Penalty Decomposition Methods for Rank Minimization *

Zhaosong Lu, Yong Zhang and Xiaorui Li

Department of Mathematics, Simon Fraser University, Canada
{zhaosong,yza30,xla97}@sfu.ca

October 5, 2013 (Revised: April 1, 2014)

Abstract

In this paper we consider general rank minimization problems with rank appearing either in the objective function or as a constraint. We first establish that a class of special rank minimization problems has closed-form solutions. Using this result, we then propose penalty decomposition methods for general rank minimization problems in which each subproblem is solved by a block coordinate descent method. Under some suitable assumptions, we show that any accumulation point of the sequence generated by the penalty decomposition methods satisfies the first-order optimality conditions of a nonlinear reformulation of the problems. Finally, we test the performance of our methods by applying them to the matrix completion and nearest low-rank correlation matrix problems. The computational results demonstrate that our methods are generally comparable or superior to the existing methods in terms of solution quality.

Key words: rank minimization, penalty decomposition methods, matrix completion, nearest low-rank correlation matrix

AMS 2000 subject classification: 15A99, 15A83, 90C99

1 Introduction

In this paper we consider the following rank minimization problems:

$$\min_X \{f(X) : \ g(X) \le 0, \ h(X) = 0, \ \mathrm{rank}(X) \le r, \ X \in \mathcal{X} \cap \Omega\}, \tag{1}$$

$$\min_X \{f(X) + \nu \ \mathrm{rank}(X) : g(X) \le 0, \ h(X) = 0, \ X \in \mathcal{X} \cap \Omega\} \tag{2}$$

for some integer $r \ge 0$ and $\nu \ge 0$ controlling the rank of the solutions, where $\mathcal{X}$ is a closed convex set, Ω is a closed unitarily invariant convex set in $\Re^{m\times n}$, and $f : \Re^{m\times n} \to \Re$, $g : \Re^{m\times n} \to \Re^p$ and $h : \Re^{m\times n} \to \Re^t$ are continuously differentiable functions (for the definition of unitarily invariant set, see Section 2). In the literature, there are numerous application problems in the form of (1) or (2). For example, several well-known combinatorial optimization problems such as maximal cut (MAXCUT) and maximal stable set can be formulated as problem (1) (see, for example, [14, 1, 5]). More generally, nonconvex quadratic

*This work was supported in part by NSERC Discovery Grant.

programming problems can also be cast into (2) (see, for example, [1]). Recently, some image recovery problems are formulated as (1) or (2) (see, for example, [36]). In addition, the problem of finding nearest low-rank correlation matrix is in the form of (1), which has important application in finance (see, for example, [4, 38, 46, 47, 33, 39, 15]).

Several approaches have recently been developed for solving problems (1) and (2) or their special cases. In particular, for those arising in combinatorial optimization (e.g., MAXCUT), one novel approach is to first solve a semidefinite programming (SDP) relaxation of (1) and then obtain an approximate solution of (1) by applying some heuristics to the solution of the SDP (see, for example, [14]). Despite the remarkable success on those problems, it is not clear about the performance of this method when extended to solve general problem (1). In addition, the nuclear norm relaxation approach has been proposed for problems (1) or (2). For example, Fazel et al. [10] considered a special case of problem (2) with $f \equiv 0$ and $\Omega = \Re^{m \times n}$. In their approach, a convex relaxation is applied to (1) or (2) by replacing the rank of X by the nuclear norm of X and numerous efficient methods can then be applied to solve the resulting convex problems. Recently, Recht et al. [36] showed that under some conditions, such a convex relaxation is tight for the case where $\mathcal{X}$ is an affine manifold. The quality of such a relaxation, however, remains unknown when applied to general problems (1) and (2). Additionally, in some applications, the nuclear norm stays constant in the feasible region. As an example, for the nearest low-rank correlation matrix problem (see Subsection 6.2), any feasible point is a symmetric positive semidefinite matrix with all diagonal entries equal to one and its nuclear norm is a constant. For those problems, nuclear norm relaxation approach is clearly inappropriate. In a recent work [13], Gao and Sun proposed a majorized penalty approach to solving this rank-constrained nearest correlation matrix problem. In their approach, the constraint $\mathrm{rank}(X) \le r$ for an $n \times n$ symmetric positive semidefinite matrix X is equivalently reformulated as $\sum_{i=r+1}^n \lambda_i(X) = 0$, where $\lambda_i(X)$ is the ith largest eigenvalue of X, and a majorization method is then applied to solve the penalty problem resulted from penalizing the constraint $\sum_{i=r+1}^n \lambda_i(X) = 0$. Another nonconvex relaxation approach was proposed to solve a special class of rank minimization problems arising in matrix completion and sensor network (see, for example, [31, 11, 18, 20, 27]), in which $\mathrm{rank}(X)$ is approximated by the Schatten p-quasi-norm of X for some $p \in (0, 1)$. In addition, Meka et al. [30] proposed a singular value projection method for solving rank constrained least squares problems arising in matrix completion. A nonlinear programming (NLP) reformulation approach has also been applied to a class of rank-constrained optimization problems arising in SDP and matrix completion (see, for example, [5, 45]). For this approach, the rank-constrained optimization problems are cast into NLP problems by replacing the constraint $\mathrm{rank}(X) \le r$ by $X = UV$ where $U \in \Re^{m \times r}$ and $V \in \Re^{r \times n}$, and some classical nonlinear optimization methods are then applied to solve the resulting NLPs. For general problem (1), the resulting NLP by this approach is usually highly nonlinear, which might be challenging for the existing numerical optimization methods. Moreover, it seems that this approach is not applicable to problem (2).

In this paper we consider general rank minimization problems (1) and (2). We first establish that a class of special rank minimization problems has closed-form solutions. Using this result, we then propose penalty decomposition methods for general rank minimization problems in which each subproblem is solved by a block coordinate descent method. Under some suitable assumptions, we show that any accumulation point of the sequence generated by the penalty decomposition methods satisfies the first-order optimality conditions of a nonlinear reformulation of the problems. Finally, we test the performance of our methods by applying them to the matrix completion and nearest low-rank correlation matrix problems. The computational results demonstrate that our methods are generally comparable or superior to the existing methods in terms of solution quality.

The rest of this paper is organized as follows. In Subsection 1.1, we introduce the notation that is used throughout the paper. In Section 2, we study a class of special rank minimization problems. We develop penalty decomposition (PD) methods for general rank minimization problems in Sections 3, 4 and 5 and establish some convergence results for them. In Section 6, we conduct numerical experiments to test the performance of our PD methods for solving matrix completion and nearest low-rank correlation matrix problems. Finally, we present some concluding remarks in Section 7.

1.1 Notation

In this paper, the symbol $\Re^n$ denotes the n-dimensional Euclidean space, and the set of all $m \times n$ matrices with real entries is denoted by $\Re^{m\times n}$. The spaces of $n \times n$ diagonal and symmetric matrices will be denoted by $\mathcal{D}^n$ and $\mathcal{S}^n$, respectively. If $X \in \mathcal{S}^n$ is positive semidefinite, we write $X \succeq 0$. The cone of positive semidefinite matrices is denoted by $\mathcal{S}^n_+$. Given matrices X and Y in $\Re^{m\times n}$, the standard inner product is defined by $\langle X, Y\rangle := \mathrm{Tr}(XY^T)$, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. The Frobenius norm of a real matrix X is defined as $\|X\|_F := \sqrt{\mathrm{Tr}(XX^T)}$, and the nuclear norm of X, denoted by $\|X\|_*$, is defined as the sum of all singular values of X. The rank of a matrix X is denoted by $\mathrm{rank}(X)$. We denote the identity matrix and the all-ones matrix by I and E, respectively, whose dimension should be clear from the context. For a real symmetric matrix X, $\lambda(X)$ denotes the vector of all eigenvalues of X arranged in nondecreasing order and $\Lambda(X)$ is the diagonal matrix whose ith diagonal entry is $\lambda_i(X)$ for all i. Similarly, for any $X \in \Re^{m\times n}$, $\sigma(X)$ denotes the q-dimensional vector consisting of all singular values of X arranged in nondecreasing order, where $q = \min(m, n)$, and $\Sigma(X)$ is the $m \times n$ matrix whose ith diagonal entry is $\sigma_i(X)$ for all i and all off-diagonal entries are 0, that is, $\Sigma_{ii}(X) = \sigma_i(X)$ for $1 \le i \le q$ and $\Sigma_{ij}(X) = 0$ for all $i \ne j$. We define the operator $\mathscr{D} : \Re^q \to \Re^{m\times n}$ as follows:

$$\mathscr{D}_{ij}(x) = \begin{cases} x_i & \text{if } i = j; \\ 0 & \text{otherwise} \end{cases} \qquad \forall x \in \Re^q,$$

where $q = \min(m, n)$. Given an $n \times n$ matrix X, $\widetilde{\mathscr{D}}(X)$ denotes a diagonal matrix whose ith diagonal element is X_{ii} for $i = 1, \ldots, n$. Given a vector $x \in \Re^n$, the nonnegative part of x is denoted by $x^+ = \max(x, 0)$, where the maximization operates entry-wise, and $\|x\|_0$, $\|x\|_1$ and $\|x\|_2$ denote the cardinality (i.e., the number of nonzero entries), the standard 1-norm and the Euclidean norm of x, respectively. Given a real vector space $\mathcal{U}$ and a closed convex set $C \subseteq \mathcal{U}$, $\mathcal{N}_C(x)$ and $\mathcal{T}_C(x)$ denote the normal and tangent cones of C at any $x \in C$, respectively, and $\mathcal{P}_C(\cdot)$ denotes the standard projection map.

2 A class of special rank minimization problems

In this section we first show that a class of matrix optimization problems can be solved as lower dimensional vector optimization problems. As a consequence, we establish that a class of special rank minimization problems has closed-form solutions, which will be used to develop penalty decomposition methods in Sections 3, 4 and 5.

Before proceeding, we introduce some definitions that will be used subsequently. Let $\mathcal{U}^n$ denote the set of all unitary matrices in $\Re^{n\times n}$.

Definition 1 *A norm $\|\cdot\|$ is a unitarily invariant norm on $\Re^{m\times n}$ if $\|UXV\| = \|X\|$ for all $U \in \mathcal{U}^m$, $V \in \mathcal{U}^n$, $X \in \Re^{m\times n}$.*

Definition 2 *A function $F : \Re^{m\times n} \to \Re$ is a unitarily invariant function if $F(UXV) = F(X)$ for all $U \in \mathcal{U}^m$, $V \in \mathcal{U}^n$, $X \in \Re^{m\times n}$.*

Definition 3 *A set* $\mathcal{X} \subseteq \Re^{m\times n}$ *is a unitarily invariant set if*

$$\{UXV : U \in \mathcal{U}^m, V \in \mathcal{U}^n, X \in \mathcal{X}\} = \mathcal{X}.$$

Definition 4 *A function* $F : \mathcal{S}^n \to \Re$ *is a unitary similarity invariant function if* $F(UXU^T) = F(X)$ *for all* $U \in \mathcal{U}^n$, $X \in \mathcal{S}^n$.

Definition 5 *A set* $\mathcal{X} \subseteq \mathcal{S}^n$ *is a unitary similarity invariant set if*

$$\{UXU^T : U \in \mathcal{U}^n, X \in \mathcal{X}\} = \mathcal{X}.$$

The following result establishes that a class of matrix optimization problems over a subset of $\Re^{m\times n}$ can be solved as lower dimensional vector optimization problems. We make use of the fact that diagonal matrices are sufficient to characterize useful properties of unitarily invariant functions. This fact has been used to derive the gradient formula for spectral functions in [22].

Proposition 2.1 *Let* $\|\cdot\|$ *be a unitarily invariant norm on* $\Re^{m\times n}$, *and let* $F : \Re^{m\times n} \to \Re$ *be a unitarily invariant function. Suppose that* $\mathcal{X} \subseteq \Re^{m\times n}$ *is a unitarily invariant set. Let* $A \in \Re^{m\times n}$ *be given,* $q = \min(m, n)$, *and let* ϕ *be a non-decreasing function on* $[0, \infty)$. *Suppose that* $U\Sigma(A)V^T$ *is the singular value decomposition of* A. *Then,* $X^* = U\mathscr{D}(x^*)V^T$ *is an optimal solution of the problem*

$$\begin{array}{ll} \min & F(X) + \phi(\|X - A\|) \\ \text{s.t.} & X \in \mathcal{X}, \end{array} \tag{3}$$

where $x^* \in \Re^q$ *is an optimal solution of the problem*

$$\begin{array}{ll} \min & F(\mathscr{D}(x)) + \phi(\|\mathscr{D}(x) - \Sigma(A)\|) \\ \text{s.t.} & \mathscr{D}(x) \in \mathcal{X}. \end{array} \tag{4}$$

Proof. Since $\|\cdot\|$ is a unitarily invariant norm, we know from Theorem 7.4.51 on page 448 of [17] that

$$\|X - A\| \geq \|\Sigma(X) - \Sigma(A)\| \quad \forall X \in \Re^{m\times n}. \tag{5}$$

It then follows from (5), the monotonicity of ϕ and the relation $\Sigma(X) = \mathscr{D}(\sigma(X))$ that

$$\phi(\|X - A\|) \geq \phi(\|\mathscr{D}(\sigma(X)) - \Sigma(A)\|) \quad \forall X \in \Re^{m\times n}.$$

Since $\mathcal{X}$ is a unitarily invariant set and F is a unitarily invariant function, we have

$$\mathscr{D}(\sigma(X)) \in \mathcal{X}, \quad F(\mathscr{D}(\sigma(X))) = F(X) \quad \forall X \in \mathcal{X}.$$

Using the above relations, we immediately obtain that

$$F(X) + \phi(\|X - A\|) \geq F(\mathscr{D}(\sigma(X))) + \phi(\|\mathscr{D}(\sigma(X)) - \Sigma(A)\|) \quad \forall X \in \mathcal{X},$$

which together with $\mathscr{D}(\sigma(X)) \in \mathcal{X}$ implies that the optimal value of problem (3) is minorized by that of problem (4). Further, by the definitions of x^* and X^*, we know that $\mathscr{D}(x^*) \in \mathcal{X}$, which along with the assumption that $\mathcal{X}$ is a unitarily invariant set, implies that $X^* \in \mathcal{X}$, that is, X^* is a feasible solution of problem (3). Moreover, we observe that

$$F(X^*) = F(\mathscr{D}(x^*)), \quad \phi(\|X^* - A\|) = \phi(\|U\mathscr{D}(x^*)V^T - A\|) = \phi(\|\mathscr{D}(x^*) - \Sigma(A)\|).$$

Thus, the objective function of (3) reaches the optimal value of problem (4) at X^*. It then immediately follows that problems (4) and (3) share the same optimal value, and hence X^* is an optimal solution of (3). ■

As some consequences of Proposition 2.1, we next state that a class of rank minimization problems on a subset of $\Re^{m\times n}$ can be solved as lower dimensional vector minimization problems.

Corollary 2.2 *Let $\nu \geq 0$ and $A \in \Re^{m\times n}$ be given, and let $q = \min(m,n)$. Suppose that $\mathcal{X} \subseteq \Re^{m\times n}$ is a unitarily invariant set, and $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem*

$$\min\{\nu \;\mathrm{rank}(X) + \frac{1}{2}\|X - A\|_F^2 : \; X \in \mathcal{X}\}, \tag{6}$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\nu\|x\|_0 + \frac{1}{2}\|x - \sigma(A)\|_2^2 : \; \mathscr{D}(x) \in \mathcal{X}\}. \tag{7}$$

Proof. Let $\|\cdot\| := \|\cdot\|_F$, $F(X) := \nu \;\mathrm{rank}(X)$ for all X, and $\phi(t) := t^2/2$ for all t. Clearly, the assumptions of Proposition 2.1 are satisfied for such $\|\cdot\|$, F and ϕ. Further, notice that $\mathrm{rank}(\mathscr{D}(x)) = \|x\|_0$ and $\|\mathscr{D}(x) - \Sigma(A)\|_F = \|x - \sigma(A)\|_2$ for all $x \in \Re^q$. It immediately follows from Proposition 2.1 that the conclusion holds. ■

Corollary 2.3 *Let $r \geq 0$ and $A \in \Re^{m\times n}$ be given, and let $q = \min(m,n)$. Suppose that $\mathcal{X} \subseteq \Re^{m\times n}$ is a unitarily invariant set, and $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem*

$$\min\{\|X - A\|_F : \; \mathrm{rank}(X) \leq r, \; X \in \mathcal{X}\}, \tag{8}$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\|x - \sigma(A)\|_2 : \; \|x\|_0 \leq r, \; \mathscr{D}(x) \in \mathcal{X}\}. \tag{9}$$

Proof. Its proof is similar to that of Corollary 2.2. ■

Remark. When $\mathcal{X}$ is simple enough, problems (6) and (8) have closed form solutions. For example, when $\mathcal{X} = \{X \in \Re^{m\times n} : a \leq \sigma_i(X) \leq b \;\forall i\}$ for some $0 \leq a < b \leq \infty$, one can see that $\mathscr{D}(x) \in \mathcal{X}$ if and only if $a \leq |x_i| \leq b$ for all i. It is not hard to observe that the associated problems (7) and (9) have closed-form solutions, and hence problems (6) and (8) also have closed-form solutions.

The following results are crucially used to develop efficient algorithms for solving the nuclear norm relaxation of the matrix completion problems (see, for example, [6, 29, 44]). They can be immediately obtained from Proposition 2.1.

Corollary 2.4 *Let $\nu \geq 0$ and $A \in \Re^{m\times n}$ be given, and let $q = \min(m,n)$. Suppose that $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem*

$$\min \nu\|X\|_* + \frac{1}{2}\|X - A\|_F^2,$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min \nu\|x\|_1 + \frac{1}{2}\|x - \sigma(A)\|_2^2.$$

Corollary 2.5 *Let $r \geq 0$ and $A \in \Re^{m\times n}$ be given, and let $q = \min(m,n)$. Suppose that $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem*

$$\min\{\|X - A\|_F : \ \|X\|_* \leq r\},$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\|x - \sigma(A)\|_2 : \ \|x\|_1 \leq r\}.$$

We next show that a class of matrix optimization problems over a subset of $\mathcal{S}^n$ can also be solved as lower dimensional vector optimization problems.

Proposition 2.6 *Let $\|\cdot\|$ be a unitarily invariant norm on $\Re^{n\times n}$, and let $F : \mathcal{S}^n \to \Re$ be a unitary similarity invariant function. Suppose that $\mathcal{X} \subseteq \mathcal{S}^n$ is a unitary similarity invariant set. Let $A \in \mathcal{S}^n$ be given, and let ϕ be a non-decreasing function on $[0,\infty)$. Suppose that $U\Lambda(A)U^T$ is the eigenvalue value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)U^T$ is an optimal solution of the problem*

$$\begin{array}{ll} \min & F(X) + \phi(\|X - A\|) \\ \text{s.t.} & X \in \mathcal{X}, \end{array}$$

where $x^* \in \Re^n$ is an optimal solution of the problem

$$\begin{array}{ll} \min & F(\mathscr{D}(x)) + \phi(\|\mathscr{D}(x) - \Lambda(A)\|) \\ \text{s.t.} & \mathscr{D}(x) \in \mathcal{X}. \end{array}$$

Proof. The conclusion of this proposition follows from the Ky Fan's inequality (e.g., see [2] (IV.62))

$$\|X - Y\| \geq \|\Lambda(X) - \Lambda(Y)\| \quad \forall X, Y \in \mathcal{S}^n,$$

and a similar argument as used in the proof of Proposition 2.1. ∎

As some consequences of Proposition 2.6, we next show that a class of rank minimization problems on a subset of $\mathcal{S}^n$ can be solved as lower dimensional vector minimization problems.

Corollary 2.7 *Let $\nu \geq 0$ and $A \in \mathcal{S}^n$ be given. Suppose that $\mathcal{X} \subseteq \mathcal{S}^n$ is a unitary similarity invariant set, and $U\Lambda(A)U^T$ is the eigenvalue decomposition of A. Then, $X^* = U\mathscr{D}(x^*)U^T$ is an optimal solution of the problem*

$$\min\{\nu \ \mathrm{rank}(X) + \frac{1}{2}\|X - A\|_F^2 : \ X \in \mathcal{X}\}, \tag{10}$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\nu\|x\|_0 + \frac{1}{2}\|x - \lambda(A)\|_2^2 : \ \mathscr{D}(x) \in \mathcal{X}\}. \tag{11}$$

Proof. The conclusion of this corollary immediately follows from Proposition 2.6, and the relations $\mathrm{rank}(\mathscr{D}(x)) = \|x\|_0$ and $\|\mathscr{D}(x) - \Lambda(A)\|_F = \|x - \lambda(A)\|$ for all $x \in \Re^n$. ∎

Corollary 2.8 *Let $r \geq 0$ and $A \in \mathcal{S}^n$ be given. Suppose that $\mathcal{X} \subseteq \mathcal{S}^n$ is a unitary similarity invariant set, and $U\Lambda(A)U^T$ is the eigenvalue decomposition of A. Then, $X^* = U\mathscr{D}(x^*)U^T$ is an optimal solution of the problem*

$$\min\{\|X - A\|_F : \ \mathrm{rank}(X) \leq r, \ X \in \mathcal{X}\}, \tag{12}$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\|x - \lambda(A)\|_2 : \ \|x\|_0 \leq r, \ \mathscr{D}(x) \in \mathcal{X}\}. \tag{13}$$

Proof. Its proof is similar to that of Corollary 2.7. ∎

Remark. When $\mathcal{X}$ is simple enough, problems (10) and (12) have closed form solutions. For example, when $\mathcal{X} = \{X \in \mathcal{S}^n : a \leq \lambda_i(X) \leq b \ \forall i\}$ for some $a < b \leq \infty$, one can see that $\mathscr{D}(x) \in \mathcal{X}$ if and only if $a \leq x_i \leq b$ for all i. It is not hard to observe that the associated problems (11) and (13) have closed-form solutions, and hence problems (10) and (12) also have closed-form solutions.

3 Penalty decomposition method for rank minimization of asymmetric matrices

In this section, we consider general rank minimization problems (1) and (2) by assuming that $\mathcal{X}$ and Ω are some subsets in $\Re^{m\times n}$. In particular, we first study the first-order optimality conditions for (1) and (2) in Subsection 3.1. We then propose penalty decomposition (PD) methods for solving (1) and (2) and establish their convergence in Subsections 3.2 and 3.3, respectively.

Throughout this section, we make the following assumption for problems (1) and (2).

Assumption 1 *Problems (1) and (2) are feasible, and moreover, at least a feasible solution, denoted by X^{feas}, is known.*

This assumption will be used to design the PD methods with nice convergence properties. It can be dropped, but the theoretical convergence of the corresponding PD methods may be weakened. We shall mention that, for numerous real applications, X^{feas} is readily available or can be observed from the physical background of problems. For example, all application problems discussed in Section 6 have a trivial feasible solution. On the other hand, for some problems which do not have a trivial feasible solution, one can always approximate them by the problems which have a trivial feasible solution. For instance, problem (1) can be approximately solved as the following problem:

$$\min_{X\in\mathcal{X}\cap\Omega}\{f(X) + \rho(\|u^+\|_2^2 + \|v\|_2^2) : g(X) - u \leq 0, \ h(X) - v = 0, \ \mathrm{rank}(X) \leq r\}$$

for some large ρ. The latter problem has a trivial feasible solution when $\mathcal{X}$ and Ω are sufficiently simple.

3.1 First-order optimality conditions

In this subsection, we establish the first-order optimality conditions for problems (1) and (2).

Theorem 3.1 *Suppose that X^* is a local minimizer of problem (1). Let $U^* \in \Re^{m\times r}$, $V^* \in \Re^{r\times n}$ be such that $(U^*)^T U^* = I$ and $X^* = U^* V^*$. Assume that the following Robinson condition holds:*

$$\left\{ \left(\begin{array}{c} g'(X^*)d_X - v \\ h'(X^*)d_X \\ d_X - d_U V^* - U^* d_V \\ d_Y - d_U V^* - U^* d_V \end{array} \right) : \begin{array}{l} d_X \in \mathcal{T}_{\mathcal{X}}(X^*), v \in \Re^p, v_i \le 0, i \in \mathcal{A}(X^*), \\ d_U \in \Re^{m\times r}, d_V \in \Re^{r\times n}, d_Y \in \mathcal{T}_{\Omega}(X^*) \end{array} \right\} = \Re^p \times \Re^t \times \Re^{m\times n} \times \Re^{m\times n}, \tag{14}$$

where $g'(X^*)$ and $h'(X^*)$ denote the Jacobian of the functions $g = (g_1, \ldots, g_p)$ and $h = (h_1, \ldots, h_t)$ at X^*, respectively, and

$$\mathcal{A}(X^*) = \{1 \le i \le p : g_i(X^*) = 0\}.$$

Then, there exist $\lambda^* \in \Re^p_+$, $\mu^* \in \Re^t$, $Z^*_X \in \Re^{m\times n}$, $Z^*_Y \in \Re^{m\times n}$ such that

$$\begin{array}{c} -\nabla f(X^*) - \nabla g(X^*)\lambda^* - \nabla h(X^*)\mu^* - Z^*_X \in \mathcal{N}_{\mathcal{X}}(X^*), \\ (Z^*_X - Z^*_Y)(V^*)^T = 0, (U^*)^T(Z^*_X - Z^*_Y) = 0, \\ \lambda^*_i \ge 0,\ \lambda^*_i g_i(X^*) = 0,\ i = 1, \ldots, p; \quad Z^*_Y \in \mathcal{N}_{\Omega}(X^*). \end{array} \tag{15}$$

Proof. Let $Y^* = X^*$. Since X^* is a local minimizer of problem (1), one can observe that (X^*, Y^*, U^*, V^*) is a local minimizer of

$$\min_{X,Y,U,V} \{f(X) :\ g(X) \le 0,\ h(X) = 0,\ X - UV = 0,\ Y - UV = 0,\ X \in \mathcal{X},\ Y \in \Omega,\ U \in \Re^{m\times r},\ V \in \Re^{r\times n}\}.$$

Using this observation, (14) and Theorem 3.25 of [40], we see that the conclusion holds. ∎

Theorem 3.2 *Suppose that X^* is a local minimizer of problem (2). Let $r = \text{rank}(X^*)$, $U^* \in \Re^{m\times r}$, $V^* \in \Re^{r\times n}$ be such that $(U^*)^T U^* = I$ and $X^* = U^* V^*$. Assume that the Robinson condition (14) holds at (X^*, U^*, V^*). Then, there exist $\lambda^* \in \Re^p_+$, $\mu^* \in \Re^t$, $Z^*_X \in \Re^{m\times n}$, $Z^*_Y \in \Re^{m\times n}$ such that (15) holds.*

Proof. By the assumption that X^* is a local minimizer of problem (2), we can observe that X^* is also a local minimizer of problem (1) for $r = \text{rank}(X^*)$. The conclusion of this theorem then immediately follows from Theorem 3.1. ∎

3.2 Penalty decomposition method for problem (1)

In this subsection, we propose a PD method for solving (1) and establish its convergence.

Clearly, problem (1) can be equivalently reformulated as

$$\min_{X,Y} \{f(X) :\ g(X) \le 0,\ h(X) = 0,\ X - Y = 0,\ X \in \mathcal{X},\ Y \in \mathcal{Y}\}, \tag{16}$$

where

$$\mathcal{Y} := \{Y \in \Omega |\ \text{rank}(Y) \le r\}. \tag{17}$$

Given a penalty parameter $\varrho > 0$, the associated quadratic penalty function for (16) is defined as

$$Q_\varrho(X, Y) := f(X) + \frac{\varrho}{2}\left(\|[g(X)]^+\|_2^2 + \|h(X)\|_2^2 + \|X - Y\|_F^2\right). \tag{18}$$

We now propose a PD method for solving problem (16) (or equivalently, (1)) in which each penalty subproblem is approximately solved by a block coordinate descent (BCD) method. BCD method is a simple but widely used method for solving numerous large-scale optimization problems. The properties and convergence of BCD for some classes of optimization problems have been well studied in literature (see, for example, [43]).

Algorithm 3.1 Penalty decomposition method for (16) (asymmetric matrices):

Let $\{\epsilon_k\}$ *be a positive decreasing sequence. Let* $\varrho_0 > 0$, $c > 1$ *be given. Choose an arbitrary* $Y_0^0 \in \mathcal{Y}$ *and a constant* $\Upsilon \geq \max\{f(X^{\rm feas}), \min_{X\in\mathcal{X}} Q_{\varrho_0}(X, Y_0^0)\}$*. Set* $k = 0$.

1) *Set* $l = 0$ *and apply the BCD method to find an approximate solution* $(X^k, Y^k) \in \mathcal{X} \times \mathcal{Y}$ *to the penalty subproblem*

$$\min\{Q_{\varrho_k}(X, Y) : \ X \in \mathcal{X}, \ Y \in \mathcal{Y}\} \tag{19}$$

by performing steps 1a)-1d):

1a) *Solve* $X_{l+1}^k \in \operatorname{Arg}\min_{X\in\mathcal{X}} Q_{\varrho_k}(X, Y_l^k)$.

1b) *Solve* $Y_{l+1}^k \in \operatorname{Arg}\min_{Y\in\mathcal{Y}} Q_{\varrho_k}(X_{l+1}^k, Y)$.

1c) *Set* $(X^k, Y^k) := (X_{l+1}^k, Y_{l+1}^k)$*. If* (X^k, Y^k) *satisfies*

$$\|\mathcal{P}_{\mathcal{X}}(X^k - \nabla_X Q_{\varrho_k}(X^k, Y^k)) - X^k\|_F \leq \epsilon_k, \tag{20}$$

then go to step 2).

1d) *Set* $l \leftarrow l + 1$ *and go to step 1a).*

2) *Set* $\varrho_{k+1} := c\varrho_k$.

3) *If* $\min_{X\in\mathcal{X}} Q_{\varrho_{k+1}}(X, Y^k) > \Upsilon$*, set* $Y_0^{k+1} := X^{\rm feas}$*. Otherwise, set* $Y_0^{k+1} := Y^k$.

4) *Set* $k \leftarrow k + 1$ *and go to step 1).*

end

Remark. The condition (20) will be used to establish the global convergence of Algorithm 3.1. It may not be easily verifiable unless $\mathcal{X}$ is simple. On the other hand, we observe that the sequence $\{Q_{\varrho_k}(X_l^k, Y_l^k)\}$ is non-increasing for any fixed k. In practice, it is thus reasonable to terminate the BCD method based on the progress of $\{Q_{\varrho_k}(X_l^k, Y_l^k)\}$. That is, one can terminate the BCD method if

$$\frac{|Q_{\varrho_k}(X_l^k, Y_l^k) - Q_{\varrho_k}(X_{l-1}^k, Y_{l-1}^k)|}{\max(|Q_{\varrho_k}(X_l^k, Y_l^k)|, 1)} \ \leq \ \epsilon_I \tag{21}$$

for some $\epsilon_I > 0$. Another practical termination criterion for the BCD method is based on the relative change of the sequence $\{(X_l^k, Y_l^k)\}$. Similarly, we can terminate the outer iterations of the PD method once

$$\max_{ij} |X_{ij}^k - Y_{ij}^k| \ \leq \ \epsilon_O \tag{22}$$

for some $\epsilon_O > 0$. Given that problem (19) is nonconvex, the BCD method may converge to a local stationary point. To enhance its practical performance, one may execute it multiple times by restarting from a suitable perturbation of the current best approximate solution. For example, at the kth outer iteration, let (X^k, Y^k) be the current best approximate solution of (19) found by the BCD method, and let $r_k = \text{rank}(Y^k)$. Assume that $r_k > 1$. Before starting the $(k + 1)$th outer iteration, one can re-apply the BCD method starting from $Y_0^k \in \operatorname{Arg}\min_{Y\in\Omega}\{\|Y - Y^k\|_F : \text{rank}(Y) \leq r_k - 1\}$ (namely, a rank-one perturbation of Y^k) and obtain a new approximate solution $(\tilde{X}^k, \tilde{Y}^k)$ of (19). If $Q_{\varrho_k}(\tilde{X}^k, \tilde{Y}^k)$ is "sufficiently" smaller than $Q_{\varrho_k}(X^k, Y^k)$, one can set $(X^k, Y^k) := (\tilde{X}^k, \tilde{Y}^k)$ and repeat the above process. Otherwise, one can terminate the kth outer iteration and start the next outer iteration. Finally, in view of Corollary 2.3, the subproblem in step 1b) can be reduced to a problem in the form of (9), which has a closed-form solution when Ω is simple enough.

The following theorem establishes the convergence of the outer iterations of Algorithm 3.1 for solving problem (1). In particular, we show that under some suitable assumptions, any accumulation point of the sequence generated by Algorithm 3.1 satisfies the first-order optimality conditions (15).

Theorem 3.3 *Let $\{(X^k, Y^k)\}$ be the sequence generated by Algorithm 3.1 satisfying (20), and let $U^k \in \Re^{m\times r}$ and $V^k \in \Re^{r\times n}$ be such that*

$$(U^k)^T U^k = I, \quad Y^k = U^k V^k. \tag{23}$$

Assume that $\epsilon_k \to 0$. Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \leq \Upsilon\}$ is compact. Then, the following statements hold:

(a) The sequence $\{(X^k, Y^k, U^k, V^k)\}$ is bounded;

(b) Suppose that a subsequence $\{(X^k, Y^k, U^k, V^k)\}_{k\in K}$ converges to (X^*, Y^*, U^*, V^*). Then, X^* is a feasible point of problem (1), and moreover, $X^* = Y^* = U^*V^*$. In addition, assume that the Robinson condition (14) *holds at (X^*, U^*, V^*) and*

$$\left\{d_Y - d_U V^k - U^k d_V : d_U \in \Re^{m\times r}, d_V \in \Re^{r\times n}, d_Y \in \mathcal{T}_\Omega(Y^k)\right\} = \Re^{m\times n} \tag{24}$$

holds for sufficiently large $k \in K$. Then, $\{(\lambda^k, \mu^k, Z_X^k)\}_{k\in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(X^k)]^+, \qquad \mu^k = \varrho_k h(X^k), \qquad Z_X^k = \varrho_k (X^k - Y^k), \tag{25}$$

and each accumulation point $(\lambda^*, \mu^*, Z_X^*)$ of $\{(\lambda^k, \mu^k, Z_X^k)\}_{k\in K}$ together with (X^*, U^*, V^*) and some $Z_Y^* \in \Re^{m\times n}$ satisfies the first-order optimality conditions (15).

Proof. In view of (18) and our choice of Y_0^k that is specified in step 3), one can observe that

$$f(X^k) + \frac{\varrho_k}{2}(\|[g(X^k)]^+\|_2^2 + \|h(X^k)\|_2^2 + \|X^k - Y^k\|_F^2) = Q_{\varrho_k}(X^k, Y^k) \leq \min_{X\in\mathcal{X}} Q_{\varrho_k}(X, Y_0^k) \leq \Upsilon \quad \forall k. \tag{26}$$

It immediately implies that $\{X^k\} \subseteq \mathcal{X}_\Upsilon$, and hence $\{X^k\}$ is bounded due to the compactness of $\mathcal{X}_\Upsilon$. Moreover, we can obtain from (26) that

$$\|X^k - Y^k\|_F^2 \leq 2[\Upsilon - f(X^k)]/\varrho_k \leq 2[\Upsilon - \min_{X\in\mathcal{X}_\Upsilon} f(X)]/\varrho_0, \tag{27}$$

which together with the boundedness of $\{X^k\}$ yields that $\{Y^k\}$ is bounded. In addition, it follows from (23) that $\{U^k\}$ is bounded and $V^k = (U^k)^T Y^k$, which implies that $\{V^k\}$ is also bounded. Thus, statement (a) holds.

We next show that statement (b) also holds. Suppose that $\{(X^k, Y^k, U^k, V^k)\}_{k\in K}$ converges to (X^*, Y^*, U^*, V^*). Notice that $\varrho_k \to \infty$ as $k \to \infty$. Upon taking limits on both sides of (27) as $k \in K \to \infty$, we have $X^* - Y^* = 0$. In addition, it is not hard to show that $\mathcal{Y}$ is closed. We then see that $X^* \in \mathcal{X}$ and $Y^* \in \mathcal{Y}$ due to the closedness of $\mathcal{X}$ and $\mathcal{Y}$. It thus follows that (X^*, Y^*) is a feasible point of problem (16) and X^* is a feasible point of (1). The equality $X^* = U^*V^*$ immediately follows from the identities $Y^k = U^k V^k$ and $Y^* = X^*$. Now, let us prove the second part of statement (b). Indeed, let S^k be the matrix such that

$$\mathcal{P}_{\mathcal{X}}(X^k - \nabla_X Q_{\varrho_k}(X^k, Y^k)) = X^k + S^k.$$

It then follows from (20) that $\|S^k\|_F \le \epsilon_k$ for all k, which together with $\lim_{k\to\infty} \epsilon_k = 0$ implies $\lim_{k\to\infty} S^k = 0$. By a well-known property of the projection map $\mathcal{P}_{\mathcal{X}}$, we have

$$\langle X - X^k - S^k, X^k - \nabla_X Q_{\varrho_k}(X^k, Y^k) - X^k - S^k \rangle \ \le \ 0, \ \forall X \in \mathcal{X}.$$

Hence, we obtain that

$$-\nabla_X Q_{\varrho_k}(X^k, Y^k) - S^k \in \mathcal{N}_{\mathcal{X}}(X^k + S^k). \tag{28}$$

Using this relation, (28), (25) and the definition of Q_ϱ, we have

$$-\nabla f(X^k) - \nabla g(X^k)\lambda^k - \nabla h(X^k)\mu^k - Z_X^k - S^k \in \mathcal{N}_{\mathcal{X}}(X^k + S^k). \tag{29}$$

By the definitions of X^k and Y^k, one can see that

$$Y^k \in \operatorname{Arg}\min_{Y\in\mathcal{Y}} Q_{\varrho_k}(X^k, Y).$$

Using this relation and the definitions of U^k, V^k and $\mathcal{Y}$, we can observe that

$$(Y^k, U^k, V^k) \in \operatorname{Arg}\min\{Q_{\varrho_k}(X^k, UV): \ Y - UV = 0, Y \in \Omega, U \in \Re^{m\times r}, V \in \Re^{r\times n}\}.$$

It follows from (24) and Theorem 3.34 of [40] that for sufficiently large $k \in K$, there exists $Z_Y^k \in \Re^{m\times n}$ such that

$$(Z_Y^k - Z_X^k)(V^k)^T = 0, \quad (U^k)^T(Z_Y^k - Z_X^k) = 0, \quad Z_Y^k \in \mathcal{N}_\Omega(Y^k), \tag{30}$$

where Z_X^k is defined in (25). We claim that $\{(\lambda^k, \mu^k, Z_X^k, Z_Y^k)\}_{k\in K}$ is bounded. Suppose not, by passing to a subsequence if necessary, we can assume that $\{(\lambda^k, \mu^k, Z_X^k, Z_Y^k)\}_{k\in K} \to \infty$. Let

$$(\bar\lambda^k, \bar\mu^k, \bar Z_X^k, \bar Z_Y^k) = (\lambda^k, \mu^k, Z_X^k, Z_Y^k)/\|(\lambda^k, \mu^k, Z_X^k, Z_Y^k)\| \quad \forall k.$$

Without loss of generality, assume that $\{(\bar\lambda^k, \bar\mu^k, \bar Z_X^k, \bar Z_Y^k)\}_{k\in K} \to (\bar\lambda, \bar\mu, \bar Z_X, \bar Z_Y)$ (otherwise, one can consider its convergent subsequence). Clearly, $\|(\bar\lambda, \bar\mu, \bar Z_X, \bar Z_Y)\| = 1$. Recall that $\{(X^k, Y^k, U^k, V^k)\}_{k\in K} \to (X^*, Y^*, U^*, V^*)$ and $Y^* = X^*$. Dividing both sides of (29) and (30) by $\|(\lambda^k, \mu^k, Z_X^k, Z_Y^k)\|$, taking limits as $k \in K \to \infty$, and using the relation $Z_Y^k \in \mathcal{N}_\Omega(Y^k)$ and the semicontinuity of $\mathcal{N}_{\mathcal{X}}(\cdot)$ and $\mathcal{N}_\Omega(\cdot)$ (see Lemma 2.42 of [40]), we obtain that

$$\begin{array}{c} -\nabla g(X^*)\bar\lambda - \nabla h(X^*)\bar\mu - \bar Z_X \in \mathcal{N}_{\mathcal{X}}(X^*), \\ (\bar Z_Y - \bar Z_X)(V^*)^T = 0, \quad (U^*)^T(\bar Z_Y - \bar Z_X) = 0, \quad \bar Z_Y \in \mathcal{N}_\Omega(X^*). \end{array} \tag{31}$$

We can see from (25) that $\bar\lambda \in \Re^m_+$ and $\bar\lambda_i = 0$ for $i \notin \mathcal{A}(X^*)$. By (14), there exist $d_X \in \mathcal{T}_{\mathcal{X}}(X^*)$, $d_Y \in \mathcal{T}_\Omega(X^*)$, $d_U \in \Re^{m\times r}$, $d_V \in \Re^{r\times n}$, $v \in \Re^p$ with $v_i \le 0$ for $i \in \mathcal{A}(X^*)$ such that

$$\begin{array}{rcl} -\bar\lambda &=& g'(X^*)d_X - v, \\ -\bar\mu &=& h'(X^*)d_X, \\ -\bar Z_X &=& d_X - d_U V^* - U^* d_V, \\ -\bar Z_Y &=& d_U V^* + U^* d_V - d_Y. \end{array}$$

Recall that $\bar\lambda \in \Re^m_+$, $\bar\lambda_i = 0$ for $i \notin \mathcal{A}(X^*)$, and $v_i \le 0$ for $i \in \mathcal{A}(X^*)$. Hence, $v^T\bar\lambda \le 0$. Using these relations, (31) and the fact that $d_X \in \mathcal{T}_{\mathcal{X}}(X^*)$ and $d_Y \in \mathcal{T}_\Omega(X^*)$, we have

$$\begin{aligned}\|(\bar\lambda,\bar\mu)\|_2^2 + \|(\bar Z_X,\bar Z_Y)\|_F^2 &= -\langle\bar\lambda,-\bar\lambda\rangle - \langle\bar\mu,-\bar\mu\rangle - \langle\bar Z_X,-\bar Z_X\rangle - \langle\bar Z_Y,-\bar Z_Y\rangle,\\ &= -\langle\bar\lambda, g'(X^*)d_X - v\rangle - \langle\bar\mu, h'(X^*)d_X\rangle - \langle\bar Z_X, d_X - d_U V^* - U^* d_V\rangle - \langle\bar Z_Y, d_U V^* + U^* d_V - d_Y\rangle,\\ &= \langle -\nabla g(X^*)\bar\lambda - \nabla h(X^*)\bar\mu - \bar Z_X, d_X\rangle + \langle\bar\lambda, v\rangle + \langle\bar Z_X - \bar Z_Y, d_U V^* + U^* d_V\rangle + \langle\bar Z_Y, d_Y\rangle \le 0,\end{aligned}$$

which implies that $\|(\bar\lambda,\bar\mu,\bar Z_X,\bar Z_Y)\| = 0$, which contradicts the identity $\|(\bar\lambda,\bar\mu,\bar Z_X,\bar Z_Y)\| = 1$. Thus, $\{(\lambda^k,\mu^k,Z_X^k,Z_Y^k)\}_{k\in K}$ is bounded. Now let $(\lambda^*,\mu^*,Z_X^*,Z_Y^*)$ be an accumulation point of $\{(\lambda^k,\mu^k,Z_X^k,Z_Y^k)\}_{k\in K}$. By passing to a subsequence if necessary, we can assume that $\{(\lambda^k,\mu^k,Z_X^k,Z_Y^k)\}_{k\in K} \to (\lambda^*,\mu^*,Z_X^*,Z_Y^*)$. Recall that $\{(X^k,U^k,V^k)\}_{k\in K} \to (X^*,U^*,V^*)$. Taking limits on both sides of (29) and (30) as $k \in K \to \infty$, and using the semicontinuity of $\mathcal{N}_{\mathcal{X}}(\cdot)$, we immediately see that the first three relations of (15) hold. In addition, we see from (25) that $\lambda_i^k \ge 0$ and $\lambda_i^k g_i(X^k) = 0$ for all i, which immediately implies that $\lambda_i^* \ge 0$, $\lambda_i^* g_i(X^*) = 0$ for $i = 1,\ldots,p$. Finally, using the semicontinuity of $\mathcal{N}_\Omega(\cdot)$, $Z_Y^k \in \mathcal{N}_\Omega(Y^k)$ and $\{(Z_Y^k,Y^k)\}_{k\in K} \to (Z_Y^*,X^*)$, we conclude that $Z_Y^* \in \mathcal{N}_\Omega(X^*)$. ∎

Before ending this subsection, we next establish a convergence result regarding the inner iterations of Algorithm 3.1. In particular, we will show that an approximate solution $(X^k,Y^k) \in \mathcal{X}\times\mathcal{Y}$ of problem (19) satisfying (20) can be found by the BCD method described in steps 1a)-1d). For convenience of presentation, we omit the index k from (19) and consider the BCD method for solving the following problem

$$\min_{X,Y}\{Q_\varrho(X,Y): \ X \in \mathcal{X}, \ Y \in \mathcal{Y}\}. \tag{32}$$

instead. Accordingly, we rename the iterates of the BCD method stated in Algorithm 3.1 and present it as follows.

Algorithm 3.2 Block coordinate descent method for (32):

Choose an arbitrary initial point $Y_0 \in \mathcal{Y}$. Set $l = 0$.

1) *Solve* $X_{l+1} \in \mathrm{Arg}\min_{X\in\mathcal{X}} Q_\varrho(X,Y_l)$.

2) *Solve* $Y_{l+1} \in \mathrm{Arg}\min_{Y\in\mathcal{Y}} Q_\varrho(X_{l+1},Y)$.

3) *Set $l \leftarrow l+1$ and go to step 1).*

end

Theorem 3.4 *Let $\{(X_l,Y_l)\} \subseteq \mathcal{X}\times\mathcal{Y}$ be generated by Algorithm 3.2, and let $\epsilon > 0$ be given. Suppose that $\{(X_l,Y_l)\}$ has at least an accumulation point. Then, there exists some $l > 0$ such that*

$$\|\mathcal{P}_{\mathcal{X}}(X_l - \nabla_X Q_\varrho(X_l,Y_l)) - X_l\|_F < \epsilon.$$

Proof. We observe from the first two steps of Algorithm 3.2 that

$$\begin{aligned} Q_\varrho(X_{l+1},Y_l) &\le Q_\varrho(X,Y_l) \quad \forall X \in \mathcal{X},\\ Q_\varrho(X_l,Y_l) &\le Q_\varrho(X_l,Y) \quad \forall Y \in \mathcal{Y}.\end{aligned} \tag{33}$$

It follows that

$$Q_\varrho(X_{l+1},Y_{l+1}) \le Q_\varrho(X_{l+1},Y_l) \le Q_\varrho(X_l,Y_l) \quad \forall l \ge 1. \tag{34}$$

Hence, the sequence $\{Q_\varrho(X_l, Y_l)\}$ is non-increasing. By the assumption, $\{(X_l, Y_l)\}$ has at least an accumulation point, denoted by (X^*, Y^*), and there exists a subsequence L such that $\lim_{l\in L\to\infty}(X_l, Y_l) = (X^*, Y^*)$. We then observe that $\{Q_\varrho(X_l, Y_l)\}_{l\in L}$ is bounded, which together with the monotonicity of $\{Q_\varrho(X_l, Y_l)\}$ implies that $\{Q_\varrho(X_l, Y_l)\}$ is bounded below and hence $\lim_{l\to\infty} Q_\varrho(X_l, Y_l)$ exists. This observation, (34) and the continuity of $Q_\varrho(\cdot,\cdot)$ yield

$$\lim_{l\to\infty} Q_\varrho(X_{l+1}, Y_l) = \lim_{l\to\infty} Q_\varrho(X_l, Y_l) = \lim_{l\in L\to\infty} Q_\varrho(X_l, Y_l) = Q_\varrho(X^*, Y^*).$$

Using these relations, the continuity of $Q_\varrho(\cdot,\cdot)$, and taking limits on both sides of the first inequality of (33) as $l \in L \to \infty$, we have

$$Q_\varrho(X^*, Y^*) \le Q_\varrho(X, Y^*) \quad \forall X \in \mathcal{X}.$$

Using this relation and the first-order optimality condition, we have

$$\|\mathcal{P}_\mathcal{X}(X^* - \nabla_X Q_\varrho(X^*, Y^*)) - X^*\|_F = 0.$$

By the continuity of $\mathcal{P}_\mathcal{X}(\cdot)$ and $\nabla_X Q_\varrho(\cdot,\cdot)$, and the relation $\lim_{l\in L\to\infty}(X_l, Y_l) = (X^*, Y^*)$, one can see that

$$\lim_{l\in L\to\infty} \|\mathcal{P}_\mathcal{X}(X_l - \nabla_X Q_\varrho(X_l, Y_l)) - X_l\|_F = 0,$$

and hence, the conclusion immediately follows. ∎

3.3 Penalty decomposition method for problem (2)

In this subsection we propose a PD method for solving problem (2) and establish some convergence results for it.

Clearly, (2) can be equivalently reformulated as

$$\min_{X,Y}\{f(X) + \nu \ \mathrm{rank}(Y): \ g(X) \le 0, \ h(X) = 0, \ X - Y = 0, \ X \in \mathcal{X}, \ Y \in \Omega\}. \tag{35}$$

Given a penalty parameter $\varrho > 0$, the associated quadratic penalty function for (35) is defined as

$$P_\varrho(X, Y) := f(X) + \nu \ \mathrm{rank}(Y) + \frac{\varrho}{2}\left(\|[g(X)]^+\|_2^2 + \|h(X)\|_2^2 + \|X - Y\|_F^2\right). \tag{36}$$

We are now ready to present a PD method for solving (35) (or, equivalently, (2)) in which each penalty subproblem is approximately solved by a BCD method.

Algorithm 3.3 Penalty decomposition method for (35):

Let $\varrho_0 > 0$, $c > 1$ be given. Choose an arbitrary $Y_0^0 \in \Omega$ and a constant Υ such that $\Upsilon \ge \max\{f(X^{\mathrm{feas}}) + \nu \ \mathrm{rank}(X^{\mathrm{feas}}), \min_{X\in\mathcal{X}} P_{\varrho_0}(X, Y_0^0)\}$. Set $k = 0$.

1) *Set $l = 0$ and apply the BCD method to find an approximate solution $(X^k, Y^k) \in \mathcal{X} \times \Omega$ to the penalty subproblem*

$$\min\{P_{\varrho_k}(X, Y): \ X \in \mathcal{X}, \ Y \in \Omega\} \tag{37}$$

by performing steps 1a)-1c):

1a) *Solve $X_{l+1}^k \in \mathrm{Arg}\min_{X\in\mathcal{X}} P_{\varrho_k}(X, Y_l^k)$.*

1b) Solve $Y^k_{l+1} \in \mathrm{Arg}\min_{Y\in\Omega} P_{\varrho_k}(X^k_{l+1}, Y)$.

1c) Set $(X^k, Y^k) := (X^k_{l+1}, Y^k_{l+1})$. *If* (X^k, Y^k) *satisfies*

$$\|\mathcal{P}_{\mathcal{X}}(X^k - \nabla_X Q_{\varrho_k}(X^k, Y^k)) - X^k\|_F \le \epsilon_k, \tag{38}$$

where $Q_\varrho(\cdot,\cdot)$ *is defined in* (18)*, then go to step 2).*

1d) Set $l \leftarrow l+1$ *and go to step 1a).*

2) Set $\varrho_{k+1} := c\varrho_k$.

3) If $\min_{X\in\mathcal{X}} P_{\varrho_{k+1}}(X, Y^k) > \Upsilon$*, set* $Y^{k+1}_0 := X^{\mathrm{feas}}$*. Otherwise, set* $Y^{k+1}_0 := Y^k$.

4) Set $k \leftarrow k+1$ *and go to step 1).*

end

Remark. The practical termination criteria proposed in Subsection 3.2 can also be applied to Algorithm 3.3. In addition, one can apply a similar strategy as mentioned in Subsection 3.2 to enhance the performance of the BCD method for solving (37). Finally, in view of Corollary 2.2, the subproblem in step 1b) can be reduced to a problem in the form of (7), which has a closed-form solution when Ω is simple enough.

We next establish a convergence result regarding the inner iterations of Algorithm 3.3. In particular, we will show that an approximate solution (X^k, Y^k) of problem (37) satisfying (38) can be found by the BCD method described in steps 1a)-1d) of Algorithm 3.3. For convenience of presentation, we omit the index k from (37) and consider the BCD method for solving the following problem:

$$\min\{P_\varrho(X, Y) : \ X \in \mathcal{X}, \ Y \in \Omega\} \tag{39}$$

instead. Accordingly, we rename the iterates of the BCD method presented in Algorithm 3.3. We can observe that the resulting BCD method is the same as the one presented in Subsection 3.2 except that P_ϱ and Ω replace Q_ϱ and $\mathcal{Y}$, respectively. For the sake of brevity, we omit the presentation of this BCD method.

Theorem 3.5 *Let* $\{(X_l, Y_l)\}$ *be the sequence generated by the BCD method applied to problem* (39)*, and let* $\epsilon > 0$ *be given. Suppose that* $\{(X_l, Y_l)\}$ *has at least an accumulation point. Then, there exists some* $l > 0$ *such that*

$$\|\mathcal{P}_{\mathcal{X}}(X_l - \nabla_X Q_\varrho(X_l, Y_l)) - X_l\|_F < \epsilon,$$

where Q_ϱ *is defined in (18).*

Proof. We first observe that

$$\begin{aligned} P_\varrho(X_{l+1}, Y_l) &\le P_\varrho(X, Y_l) \quad \forall X \in \mathcal{X}, & (40)\\ P_\varrho(X_l, Y_l) &\le P_\varrho(X_l, Y) \quad \forall Y \in \Omega. & (41)\end{aligned}$$

It then follows that

$$P_\varrho(X_{l+1}, Y_{l+1}) \le P_\varrho(X_{l+1}, Y_l) \le P_\varrho(X_l, Y_l) \quad \forall l \ge 1. \tag{42}$$

Hence, the sequence $\{P_\varrho(X_l, Y_l)\}$ is non-increasing. By the assumption, $\{(X_l, Y_l)\}$ has at least an accumulation point, denoted by (X^*, Y^*). Then, there exists a subsequence L such that $\lim_{l\in L\to\infty}(X_l, Y_l) =$

(X^*, Y^*) and moreover, $X^* \in \mathcal{X}$ due to the closedness of $\mathcal{X}$. We can observe that $\{P_\varrho(X_l, Y_l)\}_{l \in L}$ is bounded, which together with the monotonicity of $\{P_\varrho(X_l, Y_l)\}$ implies that $\{P_\varrho(X_l, Y_l)\}$ is bounded below and hence $\lim_{l\to\infty} P_\varrho(X_l, Y_l)$ exists. This observation and (42) yield

$$\lim_{l\to\infty} P_\varrho(X_l, Y_l) = \lim_{l\to\infty} P_\varrho(X_{l+1}, Y_l). \tag{43}$$

For the sake of notational convenience, let

$$F(X) := f(X) + \frac{\varrho}{2}(\|[g(X)]^+\|_2^2 + \|h(X)\|_2^2).$$

It then follows from (36) that

$$P_\varrho(X, Y) = F(X) + \nu \operatorname{rank}(Y) + \frac{\varrho}{2}\|X - Y\|_F^2. \tag{44}$$

In view of (40) and (44), we have

$$\begin{array}{rcl} F(X) + \frac{\varrho}{2}\|X - Y_l\|_F^2 & = & P_\varrho(X, Y_l) - \nu \operatorname{rank}(Y_l) \;\geq\; P_\varrho(X_{l+1}, Y_l) - \nu \operatorname{rank}(Y_l) \\ & = & F(X_{l+1}) + \frac{\varrho}{2}\|X_{l+1} - Y_l\|_F^2, \quad \forall X \in \mathcal{X}. \end{array} \tag{45}$$

Since $\{\operatorname{rank}(Y_l)\}_{l\in L}$ is bounded, there exists a subsequence $\bar{L} \subseteq L$ such that $\lim_{l\in\bar{L}\to\infty} \operatorname{rank}(Y_l)$ exists. Then we have

$$\begin{array}{l} \lim\limits_{l\in\bar{L}\to\infty} F(X_{l+1}) + \frac{\varrho}{2}\|X_{l+1} - Y_l\|_F^2 \;=\; \lim\limits_{l\in\bar{L}\to\infty} P_\varrho(X_{l+1}, Y_l) - \nu \operatorname{rank}(Y_l) \\ = \lim\limits_{l\in\bar{L}\to\infty} P_\varrho(X_{l+1}, Y_l) - \nu \lim\limits_{l\in\bar{L}\to\infty} \operatorname{rank}(Y_l) \;=\; \lim\limits_{l\in\bar{L}\to\infty} P_\varrho(X_l, Y_l) - \nu \lim\limits_{l\in\bar{L}\to\infty} \operatorname{rank}(Y_l) \\ = \lim\limits_{l\in\bar{L}\to\infty} P_\varrho(X_l, Y_l) - \nu \operatorname{rank}(Y_l) = \lim\limits_{l\in\bar{L}\to\infty} F(X_l) + \frac{\varrho}{2}\|X_l - Y_l\|_F^2 \;=\; F(X^*) + \frac{\varrho}{2}\|X^* - Y^*\|_F^2, \end{array}$$

where the third equality is due to (43). Using this relation and taking limits on both sides of (45) as $l \in \bar{L} \to \infty$, we further have

$$F(X) + \frac{\varrho}{2}\|X - Y^*\|_F^2 \;\geq\; F(X^*) + \frac{\varrho}{2}\|X^* - Y^*\|_F^2, \quad \forall X \in \mathcal{X},$$

which together with (18) yields

$$Q_\varrho(X, Y^*) \;\geq\; Q_\varrho(X^*, Y^*), \quad \forall X \in \mathcal{X}.$$

The rest of proof is similar to that of Theorem 3.4. ■

Before ending this subsection we next establish the convergence of the outer iterations of Algorithm 3.3 for solving problem (2). In particular, we show that under some suitable assumptions, any accumulation point of the sequence generated by Algorithm 3.3 satisfies the first-order optimality conditions (15).

Theorem 3.6 *Let $\{(X^k, Y^k)\}$ be the sequence generated by Algorithm 3.3 satisfying (38), and let $\{(\lambda^k, \mu^k, Z_X^k)\}$ be the associated sequence defined according to (25). Assume that $\epsilon_k \to 0$. Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \leq \Upsilon\}$ is compact. Then, the following statements hold:*

(a) The sequence $\{(X^k, Y^k)\}$ is bounded. Moreover, any accumulation point (X^*, Y^*) of the sequence $\{(X^k, Y^k)\}$ is a feasible point of problem (35).

(b) Suppose that a subsequence $\{(X^k,Y^k)\}_{k\in\bar{K}}$ *converges to* (X^*,Y^*) *and* $\mathrm{rank}(Y^k) = r$ *for all* $k\in\bar{K}$*, where* $r = \mathrm{rank}(Y^*)$*. Let* $U^k \in \Re^{m\times r}$ *and* $V^k \in \Re^{r\times n}$ *be such that* (23) *holds. Then,* $\{(X^k,Y^k,U^k,V^k)\}_{k\in\bar{K}}$ *is bounded. Further, let* $K\subseteq \bar{K}$ *be a subsequence such that* $\{(X^k,Y^k,U^k,V^k)\}_{k\in K}$ *converges to* (X^*,Y^*,U^*,V^*)*. Assume that the Robinson condition* (14) *holds at* (X^*,U^*,V^*) *and* (24) *holds for sufficiently large* $k\in K$*. Then,* $\{(\lambda^k,\mu^k,Z_X^k)\}_{k\in K}$ *is bounded, and each accumulation point* (λ^*,μ^*,Z_X^*) *of* $\{(\lambda^k,\mu^k,Z_X^k)\}_{k\in K}$ *together with* (X^*,U^*,V^*) *and some* $Z_Y^*\in\Re^{m\times n}$ *satisfies the first-order optimality conditions* (15)*.*

Proof. Statement (a) and the first part of statement (b) can be similarly proved as in Theorem 3.3. We now prove the second part of statement (b). Indeed, by the definition of Y^k, one can see that

$$Y^k \in \mathrm{Arg}\min_{Y\in\Omega} P_{\varrho_k}(X^k,Y),$$

which, together with (36), (18) and the assumption that $\mathrm{rank}(Y^k)=r$ for all $k\in\bar{K}$, implies that

$$Y^k \in \mathrm{Arg}\min_{Y\in\Omega}\{Q_{\varrho_k}(X^k,Y) : \mathrm{rank}(Y)\le r\}, \quad \forall k\in\bar{K}.$$

Using this relation and (23), we can observe that

$$(Y^k,U^k,V^k)\in \mathrm{Arg}\min\{Q_{\varrho_k}(X^k,UV):\ Y-UV=0, Y\in\Omega, U\in\Re^{m\times r}, V\in\Re^{r\times n}\}, \quad \forall k\in\bar{K}.$$

By virtue of this relation and (38), the rest of proof follows similarly as that of Theorem 3.3. ■

4 Penalty decomposition method for rank minimization of general symmetric matrices

In this section, we consider rank minimization problems (1) and (2) by assuming that $\mathcal{X}$ is a closed convex set in $\mathcal{S}^n$, Ω is a closed unitarily invariant convex set in $\mathcal{S}^n$, and $f:\mathcal{S}^n\to\Re$, $g:\mathcal{S}^n\to\Re^p$ and $h:\mathcal{S}^n\to\Re^t$ are continuously differentiable functions. In particular, we first study the first-order optimality conditions for (1) and (2) in the context of symmetric matrix space. We then discuss the convergence of the PD methods for solving these problems. As in Section 3, we also assume that problems (1) and (2) are feasible, and moreover, at least a feasible solution, denoted by X^{feas}, is known.

We first study the first-order optimality conditions for problems (1) and (2) in the context of symmetric matrix space.

Theorem 4.1 *Suppose that* X^* *is a local minimizer of problem (1) in the context of symmetric matrix space. Let* $U^*\in\Re^{n\times r}$*,* $D^*\in\mathcal{D}^r$ *be such that* $(U^*)^TU^*=I$ *and* $X^*=U^*D^*(U^*)^T$*. Assume that the following Robinson condition holds:*

$$\left\{\begin{pmatrix} g'(X^*)d_X - v \\ h'(X^*)d_X \\ d_X - d_U D^*(U^*)^T - U^*d_D(U^*)^T - U^*D^*d_U^T \\ d_Y - d_U D^*(U^*)^T - U^*d_D(U^*)^T - U^*D^*d_U^T \end{pmatrix} : \begin{array}{l} d_X\in\mathcal{T}_{\mathcal{X}}(X^*), v\in\Re^p, v_i\le 0, i\in\mathcal{A}(X^*), \\ d_U\in\Re^{n\times r}, d_D\in\mathcal{D}^r, d_Y\in\mathcal{T}_\Omega(X^*) \end{array}\right\} = \Re^p\times\Re^t\times\mathcal{S}^n\times\mathcal{S}^n. \tag{46}$$

Then, there exist $\lambda^*\in\Re^p_+$*,* $\mu^*\in\Re^t$*,* $Z_X^*\in\mathcal{S}^n$*,* $Z_Y^*\in\mathcal{S}^n$ *such that*

$$\begin{array}{l} -\nabla f(X^*) - \nabla g(X^*)\lambda^* - \nabla h(X^*)\mu^* - Z_X^* \in \mathcal{N}_{\mathcal{X}}(X^*), \\ (Z_X^*-Z_Y^*)U^*D^* = 0, \ \ \widetilde{\mathscr{D}}\left((U^*)^T(Z_X^*-Z_Y^*)U^*\right) = 0, \\ \lambda_i^*\ge 0, \ \lambda_i^* g_i(X^*) = 0, \ i=1,\ldots,p; \quad Z_Y^*\in\mathcal{N}_\Omega(X^*). \end{array} \tag{47}$$

Proof. Let $Y^* = X^*$. Since X^* is a local minimizer of problem (1) in the context of symmetric matrix space, one can observe that (X^*, Y^*, U^*, D^*) is a local minimizer of

$$\min_{X,Y,U,D} \{f(X) : \, g(X) \le 0, \; h(X) = 0, \; X - UDU^T = 0, \; Y - UDU^T = 0, \; X \in \mathcal{X}, \; Y \in \Omega, \; U \in \Re^{n\times r}, \; D \in \mathcal{D}^r\}.$$

Using this observation, (46) and Theorem 3.25 of [40], we see that the conclusion holds. ∎

Theorem 4.2 *Suppose that X^* is a local minimizer of problem (2) in the context of symmetric matrix space. Let $r = \mathrm{rank}(X^*)$, $U^* \in \Re^{n\times r}$, $D^* \in \mathcal{D}^r$ be such that $(U^*)^T U^* = I$ and $X^* = U^* D^* (U^*)^T$. Assume that the Robinson condition (46) holds. Then, there exist $\lambda^* \in \Re^p_+$, $\mu^* \in \Re^t$, $Z^*_X \in \mathcal{S}^n$, $Z^*_Y \in \mathcal{S}^n$ such that (47) holds.*

Proof. By the assumption that X^* is a local minimizer of problem (2), we observe that X^* is a local minimizer of problem (1) with $r = \mathrm{rank}(X^*)$. The conclusion of this theorem then immediately follows from Theorem 4.1. ∎

Clearly, the PD methods (namely, Algorithms 3.1 and 3.3) proposed in Section 3 can be directly applied to solve problems (1) and (2) in the context of symmetric matrix space detailed in the beginning of this section. We now state their convergence results in the following theorems. Their proofs are similar to those of Theorems 3.3 and 3.6.

Theorem 4.3 *Let $\{(X^k, Y^k)\}$ be the sequence generated by Algorithm 3.1 applied to problem (1) in the context of symmetric matrix space, and let $\{(\lambda^k, \mu^k, Z^k_X)\}$ be the associated sequence defined according to (25). In addition, let $U^k \in \Re^{n\times r}$ and $D^k \in \mathcal{D}^r$ be such that*

$$(U^k)^T U^k = I, \quad Y^k = U^k D^k (U^k)^T. \tag{48}$$

Assume that $\{(X^k, Y^k)\}$ satisfies (20) and $\epsilon_k \to 0$. Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \le \Upsilon\}$ is compact. Then, the following statements hold:

(a) The sequence $\{(X^k, Y^k, U^k, D^k)\}$ is bounded;

(b) Suppose that a subsequence $\{(X^k, Y^k, U^k, D^k)\}_{k\in K}$ converges to (X^*, Y^*, U^*, D^*). Then, X^* is a feasible point of problem (1), and moreover, $X^* = Y^* = U^* D^* (U^*)^T$. In addition, assume that the Robinson condition (46) holds at (X^*, U^*, D^*) and

$$\left\{ d_Y - d_U D^k (U^k)^T - U^k d_D (U^k)^T - U^k D^k d_U^T : d_U \in \Re^{n\times r}, d_D \in \mathcal{D}^r, d_Y \in \mathcal{T}_\Omega(Y^k) \right\} = \mathcal{S}^n \tag{49}$$

holds for sufficiently large $k \in K$. Then, $\{(\lambda^k, \mu^k, Z^k_X)\}_{k\in K}$ is bounded, and each accumulation point $(\lambda^*, \mu^*, Z^*_X)$ of $\{(\lambda^k, \mu^k, Z^k_X)\}_{k\in K}$ together with (X^*, U^*, D^*) and some $Z^*_Y \in \mathcal{S}^n$ satisfies the first-order optimality conditions (47).

Theorem 4.4 *Let $\{(X^k, Y^k)\}$ be the sequence generated by Algorithm 3.3 applied to problem (2) in the context of symmetric matrix space, and let $\{(\lambda^k, \mu^k, Z^k_X)\}$ be the associated sequence defined according to (25). Assume that $\{(X^k, Y^k)\}$ satisfies (38) and $\epsilon_k \to 0$. Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \le \Upsilon\}$ is compact. Then, the following statements hold:*

(a) The sequence $\{(X^k, Y^k)\}$ is bounded. Moreover, any accumulation point (X^*, Y^*) of the sequence $\{(X^k, Y^k)\}$ is a feasible point of problem (35).

(b) Suppose that a subsequence $\{(X^k, Y^k)\}_{k\in\bar{K}}$ converges to (X^*, Y^*) and $\mathrm{rank}(Y^k) = r$ for all $k \in \bar{K}$, where $r = \mathrm{rank}(Y^*)$. Let $U^k \in \Re^{n\times r}$ and $D^k \in \mathcal{D}^r$ be such that (48) holds. Then, $\{(X^k, Y^k, U^k, D^k)\}_{k\in\bar{K}}$ is bounded. Further, let $K \subseteq \bar{K}$ be a subsequence such that $\{(X^k, Y^k, U^k, D^k)\}_{k\in K}$ converges to (X^*, Y^*, U^*, D^*). Assume that the Robinson condition (46) holds at (X^*, U^*, D^*) and (49) holds for sufficiently large $k \in K$. Then, $\{(\lambda^k, \mu^k, Z^k_X)\}_{k\in K}$ is bounded, and each accumulation point $(\lambda^*, \mu^*, Z^*_X)$ of $\{(\lambda^k, \mu^k, Z^k_X)\}_{k\in K}$ together with (X^*, U^*, D^*) and some $Z^*_Y \in \mathcal{S}^n$ satisfies the first-order optimality conditions (47).

5 Penalty decomposition method for rank minimization of positive semidefinite matrices

In this section, we consider rank minimization problems (1) and (2) by assuming that $\mathcal{X}$ is a closed convex set in $\mathcal{S}^n$, Ω is a closed unitarily invariant convex set in $\mathcal{S}^n_+$, and $f : \mathcal{S}^n \to \Re$, $g : \mathcal{S}^n \to \Re^p$ and $h : \mathcal{S}^n \to \Re^t$ are continuously differentiable functions. As in Section 3, we also assume that problems (1) and (2) are feasible, and moreover, at least a feasible solution, denoted by X^{feas}, is known. Since $\Omega \subseteq \mathcal{S}^n_+$, it can be represented as

$$\Omega = \mathcal{S}^n_+ \cap \widetilde{\Omega}$$

for some unitarily invariant convex set $\widetilde{\Omega} \subseteq \mathcal{S}^n$. For example, when $\Omega = \{X \in \mathcal{S}^n_+ : \mathrm{Tr}(X) = 1\}$, one can choose $\widetilde{\Omega} = \{X \in \mathcal{S}^n : \mathrm{Tr}(X) = 1\}$. In contrast with the Ω considered in Section 4, the above Ω is more structured. As shown below, by exploiting the structure of Ω, we are able to derive simpler first-order optimality conditions for (1) and (2) in the context of positive semidefinite matrices than those given in Theorems 4.1 and 4.2. We can also establish convergence of the PD methods for solving these problems under simpler conditions.

We first study the first-order optimality conditions for problems (1) and (2) in the context of positive semidefinite matrices.

Theorem 5.1 *Suppose that X^* is a local minimizer of problem (1) in the context of positive semidefinite matrices. Let $W^* \in \Re^{n\times r}$ be such that $X^* = W^*(W^*)^T$. Assume that the following Robinson condition holds:*

$$\left\{\begin{pmatrix} g'(X^*)d_X - v \\ h'(X^*)d_X \\ d_X - d_W(W^*)^T - W^*d_W^T \\ d_Y - d_W(W^*)^T - W^*d_W^T \end{pmatrix} : \begin{array}{l} d_X \in \mathcal{T}_{\mathcal{X}}(X^*), v \in \Re^p, v_i \le 0, i \in \mathcal{A}(X^*), \\ d_W \in \Re^{n\times r}, d_Y \in \mathcal{T}_{\widetilde{\Omega}}(X^*) \end{array}\right\} = \Re^p \times \Re^t \times \mathcal{S}^n \times \mathcal{S}^n. \tag{50}$$

Then, there exist $\lambda^* \in \Re^p_+$, $\mu^* \in \Re^t$, $Z^*_X \in \mathcal{S}^n$, $Z^*_Y \in \mathcal{S}^n$ such that

$$\begin{array}{c} -\nabla f(X^*) - \nabla g(X^*)\lambda^* - \nabla h(X^*)\mu^* - Z^*_X \in \mathcal{N}_{\mathcal{X}}(X^*), \\ (Z^*_X - Z^*_Y)W^* = 0, \\ \lambda^*_i \ge 0,\ \lambda^*_i g_i(X^*) = 0,\ i = 1, \ldots, p; \quad Z^*_Y \in \mathcal{N}_{\widetilde{\Omega}}(X^*). \end{array} \tag{51}$$

Proof. Let $Y^* = X^*$. Since X^* is a local minimizer of problem (1) in the context of positive semidefinite matrices, one can observe that (X^*, Y^*, W^*) is a local minimizer of

$$\min_{X,Y,W} \{f(X): \ g(X) \leq 0, \ h(X) = 0, \ X - WW^T = 0, \ Y - WW^T = 0, \ X \in \mathcal{X}, \ Y \in \widetilde{\Omega}, \ W \in \Re^{n \times r}\}.$$

Using this observation, (50) and Theorem 3.25 of [40], we see that the conclusion holds. ■

Theorem 5.2 *Suppose that X^* is a local minimizer of problem (2) in the context of positive semidefinite matrices. Let $r = \text{rank}(X^*)$, $W^* \in \Re^{n \times r}$ be such that $X^* = W^*(W^*)^T$. Assume that the Robinson condition (50) holds. Then, there exist $\lambda^* \in \Re^p_+$, $\mu^* \in \Re^t$, $Z_X^* \in \mathcal{S}^n$, $Z_Y^* \in \mathcal{S}^n$ such that (51) holds.*

Proof. By the assumption that X^* is a local minimizer of problem (2), we observe that X^* is a local minimizer of problem (1) with $r = \text{rank}(X^*)$. The conclusion of this theorem then immediately follows from Theorem 5.1. ■

Clearly, the PD methods (namely, Algorithms 3.1 and 3.3) proposed in Section 3 can be directly applied to solve problems (1) and (2) in the context of positive semidefinite matrices detailed in the beginning of this section. We now state their convergence results as follows. Their proofs are similar to those of Theorems 3.3 and 3.6.

Theorem 5.3 *Let $\{(X^k, Y^k)\}$ be the sequence generated by Algorithm 3.1 applied to problem (1) in the context of positive semidefinite matrices, and let $\{(\lambda^k, \mu^k, Z_X^k)\}$ be the associated sequence defined according to (25). In addition, let $W^k \in \Re^{n \times r}$ be such that $Y^k = W^k(W^k)^T$. Assume that $\{(X^k, Y^k)\}$ satisfies (20) and $\epsilon_k \to 0$. Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \leq \Upsilon\}$ is compact. Then, the following statements hold:*

(a) The sequence $\{(X^k, Y^k, W^k)\}$ is bounded;

(b) Suppose that a subsequence $\{(X^k, Y^k, W^k)\}_{k \in K}$ converges to (X^*, Y^*, W^*). Then, X^* is a feasible point of problem (1), and moreover, $X^* = Y^* = W^*(W^*)^T$. In addition, assume that the Robinson condition (50) *holds at (X^*, W^*) and*

$$\left\{ d_Y - d_W(W^k)^T - W^k d_W^T : d_W \in \Re^{n \times r}, d_Y \in \mathcal{T}_{\widetilde{\Omega}}(Y^k) \right\} = \mathcal{S}^n \quad (52)$$

holds for sufficiently large $k \in K$. Then, $\{(\lambda^k, \mu^k, Z_X^k)\}_{k \in K}$ is bounded, and each accumulation point $(\lambda^*, \mu^*, Z_X^*)$ of $\{(\lambda^k, \mu^k, Z_X^k)\}_{k \in K}$ together with (X^*, W^*) and some $Z_Y^* \in \mathcal{S}^n$ satisfies the first-order optimality conditions (51).

Theorem 5.4 *Let $\{(X^k, Y^k)\}$ be the sequence generated by Algorithm 3.3 applied to problem (2) in the context of positive semidefinite matrices, and let $\{(\lambda^k, \mu^k, Z_X^k)\}$ be the associated sequence defined according to (25). Assume that $\{(X^k, Y^k)\}$ satisfies (38) and $\epsilon_k \to 0$. Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \leq \Upsilon\}$ is compact. Then, the following statements hold:*

(a) The sequence $\{(X^k, Y^k)\}$ is bounded. Moreover, any accumulation point (X^*, Y^*) of the sequence $\{(X^k, Y^k)\}$ is a feasible point of problem (35).

(b) *Suppose that a subsequence* $\{(X^k, Y^k)\}_{k\in\bar{K}}$ *converges to* (X^*, Y^*) *and* $\mathrm{rank}(Y^k) = r$ *for all* $k \in \bar{K}$, *where* $r = \mathrm{rank}(Y^*)$. *Let* $W^k \in \Re^{n\times r}$ *be such that* $Y^k = W^k(W^k)^T$. *Then,* $\{(X^k, Y^k, W^k)\}_{k\in\bar{K}}$ *is bounded. Further, let* $K \subseteq \bar{K}$ *be a subsequence such that* $\{(X^k, Y^k, W^k)\}_{k\in K}$ *converges to* (X^*, Y^*, W^*). *Assume that the Robinson condition* (50) *holds at* (X^*, W^*) *and* (52) *holds for sufficiently large* $k \in K$. *Then,* $\{(\lambda^k, \mu^k, Z_X^k)\}_{k\in K}$ *is bounded, and each accumulation point* $(\lambda^*, \mu^*, Z_X^*)$ *of* $\{(\lambda^k, \mu^k, Z_X^k)\}_{k\in K}$ *together with* (X^*, W^*) *and some* $Z_Y^* \in \mathcal{S}^n$ *satisfies the first-order optimality conditions* (51).

6 Numerical results

In this section we conduct numerical experiments to test the performance of the PD methods proposed in Sections 3 and 5 by applying them to solve matrix completion and nearest low-rank correlation matrix problems. The codes of all the methods implemented in this section are written in Matlab and all experiments are performed in Matlab 7.11.0 (2010b) on a workstation with an Intel Xeon E5410 CPU (2.33 GHz) and 8GB RAM running Red Hat Enterprise Linux (kernel 2.6.18).

6.1 Matrix completion problem

In this subsection we apply our PD method proposed in Section 3 to the matrix completion problem, which has numerous applications in control and systems theory, image recovery and data mining (see, for example, [42, 32, 9, 24]). It can be formulated as

$$\begin{array}{ll} \min\limits_{X\in\Re^{m\times n}} & \mathrm{rank}(X) \\ \mathrm{s.t.} & X_{ij} = M_{ij},\ (i,j) \in \Theta, \end{array} \tag{53}$$

where $M \in \Re^{m\times n}$ and Θ is a subset of index pairs (i, j). Recently, numerous methods were proposed to solve the nuclear norm relaxation or the variant of (53) (see, for example, [26, 6, 29, 8, 19, 21, 28, 30, 41, 25, 45]).

It is not hard to see that problem (53) is a special case of the general rank minimization problem (2) with $f(X) \equiv 0$, $p = q = 0$, $\nu = 1$, $\Omega = \Re^{m\times n}$, and

$$\mathcal{X} = \{X \in \Re^{m\times n} : X_{ij} = M_{ij},\ (i,j) \in \Theta\}.$$

Thus, the PD method proposed in Section 3 for problem (2) can be suitably applied to (53). Recall that the main computational parts of this method lie in solving the subproblems in steps 1a) and 1b). In the context of (53), these subproblems are in the form of

$$\min_X \{\|X - A\|_F^2 : X \in \mathcal{X}\}, \tag{54}$$

$$\min_Y \{\mathrm{rank}(Y) + \varrho\|Y - B\|_F^2 : Y \in \Re^{m\times n}\} \tag{55}$$

for some $\varrho > 0$, $A, B \in \Re^{m\times n}$, respectively. From the above definition of $\mathcal{X}$, we observe that problem (54) has a closed-form solution. In addition, it follows from Corollary 2.2 that problem (55) also has a closed-form solution.

We now address the initialization and the termination criteria for our PD method when applied to (53). In particular, we choose X^{feas} to be the $m \times n$ matrix satisfying $X_{ij}^{\text{feas}} = M_{ij}$ for all $(i,j) \in \Theta$ and $X_{ij}^{\text{feas}} = 0$ for all $(i,j) \notin \Theta$, and set $Y_0^0 = X^{\text{feas}}$. In addition, we choose the initial penalty parameter ϱ_0 to

Table 1: Computational results for $m = 500$, $n = 500$ and $p = 125000$

Rank	FPCA			LMaFit			PD		
r	NS	rel_err	Time	NS	rel_err	Time	NS	rel_err	Time
10	50	7.80e−6	6.7	50	1.21e−4	0.2	50	6.35e−5	8.7
20	50	2.07e−5	6.9	50	1.33e−4	0.3	50	9.16e−5	10.1
30	50	3.26e−5	7.8	50	1.72e−4	0.6	50	9.34e−5	11.9
40	50	6.57e−5	8.4	50	2.10e−4	0.9	50	1.31e−4	14.4
50	50	1.16e−4	8.9	50	2.29e−4	1.3	50	1.34e−4	17.1

be 0.1, and set the parameter $c = 5$. We use (21) and (22) as the inner and outer termination criteria for the PD method and set their associated accuracy parameters $\epsilon_I = 10^{-4}$ and $\epsilon_O = \sqrt{\sum_{(i,j)\in\Theta} M_{ij}^2} \times 10^{-4}$.

Next we conduct numerical experiments to test the performance of our PD method for solving (53) on random data. We also compare the results of our method with other two related methods, that is, FPCA [29] and LMaFit [45]. It shall be mentioned that FPCA solves a (convex) nuclear norm relaxation of (53) and LMaFit solves a nonconvex smooth reformulation of (53).

In the first experiment, we aim to recover a random matrix $M \in \Re^{m\times n}$ with rank r based on a subset of entries $\{M_{ij}\}_{(i,j)\in\Theta}$. For this purpose, we randomly generate M and Θ by a similar procedure as described in [29]. In detail, we first generate random matrices $M_L \in \Re^{m\times r}$ and $M_R \in \Re^{r\times n}$ with i.i.d. standard Gaussian entries and let $M = M_L M_R^T$. We then sample a subset Θ of s entries uniformly at random. In our experiment, we set $m = n = 500$, $s = 125,000$, and randomly generate 50 copies of M for five different values of r.

Given an approximate recovery X^* for M, we define the relative error as

$$\text{rel\_err} := \frac{\|X^* - M\|_F}{\|M\|_F}.$$

We adopt the same criterion as used in [36, 7], and say a matrix M is *successfully recovered* by X^* if the corresponding relative error is less than 10^{-3}. For each rank r, we apply our PD method and the aforementioned methods FPCA and LMaFit to recover M on 50 instances that are randomly generated above. In particular, we set the parameters tol $= 10^{-4}$, $K = \lfloor 1.25r \rfloor$, est_rank = 1 and rank_max = 500 for LMaFit, and set tol $= 10^{-4}$, $\mu = 10^{-6}$ and maxr $= \lfloor 1.25r \rfloor$ for FPCA. All other parameters of these two methods are chosen by default. For convenience of presentation, we use NS to denote the number of matrices that are successfully recovered. The computational results are presented in Table 1. In detail, the rank r of the problems is given in the first column. The results of all three methods in terms of NS, the average rel_err and the average CPU time on the successfully recovered instances are reported in columns two to ten, respectively. Table 1 shows that the recoverability of three methods is comparable for all the instances. In addition, we observe that FPCA and PD are slower than LMaFit because they require a full or partial singular value decomposition at each inner iteration while LMaFit does not need any singular value decomposition. Finally, we remark that the recoverability of LMaFit and FPCA depends on the upper bounds of the rank r of the matrix M, which are the parameters K and maxr, respectively. We observed in our experiments that for larger values of K and maxr, the recoverability of LMaFit and FPCA becomes worse. Therefore, a good upper bound on r seems to be crucial for these methods, but it is not needed for our PD method.

Our second experiment is similar to the one conducted in [45]. Its goal is to recover a high-rank matrix $M \in \Re^{n\times n}$, where most of the singular values of M are close to zero, by a low-rank matrix based on a subset of entries $\{M_{ij}\}_{(i,j)\in\Theta}$. To this end, we randomly generate M and Θ by a similar procedure as described in [45]. In particular, we first generate random matrices $M_L \in \Re^{n\times n}$ and $M_R \in \Re^{n\times n}$ with i.i.d. standard Gaussian entries. Then we obtain matrices U and V by orthonormalizing the columns of M_L and M_R, respectively. We generate two types of diagonal matrix Σ: one has diagonal elements $\sigma_i = i^{-5}$ for all i while another has diagonal elements $\sigma_i = 9.9^{-(i-1)}$ for all i. Finally, we set $M = U\Sigma V^T$ and sample a subset Θ of s entries uniformly at random. We generate 50 instances with $n = 500$ and a sample ratio SR varying from 0.1 to 0.9, where $SR = s/(mn)$.

For each sample ratio SR, we apply PD, FPCA and LMaFit methods to recover M on 50 instances that are randomly generated above. As in [45], we set tol $= 10^{-4}$, $K = 1$, est_rank $= 2$, rk_inc $= 1$, rank_max $= 500$ for LMaFit, while the parameters for the other two methods are the same as the ones above except maxr $= 500$ for FPCA. We notice that LMaFit solves a sequence of nonlinear least squares problems by gradually increasing the trial rank for the unknown recovery while FPCA solves a sequence of nuclear norm relaxation problems of (53) by gradually reducing the associated regularization parameter which essentially also increases the trial rank for the unknown recovery. As our aim is to find a successful recovery with the lowest possible rank, we immediately terminate LMaFit, FPCA and PD once a recovery with rel_err less than 10^{-3} is obtained. The computational results are presented in Tables 2 and 3. In particular, Table 2 reports the results for the instances with $\sigma_i = i^{-5}$ for all i while Table 3 presents the results for the instances with $\sigma_i = 9.9^{-(i-1)}$ for all i. In each table, the sample ratio SR of the test problems is given in the first column. The average rank of the solution matrices, the average rel_err and the average CPU time for these three methods over each group of 50 randomly generated instances are reported in columns two to ten, respectively. From Table 2, we observe that PD and FPCA are slower than LMaFit. In addition, M is successfully recovered by PD and LMaFit for all instances as their average rel_err is below 10^{-3} while FPCA only successfully recovers M for the instances with the sample ratios larger than 0.2. We also observe that PD generally provides smaller average rank than LMaFit and FPCA for all instances. Now, one natural question is whether there exists a matrix X^* with a smaller rank than the one given by PD for successfully recovering such a M. The answer is actually not. Indeed, let X^* be a matrix of rank at most three with smallest rel_err, that is,

$$X^* \in \operatorname{Arg\,min}\{\|X - M\|_F : \ \operatorname{rank}(X) \le 3\}.$$

Using Corollary 2.3 and the fact that $\sigma_i(M) = i^{-5}$ for all i, we have

$$\frac{\|X^* - M\|_F}{\|M\|_F} = \frac{\sqrt{\sum_{i=4}^{500} \sigma_i^2(M)}}{\sqrt{\sum_{i=1}^{500} \sigma_i^2(M)}} = \frac{\sqrt{\sum_{i=4}^{500} i^{-10}}}{\sqrt{\sum_{i=1}^{500} i^{-10}}} \approx 1.04\mathrm{e}-3 > 10^{-3}.$$

Therefore, according to the above criterion, any matrix of rank at most three cannot successfully recover such a M. It follows that our PD method is capable of producing a matrix with smallest possible rank to successfully recover M, but the other two methods cannot. A similar phenomenon can also be observed in Table 3. Though LMaFit generally outperforms our PD method in terms of speed, it is a specifically developed method for solving matrix completion problems while our PD method can be applied to much broader class of problems.

6.2 Nearest low-rank correlation matrix problem

In this subsection we apply our PD method proposed in Section 5 to find the nearest low-rank correlation matrix, which has important applications in finance (see, for example, [4, 38, 46, 47, 39]). This problem

Table 2: Computational results for $n = 500$ and $\sigma_i = i^{-5}$

	FPCA			LMaFit			PD		
SR	Rank	rel_err	Time	Rank	rel_err	Time	Rank	rel_err	Time
0.1	11.9	3.40e−2	32.6	4.6	9.11e−4	0.1	4.0	9.67e−4	54.2
0.2	5.0	5.16e−4	17.2	4.6	8.24e−4	0.1	4.0	9.04e−4	20.9
0.3	5.0	3.97e−4	14.6	4.6	7.73e−4	0.1	4.0	8.01e−4	16.1
0.4	5.0	3.66e−4	11.2	4.5	7.23e−4	0.1	4.0	7.09e−4	13.6
0.5	5.5	3.26e−4	7.8	4.5	6.15e−4	0.2	4.0	6.23e−4	8.4
0.6	5.6	3.22e−4	6.9	4.3	6.03e−4	0.2	4.0	5.42e−4	6.4
0.7	5.7	3.01e−4	6.4	5.0	5.95e−4	0.2	4.0	4.70e−4	5.7
0.8	6.2	2.88e−4	6.2	4.0	5.93e−4	0.2	4.0	4.12e−4	4.9
0.9	6.5	2.69e−4	5.9	4.0	6.04e−4	0.2	4.0	3.73e−4	3.8

Table 3: Computational results for $n = 500$ and $\sigma_i = 9.9^{-(i-1)}$

	FPCA			LMaFit			PD		
SR	Rank	rel_err	Time	Rank	rel_err	Time	Rank	rel_err	Time
0.1	12.2	3.11e−2	32.3	4.6	9.23e−4	0.1	4.0	9.41e−4	52.8
0.2	5.0	4.09e−4	12.9	4.5	7.92e−4	0.1	4.0	8.72e−4	28.7
0.3	5.0	1.94e−4	11.4	4.6	6.84e−4	0.1	4.0	7.58e−4	21.3
0.4	5.0	1.25e−4	10.7	4.6	6.27e−4	0.1	4.0	6.49e−4	14.9
0.5	5.0	1.12e−4	9.2	5.1	5.91e−4	0.2	4.0	5.45e−4	10.5
0.6	5.0	1.09e−4	8.4	4.1	5.10e−4	0.2	4.0	4.42e−4	8.1
0.7	5.5	1.07e−4	8.1	5.0	4.66e−4	0.2	4.0	3.41e−4	6.9
0.8	5.6	1.03e−5	7.5	4.2	4.19e−4	0.3	4.0	2.44e−4	6.7
0.9	5.6	1.03e−5	7.1	4.0	4.31e−4	0.3	4.0	1.56e−4	6.4

can be formulated as

$$\begin{array}{rl} \min\limits_{X \in S^n} & \frac{1}{2}\|H \circ (X - C)\|_F^2 \\ \text{s.t.} & \mathrm{diag}(X) = e, \\ & \mathrm{rank}(X) \le r, \ X \succeq 0 \end{array} \tag{56}$$

for some weight matrix $H \in \mathcal{S}^n$, some correlation matrix $C \in S^n$ and some integer $r \in [1, n]$, where $\mathrm{diag}(X)$ denotes the vector consisting of the diagonal entries of X, e is the all-ones vector and "$\circ$" denotes the Hadamard product (i.e., $(A \circ B)_{ij} = A_{ij}B_{ij}$, $i, j = 1, ..., n$). Recently, several methods have been proposed for solving problem (56) in the literature (see, for example, [37, 35, 3, 33, 15, 23]).

It is not hard to see that problem (56) is a special case of the general rank constraint problem (1) with $f(X) = \frac{1}{2}\|H \circ (X - C)\|_F^2$, $p = q = 0$, $\Omega = \mathcal{S}^n_+$, and

$$\mathcal{X} = \{X \in \mathcal{S}^n : \mathrm{diag}(X) = e\}.$$

Thus, the PD method proposed in Section 5 for problem (1) can be suitably applied to (56). Recall that the main computational parts of this method lie in solving the subproblems in steps 1a) and 1b). In the context of (56), these subproblems are in the form of

$$\begin{array}{l} \min\limits_{X}\{\|W \circ (X - A)\|_F^2 : X \in \mathcal{X}\}, \\ \min\limits_{Y}\{\|Y - B\|_F^2 : \mathrm{rank}(Y) \le r, \ Y \succeq 0\} \end{array} \tag{57}$$

for some A, B, $W \in \mathcal{S}^n$, respectively. In view of the above definition of $\mathcal{X}$ and Corollary 2.8, we can see that the above two problems have closed-form solutions.

We now address the initialization and termination criteria for our PD method when applied to (56). In particular, we choose $X^{\rm feas} = ee^T$, and Y_0^0 to be the solution of (57) by replacing B by C. In addition, we choose the initial penalty parameter ϱ_0 to be 1, and set the parameter $c = \sqrt{10}$. And we use (21) and

$$\frac{\|X^k - Y^k\|_F}{\max\{|Q_{\rho_k}(X^k, Y^k)|, 1\}} \le \epsilon_O$$

as the inner and outer termination criteria for the PD method and set their associated accuracy parameters $\epsilon_O = 10^{-5}$ and $\epsilon_I = \epsilon_O/2$.

Next we conduct numerical experiments to test the performance of our method for solving (56) on five classes of testing problems. These problems are widely used in literature (see, for example, [3, 23, 13, 34]). Their corresponding data matrix C and weight matrix H are defined as follows:

(P1) Set $n = 500$, $C_{ij} = 0.5 + 0.5\exp(-0.05|i-j|)$ for all i, j and $H = E$, where E is the all-ones matrix (see [3]).

(P2) The matrix C is a 1122×1122 correlation matrix corresponding to the 2611-days return (from December 31st, 2000 to January 2nd, 2011) extracted from the equities' data of Bloomberg (*http://www.bloomberg.com/professional/equities/*) and $H = E$.

(P3) The matrix C is the same as in (P1). The weight matrix H is generated in the same way as in [13, 34] such that all its entries are uniformly distributed in $[0.1, 10]$ except that the 2×100 submatrix in northwest corner has entries uniformly distributed $[0.01, 100]$.

(P4) The matrix C is the same as in (P2). The weight matrix H is generated in the same way as in (P3).

(P5) The matrix C is a 943×943 correlation matrix extracted from "movie-100K" data in the MovieLens data sets [16]. This data consists of 100,000 ratings from 943 users on 1682 movies and has been used in [13]. It shall be noted that such C may not be positive semidefinite because of missing data (see [12]). The weight matrix H is provided by T. Fushiki at the Institute of statistical Mathematics, Japan.

We now apply our PD method, the method Major[1] [33] and the method PenCorr [13] to solve problem (56) on the instances mentioned above. We set the parameters gradtol $= 10^{-5}$, tolrel $= 10^{-5}$ and ftol $= 10^{-6}$ for Major and set tolrel $= 10^{-5}$ for PenCorr. The computational results of all methods on the aforementioned instances are presented in Tables 4-8. It shall be mentioned that we choose to terminate the method Major when its CPU time exceeds 18000 seconds. In each table, the values of r are listed in the first column and the CPU time and the residue $\sqrt{2f(X)}$ of an approximate solution X for three methods are reported in the rest of columns, respectively. We observe that the residues $\sqrt{2f(X)}$ for PD and PenCorr are comparable to and generally smaller than those for Major. In addition, when $H = E$, PenCorr is generally the fastest method among these three methods. For relatively small r (say, $r \le 30$), Major generally outperforms PD in terms of speed, but PD substantially outperforms Major as r becomes larger. When $H \ne E$, both PenCorr and PD are faster than Major. Moreover, PenCorr is generally faster than PD, but for some instances the speed of PD is comparable or superior to that of PenCorr. Though PenCorr generally outperforms our PD method in terms of speed, it can only solve the rank minimization problems in the form of (1) while our PD method can be applied to both problems (1) and (2).

[1]The code for Major used in our paper is the one modified by Defeng Sun, Department of Mathematics, National University of Singapore.

Table 4: Comparison on problem (P1) with $n = 500$

Rank	Major		PenCorr		PD	
r	Time	Residue	Time	Residue	Time	Residue
5	3.9	7.883e+1	7.6	7.883e+1	22.6	7.883e+1
10	4.4	3.869e+1	4.8	3.869e+1	21.7	3.869e+1
15	6.0	2.325e+1	4.1	2.325e+1	19.3	2.325e+1
20	10.4	1.571e+1	4.2	1.571e+1	17.5	1.571e+1
25	16.6	1.145e+1	3.6	1.145e+1	16.8	1.145e+1
30	26.0	8.797e+0	3.4	8.796e+0	33.2	8.796e+0
35	44.6	7.020e+0	5.0	7.019e+0	29.7	7.019e+0
40	59.4	5.766e+0	3.3	5.765e+0	27.1	5.765e+0
45	83.6	4.843e+0	3.4	4.841e+0	24.6	4.841e+0
50	108.7	4.141e+0	1.8	4.139e+0	24.0	4.139e+0
60	181.7	3.156e+0	1.8	3.154e+0	19.2	3.154e+0
70	222.1	2.507e+0	1.7	2.504e+0	16.9	2.504e+0
80	329.2	2.053e+0	1.7	2.050e+0	15.1	2.050e+0
90	496.4	1.722e+0	1.8	1.718e+0	13.6	1.718e+0
100	664.2	1.471e+0	1.8	1.467e+0	12.8	1.466e+0
125	1364.5	1.055e+0	1.8	1.048e+0	8.9	1.048e+0

Table 5: Comparison on problem (P2)

Rank	Major		PenCorr		PD	
r	Time	Residue	Time	Residue	Time	Residue
10	58.5	2.000e+2	150.5	2.000e+2	360.3	2.000e+1
20	57.9	1.318e+2	125.3	1.318e+2	230.4	1.318e+2
30	115.0	1.025e+2	83.7	1.025e+2	183.0	1.025e+2
40	138.1	8.537e+1	69.0	8.535e+1	163.8	8.536e+1
50	215.3	7.383e+1	63.1	7.380e+1	159.3	7.381e+1
60	427.1	6.533e+1	57.5	6.531e+1	160.7	6.531e+1
70	404.7	5.874e+1	48.8	5.871e+1	163.4	5.871e+1
80	493.4	5.340e+1	48.4	5.338e+1	170.4	5.339e+1
90	665.7	4.899e+1	46.3	4.897e+1	176.7	4.897e+1
100	826.7	4.524e+1	41.6	4.523e+1	185.2	4.523e+1
120	1049.5	3.921e+1	39.8	3.919e+1	200.0	3.920e+1
140	1266.6	3.452e+1	37.1	3.450e+1	219.2	3.450e+1
160	1536.6	3.074e+1	35.2	3.071e+1	254.5	3.071e+1
180	1822.2	2.759e+1	35.4	2.756e+1	273.2	2.756e+1
200	2176.3	2.492e+1	36.4	2.490e+1	310.7	2.490e+1
250	2516.4	1.976e+1	36.1	1.973e+1	386.2	1.972e+1

7 Concluding remarks

In this paper we proposed penalty decomposition methods for general rank minimization problems in which each subproblem is solved by a block coordinate descent method. Under some suitable assumptions, we showed that any accumulation point of the sequence generated by the penalty decomposition methods satisfies the first-order optimality conditions of a nonlinear reformulation of the problems. The computational results on matrix completion and nearest low-rank correlation matrix problems demonstrate that our methods are generally comparable or superior to the existing methods in terms of solution quality. Though the speed of our methods is generally slower than the methods LMaFit [45] and PenCorr [13] on these problems, the latter methods are specifically developed for solving these problems while our methods are general solvers and can be applied to much broader class of problems.

Table 6: Comparison on problem (P3) with $n = 500$

Rank r	Major Time	Major Residue	PenCorr Time	PenCorr Residue	PD Time	PD Residue
5	56.1	8.971e+1	155.4	8.971e+1	171.5	8.972e+1
10	80.5	4.390e+1	103.8	4.393e+1	142.7	4.394e+1
15	128.6	2.622e+1	90.5	2.623e+1	134.5	2.623e+1
20	166.9	1.761e+1	86.2	1.762e+1	54.6	1.763e+1
25	245.2	1.275e+1	80.0	1.275e+1	64.6	1.275e+1
30	297.8	9.688e+0	78.9	9.689e+0	224.2	9.682e+0
35	386.1	7.676e+0	70.4	7.678e+0	207.9	7.667e+0
40	468.8	6.258e+0	81.3	6.252e+0	204.2	6.240e+0
45	639.4	5.225e+0	84.5	5.190e+0	236.3	5.180e+0
50	852.1	4.412e+0	86.9	4.386e+0	258.1	4.375e+0
60	1005.2	3.258e+0	81.1	3.254e+0	409.5	3.236e+0
70	1677.8	2.521e+0	86.0	2.503e+0	492.8	2.473e+0
80	2165.5	2.005e+0	83.4	1.991e+0	137.6	1.988e+0
90	2852.6	1.705e+0	90.3	1.606e+0	162.5	1.585e+0
100	3675.6	1.438e+0	82.6	1.328e+0	146.7	1.293e+0
125	5109.3	9.678e−1	98.4	8.664e−1	91.1	8.165e−1

Table 7: Comparison on problem (P4)

Rank r	Major Time	Major Residue	PenCorr Time	PenCorr Residue	PD Time	PD Residue
20	4225.3	1.378e+2	1263.9	1.380e+2	1258.6	1.384e+2
30	5103.5	1.025e+2	1903.9	1.029e+2	1519.8	1.030e+2
40	6151.4	8.252e+2	1759.8	8.263e+1	1484.7	8.244e+1
50	7261.4	6.905e+2	1853.3	6.909e+1	2033.2	6.873e+1
60	8548.9	6.935e+2	1888.7	5.932e+1	5463.7	5.900e+1
70	10025.6	5.163e+2	1662.8	5.161e+1	1619.2	5.177e+1
80	11426.1	4.569e+2	1885.1	4.561e+1	1853.5	4.553e+1
90	13008.9	4.098e+1	1477.7	4.092e+1	1798.6	4.053e+1
100	14234.7	3.809e+1	1715.5	3.683e+1	1873.6	3.651e+1
120	16012.7	3.227e+1	1554.6	3.058e+1	4130.9	2.942e+1
140	17730.0	2.808e+1	1435.9	2.580e+1	6275.8	2.341e+1
160	18000.0	2.470e+1	1556.5	2.203e+1	4964.3	1.978e+1
180	18000.0	2.195e+1	1446.6	1.906e+1	6548.9	1.662e+1
200	18000.0	1.965e+1	1392.1	1.661e+1	3621.5	1.466e+1
250	18000.0	1.558e+1	1222.6	1.205e+1	3302.1	1.065e+1

Table 8: Comparison on problem (P5)

Rank r	Major Time	Major Residue	PenCorr Time	PenCorr Residue	PD Time	PD Residue
20	3088.1	2.388e+2	1242.9	2.398e+2	1039.7	2.398e+2
40	5427.3	1.798e+2	725.2	1.803e+2	653.6	1.802e+2
60	8532.6	1.660e+2	584.7	1.659e+2	1430.2	1.658e+2
80	13168.4	1.632e+2	467.0	1.620e+2	1376.6	1.620e+2
100	18000.0	1.611e+2	449.7	1.611e+2	1291.4	1.610e+2
140	18000.0	1.611e+2	523.4	1.610e+2	1247.7	1.610e+2
180	18000.0	1.611e+2	525.7	1.610e+2	1194.8	1.610e+2
250	18000.0	1.610e+2	527.5	1.610e+2	1219.5	1.610e+2

We shall remark that the augmented Lagrangian decomposition methods can be developed for solving general rank minimization problems (1) and (2) simply by replacing the quadratic penalty functions in the PD methods by the augmented Lagrangian functions. Nevertheless, as observed in our experiments, their practical performance is generally inferior to the PD methods.

Acknowledgment

The authors would like to thank the two anonymous referees for their constructive comments which substantially improved the presentation of the paper.

References

[1] A. Ben-Tal and A. Nemirovski. Lectures on Modern Convex Optimization: Analysis, algorithms, Engineering Applications. *MPS-SIAM Series on Optimization*, SIAM, Philadelphia, PA, USA, 2001.

[2] R. Bhatia. Matrix Analysis. Springer, New York, 1997.

[3] D. Brigo. A note on correlation and rank reduction. Available at *www.damianobrigo.it*, 2002.

[4] D. Brigo and F. Mercurio. Interest Rate Models: Theory and Practice. Springer-Verlag, Berlin, 2001.

[5] S. Burer, R. D. C. Monteiro, and Y. Zhang. Maximum stable set formulations and heuristics based on continuous optimization. *Math. Program.*, 94:137-166, 2002.

[6] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optimiz*, 20(4):1956-1982, 2010.

[7] E. J. Candés and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 2009.

[8] W. Dai and O. Milenkovic. SET: an algorithm for consistent matrix completion. *Int. Conf. Acoust. Spee.*, 2010.

[9] L. Eldén. Matrix methods in data mining and pattern recognition (fundamentals of algorithms). SIAM, Philadelphia, PA, USA, 2009.

[10] M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. *P. Amer. Contr. Conf.*, 6:4734-4739, 2001.

[11] M. Fornasier, H. Rauhut, and R. Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM J. Optim.*, 21:1614–1640, 2011.

[12] T. Fushiki. Estimation of positive semidefinite correlation matrices by using convex quadratic semidefinite programming. *Neural Comput.*, 21:2028-2048, 2009.

[13] Y. Gao and D. Sun. A majorized penalty approach for calibrating rank constrained correlation matrix problems. Technical report, Department of Mathematics, National University of Singapore, 2010.

[14] M. X. Goemans and D. P. Williamson. .878-approximation algorithms for MAX CUT and MAX 2SAT. *Lect. Notes Comput. Sc.*, 422-431, 1994.

[15] I. Grubiŝić and R. Pietersz. Efficient rank reduction of correlation matrices. *Linear Algebra Appl.*, 422:629-653, 2007.

[16] J. L. Herlocker, J. A. Konstan, A. Borchers and J. Riedl. An algorithmic framework for performing collaborative filtering. SIGIR Forum., New York, NY, USA, 230-237, 1999.

[17] R. A. Horn and C. R. Johnson. Matrix Analysis. Cambridge University Press, New York, NY, USA, 1990.

[18] S. Ji, K.-F. Sze, Z. Zhou, A. So, and Y. Ye. Beyond convex relaxation: A polynomialtime nonconvex optimization approach to network localization. INFOCOM 2013.

[19] R. H. Keshavan and S. Oh. A gradient descent algorithm on the Grassman manifold for matrix completion. Technical report, Department of Electrical Engineering, Stanford University, 2009.

[20] M. Lai, Y. Xu, and W. Yin. Improved iteratively reweighted least squares for unconstrained smoothed l_q minimization. *SIAM J. Numer. Anal.*, 5(2):927–957, 2013.

[21] K. Lee and Y. Bresler. Admira: Atomic decomposition for minimum rank approximation. *IEEE T. Inform. Theory*, 56(9):4402-4416, 2010.

[22] A. S. Lewis. Derivatives of spectral functions. *Math. Oper. Res*, 21(3):576-588, 1996.

[23] Q. Li and H. Qi. A sequential semismooth Newton method for the nearest low-rank correlation matrix problem. *SIAM J. Optimiz*, 21:1641-1666, 2011.

[24] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM J. Matrix Anal. A.*, 31:1235-1256, 2009.

[25] Y. Liu, D. Sun, and K. C. Toh. An implementable proximal point algorithmic framework for nuclear norm minimization. *Math. Program.*, 1-38, 2009.

[26] Z. Lu, R. D. C. Monteiro, and M. Yuan. Convex optimization methods for dimension reduction and coefficient estimation in multivariate linear regression. *Math. Program.*, 131:163-194, 2012.

[27] Z. Lu and Y. Zhang. Iterative reweighted singular value minimization methods for l_p regularized unconstrained matrix minimization. Preprint, 2014.

[28] R. Mazumder, T. Hastie, and R. Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.*, 11:2287-2232, 2010.

[29] S. Ma, D. Goldfarb, and L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization. *Math. Program.*, 128(1):321-353, 2011.

[30] R. Meka, P. Jain and I. S. Dhillon. Guaranteed rank minimization via singular value projection. *NIPS*, 2010.

[31] K. Mohan and M. Fazel. Iterative reweighted least squares for matrix rank minimization. *48th Annual Allerton Conference on Communication, Control, and Computing*, 653-661, 2010.

[32] T. Mrita and T. Kanade. A sequential factorization method for recovering shape and motion from image streams. *IEEE T. Pattern Anal.*, 19:858-867, 1997.

[33] R. Pietersz and I. Grubiŝić. Rank reduction of correlation matrices by majorization. *Quant. Financ.*, 4:649-662, 2004.

[34] H. Qi and D. Sun. An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem. *IMA J. Numer. Anal.*, 31(2):491-511, 2011.

[35] F. Rapisarda, D. Brigo and F. Mercurio. Parameterizing correlations: a geometric interpretation. IMA J. Manag. Math., 18(1):55-73, 2007.

[36] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471-501, 2010.

[37] R. Rebonato. On the simultaneous calibration of multifactor lognormal interest rate models to Black volatilities and to the correlation matrix. *J. Comput. Financ.*, 2:5-27, 1999.

[38] R. Rebonato. Modern Pricing and Interest-Rate Derivatives. Princeton University Press, New Jersey, 2002.

[39] R. Rebonato. Interest-rate term-structure pricing models: a review. *P. R. Soc. Lond. A-Conta.*, 460:667-728, 2004.

[40] A. Ruszczyński. Nonlinear Optimization. Princeton University Press, 2006.

[41] K. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pac. J. Optim.*, 6:615-640, 2010.

[42] C. Tpmasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vision*, 9:137-154, 1992.

[43] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optimiz. Theory App.*, 109:475-494, 2001.

[44] E. van den Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. *SIAM J. Optimiz*, 21(4):1201-1229, 2011.

[45] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Math. Prog. Comp.*, 4:333-361, 2012.

[46] L. Wu. Fast at-the-money calibration of the LIBOR market model using Lagrangian multipliers. *J. Comput. Financ.*, 6:39-77, 2003.

[47] Z. Zhang and L. Wu. Optimal low-rank approximation to a correlation matrix. *Linear Algebra Appl.*, 364:161-187, 2003.