \end{bmatrix} \tag{11}$$

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

Proposition 2.4 *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) $\nabla_w \ell(w, y, \lambda) = 2 \, diag(XL)$;

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

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

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

Proof. The items (a), (b) and (d) have been proved in Corollary 4.4 and Theorem 4.8 of [4]. Item (c) is obvious. ∎

We mention that had we assumed linear independence of $\{e_i e_i^T\}_{i=1}^n \cup \{A_j\}_{j=1}^m$ where e_i is the i-th column of the identity matrix, as we did in [5], 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 A.2, the claim does not necessarily hold.

Later in the paper, we will make use of the Jacobian $\nabla z(w, y)$, where the function $z(w, y)$ is defined in Theorem 2.1 and $\nabla z(w, y)$ is the $(n + m) \times n$ matrix whose i-th column is $\nabla z_i(w, y)$, $i = 1, 2, \cdots, n$. By the definition (4) of $\ell(w, y, \lambda)$, the chain rule and Proposition 2.4(a)-(b), we have

$$\begin{aligned} \nabla_w \ell(w, y, \lambda) &= [\nabla_w z(w, y)] \, (d + \lambda) = 2 \, \mathrm{diag}(XL), \\ \nabla_y \ell(w, y, \lambda) &= [\nabla_y z(w, y)] \, (d + \lambda) + b = b - \mathbf{A}(X). \end{aligned}$$

where $L \equiv L(w, y)$ and $X \equiv X[L](d+\lambda)$. It is straightforward to verify that one can replace $d + \lambda$ by any $v \in \Re^n$ and obtain the following general result.

Proposition 2.5 *For all* $(w, y) \in \Re^n_{++} \times \Re^m$ *and all* $v \in \Re^n$,

$$[\nabla_w z(w, y)]v = 2\, diag(XL) \quad and \quad [\nabla_y z(w, y)]v = -\mathbf{A}(X),$$

where $L \equiv L(w, y)$ *and* $X \equiv X[L](v)$.

We now establish several important properties of the function $X(w, y, \lambda)$ in the following proposition, which is an analogy of Lemma 2.3 in [5].

Lemma 2.6 *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,* $L^T XL$ *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) $w * \nabla_w \ell = 2\, diag(L^T XL)$, *hence* $w^T \nabla_w \ell = 2\, \mathrm{tr}(L^T XL) = 2X \bullet S$.

Proof. The upper triangularity of XL follows directly from (9). Since L^T and XL are both upper triangular, so is the product $L^T XL$ which is also symmetric. Hence $L^T XL$ must be diagonal. On the other hand, if $L^T XL$ 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 $L^T XL \succeq 0$, but since $L^T XL$ is diagonal by (a), $L^T XL \succeq 0$ if and only if $\mathrm{diag}(L^T XL) \geq 0$. Given that both L^T and XL are upper triangular matrices, it is easy to see that $\mathrm{diag}(L^T XL)$ is the Hadamard product of $\mathrm{diag}(L^T)$ and $\mathrm{diag}(XL)$. Since $\mathrm{diag}(L^T) > 0$, it follows that $\mathrm{diag}(L^T XL) \geq 0$ 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 $\nabla_w f = 2\, \mathrm{diag}(XL)$ by Proposition 2.4(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 (8), Proposition 2.4(a), and the simple observation that the diagonal of $L^T XL$ 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 $(w, y, \lambda) \in \Re^n_{++} \times \Re^m \times \Re^n_+$. It gives necessary and sufficient conditions for $X(w, y, \lambda)$ to be a feasible or strictly feasible solution for (P). These conditions are based entirely on the gradient of the Lagrangian function $\ell(w, y, \lambda)$.

Proposition 2.7 *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 definition of X, we have $\operatorname{diag}(X) = d + \lambda \geq d$ and $X \in \mathcal{S}^n$. The theorem is an easy consequence of Proposition 2.4(b) and Lemma 2.6(b). ∎

Given feasibility, the following proposition provides a measure of the duality gap, or optimality.

Proposition 2.8 *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 2.7, 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 $C = \operatorname{Diag}(z) + \mathbf{A}^*(y) - S$ and $\operatorname{diag}(X) = d + \lambda$, and Lemma 2.6(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 a convex SDP problem 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 be introduced to the transformed log-barrier function by the nonlinear transformations applied. In this section, we show that for linear SDPs the transformed log-barrier functions have no such non-extreme stationary points.

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, 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

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

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 $\lambda \equiv -\nu\, z(w, y)^{-1}$, $z \equiv z(w, y)$, $L \equiv L(w, y)$, $S \equiv S(w, y)$, $X \equiv X(w, y, \lambda)$, $\nabla \ell \equiv \nabla \ell(w, y, \lambda)$, and $\nabla z \equiv \nabla z(w, y)$. 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 λ and some basic algebra, these optimality conditions can be written more simply as the equations $\nabla_w z(d + \lambda) * w = 2\,\nu\, e$ and $\nabla_y z(d + \lambda) + b = 0$, where $e \in \Re^n$ is the vector of all ones. Now noting that $\nabla_w \ell = \nabla_w z(d + \lambda)$ and that $\nabla_y \ell = \nabla_y z(d + \lambda) + b$, the optimality conditions can be written even more simply as

$$\nabla_w \ell * w = 2\,\nu\, e, \qquad \nabla_y \ell = 0, \tag{12}$$

It is now clear from (12) and Proposition 2.7 that X is a strictly feasible solution of (P).

Corollary 2.4(a) implies that the first equation of (12) 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 definition of λ, $[\mathrm{diag}(X) - d] * (-z) = \nu\, e$. 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 A.1, there exists a point $\bar{X}^0 \in \mathcal{F}^0(P)$. Define $\bar{\lambda}^0 \equiv \mathrm{diag}(\bar{X}^0) - d$. By the definition of $\mathcal{F}^0(D)$, we have $\bar{X}^0 \succ 0$ and $\bar{\lambda}^0 > 0$, and since $\bar{X}^0 \succ 0$, there exists $\eta > 0$ such that $\eta^{-1} I \succ \bar{X}^0 \succ \eta I$. Define

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

Clearly, $\mathcal{N}^0$ is an open set containing $\bar{X}^0$. Hence, by the linearity of $\mathbf{A}$, $\mathbf{A}(\mathcal{N}^0)$ is an open set containing $b = \mathbf{A}(\bar{X}^0)$. By Corollary 2.4(b) and the assumption that $\lim_{k \to \infty} \nabla_y \ell^k = 0$, we conclude that $\lim_{k \to \infty} \mathbf{A}(X^k) = b$, and hence that $\mathbf{A}(X^k) \in \mathbf{A}(\mathcal{N}^0)$ for all k sufficiently large, say $k \geq k_0$. Hence, there exists $\tilde{X}^k \in \mathcal{N}^0$ such that $\mathbf{A}(X^k - \tilde{X}^k) = 0$ for all $k \geq k_0$. We define $\tilde{\lambda}^k \equiv \mathrm{diag}(\tilde{X}^k) - d$ 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 $\mathbf{A}$ are the adjoints of Diag and $\mathbf{A}^*$, respectively, and the inequalities $(\lambda^k)^T\bar{z}^0 < 0$ and $(\tilde{\lambda}^k)^T z^k < 0$ to obtain the following inequality:

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

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 2.7(b) and the assumption that $\nabla_w \ell \geq 0$. Proposition 2.6 shows that 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 A.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*

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

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 2.7 imply that

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

$X^k \succeq 0$ and $\mathrm{diag}(X^k) = d + \lambda^k$ for all $k \geq 0$. This clearly implies that $X^\infty \succeq 0$, $\mathbf{A}(X^\infty) = b$ 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 2.6, 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 (12) 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 2.6(b), the algorithm clearly produces a sequence of points $\{(w^k, y^k, \lambda^k)\}_{k\geq 0}$ which 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 which 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

$$\Xi \equiv \left\{(w, y, \lambda) \in \Omega : \nabla_w \ell(w, y, \lambda) > 0,\ z(w, y) < 0,\ f(w, y) < f^+\right\}. \tag{13}$$

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

$$F(w, y, \lambda) \equiv \begin{bmatrix} w * \nabla_w \ell(w, y, \lambda) \\ \nabla_y \ell(w, y, \lambda) \\ -\lambda * z(w, y) \end{bmatrix} \equiv \begin{bmatrix} w \\ e \\ -\lambda \end{bmatrix} * \nabla \ell(w, y, \lambda), \tag{14}$$

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

$$\lim_{k\to\infty} F(w^k, y^k, \lambda^k) = 0. \tag{15}$$

The primary tool which allows us to accomplish (15) is the merit function $\mathcal{M} : \Xi \to \Re$ given by

$$\mathcal{M}(w, y, \lambda) \equiv \zeta \log \|F\|^2 - \sum_{i\in N} \log F_i - \log(f^+ - f), \tag{16}$$

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 (15) 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

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

Straightforward differentiation of $F(w, y, \lambda)$ yields

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

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

$$P(w, y, \lambda) \equiv \mathrm{Diag}\left(\begin{bmatrix} w^{-1} * \nabla_w\ell \\ 0 \\ -\lambda^{-1} * z \end{bmatrix}\right) + \begin{bmatrix} \hat{\nabla}^2\ell & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \nabla z \\ -\nabla z^T & 0 \end{bmatrix}. \tag{17}$$

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

$$P(w, y, \lambda)\begin{bmatrix} \Delta w \\ \Delta y \\ \Delta\lambda \end{bmatrix} = -\begin{bmatrix} w^{-1} \\ e \\ \lambda^{-1} \end{bmatrix} * F(w, y, \lambda). \tag{18}$$

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 $F'(w, y, \lambda)$.

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 2.6(b) and Proposition 2.4(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 Proposition 2.4). Moreover, we have

$$w^{-1} * \nabla_w \ell > 0 \quad \text{and} \quad -\lambda^{-1} * z > 0. \tag{19}$$

Hence, we conclude from (17) 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

$$P(w, y, \lambda) \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in \Re^{n+m+n}, \tag{20}$$

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 (20). Pre-multiplying both sides of (20) by the row vector $[\Delta w^T, \Delta y^T, \Delta \lambda^T]$ and using (17), we obtain a sum of three terms corresponding to the three matrices in (17) that add to zero. By skew-symmetry of the third matrix in (17), 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 (17) 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 (19), implies that $(\Delta w, \Delta \lambda) = (0, 0)$. Rewriting (20) to reflect this information, we obtain the equations

$$\hat{\nabla}^2 \ell \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \Re^{n+m}, \qquad \nabla z^T \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = 0 \in \Re^n, \tag{21}$$

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

The first equation of (21) implies that

$$[0^T \Delta y^T] \hat{\nabla}^2 \ell \begin{bmatrix} 0 \\ \Delta y \end{bmatrix} = \Delta y^T \left[\nabla^2_{yy} \ell \right] \Delta y = 0,$$

where $\nabla^2_{yy} \ell \equiv \nabla^2_{yy} \ell(w, y, \lambda)$ has a similar form as $\nabla^2_{yy} f(w, y)$ given in Corollary 4.7 of [4] in a more general setting. It follows from a similar calculation as in the proof of Theorem 4.8 of [4] that

$$\Delta y^T \left[\nabla^2_{yy} \ell \right] \Delta y = \left\| X^{1/2} \Phi \left(\mathbf{A}^*(\Delta y) - \mathrm{Diag}(\mathrm{diag}(\mathbf{A}^*(\Delta y))) \right) \right\|_F^2 = 0,$$

where $X = X(w, y, \lambda) \succ 0$ in Ξ, $\|\cdot\|_F$ is the Frobenius norm for matrices and Φ is an invertible linear operator (defined in [4] under another name and in a more general setting) which maps the set of symmetric, zero-diagonal matrices to itself. The above equation clearly implies that

$$\mathbf{A}^*(\Delta y) = \mathrm{Diag}(\mathrm{diag}(\mathbf{A}^*(\Delta y))), \tag{22}$$

namely, $\mathbf{A}^*(\Delta y)$ is a diagonal matrix.

The second equation of (21) can be rewritten as $\nabla_y z^T \Delta y = 0$ or equivalently that $v^T(\nabla_y z^T \Delta y) = 0$ for all $v \in \Re^n$. In view of Proposition 2.5, (22) and Definition 2.3, we derive that for all $v \in \Re^n$,

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

where $L \equiv L(w, y)$. This implies that $\mathrm{diag}(\mathbf{A}^*(\Delta y)) = 0$. Since we already know that $\mathbf{A}^*(\Delta y)$ is diagonal, we conclude that $\mathbf{A}^*(\Delta y) = 0$, and by Assumption A.2, we conclude that $\Delta y = 0$. This completes the proof that P is positive definite. ∎

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

$$F'(w, y, \lambda)[\Delta w, \Delta y, \Delta\lambda] = -F(w, y, \lambda) \tag{23}$$

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 (23). 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,

$$\hat{P} \equiv \mathrm{Diag}\left(\begin{bmatrix} w^{-1} \\ 0 \end{bmatrix} * \hat{\nabla}\ell \right) + \hat{\nabla}^2 \ell, \tag{24}$$

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 (23) implies that (18) holds. Using (14), (17), and (24), it is easy to see that (18) 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 (23). 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 (23) implies that

$$F^T\nabla F^T\begin{bmatrix}\Delta w\\ \Delta y\\ \Delta\lambda\end{bmatrix} = -F^TF = -\|F\|^2 \qquad \text{and} \qquad \nabla F_i^T\begin{bmatrix}\Delta w\\ \Delta y\\ \Delta\lambda\end{bmatrix} = -F_i \quad \forall\ i \in N. \tag{25}$$

We have from Lemma 4.2 that

$$\nabla f^T\begin{bmatrix}\Delta w\\ \Delta y\end{bmatrix} < 0. \tag{26}$$

In addition, using (16), we have that

$$\nabla\mathcal{M} = 2\,\zeta\,\frac{\nabla F F}{\|F\|^2} - \sum_{i\in N}\frac{\nabla F_i}{F_i} + \frac{\nabla f}{f^+ - f}. \tag{27}$$

Now, using (25), (26), (27) and the inequalities $\zeta > n$ and $f(w,y) < f^+$, we have

$$\nabla\mathcal{M}^T\begin{bmatrix}\Delta w\\ \Delta y\\ \Delta\lambda\end{bmatrix} < 2\,\zeta\,\frac{(-\|F\|^2)}{\|F\|^2} - \sum_{i\in N}\frac{(-F_i)}{F_i} = -2\zeta + 2n < 0, \tag{28}$$

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 (23). 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

$$\mathcal{M}_\alpha - \mathcal{M} \leq \sigma\alpha\nabla\mathcal{M}^T \begin{bmatrix} \Delta w \\ \Delta y \\ \Delta\lambda \end{bmatrix}, \tag{29}$$

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 (23) with respect to (w^k, y^k, λ^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 (29) 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 $f^k \equiv f(w^k, y^k)$, $F^k \equiv F(w^k, y^k)$, $L^k \equiv L(w^k, y^k)$, $X^k \equiv X(w^k, y^k)$, $S^k \equiv S(w^k, y^k)$, 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 $\{\mathbf{A}(X^k) - b\}$ 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 (16). It is not difficult to verify (see Monteiro and Pang [10], 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

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

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

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

$$\mathcal{M}(w, y, \lambda) = p(F(w, y, \lambda)) - \log(f^+ - f(w, y)). \tag{30}$$

By the assumption that both of the original SDPs (P) and (D) are feasible, weak duality implies that there exists a constant f^- such that $f(w, y) \geq f^-$ for all $(w, y) \in \Re^n_{++} \times \Re^m$. Hence, (30) implies that

$$p(F(w, y, \lambda)) \leq \mathcal{M}(w, y, \lambda) + \log(f^+ - f^-) \tag{31}$$

for all $(w, y, \lambda) \in \Xi$. Now combining (31) with the fact that the potential reduction algorithm decreases the merit function $\mathcal{M}$ in each iteration, we see that

$$p(F^k) \leq \bar{\gamma} \equiv \mathcal{M}(w^0, y^0, \lambda^0) + \log(f^+ - f^-) \tag{32}$$

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 $\{\mathbf{A}(X^k) - b\}$ now follows immediately from (14), Lemma 2.6(c) and Proposition 2.4(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 $\{w^k\}$. Indeed, 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 $k \geq 0$,

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

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 $k \geq 0$:

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

Since $\{f^k\}$ is bounded and since $\{\mathbf{A}(X^k) - b\}$ 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 $k \geq 0$:

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

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

(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

$$\eta_1 < \lim_{k \in K} \left(\zeta \log \|F^k\|^2 \right), \quad \eta_2 < \lim_{k \in K} \left(-\sum_{i \in N} \log F_i^k \right), \quad \eta_3 < \lim_{k \in K} \left(-\log(f^+ - f^k) \right). \quad (33)$$

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 (23) 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 $(\Delta w^\infty, \Delta y^\infty, \Delta \lambda^\infty)$. By equation (28) found in the proof of Lemma 4.3, we have moreover that

$$\nabla \mathcal{M}(w^\infty, y^\infty, \lambda^\infty)^T \begin{bmatrix} \Delta w^\infty \\ \Delta y^\infty \\ \Delta \lambda^\infty \end{bmatrix} < 2n - 2\zeta < 0. \quad (34)$$

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 (29) and (28), 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 (29) is satisfied. Using that $\{(w^k, y^k, \lambda^k)\}_{k \in K}$ 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

$$(w^k(\rho^j), y^k(\rho^j), \lambda^k(\rho^j)) \in \Xi \quad (35)$$

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 (35) holds with $j = j_k - 1$ but (29) is not satisfied for the step-size ρ^{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 (34) 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 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 [4] 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 have indicated that the nonlinear programming 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. A comprehensive numerical study is currently under way and results will be reported in a subsequent paper.

References

[1] S. Benson, Y. Ye, and X. Zhang. Solving large-scale sparse semidefinite programs for combinatorial optimization. Research Report, Department of Management Science, University of Iowa, Iowa, 1998.

[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. An Efficient Algorithm for Solving the MAXCUT SDP Relaxation. Working paper, School of ISyE, Georgia Tech, USA, December 1998.

[4] S. Burer, R. D. C. Monteiro, and Y. Zhang. Solving Semidefinite Programs via Nonlinear Programming. Part I: Transformations and Derivatives. Technical Report TR99-17, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA, September 1999.

[5] S. Burer, R. D. C. Monteiro, and Y. Zhang. Solving Semidefinite Programs via Nonlinear Programming. Part II: Interior Point Methods for a Subclass of SDPs. Technical Report TR99-23, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, USA, October 1999.

[6] C. Choi and Y. Ye. Application of Semidefinite Programming to Circuit Partitioning. Working paper, Department of Management Science, University of Iowa, Iowa, August 1999.

[7] 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).

[8] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. Manuscript, Konrad-Zuse-Zentrum fuer Informationstechnik Berlin, Takustrasse 7, D-14195 Berlin, Germany, 1997.

[9] S. Homer and M. Peinado. Design and performance of parallel and distributed approximation algorithms for maxcut. Manuscript, Dept. of Computer Science and Center for Computational Science, Boston University, 111 Cummington Street, Boston, MA, 02215, USA, 1995.

optimization.

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

[11] 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.

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