Math. Program., Ser. A 103, 427–444 (2005)

Digital Object Identifier (DOI) 10.1007/s10107-004-0564-1

Samuel Burer · Renato D.C. Monteiro

Local Minima and Convergence in Low-Rank Semidefinite Programming

Received: September 30, 2003 / Accepted: November 1, 2004
Published Online: December 29, 2004 –

Abstract. The low-rank semidefinite programming problem LRSDP_r is a restriction of the semidefinite programming problem SDP in which a bound r is imposed on the rank of X, and it is well known that LRSDP_r is equivalent to SDP if r is not too small. In this paper, we classify the local minima of LRSDP_r and prove the optimal convergence of a slight variant of the successful, yet experimental, algorithm of Burer and Monteiro [5], which handles LRSDP_r via the nonconvex change of variables $X = RR^T$. In addition, for particular problem classes, we describe a practical technique for obtaining lower bounds on the optimal solution value during the execution of the algorithm. Computational results are presented on a set of combinatorial optimization relaxations, including some of the largest quadratic assignment SDPs solved to date.

Key words. Semidefinite programming – Low-rank matrices – Vector programming – Combinatorial optimization – Nonlinear programming – Augmented Lagrangian – Numerical experiments

1. Introduction

We study the standard-form semidefinite programming problem

$$\begin{array}{llll} \text{SDP} & \min & C \bullet X & \\ & \text{s.t.} & A_i \bullet X = b_i, & i = 1, \dots, m \\ & & X \succeq 0 & \end{array}$$

and its dual

$$\begin{array}{lll} \text{DSDP} & \max & b^T y \\ & \text{s.t.} & \sum_{i=1}^{m} y_i A_i + S = C \\ & & S \succeq 0, \end{array}$$

where the matrices $C, A_1, \dots, A_m$ and the vector b are the data and the matrices X, S and the vector y are the variables. Each matrix is $n \times n$ symmetric (i.e., an element of $\mathcal{S}^n$); $M \bullet N = \mathrm{trace}(MN)$; and $M \succeq 0$ (or $M \in \mathcal{S}^n_+$) indicates that M is symmetric and positive semidefinite. We assume that SDP has an interior feasible solution, but note that we do not assume the same of DSDP. In addition, we make the assumption that both

S. Burer: Department of Management Sciences, University of Iowa, Iowa City, IA 52242-1000, USA. e-mail: samuel-burer@uiowa.edu. This author was supported in part by NSF Grant CCR-0203426.

R.D.C. Monteiro: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. e-mail: monteiro@isye.gatech.edu. This author was supported in part by NSF Grants CCR-0203113 and INT-9910084 and ONR grant N00014-03-1-0401.

problems attain their optimal value with zero duality gap, i.e., there exist feasible X and (S, y) such that $X \bullet S = 0$.

There are many varied algorithms for solving SDP and DSDP, and it is convenient to divide the methods into three groups according to their methodology and their effectiveness on problems of different size. The first group is the second-order primal-dual interior-point methods which use Newton's method to solve SDP and DSDP simultaneously (for example, see [1, 10, 13, 16, 17, 26]). These methods are capable of solving small- to medium-sized problems very accurately but have difficulty on large, sparse problems because of their inherent high demand for storage and computation. The second group is similar to the first, but instead of solving for the Newton direction exactly at each iteration, an iterative solver is used to find the direction instead (for example, see [4, 14, 18, 24, 25]). This approach allows large-scale problems to be solved to a medium amount of accuracy. The final group consists of the first-order nonlinear programming algorithms (for example, see [5, 6, 9]), which use fast, gradient-based techniques to solve a nonlinear reformulation of either SDP or DSDP. Strong computational results, obtaining medium accuracy on large problems, have been reported for these algorithms, especially on the class of semidefinite relaxations of combinatorial problems. A comprehensive survey of all three of these groups of algorithms can be found in [15].

This paper investigates the first-order nonlinear programming algorithm introduced by Burer and Monteiro in [5]. The algorithm is motivated by the following two results, the first of which establishes the existence of extreme points for SDP (e.g., see Rockafellar [22]):

Theorem 1.1. *A nonempty closed convex set with no lines has an extreme point.*

The second result, which provides a bound on the rank of extreme points for SDP (Pataki [19]), uses the following definition: for any positive integer ℓ,

$$r_\ell := \max\{r \text{ integer} : r(r+1)/2 \le \ell, \}. \qquad (1)$$

Theorem 1.2. *If $\bar{X}$ is an extreme point of SDP, then rank$(\bar{X}) \le r_m$.*

Since the optimal value of SDP is attained at an extreme point, the following low-rank semidefinite programming problem is equivalent to SDP for any integer $r \ge r_m$:

$$\begin{array}{llll} \text{LRSDP}_r & \min & C \bullet X & \\ & \text{s.t.} & A_i \bullet X = b_i, & i = 1, \dots, m \\ & & X \succeq 0,\ \text{rank}(X) \le r & \end{array}$$

Unless otherwise stated, we assume throughout that the integer r has been chosen large enough so that the two problems are indeed equivalent.

Since the constraint rank$(X) \le r$ is difficult to handle directly, Burer and Monteiro [5] propose to use the fact that any $X \succeq 0$ with rank$(X) \le r$ may be written as $X = RR^T$ for some $R \in \Re^{n\times r}$ to reformulate LRSDP$_r$ as the nonlinear program

$$\begin{array}{llll} \text{NSDP}_r & \min & C \bullet RR^T & \\ & \text{s.t.} & A_i \bullet RR^T = b_i, & i = 1, \dots, m. \end{array}$$

An immediate benefit of NSDP_r is the reduced number of variables as compared with LRSDP_r. Burer and Monteiro [5] then use a first-order augmented Lagrangian algorithm to solve NSDP_r on the relaxations of some large-scale combinatorial optimization problems such as maximum cut and maximum stable set. They report strong computational results, including speed-up factors of nearly 500 over the second fastest algorithm on some problems (see section 4.2 of [5]), based on the fact that: (i) the function and gradient evaluations of the augmented Lagrangian function are extremely quick, especially when the A_i's are sparse or low-rank and m and r are small; and (ii) even though NSDP_r is nonconvex, an optimal solution to NSDP_r, and hence SDP, is always achieved experimentally. Although Burer and Monteiro [5] provide some insight as to why (ii) occurs, a formal convergence proof for their method is not established.

In this paper, we study LRSDP_r and NSDP_r in an effort to shed some theoretical light on the intriguing practical behavior (ii) observed in [5]. In Section 2, we show some basic facts relating LRSDP_r and NSDP_r, including an explicit correspondence between the local minima of the two problems. In particular, we show that the change of variables does not introduce any extraneous local minima. Then, in Section 3, we provide the following classification of the local minima of LRSDP_r: if X is a local minimum, then either X is an optimal extreme point for SDP, or X is contained in the relative interior of a face of the feasible set of SDP which is constant with respect to the objective function.

In Section 4, we study the theoretical properties of sequences $\{R^k\}$ produced by augmented Lagrangian algorithms applied to NSDP_r. Then in Section 5 we use these properties to investigate a slight variant of the augmented Lagrangian algorithm proposed by Burer and Monteiro [5] for solving NSDP_r, which differs only in the addition of the term $\mu \det(R^T R)$ to the augmented Lagrangian function, where $\mu > 0$ is a scaling parameter of arbitrarily small magnitude which is simply required to go to zero as the algorithm progresses. Assuming that a local minimum is obtained at each stage of the algorithm, we show that any accumulation point $\bar{R}$ of the resulting sequence is an optimal solution of NSDP_r, and hence $\bar{X} = \bar{R}\bar{R}^T$ is an optimal solution of SDP. Moreover, we show that the algorithm produces an optimal dual $\bar{S}$ as well.

Finally in Section 6, we discuss some computational issues, including how, for special problem classes, one can calculate lower bounds on the optimal value of SDP during the execution of the algorithm. From a practical point of view, this addresses a key drawback of the algorithm of Burer and Monteiro [5] in which lower bounds were not available. We then provide computational results on the SDP relaxations of some large-scale maximum cut, maximum stable set, and quadratic assignment problems. The first two classes of problems are also considered in [5], while for the third class, we report here some of the largest quadratic assignment SDP relaxations solved to date.

2. Some Facts Concerning the Change of Variables

In this section, we establish some basic facts concerning the change of variables $X = RR^T$. Note that each of these results is valid for any r.

At first glance, it is unclear how the local minima of LRSDP_r relate to the local minima of NSDP_r. By continuity, we know that if X is a local minimum then each R satisfying $X = RR^T$ is a local minimum, though it may be the case that X is not a local minimum when R is. In other words, the change of variables may introduce extraneous local minima. In actuality, however, the results below show that this cannot happen.

The following lemma establishes a simple correspondence between any R and S such that $RR^T = SS^T$.

Lemma 2.1. *Suppose $R, S \in \Re^{n\times r}$ satisfy $RR^T = SS^T$. Then $S = RQ$ for some orthogonal matrix $Q \in \Re^{r\times r}$.*

Proof. Let $q = \text{rank}(RR^T)$. By considering the eigenvalue decomposition of RR^T, it is easy to see that there exists $U \in \Re^{n\times r}$ such that $RR^T = UU^T$ and the last $r - q$ columns of U are zero. To prove the lemma, we exhibit an orthogonal Q_1 such that $R = UQ_1$, which similarly implies the existence of Q_2 such that $S = UQ_2$. Hence, $Q = Q_1^T Q_2$ satisfies $S = RQ$.

Using that $UU^T = RR^T$ is positive semidefinite, it is straightforward to argue $\text{Null}(U^T) = \text{Null}(R^T)$, which implies $\text{Range}(U) = \text{Range}(R)$. Hence, if we write

$$U = \left(\tilde{U}\ 0\right),$$

so that $\tilde{U} \in \Re^{n\times q}$ denotes the nonzero part of U, there exists a unique $\tilde{H} \in \Re^{q\times r}$ such that $\tilde{U}\tilde{H} = R$. Hence,

$$\tilde{U}(I_q - \tilde{H}\tilde{H}^T)\tilde{U}^T = 0.$$

Since $\tilde{U}$ is full rank, this implies $\tilde{H}\tilde{H}^T = I_q$, i.e., the rows of $\tilde{H}$ are orthonormal. Extending $\tilde{H}$ to an orthogonal matrix $Q_1 \in \Re^{r\times r}$, we have $UQ_1 = R$, as desired. □

The next lemma is a fundamental observation about the local minima of NSDP$_r$ — namely that the local minima occur as sets parameterized by multiplication by an orthogonal matrix. The proof is straightforward based on the fact that $RR^T = RQQ^TR^T$ for all R and all orthogonal Q.

Lemma 2.2. *$\bar{R}$ is a local minimum of NSDP$_r$ if and only if $\bar{R}Q$ is a local minimum for all orthogonal $Q \in \Re^{n\times r}$.*

By combining Lemmas 2.1 and 2.2, we now show that the change of variables $X = RR^T$ does not introduce any extraneous local minima.

Proposition 2.3. *Suppose $\bar{X} = \bar{R}\bar{R}^T$, where $\bar{X}$ is feasible for LRSDP$_r$ and hence $\bar{R}$ is feasible for NSDP$_r$. Then $\bar{X}$ is a local minimum of LRSDP$_r$ if and only if $\bar{R}$ is a local minimum of NSDP$_r$.*

Proof. As discussed above, continuity of the map $R \mapsto RR^T$ implies that if $\bar{X}$ is a local minimum, then so is $\bar{R}$. In fact, any R such that $\bar{X} = RR^T$ is a local minimum.

Now suppose that $\bar{X}$ is not a local minimum of LRSDP$_r$. Then there exists a sequence of feasible solutions $\{X^k\}$ of LRSDP$_r$ converging to $\bar{X}$ such that $C \bullet X^k < C \bullet \bar{X}$ for all k. For each k, choose R^k such that $X^k = R^k(R^k)^T$. Since $\{X^k\}$ is bounded, it follows that $\{R^k\}$ is bounded and hence has a subsequence $\{R^k\}_{k\in\mathcal{K}}$ converging to some R such that $\bar{X} = RR^T$. Since $C \bullet R^k(R^k)^T = C \bullet X^k < C \bullet \bar{X} = C \bullet RR^T$, we see that R is not a local minimum of NSDP$_r$. Using the fact that $\bar{X} = \bar{R}\bar{R}^T = RR^T$ together with Lemmas 2.1 and 2.2, we conclude that $\bar{R}$ is not a local minimum of NSDP$_r$. □

We remark that arguments similar to those in this section can be used to show that the local minima of any continuous optimization problem over the set $\{X : X \succeq 0,\ \mathrm{rank}(X) \le r\}$ and the local minima of its corresponding reformulation by the change of variables $X = RR^T$ are related according to Proposition 2.3.

3. Local Minima Classification

In this section, we provide a classification of the local minima of LRSDP$_r$. By Proposition 2.3, this also serves to classify the local minima of NSDP$_r$.

Given a point $\bar{X} \in \mathcal{S}^n_+$, we define the following convex set, which consists of all matrices in $\mathcal{S}^n_+$ having the same inner product with $C, A_1, \ldots, A_m$ as $\bar{X}$:

$$\mathcal{X}(\bar{X}) := \{X \in \mathcal{S}^n_+ : C \bullet X = C \bullet \bar{X},\ A_i \bullet X = A_i \bullet \bar{X},\ \forall\, i = 1, \ldots, m\}.$$

When $\bar{X}$ is feasible for SDP, and hence satisfies $A_i \bullet \bar{X} = b_i$ for $i = 1, \ldots, m$, one can think of $\mathcal{X}(\bar{X})$ as an "iso-cost" slice of the feasible region. The following well-known result (see [20], for example) characterizes when $\bar{X}$ is an extreme point of $\mathcal{X}(\bar{X})$.

Proposition 3.1. *Let $\bar{X} \in \mathcal{S}^n_+$ be given and let $R \in \Re^{n\times r}$ be a matrix with full-column rank satisfying $\bar{X} = RR^T$. Then $\bar{X}$ is an extreme point of $\mathcal{X}(\bar{X})$ if and only if the system of linear equations*

$$\phi(R) \qquad \begin{aligned} C \bullet R\Delta R^T &= 0 \\ A_i \bullet R\Delta R^T &= 0, \quad \forall\, i = 1, \ldots, m, \end{aligned}$$

where $\Delta \in \mathcal{S}^r$ is the unknown, has a unique solution (i.e., the trivial solution $\Delta = 0$).

An observation, which will be used in Section 4, is that $\phi(R)$ has a nontrivial solution when $r(r+1)/2 > m + 1$, i.e., the number of unknowns is larger than the number of equations, or equivalently when $r > r_{m+1}$. Accordingly, we see that if $\bar{X}$ is an extreme point of $\mathcal{X}(\bar{X})$, then $r \le r_{m+1}$. Note that Theorem 1.2 follows as an immediate consequence of this observation (with C removed from consideration and $m + 1$ replaced by m).

The following lemma is the key result which serves to classify the local minima of LRSDP$_r$. The basic idea is a "rank-reduction" technique proposed by Pataki [20] (also easily derived from [19]), in which, if the rank of X is large enough, then X may be moved to a matrix of lower rank without changing its inner product with $C, A_1, \ldots, A_m$. The lemma can be seen as an application of this rank-reduction technique to a sequence of points. In its proof, we use the following notation: for $E \in \mathcal{S}^n$, we let $\lambda_{\min}(E)$ and $\lambda_{\max}(E)$ denote the minimum and maximum eigenvalue of E and $\|E\|$ denote the operator norm of E defined as $\|E\| := \max\{-\lambda_{\min}(E), \lambda_{\max}(E)\}$.

Lemma 3.2. *Assume that $\bar{X}$ is an extreme point of the set $\mathcal{X}(\bar{X})$ and that $\{X^k\} \subset \mathcal{S}^n_+$ is a sequence converging to $\bar{X}$. Then there exists a sequence $\{Y^k\} \subset \mathcal{S}^n_+$ satisfying the following two conditions:*

(a) $\{Y^k\}$ converges to $\bar{X}$;
(b) Y^k is an extreme point of $\mathcal{X}(X^k)$ for all k.

As a consequence, each Y^k *satisfies* $\text{rank}\,(Y^k) \le r_{m+1}$.

Proof. If X^k is an extreme point of the set $\mathcal{X}(X^k)$ for every k, then we may simply define $Y^k = X^k$ for all k. Otherwise, we will prove that $\{X^k\}$ can be "updated" to a sequence $\{Y^k\}$ converging to $\bar{X}$ and having the properties that $Y^k \in \mathcal{X}(X^k)$ for all k and $\text{rank}\,(Y^k) \le \text{rank}\,(X^k) - 1$ for all $k \in \mathcal{K}$, where $\mathcal{K}$ is the set of indices k for which X^k is not an extreme point of $\mathcal{X}(X^k)$. A simple argument shows that, after performing a finite number of these updates, we obtain a sequence $\{Y^k\}$ satisfying (a) and (b).

To prove the above claim, we factor each X^k as $X^k = R^k(R^k)^T$, where $R^k \in \Re^{n \times r_k}$ has full-column rank, or equivalently, $r_k = \text{rank}\,(R^k)$. Since $\{X^k\}$ is bounded, the sequence $\{\|R^k\|\}$ is also bounded. We next build the sequence $\{Y^k\}$ and an auxiliary sequence $\{Z^k\}$ as follows. If X^k is an extreme point of $\mathcal{X}(X^k)$, then we define $Y^k := X^k$ and $Z^k := X^k$. Now suppose X^k is not an extreme point of $\mathcal{X}(X^k)$. By Proposition 3.1, the system of equations $\phi(R^k)$ has a nontrivial solution $\Delta^k \in \mathcal{S}^{r_k}$. We assume without loss of generality that $\|\Delta^k\| = \lambda_{\max}(\Delta^k) = 1$; otherwise, we can scale Δ^k and/or take $-\Delta^k$. We then define

$$
\begin{aligned}
Y^k &:= X^k - R^k\Delta^k(R^k)^T = R^k(I - \Delta^k)(R^k)^T, \\
Z^k &:= X^k + R^k\Delta^k(R^k)^T = R^k(I + \Delta^k)(R^k)^T.
\end{aligned}
$$

Using the fact that $\|\Delta^k\| = \lambda_{\max}(\Delta^k) = 1$ and Δ^k is a solution of $\phi(R^k)$, we easily see that $\{Y^k\}$ and $\{Z^k\}$ are sequences of positive semidefinite matrices such that $Y^k, Z^k \in \mathcal{X}(X^k)$ for all k and $\text{rank}\,(Y^k) \le \text{rank}\,(X^k) - 1$ for all $k \in \mathcal{K}$. It remains to show that $\{Y^k\}$ converges to $\bar{X}$. Indeed, let $\{U^k\}$ be the sequence defined as $U^k = R^k\Delta^k(R^k)^T$ if $k \in \mathcal{K}$ and $U^k = 0$ if $k \notin \mathcal{K}$. Note that $\{U^k\}$ is clearly bounded. Now let $\bar{U}$ be an accumulation point of $\{U^k\}$. Since $Y^k = X^k - U^k$ and $Z^k = X^k + U^k$ for all k, it follows that $\bar{X} - \bar{U}$ and $\bar{X} + \bar{U}$ are accumulation points of $\{Y^k\}$ and $\{Z^k\}$, respectively. Using this fact, we can easily see that $\bar{Y} := \bar{X} - \bar{U}$ and $\bar{Z} := \bar{X} + \bar{U}$ are both in $\mathcal{X}(\bar{X})$. Since $\bar{X} = (\bar{Y} + \bar{Z})/2$ and $\bar{X}$ is an extreme point of $\mathcal{X}(\bar{X})$, it follows that $\bar{Y} = \bar{Z} = \bar{X}$, or equivalently, $\bar{U} = 0$. We have thus proved that $\{U^k\}$ converges to 0, and hence that Y^k converges to $\bar{X}$. □

The following result follows as an immediate consequence of the above lemma.

Proposition 3.3. *Let* $\bar{X}$ *be an extreme point (of the feasible region) of SDP and assume that* $\{X^k\}$ *is a sequence of feasible points for SDP converging to* $\bar{X}$. *Then there exists a sequence* $\{Y^k\}$ *of feasible points for SDP converging to* $\bar{X}$ *and satisfying* $C \bullet Y^k = C \bullet X^k$ *and* $\text{rank}\,(Y^k) \le r_{m+1}$ *for all* k.

We now are able to provide a classification of the local minima of LRSDP_r for r sufficiently large.

Theorem 3.4. *Suppose* $\bar{X}$ *is a local minimum of* LRSDP_r *for some* $r \ge r_{m+1}$. *Then,* $\bar{X}$ *is contained in the relative interior of a face* $\bar{F}$ *of SDP over which the objective function is constant. Moreover, if the dimension of* $\bar{F}$ *is zero then* $\bar{X}$ *is an optimal extreme point of SDP.*

Proof. Let $\bar{F}$ be the minimal face of SDP containing $\bar{X}$. It is well-known that $\bar{X} \in \operatorname{ri} \bar{F}$ (e.g., see Theorem 18.2 of [22]) and that

$$\bar{F} = \{X \succeq 0 : \operatorname{Range}(X) \subseteq \operatorname{Range}(\bar{X})\} \cap \{X \in \mathcal{S}^n : A_i \bullet X = b, \ i = 1, \ldots, m\}.$$

(e.g., see [3, 20]). The second fact then implies that every $X \in \bar{F}$ is feasible for LRSDP$_r$. The assumption that $\bar{X}$ is a local minimum of LRSDP$_r$ then implies that $\bar{X}$ is also a local minimum of the problem $\min\{C \bullet X : X \in \bar{F}\}$. Since $\bar{X} \in \operatorname{ri} \bar{F}$, it is easy to see that this implies that the objective function $C \bullet X$ is constant on $\bar{F}$.

If the dimension of $\bar{F}$ is zero, then $\bar{F} = \{\bar{X}\}$ and $\bar{X}$ is clearly an extreme point of SDP (see Section 2.3 of [11]). Suppose that $\bar{X}$ is not an optimal solution of SDP so that there exists a sequence $\{X^k\}$ of SDP-feasible points converging to $\bar{X}$ such that $C \bullet X^k < C \bullet \bar{X}$ for all k. Then, by Proposition 3.3 and the assumption that $r \geq r_{m+1}$, there exists a sequence $\{Y^k\}$ converging to $\bar{X}$ such that Y^k is feasible for LRSDP$_r$ and $C \bullet Y^k = C \bullet X^k < C \bullet \bar{X}$ for every k. This implies that $\bar{X}$ is not a local minimum of LRSDP$_r$, which is a contradiction. Thus, $\bar{X}$ is in fact an optimal solution of SDP. □

Roughly speaking, one can also interpret the above theorem as providing an answer to the following question: how large must we take r so that the local minima of LRSDP$_r$ are guaranteed to be global minima of SDP? The theorem asserts that we need only $r \geq r_{m+1}$ (with the important caveat that positive-dimensional faces of SDP, which are "flat" with respect to the objective function, can harbor non-global local minima). One might suspect from Theorem 1.2 that taking $r \geq r_m$ would suffice, but this is indeed not the case, in particular due to the critical role the objective function plays in Proposition 3.3. It is interesting to note that the implementation of Burer and Monteiro [5] required only $r \geq r_m$, but now with the insight provided by Theorem 3.4, our implementation in Section 6 implements $r \geq r_{m+1}$.

4. Analysis of Augmented-Lagrangian Sequences

In this section we analyze some properties of the augmented Lagrangian method in connection with problem NSDP$_r$.

For notational convenience, we define $\mathcal{A} : \mathcal{S}^n \to \Re^m$ to be the linear operator defined by $[\mathcal{A}(X)]_i = A_i \bullet X$ for all $X \in \mathcal{S}^n$ and $i = 1, \ldots, m$, so that the linear constraints of SDP can be stated compactly as $\mathcal{A}(X) = b$. It turns out that the adjoint operator $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ is given by $\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i$ for all $y \in \Re^m$, and hence the linear constraints of DSDP can be compactly written as $S \in C + \operatorname{Im} \mathcal{A}^*$.

Given sequences $\{y^k\} \subset \Re^m$ and $\{\sigma_k\} \subset \Re_{++}$, the general augmented Lagrangian approach applied to NSDP$_r$ consists of finding approximate stationary points R^k of the sequence of subproblems

$$\min_{R \in \Re^{n \times r}} \quad \mathcal{L}_k(R) := C^k \bullet RR^T + \frac{\sigma_k}{2} \|\mathcal{A}(RR^T) - b\|^2, \tag{2}$$

where $C^k := C + \mathcal{A}^* y^k$. Clearly, if we take $y^k = 0$ and allow $\sigma_k \to \infty$, then the method becomes a standard penalty method. More typically, y^k and σ_k are chosen dynamically.

Of course, one natural requirement of any variation of the method is that any accumulation point $\bar{R}$ of the sequence of approximate solutions $\{R^k\}$ is feasible for NSDP_r.

It can be easily seen that

$$\nabla \mathcal{L}_k(R^k) = 2\, S^k R^k \tag{3}$$

$$\mathcal{L}_k''(R^k)(H, H) = 2\, S^k \bullet H H^T + 4\, \sigma_k \left\| \mathcal{A}\left(R^k H^T\right) \right\|^2, \quad \forall\, H \in \Re^{n\times r}, \tag{4}$$

where

$$S^k := C^k + \sigma_k\, \mathcal{A}^*\left(\mathcal{A}\left(R^k (R^k)^T\right) - b\right) = C + \mathcal{A}^*\left(y^k + \sigma_k\left(\mathcal{A}\left(R^k (R^k)^T\right) - b\right)\right). \tag{5}$$

It is well-known that necessary conditions for R^k to be a local minimum of $\mathcal{L}_k(R)$ are that $\nabla \mathcal{L}_k(R^k) = 0$ and $\mathcal{L}_k''(R^k)(H, H) \geq 0$ for all $H \in \Re^{n\times r}$.

We now state our first result concerning sequences of points R^k arising as approximate stationary points of the sequence of subproblems (2).

Theorem 4.1. *Let $\{R^k\} \subset \Re^{n\times r}$ be a bounded sequence satisfying the following conditions:*

(a) $\lim_{k\to\infty} \mathcal{A}\left(R^k (R^k)^T\right) = b$;
(b) $\lim_{k\to\infty} \nabla \mathcal{L}_k(R^k) = 0$;
(c) $\liminf_{k\to\infty} \mathcal{L}_k''(R^k)(H^k, H^k) \geq 0$ *for all bounded sequences* $\{H^k\} \subset \Re^{n\times r}$;
(d) rank $(R^k) < r$ *for all* k.

Then the following statements hold:

(i) every accumulation point of $\{R^k (R^k)^T\}$ is an optimal solution of SDP;
(ii) the sequence $\{S^k\}$ is bounded and any of its accumulation points is an optimal dual slack for DSDP.

Proof. Let $X^k := R^k (R^k)^T$ for all k. Clearly, (3) and condition (b) together imply that

$$\lim_{k\to\infty} S^k X^k = 0. \tag{6}$$

Also, condition (d) implies that for each k there exists an orthogonal matrix $Q^k \in \Re^{r\times r}$ such that the last column of $R^k Q^k$ is zero. Now, let $h \in \Re^n$ be given and define

$$H^k := [0, \ldots, 0, h](Q^k)^T \in \Re^{n\times r}.$$

Using (4) together with the equalities $H^k (H^k)^T = h h^T$ and $R^k (H^k)^T = 0$, we conclude from condition (c) that

$$\liminf_{k\to\infty} S^k \bullet h h^T \geq 0. \tag{7}$$

We will now show that $\{S^k\}$ is bounded. Indeed, assume for contradiction that, for some subsequence $\{S^k\}_{k\in\mathcal{K}}$, we have $\lim_{k\in\mathcal{K}\to\infty} \|S^k\| = \infty$, and let $(\bar{X}, \bar{S})$ be an accumulation point of $\{(X^k, S^k/\|S^k\|)\}_{k\in\mathcal{K}}$. Using condition (a), relations (5), (6) and (7) and the fact that $\lim_{k\in\mathcal{K}\to\infty} \|S^k\| = \infty$, we easily see that $\mathcal{A}(\bar{X}) = b$, $0 \neq \bar{S} \in \text{Im}\,(\mathcal{A}^*)$,

$\bar{S} \succeq 0$, and $\bar{S} \bullet \bar{X} = 0$. It is now easy to see that these conclusions imply that $\bar{S}$ is a nontrivial direction of recession for the set of feasible dual slacks of DSDP. This violates the assumption that SDP has an interior feasible solution, however, yielding the desired contradiction. Hence $\{S^k\}$ must be bounded.

Again, using (5), (6) and (7), it is straightforward to verify (i) and the remaining part of (ii). □

Observe that if R^k is a local minimum of $\mathcal{L}_k(R)$, then the sequence $\{R^k\}$ obviously satisfies conditions (b) and (c) of Theorem 4.1. However, there is no reason for this sequence to satisfy condition (d). In the next section, we show how to obtain a sequence $\{R^k\}$ satisfying all conditions simultaneously, simply by taking R^k to be a local minimizer of a function obtained by adding an extra term to the augmented Lagrangian function $\mathcal{L}_k$.

A disadvantage of Theorem 4.1 is that the boundedness of the sequence $\{R^k\}$ must be assumed. We will now study some properties of approximate stationary points R^k for the sequence of subproblems obtained by adding the constraint $\|R\|_F^2 \leq M$ to the subproblems (2), where $M > 0$ is some large constant. This approach has the advantage that $\{R^k\}$ will be automatically bounded.

We assume that $M > 0$ is such that $I \bullet X^* < M$ for some optimal solution X^* of SDP. Then we may add the constraint $I \bullet X \leq M$ to SDP, obtaining the equivalent semidefinite programming problem

$$\begin{array}{lll} \text{SDP}' & \min & C \bullet X \\ & \text{s.t.} & \mathcal{A}(X) = b \\ & & I \bullet X \leq M \\ & & X \succeq 0, \end{array}$$

whose dual can be written in nonstandard format as

$$\begin{array}{lll} \text{DSDP}' & \max & b^T y - M\theta \\ & \text{s.t.} & \mathcal{A}^*(y) + S = C \\ & & \theta \geq 0, \quad S + \theta I \succeq 0. \end{array}$$

Note that any optimal solution of DSDP′ must have $\theta = 0$ so that S is an optimal dual slack for DSDP. Applying the low-rank change of variables $X = RR^T$ to SDP′, we obtain the nonlinear programming formulation

$$\begin{array}{lll} \text{NSDP}'_r & \min & C \bullet RR^T \\ & \text{s.t.} & \mathcal{A}(RR^T) = b \\ & & \|R\|_F^2 \leq M \end{array}$$

A partial augmented Lagrangian approach applied to this problem consists of finding approximate stationary points R^k for the sequence of subproblems

$$\begin{array}{ll} \min\limits_{R \in \Re^{n \times r}} & \mathcal{L}_k(R) \\ \text{s.t.} & \|R\|_F^2 \leq M \end{array} \tag{8}$$

A necessary condition for R^k to be a local minimum of the k-th subproblem of (8) is the existence of $\theta_k \geq 0$ such that

$$\nabla \mathcal{L}_k(R^k) + \theta_k R = 0, \quad \theta_k(M - \|R^k\|_F^2) = 0, \tag{9}$$

$$\mathcal{L}_k''(R^k)(H, H) + \theta_k I \bullet HH^T \geq 0, \quad \forall H \in \Re^{n\times r} \text{ such that } R^k \bullet H = 0. \tag{10}$$

We now state our second result regarding approximate stationary points R^k of the sequence of subproblems (8). The proof, which is an extension of the proof of Theorem 4.1, is left to the reader.

Theorem 4.2. *Let $M > 0$ be a constant large enough so that $I \bullet X^* < M$ for some optimal solution X^* of SDP. In addition, let $\{R^k\} \subset \Re^{n\times r}$ and $\{\theta_k\} \subset \Re_+$ be sequences such that $\|R^k\|_F^2 \leq M$ and which also satisfy the following conditions:*

(a) $\lim_{k\to\infty} \mathcal{A}\left(R^k(R^k)^T\right) = b$;
(b) $\lim_{k\to\infty} \nabla \mathcal{L}_k(R^k) + \theta_k R^k = 0$ *and* $\lim_{k\to\infty} \theta_k(M - \|R^k\|_F^2) = 0$;
(c) $\liminf_{k\to\infty} \mathcal{L}_k''(R^k)(H, H) + \theta_k I \bullet HH^T \geq 0$ *for all bounded sequences* $\{H^k\} \subset \Re^{n\times r}$ *such that* $R^k \bullet H^k = 0$ *for all k;*
(d) rank $(R^k) < r$ *for all k.*

Then the following statements hold:

(i) every accumulation point of $\{R^k(R^k)^T\}$ is an optimal solution of SDP;
(ii) the sequence $\{S^k\}$ defined by (5) is bounded and any of its accumulation points is an optimal dual slack for DSDP, in which case $\lim_{k\to\infty} \theta_k = 0$.

5. A Perturbed Augmented Lagrangian Algorithm

We now consider a perturbed version of the augmented Lagrangian algorithm considered in Section 4. For eack k, the method consists of finding a stationary point R^k of the following subproblem:

$$\min_{R\in\Re^{n\times r}} \quad f_k(R) := \mathcal{L}_k(R) + \mu_k \det(R^T R), \tag{11}$$

where $\mathcal{L}_k$ is the function defined in (2) and $\{\mu_k\} \subset \Re_{++}$ is a sequence converging to 0. Under mild conditions, we will show below that any accumulation point of the sequence $\{R^k(R^k)^T\}$ is an optimal solution of SDP. Our strategy will be to show that $\{R^k\}$ satisfies the conditions of Theorem 4.1.

The following two lemmas essentially show that $\{R^k\}$ satisfies condition (d) of Theorem 4.1.

Lemma 5.1. *Let $0 \neq \Delta \in \mathcal{S}^r$ be given and define $d(\delta) = \det(I_r + \delta\,\Delta)$ for all $\delta \in \Re$. Then $\delta = 0$ is not a local minimum of $d(\delta)$.*

Proof. Let $\lambda = (\lambda_j) \neq 0$ denote the vector of eigenvalues of Δ, in which case $d(\delta) = \Pi_{j=1}^r (1 + \delta\,\lambda_j)$. It is not difficult to see that

$$d'(0) = e^T \lambda$$
$$d''(0) = \left(e^T \lambda\right)^2 - e^T(\lambda^2),$$

where e is the vector of all ones and $\lambda^2 = (\lambda_j^2)$. If $d'(0) \neq 0$, then the result follows. On the other hand, if $d'(0) = 0$, then $d''(0) < 0$, showing that $\delta = 0$ is a strict local maximum, from which the result follows. □

Lemma 5.2. *Assume that $r > r_{m+1}$. If R^k is a local minimum of $f_k(R)$, then rank$(R^k) < r$.*

Proof. Suppose for contradiction that $\text{rank}(R^k) = r$, and for notational convenience let $R = R^k$. Note that $\det(R^T R) > 0$. Because the assumption on r implies that $r(r+1)/2 > m+1$, we conclude that system $\phi(R)$ with $C = C^k$ in Proposition 3.1 has a nontrivial solution Δ. For any δ such that $I + \delta\,\Delta \succ 0$, define

$$R_\delta = R\,\text{chol}(I_r + \delta\,\Delta),$$

where $\text{chol}(\cdot)$ denotes the lower Cholesky factor of $(\cdot)$. Note that R_δ is well-defined on an open interval of δ containing 0 and that $M \bullet R_\delta R_\delta^T = M \bullet RR^T$ for all $M \in \{C^k, A_1, \ldots, A_m\}$. This implies that $\mathcal{L}_k(R_\delta) = \mathcal{L}_k(R)$, and hence

$$\begin{aligned} f_k(R) - f_k(R_\delta) &= \mu_k \left(\det(R^T R) - \det(R_\delta^T R_\delta)\right) \\ &= \mu_k \det(R^T R)\,(1 - \det(I_r + \delta\,\Delta))\,, \end{aligned}$$

where the second equality follows from standard properties of the determinant. By Lemma 5.1, $\delta = 0$ is not a local minimum of $\det(I_r + \delta\,\Delta)$, i.e., there exists arbitrarily small $\delta \neq 0$ such that $\det(I_r + \delta\,\Delta) < 1$, which when combined with the above equality and the fact that $\mu_k \det(R^T R) > 0$ imply that R is not a local minimum of $f_k(R)$. Since this contradicts the definition of R, we must have rank $(R) < r$. □

We remark that the main point of Lemma 5.2 can also be achieved by analyzing the behavior of $\det(R^T R)^{1/r}$. The key observation is that $\det(\cdot)^{1/r}$ is a concave function over the set of $r \times r$ positive definite matrices and is actually strictly concave over line segments between linearly independent matrices (see section 7.8 of Horn and Johnson [12]).

Theorem 5.3. *Assume that $r > r_{m+1}$ and that $\{\mu_k\} \subset \Re_{++}$ is a sequence converging to 0. For each k, let R^k be a local minimum of $f_k(R)$ and let S^k be given by (5). Moreover, assume that:*

(a) $\lim_{k\to\infty} \mathcal{A}\left(R^k (R^k)^T\right) = b$;
(b) the sequence $\{R^k\} \subset \Re^{n\times r}$ is bounded.

Then the following statements hold:

(i) every accumulation point of $\{R^k (R^k)^T\}$ is an optimal solution of SDP;
(ii) the sequence $\{S^k\}$ defined by (5) is bounded and any of its accumulation points is an optimal dual slack for DSDP.

Proof. The result follows immediately by verifying that $\{R^k\}$ satisfies conditions (b) to (d) of Theorem 4.1. Condition (d) of Theorem 4.1 follows from Lemma 5.2. To verify

(b) and (c) of Theorem 4.1, define $d(R) = \det(R^T R)$ for all $R \in \Re^{n\times r}$. Since R^k is a local minimum of $f_k(R)$, we must have

$$\nabla f_k(R^k) = \nabla \mathcal{L}_k(R^k) + \mu_k \nabla d(R^k) \; = \; 0,$$
$$f_k''(R^k)(H, H) = \mathcal{L}_k''(R^k)(H, H) + \mu_k \, d''(R^k)(H, H) \; \geq \; 0, \;\; \forall \, H \in \Re^{n\times r}.$$

Since $\{\mu_k\}$ converges to 0 and the derivatives of d are uniformly bounded over compact sets, it follows that $\lim_{k\to\infty} \nabla \mathcal{L}_k(R^k) = 0$ and $\liminf_{k\to\infty} \mathcal{L}_k''(R^k)(H^k, H^k) \geq 0$ for all bounded sequences $\{H^k\} \subset \Re^{n\times r}$, showing that $\{R^k\}$ also satisfies conditions (b) and (c) of Theorem 4.1. □

Similar to Theorem 4.1, one drawback of the above theorem is that the boundedness of $\{R^k\}$ must be assumed, and similar to Theorem 4.2, the next theorem addresses this issue by considering the sequence of stationary points $\{R^k\}$ of the sequence of subproblems

$$\min_{R\in\Re^{n\times r}} \quad f_k(R)$$
$$\text{s.t.} \quad \|R\|_F^2 \leq M,$$

which automatically enforce that the sequence $\{R^k\}$ is bounded. Its proof, which is based on Theorem 4.2, is quite similar to the one of Theorem 5.3.

Theorem 5.4. *Let $M > 0$ be a constant large enough so that $I \bullet X^* < M$ for some optimal solution X^* of SDP. Assume that $r > r_{m+2}$ and that $\{\mu_k\} \subset \Re_{++}$ is a sequence converging to* 0*. For each k, let R^k be a local minimum of the subproblem* $\min\{f_k(R) : \|R\|_F^2 \leq M\}$ *and let S^k be given by (5). Then, the following statements hold:*

(i) if $\lim_{k\to\infty} \mathcal{A}(R^k(R^k)^T) = b$ then any accumulation point of $\{R^k(R^k)^T\}$ is an optimal solution of SDP, the sequence $\{S^k\}$ is bounded and any accumulation point of $\{S^k\}$ is an optimal dual slack of SDP;

(ii) if $\lim_{k\to\infty} \sigma_k = \infty$ and the sequences $\{y^k\}$ and $\{S^k\}$ are bounded then $\lim_{k\to\infty} \mathcal{A}(R^k(R^k)^T) = b$.

Proof. To prove (i), assume that $\lim_{k\to\infty} \mathcal{A}(R^k(R^k)^T) = b$. Let $\theta_k \in \Re_+$ denote the Lagrange multiplier corresponding to the constraint $\|R\|_F^2 \leq M$ of the k-th subproblem. Using the fact that (R^k, θ_k) satisfies $\lim_{k\in\to\infty} \mathcal{A}(R^k(R^k)^T) = b$ and relations (9) and (10), it is possible to show that the sequences $\{R^k\}$ and $\{\theta_k\}$ satisfy all the conditions of Theorem 4.2, from which (i) immediately follows. (We remark that a variation of Lemma 5.2 is needed in order to guarantee that rank $(R^k) < r$. In this variation, it is necessary to assume that $r > r_{m+2}$, or equivalently that $r(r+1)/2 > m+2$, which allows the matrix Δ in the proof of Lemma 5.2 to be chosen so as to ensure that $\|R_\delta\|_F^2 = I \bullet R_\delta R_\delta^T$ is a constant function of δ.)

We now prove (ii). Using (5), the assumption that $\{y^k\}$ and $\{S^k\}$ are bounded and $\mathcal{A}^*$ is one-to-one, we easily see that $\{\sigma_k(\mathcal{A}(R^k(R^k)^T) - b)\}$ is bounded. Since $\lim_{k\to\infty} \sigma_k = \infty$, this implies that $\lim_{k\to\infty} \mathcal{A}(R^k(R^k)^T) = b$. □

Observe that Theorem 5.4 establishes, under the assumption that $\lim_{k\to\infty} \sigma_k = \infty$ and $\{y^k\}$ is bounded, that the condition $\lim_{k\to\infty} \mathcal{A}(R^k(R^k)^T) = b$ is equivalent to the boundedness of $\{S^k\}$. Unfortunately, we do not know whether one of these two conditions will always hold, even though they are always observed in our practical experiments.

6. Computational Results

The algorithm of the previous section, whose convergence is proven in Theorems 5.3 and 5.4, differs only slightly from the practical algorithm of [5] in that the extra term $\mu_k \det(R^T R)$ is added to the augmented Lagrangian function. While it seems that the extra term is necessary for theoretical convergence, it does not appear to be necessary for practical convergence. Indeed, the practical convergence observed in [5] has served as the main motivation for the theoretical investigations of the current paper. Informally, one can also see that the theoretical and practical versions are not extremely different since one may theoretically choose $\mu_k > 0$ as small as one wishes, with the only requirement being that $\mu_k \to 0$.

Another reason for favoring the practical algorithm is the difficulty of calculating the derivative of $d(R) = \det(R^T R)$, which in particular would need to be calculated for any R such that $\operatorname{rank}(R) < r$, or equivalently, when $R^T R$ is singular. It is not difficult to see that

$$\nabla d(R) = R \operatorname{cofactor}(R^T R),$$

where $\operatorname{cofactor}(R^T R)$ denotes the matrix of cofactors of $(R^T R)_{ij}$ in $R^T R$. The authors are not aware of any quick, numerically stable way of calculating $\operatorname{cofactor}(R^T R)$. For these reasons, the numerical results that we present are based on the same algorithm as introduced in [5].

These things being said, however, it is reasonable to expect the practical algorithm to deliver a certificate of optimality, at least asymptotically. Letting $\{R^k\}$ and $\{S^k\}$ be the sequences generated by the algorithm, the relevant measurements are

$$\left(\sum_{i=1}^{m}(A_i \bullet R^k(R^k)^T - b_i)^2\right)^{1/2}, \quad \|S^k R^k\|_F, \quad \lambda_{\min}(S^k),$$

which monitor primal feasibility, complementarity (which also corresponds to the norm of the gradient of the augmented Lagrangian function), and dual feasibility, respectively. In expectation that each of these quantities will go to zero during the execution of the algorithm, we implement the following strategy. Given parameters $\rho_f, \rho_c > 0$:

- the k-th subproblem is terminated with R^k and S^k once

$$\frac{\|S^k R^k\|_F}{\|C\|_F + 1} < \frac{\rho_c}{\sigma_k};$$

- the entire algorithm is terminated with $\bar{R} = R^k$ and $\bar{S} = S^k$ once R^k is obtained such that

$$\frac{\left(\sum_{i=1}^{m}(A_i \bullet R^k(R^k)^T - b_i)^2\right)^{1/2}}{\|b\| + 1} < \rho_f.$$

On all test problems, these termination criteria were realized (see below). In addition, although we cannot exercise as much control over $\lambda_{\min}(S^k)$, we have found that $\lambda_{\min}(\bar{S})$ is typically slightly negative, which matches the theoretical prediction of Section 4.

For other implementation details regarding the augmented Lagrangian algorithm, we refer the reader to [5]. We mention briefly that the penalty parameter σ is updated by a fairly conservative factor of $\sqrt{10}$ every so often, typically after the solution of each third or fourth subproblem.

In the following two subsections, we demonstrate the performance of the low-rank algorithm on three classes of SDP relaxations of combinatorial optimization problems. We remark that a common feature of the three classes of problems we solve is that the constraints $A_i \bullet X = b_i$, $i = 1, \ldots, m$, impose an upper bound on the trace of X and hence a bound on the norm of any feasible R. Hence, in accordance with Theorem 5.3, we can expect the sequences generated by the algorithm to be bounded.

The implementation of the low-rank algorithm was written in ANSI C, and all computational results were performed on a Pentium 2.4 GHz having 1 Gb of RAM.

6.1. Maximum cut and maximum stable set relaxations

We consider ten test problems which were used in [5]; see [5] for a careful description. In particular, we have chosen five of the largest maximum cut SDP relaxations and five of the largest maximum stable set SDP relaxations, whose results are shown in Table 1. The parameters chosen for the test runs were $\rho_f = 10^{-5}$ for primal feasibility and $\rho_c = 10^{-1}$ for complementarity. The first three columns of Table 1 give basic problem information; the fourth gives the final objective value achieved by the algorithm; the fifth gives a lower bound on the optimal value of SDP; the sixth gives the minimum eigenvalue of the final dual matrix; and the last gives the total time required in seconds.

The lower bounds given in Table 1 were computed by perturbing the final dual matrix $\bar{S}$ in order to achieve dual feasibility and then reporting the corresponding dual objective value. In particular, both the maximum cut and maximum stable set SDPs share the property that the identity matrix I can be written as a known linear combination of the matrices $A_1, \ldots, A_m$, which makes it straightforward to perturb $\bar{S}$ as long as $\lambda_{\min}(\bar{S})$ is available. The minimum eigenvalue of $\bar{S}$ was computed with the Lanczos-based package LASO available from the Netlib Repository.

The computational results demonstrate that the low-rank algorithm with the described parameters is able to solve the the maximum cut problems to several digits of accuracy

Table 1. Results of the low-rank algorithm on five maximum-cut and five maximum-stable-set SDP relaxations (see [5]). Parameters are $\rho_f = 10^{-5}$ and $\rho_c = 10^{-1}$, and lower bounds are calculated by shifting $\bar{S}$ to dual feasibility. Times are given in seconds.

problem	n	m	$C \bullet \bar{R}\bar{R}^T$	lower bd	$\lambda_{\min}(\bar{S})$	time
G67	10000	10000	-7.744e+03	-7.745e+03	-1.8e−04	595
G70	10000	10000	-9.861e+03	-9.863e+03	-1.4e−04	517
G72	10000	10000	-7.808e+03	-7.809e+03	-4.7e−05	787
G77	14000	14000	-1.104e+04	-1.105e+04	-1.6e−04	865
G81	20000	20000	-1.565e+04	-1.567e+04	-6.7e−04	2433
G43	5000	9991	-2.806e+02	-2.833e+02	-2.7e+00	1709
G51	3000	6001	-3.490e+02	-3.503e+02	-1.3e+00	3265
brock400-4.co	400	20078	-3.970e+01	-4.066e+01	-9.7e−01	768
c-fat200-1.co	200	18367	-1.200e+01	-1.229e+01	-2.9e−01	260
p-hat300-1.co	300	33918	-1.007e+01	-1.199e+01	-1.9e+00	4948

in a small amount of time. In particular, approximate primal and dual optimal solutions are produced by the algorithm as indicated by the achieved feasibility tolerance ρ_f, the small minimum eigenvalues of $\bar{S}$, and the associated duality gap.

The results for the maximum stable set relaxations do not appear as strong, however, since the minimum eigenvalues and lower bounds are not quite as accurate. Upon further investigation, we found that by tightening the complementarity parameter ρ_c to values such as 10^{-2} or 10^{-3}, we could significantly improve these metrics, but a fair amount of additional computation time was required. Moreover, the primal matrix $\bar{R}$ improved only incrementally under these scenarios. (These observations are demonstrated in Table 2.) Hence, with regard to the maximum stable set SDP, the results of Table 1 present a balance between good progress in the primal with the time required to achieve good progress in the dual.

6.2. *Quadratic assignment relaxations*

The results of the previous subsection highlight a capability of the low-rank algorithm — namely that it can be used to obtain lower bounds on the optimal value of SDP whenever I is in the subspace generated by $A_1, \ldots, A_m$ or, equivalently, when the constraints of SDP imply a constant trace over all feasible X. This class of SDPs includes the relaxations of many combinatorial optimization problems (e.g., maximum cut and maximum stable set) and has been studied extensively in [9]. In such cases, since the optimal value of the SDP relaxation is itself a lower bound on the optimal value of the underlying combinatorial problem, the low-rank algorithm can be used as a tool to obtain bounds for combinatorial optimization problems also.

Given a general 0-1 quadratic program, its standard SDP relaxation does not satisfy the condition of the previous paragraph, i.e., I is not in the subspace generated by $A_1, \ldots, A_m$. There is, however, a simple, easily computable scaling PA_iP^T of the matrices A_i such that I is generated by $PA_1P^T, \ldots, PA_mP^T$ (see [23, 8]). Hence, this scaling can be used in conjunction with the low-rank algorithm to compute lower bounds on the optimal value of 0-1 quadratic programs.

The quadratic assignment problem (QAP) is a 0-1 quadratic program arising in location theory that has proven to be extremely difficult to solve to optimality, due in no small part to its large size even for moderate numbers of decision variables. In particular, a QAP with ℓ facilities and ℓ locations yields a quadratic program with ℓ^2 binary variables and 2ℓ linear constraints. In terms of optimizing QAP using an implicit enumeration scheme such as branch-and-bound, a key ingredient in any such scheme is the bounding

Table 2. Results of the low-rank algorithm on the five maximum-stable-set SDP relaxations of Table 1. Here, parameters are $\rho_f = 10^{-5}$ and $\rho_c = 10^{-3}$.

problem	$C \bullet \bar{R}\bar{R}^T$	lower bd	$\lambda_{\min}(\bar{S})$	time
G43	-2.806e+02	-2.817e+02	-1.1e+00	12440
G51	-3.490e+02	-3.491e+02	-1.3e−01	34972
brock400-4.co	-3.970e+01	-4.066e+01	-9.6e−01	3476
c-fat200-1.co	-1.200e+01	-1.200e+01	-2.1e−03	977
p-hat300-1.co	-1.007e+01	-1.025e+01	-1.8e−01	15514

technique used to obtain lower bounds on the optimal value of QAP, and for this, many bounds based on convex optimization have been proposed, including ones based on linear programming, convex quadratic programming, and semidefinite programming. A recent survey on progress made towards solving QAP is given by Anstreicher [2].

SDP relaxations of QAP have been studied in [14, 21, 27] and are most notable for the fact that, even though the quality of bounds is usually quite good, the huge size of the SDPs makes the calculation of these bounds very difficult. In [14, 27], four successively larger SDP relaxations are introduced, and generally speaking, the bound is improved as the size of the relaxation is increased. Table 3 gives basic information on the size of these relaxations in terms of the number ℓ of facilities and locations; we refer the reader to [14, 27] for a full description.

Lin and Saigal [14] give computational results on solving the relaxation QAP_{R_0} of Table 3 for several problems of size up to $\ell = 30$. Likewise, Zhao et al. [27] investigate QAP_{R_1} and QAP_{R_2} for problems up to size $\ell = 30$ and QAP_{R_3} for problems up to size $\ell = 22$ with at most 2,000 linear inequalities. Most recently, Rendl and Sotirov [21] have used the bundle method to compute bounds provided by QAP_{R_2} and QAP_{R_3} (with all inequality constraints included) for instances up to $\ell = 30$.

For the algorithm of this paper, we provide computational results for computing bounds provided by QAP_{R_1} and QAP_{R_2} for instances of size up to $\ell = 40$. In particular, we do not include any problems with $\ell < 30$ since we wish to concentrate on problems of larger size. Also, we do not test QAP_{R_3} for two primary reasons. First, it is not clear at this moment the best way to incorporate linear inequality constraints into the low-rank algorithm. Second, since it makes sense to solve QAP_{R_3} with only a few important inequalities and since choosing such inequalities is itself a difficult task, we would like instead to study the performance of the low-rank algorithm on the well-defined problem classes QAP_{R_1} and QAP_{R_2}.

Our test problems come from QAPLIB [7], and we have selected a representative sample of all problems in QAPLIB with $30 \le \ell \le 40$. The results of the problems are shown in Table 4. The feasibility and centrality parameters are taken to be $\rho_f = 10^{-3}$ and $\rho_c = 10^2$, respectively. In contrast with Table 1, we do not report any information concerning the primal objective value or the minimum eigenvalue of $\bar{S}$, since primal and dual solutions of high accuracy are not necessarily of interest here. Instead, we wish to demonstrate that reasonably good bounds for QAP can be computed using the low-rank algorithm. To judge the quality of the bounds, we also include the objective value of the best known integer feasible solution of QAP as well. In particular, those problems for which the best known integer feasible value is also optimal are indicated by a prefixed asterisk ($*$). We remark that, if the reader is further interested in the quality of the bounds, the papers [2, 21, 27] discuss such issues in detail.

Table 3. Size comparison of four SDP relaxations of QAP. Here, ℓ is the basic dimension of the QAP; n gives the size of the semidefinite matrix; and m gives the number of equality constraints.

	n	m	linear inequalities
QAP_{R_0}	$\ell^2 + 1$	$\ell^2 + 3$	0
QAP_{R_1}	$(\ell - 1)^2 - 1$	$2\ell^2 + \ell + 1$	0
QAP_{R_2}	$(\ell - 1)^2 - 1$	$\ell^3 - 2\ell^2 + 1$	0
QAP_{R_3}	$(\ell - 1)^2 - 1$	$\ell^3 - 2\ell^2 + 1$	$\le \frac{1}{2}\ell^4 - \ell^3 + \frac{5}{2}\ell^2 + 1$

A few comments regarding the results presented in Table 4 are in order. First of all, the low-rank algorithm was able to successfully solve all instances to the desired accuracy, delivering bounds of roughly the same quality as documented in other investigations of SDP bounds for QAP; see [21, 27].

In terms of computation times, it is clear that the low-rank algorithm can take a significant amount of time on some problems (for example, the maximum time was approximately 6.4 days for ste36b). However, we stress that these times, although large in some cases, compare very favorably to other investigations. Moreover, to our knowledge, no computational results for SDP relaxations having $\ell > 30$ have been reported in the literature. As an example, Rendl and Sotirov [21] report that their bundle method requires approximately 10 hours to deliver a bound of 5651 on nug30 via QAP_{R_2} on an Athlon XP running at 1.8 GHz. As shown in Table 4, we were able to achieve a comparable bound of 5629 in approximately 36 minutes.

In addition, the computational results demonstrate that solving QAP_{R_2} requires much more time than QAP_{R_1}. Moreover, it seems difficult to predict an expected increase of time between QAP_{R_1} and QAP_{R_2}, as the factors of increase range from a low of 4.7 for esc32a to a high of 74.1 for ste36b. For classes of problems for which the bound does not improve dramatically from QAP_{R_1} to QAP_{R_2}, it thus may be reasonable to solve only QAP_{R_1}.

Finally, Table 4 illustrates a phenomenon that many authors have recognized in working with QAP, namely that problems of similar size have varying degrees of difficulty. In other words, the data of the QAP can greatly affect the difficulty of the instance. This is evidenced in the table, for example, by lipa30a and tho30. Although each is of the same size, tho30 takes about 4 times longer to solve for QAP_{R_1} and about 36 times longer to solve for QAP_{R_2}.

Table 4. Results of the low-rank algorithm for QAP_{R_1} and QAP_{R_2} on seventeen problems from QAPLIB; subscripts indicate the relevant relaxation. Parameters are $\rho_f = 10^{-3}$ and $\rho_c = 10^2$, and lower bounds are rounded up to nearest integer due to integral data for underlying QAP. Times are in seconds.

problem	feasible val	$n_{\{1,2\}}$	m_1	m_2	lower bd$_1$	lower bd$_2$	time$_1$	time$_2$
esc32a	130	960	2081	30721	−326	−144	103	480
esc32h	438	960	2081	30721	176	225	111	527
kra30a	*88900	840	1831	25201	69509	78255	3274	58359
kra30b	*91420	840	1831	25201	70096	79165	2602	48846
kra32	*88700	960	2081	30721	65605	76669	2894	58103
lipa30a	*13178	840	1831	25201	12765	12934	439	2294
lipa30b	*151426	840	1831	25201	151133	151357	582	14862
lipa40a	*31538	1520	3241	60801	30575	30560	889	8753
lipa40b	*476581	1520	3241	60801	474875	476417	4747	93621
nug30	*6124	840	1831	25201	5311	5629	359	2161
ste36a	*9526	1224	2629	44065	−9452	7156	2963	25703
ste36b	*15852	1224	2629	44065	−115816	10350	7464	552860
tai30a	1818146	840	1831	25201	1528834	1577013	3216	72911
tai35a	2422002	1155	2486	40426	1970071	2029376	6775	155143
tai40a	3139370	1520	3241	60801	2519257	2592756	11938	421348
tho30	*149936	840	1831	25201	125846	135535	1921	81454
tho40	240516	1520	3241	60801	199680	214593	7384	219336

Acknowledgements. The authors would like to thank Kurt Anstreicher for many helpful discussions regarding the quadratic assignment problem and Gábor Pataki for his insightful comments on the paper. The authors are also in debt to two anonymous referees, whose comments and suggestions greatly improved the paper.

References

1. Alizadeh, F., Haeberly, J.-P.A., Overton. M.L.: Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. SIAM Journal on Optimization, **8**, 746–768 (1998)
2. K.M. Anstreicher. Recent advances in the solution of quadratic assignment problems. Mathematical Programming (Series B), **97** (1-2), 27–42 (2003)
3. Barker, G.P., Carlson, D.: Cones of diagonally dominant matrices. Pacific Journal of Mathematics, **57** (1), 15–32 (1975)
4. Burer, S.: Semidefinite programming in the space of partial positive semidefinite matrices. SIAM Journal on Optimization, **14** (1), 139–172 (2003)
5. Burer, S., Monteiro, R.D.C.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical Programming (Series B), **95**, 329–357 (2003)
6. Burer, S., Monteiro, R.D.C., Zhang. Y.: Solving a class of semidefinite programs via nonlinear programming. Mathematical Programming, **93**, 97–122 (2002)
7. Burkard, R.E., Karisch, S., Rendl, F.: QAPLIB — a quadratic assignment problem library. European Journal of Operational Research, **55**,115–119 (1991)
8. Helmberg, C.: Semidefinite programming for combinatorial optimization. ZIB-Report 00-34, Konrad-Zuse-Zentrum für Informationstechnik Berlin, October 2000.
9. Helmberg, C., Rendl, F.: A spectral bundle method for semidefinite programming. SIAM Journal on Optimization, **10**, 673–696 (2000)
10. Helmberg, C., Rendl, F., Vanderbei, R.J., Wolkowicz, H.: An interior-point method for semidefinite programming. SIAM Journal on Optimization, **6**, 342–361 (1996)
11. Hiriart-Urruty, J.-B., Lemaréchal, C., Convex analysis and minimization algorithms. I, volume 305 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1993.
12. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, New York, 1985.
13. Kojima, M., Shindoh, S., Hara, S.: Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. SIAM Journal on Optimization, **7**, 86–125 (1997)
14. Lin, C-J., Saigal, R.: On solving large scale semidefinite programming problems: a case study of quadratic assigment problem. Technical Report, Dept. of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2177, 1997.
15. Monteiro, R.D.C.: First- and second-order methods for semidefinite programming. Mathematical Programming, **97**, 209–244 (2003)
16. Monteiro, R.D.C.: Primal-dual path following algorithms for semidefinite programming. SIAM Journal on Optimization, **7**, 663–678 (1997)
17. Monteiro, R.D.C., Zhang, Y.: A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. Mathematical Programming, **81**, 281–299 (1998)
18. Nakata, K., Fujisawa, K., Kojima, M.: Using the conjugate gradient method in interior-points for semidefinite programs. Proceedings of the Institute of Statistical Mathematics, **46**, 297–316 (1998) In Japanese.
19. Pataki, G.: On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. Mathematics of Operations Research, **23**, 339–358 (1998)
20. Pataki, G.: The geometry of semidefinite programming. In H. Wolkowicz, R. Saigal, L. Vandenberghe, (eds.), Handbook of Semidefinite Programming: Theory, Algorithms, and Applications. Kluwer Academic Publishers, 2000.
21. Rendl, F., Sotirov, R.: Bounds for the quadratic assignment problem using the bundle method. Manuscript, University of Klagenfurt, August 2003.
22. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ, 1970.
23. De Simone, C.: The cut polytope and the boolean quadric polytope. Discrete Mathematics, **79**, 71–75 (1989)
24. Toh, K.C.: Solving large scale semidefinite programs via an iterative solver on the agumented systems. SIAM Journal on Optimization, **14**, 670–698 (2004)
25. Toh, K.C., Kojima, M.: Solving some large scale semidefinite programs via the conjugate residual method. SIAM Journal on Optimization, **12**, 669–691 (2002)
26. Zhang, Y.: On extending some primal–dual interior–point algorithms from linear programming to semidefinite programming. SIAM Journal on Optimization, **8**, 365–386 (1998)
27. Zhao, Q., Karisch, S.E., Rendl, F., Wolkowicz, H.: Semidefinite programming relaxations for the quadratic assignment problem. Journal of Combinatorial Optimization, **2**, 71–109 (1998)