Solving Semidefinite Programs via Nonlinear Programming Part II: Interior Point Methods for a Subclass of SDPs

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

October, 1999

Abstract

In Part I of this series of papers, we have introduced transformations which convert a large class of linear and nonlinear semidefinite programs (SDPs) into nonlinear optimization problems over "orthants" of the form $\Re^n_{++} \times \Re^N$, where n is the size of the matrices involved in the problem and N is a nonnegative integer dependent upon the specific problem. In doing so, we have effectively reduced the number of variables and constraints. In this paper, we develop interior point methods for solving a subclass of the transformable linear SDP problems where the diagonal of a matrix variable is given. These new interior point methods have the advantage of working entirely within the space of the transformed problem while still maintaining close ties with the original SDP. Under very mild and reasonable assumptions, global convergence of these methods is proved.

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

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

1 Introduction

The primary motivation for this research has been the challenge of solving large-scale semidefinite programs (SDPs). This paper is the second in a series of papers concerning the solution of large-scale SDPs via nonlinear programming methods.

---

*School of Mathematics, Georgia Institute of Technology, 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 Institute of Technology, Atlanta, Georgia 30332, USA. This author was supported in part by NSF Grant 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).

Though primal-dual interior point methods have proven to be reliable and effective in the solution of small to moderate-sized problems, their capability on large-scale problems is severely limited due to their high demand for storage and computation, in spite of efforts in exploiting sparsity (see [4], for example) in SDPs. Recently, Benson *et al* [1] proposed a potential-reduction dual-scaling interior-point method that can better take advantage of the special structure of SDP relaxations of certain combinatorial optimization problems.

In addition to interior-point methods, nonlinear programming methods which are able to exploit the sparsity of certain specially structured SDPs have been recently proposed (see [2, 6, 7]). In particular the MAXCUT SDP relaxation of Goemans and Williamson [5] has received considerable attention. The approach used in Helmberg and Rendl [6] to solve the MAXCUT SDP relaxation consists of solving a certain partial Lagrangian dual problem, whose objective function is nondifferentiable, using the usual bundle method for convex programming. On the other hand, Homer and Peinado [7] use the change of variables $X = VV^T$, $V \in \Re^{n\times n}$, where X is the primal matrix variable of the MAXCUT SDP relaxation, to transform it into a constrained nonlinear programming problem in the new variable V. More recently, Burer and Monteiro [2] proposed a variant of Homer and Peinado's method based on the constrained nonlinear programming reformulation of the MAXCUT SDP relaxation obtained by using the change of variables $X = LL^T$, where L is a lower triangular matrix. We note that the substitutions $X = VV^T$ and $X = LL^T$ can be viewed as a matrix analogy to the well-known, but rarely utilized, square-slack variable substitution for scalar inequality constraints.

In our first paper [3], we introduced transformations that convert linear and nonlinear SDPs of a certain form into equivalent nonlinear programming problems over "orthants" of the form $\Re^n_{++} \times \Re^N$, where n is the size of the matrices involved in the problem and N is a nonnegative integer dependent upon the specific problem. These transformations significantly reduce the number of variables and constraints in SDPs and, at the same time, facilitate the exploitation of sparsity in SDPs. We have also developed formulas for the first and second derivatives of the transformations, hence enabling the effective application of existing nonlinear optimization techniques to the solution of large-scale SDP problems. In developing our transformations, we used two fundamental ideas: the square-slack variable substitution $S = LL^T$ and a novel elimination scheme. Interestingly, both ideas are based on the Cholesky factorization.

In this paper, we continue the research and develop new interior-point methods for a subclass of linear SDPs which contains the SDP relaxations of binary combinatorial optimization problems, including the MAXCUT SDP relaxation. The paper is organized as follows. In Section 2, we present the subclass of transformable linear SDPs which we study in this paper and restate the key results of the first paper in a form specialized to this subclass. We also establish several important properties of the subclass under our transformations, allowing us to characterize the transformed central path associated with each problem in the subclass. In Section 3, we state sufficient conditions for a sequence of points in the transformed space to asymptotically yield an optimal solution to the original problem and propose a convergent log-barrier algorithmic framework for the new nonlinear optimization problem. In Section 4, we propose and prove the convergence of a potential reduction method whose search directions are Newton directions for a certain optimality system of nonlinear equations. In Section 5, we conclude with a few final remarks.

1.1 Basic notation

By $\Re$, $\Re^n$, and $\Re^{n\times n}$ we denote the space of real numbers, real n-dimensional column vectors, and real $n \times n$ matrices, 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 $\mathrm{tr}(A)$ 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 $u, v \in \Re^n$, $u \circ v \in \Re^n$ denotes the Hadamard product of u and v, i.e., the entry-wise multiplication of u and v. We also define $\mathrm{Diag} : \Re^n \to \Re^{n\times n}$ to be the function which sends a vector u to the diagonal matrix U having $U_{ii} = u_i$ for all $i = 1, \ldots, n$. In addition, we define $\mathrm{diag} : \Re^{n\times n} \to \Re^n$ by $\mathrm{diag}(U) = u$, where $u_i = U_{ii}$ for all $i = 1, \ldots, n$. Finally, if $\mathcal{I}$ is a finite set, we let $|\mathcal{I}|$ denote its cardinality, that is, the number of elements of $\mathcal{I}$.

2 The Subclass and Key Results

In this section, we introduce the special subclass of linear SDP problems for which interior point methods are developed in this paper, and we restate some of the main results of the first paper specialized to this subclass. We then derive some important properties of this subclass which will be used in the next section to characterize the central path in the transformed space associated with each problem in the subclass.

2.1 The subclass and corresponding assumptions

Consider the primal SDP

$$(P) \qquad \max \left\{ C \bullet X : \mathrm{diag}(X) = d;\ \mathcal{A}(X) = b;\ X \succeq 0 \right\},$$

where $\mathcal{A} : \mathcal{S}^n \to \Re^m$ is defined by $[\mathcal{A}(X)]_j = A_j \bullet X$ for all $X \in \mathcal{S}^n$ and $j = 1, \ldots, m$, and where $C, A_1, \ldots, A_m \in \mathcal{S}^n$, $d \in \Re^n_{++}$, and $b \in \Re^m$ are the data of the problem. Its dual is

$$(D) \qquad \min \left\{ d^T z + b^T y : \mathrm{Diag}(z) + \mathcal{A}^*(y) - C = S; S \succeq 0;\ (z, y) \in \Re^n \times \Re^m \right\},$$

where $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ is the adjoint of the operator $\mathcal{A}$, which is given by $\mathcal{A}^*(y) = \sum_{j=1}^m y_j A_j$ for all $y \in \Re^m$. The special subclass which we study in this paper consists of all SDPs which can be put in the form of problem (D). We remark that this subclass contains the SDP relaxations of most of, if not all, the binary combinatorial optimization problems (see section 2 of [3]).

We denote by $\mathcal{F}^0(P)$ and $\mathcal{F}^0(D)$ the sets of interior feasible solutions for problems (P) and (D), respectively, i.e.,

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

Note that $\mathcal{F}^0(D) \neq \emptyset$. We also note that the term $\mathrm{Diag}(z)$ found in the statement of problem (D) can be rewritten as $\sum_{i=1}^n z_i(e_i e_i^T)$, where $e_i \in \Re^n$ is the vector having all zeros except a one in the i-th position. We make the following assumptions throughout our presentation.

Assumption A.1: $\mathcal{F}^0(P) \neq \emptyset$, i.e., (P) has a feasible solution that is positive definite.

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

2.2 The transformation and some key results

We now introduce some definitions and notation which help us describe the transformation of (D) into an equivalent nonlinear program. Define $\mathcal{K} \equiv \{(i, j) : 1 \leq j \leq i \leq n\}$, $\mathcal{D} \equiv \{(i, i) : 1 \leq i \leq n\}$, and $\mathcal{J} \equiv \mathcal{K} \setminus \mathcal{D}$. We think of $\mathcal{K}$ as consisting of index pairs corresponding to the lower triangular entries of an $n \times n$ matrix, and $\mathcal{D}$ and $\mathcal{J}$ form a partition of $\mathcal{K}$, where $\mathcal{D}$ is the set of diagonal entries.

For any subset $\mathcal{P} \subseteq \mathcal{K}$, we define

$$\mathcal{S}^{\mathcal{P}} \equiv \{X \in \mathcal{S}^n : X_{ij} = 0 \text{ if } i \geq j \text{ and } (i, j) \notin \mathcal{P}\},$$
$$\mathcal{L}^{\mathcal{P}} \equiv \{L \in \mathcal{L}^n : L_{ij} = 0 \text{ if } (i, j) \notin \mathcal{P}\}.$$

In addition, we define $\mathcal{L}^{\mathcal{P}}_{++} \equiv \mathcal{L}^{\mathcal{P}} \cap \mathcal{L}^n_{++}$. Note that, if $\mathcal{P}$ does not contain $\mathcal{D}$, then $\mathcal{L}^{\mathcal{P}}_{++} = \emptyset$. Furthermore, we let $\pi_{\mathcal{P}} : \Re^{n \times n} \to \mathcal{S}^{\mathcal{P}}$ and $\pi^l_{\mathcal{P}} : \Re^{n \times n} \to \mathcal{L}^{\mathcal{P}}$ denote the orthogonal projection operators from $\Re^{n \times n}$ onto the subspaces $\mathcal{S}^{\mathcal{P}}$ and $\mathcal{L}^{\mathcal{P}}$, respectively.

The following theorem is a consequence of the results established in [3].

Theorem 2.1 *Under Assumption A.2, there hold:*

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

$$Diag(z) + \mathcal{A}^*(y) - C = \left(Diag(w) + L^{\mathcal{J}}\right)\left(Diag(w) + L^{\mathcal{J}}\right)^T; \tag{1}$$

(b) *the functions* $L^{\mathcal{J}}(w,y)$ *and* $z(w,y)$ *defined according to (1) are each infinitely differentiable on their domain* $\Re^n_{++} \times \Re^m$;

(c) *the spaces* $\Re^n_{++} \times \Re^m$ *and* $\mathcal{F}^0(D)$ *are in bijective correspondence according to the assignment* $(w,y) \mapsto \left(z(w,y), y, LL^T\right)$, *where* $L \equiv Diag(w) + L^{\mathcal{J}}(w,y)$.

As a consequence of Theorem 2.1, the problem obtained from (D) by restricting the feasible region to the set $\mathcal{F}^0(D)$ can be recast as the nonlinear program

$$(LD) \qquad \inf\left\{f(w,y) : (w,y) \in \Re^n_{++} \times \Re^m\right\}.$$

where $f : \Re^n_{++} \times \Re^m \to \Re$ is given by

$$f(w,y) = d^T z(w,y) + b^T y, \quad \forall (w,y) \in \Re^n_{++} \times \Re^m.$$

We note that problems (D) and (LD) have the same optimal values as was shown in [3]. However, (LD) has an open feasible set and in general does not have an optimal solution. In fact, it can be easily seen that if $(b,d) \neq 0$ then all optimal solutions of (D) lie in the boundary of $\mathcal{F}^0(D)$, and in this case (LD) does not have an optimal solution.

We next derive some properties of the gradient and Hessian of the function f which will be used throughout our presentation. We first need to introduce some notation and definitions. We define the linear operator sim : $\Re^{n\times n} \to \mathcal{S}^n$ by $\mathrm{sim}(P) = P + P^T$. In addition, given subsets $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{K}$ and $L \in \mathcal{L}^n_{++}$, we define the linear operators $\Omega[L]_{\mathcal{QP}} : \mathcal{L}^{\mathcal{P}} \to \mathcal{S}^{\mathcal{Q}}$ and $\Omega[L]^a_{\mathcal{PQ}} : \mathcal{S}^{\mathcal{Q}} \to \mathcal{L}^{\mathcal{P}}$ by

$$\Omega[L]_{\mathcal{QP}}(L^{\mathcal{P}}) \equiv \pi_{\mathcal{Q}}\left(\mathrm{sim}(L^{\mathcal{P}} L^T)\right) \tag{2}$$
$$\Omega[L]^a_{\mathcal{PQ}}(M^{\mathcal{Q}}) \equiv 2\,\pi^l_{\mathcal{P}}(M^{\mathcal{Q}} L). \tag{3}$$

When the matrix L is understood from the context, we will denote these two operators simply as $\Omega_{\mathcal{QP}}$ and $\Omega^a_{\mathcal{PQ}}$. It is not difficult to verify that, for a fixed L, $\Omega^a_{\mathcal{PQ}}$ is the adjoint operator of $\Omega_{\mathcal{QP}}$. We also remark that $\Omega_{\mathcal{JJ}}$ and $\Omega^a_{\mathcal{JJ}}$ are invertible operators (see Proposition 4.1 of [3]).

Associated with a point $(w,y) \in \Re^n_{++} \times \Re^m$, define

$$L(w,y) \equiv \mathrm{Diag}(w) + L^{\mathcal{J}}(w,y) \in \mathcal{L}^n, \tag{4}$$
$$X(w,y) \equiv \left[I - (\Omega^a_{\mathcal{JJ}})^{-1}\Omega^a_{\mathcal{JD}}\right](\mathrm{Diag}(d)) \in \mathcal{S}^n, \tag{5}$$
$$S(w,y) \equiv L(w,y)L(w,y)^T \in \mathcal{S}^n_{++}, \tag{6}$$

where I denotes the identity operator and where the Ω operators are implicitly dependent upon $L(w, y)$. The following two theorems are restatements of two important results from our first paper.

Theorem 2.2 *Let $(w, y) \in \Re^n_{++} \times \Re^m$, and define $L \equiv L(w, y)$ and $X \equiv X(w, y)$. Then:*

(a) $\nabla_w f(w, y) = 2\, diag(XL)$;

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

Proof. Let $\hat{f} : \mathcal{L}^{\mathcal{D}}_{++} \times \Re^m$ be defined by $\hat{f}(L^{\mathcal{D}}, y) = f(\mathrm{diag}(L^{\mathcal{D}}), y)$. Using the fact that the objective function of (D) is linear, we see from theorem 4.5 of [3] that

$$\nabla_{L^{\mathcal{D}}} \hat{f}(\mathrm{Diag}(w), y) = 2\pi^l_{\mathcal{D}}(XL) \qquad \text{and} \qquad \nabla_y \hat{f}(\mathrm{Diag}(w), y) = b - \mathcal{A}(X).$$

The theorem now follows from the straightforward isomorphism between $\mathcal{L}^{\mathcal{D}}_{++}$ and $\Re^n_{++}$. ■

Theorem 2.3 *Let $(w, y) \in \Re^n_{++} \times \Re^m$, and define $X \equiv X(w, y)$. Then $\nabla^2 f(w, y)$ is positive semidefinite if $X \succeq 0$. Moreover, $\nabla^2 f(w, y)$ is positive definite if $X \succ 0$.*

Proof. Let $(L^{\mathcal{D}}, y) \in \mathcal{L}^{\mathcal{D}}_{++} \times \Re^m$, and define $\hat{f}(L^{\mathcal{D}}, y)$ as in the proof of Theorem 2.2. Theorem 4.10 of [3] implies that $\nabla^2 \hat{f}(L^{\mathcal{D}}, y)$ is positive semidefinite if $X \succeq 0$ and positive definite if $X \succ 0$. (Note that the matrix G used in the theorem of the previous paper is precisely X.) Using this result, the fact that $\Re^n_{++} \times \Re^m$ and $\mathcal{L}^{\mathcal{D}}_{++} \times \Re^m$ are isomorphic, and the observation that Assumption A.2 is equivalent to the injectivity of the operator $\pi_{\mathcal{J}} \circ \mathcal{A}^*$, the statement of the theorem easily follows. ■

Theorems 2.2 and 2.3 demonstrate that, for each $(w, y) \in \Re^n_{++} \times \Re^m$, the gradient $\nabla f(w, y)$ and the Hessian $\nabla^2 f(w, y)$ are tightly linked to the matrix $X(w, y)$. In the following two propositions, we establish several important properties of the function $X(w, y)$.

Lemma 2.4 *Let $(w, y) \in \Re^n_{++} \times \Re^m$, and define $L \equiv L(w, y)$, $X \equiv X(w, y)$, and $S \equiv S(w, y)$. Then:*

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

(b) $w \circ \nabla_w f(w, y) = 2\, diag(L^T XL)$, and hence $w^T \nabla_w f(w, y) = 2\,\mathrm{tr}(L^T XL) = 2\, X \bullet S$.

Proof. Define $D \equiv \mathrm{diag}(d)$. To show that XL is upper triangular, it suffices to show that $\pi^l_{\mathcal{J}}(XL) = 0$. Indeed, using (3) and (5), we obtain

$$\begin{aligned} 2\,\pi^l_{\mathcal{J}}(XL) &= \Omega^a_{\mathcal{J}\mathcal{K}}(X) = \Omega^a_{\mathcal{J}\mathcal{K}}\left(\left[I - (\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{D}}\right](D)\right) \\ &= \Omega^a_{\mathcal{J}\mathcal{K}}(D) - \left[\Omega^a_{\mathcal{J}\mathcal{K}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{D}}\right](D) \\ &= \Omega^a_{\mathcal{J}\mathcal{D}}(D) - \left[\Omega^a_{\mathcal{J}\mathcal{J}}(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{D}}\right](D) = 0, \end{aligned}$$

where the next-to-last inequality follows from the fact that $[(\Omega^a_{\mathcal{J}\mathcal{J}})^{-1}\Omega^a_{\mathcal{J}\mathcal{D}}](D) \in S^{\mathcal{J}}$ and $D \in S^{\mathcal{D}}$. Noting that L^T is an invertible upper triangular matrix and that $L^T XL$ is a symmetric matrix, it is easy to see that XL is upper triangular if and only if $L^T XL$ is diagonal, and so statement (a) of the proposition follows.

Statement (b) follows from (6), Theorem 2.2(a), and the simple observation that the diagonal of $L^T XL$ is the Hadamard product of $\text{diag}(L^T) = w$ and $\text{diag}(XL)$ since both L^T and XL are upper triangular. ∎

Lemma 2.5 *Let $(w, y) \in \Re^n_{++} \times \Re^m$, and define $L \equiv L(w, y)$, $X \equiv X(w, y)$, $\nabla f_w \equiv \nabla f_w(w, y)$, and $\nabla f_y \equiv \nabla f_y(w, y)$. Then:*

(a) $X \in \mathcal{S}^n$ and diag$(X) = d$;

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

(c) $\mathcal{A}(X) = b$ if and only if $\nabla_y f = 0$.

Proof. Statement (a) follows directly from the definition of X, and statement (c) is an easy consequence of Theorem 2.2(b).

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 Lemma 2.4, $L^T XL \succeq 0$ if and only if $\text{diag}(L^T XL) \geq 0$. Given that both L^T and XL are upper triangular matrices, it is easy to see that $\text{diag}(L^T XL)$ is the Hadamard product of $\text{diag}(L^T)$ and $\text{diag}(XL)$. Since $\text{diag}(L^T) > 0$, it follows that $\text{diag}(L^T XL) \geq 0$ if and only if $\text{diag}(XL) \geq 0$. The first statement of (b) now follows from the sequence of implications just derived and the fact that $\nabla_w f = 2\,\text{diag}(XL)$ by Theorem 2.2(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. ∎

Notice that, for each $(w, y) \in \Re^n_{++} \times \Re^m$, Lemma 2.5 gives necessary and sufficient conditions for $X(w, y)$ to be a feasible solution for (P). One can think of $X(w, y)$ as a, possibly infeasible, primal estimate associated with (w, y).

3 The Central Path and 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 minimizers (maximizers) of f in the range of ξ are mapped into corresponding minimizers (maximizers) of $f \circ \xi$ 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 $f \circ \xi$ in the domain of ξ.

In view of these observations, it is easy to see that under the transformation introduced in our first paper [3], the central path of a convex SDP problem is transformed into a new "central path" in the "orthants" of 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 at least for the subclass of SDPs defined in Section 2, the transformed log-barrier functions do not have any such non-extreme stationary point.

3.1 The central path and its properties for the subclass

Since the variable w of (LD) is constrained to be positive, a natural problem to consider is the following log-barrier problem associated with (LD), which depends on the choice of a constant $\nu > 0$:

$$(LD_\nu) \qquad \min \left\{ f(w,y) - 2\nu \sum_{i=1}^{n} \log w_i : (w,y) \in \Re^n_{++} \times \Re^m \right\}.$$

(The reason for the factor 2 will become apparent shortly.) Not surprisingly, (LD_ν) is nothing but the transformed form of the standard dual log-barrier problem

$$(D_\nu) \quad \min \left\{ d^T z + b^T y - \nu \log(\det S) : \mathrm{Diag}(z) + \mathcal{A}^*(y) - C = S;\ S \succ 0 \right\},$$

under the bijective transformation given in Theorem 2.1 between $\Re^n_{++} \times \Re^m$ and $\mathcal{F}^0(D)$. Indeed, (D_ν) is transformed into the nonlinear program

$$\min \left\{ f(w,y) - \nu \log(\det(S(w,y))) : (w,y) \in \Re^n_{++} \times \Re^m \right\},$$

which is exactly (LD_ν) after the simplification

$$\log(\det S) = \log(\det(LL^T)) = 2\log(\det L) = 2\log\left(\prod_{i=1}^{n} w_i\right) = 2\sum_{i=1}^{n} \log w_i,$$

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

Recall that (D) has interior feasible solutions. Hence, under Assumptions A.1 and A.2, (D_ν) and the corresponding standard primal log-barrier problem

$$(P_\nu) \qquad \max \{ C \bullet X + \nu \log(\det X) : \mathrm{diag}(X) = d;\ \mathcal{A}(X) = b;\ X \succ 0 \}$$

each have unique solutions (z_ν, y_ν, S_ν) and X_ν, respectively, such that $X_\nu S_\nu = \nu I$. One can ask whether (LD_ν) has a unique stationary point, and if so, how this unique stationary point (i.e., global minimum) relates to X_ν and (z_ν, y_ν, S_ν). The following theorem establishes that (LD_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(c).

Theorem 3.1 *For each $\nu > 0$, problem (LD_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(c). In particular, we have $z(w_\nu, y_\nu) = z_\nu$ and $S(w_\nu, y_\nu) = S_\nu$. Moreover, $X(w_\nu, y_\nu) = X_\nu$.*

Proof. Let (w, y) be a stationary point of (LD_ν), and define $\nabla f \equiv \nabla f(w, y)$, $L \equiv L(w, y)$, $X \equiv X(w, y)$, $z \equiv z(w, y)$, and $S \equiv S(w, y)$. Since (w, y) is a stationary point of (LD_ν), it satisfies the first-order optimality conditions of (LD_ν)

$$\nabla_w f \circ w = 2\,\nu e, \qquad \nabla_y f = 0, \tag{7}$$

where $e \in \Re^n$ is the vector of all ones. Notice that (7), together with Lemma 2.5 and Theorem 2.3, implies that X is strictly feasible for (P) and f is strictly convex at (w, y).

Theorem 2.2(a) implies that the first equation of (7) is equivalent to $\operatorname{diag}(XL) \circ w = \nu e$, and the equality $w = \operatorname{diag}(L^T)$ implies that $\operatorname{diag}(XL) \circ \operatorname{diag}(L^T) = \nu e$, which in turn implies that $\operatorname{diag}(L^T X L) = \nu e$ since XL is upper triangular by Lemma 2.4. Since $L^T X L$ is diagonal by the same proposition, it follows that $L^T X L = \nu I$, and hence that $XS = \nu I$. This clearly implies that $X = X_\nu$ and $(z, y, S) = (z_\nu, y_\nu, S_\nu)$. This conclusion shows that (LD_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 a global minimum of (D_ν). ∎

3.2 Sufficient conditions for convergence

Given a sequence of points $\{(w^k, y^k)\}_{k\geq 0} \subset \Re^n_{++} \times \Re^m$, we define $L^k \equiv L(w^k, y^k)$, $X^k \equiv X(w^k, y^k)$, $S^k \equiv S(w^k, y^k)$, $z^k \equiv z(w^k, y^k)$, $g^k_w \equiv \nabla_w f(w^k, y^k)$ and $g^k_y \equiv \nabla_y f(w^k, y^k)$ for all $k \geq 0$. The following results provide sufficient conditions for the sequences $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$ to be bounded and/or for their accumulation points to be optimal solutions of problems (P) and (D), respectively.

Lemma 3.2 *Let $\{(w^k, y^k)\}_{k\geq 0} \subset \Re^n_{++} \times \Re^m$ be a sequence of points such that $g^k_w \geq 0$ for all $k \geq 0$. Then the sequence $\{X^k\}$ is bounded.*

Proof. By Lemma 2.5(a), we have that $\operatorname{diag}(X^k) = d$ for all $k \geq 1$ and, by Lemma 2.5(b), we have that $X^k \succeq 0$ for all $k \geq 1$. Clearly, these two conclusions together imply that $\{X^k\}$ is bounded. ∎

Lemma 3.3 *Let $\{(w^k, y^k)\}_{k\geq 0} \subset \Re^n_{++} \times \Re^m$ be a sequence of points such that $g_w^k \geq 0$ for all $k \geq 0$, $\lim_{k\to\infty} g_y^k = 0$ and the sequence $\{(w^k)^T g_w^k\}$ is bounded. Then, the sequences $\{(z^k, y^k, S^k)\}$ and $\{L^k\}$ are bounded. Consequently, $\{w^k\} \equiv \{diag(L^k)\}$ is also bounded.*

Proof. Let $\tilde{\mathcal{A}} : \mathcal{S}^n \to \Re^n \times \Re^m$ be the linear map defined by $\tilde{\mathcal{A}}(X) = (\mathrm{diag}(X), \mathcal{A}(X))$ for all $X \in \mathcal{S}^n$. Clearly, its adjoint $\tilde{\mathcal{A}}^* : \Re^n \times \Re^m \to \mathcal{S}^n$ is the map defined by $\tilde{\mathcal{A}}^*(z, y) = \mathrm{diag}(z) + \mathcal{A}^*(y)$ for all $(z, y) \in \Re^n \times \Re^m$. It is easy to see that Assumption A.2 implies that $\tilde{\mathcal{A}}^*$ is an injective map. Hence, it follows that $\tilde{\mathcal{A}}$ is an onto map. Since $\tilde{\mathcal{A}}$ is linear, this implies that $\tilde{\mathcal{A}}$ is an open map, that is, it maps open sets onto open sets. Now consider a strictly feasible point $\bar{X}$ of (P), and a strictly feasible point $(\bar{z}, \bar{y}, \bar{S})$ of (D). Since $\bar{X} \succ 0$, there exists $\eta > 0$ such that $\eta^{-1} I \succ \bar{X} \succ \eta I$. Define $\mathcal{N}^0 \equiv \{X \in \mathcal{S}^n : \eta^{-1} I \succ X \succ \eta I\}$. Clearly, $\mathcal{N}^0$ is an open set containing $\bar{X}$. Hence, we conclude that $\tilde{\mathcal{A}}(\mathcal{N}^0)$ is an open set containing $(d, b) = \tilde{\mathcal{A}}(\bar{X})$. By Theorem 2.2(b) and the assumption that $\lim_{k\to\infty} g_y^k = 0$, we conclude that $\lim_{k\to\infty} \tilde{\mathcal{A}}(X^k) = (d, b)$ and that $\tilde{\mathcal{A}}(X^k) \in \tilde{\mathcal{A}}(\mathcal{N}^0)$ 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 have

$$(X^k - \tilde{X}^k) \bullet (S^k - \bar{S}) = (X^k - \tilde{X}^k) \bullet \tilde{\mathcal{A}}^*(y^k - \bar{y}, z^k - \bar{z}) = \tilde{\mathcal{A}}(X^k - \tilde{X}^k)^T \begin{bmatrix} y^k - \bar{y} \\ z^k - \bar{z} \end{bmatrix} = 0$$

for all $k \geq k_0$. Using the above relation and the fact that $\tilde{X}^k \in \mathcal{N}^0$, we obtain

$$X^k \bullet S^k + \eta^{-1}(I \bullet \bar{S}) \geq X^k \bullet S^k + X^k \bullet \bar{S} = X^k \bullet \bar{S} + \tilde{X}^k \bullet S^k \geq \tilde{X}^k \bullet S^k \geq \eta(I \bullet S^k),$$

for all $k \geq k_0$. By Lemma 2.4(b) and the boundedness assumption on $\{(w^k)^T g_w^k\}$, the sequence $\{X^k \bullet S^k \equiv (w^k)^T g_w^k\}$ is bounded. Hence, the above inequality implies that the sequence $\{S^k\}$ is bounded. Observe that $S^k = L^k (L^k)^T$ and $\mathrm{tr}(S^k) = \|L^k\|_F^2$. Hence the boundedness of $\{S^k\}$ implies that of $\{L^k\}$, and in turn that of $\{w^k\} = \{\mathrm{diag}(L^k)\}$. Also, we have $\tilde{\mathcal{A}}^*(z^k, y^k) = C - S^k$, from which it follows that that $\{\tilde{\mathcal{A}}^*(z^k, y^k)\}$ is bounded. Since $\tilde{\mathcal{A}}^*$ is injective, this clearly implies that $\{(y^k, z^k)\}$ is bounded. ∎

Theorem 3.4 *Let $\{(w^k, y^k)\}_{k\geq 0} \subset \Re^n_{++} \times \Re^m$ be a sequence of points such that $g_w^k \geq 0$ for all $k \geq 0$, $\lim\limits_{k\to\infty} (w^k)^T g_w^k = 0$, and $\lim\limits_{k\to\infty} g_y^k = 0$. 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 Lemmas 3.2 and 3.3. To prove (b), let X^∞ and $(z^\infty, y^\infty, S^\infty)$ be accumulation points of the sequences $\{X^k\}$ and $\{(z^k, y^k, S^k)\}$, respectively. The assumptions and Lemma 2.5 imply that

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

and that $X^k \succeq 0$ and $\operatorname{diag}(X^k) = d$ for all $k \geq 0$. This clearly implies that $X^\infty \succeq 0$, $\mathcal{A}(X^\infty) = b$ and $\operatorname{diag}(X^\infty) = d$, 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, we have $(w^k)^T g_w^k = X^k \bullet S^k$, for all $k \geq 0$, from which it follows that $X^\infty \bullet S^\infty = 0$. 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 subsection, we introduce a straightforward barrier algorithm for solving (LD). The convergence of the algorithm is a simple consequence of Theorem 3.4.

Let constants $\gamma_1 \in [0,1)$, $\gamma_2 > 1$, and $\Gamma > 0$ be given, and for each $\nu > 0$, define

$$\mathcal{N}(\nu) \equiv \left\{ (w,y) \in \Re^n_{++} \times \Re^m : 2\gamma_1 \nu\, e \leq w \circ \nabla_w f(w,y) \leq 2\gamma_2 \nu\, e, \|\nabla_y f(w,y)\| \leq \Gamma \nu \right\}, \quad (8)$$

where $e \in \Re^n$ is the vector of all ones. Note that each $(w,y) \in \mathcal{N}(\nu)$ satisfies $\nabla_w f(w,y) > 0$ and that the unique minimizer (w_ν, y_ν) of (LD_ν) is in $\mathcal{N}(\nu)$, as can be seen from equation (7). 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 a minimization method to solve (LD_{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 (LD_{μ_k}) has a unique stationary point (w_ν, y_ν) 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 $\mathcal{N}(\mu_k)$.

Based on (8) and Proposition 2.4(b), the algorithm clearly produces a sequence of points $\{(w^k, y^k)\}_{k \geq 0}$ which satisfies the hypotheses of Theorem 3.4. 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 (LD). The basic idea is to produce a sequence of points $\{(w^k, y^k)\}_{k \geq 0}$ satisfying the hypotheses of Theorem 3.4 via the minimization of a special merit function defined inside $\Re^n_{++} \times \Re^m$. 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 $(w^0, y^0) \in \Re^n_{++} \times \Re^m$ satisfying $\nabla_w f(w^0, y^0) > 0$ is given.

4.1 Definitions, technical results, and the algorithm

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

$$\Xi \equiv \left\{(w, y) \in \Re^n_{++} \times \Re^m : \nabla_w f(w, y) > 0 \text{ and } f(w, y) < f^+\right\}. \tag{9}$$

Note that $(w^0, y^0) \in \Xi$. The potential reduction algorithm, which we state explicitly at the end of this subsection, will be initialized with the point (w^0, y^0) and will subsequently produce a sequence of points $\{(w^k, y^k)\}_{k \geq 0}$ such that $(w^k, y^k) \in \Xi$ for all $k \geq 0$. The requirement that $\nabla_w f(w^k, y^k)$ be positive for all k is reasonable in light of our goal of producing a sequence satisfying the hypotheses of Theorem 3.4, but this requirement has the added benefit of ensuring that $\nabla^2 f(w^k, y^k)$ is positive definite for all k (see Lemma 2.5(b) and Theorem 2.3). The significance of $\nabla^2 f(w^k, y^k)$ being positive definite will be made clear shortly. The second 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 define $F : \Xi \to \Re^n_{++} \times \Re^m$ by

$$F(w, y) \equiv \begin{bmatrix} w \circ \nabla_w f(w, y) \\ \nabla_y f(w, y) \end{bmatrix}. \tag{10}$$

With this definition, our goal of producing a sequence of points satisfying the hypotheses of Theorem 3.4 can be stated more simply as the goal of producing a sequence $\{(w^k, y^k)\}_{k \geq 0} \subset \Xi$ such that

$$\lim_{k \to \infty} F(w^k, y^k) = 0. \tag{11}$$

The primary tool for accomplishing (11) is the merit function $\mathcal{M} : \Xi \to \Re$ given by

$$\mathcal{M}(w, y) \equiv \zeta \log \|F(w, y)\|^2 - \sum_{i=1}^{n} \log F_i(w, y) - \log(f^+ - f(w, y)), \tag{12}$$

where ζ is an arbitrary constant satisfying $2\zeta > n$ and $F_i : \Xi \to \Re$ is the i-th coordinate function of F, that is, $F_i(w, y) = w_i[\nabla_w f(w, y)]_i$ for $i = 1, \ldots, n$. The usefulness of this merit function comes from the fact that (11) can be accomplished via the iterative minimization of $\mathcal{M}$ by taking a step along the Newton direction of the system $F(w, y) = 0$ at the current point. The following three results are related to this minimization scheme.

Lemma 4.1 *Let* $(w,y) \in \Re^n_{++} \times \Re^m$*, and suppose that* $(\Delta w, \Delta y), (r,s) \in \Re^n \times \Re^m$ *satisfy* $F'(w,z)[\Delta w, \Delta y] = (r,s)$*. Then the following equation holds:*

$$\begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix}^T \nabla^2 f(w,y) \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} + \Delta w^T P(w,y) \Delta w = (w^{-1} \circ r)^T \Delta w + s^T \Delta y, \tag{13}$$

where $P(w,y) \equiv \mathit{Diag}\left(w^{-1} \circ \nabla_w f(w,y)\right)$ *and* $w^{-1} \equiv (w_1^{-1}, \ldots, w_n^{-1})^T$*.*

Proof. Note that (10) can be rewritten as

$$F(w,y) = \begin{bmatrix} w \\ e \end{bmatrix} \circ \nabla f(w,y),$$

where $e \in \Re^m$ is the vector of all ones. Hence, the equation $F'(w,z)[\Delta w, \Delta y] = (r,s)$ can be written as the system

$$\begin{bmatrix} \Delta w \\ 0 \end{bmatrix} \circ \nabla f(w,y) + \begin{bmatrix} w \\ e \end{bmatrix} \circ \left(\nabla^2 f(w,y) \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} \right) = \begin{bmatrix} r \\ s \end{bmatrix}.$$

Taking the Hadamard product of each side with $[w^{-T}, e^T]^T \in \Re^n_{++} \times \Re^m$, then multiplying on the left by the row vector $[\Delta w^T, \Delta y^T]$, and finally using the definition of $P(w,y)$, we obtain (13). ■

Lemma 4.2 $F'(w,y)$ *is nonsingular for all* $(w,y) \in \Xi$*.*

Proof. To prove that $F'(w,y)$ is nonsingular, it suffices to show that $(\Delta w, \Delta y) = (0,0)$ is the unique solution of the system $F'(w,y)[\Delta w, \Delta y] = (0,0)$. Indeed, let $(\Delta w, \Delta y)$ be such a solution; we will prove that $(\Delta w, \Delta y) = 0$.

We first note that the assumption that $(w,y) \in \Xi$ implies, by Lemma 2.5(b), that $X(w,y) \succ 0$, which in turn implies that $\nabla^2 f(w,y) \succ 0$ by Theorem 2.3. In addition, since both w and $\nabla_w f(w,y)$ are positive, we see that $P(w,y) \succ 0$ where $P(w,y)$ is defined as in Lemma 4.1.

By Lemma 4.1, equation (13) holds with $(r,s) = (0,0)$. Using the positive definiteness of $\nabla^2 f(w,y)$ and $P(w,y)$, it is now straightforward to see that $(\Delta w, \Delta y) = (0,0)$. ■

A direct consequence of Lemma 4.2 is that, for each $(w,y) \in \Xi$, the Newton direction $(\Delta w, \Delta y)$ for the system $F(w,y) = 0$ exists at (w,y). Stated differently, Lemma 4.2 shows that the system

$$F'(w,y)[\Delta w, \Delta y] = -F(w,y) \tag{14}$$

has a unique solution for all $(w,y) \in \Xi$.

Lemma 4.3 *Let $(w, y) \in \Xi$, and let $(\Delta w, \Delta y)$ be the Newton direction at (w, y) given by (14). Then $(\Delta w, \Delta y)$ is a descent direction for $\mathcal{M}$ at (w, y).*

Proof. We first prove four simple results which combine to prove the Lemma. Let $f \equiv f(w, y)$, $\nabla f \equiv \nabla f(w, y)$, $F \equiv F(w, y)$, $\nabla F \equiv \nabla F(w, y)$, and $\nabla M \equiv \nabla M(w, y)$. Then equation (14) implies that

$$F^T \nabla F^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} = -F^T F = -\|F\|^2 \quad \text{and} \quad \nabla F_i^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} = -F_i \;\; \forall \;\; i = 1, \ldots, n. \quad (15)$$

We note that $\nabla^2 f(w, y) \succ 0$, $P(w, y) \succ 0$, as mentioned in the proof of Lemma 4.2, and $(\Delta w, \Delta y) \neq 0$ since $(w, y) \in \Xi$. Applying Lemma 4.1 to (14) with $(r, s) = -(w \circ \nabla_w f, \nabla_y f)$, we derive that

$$\nabla f^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} \equiv \nabla_w f^T \Delta w + \nabla_y f^T \Delta y < 0. \quad (16)$$

In addition, differentiating (12), we obtain

$$\nabla \mathcal{M} = 2\,\zeta \, \frac{\nabla F F}{\|F\|^2} - \sum_{i=1}^{n} \frac{\nabla F_i}{F_i} + \frac{\nabla f}{f^+ - f}. \quad (17)$$

Now, using (15), (16), (17), the fact that $2\,\zeta > n$, and the fact that $f(w, y) < f^+$, we finally have

$$\nabla \mathcal{M}^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix} < 2\,\zeta \, \frac{(-\|F\|^2)}{\|F\|^2} - \sum_{i=1}^{n} \frac{(-F_i)}{F_i} = -2\zeta + n < 0, \quad (18)$$

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

Given $(w, y) \in \Re^n_{++} \times \Re^m$ and $\alpha > 0$, we define

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

where $(\Delta w, \Delta y)$ is the Newton direction given by (14). An important step in the potential reduction algorithm is the Armijo line search: given $(w, y) \in \Xi$ and $\sigma \in (0, 1)$, the line search selects a step size $\alpha > 0$ such that $(w(\alpha), y(\alpha)) \in \Xi$ and

$$\mathcal{M}(w(\alpha), y(\alpha)) - \mathcal{M}(w, y) \leq \sigma \alpha \nabla \mathcal{M}(w, y)^T \begin{bmatrix} \Delta w \\ \Delta y \end{bmatrix}. \quad (19)$$

Due to the fact that Ξ is an open set and also due to Lemma 4.3, such an α can be found by back-tracking in a finite number of steps.

We are now ready to state the potential reduction algorithm.

Potential Reduction (PR) Algorithm:

Let $\zeta > n/2$ and $\sigma, \rho \in (0,1)$ be given, and set $k = 0$.

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

1. Solve system (14) with respect to (w^k, y^k) to obtain the Newton direction $(\Delta w^k, \Delta y^k)$.
2. Let ℓ_k be the smallest nonnegative integer such that $(w^k(\rho^{\ell_k}), y^k(\rho^{\ell_k})) \in \Xi$ and such that (19) holds with $\alpha = \rho^{\ell_k}$.
3. Set $(w^{k+1}, y^{k+1}) = (w^k(\rho^{\ell_k}), y^k(\rho^{\ell_k}))$, increment k by 1, and return to step 1.

End

4.2 Convergence of the algorithm

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 Lemma 4.4.

Let $\{(w^k, y^k)\}_{k \geq 0}$ 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 $k \geq 0$. We have the following important result.

Lemma 4.4 *The sequences $\{(z^k, y^k, S^k)\}$ and $\{L^k\}$ are bounded. Consequently, so is $\{w^k\}$.*

Proof. As in the proof of Lemma 3.3, it suffices to show that the sequence $\{S^k\}$ is bounded. Indeed, let $\bar{X}$ be as in Assumption A.1 and observe that

$$\begin{aligned} \bar{X} \bullet S^k &= \bar{X} \bullet \left[\operatorname{diag}(z^k) + \mathcal{A}^*(y^k) - C\right] \\ &= \operatorname{diag}(\bar{X})^T z^k + \mathcal{A}(\bar{X})^T y^k - C \bullet \bar{X} \\ &= d^T z^k + b^T y^k - C \bullet \bar{X} = f(w^k, y^k) - C \bullet \bar{X}. \end{aligned}$$

Since $\bar{X} \succ 0$, $S^k \succeq 0$ for all $k \geq 0$ and since $f(w^k, y^k) < f^+$ for all $k \geq 0$, it follows from the above relation that $\{S^k\}$ is bounded. ∎

The reader may have noticed that the conclusions of Lemma 4.4 are identical to those of Lemma 3.3, namely that the sequences $\{(z^k, y^k, S^k)\}$, $\{L^k\}$, and $\{w^k\}$ are bounded. The lemmas differ, however, in the assumptions placed on the sequence $\{w^k, y^k\}$. Lemma 3.3 assumes that $\nabla_w f(w^k, y^k) \geq 0$ for all $k \geq 0$, that $\{(w^k)^T \nabla_w f^k\}$ is bounded, and that $\lim_{k \to \infty} \nabla_y f^k = 0$, whereas Lemma 4.4 simply uses that $\{f^k\}$ is bounded from above. This difference highlights the special characteristics of the potential reduction algorithm when compared with, say, the log-barrier algorithm presented in Subsection 3.3. In particular, we wish to prove that algorithm PR produces a sequence of points satisfying (11), and for this we require the boundedness of the various sequences given in Lemma 4.4. (This requirement is detailed below in Theorem 4.6). Hence, the boundedness of $\{f^k\}$ is critical in that it guarantees the result of Lemma 4.4 which in turn guarantees the convergence of the algorithm. This provides a motivation for the atypical term $-\log(f^+ - f(w,y))$ found in the merit function $\mathcal{M}$ since this term is precisely what ensures the boundedness of $\{f^k\}$.

Also important to the convergence of the potential reduction algorithm is the following lemma.

Lemma 4.5 *The sequence $\{F^k\}$ is bounded.*

Proof. Consider the function $p : \Re^n_{++} \times \Re^m \to \Re$ defined by

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

where ζ is the same constant appearing in (12). It is not difficult to verify (see Monteiro and Pang [8]) that p is coercive, i.e., for every sequence $\{(r^k, s^k)\} \subset \Re^n_{++} \times \Re^m$, if $\lim_{k\to\infty} \|(r^k, s^k)\| = \infty$, then $\lim_{k\to\infty} p(r^k, s^k) = \infty$. This property implies that

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

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

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

$$\mathcal{M}(w,y) = p(F(w,y)) - \log(f^+ - f(w,y)). \tag{20}$$

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 $f(w,y) \ge f^-$ for all $(w,y) \in \Re^n_{++} \times \Re^m$. Hence, (20) implies that

$$p(F(w,y)) \le \mathcal{M}(w,y) + \log(f^+ - f^-) \tag{21}$$

for all $(w,y) \in \Xi$. Now combining (21) with the fact that the potential reduction algorithm decreases the merit function $\mathcal{M}$ in each iteration, we see that

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

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

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

We conclude that $\{F^k\}$ is contained in the compact set $\Lambda(\bar{\gamma})$ and hence 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 3.2 and 4.4; and once (a) is established, part (c) follows immediately from Theorem 3.4.

Theorem 4.6 *Let $\{(w^k, y^k)\}_{k\geq 0}$ 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 4.5 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 Lemma 4.4, we may also assume, without loss of generality, that the sequence $\{(w^k, y^k)\}_{k\in K}$ converges to a point (w^∞, y^∞).

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

$$\eta_1 < \lim_{k\in K}\left(\zeta \log \|F^k\|^2\right), \quad \eta_2 < \lim_{k\in K}\left(-\sum_{i=1}^{n} \log F_i^k\right), \quad \eta_3 < \lim_{k\in K}\left(-\log(f^+ - f^k)\right). \tag{22}$$

These three inequalities together imply that

$$\lim_{k\in K} f^k \neq f^+ \qquad \text{and} \qquad F_i^\infty \neq 0 \;\; \forall \;\; i = 1, \ldots, n$$

since otherwise, $\mathcal{M}(w^k, y^k)$ 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 $F^\infty \in \Re^n_{++} \times \Re^m$, and so we clearly have that $(w^\infty, y^\infty) \in \Xi$. It follows that the Newton direction $(\Delta w^\infty, \Delta y^\infty)$ of system (14) exists at (w^∞, y^∞), and in addition, the sequence $\{(\Delta w^k, \Delta y^k)\}_{k\in K}$ of Newton directions converges to $(\Delta w^\infty, \Delta y^\infty)$. By equation (18) found in the proof of Lemma 4.3, we have moreover that

$$\nabla \mathcal{M}(w^\infty, y^\infty)^T \begin{bmatrix} \Delta w^\infty \\ \Delta y^\infty \end{bmatrix} < n - 2\zeta < 0. \tag{23}$$

Since $\{(w^k, y^k)\}_{k\in K}$ converges in Ξ and since $\mathcal{M}$ is continuous on Ξ, it follows that $\{\mathcal{M}(w^k, y^k)\}_{k\in K}$ converges, and since the entire sequence $\{\mathcal{M}(w^k, y^k)\}$ is monotonically decreasing and bounded below due to (22), we conclude that it converges as well. Using the relation

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

for all $k \geq 0$, where the first and second inequalities follow from (19) and (18), respectively, we clearly see that $\lim_{k\in K} \rho^{\ell_k} = 0$ since the left-hand side tends to zero as $k \in K$ tends to infinity. This implies $\lim_{k\in K} \ell_k = \infty$.

Since ℓ_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 $(w, y) \in \Xi$, the line search finds a step size $\alpha > 0$ such that $(w(\alpha), y(\alpha)) \in \Xi$ and such that relation (19) is satisfied. Using that $\{(w^k, y^k)\}_{k \in K}$ and $\{(\Delta w^k, \Delta y^k)\}_{k \in K}$ converge to $(w^\infty, y^\infty) \in \Xi$ and $(\Delta w^\infty, \Delta y^\infty)$, respectively, where Ξ is an open set, it is straightforward to see that there exist $\tilde{\ell} \geq 0$ and $\tilde{k} \in K$ such that

$$(w^k(\rho^\ell), y^k(\rho^\ell)) \in \Xi$$

for all $\ell \geq \tilde{\ell}$ and all $k \in K$ such that $k \geq \tilde{k}$. Hence, there exists $\hat{k} > \tilde{k}$ such that, for all $k \geq \hat{k}$, we have $\ell_k > \tilde{\ell}$, which implies $(w^k(\rho^{\ell_k - 1}), y^k(\rho^{\ell_k - 1})) \in \Xi$ but (19) is not satisfied for the step size $\rho^{\ell_k - 1}$, i.e.,

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

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

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

which contradicts (23) 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 Final Remarks

Since the existence of the Newton directions in the potential reduction algorithm in Section 4 are insured only in a certain neighborhood of the central path lying in the feasible region $\Re^n_{++} \times \Re^N$, the potential reduction algorithm starts and generates its iterates only in this neighborhood. This is in contrast with the log-barrier algorithm in Section 3 which can start from any point in $\Re^n_{++} \times \Re^N$ and can use any descent search directions to approximately minimize the log-barrier subproblems.

Interior-point methods for general linear SDPs in the transformed space will be presented and analyzed in the third part of this series of papers.

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. 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.

[3] 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.

[4] K. Fujisawa, M. Kojima, and K. Nakata. Exploiting sparsity in primal-dual interior-point methods for semidefinite programming. *Mathematical Programming*, 79:235–253, 1997.

[5] 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).

[6] 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.

[7] 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.

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