An exact penalty method for semidefinite-box constrained low-rank matrix optimization problems

Tianxiang Liu[*] Zhaosong Lu[†] Xiaojun Chen[‡] Yu-Hong Dai[§]

October 11, 2018

Abstract

This paper considers a matrix optimization problem where the objective function is continuously differentiable and the constraints involve a semidefinite-box constraint and a rank constraint. We first replace the rank constraint by adding a non-Lipschitz penalty function in the objective and prove that this penalty problem is exact with respect to the original problem. Next, for the penalty problem, we present a nonmonotone proximal gradient (NPG) algorithm whose subproblem can be solved by Newton's method with globally quadratic convergence. We also prove the convergence of the NPG algorithm to a first-order stationary point of the penalty problem. Furthermore, based on the NPG algorithm, we propose an adaptive penalty method (APM) for solving the original problem. Finally, the efficiency of APM is shown via numerical experiments for the sensor network localization (SNL) problem and the nearest low-rank correlation matrix problem.

Keywords: rank constrained optimization, non-Lipschitz penalty, nonmonotone proximal gradient, penalty method.

1 Introduction

In this paper we consider the following constrained problem

$$\begin{array}{ll} \min & f(X) \\ \text{s.t.} & 0 \preceq X \preceq I, \ \operatorname{rank}(X) \leq r, \end{array} \tag{1.1}$$

where $f : \mathcal{S}_+^n \to \Re$ is continuously differentiable with gradient ∇f being Lipschitz continuous, and $r < n$ is a given positive integer. Here, $\mathcal{S}_+^n$ denotes the cone of $n \times n$ positive semidefinite symmetric matrices, I is the $n \times n$ identity matrix, and $0 \preceq X \preceq I$ means $X \in \mathcal{S}_+^n$ and $I - X \in \mathcal{S}_+^n$,

[*] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (tiskyliu@polyu.edu.hk). This author was supported partly by the AMSS-PolyU Joint Research Institute Postdoctoral Scheme.

[†] Department of Mathematics, Simon Fraser University, Canada (zhaosong@sfu.ca). This author was supported in part by NSERC Discovery Grant.

[‡] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (xiaojun.chen@polyu.edu.hk). This author was supported in part by NSFC/Hong Kong Research Grant Council N-PolyU504/14.

[§] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China (dyh@lsec.cc.ac.cn). This author was supported in part by the Chinese Natural Science Foundation (Nos. 11631013 and 11331012) and the National 973 Program of China (Nos. 2015CB856002).

which is referred to as a semidefinite-box constraint. Many application problems can be modeled by (1.1), including the wireless sensor network localization problem [4, 14] and the nearest low-rank correlation matrix problem [5, 12, 21].

Problem (1.1) is generally difficult to solve, due to the discontinuity and nonconvexity of the rank function. Recently, approximations of the rank function have been extensively studied. One well-known convex approximation is the *nuclear norm* $\|X\|_*$, namely, the sum of singular values of X (see for example [10]). For other research works involving this approximation, see for example [7, 22, 23]. Besides, a nonconvex and nonsmooth approximation, the so-called *Schatten p-norm* $\|X\|_p^p = \sum_{i\geq 1} \sigma_i(X)^p$ ($p \in (0,1), \sigma_i(X)$ is the i-th largest singular value), has attracted a lot of attention due to its good computational performance (see for example [14, 20, 17]). However, simply adding these approximations into the objective genenerally cannot guarantee to produce a solution satisfying the rank constraint $\text{rank}(X) \leq r$ since they are not the exact penalty function for this constraint. Inspired by the relation

$$\text{rank}(X) \leq r \Leftrightarrow \sum_{i=r+1}^{n} \lambda_i^p(X) = 0 \quad \text{for} \quad X \succeq 0$$

and good computational performance of the p-norm with $p \in (0,1]$ for sparsity, we propose the following penalty model for problem (1.1):

$$\min_{0 \preceq X \preceq I} F_\mu(X) := f(X) + \mu \sum_{i=r+1}^{n} \lambda_i^p(X), \tag{1.2}$$

where $\mu > 0$ and $\lambda_i(X)$ $(i = 1, ..., n)$ is the i-th largest eigenvalue of X. Such a penalty term with $p = 1$ has been used in [11] for solving a nearest low-rank correlation matrix problem. Nevertheless, we observe in numerical experiments that the penalty term with $p \in (0,1)$ is generally more efficient than $p = 1$ in producing a low-rank solution of problem (1.1). The main contributions of this paper are as follows.

- We propose a new penalty model (1.2) for the low-rank constrained problem (1.1) and prove that (1.2) is an exact penalty reformulation for (1.1) in the sense: there exists some $\bar{\mu} > 0$ such that for any $\mu > \bar{\mu}$, X^* is a global minimizer of problem (1.1) if and only if it is a global minimizer of problem (1.2). Furthermore, for any $\mu \geq \bar{\mu}$, any local minimizer of problem (1.1) is a local minimizer of problem (1.2).

- We propose a nonmonotone proximal gradient (NPG) method for solving the penalty model (1.2). Although the associated proximal subproblem is sophisticated and challenging due to the partial set of eigenvalues, we reduce it into a set of univariate root-finding problems and show that they can be suitably solved by Newton's method with globally quadratic convergence.

- We propose an adaptive penalty method (APM) for (1.1) with a suitable updating scheme on penalty parameter in which each penalty subproblem is solved by the aforementioned NPG. We establish its global convergence and also provide an estimate on iteration complexity for finding an approximated stationary point of (1.1).

The rest of this paper is organized as follows. In Section 2, notation and preliminaries are given. In Section 3, we show that the penalty model (1.2) is an exact penalty reformulation of problem (1.1). In Section 4, we present an NPG algorithm for solving the penalty problem (1.2). In Section 5, we propose an APM for solving problem (1.1). In Section 6, we present numerical experiments for solving a sensor network localization problem and a nearest low-rank correlation matrix problem.

2 Notation and preliminaries

The following notation will be used throughout this paper. Given any $x \in \Re^n$, $x_{[i]}$ denotes the ith largest entry of x and $\mathrm{supp}(x)$ denotes the support of x, namely, $\mathrm{supp}(x) = \{i : x_i \neq 0\}$. The symbol $\mathbf{1}_n$ denotes the all-ones vector of dimension n. Given $x, y \in \Re^n$ and $\Omega \subseteq \Re^n$, $x \leq y$ means $x_i \leq y_i$ for all i and $\delta_\Omega(\cdot)$ is the indicator function of Ω, i.e., $\delta_\Omega(x) = 0$ if $x \in \Omega$, otherwise $\delta_\Omega(x) = \infty$. For $x \in \Re^n$ and a closed convex set $\Omega \subseteq \Re^n$, $\mathcal{P}_\Omega(x)$ is the projection of x onto Ω. The space of symmetric $n \times n$ matrices is denoted by $\mathcal{S}^n$. If $X \in \mathcal{S}^n$ is positive semidefinite, we write $X \succeq 0$. Given any X and Y in $\mathcal{S}^n$, $X \preceq Y$ means $Y - X$ is postive semidefinite. In addition, 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 $\|X\|_F := \sqrt{\mathrm{tr}(XX^T)}$. The identity matrix is denoted by I, and all-ones matrix is denoted by E, whose dimensions shall be clear from the context. For any $A, B \in \Re^{n \times n}$, "$\circ$" denotes the Hadamard product, i.e., $(A \circ B)_{ij} = A_{ij} B_{ij}$, $i, j = 1, ..., n$. For any $X \in \mathcal{S}^n$, we denote by $\lambda_i(X)$ $(i = 1, ..., n)$ the ith largest eigenvalue of X and write $\lambda(X) = (\lambda_1(X), ..., \lambda_n(X))^T$. We use $\|\cdot\|_F$ and $\|\cdot\|_2$ to denote the Frobenius norm and the Euclidean norm, respectively. In addition, $\mathcal{B}(X; \epsilon)$ stands for a ball in $\mathcal{S}^n$ centered at X with radius ϵ, that is, $\mathcal{B}(X; \epsilon) := \{Y \in \mathcal{S}^n : \|Y - X\|_F \leq \epsilon\}$.

Given $x \in \Re^n$ and $X \in \Re^{n \times n}$, $\mathrm{Diag}(x)$ and $\mathrm{diag}(X)$ denote an $n \times n$ diagonal matrix whose diagonal is formed by the vector x and vector extracted from the diagonal of X, respectively. For the sake of convenience, we use

$$\mathcal{C} := \{X \in \mathcal{S}^n : 0 \preceq X \preceq I\}, \qquad \Omega := \{X \in \mathcal{C} : \ \mathrm{rank}(X) \leq r\}, \tag{2.1}$$

to denote the feasible regions of problems (1.2) and (1.1), respectively. Given any $X \in \mathcal{S}^n$, let X_Ω be a projection of X onto Ω, that is, $X_\Omega \in \Omega$ and

$$\|X - X_\Omega\|_F = \min_{Z \in \Omega} \|X - Z\|_F. \tag{2.2}$$

Recall that f is assumed to be continuously differentiable in $\mathcal{C}$. It follows that f is Lipschitz continuous in $\mathcal{C}$, that is, there exists some constant $L_f > 0$ such that

$$|f(X) - f(Y)| \leq L_f \|X - Y\|_F, \qquad \forall X, Y \in \mathcal{C}. \tag{2.3}$$

Before ending this section, we present some preliminary technical results that will be used subsequently.

Lemma 2.1. *Let $p \in (0, 1]$ and X_Ω be a projection of X onto Ω. Then it holds*

$$\|X - X_\Omega\|_F \leq \sum_{i=r+1}^{n} \lambda_i^p(X), \qquad \forall X \in \mathcal{C}. \tag{2.4}$$

Proof. By Proposition 2.6 of [19], it is not hard to show that

$$\|X - X_\Omega\|_F = \sqrt{\sum_{i=r+1}^{n} \lambda_i^2(X)}, \quad \forall X \in \mathcal{C}. \tag{2.5}$$

Notice from (2.1) that $0 \leq \lambda_i(X) \leq 1$ for all i and $X \in \mathcal{C}$. In view of this fact and $p \in (0, 1]$, one can observe that

$$\sqrt{\sum_{i=r+1}^{n} \lambda_i^2(X)} \leq \sum_{i=r+1}^{n} \lambda_i(X) \leq \sum_{i=r+1}^{n} \lambda_i^p(X), \qquad \forall X \in \mathcal{C}. \tag{2.6}$$

It then follows from this relation and (2.5) that (2.4) holds as desired. This completes the proof. □

3 Exact penalty reformulation

In this section we study the relationship between the penalty model (1.2) and problem (1.1). The following theorem shows that (1.2) is an exact penalty reformulation of (1.1), in terms of global minimizers.

Theorem 3.1. *Let $p \in (0,1]$. For any $\mu \geq L_f$, any global minimizer of problem* (1.1) *is a global minimizer of problem* (1.2)*. Conversely, for any $\mu > L_f$, any global minimizer of* (1.2) *is also a global minimizer of* (1.1)*.*

Proof. For the first part, let X^* be a global minimizer of (1.1) and X be an arbitrary matrix in $\mathcal{C}$. We let X_Ω denote a projection of X onto Ω. Thus, we know from the global optimality of X^* that $f(X_\Omega) \geq f(X^*)$. Using this relation and (2.3), we have

$$\begin{aligned} f(X^*) - f(X) &= f(X^*) - f(X_\Omega) + f(X_\Omega) - f(X) \\ &\leq f(X_\Omega) - f(X) \ \leq \ L_f \|X - X_\Omega\|_F. \end{aligned} \tag{3.1}$$

This together with (2.4), $\mu \geq L_f$ and $\operatorname{rank}(X^*) \leq r$ implies that

$$f(X) + \mu \sum_{i=r+1}^{n} \lambda_i^p(X) \geq f(X) + L_f \|X - X_\Omega\|_F \geq \ f(X^*) = f(X^*) + \mu \sum_{i=r+1}^{n} \lambda_i^p(X^*),$$

which together with the arbitrariness of $X \in \mathcal{C}$ and $X^* \in \mathcal{C}$ implies that X^* is a global minimizer of (1.2).

For the second part, assume $\mu > L_f$. Let X^* be a global minimizer of problem (1.2) and X^*_Ω be a projection of X^* onto Ω. It is easy to observe that if $X^* \in \Omega$, then it is a global minimizer of problem (1.1). Thus it suffices to show that $X^* \in \Omega$. Suppose for contradiction that $X^* \notin \Omega$. Then we have $\|X^* - X^*_\Omega\|_F > 0$, and hence

$$\begin{aligned} f(X^*_\Omega) &\leq f(X^*) + L_f \|X^* - X^*_\Omega\|_F < f(X^*) + \mu \|X^* - X^*_\Omega\|_F \\ &\leq f(X^*) + \mu \sum_{i=r+1}^{n} \lambda_i^p(X^*) < f(X^*_\Omega), \end{aligned}$$

where the first inequality follows from (2.3), the second inequality is due to $\mu > L_f$, the third inequality is due to (2.4), and the last inequality follows from the global optimality of X^*. These inequalities immediately lead to a contradiction $f(X^*_\Omega) < f(X^*_\Omega)$. This completes the proof. □

We show in the next theorem that any local minimizer of problem (1.1) is also that of problem (1.2), provided $\mu \geq L_f$.

Theorem 3.2. *Let $p \in (0,1]$. For any $\mu \geq L_f$, any local minimizer of problem* (1.1) *is a local minimizer of problem* (1.2)*.*

Proof. Suppose that X^* is an arbitrary local minimizer of problem (1.1) with $\mu \geq L_f$. Then there exists some $\varepsilon > 0$ such that

$$f(X) \geq f(X^*), \quad \forall X \in \mathcal{B}(X^*; \varepsilon) \cap \Omega. \tag{3.2}$$

It follows from (2.2) that for every $X \in \mathcal{B}(X^*; \varepsilon/2)$,

$$\|X_\Omega - X^*\|_F \leq \|X_\Omega - X\|_F + \|X - X^*\|_F \leq 2\|X - X^*\|_F \leq \varepsilon,$$

where X_Ω is a projection of X onto Ω. This implies $X_\Omega \in \mathcal{B}(X^*;\varepsilon)\cap\Omega$ for every $X \in \mathcal{B}(X^*;\varepsilon/2)$. It follows from this and (3.2) that $f(X_\Omega) \geq f(X^*)$ for any $X \in \mathcal{B}(X^*;\varepsilon/2)$. Using this relation and (2.3), we see that (3.1) also holds for every $X \in \mathcal{B}(X^*;\varepsilon/2)\cap\mathcal{C}$. In view of (2.4), (3.1) and an argument similar to the proof of Theorem 3.1, one can obtain that for every $X \in \mathcal{B}(X^*;\varepsilon/2)\cap\mathcal{C}$,

$$f(X) + \mu\sum_{i=r+1}^{n}\lambda_i^p(X) \geq f(X) + L_f\|X - X_\Omega\|_F \geq f(X^*) = f(X^*) + \mu\sum_{i=r+1}^{n}\lambda_i^p(X^*),$$

where the equality is due to $\operatorname{rank}(X^*) \leq r$. Hence, X^* is a local minimizer of problem (1.2). This completes the proof. □

4 A nonmonotone proximal gradient method for solving (1.2)

In this section, we present a nonmonotone proximal gradient (NPG) method for solving problem (1.2), which is similar to the one proposed by Wright et al. [25]. We show that the subproblems arising in NPG can be efficiently solved. Also, we establish convergence for this method.

4.1 NPG algorithm and convergence

We first present an NPG method for solving problem (1.2).

Algorithm 1 Nonmonotone proximal gradient (NPG) method for (1.2)

Initialization. Let $0 < L_{\min} < L_{\max}$, $\gamma > 1$, $c > 0$, integer $N \geq 0$ be given. Choose an arbitrary $0 \preceq X^0 \preceq I$ and set $k = 0$.

Step 1. Choose $L_k^0 \in [L_{\min}, L_{\max}]$ arbitrarily. Set $L_k = L_k^0$.

(**1a**) Solve the subproblem

$$X^{k+1} \in \operatorname*{Arg\,min}_{0\preceq X\preceq I}\left\{\langle\nabla f(X^k), X - X^k\rangle + \frac{L_k}{2}\|X - X^k\|_F^2 + \mu\sum_{i=r+1}^{n}\lambda_i^p(X)\right\}. \tag{4.1}$$

(**1b**) Go to **Step 2** if

$$F_\mu(X^{k+1}) \quad \leq \max_{[k-N]_+\leq i\leq k} F_\mu(X^i) - \frac{c}{2}\|X^{k+1} - X^k\|_F^2. \tag{4.2}$$

(**1c**) Set $L_k \leftarrow \gamma L_k$ and go to (**1a**).

Step 2. Set $k \leftarrow k+1$ and go to **Step 1**.

Remark 4.1. *(i) When $N = 0$, the sequence $\{F_\mu(X^k)\}$ is monotonically decreasing. Otherwise, it may increase at some iterations and thus the above method is generally a nonmonotone method.*

(ii) A popular choice of L_k^0 is by the following formula proposed by Barzilai and Borwein [1] (see also [2]):

$$L_k^0 = \max\left\{L_{\min}, \min\left\{L_{\max}, \frac{\langle S^k, Y^k\rangle}{\|S^k\|_F^2}\right\}\right\}, \tag{4.3}$$

where $S^k = X^k - X^{k-1}$ and $Y^k = \nabla f(X^k) - \nabla f(X^{k-1})$.

We next study the convergence of the NPG method for solving problem (1.2). Before proceeding, we introduce two definitions as follows, which can be found in [24].

Definition 4.1 (limiting subdifferential)**.** *For a lower semi-continuous function g in $\mathcal{S}^n$, the limiting subdifferential of g at $X \in \mathcal{S}^n$ is defined as*

$$\partial g(X) := \left\{ V : \exists Z^k \xrightarrow{g} X, V^k \to V \text{ with } \liminf_{Z \to Z^k} \frac{g(Z) - g(Z^k) - \langle V^k, Z - Z^k \rangle}{\|Z - Z^k\|_F} \geq 0 \ \forall k \right\},$$

where $Z^k \xrightarrow{g} X$ means $Z^k \to X$ and $g(Z^k) \to g(X)$.

Definition 4.2 (first-order stationary point)**.** *We say that X^* is a first-order stationary point of* (1.2) *if $X^* \in \mathcal{C}$ and*

$$0 \in \nabla f(X^*) + \partial(\mu\Theta(X^*) + \delta_{\mathcal{C}}(X^*)), \tag{4.4}$$

where $\Theta(X) := \sum_{i=r+1}^{n} \lambda_i^p(X)$, $\mathcal{C}$ is defined in (2.1) *and $\partial(\cdot)$ is given in Definition* 4.1.

Notice from [24, Theorem 10.1] and [24, Exercise 10.10] that any local minimizer $\bar{X} \in \mathcal{C}$ of (1.2) is a first-order stationary point of (1.2). The following theorem states that at each outer iteration of Algorithm 1, the number of its inner iterations is uniformly bounded. Its proof is similar to that of [19, Theorem 4.2].

Theorem 4.1. *For each $k \geq 0$, the inner termination criterion* (4.2) *is satisfied after at most*

$$\max\left\{ \left\lfloor \frac{\log(L_{\nabla f} + c) - \log(L_{\min})}{\log \gamma} + 1 \right\rfloor, 1 \right\}$$

inner iterations, where $L_{\nabla f}$ is the Lipschitz constant associated with ∇f.

We next show that any accumulation point of $\{X^k\}$ is a first-order stationary point of problem (1.2).

Theorem 4.2. *Let the sequence $\{X^k\}$ be generated by Algorithm 1. The following statements hold:*

(i) $\|X^{k+1} - X^k\|_F \to 0$ as $k \to \infty$;

(ii) Any accumulation point of $\{X^k\}$ is a first-order stationary point of (1.2).

Proof. (i) The proof is similar to that of [25, Lemma 4].

(ii) Let $\bar{L}_k$ be the final value of L_k at the kth outer iteration. It follows from (4.1) that $\{\bar{L}_k\}$ is bounded. By the first-order optimality condition of (4.1), we have $X^{k+1} \in \mathcal{C}$ and

$$0 \in \nabla f(X^k) + \bar{L}_k(X^{k+1} - X^k) + \partial(\mu\Theta(X^{k+1}) + \delta_{\mathcal{C}}(X^{k+1})), \tag{4.5}$$

where $\Theta(X) := \sum_{i=r+1}^{n} \lambda_i^p(X)$. Notice that $\{X^k\} \subset \mathcal{C}$ and $\mathcal{C}$ is bounded. Hence, $\{X^k\}$ is bounded and it has at least an accumulation point, say X^*. Let $\mathcal{K}$ be a subsequence index such that $\{X^k\}_{\mathcal{K}} \to X^*$, which together with $\{X^k\} \subset \mathcal{C}$ and $\|X^{k+1} - X^k\|_F \to 0$ implies that $X^* \in \mathcal{C}$ and $\{X^{k+1}\}_{\mathcal{K}} \to X^*$. Using this, the boundedness of $\{\bar{L}_k\}$, the continuity of ∇f, and the outer semi-continuity of $\partial(\mu\Theta + \delta_{\mathcal{C}})$ [24, Proposition 8.7], and taking limits on both sides of (4.5) as $k \in \mathcal{K} \to \infty$, we have

$$0 \in \nabla f(X^*) + \partial(\mu\Theta(X^*) + \delta_{\mathcal{C}}(X^*)).$$

Hence, X^* is a first-order stationary point of (1.2). This completes the proof. □

4.2 An efficient algorithm for solving subproblem (4.1)

In this subsection we propose an efficient algorithm for solving subproblem (4.1). To proceed, we first consider the parametric univariate optimization problem

$$\min_{0\leq z\leq 1} \left\{ \Phi(z,t) := \frac{1}{2}(z-t)^2 + \nu z^p \right\} \tag{4.6}$$

for $\nu > 0$ and $p \in (0,1]$. Clearly, problem (4.6) has at least one optimal solution z^* and $\Phi(z^*,t)$ is well defined for any $t \in (-\infty,\infty)$. In addition, it is not hard to see that for $p=1$, problem (4.6) has a unique optimal solution $z^* = \min(1,\max(t-\nu,0))$. In what follows, we study some properties of the optimal solution set of (4.6) for $p \in (0,1)$.

Lemma 4.1. *Let $\mathcal{Z}^*(t)$ denote the set of optimal solutions of problem* (4.6) *for $t \in (-\infty,\infty)$ and $p \in (0,1)$. Let*

$$\alpha := \min\left\{[2(1-p)\nu]^{\frac{1}{2-p}}, 1\right\}, \qquad \beta := [\nu p(1-p)]^{\frac{1}{2-p}}, \tag{4.7}$$

$$t_1 := \frac{\alpha}{2} + \nu\alpha^{p-1}, \qquad t_2 := \max\left\{\frac{1}{2}+\nu, 1+\nu p\right\}. \tag{4.8}$$

Then the following statements hold:

(i) $0 \in \mathcal{Z}^*(t)$ *if and only if* $t \leq t_1$*;*

(ii) $1 \in \mathcal{Z}^*(t)$ *if and only if* $t \geq t_2$*;*

(iii) $\mathcal{Z}^*(t) = \{z^*\} \subseteq [\beta, \min\{t,1\})$ *if and only if* $t \in (t_1,t_2)$*, where z^* is the unique root of the equation*

$$g(z) := z - t + \nu p z^{p-1} = 0 \tag{4.9}$$

in the interval $[\beta,\infty)$*.*

The proof of this Lemma is given in Appendix. As an immediate consequence of Lemma 4.1, we obtain the following formula for computing an optimal solution of problem (4.6) for $p \in (0,1)$.

Corollary 4.1. *Let $\mathcal{Z}^*(t)$ denote the set of optimal solutions of problem* (4.6) *for $t \in (-\infty,\infty)$ and $p \in (0,1)$. Let β, t_1 and t_2 be defined in* (4.7) *and* (4.8)*, respectively. Then we have $z^*(t) \in \mathcal{Z}^*(t)$, where $z^* : \Re \to [0,1]$ is defined as follows:*

$$z^*(t) = \begin{cases} 0 & \text{if } t \leq t_1, \\ \tilde{z}^* & \text{if } t_1 < t < t_2, \\ 1 & \text{otherwise,} \end{cases} \tag{4.10}$$

where $\tilde{z}^*$ is the unique root of equation (4.9) *in* $[\beta,\infty)$*.*

As seen from (4.10), the value of $z^*(t)$ is precisely known for $t \leq t_1$ or $t \geq t_2$. Nevertheless, for $t \in (t_1,t_2)$, the exact value of $z^*(t)$ is typically unknown since equation (4.9) generally does not have a closed-form root. We next present an efficient numerical scheme for estimating the root $\tilde{z}^*$ of equation (4.9) by Newton's method.

Newton's method for solving (4.9):

Let β, t_1, t_2 and $g(\cdot)$ be defined in (4.7), (4.8) and (4.9), respectively. Let $t \in (t_1,t_2)$ be given. If $g(\beta)=0$, set $\tilde{z}^* = \beta$. Otherwise choose $z_0 \in (\beta,\infty)$ and perform

$$z_{k+1} = z_k - g(z_k)/g'(z_k) \quad \text{for} \quad k \geq 0. \tag{4.11}$$

Remark 4.2. *Recall from Lemma 4.1 (iii) that the unique root $\tilde{z}^*$ of equation* (4.9) *in* $[\beta,\infty)$ *lies in* $[\beta,\min\{t,1\})$. *Therefore, for practical efficiency, it is natural to choose* $z_0 = (\beta+\min\{t,1\})/2$.

The following theorem shows that the above Newton's method is able to find an approximate root in $[\beta,\infty)$ to equation (4.9), and moreover, it is globally and quadratically convergent.

Theorem 4.3. *Let β, t_1 and t_2 be defined in* (4.7) *and* (4.8)*, respectively. Then for any $t \in (t_1, t_2)$ and $p \in (0,1)$, Newton's method given above either finds the root $\tilde{z}^*$ of equation* (4.9) *or generates a sequence $\{z_k\}$ which is globally and quadratically convergent to $\tilde{z}^*$, and in particular,*

$$0 \le z_{k+1} - \tilde{z}^* \le \frac{\nu p(1-p)(2-p)(\tilde{z}^*)^{p-3}}{1-\nu p(1-p)(\tilde{z}^*)^{p-2}}(z_k - \tilde{z}^*)^2, \qquad \forall k \ge 1.$$

Proof. In view of Corollary 4.1, we know that for any $t \in (t_1,t_2)$ and $p \in (0,1)$, equation (4.9) has a unique root $\tilde{z}^*$ in $[\beta,\infty)$. Therefore, if $g(\beta)=0$, then $\tilde{z}^*=\beta$. Otherwise, $\tilde{z}^*$ is the unique root of (4.9) in (β,∞) and Newton's iterartion (4.11) generates a sequence $\{z_k\}$. We have from (4.9) that

$$g'(z) = 1-\nu p(1-p)z^{p-2}, \qquad g''(z) = \nu p(1-p)(2-p)z^{p-3}. \tag{4.12}$$

Notice that $g'(\beta)=0$ and $\beta>0$. It is easy to see that $g'(z)>0$, $g''(z)>0$ and $g''(z)$ is continuous for every $z \in (\beta,\infty)$. Hence, the assumptions of Lemma A.1 hold for $q=g$, $a=\beta$ and $z_*=\tilde{z}^*$. In addition, one can observe from (4.12) that $g'(\tilde{z}^*) = 1-\nu p(1-p)(\tilde{z}^*)^{p-2}$ and $\max\limits_{z\in[\tilde{z}^*,z_1]} g''(z) = \nu p(1-p)(2-p)(\tilde{z}^*)^{p-3}$. Therefore, the conclusion follows directly from Lemma A.1. This completes the proof. □

The proof of the following lemma is by a similar approach as proposed in [18].

Lemma 4.2. *Let $p \in (0,1]$ and $\nu > 0$ be given, and let*

$$V(t) := \underbrace{\min_{0\le z\le 1}\left\{\frac{1}{2}(z-t)^2+\nu z^p\right\}}_{V_1(t)} - \underbrace{\min_{0\le z\le 1}\left\{\frac{1}{2}(z-t)^2\right\}}_{V_2(t)}, \quad \forall t \in \Re. \tag{4.13}$$

Then $V(t)$ is increasing in $(-\infty,\infty)$.

Proof. It is not hard to observe from (4.13) that V_1, V_2 and V are well defined. Also, we see from (4.6) that

$$V_1(t) = \min_{0\le z\le 1}\Phi(z,t), \tag{4.14}$$

where Φ is defined in (4.6). Notice that Φ and $\nabla_t\Phi$ are continuous in $[0,\infty)\times\Re$. Moreover, $|\nabla_t\Phi(z,t)| = |t-z| \le |t|+1, \forall z \in [0,1]$. Hence, Φ is locally Lipschitz in t, uniformly for all $z \in [0,1]$. Using these facts, one can observe that the assumptions of [9, Theorem 2.1] hold for $g=\Phi$ and $U=[0,1]$. Thus it follows from [9, Theorem 2.1] that V_1 is locally Lipschitz continuous in $\Re$ and moreover

$$\partial V_1(t) = \mathrm{conv}(\{\nabla_t\Phi(z,t) : z \in \mathcal{Z}^*(t)\}) = \mathrm{conv}(t-\mathcal{Z}^*(t)), \tag{4.15}$$

where ∂V_1 denotes the Clarke subdifferential of V_1, $\mathrm{conv}(\cdot)$ is the convex hull of the associated set, and $\mathcal{Z}^*(t)$ denotes the set of optimal solutions of (4.14). Notice that V_2 is differentiable and moreover $\nabla V_2(t) = t-\mathcal{P}_{[0,1]}(t), \forall t \in \Re$. It then follows from the above two equalities that

$$\partial V(t) = \partial V_1(t) - \nabla V_2(t) = \mathrm{conv}\left(\mathcal{P}_{[0,1]}(t)-\mathcal{Z}^*(t)\right). \tag{4.16}$$

Since V_1 is locally Lipschitz continuous in $\Re$, V_1 is differentiable almost everywhere and so is V. Let $t \in \Re$ be such that V_1 is differentiable at t. It is not hard to observe from (4.15) that $\mathcal{Z}^*(t)$ contains a singleton. Moreover, V is also differentiable at t. We next show that $\nabla V(t) \geq 0$ by considering three separate cases as follows.

Case 1): $t \leq 0$. It follows from Lemma 4.1 (i) that $0 \in \mathcal{Z}^*(t)$. Also, $\mathcal{P}_{[0,1]}(t) = 0$ for $t \leq 0$.

Case 2): $t \in (0,1)$. This together with the definition of $\mathcal{Z}^*(t)$ implies that for any $z^* \in \mathcal{Z}^*(t)$, one has $\nu(z^*)^p \leq \frac{1}{2}(z^* - t)^2 + \nu(z^*)^p \leq \frac{1}{2}(t-t)^2 + \nu t^p = \nu t^p$, which yields $z^* \leq t$. Hence, $\mathcal{Z}^*(t) \subset [t,1]$. Also, $\mathcal{P}_{[0,1]}(t) = t$ for $t \in (0,1)$.

Case 3): $t > 1$. Clearly $\mathcal{Z}^*(t) \subseteq [0,1]$ and $\mathcal{P}_{[0,1]}(t) = 1$ for such a t.

In view of these observations, (4.16) and the differentiability of V at t, one can see that $\nabla V(t) \geq 0$. Hence, V has nonnegative derivative almost everywhere. Since V_1 is locally Lipschitz continuous and V_2 is differentiable in $\Re$, V is locally Lipschitz continuous in $\Re$. Thus V is absolutely continuous in any compact set. It follows from this and the fact that V has nonnegative derivative almost everywhere that V is increasing in $\Re$ (see, for example, [6, p. 120]). This completes the proof. □

Lemma 4.3. *Let $p \in (0,1]$, $d \in \Re^n$ and $\nu > 0$ be given. Consider the problem*

$$\vartheta^* = \min_{\mathbf{0} \leq x \leq \mathbf{1}_n} \frac{1}{2}\|x - d\|_2^2 + \nu \sum_{i=r+1}^{n} x_{[i]}^p. \tag{4.17}$$

Let Γ be an index set in $\{1,\ldots,n\}$ of size $n-r$ corresponding to the $n-r$ smallest entries of d. In addition, let $z^*(\cdot)$ be defined in Corollary 4.1, and $x^* \in \Re^n$ be defined as follows:

$$x_i^* = \begin{cases} z^*(d_i) & \text{if } i \in \Gamma, \\ \mathcal{P}_{[0,1]}(d_i) & \text{otherwise.} \end{cases} \tag{4.18}$$

Then x^* is an optimal solution of problem (4.17).

Proof. Let $S = \{s \in \{0,1\}^n : \sum_{i=1}^n s_i = n-r\}$. Observe that $\sum_{i=r+1}^n x_{[i]}^p = \min\limits_{s \in S} \sum_{i=1}^n s_i x_i^p$. It follows from this and (4.17) that

$$\vartheta^* = \min_{\mathbf{0} \leq x \leq \mathbf{1}_n} \min_{s \in S} \left\{ \frac{1}{2}\|x - d\|_2^2 + \nu \sum_{i=1}^n s_i x_i^p \right\} = \min_{s \in S} \underbrace{\min_{\mathbf{0} \leq x \leq \mathbf{1}_n} \left\{ \frac{1}{2}\|x - d\|_2^2 + \nu \sum_{i=1}^n s_i x_i^p \right\}}_{\psi(s)}. \tag{4.19}$$

Observe from (4.13) and (4.19) that

$$\psi(s) = \sum_{i \in \mathrm{supp}(s)} V_1(d_i) + \sum_{i \notin \mathrm{supp}(s)} V_2(d_i), \qquad \forall s \in S. \tag{4.20}$$

Let $s^* \in \{0,1\}^n$ be defined as follows:

$$s_i^* = \begin{cases} 1 & \text{if } i \in \Gamma, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $s^* \in S$. We first show that s^* is an optimal solution of problem (4.19). Let $\tilde{s}^* \in S$ be an arbitrary optimal solution of (4.19). We divide the rest of the proof into two separate cases as follows.

Case 1): $d_j \leq d_{[r+1]}$ for every $j \in \mathrm{supp}(\tilde{s}^*)$, that is, $\mathrm{supp}(\tilde{s}^*)$ is an index set corresponding to the $n-r$ smallest entries of d. It is not hard to observe from (4.20) that $\psi(\tilde{s}^*) = \psi(s^*)$. Hence, s^* is an optimal solution of (4.19).

Case 2): $d_j > d_{[r+1]}$ for some $j \in \mathrm{supp}(\tilde{s}^*)$. Let $\ell \in \mathrm{Arg\,min}\{d_i : i \notin \mathrm{supp}(\tilde{s}^*)\}$. It is not hard to observe that $d_\ell < d_j$. Let $\hat{s}^* \in \{0,1\}^n$ be defined as follows:

$$\hat{s}_i^* = \begin{cases} 1 & \text{if } i \in \mathrm{supp}(\tilde{s}^*) \cup \{\ell\} \setminus \{j\}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.21}$$

It follows from (4.13), (4.20) and (4.21) that

$$\begin{aligned} \psi(\hat{s}^*) &= \sum_{i \in \mathrm{supp}(\tilde{s}^*)} V_1(d_i) + \sum_{i \notin \mathrm{supp}(\tilde{s}^*)} V_2(d_i) + [V_1(d_\ell) - V_1(d_j) + V_2(d_j) - V_2(d_\ell)], \\ &= \psi(\tilde{s}^*) + V(d_\ell) - V(d_j), \end{aligned}$$

where V is defined in (4.13). This relation together with $d_\ell < d_j$ and Lemma 4.2 implies $\psi(\hat{s}^*) \le \psi(\tilde{s}^*)$. Using this and the fact that $\tilde{s}^*$ is an optimal solution of (4.19), we see that $\hat{s}^*$ is also an optimal solution of (4.19). Repeating the above process by replacing $\tilde{s}^*$ by $\hat{s}^*$ for a finite number of times, we reach an optimal solution $\bar{s}^*$ of (4.19) for which $d_j \le d_{[r+1]}$ for every $j \in \mathrm{supp}(\bar{s}^*)$. This means that Case 1) holds at $\bar{s}^*$. Thus the conclusion also holds due to Case 1).

Finally, since s^* is an optimal solution of (4.19), we have $\psi(s^*) = \vartheta^*$. By Corollary 4.1 and the definitions of x^* and s^*, one can observe that

$$x^* \in \mathop{\mathrm{Arg\,min}}_{\mathbf{0} \le x \le \mathbf{1}_n} \left\{ \frac{1}{2}\|x - d\|_2^2 + \nu \sum_{i=1}^n s_i^* x_i^p \right\},$$

which together with (4.19), $\psi(s^*) = \vartheta^*$ and $s^* \in S$ implies that

$$\vartheta^* = \psi(s^*) = \frac{1}{2}\|x^* - d\|_2^2 + \nu \sum_{i=1}^n s_i^* (x_i^*)^p \ \ge \ \frac{1}{2}\|x^* - d\|_2^2 + \nu \sum_{i=r+1}^n (x_{[i]}^*)^p.$$

It follows from this, $\mathbf{0} \le x^* \le \mathbf{1}_n$ and the definition of ϑ^* that x^* is an optimal solution of (4.17). This completes the proof. □

We are now ready to show how subproblem (4.1) arising in Algorithm 1 is solved. For convenience, we define the set-valued proximal operator as follows:

$$\mathrm{Prox}_{\nu\Theta}(Y) := \mathop{\mathrm{Arg\,min}}_{0 \preceq X \preceq I} \ \frac{1}{2}\|X - Y\|_F^2 + \nu \sum_{i=r+1}^n \lambda_i^p(X), \tag{4.22}$$

where $\Theta(X) = \sum_{i=r+1}^n \lambda_i^p(X)$. One can observe that (4.1) can be rewritten as

$$X^{k+1} \in \mathop{\mathrm{Arg\,min}}_{0 \preceq X \preceq I} \ \frac{1}{2}\left\| X - \left(X^k - \frac{\nabla f(X^k)}{L_k} \right) \right\|_F^2 + \frac{\mu}{L_k} \sum_{i=r+1}^n \lambda_i^p(X), \tag{4.23}$$

which is a special case of (4.22). It then follows that

$$X^{k+1} \in \mathrm{Prox}_{\frac{\mu}{L_k}\Theta}\left(X^k - \frac{\nabla f(X^k)}{L_k} \right).$$

In order to solve (4.1), it thus suffices to solve (4.22). We next show that (4.22) can be solved by a vector optimization problem.

Theorem 4.4. *Given $Y \in \mathcal{S}^n$, let $U\mathrm{Diag}(d)U^T$ be the eigenvalue decomposition of Y, and let x^* be an optimal solution to problem* (4.17)*. Then $U\mathrm{Diag}(x^*)U^T$ is an optimal solution to problem* (4.22).

Proof. Observe that $\|\cdot\|_F$ is a unitarily invariant norm, $\sum_{i=r+1}^n \lambda_i^p(\cdot)$ is a unitary similarity invariant function in $\mathcal{S}^n$, and $\{X : 0 \preceq X \preceq I\}$ is a unitary similarity invariant set. In addition, $t^2/2$ is an increasing function in $[0, \infty)$. Therefore, the assumptions of [19, Proposition 2.6] hold with $\|\cdot\| = \|\cdot\|_F$, $A = Y$, and

$$F(\cdot) = \sum_{i=r+1}^{n} \lambda_i^p(\cdot), \quad \mathcal{X} = \{X : 0 \preceq X \preceq I\}, \quad \phi(\cdot) = (\cdot)^2/2.$$

It then follows from [19, Proposition 2.6] that $U\mathrm{Diag}(\tilde{x}^*)U^T$ is an optimal solution of problem (4.22), where $\tilde{x}^*$ is any optimal solution of the problem

$$\begin{array}{ll} \min\limits_{x\in R^n} & \frac{1}{2}\|\mathrm{Diag}(x) - \mathrm{Diag}(d)\|_F^2 + \nu \sum\limits_{i=r+1}^{n} \lambda_i^p(\mathrm{Diag}(x)) \\ \text{s.t.} & 0 \preceq \mathrm{Diag}(x) \preceq I. \end{array} \tag{4.24}$$

It is not hard to observe that problem (4.24) is equivalent to (4.17), namely, they share exactly the same optimal solutions. Since x^* is an optimal solution of (4.17), x^* is also that of (4.24). The conclusion of this theorem thus holds due to the above observation with $\tilde{x}^* = x^*$. This completes the proof. □

Based on the above discussion, we now present an algorithm for finding an element in $\mathrm{Prox}_{\nu\Theta}(Y)$ for a given $Y \in \mathcal{S}^n$.

Algorithm 2 Algorithm for finding an element in $\mathrm{Prox}_{\nu\Theta}(Y)$

```
Input: $\nu$, $Y$.
Output: $X^* \in \mathrm{Prox}_{\nu\Theta}(Y)$.
Step 1. Do eigenvalue decomposition: $Y = U\mathrm{Diag}(d)U^T$.
Step 2. Use (4.18) in Lemma 4.3 to find
        $x^* \in \mathrm{Arg}\min_{\mathbf{0}\le x\le \mathbf{1}_n} \frac{1}{2}\|x - d\|_2^2 + \nu \sum_{i=r+1}^{n} x_{[i]}^p.$
Step 3. Let $X^* = U\mathrm{Diag}(x^*)U^T$.
```

Thus, we can find X^{k+1} in (4.1) in Algorithm 1 by Algorithm 2 with $\nu = \frac{\mu}{L_k}$ and $Y = X^k - \frac{\nabla f(X^k)}{L_k}$.

5 An adaptive penalty method for solving problem (1.1)

In this section we propose an adaptive penalty method for solving problem (1.1). Recall from Theorem 3.1, a global minimizer of (1.1) can be obtained by finding a global minimizer of (1.2) for a sufficiently large μ. Though an upper bound for such a μ is estimated in Theorem 3.1, it may be computationally inefficient to solve (1.2) once by choosing μ as this upper bound. Instead, it is natural to solve a sequence of problems in the form of (1.2) in which μ gradually increases. This scheme is commonly used in the classical penalty method and also a penalty method recently proposed in [8] for a non-Lipschitz optimization problem. We now present this scheme for solving problem (1.1) as follows.

Algorithm 3 An adaptive penalty method (APM) for problem (1.1)

Initialization. Let $p \in (0,1]$, $\epsilon > 0$ be given and X^{feas} be an arbitrary feasible point of problem (1.1). Choose $0 \preceq X^0 \preceq I$, $\mu_0 > 0$ and $\tau > 1$ arbitrarily. Set $k = 0$.

Step 1. If $F_{\mu_k}(X^k) > F_{\mu_k}(X^{\text{feas}})$, set $X^{k,0} = X^{\text{feas}}$. Otherwise, set $X^{k,0} = X^k$.

Step 2. Apply Algorithm 1 to (1.2) with $\mu = \mu_k$ starting with $X^{k,0}$ to generate $\{X^{k,j}\}$ until finding some X^{k,n_k} such that

$$\|\nabla f(X^{k,n_k-1}) - \nabla f(X^{k,n_k}) + L_{k,n_k-1}(X^{k,n_k} - X^{k,n_k-1})\|_F \leq \epsilon. \tag{5.1}$$

Set $X^{k+1} = X^{k,n_k}$.

Step 3. If $\sum_{i=r+1}^{n} \lambda_i^p(X^{k+1}) \leq \epsilon$, terminate the algorithm.

Step 4. Set $\mu_{k+1} = \tau\mu_k$, $k \leftarrow k+1$ and go to **Step 1**.

Remark 5.1. *Notice from Theorem 4.2 that $\|X^{k,j+1} - X^{k,j}\|_F \to 0$ as $j \to \infty$. In addition, one can observe from Theorem 4.1 that $\{L_{k,j}\}$ is bounded. Also, by the Lipschitz continuity of ∇f, one has*

$$\|\nabla f(X^{k,j+1}) - \nabla f(X^{k,j})\|_F \leq L_{\nabla f}\|X^{k,j+1} - X^{k,j}\|_F.$$

It then follows that inequality (5.1) *must hold for at some $j = n_k$.*

We next establish some convergence properties of Algorithm 3.

Theorem 5.1. *Suppose that the sequence $\{X^k\}$ is generated by Algorithm 3. Then the following statements hold.*

(i) After at most $\max\left\{\left\lfloor \frac{\log(f(X^{\text{feas}})-\underline{f})-\log(\mu_0\epsilon)}{\log\tau} + 1\right\rfloor, 1\right\}$ *iterations, Algorithm 3 generates some X^k satisfying*

$$\sum_{i=r+1}^{n} \lambda_i^p(X^k) \leq \epsilon, \qquad \operatorname{dist}(0, \nabla f(X^k) + \partial(\mu_{k-1}\Theta(X^k) + \delta_{\mathcal{C}}(X^k))) \leq \epsilon \tag{5.2}$$

for some $\mu_{k-1} > 0$, where $\Theta(X) = \sum_{i=r+1}^{n} \lambda_i^p(X)$ and $\mathcal{C}$ is defined in (2.1).

(ii) Let X_Ω^k be a projection of the above X^k onto Ω, where Ω is the feasible region of (1.1). *Then X_Ω^k satisfies*

$$\|X^k - X_\Omega^k\|_F \leq \epsilon, \qquad f(X_\Omega^k) \;\leq\; f(X^k) + L_f\epsilon. \tag{5.3}$$

Proof. (i) One can observe from Algorithm 1 that $F_{\mu_k}(X^{k,j}) \leq F_{\mu_k}(X^{k,0}), \forall k, j$. By the specific choice of $X^{k,0}$, we know that $F_{\mu_k}(X^{k,0}) \leq F_{\mu_k}(X^{\text{feas}})$. Since X^{feas} is a feasible point of (1.1), one can see that $F_{\mu_k}(X^{\text{feas}}) = f(X^{\text{feas}})$. It then follows that

$$F_{\mu_k}(X^{k,j}) \leq f(X^{\text{feas}}), \qquad \forall k, j,$$

which together with (1.2) yields $f(X^{k,j}) + \mu_k \sum_{i=r+1}^{n} \lambda_i^p(X^{k,j}) \leq f(X^{\text{feas}}), \forall k, j$. Using this and the fact $\mu_k = \mu_0\tau^k$, we obtain that

$$\sum_{i=r+1}^{n} \lambda_i^p(X^{k,j}) \;\leq\; \frac{f(X^{\text{feas}}) - \underline{f}}{\mu_k} \;=\; \frac{f(X^{\text{feas}}) - \underline{f}}{\mu_0\tau^k}, \qquad \forall k, j,$$

where $\underline{f} = \min\{f(X) : 0 \preceq X \preceq I\}$. It follows from this and $X^{k+1} = X^{k,n_k}$ that $\sum_{i=r+1}^n \lambda_i^p(X^{k+1}) \leq \epsilon$ when

$$k \geq \frac{\log(f(X^{\text{feas}}) - \underline{f}) - \log(\mu_0 \epsilon)}{\log \tau}.$$

We next show that the second relation of (5.2) holds for any $k \geq 1$. Since X^{k,n_k} is an optimal solution of problem (4.1) with μ, X^k and L_k replaced by μ_k, X^{k,n_k-1} and L_{k,n_k-1}, by the first-order optimality condition we have

$$0 \in \nabla f(X^{k,n_k-1}) + L_{k,n_k-1}(X^{k,n_k} - X^{k,n_k-1}) + \partial(\mu_k \Theta(X^{k,n_k}) + \delta_{\mathcal{C}}(X^{k,n_k})), \quad \forall k \geq 0,$$

where $\mathcal{C}$ is defined in (2.1). This together with (5.1) and $X^{k+1} = X^{k,n_k}$ yields

$$\text{dist}(0, \nabla f(X^{k+1}) + \partial(\mu_k \Theta(X^{k+1}) + \delta_{\mathcal{C}}(X^{k+1})) \leq \epsilon, \quad \forall k \geq 0.$$

This proves statement (i).

(ii) Notice that $X^k \in \mathcal{C}$, where $\mathcal{C}$ is defined in (2.1). This together with the definition of X_Ω^k, (2.4) and (5.2) yields $\|X^k - X_\Omega^k\|_F \leq \sum_{i=r+1}^n \lambda_i^p(X^k) \leq \epsilon$. Hence, the first relation of (5.3) holds. It follows from this relation and (2.3) that

$$f(X_\Omega^k) \ \leq \ f(X^k) + L_f \|X^k - X_\Omega^k\|_F \ \leq \ f(X^k) + L_f \epsilon.$$

The second relation of (5.3) thus holds. □

Remark 5.2. *Observe that problem* (1.1) *is equivalent to*

$$\min_X \{f(X) : \Theta(X) \leq 0, \ X \in \mathcal{C}\}. \tag{5.4}$$

The point X^k satisfying (5.2) *can be viewed as an approximate KKT point to* (5.4)*. Since X_Ω^k is a feasible point of* (1.1)*, and moreover $\|X^k - X_\Omega^k\|_F \leq \epsilon$ and $f(X_\Omega^k) \leq f(X^k) + L_f \epsilon$, then X_Ω^k can be viewed as a feasible approximate "KKT" point to problem* (5.4)*.*

6 Numerical Simulations

In this section, we apply aforementioned methods to the spherical sensor localization problem [13, 26] and the nearest low-rank correlation matrix problem [5, 12, 15, 16, 21]. All the numerical experiments are performed in Matlab R2016a on a 64-bit PC with an Intel(R) Core(TM) i7-6700 CPU (3.41GHz) and 32GB of RAM.

6.1 Spherical sensor localization

Suppose that there are n sensor points $\mathbf{x}_i \in \mathbb{S}^2$ $(i = 1, ..., n)$, where $\mathbb{S}^2 = \{x \in \Re^3 : \|x\|_2 = 1\}$ is the unit sphere. The last m sensors points are called anchors, whose positions are known. We denote these anchor points as $\mathbf{x}_i = \mathbf{a}_{i-n+m}$ $(i = n-m+1, ..., n)$. The spherical sensor localization problem is to locate the first $n-m$ unknown sensors $\mathbf{x}_i \in \mathbb{S}^2$ $(i = 1, ..., n-m)$ according to anchors' positions $\mathbf{a}_1, ..., \mathbf{a}_m$ and some approximated spherical distances $d_{ij} \approx d_s(\mathbf{x}_i, \mathbf{x}_j), (i,j) \in \mathcal{N}_x$ and $\bar{d}_{ik} = d_s(\mathbf{x}_i, \mathbf{a}_k), (i,k) \in \mathcal{N}_a$ (see, for example, [13, 26]). Here, $d_s(\cdot, \cdot)$ denotes the spherical distance (namely, $d_s(x,y) = \arccos\langle x, y\rangle$ for any $x, y \in \mathbb{S}^2$) and $\mathcal{N}_x, \mathcal{N}_a$ are known, denoting index sets of some sensor-sensor pairs and sensor-anchor pairs, respectively.

Model formulation. To solve the spherical sensor localization problem, we first provide a formulation for it. To this end, let

$$\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_{n-m}]^T, \quad \mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_m]^T, \quad \bar{\mathbf{X}} = \left[\begin{array}{c} \mathbf{X} \\ \mathbf{A} \end{array}\right], \quad \mathbf{Y} = \bar{\mathbf{X}}\bar{\mathbf{X}}^T.$$

It follows from the definition of $\mathbf{Y}$ and the fact that $d_s(x,y) = \arccos\langle x, y\rangle$ for any $x, y \in \mathbb{S}^2$, $d_s(\mathbf{x}_i, \mathbf{x}_j) \approx d_{ij}, (i,j) \in \mathcal{N}_x$ and $d_s(\mathbf{x}_i, \mathbf{a}_k) \approx \bar{d}_{ik}, (i,k) \in \mathcal{N}_a$ that

- $\mathbf{Y}_{ij} = \mathbf{x}_i^T \mathbf{x}_j = \cos(d_s(\mathbf{x}_i, \mathbf{x}_j)) \approx \cos(d_{ij})$ if $(i,j) \in \mathcal{N}_x$;
- $\mathbf{Y}_{ij} = \mathbf{x}_i^T \mathbf{a}_{j-n+m} = \cos(d_s(\mathbf{x}_i, \mathbf{a}_{j-n+m})) \approx \cos(\bar{d}_{i(j-n+m)})$ if $(i, j-n+m) \in \mathcal{N}_a$;
- $\mathbf{Y}_{ij} = \mathbf{a}_{i-n+m}^T \mathbf{x}_j = \cos(d_s(\mathbf{a}_{i-n+m}, \mathbf{x}_j)) \approx \cos(\bar{d}_{(i-n+m)j})$ if $(i-n+m, j) \in \mathcal{N}_a$;
- $\mathbf{Y}_{ij} = \mathbf{a}_{i-n+m}^T \mathbf{a}_{j-n+m}$ if $n-m+1 \le i \ne j \le n$;
- $\mathbf{Y}_{ii} = 1$ if $1 \le i \le n$.

For convenience, we define the matrix $M \in \Re^{n\times n}$ and the index sets Ω_1 and Ω_2 as follows:

$$\begin{array}{rcl} M_{ij} & := & \left\{\begin{array}{ll} \cos(d_{ij}) & \text{if } (i,j) \in \mathcal{N}_x, \\ \cos(\bar{d}_{i(j-n+m)}) & \text{if } (i, j-n+m) \in \mathcal{N}_a, \\ \cos(\bar{d}_{j(i-n+m)}) & \text{if } (j, i-n+m) \in \mathcal{N}_a, \\ \mathbf{a}_{i-n+m}^T \mathbf{a}_{j-n+m} & \text{if } n-m+1 \le i \ne j \le n, \\ 1 & \text{if } 1 \le i = j \le n, \\ 0 & \text{otherwise}, \end{array}\right. \\ \Omega_1 & := & \{(i,j) |\ (i,j) \in \mathcal{N}_x\} \cup \{(i,j) |\ (i, j-n+m) \in \mathcal{N}_a\} \cup \{(i,j) |\ (j, i-n+m) \in \mathcal{N}_a\}, \\ \Omega_2 & := & \{(i,j) |\ n-m+1 \le i \ne j \le n\} \cup \{(i,i) |\ 1 \le i \le n\}. \end{array}$$

From the above deinition one can observe that $\mathbf{Y}_{ij} \approx M_{ij}$ for $(i,j) \in \Omega_1$ and $\mathbf{Y}_{ij} = M_{ij}$ for $(i,j) \in \Omega_2$. It then follows that

$$H_1 \circ (\mathbf{Y} - M) \approx 0, \qquad H_2 \circ (\mathbf{Y} - M) = 0, \tag{6.1}$$

where

$$(H_1)_{ij} = \left\{\begin{array}{ll} 1 & \text{if } (i,j) \in \Omega_1, \\ 0 & \text{otherwise.} \end{array}\right. \quad \text{and} \quad (H_2)_{ij} = \left\{\begin{array}{ll} 1 & \text{if } (i,j) \in \Omega_2, \\ 0 & \text{otherwise.} \end{array}\right.$$

In addition, notice from $\mathbf{x}_i \in \mathbb{S}^2, i = 1, ..., n$, $\bar{\mathbf{X}} = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]^T$ and $\mathbf{Y} = \bar{\mathbf{X}}\bar{\mathbf{X}}^T$ that $\|\mathbf{Y}\|_F = \|\bar{\mathbf{X}}\bar{\mathbf{X}}^T\|_F \le \|\bar{\mathbf{X}}\|_F^2 = n$, which implies that $0 \preceq \mathbf{Y} \preceq nI$ and $\text{rank}(\mathbf{Y}) \le 3$. In view of these and (6.1), one can see that $\mathbf{Y}$ is an approximate solution of the following problem:

$$\begin{array}{rl} \min & \frac{1}{2}\|H_1 \circ (Z - M)\|_F^2 \\ \text{s.t.} & H_2 \circ (Z - M) = 0, \\ & 0 \preceq Z \preceq nI, \ \text{rank}(Z) \le 3. \end{array} \tag{6.2}$$

One approach to locating the spherical sensors $\mathbf{x}_i, i = 1, ..., n-m$ is by first finding an approximate solution Z of (6.2) and then applying a suitable post-processing procedure to obtain an estimation of $\mathbf{x}_i, i = 1, ..., n-m$.

A penalty method. We now propose a penalty method for solving problem (6.2). Upon changing the variable $Y = Z/n$, problem (6.2) is reduced to the problem

$$\begin{array}{rl} \min & \frac{1}{2}\|H_1 \circ (nY - M)\|_F^2 \\ \text{s.t.} & H_2 \circ (nY - M) = 0, \\ & 0 \preceq Y \preceq I, \ \text{rank}(Y) \le 3. \end{array} \tag{6.3}$$

Inspired by Sections 3 and 5, we can solve (6.3) by a penalty scheme that solves a sequence of subproblems in the form of

$$\min_{0\preceq Y\preceq I} \frac{1}{2}\|H_1\circ(nY-M)\|_F^2+\mu_{1,k}\|H_2\circ(nY-M)\|_F^2+\mu_{2,k}\sum_{i=4}^{n}\lambda_i^p(Y) \tag{6.4}$$

for $k=1,2,\cdots$, where $\mu_{1,k},\ \mu_{2,k}>0$ are penalty parameters, and $\|H_2\circ(Y-M/n)\|_F^2$ and $\sum_{i=4}^{n}\lambda_i^p(Y)$ are the penalty functions for the constraints $H_2\circ(Y-M/n)=0$ and $\text{rank}(Y)\leq 3$, respectively.

We apply Algorithm 1 to solve (6.4). For Algorithm 1, we set $L_{\min}=10^{-8}$, $L_{\max}=10^{8}$, $\gamma=2$, $c=10^{-4}$, $N=4$, $p=0.5$, $r=3$, and choose L_k^0 according to (4.3). Let $\{Y^{k,j}\}$ be the sequence generated by Algorithm 1 applied to (6.4). We terminate Algorithm 1 once

$$\frac{\|Y^{k,j}-Y^{k,j-1}\|_F}{\max(\|Y^{k,j}\|_F,1)}\leq\epsilon_k$$

holds for some j and set $Y^k=Y^{k,j}$, where $\{\epsilon_k\}$ is updated as follows:

$$\epsilon_0=10^{-6},\quad \epsilon_k=\max(0.5\epsilon_{k-1},10^{-4})\text{ for }k>0.$$

In addition, the penalty parameters $\mu_{1,k}$ and $\mu_{2,k}$ are updated by setting $\mu_{1,1}=\mu_{2,1}=1$ and for $k\geq 1$,

$$\begin{aligned}\mu_{1,k+1}&=2\mu_{1,k} &&\text{when } \frac{\|H_2\circ(Y^k-M/n)\|_F}{\max(\|Y^k\|_F,1)}>10^{-3},\\ \mu_{2,k+1}&=2\mu_{2,k} &&\text{when } \sum_{i=r+1}^{n}\lambda_i^p(Y^k)>10^{-5}.\end{aligned}$$

We terminate the penalty method once

$$\frac{\|H_2\circ(Y^k-M/n)\|_F}{\max(\|Y^k\|_F,1)}\leq 10^{-3}\text{ and }\sum_{i=r+1}^{n}\lambda_i^p(Y^k)\leq 10^{-5}.$$

Let $Y^*\in\Re^{n\times n}$ be an approximate solution of (6.3) found by the above penalty method. To obtain an approximate location of the sensors $\mathbf{x}_i, i=1,\ldots,n-m$, we adopt the following post-processing strategy, written as pseudo Matlab code, which makes use of the anchors' positions to find an orthogonal matrix (see [27, Appendix C]):

$$\begin{gathered}[U,D]=svd(nY^*);\quad G=U(:,1:3)*sqrt(D(1:3,1:3));\quad G=\mathcal{P}_{\mathbb{S}^2}(G);\\ [\tilde{U},\sim,\tilde{V}]=svd([\mathbf{a}_1,...,\mathbf{a}_m]*G(n-m+1:n,:));\quad X^*=G(1:n-m,:)*\tilde{V}*\tilde{U}'.\end{gathered}$$

Here, $\mathcal{P}_{\mathbb{S}^2}(G)$ denotes the matrix obtained by projecting the row vectors of G onto the sphere $\mathbb{S}^2$.

A SDP relaxation approach. Let

$$W=\begin{bmatrix}I_3\\X\end{bmatrix}\begin{bmatrix}I_3 & X^T\end{bmatrix}=\begin{bmatrix}I_3 & X^T\\X & XX^T\end{bmatrix}.$$

By a similar technique as in [3], one can relax the spherical sensor localization problem into the following optimization problem

$$\begin{array}{ll}\min & \sum_{(i,j)\in\mathcal{N}_x}|W_{i+3,j+3}-\cos(d_{ij})|+\sum_{(i,k)\in\mathcal{N}_a}\left|a_k^TW_{1:3,i+3}-\cos(\bar{d}_{ik})\right|\\ \text{s.t.} & W_{1:3,1:3}=I_3,\\ & W_{\ell,\ell}=1,\quad \ell=4,\ldots,n-m+3,\\ & W\succeq 0.\end{array} \tag{6.5}$$

Notice that this problem can be rewritten as a semidefinite programming problem and solved by SDPT3 [?]. Let W^* be an approximate solution of (6.5) obtained from SDPT3 with the default settings. Finally, we set $\mathcal{P}_{\mathbb{S}^2}(W^*(4:n-m+3,1:3))$ as an approximate location of the sensors.

Performance comparison. In what follows, we compare the performance of the above penalty method (PM) with the above SDP relaxation approach on randomly generated instances. To this end, we choose $n=100, m=4$, the noise factor $\delta=0.001, 0.01, 0.05, 0.1$, and the radio range $R=1.0, 1.1, 1.2, 1.3, 1.4$. For each pair (δ, R), we generate 20 instances in which the sensor points $\mathbf{x}_1, ..., \mathbf{x}_n$ are randomly generated on $\mathbb{S}^2$ with known anchors' positions $\mathbf{x}_i = \mathbf{a}_{i-n+m}, i = n-m+1, ..., n$ and known noisy distances

$$d_{ij} = d_s(\mathbf{x}_i, \mathbf{x}_j)\cdot|1+\delta\cdot\xi_{ij}|,\ \forall(i,j)\in\mathcal{N}_x, \quad \bar{d}_{ik} = d_s(\mathbf{x}_i, \mathbf{a}_k)\cdot|1+\delta\cdot\bar{\xi}_{ik}|,\ \forall(i,k)\in\mathcal{N}_a,$$

where ξ_{ij} and $\bar{\xi}_{ik}$ are randomly generated according to the standard normal distribution $\mathcal{N}(0,1)$ and $\mathcal{N}_x, \mathcal{N}_a$ are defined as

$$\begin{aligned} \mathcal{N}_x &= \{(i,j) : d_s(\mathbf{x}_i, \mathbf{x}_j) \le R, 1\le i,j\le n-m\}, \\ \mathcal{N}_a &= \{(i,k) : d_s(\mathbf{x}_i, \mathbf{a}_k) \le R, 1\le i\le n-m, 1\le k\le m\}. \end{aligned}$$

To evaluate the performance of the above two methods, similar to sensor localization in Euclidean space, we define the root mean square deviation (RMSD) for the spherical localization problem as follows:

$$\text{RMSD} \quad = \quad \sqrt{\frac{1}{n-m}\sum_{i=1}^{n-m} d_s(\mathbf{x}_i^{\text{comp}}, \mathbf{x}_i)^2},$$

where $\mathbf{x}_i^{\text{comp}}$ and $\mathbf{x}_i$ stand for the ith sensor's estimated position and its true position, respectively.

In Table 6.1, we report the averaged RMSD (RMSD) and the averaged CPU time (CPU) over 20 instances for PM and SDP. We also present the averaged number of SVD used ($\sharp_{svd}$) over 20 instances for PM. One can see that SDP is faster than PM when the noise is large, while PM generally outperforms SDP in terms of localization accuracy.

6.2 Nearest low-rank correlation matrix problem

The nearest low-rank correlation problem can be formulated as

$$\begin{array}{rl} \min & \frac{1}{2}\|H\circ(X-C)\|_F^2 \\ \text{s.t.} & \text{diag}(X)=e, \\ & X\succeq 0,\ \text{rank}(X)\le r, \end{array} \tag{6.6}$$

where $H\in\mathcal{S}^n$ is a given weight matrix, $C\in\mathcal{S}^n$ is a given correlation matrix, $r\in[1,n]$ is a given integer, and e is the all-ones vector (see, for example, [5, 12, 21]). Notice that for any $X\in\mathcal{S}^n$ such that $\text{diag}(X)=e$ and $X\succeq 0$, we have $X\preceq nI$. Problem (6.6) is thus equivalent to

$$\begin{array}{rl} \min & \frac{1}{2}\|H\circ(X-C)\|_F^2 \\ \text{s.t.} & \text{diag}(X)=e, \\ & 0\preceq X\preceq nI,\ \text{rank}(X)\le r. \end{array}$$

Upon changing the variable $Y=X/n$, this problem can be reduced to

$$\begin{array}{rl} \min & \frac{1}{2}\|H\circ(Y-C/n)\|_F^2 \\ \text{s.t.} & \text{diag}(Y)=e/n, \\ & 0\preceq Y\preceq I,\ \text{rank}(Y)\le r. \end{array} \tag{6.7}$$

Table 6.1: Numerical results of PM and SDP for $n = 100, m = 4$.

δ	R	RMSD		CPU		$\sharp_{svd}$
		PM	SDP	PM	SDP	
0.001	1.0	1.431e-01	3.470e-03	2.1	1.9	2363
	1.1	5.011e-04	2.850e-03	0.8	2.4	939
	1.2	4.999e-04	2.395e-03	0.7	3.1	760
	1.3	5.034e-04	2.224e-03	0.6	3.9	611
	1.4	4.919e-04	1.521e-03	0.4	4.8	468
0.01	1.0	2.383e-01	3.190e-02	2.3	1.6	2599
	1.1	5.284e-03	3.082e-02	0.9	2.1	1033
	1.2	5.503e-03	2.395e-02	0.8	2.7	822
	1.3	5.088e-03	2.133e-02	0.6	3.4	686
	1.4	5.273e-03	1.634e-02	0.5	4.2	562
0.05	1.0	2.613e-01	2.119e-01	10.5	1.5	11625
	1.1	2.528e-02	1.452e-01	3.5	1.9	3807
	1.2	2.646e-02	1.335e-01	3.5	2.4	3806
	1.3	2.429e-02	9.797e-02	2.9	3.1	3201
	1.4	2.492e-02	8.746e-02	2.6	3.8	2708
0.1	1.0	4.611e-01	4.130e-01	18.0	1.4	19524
	1.1	2.307e-01	3.837e-01	7.9	1.8	8504
	1.2	5.019e-02	2.783e-01	5.5	2.3	5835
	1.3	5.031e-02	2.207e-01	4.6	2.9	4919
	1.4	4.923e-02	1.887e-01	2.8	3.7	2922

A penalty method. In a similar vein as for (6.3), we solve (6.7) by a penalty method that solves a sequence of subproblem in the form of

$$\min_{0 \preceq Y \preceq I} \frac{1}{2}\|H \circ (Y - C/n)\|_F^2 + \mu_{1,k} \|\text{diag}(Y) - e/n\|^2 + \mu_{2,k} \sum_{i=r+1}^{n} \lambda_i^p(Y), \tag{6.8}$$

for $k = 1, 2, ...$, where $\mu_{1,k},\ \mu_{2,k} > 0$ are penalty parameters, and $\|\text{diag}(Y) - e/n\|^2$ and $\sum_{i=r+1}^{n} \lambda_i^p(Y)$ are the penalty functions for the constraints $\text{diag}(Y) = e/n$ and $\text{rank}(Y) \leq r$, respectively.

We apply Algorithm 1 to solve (6.8). The parameters for Algorithm 1 are the same as those used for solving (6.3). Let $\{Y^{k,j}\}$ be the sequence generated by Algorithm 1 applied to (6.8). We terminate Algorithm 1 when

$$\frac{\|Y^{k,j} - Y^{k,j-1}\|_F}{\max(\|Y^{k,j}\|_F, 1)} \leq \epsilon_k$$

holds for some j and set $Y^k = Y^{k,j}$, where $\{\epsilon_k\}$ is updated according to:

$$\epsilon_0 = 10^{-3}, \quad \epsilon_k = \max(0.2\epsilon_{k-1}, 10^{-4}) \text{ for } k > 0.$$

In addition, the penalty parameters $\mu_{1,k}$ and $\mu_{2,k}$ are updated by setting $\mu_{1,1} = \mu_{2,1} = 0.5$ and for $k \geq 1$,

$$\mu_{1,k+1} = 5\mu_{1,k} \qquad \text{when } \frac{\|\text{diag}(Y^k) - e/n\|}{\max(\|Y^k\|_F, 1)} > 10^{-4},$$

$$\mu_{2,k+1} = 5\mu_{2,k} \qquad \text{when } \sum_{i=r+1}^{n} \lambda_i^p(Y^k) > 10^{-4}.$$

We terminate the penalty method once

$$\frac{\|\text{diag}(Y^k) - e/n\|}{\max(\|Y^k\|_F, 1)} \leq 10^{-4} \text{ and } \sum_{i=r+1}^{n} \lambda_i^p(Y^k) \leq 10^{-4}.$$

Let Y^* be an approximation solution of (6.7) obtained by the above penalty method. We use the following post-processing strategy to further obtain an approximation solution X^* of problem (6.6): let $D \in \mathcal{S}^n$ be a diagonal matrix with $D_{ii} = 1/\sqrt{nY^*_{ii}},\ i = 1, ..., n$ and $X^* = n(D * Y^* * D)$. One can observe that the resulting X^* preserves the rank of Y^* while having all ones in its diagonal.

Performance comparison. We now compare the performance of the above penalty method (PM) with a method called PenCorr [11] that is implemented in Matlab with the default parameters. To this aim, we choose $H = E$, $n = 500, 1000, 1500, 2000$, $r = 2, 5, 10, 15, 20$, and C with $C_{ij} = 0.5 + 0.5e^{-0.05|i-j|}$ for $i, j = 1, ..., n$, where E is the all-ones matrix. It shall be mentioned that such a instance with $n = 500$ has been used in [11, Example 5.1]. To evaluate the performance of these two methods, we adopt the same quantity *residue* $= \|H \circ (X^* - C)\|_F$ as in [11], where X^* is an approximation solution of (6.6).

In Table 6.2, we report CPU time and *residue* for our method and PenCorr. In particular, the penalty method with $p = 0.5$ and $p = 1$ are named as $\text{PM}_{0.5}$ and PM_1, respectively. One can see that $\text{PM}_{0.5}$ outperforms PenCorr in terms of CPU time, while it returns similar *residue* as PenCorr. Besides, the performance of PM_1 is comparable to $\text{PM}_{0.5}$ except it sometimes obtains much larger *residue*.[1]

Table 6.2: Nearest low-rank correlation matrix.

n	rank	$\text{PM}_{0.5}$		PenCorr		PM_1	
		CPU	*residue*	CPU	*residue*	CPU	*residue*
500	2	1.6	156.4053	6.3	156.4172	2.4	234.9469
	5	1.1	78.8307	1.9	78.8342	1.1	78.8307
	10	1.1	38.6845	1.2	38.6852	1.1	38.6845
	15	0.8	23.2497	1.0	23.2463	0.8	23.2497
	20	0.7	15.7106	1.2	15.7080	0.9	15.7106
1000	2	7.2	332.7649	30.4	332.8054	10.8	332.7803
	5	5.3	189.3868	9.8	189.3978	5.4	189.3868
	10	4.2	110.7867	8.7	110.7868	4.2	110.7867
	15	5.1	74.7463	7.2	74.7494	5.0	74.7463
	20	4.8	54.1675	5.5	54.1680	4.8	54.1675
1500	2	25.6	509.4009	84.6	509.4665	40.9	617.2919
	5	17.8	301.1784	34.8	301.1892	18.1	301.1784
	10	12.9	188.5594	34.1	188.5554	12.9	188.5594
	15	11.9	135.3811	26.2	135.3820	11.9	135.3811
	20	12.8	103.1023	19.9	103.1043	12.8	103.1023
2000	2	56.1	686.1070	196.9	686.1815	82.9	686.1731
	5	43.5	413.0689	74.6	413.0763	43.9	413.0689
	10	39.2	267.3751	96.7	267.3920	39.0	267.3751
	15	32.8	198.6823	73.5	198.6795	32.9	198.6823
	20	30.4	156.1624	46.5	156.1522	30.0	156.1624

[1] All solutions obtained from the three methods $\text{PM}_{0.5}$, PenCorr and PM_1 satisfy the constraints in (6.6). Thus, we only need to compare the *residue*, or equivalently, the objective function value.

Note that replacing $\sum_{i=r+1}^{n} \lambda_i^p(Y)$ in (6.8) by the convex nuclear norm regularization $\|Y\|_*$ gives a convex problem, but it is not a penalty term for $\text{rank}(Y) \leq r$. We could not find a low-rank solution by using nuclear norm regularization.

A Appendix

Proof of Lemma 4.1.

Proof. We know from the assumption that $p \in (0,1)$.

(i) One can observe from (4.6) that

$$\begin{aligned} 0 \in \mathcal{Z}^*(t) \quad &\Leftrightarrow \quad \Phi(z,t) \geq \Phi(0,t),\ \forall z \in [0,1] \Leftrightarrow \frac{1}{2}(z-t)^2 + \nu z^p \geq \frac{1}{2}t^2,\ \forall z \in [0,1] \\ &\Leftrightarrow \quad z^2 - 2tz + 2\nu z^p \geq 0,\ \forall z \in [0,1] \Leftrightarrow t \leq \inf_{z\in(0,1]} \underbrace{\frac{z}{2} + \nu z^{p-1}}_{u(z)}. \end{aligned} \tag{A.1}$$

It is not hard to observe that $u(\cdot)$ is convex in $(0,\infty)$ and moreover

$$\lim_{z\to 0^+} u(z) = \infty, \qquad u'\left([2(1-p)\nu]^{\frac{1}{2-p}}\right) = 0.$$

Using these facts, (4.8) and (4.7), one can easily see that

$$t_1 = u(\alpha) = \min_{z\in(0,1]} u(z). \tag{A.2}$$

This together with (A.1) implies that statement (i) holds.

(ii) In view of (4.6), one can observe that for arbitrary fixed $t \in [0,1]$,

$$\begin{aligned} 1 \in \mathcal{Z}^*(t) \quad &\Leftrightarrow \quad \Phi(z,t) \geq \Phi(1,t),\ \forall z \in [0,1] \Leftrightarrow \frac{1}{2}(z-t)^2 + \nu z^p \geq \frac{1}{2}(1-t)^2 + \nu,\ \forall z \in [0,1] \\ &\Leftrightarrow \quad z^2 + 2\nu z^p - 1 - 2\nu \geq 2t(z-1),\ \forall z \in [0,1] \\ &\Leftrightarrow \quad t \geq \sup_{z\in[0,1)} \underbrace{\frac{z^2 + 2\nu z^p - 1 - 2\nu}{2(z-1)}}_{w(z)}. \end{aligned} \tag{A.3}$$

By the expression of $w(\cdot)$, one can define

$$w(1) := \lim_{z\to 1^-} w(z) = 1 + \nu p. \tag{A.4}$$

We then observe that w is continuous in $[0,1]$ and moreover it is differentiable in $(0,1)$. Next we show that

$$\sup_{z\in[0,1)} w(z) = \max_{z\in[0,1]} w(z) = \max\{w(0), w(1)\} = t_2, \tag{A.5}$$

where t_2 is defined in (4.8). Indeed, the first equality of (A.5) holds due to the continuity of w in $[0,1]$ while the last equality follows from (4.8), (A.4) and $w(0) = 1/2 + \nu$. It remains to show that the second equality of (A.5) holds, that is, the maximum value of w over $[0,1]$ attains at $z = 0$ or 1. To this end, let

$$h(z) = z^2 - 2z + 2\nu(p-1)z^p - 2\nu p z^{p-1} + 1 + 2\nu.$$

It is easy to verify that

$$\lim_{z\to 0^+} h(z) = -\infty, \quad h(1) = 0, \tag{A.6}$$

$$h'(z) = 2\left[1+\nu p(p-1)z^{p-2}\right](z-1), \tag{A.7}$$

$$w'(z) = \frac{h(z)}{2(z-1)^2}. \tag{A.8}$$

We divide the rest of the proof into two separate cases as follows.

Case 1): $1+\nu p(p-1)\le 0$. It follows from (A.7) that $h'(z)\ge 0$ for every $z\in(0,1]$, which together with (A.6) implies $h(z)\le 0$ for all $z\in(0,1]$. In view of this and (A.8), one can see that $w'(z)\le 0$ for every $z\in(0,1)$. Hence, w is decreasing in $[0,1]$ and the maximum value of w over $[0,1]$ attains at $z=0$.

Case 2): $1+\nu p(p-1)>0$. This together with (4.7) implies $\beta\in(0,1)$. Let

$$\tilde{h}(z) := 1+\nu p(p-1)z^{p-2}.$$

One can observe that $\tilde{h}(\beta)=0$ and moreover $\tilde{h}$ is strictly increasing in $(0,\infty)$. Hence, $\tilde{h}(z)<0$ for $z\in(0,\beta)$ and $\tilde{h}(z)>0$ for $z\in(\beta,\infty)$. This together with (A.7) implies that $h'(z)>0$ for $z\in(0,\beta)$ and $h'(z)<0$ for $z\in(\beta,1]$. Using this and continuity of h in $(0,1]$, one can see that h is strictly increasing in $(0,\beta]$ and strictly decreasing in $(\beta,1]$. It follows from this fact and (A.6) that there exists some $\gamma\in(0,\beta)$ such that $h(z)\le 0$ for $z\in(0,\gamma]$ and $h(z)>0$ for $z\in(\gamma,1]$. This together with (A.8) implies that $w'(z)\le 0$ for $z\in(0,\gamma]$ and $w'(z)>0$ for $z\in(\gamma,1]$. Hence, w is decreasing in $(0,\gamma]$ and increasing in $(\gamma,1]$. Clearly, the maximum value of w over $[0,1]$ attains at $z=0$ or 1.

Combining the above two cases, we see that (A.5) holds. The conclusion of statement (ii) immediately follows from (A.3) and (A.5).

(iii) One can see from (A.2) that $t_1\le u(1)=1/2+\nu$, which together with (4.8) implies that $t_1\le t_2$. Notice from (4.7) that $\beta>0$. Suppose that $\mathcal{Z}^*(t)=\{z^*\}\subseteq[\beta,\min\{t,1\})$ for some z^*. It then follows that $0\notin\mathcal{Z}^*(t)$ and $1\notin\mathcal{Z}^*(t)$, which together with statements (i) and (ii) implies $t\in(t_1,t_2)$. Hence, the "only if" part of this statement holds. We next show that the "if" part also holds. Let $t\in(t_1,t_2)$ be arbitrarily chosen. Using this, $\mathcal{Z}^*(t)\in[0,1]$ and statements (i) and (ii), we know that $\mathcal{Z}^*(t)\in(0,1)$. Let $z^*\in\mathcal{Z}^*(t)\subset(0,1)$ be arbitrarily chosen. By the optimality conditions of (4.6), we have $\nabla_z\Phi(z^*,t)=0$ and $\nabla^2_{zz}\Phi(z^*,t)\ge 0$, that is,

$$z^*-t+\nu p(z^*)^{p-1}=0, \qquad 1+\nu p(p-1)(z^*)^{p-2}\ge 0. \tag{A.9}$$

The first relation of (A.9) and $z^*\in(0,1)$ yields $z^*<\min\{t,1\}$. The second relation of (A.9), $p\in(0,1)$ and the definition of β implies $z^*\ge\beta$. Hence, $z^*\in[\beta,\min\{t,1\})$. This together with (A.9) implies that z^* is a root of equation (4.9) in $[\beta,\infty)$. It remains to show that z^* is the unique root of (4.9) in $[\beta,\infty)$. Let g be defined in (4.9). Notice from (4.7) and (4.9) that $g'(\beta)=0$ and g' is strictly increasing in $(0,\infty)$. Hence, $g'(z)>0$ for every $z\in(\beta,\infty)$. It follows that g is strictly increasing in $[\beta,\infty)$, which implies that z^* is the unique root of (4.9) in $[\beta,\infty)$. This completes the proof. □

Lemma A.1. *Consider a univariate equation $q(z)=0$. Assume that (a) q has a unique root $z_*\in(a,\infty)$; (b) q' and q'' are positive and continuous in (a,∞). Let $\{z_k\}$ be a sequence generated by Newton's iteration $z_{k+1}=z_k-q(z_k)/q'(z_k)$, $\forall k\ge 0$ with a starting point $z_0\in(a,\infty)$. Then $\{z_k\}$ quadratically and globally converges to z_*, and*

$$0\le z_{k+1}-z_*\le\left\{\frac{1}{q'(z_*)}\max_{z\in[z_*,z_1]}q''(z)\right\}(z_k-z_*)^2, \qquad \forall k\ge 1.$$

Proof. From assumption (b), for any $z \in (a, \infty)$, we have

$$0 = q(z_*) = q(z) + q'(z)(z_* - z) + q''(\xi_z)(z_* - z)^2/2 \geq q(z) + q'(z)(z_* - z),$$

where ξ_z is between z and z_*. Hence, $z - q(z)/q'(z) \geq z_*$, $\forall z \in (a, \infty)$. This, together with $z_0 \in (a, \infty)$ and $z_{k+1} = z_k - q(z_k)/q'(z_k)$, $\forall k \geq 0$, implies that $z_k \geq z_*$ holds for all $k \geq 1$. Moreover, from $q' > 0$ in (a, ∞), we have $q(z_k) = q(z_*) + q'(\eta_k)(z_k - z_*) \geq 0$ for $k \geq 1$, where $\eta_k \in (z_*, z_k)$, which implies that $\{z_k\}_{k \geq 1}$ is nonincreasing. Therefore, the sequence $\{z_k\}$ converges. Taking limits on both sides of Newton's iteration as $k \to \infty$, we have that $\{z_k\}$ converges to z_*. Finally, using the mean value theorem, we have for all $k \geq 1$,

$$\begin{aligned} z_{k+1} - z_* &= z_k - z_* - \frac{q(z_k) - q(z_*)}{q'(z_k)} = z_k - z_* - \frac{q'(\xi_k)}{q'(z_k)}(z_k - z_*) = \frac{q'(z_k) - q'(\xi_k)}{q'(z_k)}(z_k - z_*) \\ &= \frac{q''(\eta_k)}{q'(z_k)}(z_k - \xi_k)(z_k - z_*) \leq \left\{ \frac{1}{q'(z_*)} \max_{z \in [z_*, z_1]} q''(z) \right\} (z_k - z_*)^2 \end{aligned}$$

for some $\xi_k \in (z_*, z_k)$ and $\eta_k \in (\xi_k, z_k)$. This completes the proof. □

Acknowledgments

We are grateful to the Associate Editor and two Referees for their helpful comments and suggestions. We would like to thank Defeng Sun, Ting Kei Pong, Ya-Feng Liu and Zaikun Zhang for their helpful comments and Xudong Li for his help in Matlab code.

References

[1] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[2] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10:1196–1211, 2000.

[3] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye. Semidefinite programming based algorithms for sensor network localization. *ACM TOSN*, 2:188–220, 2006.

[4] P. Biswas and Y. Ye. Semidefinite programming for ad hoc wireless sensor network localization. In *Proceedings of IPSN*, pages 46–54. ACM, 2004.

[5] R. Borsdorf, N. J. Higham, and M. Raydan. Computing a nearest correlation matrix with factor structure. *SIAM J. Matrix Anal. Appl.*, 31:2603–2622, 2010.

[6] A.-M. Bruckner. *Differentiation of real functions.* Springer, Collier-Macmillan Publishers, 1978.

[7] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9:717–772, 2009.

[8] X. Chen, Z. Lu, and T. K. Pong. Penalty methods for a class of non-lipschitz optimization problems. *SIAM J. Optim.*, 26:1465–1492, 2016.

[9] F. H. Clarke. Generalized gradients and applications. *Trans. Amer. Math. Soc.*, 205:247–262, 1975.

[10] M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of ACC*, pages 4734–4739. IEEE, 2001.

[11] Y. Gao and D. Sun. A majorized penalty approach for calibrating rank constrained correlation matrix problems. *Preprint available at http://www. math. nus. edu. sg/˜ matsundf/MajorPen. pdf*, 2010.

[12] N. J. Higham. Computing the nearest correlation matrix: a problem from finance. *IMA J. Numer. Anal.*, 22:329–343, 2002.

[13] B. Huang, C. Yu, and B. D. Anderson. Noisy localization on the sphere: Planar approximation. In *ISSNIP*, pages 49–54. IEEE, 2008.

[14] S. Ji, K.-F. Sze, Z. Zhou, A. M.-C. So, and Y. Ye. Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In *INFOCOM*, pages 2499–2507. IEEE, 2013.

[15] Q. Li, D. Li, and H.-D. Qi. Newtons method for computing the nearest correlation matrix with a simple upper bound. *J. Optim. Theory Appl.*, 147:546–568, 2010.

[16] Q. Li and H.-D. Qi. A sequential semismooth newton method for the nearest low-rank correlation matrix problem. *SIAM J. Optim.*, 21:1641–1666, 2011.

[17] Y. Lu, L. Zhang, and J. Wu. A smoothing majorization method for matrix minimization. *Optim. Methods Softw.*, 30:682–705, 2015.

[18] Z. Lu and X. Li. Sparse recovery via partial regularization: Models, theory and algorithms. *Math. Oper. Res.*, 2015. To appear.

[19] Z. Lu, Y. Zhang, and X. Li. Penalty decomposition methods for rank minimization. *Optim. Methods Softw.*, 30:531–558, 2015.

[20] Z. Lu, Y. Zhang, and J. Lu. ℓ_p regularized low-rank approximation via iterative reweighted singular value minimization. *Comput. Optim. Appl.*, 68:619–642, 2017.

[21] H. Qi and D. Sun. A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM J. Matrix Anal. Appl.*, 28:360–385, 2006.

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

[23] B. Recht, W. Xu, and B. Hassibi. Null space conditions and thresholds for rank minimization. *Math. Program.*, 127:175–202, 2011.

[24] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[25] S. J. Wright, R. D. Nowak, and M. A. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.*, 57:2479–2493, 2009.

[26] C. Yu, H. Chee, and B. D. Anderson. Noisy localization on the sphere: A preliminary study. In *ISSNIP*, pages 43–48. IEEE, 2008.

[27] Z. Zhang. A flexible new technique for camera calibration. *IEEE, TPAMI*, 22:1330–1334, 2000.