STRUCTURED LOW RANK MATRIX OPTIMIZATION PROBLEMS: A PENALTY APPROACH

GAO YAN

(B.Sc., ECNU)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

AUGUST 2010

This thesis is dedicated to

my parents

Acknowledgements

With great pleasure, I would like to express my sincere gratitude to Professor Sun Defeng, my dissertation supervisor, for his guidance and support during my graduate study at National University of Singapore. As his student, I have been privileged to see and learn many wonderful mathematical insights from him. I am thankful for the time and effort he extended to further my education and research ability and the opportunity to perform the research constituting this dissertation in a most independent fashion. He is such a nice mentor, besides being a well-known, energetic and insightful research. Conversations with him were always enjoyable and enlightening.

I would also like to express my sincere respect to Professor Toh Kim Chuan and Professor Zhao Gongyun for offering optimization courses, which I found particularly valuable in my research.

I am greatly indebted to National University of Singapore for providing me a full scholarship and a pleasant study environment. Also, I would like to thank the

Faculty of Science for awarding me Best Graduate Researcher Award and Department of Mathematics for providing partial financial support for my attendance of ISMP conference at Chicago and providing such a nice academic environment.

An important note of thanks goes to my family for their endless love, constant encouragement and unconditional support throughout the duration of my PhD, without which I would have been nowhere. They are the sources of my courage.

Finally, I am deeply thankful to all the friends I met in Singapore. It is their companion and emotional support that made my life at NUS much easier and this place such a fun. They are too many to name and too good to be forgotten. I also want to say thank you to all team members in our optimization group and I have benefited a lot from them.

Gao Yan

August 2010

Contents

Summary

This thesis focuses on a class of structured low rank matrix optimization problems (SLR-MOPs) which aim at finding an approximate matrix of certain specific structures and whose rank is no more than a prescribed number. This kind of approximation is needed in many important applications arising from a wide range of fields, such as finance/risk management, images compression, noise reduction, and so on. The SLR-MOPs are in general non-convex and thus difficult to solve due to the presence of the rank constraint.

In this thesis, we propose a penalty approach to deal with this difficulty. Some rationale to motivate this penalty technique is also addressed. For example, one interesting result says that an ε-optimal solution to the original SLR-MOP is guaranteed by solving the penalized problem as long as the penalty parameter c is above some ε-dependent number. We further present a general proximal subgradient method for the purpose of solving the penalized problem which is still non-convex. When using the proposed proximal subgradient method, one eventually

needs to solve a sequence of least squares nuclear norm problems. For this purpose, we design a quadratically convergent smoothing Newton-BiCGStab method to solve these least squares subproblems. Essentially, our approach transforms the structured low rank matrix problem into a sequence of least squares nuclear norm problems. One remarkable feature of our method is that it can continue to search for a better low rank solution by iteratively solving a new least squares nuclear norm problem when the initial nuclear norm convex relaxation fails to produce a satisfactory solution.

Furthermore, we also investigate the Lagrangian dual of the structured low rank matrix optimization problem and show some globalization checking results which are seldom available for the non-convex optimization problems. As a byproduct, we fully characterize the metric projection over three non-convex rank constrained sets, respectively.

Numerical results on a variety of low rank matrix problems indicate that our proposed method is able to handle both the rank and the linear constraints effectively, in particular in the situations when the rank is not very small. The numerical results also imply the efficiency and robustness of the smoothing Newton-BiCGStab method which is applied to solve the subproblems.

Chapter 1

Introduction

To approximate a given matrix by a low rank matrix has a long history in mathematics. For example, Schmidt [108], Eckart and Young [31] considered the following low rank approximation problem

$$\begin{aligned} \min \quad & \|X - C\| \\ \text{s.t.} \quad & \operatorname{rank}(X) \leq r\,. \end{aligned} \tag{1.1}$$

Here we use $\|\cdot\|$ to denote the Frobenius norm in $\Re^{n_1 \times n_2}$ (assuming $n_1 \leq n_2$ without loss of generality). Let the given matrix $C \in \Re^{n_1 \times n_2}$ have the following singular value decomposition (SVD):

$$C = U[\Sigma(C)\ \mathbf{0}]V^T\,,$$

where $U \in \mathcal{O}^{n_1}$ and $V \in \mathcal{O}^{n_2}$ are orthogonal matrices, $\sigma_1(C) \geq \cdots \geq \sigma_{n_1}(C) \geq 0$ are the singular values of C being arranged in the non-increasing order and $\Sigma(C) := \operatorname{diag}(\sigma(C))$ is the n_1 by n_1 diagonal matrix whose i-th diagonal entry is $\sigma_i(C)$, $i = 1, \ldots, n_1$. Independently, Schmidt [108] and Eckart and Young [31] proved that

$$X^* = \sum_{i=1}^{r} \sigma_i(C) U_i V_i^T$$

is an optimal solution to (1.1). A more general problem than (1.1) is the so-called weighted low rank approximation problem:

$$\begin{array}{ll} \min & \|H \circ (X - C)\| \\ \text{s.t.} & \text{rank}(X) \leq r\,, \end{array} \tag{1.2}$$

where $H \in \Re^{n_1 \times n_2}$ a given weight matrix whose entries are nonnegative and "$\circ$" denotes the Hadamard product. Unlike the case for problem (1.1), the weighted problem (1.2) no longer admits an analytic solution. Actually, problem (1.2) is known to be NP-hard in general [43]. Of course, one may use other functions to measure the distance between X and C. Moreover, in practice we not only seek a low rank matrix X, but also we want X to have certain desirable properties such as:

- X is symmetric and positive semidefinite;
- Some components of X are required to satisfy some equalities and inequalities;
- X is in a special class of matrices, e.g., correlation, Hankel, Toeplitz, tridiagonal matrices, and so on.

Such problems in the literature are called structured low rank matrix approximation problems [17].

In her PhD thesis, Fazel [33] considered the following matrix rank minimization problem (RMP)

$$\begin{array}{ll} \min & \text{rank}(X) \\ \text{s.t.} & X \in \mathcal{C}\,, \end{array} \tag{1.3}$$

where $\mathcal{C}$ is a closed convex set in $\Re^{n_1 \times n_2}$. Since the RMP is difficult to solve, Fazel suggested to use $\|X\|_*$, the sum of all the singular values of X, to replace $\text{rank}(X)$

in the objective function in (1.3). That is, she proposed to solve the following convex optimization problem

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & X \in \mathcal{C} \end{array} \tag{1.4}$$

to get an approximate solution to the RMP problem (1.3). See also [34]. Though simple, this strategy works very well in many occasions. One particular example is the so-called matrix completion problem. Given a matrix $M \in \Re^{n_1 \times n_2}$ with entries in the index set Ω given, the matrix completion problem seeks to find a low rank matrix X such that $X_{ij} \approx M_{ij}$ for all $(i,j) \in \Omega$. In [13], [14], [45], [56], [102], [103], etc., the authors made some landmark achievements: for certain stochastic models, an $n_1 \times n_2$ matrix of rank r can be recovered with high probability from a random uniform sample of size slightly larger than $O((n_1+n_2)r)$ via solving the following nuclear norm minimization problem:

$$\begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & X_{ij} \approx M_{ij} \quad \forall\ (i,j) \in \Omega\,. \end{array}$$

The breakthrough achieved in the above mentioned papers and others has not only given a theoretical justification of relaxing the nonconvex RMP problem (1.3) to its convex counterpart (1.4), but also has accelerated the development on adopting the nuclear norm minimization approach to model many more application problems that go beyond the matrix completion problem. However, since the nuclear norm convex relaxation does not take the prescribed number r into consideration, the solution obtained by solving the relaxed problem may not satisfy the required rank constraint. Moreover, the nuclear norm convex relaxation may not work at all if the low rank matrix has to possess certain structures. Next, we shall take a recently intensively studied financial problem to illustrate this situation.

Let $\mathcal{S}^n$ and $\mathcal{S}^n_+$ denote, respectively, the space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$. Denote the Frobenius norm induced

by the standard trace inner product $\langle \cdot, \cdot \rangle$ in $\mathcal{S}^n$ by $\| \cdot \|$. Let C be a given matrix in $\mathcal{S}^n$ and $H \in \mathcal{S}^n$ a given weight matrix whose entries are nonnegative. Now we consider the following rank constrained nearest correlation matrix[1] problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} \quad & X_{ii} = 1, \ i = 1, \ldots, n\,, \\ & X_{ij} = e_{ij}, \ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \geq l_{ij}, \ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \leq u_{ij}, \ (i,j) \in \mathcal{B}_u\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \mathrm{rank}(X) \leq r\,, \end{aligned} \tag{1.5}$$

where $1 \leq r \leq n$ is a given integer, $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are three index subsets of $\{(i,j) \,|\, 1 \leq i < j \leq n\}$ satisfying $\mathcal{B}_e \cap \mathcal{B}_l = \emptyset$, $\mathcal{B}_e \cap \mathcal{B}_u = \emptyset$, $-1 \leq e_{ij}, l_{ij}, u_{ij} \leq 1$ for any $(i,j) \in \mathcal{B}_e \cup \mathcal{B}_l \cup \mathcal{B}_u$, and $-1 \leq l_{ij} < u_{ij} \leq 1$ for any $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$. Denote the cardinalities of $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ by q_e, q_l, and q_u, respectively. Let $p := n + q_e$ and $m := p + q_l + q_u$. The weight matrix H is introduced by adding larger weights to correlations that are better estimated or are of higher confidence in their correctness. Zero weights are usually assigned to those correlations that are missing or not estimated. See [87] for more discussions.

This kind of problems has many applications among a variety of fields, in particular, in the quantitative finance field. Wu [121], Zhang and Wu [123], and Brigo and Mercurio [9] considered such a problem for pricing interest rate derivatives under the LIBOR and swap market models. The factor models of basket options, collateralized debt obligations (CDOs), portfolio risk models (VaR), and multivariate time series discussed by Lillo and Mantegna [69] rely on low rank nearest correlation matrices. A correlation matrix of low rank is particularly useful in the

[1] A correlation matrix, a commonly used concept in statistics, is a real symmetric and positive semidefinite matrix whose diagonal entries are all ones.

Monte Carlo simulation for solving derivatives pricing problems as a model with low factors can significantly reduce the cost of drawing random numbers. Beyond quantitative finance, the rank constrained nearest correlation matrix problems also occur in many engineering fields, see for examples, [11, 20, 53, 110].

Notice that for a correlation matrix $X \in \mathcal{S}^n_+$,

$$||X||_* = \text{trace}(X) = n\,.$$

This implies that any convex relaxation of using the nuclear norm directly is doomed as one will simply add a constant term if one does so.

In this thesis, we shall propose a penalty method to solve problem (1.5) and its more general form

$$\begin{array}{ll} \min & \dfrac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \text{rank}(X) \le r\,, \end{array} \tag{1.6}$$

where $\mathcal{A} : \mathcal{S}^n \to \Re^m$ is a linear operator, $\mathcal{Q} := \{0\}^p \times \Re^q_+$ and $m := p+q$. Moreover, since in many situations the matrix X is not necessarily required to be symmetric, we also consider the nonsymmetric counterpart of problem (1.6).

Let $\rho \ge 0$ be a given parameter. The penalty method proposed in this thesis is also strongly motivated to solve the following structured low rank matrix, not necessarily symmetric, approximation problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\|H \circ (X - C)\|^2 + \rho\|X\|_* \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & \text{rank}(X) \le r\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{1.7}$$

where $m = p + q$, $\mathcal{A} : \Re^{n_1 \times n_2} \to \Re^m$ is a linear operator, $\widehat{\mathcal{Q}} \in \Re^q$ is a closed convex cone with nonempty interior and $\mathcal{Q} := \{0\}^p \times \widehat{\mathcal{Q}}$. In this thesis, we shall address

some theoretical and numerical issues involved in problems (1.7) and (1.6).

Our main idea is to deal with the non-convex rank constraint via a penalty technique. The rationale for using the penalty approach is explained in later chapters. It is worth noting that an ε-optimal solution to the original problem is always guaranteed by solving the penalized problem as long as the penalty parameter is larger than some ε-dependent number. The penalized problem, however, is still not convex and no existing methods can be directly applied to solve it. Thus, we further propose a proximal subgradient method to solve the penalized problem. When using the proposed proximal subgradient method, one eventually needs to solve a sequence of least squares nuclear norm problems. Notice that the efficiency of the whole approach heavily relies on the method used for solving the subproblems. For this purpose, we design a smoothing Newton-BiCGStab method to solve these least squares subproblems.

Essentially, our approach transforms the structured low rank matrix problem into a sequence of least squares nuclear norm problems. In this sense, the popular nuclear norm relaxation may be regarded as the first step of our approach if we choose the starting point properly. Different from the nuclear norm relaxation approach, our method can continue to search for a better low rank solution by iteratively solving a new least squares nuclear norm problem when the former fails to generate a satisfactory solution.

Finally, it should be emphasized that our proposed approach here is quite flexible. It can be used to solve problems beyond the ones described in (1.6) and (1.7). For examples, we can easily adopt our approach to solve the portfolio selection problem with the cardinality constraint [67, 70] and the weighted version of the $(P^1_{F,K})$ problem introduced by Werner and Schöttle in [120].

1.1 Outline of the thesis

The thesis is organized as follows: in Chapter 2, we give some preliminaries to facilitate later discussions. Chapter 3 presents a general framework of the proximal subgradient approach. In Chapter 4, we introduce a penalty approach to tackle the rank constraint and then apply the general proximal subgradient method to the penalized problem. We also offer some theoretical justifications for using this penalty technique. The Lagrangian dual reformulation and the global optimality checking results are also presented in this chapter. In Chapter 5, we design a quadratically convergent inexact smoothing Newton-BiCGStab method and then apply it to solve the subproblems resulted in Chapter 4. We demonstrate the effectiveness of our method by conducting some numerical experiments on both nonsymmetric and symmetric cases on a variety of problems in Chapter 6. Finally, we conclude the thesis and point out some future research directions in Chapter 7.

Chapter 2

Preliminaries

2.1 Notations

Let m and n be positive integers. We use $\mathcal{O}^m$ to denote the set of all orthogonal matrices in $\Re^{m\times m}$, i.e.,

$$\mathcal{O}^m = \{Q \in \Re^{m\times m} \mid Q^T Q = I\},$$

where I is the identity matrix with appropriate dimension. For any symmetric matrix X, Y and Z in $\mathcal{S}^n$, we write $X \succeq 0$ ($\succ 0$) to represent that X is positive semidefinite (positive definite) and $Z \succeq X \succeq Y$ to represent that $X - Y \succeq 0$ and $Z - X \succeq 0$. Let $\alpha \subseteq \{1, \ldots, m\}$ and $\beta \subseteq \{1, \ldots, n\}$ be index sets, and M be an $m \times n$ matrix. The cardinality of α is denoted by $|\alpha|$. We write M_α for the matrix containing the columns of M indexed by α and $M_{\alpha\beta}$ for the $|\alpha| \times |\beta|$ submatrix of M formed by selecting the rows and columns from M indexed by α and β, respectively. The Frobenius norm in $\Re^{m\times n}$ is denoted by $\|\cdot\|$. For any $v \in \Re^m$, we use $\mathrm{diag}(v)$ to denote the $m \times m$ diagonal matrix whose ith diagonal entry is v_i, $i = 1, \ldots, m$, $\|v\|$ to denote the 2-norm of v, and $\|v\|_0$ to denote the cardinality of the set $\{i \,|\, v_i \neq 0,\ i = 1, \ldots, m\}$. We also use $|v|$ to denote the column vector

in $\Re^m$ such that its ith component is defined by $|v|_i = |v_i|$, $i = 1, \ldots, m$ and $v_+ := \max(0, v)$. For any set $\mathcal{W}$, the convex hull of $\mathcal{W}$ is denoted by $\operatorname{conv} \mathcal{W}$.

2.2 Matrix valued function and Löwner's operator

Let $X \in \mathcal{S}^n$ admit the following spectral decomposition

$$X = P\Lambda(X)P^T, \tag{2.1}$$

where $\Lambda(X) := \operatorname{diag}(\lambda(X))$, $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$ are the eigenvalues of X being arranged in the non-increasing order and $P \in \mathcal{O}^n$ is a corresponding orthogonal matrix of orthonormal eigenvectors of X. Let $f : \Re \to \Re$ be a scalar function. The corresponding Löwner's symmetric matrix function at X is defined by [71]

$$F(X) := P\operatorname{diag}\big(f(\lambda_1(X)), f(\lambda_2(X)), \ldots, f(\lambda_n(X))\big)P^T = \sum_{i=1}^{n} f(\lambda_i(X))P_iP_i^T\,. \tag{2.2}$$

Let $\mu \in \Re^n$ is a given vector. Assume that the scalar function $f(\cdot)$ is differentiable at each μ_i with the derivatives $f'(\mu_i)$, $i = 1, \ldots, n$. Let $f^{[1]}(\mu) \in \mathcal{S}^n$ be the first divided difference matrix whose (i, j)-th entry is given by

$$[f^{[1]}(\mu)]_{ij} = \begin{cases} \dfrac{f(\mu_i) - f(\mu_j)}{\mu_i - \mu_j}, & \text{if } \mu_i \neq \mu_j, \\ f'(\mu_i), & \text{if } \mu_i = \mu_j, \end{cases} \qquad i, j = 1, \ldots, n. \tag{2.3}$$

The following proposition concerning the differentiability of the symmetric matrix function F defined in (2.2) can be largely derived from [60].

Proposition 2.1. *Let $X \in \mathcal{S}^n$ have the spectral decomposition as in (2.1). Then, the symmetric matrix function $F(\cdot)$ is (continuously) differentiable at X if and*

only for each $i \in \{1, \ldots, n\}$, $f(\cdot)$ *is (continuously) differentiable at* $\lambda_i(X)$. *In this case, the Fréchet derivative of* $F(\cdot)$ *at* X *is given by*

$$F'(X)H = P\left[f^{[1]}(\lambda(X)) \circ (P^T H P)\right] P^T \quad \forall\ H \in \mathcal{S}^n\,. \tag{2.4}$$

2.3 Semismoothness and the generalized Jacobian

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite-dimensional real Hilbert spaces equipped with an inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$, respectively. Let $\mathcal{O}$ be an open set in $\mathcal{X}$ and $\Xi : \mathcal{O} \subseteq \mathcal{X} \to \mathcal{Y}$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. The well known Rademacher's theorem [107, Section 9.J] says that Ξ is almost everywhere F(réchet)-differentiable in $\mathcal{O}$. Let $\mathcal{D}_\Xi$ denote the set of F(réchet)-differentiable points of Ξ in $\mathcal{O}$. Then, the Bouligand subdifferential of Ξ at x, denoted by $\partial_B\Xi(x)$, is

$$\partial_B\Xi(x) := \big\{ \lim_{k\to\infty} \Xi'(x^k) \mid x^k \to x\,,\ x^k \in \mathcal{D}_\Xi \big\},$$

where $\Xi'(x)$ denotes the F-derivative of Ξ at x. Then Clarke's generalized Jacobian of Ξ at x [18] is the convex hull of $\partial_B\Xi(x)$, i.e.,

$$\partial\Xi(x) := \operatorname{conv} \partial_B\Xi(x)\,.$$

The concept of semismoothness plays an important role in convergence analysis of generalized Newton methods for nonsmooth equations. It was first introduced by Mifflin [76] for functionals, and was extended by Qi and Sun [93], for cases when a vector-valued function is not differentiable, but locally Lipschitz continuous.

Definition 1. $\Xi : \mathcal{O} \subseteq \mathcal{X} \to \mathcal{Y}$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$. We say that Ξ is semismooth at a point $x \in \mathcal{O}$ if

(i) Ξ is directionally differentiable at x; and

(ii) for any $\Delta x \in \mathcal{X}$ and $V \in \partial\Xi(x + \Delta x)$ with $\Delta x \to 0$,

$$\Xi(x + \Delta x) - \Xi(x) - V\Delta x = o(\|\Delta x\|).$$

Furthermore, Ξ is said to be strongly semismooth at $x \in \mathcal{X}$ if Ξ is semismooth at x and for any $\Delta x \in \mathcal{X}$ and $V \in \partial\Xi(x + \Delta x)$ with $\Delta x \to 0$,

$$\Xi(x + \Delta x) - \Xi(x) - V\Delta x = O(\|\Delta x\|^2).$$

2.4 Metric projection operators

In this section, we shall introduce three metric projections over three nonconvex sets which are defined by

$$\mathcal{S}^n_+(r) := \{Z \in \mathcal{S}^n \,|\, Z \succeq 0, \, \mathrm{rank}(Z) \leq r\}\,, \tag{2.5}$$

$$\mathcal{S}^n(r) := \{Z \in \mathcal{S}^n \,|\, \mathrm{rank}(Z) \leq r\}\,, \tag{2.6}$$

$$\Re^{n_1 \times n_2}_r := \{Z \in \Re^{n_1 \times n_2} \,|\, \mathrm{rank}(Z) \leq r\}\,. \tag{2.7}$$

In order to study the metric projections over the above sets, which will be used in the Lagrangian dual formulation in Chapter 4, much more analysis is involved due to the non-convex nature of these sets.

We first discuss two metric projections over the sets $\mathcal{S}^n_+(r)$ and $\mathcal{S}^n(r)$ in the symmetric case. Let $X \in \mathcal{S}^n$ be arbitrarily chosen. Suppose that X has the spectral decomposition

$$X = P\Lambda(X)P^T\,, \tag{2.8}$$

where $\Lambda(X) := \mathrm{diag}(\lambda(X))$, $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$ are the eigenvalues of X being arranged in the non-increasing order and $P \in \mathcal{O}^n$ is a corresponding orthogonal matrix of orthonormal eigenvectors of X. In order to characterize the following

metric projections, we need the following Proposition. Suppose that $X \in \mathcal{S}^n$ has the spectral decomposition as in (2.8) and let $\mu_1 > \mu_2 > \cdots > \mu_s$ be the distinct eigenvalues of X. Define the following subsets of $\{1, \ldots, n\}$

$$\bar{\tau}_k := \{\, i \,|\, \lambda_i(X) = \mu_k \,\}, \quad k = 1, \ldots, s. \tag{2.9}$$

Proposition 2.2. *Let* $\Lambda(X) = \mathrm{diag}(\lambda_1(X), \lambda_2(X), \ldots, \lambda_n(X))$ *with* $\lambda_1(X) \geq \lambda_2(X) \geq \ldots \geq \lambda_n(X)$. *Let* $\bar{\tau}_k$, $k = 1, \ldots, s$ *be the corresponding subsets given by (2.9). Let* Q *be an orthogonal matrix such that* $Q^T \Lambda(X) Q = \Lambda(X)$. *Then, we have*

$$\begin{cases} Q_{\bar{\tau}_k \bar{\tau}_l} = \mathbf{0}_{\bar{\tau}_k \bar{\tau}_l}, & k, l = 1, \cdots, s, \ k \neq l, \\ Q_{\bar{\tau}_k \bar{\tau}_k} Q^T_{\bar{\tau}_k \bar{\tau}_k} = Q^T_{\bar{\tau}_k \bar{\tau}_k} Q_{\bar{\tau}_k \bar{\tau}_k} = I_{|\bar{\tau}_k|}, & k = 1, \cdots, s. \end{cases} \tag{2.10}$$

2.4.1 Projection onto the nonconvex set $\mathcal{S}^n_+(r)$

Let $X \in \mathcal{S}^n$ have the spectral decomposition as in (2.8), i.e., $X = P\Lambda(X)P^T$. Define

$$\alpha := \{i \mid \lambda_i(X) > \lambda_r(X)\}, \ \beta := \{i \mid \lambda_i(X) = \lambda_r(X)\}, \text{ and } \gamma := \{i \mid \lambda_i(X) < \lambda_r(X)\}$$

and write $P = [P_\alpha \ P_\beta \ P_\gamma]$.

Denote

$$\begin{array}{rll} \Psi^s_r(X) := & \min & \dfrac{1}{2}\|Z - X\|^2 \\ & \text{s.t.} & Z \in \mathcal{S}^n_+(r)\,. \end{array} \tag{2.11}$$

Denote the set of optimal solutions to (2.11) by $\Pi_{\mathcal{S}^n_+(r)}(X)$, which is called the metric projection of X over $\mathcal{S}^n_+(r)$.

In order to characterize the solution set $\Pi_{\mathcal{S}^n_+(r)}(X)$, we need the Ky Fan's inequality given in the following lemma (e.g., see [3, (IV.62)]).

Lemma 2.3. *Any matrices* X *and* Y *in* $\mathcal{S}^n$ *satisfy the inequality*

$$\|X - Y\| \geq \|\lambda(X) - \lambda(Y)\|\,, \tag{2.12}$$

where the equality holds if and only if X and Y have a simultaneous ordered spectral decomposition.

Define $Z^* \in \mathcal{S}^n$ by

$$Z^* = \sum_{i=1}^{r} \big(\lambda_i(X)\big)_+ P_i P_i^T \,. \tag{2.13}$$

Thus, from the Ky Fan's inequality (2.12) and the fact that $Z^* \in \mathcal{S}^n_+(r)$, we obtain that $Z^* \in \Pi_{\mathcal{S}^n_+(r)}(X)$ and

$$\Psi^s_r(X) = \frac{1}{2}\sum_{i=1}^{r} \big((\lambda_i(X))_+ - \lambda_i(X)\big)^2 + \frac{1}{2}\sum_{i=r+1}^{n} \lambda_i^2(X) = \frac{1}{2}\sum_{i=t+1}^{n} \lambda_i^2(X)\,, \tag{2.14}$$

where k denotes the number of positive eigenvalues of X, i.e., $k := \big|\{i \,|\, \lambda_i(X) > 0\}\big|$ and $t := \min(r, k)$.

Lemma 2.4. *Let $X \in \mathcal{S}^n$ have the spectral decomposition as in (2.8). Then the solution set $\Pi_{\mathcal{S}^n_+(r)}(X)$ to problem (2.11) can be characterized as follows*

$$\Pi_{\mathcal{S}^n_+(r)}(X) = \Big\{ [\, P_\alpha \;\; P_\beta Q_\beta \;\; P_\gamma \,]\mathrm{diag}(v)\, [\, P_\alpha \;\; P_\beta Q_\beta \;\; P_\gamma \,]^T \Big|\; Q_\beta \in \mathcal{O}^{|\beta|} \Big\}\,, \tag{2.15}$$

where $v = \big((\lambda_1(X))_+, \ldots, (\lambda_r(X))_+, 0, \ldots, 0\big)^T \in \Re^n$.

Proof. By employing the Ky Fan's inequality and noting (2.14), we have for any $\overline{Z} \in \Pi_{\mathcal{S}^n_+(r)}(X)$, that

$$\|\overline{Z} - X\|^2 \geq \|\lambda(\overline{Z}) - \lambda(X)\|^2 \geq \sum_{i=1}^{r} \big((\lambda_i(X))_+ - \lambda_i(X)\big)^2 + \sum_{i=r+1}^{n} \lambda_i^2(X)\,, \tag{2.16}$$

which implies that there exists $U \in \mathcal{O}^n$ such that X and $\overline{Z}$ admit a simultaneous ordered spectral decomposition as

$$X = U\Lambda(X)U^T \quad \text{and} \quad \overline{Z} = U\Lambda(\overline{Z})U^T. \tag{2.17}$$

As $\Lambda(X)$ is arranged in the non-increasing order and from (2.16), we obtain that

$$\lambda(\overline{Z}) := v = \big((\lambda_1(X))_+, \ldots, (\lambda_r(X))_+, 0, \ldots, 0\big)^T. \tag{2.18}$$

Thus, by noting that $X = P\Lambda(X)P^T$ and then applying Proposition 2.2, we obtain that

$$\overline{Z} = U\Lambda(\overline{Z})U^T = (PQ)\mathrm{diag}(v)(PQ)^T\,,$$

where $Q \in \mathcal{O}^n$ takes the form as in (2.10). Notice that the two matrices $\sum_{i\in\alpha} v_i U_i U_i^T$ and $\sum_{i\in\gamma} v_i U_i U_i^T$ are independent of the choices of $Q \in \mathcal{O}^n$ satisfying (2.10). Thus, we can easily derive the conclusion (2.15) and complete the proof. □

Since $\Psi_r^s(X)$ takes the same value as in (2.14) for any element in $\Pi_{\mathcal{S}_+^n(r)}(X)$, for notational convenience, with no ambiguity, we use $\frac{1}{2}\|\Pi_{\mathcal{S}_+^n(r)}(X) - X\|^2$ to represent $\Psi_r^s(X)$. Define $\Xi_r^s : \mathcal{S}^n \to \Re$ by

$$\Xi_r^s(Z) = -\frac{1}{2}\|\Pi_{\mathcal{S}_+^n(r)}(Z) - Z\|^2 + \frac{1}{2}\|Z\|^2\,,\ Z \in \mathcal{S}^n\,. \tag{2.19}$$

Then we have

$$\Xi_r^s(X) = \frac{1}{2}\sum_{i=1}^{t}\lambda_i^2(X) = \frac{1}{2}\sum_{i=1}^{r}\big(\lambda_i(X)\big)_+^2 = \frac{1}{2}\|\Pi_{\mathcal{S}_+^n(r)}(X)\|^2\,,$$

where $\|\Pi_{\mathcal{S}_+^n(r)}(X)\|$ is interpreted as $\|\overline{Z}\|$ for any $\overline{Z} \in \Pi_{\mathcal{S}_+^n(r)}(X)$, e.g., the matrix Z^* defined by (2.13). By noting that for any $Z \in \mathcal{S}^n$, $\Xi_r^s(Z)$ can be reformulated as

$$\begin{aligned}\Xi_r^s(Z) &= \max_{Y\in\mathcal{S}_+^n(r)}\Big\{\frac{1}{2}\|Z\|^2 - \frac{1}{2}\|Y - Z\|^2\Big\}\\ &= \max_{Y\in\mathcal{S}_+^n(r)}\Big\{\langle Y, Z\rangle - \frac{1}{2}\|Y\|^2\Big\}\,,\end{aligned} \tag{2.20}$$

we know that $\Xi_r^s(\cdot)$ is a convex function as it is the maximum of infinitely many affine functions.

Proposition 2.5. *Let $X \in \mathcal{S}^n$ have the spectral decomposition as in (2.8). Then*

$$\partial\Xi_r^s(X) = \mathrm{conv}\,\Pi_{\mathcal{S}_+^n(r)}(X)\,. \tag{2.21}$$

Proof. For any $z \in \Re^n$, define

$$\xi_r(z) = \max_{y \in \mathcal{F}_r} \left\{ \frac{1}{2}\|z\|^2 - \frac{1}{2}\|y - z\|^2 \right\} = \max_{y \in \mathcal{F}_r} \left\{ \langle y, z \rangle - \frac{1}{2}\|y\|^2 \right\}, \tag{2.22}$$

where $\mathcal{F}_r := \{y \in \Re^n \,|\, \|y\|_0 \leq r, y \geq 0\}$. Then $\xi_r(\cdot)$ is a convex function and its sub-differential is well defined. Let $x = \lambda(X)$. Thus

$$\xi_r(x) = \max_{y \in \mathcal{F}_r} \left\{ \frac{1}{2}\|x\|^2 - \frac{1}{2}\|y - x\|^2 \right\} = \frac{1}{2}\|\lambda(X)\|^2 - \min_{y \in \mathcal{F}_r} \frac{1}{2}\|y - x\|^2 . \tag{2.23}$$

Denote the solution set of (2.23) by $\mathcal{F}_r^*$. Thus, from the non-increasing order of $\lambda(X)$, one can easily show that

$$\xi_r(x) = \frac{1}{2}\sum_{i=1}^{r}(x_i)_+^2 \quad \text{and} \quad \mathcal{F}_r^* = \mathcal{V}, \tag{2.24}$$

where

$$\begin{aligned} \mathcal{V} := \quad \big\{ v \in \Re^n \,|\quad & v_i = \big(\lambda_i(X)\big)_+ \text{ for } i \in \alpha \cup \beta_1,\ v_i = 0 \text{ for } i \in (\beta \setminus \beta_1) \cup \gamma\,, \\ & \text{where } \beta_1 \subseteq \beta \text{ and } |\beta_1| = r - |\alpha| \ \big\}. \end{aligned} \tag{2.25}$$

From convex analysis [105], we can easily derive that

$$\partial \xi_r(x) = \operatorname{conv} \mathcal{V}$$

and that $\xi_r(\cdot)$ is differentiable at x if and only if $\lambda_r(X) > \lambda_{r+1}(X) > 0$ or $\lambda_{r+1}(X) \leq 0$. In the latter case,

$$\partial \xi_r(x) = \{\nabla \xi_r(x)\} = \{v\},$$

where v is defined in (2.18), i.e., $v = \big((\lambda_1(X))_+, \ldots, (\lambda_r(X))_+, 0, \ldots, 0\big)^T \in \Re^n$. Since the convex function $\xi_r(\cdot)$ is symmetric, i.e., $\xi_r(z) = \xi_r(Sz)$ for any $z \in \Re^n$ and any permutation matrix S, from [64, Theorem 1.4], we know that $\Xi_r^s(\cdot)$ is differentiable at $X \in \mathcal{S}^n$ if and only if $\xi_r(\cdot)$ is differentiable at $\lambda(X)$ and

$$\partial \Xi_r^s(X) = \big\{ P \operatorname{diag}(v) P^T \,|\, v \in \partial \xi_r(\lambda(X))\,,\ P \Lambda(X) P^T = X\,,\ P \in \mathcal{O}^n \big\}.$$

Thus $\Xi_r^s(\cdot)$ is differentiable at X if and only if $\lambda_r(X) > \lambda_{r+1}(X) > 0$ or $\lambda_{r+1}(X) \leq 0$. In the latter case,

$$\partial \Xi_r^s(X) = \{\, (\Xi_r^s)'(X) \,\} = \{\, P\mathrm{diag}(v)P^T \,\}.$$

Let the B-subdifferential of $\Xi_r^s(\cdot)$ at X be defined by

$$\partial_B \Xi_r^s(X) = \Big\{ \lim_{X^k \to X} (\Xi_r^s)'(X^k)\,,\ \Xi_r^s(\cdot) \text{ is differentiable at } X^k \Big\}.$$

Then we can easily check that

$$\partial_B \Xi_r^s(X) = \Pi_{\mathcal{S}_+^n(r)}(X)\,, \tag{2.26}$$

where we use the fact that the two matrices $\sum_{i\in\alpha} \lambda_i(X) P_i P_i^T$ and $\sum_{i\in\gamma} \lambda_i(X) P_i P_i^T$ are independent of the choices of $P \in \mathcal{O}^n$ satisfying (2.8). Thus, by Theorem 2.5.1 in [18], one has

$$\partial \Xi_r^s(X) = \mathrm{conv}\, \partial_B \Xi_r^s(X) = \mathrm{conv}\, \Pi_{\mathcal{S}_+^n(r)}(X)\,.$$

The proof is completed. □

Remark 2.6. *Proposition 2.5 implies that when $\lambda_r(X) > \lambda_{r+1}(X) > 0$ or $\lambda_{r+1}(X) \leq 0$, $\Xi_r^s(\cdot)$ is continuously differentiable near X and $(\Xi_r^s)'(X) = \Pi_{\mathcal{S}_+^n(r)}(X) = \{Z^*\}$, where Z^* is defined in (2.13).*

Remark 2.7. *Since, for a given symmetric positive definite matrix $W \in \mathcal{S}^n$, the following W-weighted problem*

$$\begin{array}{ll} \min & \dfrac{1}{2}\|W^{1/2}(Z - X)W^{1/2}\|^2 \\ \text{s.t.} & Z \in \mathcal{S}_+^n(r) \end{array} \tag{2.27}$$

admits the solution set as $W^{-\frac{1}{2}} \Pi_{\mathcal{S}_+^n(r)}(W^{\frac{1}{2}} X W^{\frac{1}{2}}) W^{-\frac{1}{2}}$, there is no difficulty to work out the corresponding results presented in Lemma 2.4 and Proposition 2.5 for this more general case.

Remark 2.8. *When $r = n$, the metric projection operator $\Pi_{\mathcal{S}^n_+(r)}(\cdot)$ reduces to the projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ over the closed convex cone $\mathcal{S}^n_+$. Given $X \in \mathcal{S}^n$, $\Pi_{\mathcal{S}^n_+}(X)$ is the unique optimal solution to the following problem*

$$\begin{array}{ll} \min & \frac{1}{2}\|Z - X\|^2 \\ \text{s.t.} & Z \in \mathcal{S}^n_+ \,. \end{array} \tag{2.28}$$

It has long been known that $\Pi_{\mathcal{S}^n_+}(X)$ can be computed analytically (e.g., [109])

$$\Pi_{\mathcal{S}^n_+}(X) = P\mathrm{diag}\big((\lambda_1(X))_+, \ldots, (\lambda_n(X))_+\big)P^T \,. \tag{2.29}$$

For more properties about the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$, see [122, 75, 113] and references therein.

2.4.2 Projection onto the nonconvex set $\mathcal{S}^n(r)$

Let $Y \in \mathcal{S}^n$ be arbitrarily chosen. Suppose that Y has the spectral decomposition

$$Y = U\widehat{\Lambda}(Y)U^T, \tag{2.30}$$

where $U \in \mathcal{O}^n$ is a corresponding orthogonal matrix of orthonormal eigenvectors of Y and $\widehat{\Lambda}(Y) := \mathrm{diag}(\hat{\lambda}(Y))$ where $\hat{\lambda}(Y) = (\hat{\lambda}_1(Y), \ldots, \hat{\lambda}_n(Y))^T$ is the column vector containing all the eigenvalues of Y being arranged in the non-increasing order in terms of their absolute values, i.e.,

$$|\hat{\lambda}_1(Y)| \geq \cdots \geq |\hat{\lambda}_n(Y)| \,,$$

and whenever the equality holds, the larger one comes first, i.e.,

$$\text{if } |\hat{\lambda}_i(Y)| = |\hat{\lambda}_j(Y)| \text{ and } \hat{\lambda}_i(Y) > \hat{\lambda}_j(Y), \text{ then } i < j \,.$$

Define

$$\hat{\alpha} := \{i \mid |\hat{\lambda}_i(Y)| > |\hat{\lambda}_r(Y)|\}, \;\; \hat{\beta} := \{i \mid |\hat{\lambda}_i(Y)| = |\hat{\lambda}_r(Y)|\}, \;\; \hat{\gamma} := \{i \mid |\hat{\lambda}_i(Y)| < |\hat{\lambda}_r(Y)|\},$$

$$\text{and} \quad \hat{\beta}^+ := \{i \mid \hat{\lambda}_i(Y) = |\hat{\lambda}_r(Y)|\}, \quad \hat{\beta}^- := \{i \mid \hat{\lambda}_i(Y) = -|\hat{\lambda}_r(Y)|\}.$$

Write $U = [U_{\hat{\alpha}} \; U_{\hat{\beta}} \; U_{\hat{\gamma}}]$. Denote

$$\begin{array}{rll} \Psi_r^s(Y) := & \min & \frac{1}{2}\|Z - Y\|^2 \\ & \text{s.t.} & Z \in \mathcal{S}^n(r)\,. \end{array} \tag{2.31}$$

Denote the set of optimal solutions to (2.31) by $\Pi_{\mathcal{S}^n(r)}(Y)$, which is called the metric projection of Y over $\mathcal{S}^n(r)$. Define $V \in \mathcal{O}^n$ by

$$V = U\mathrm{diag}(v),$$

where for each $i \in \{1, \ldots, n\}$, $v_i = \hat{\lambda}_i(Y)/|\hat{\lambda}_i(Y)|$ if $\hat{\lambda}_i(Y) \neq 0$ and $v_i = 1$ otherwise. Then, we have

$$Y = U\mathrm{diag}(|\hat{\lambda}(Y)|)V^T.$$

Define $Z^* \in \mathcal{S}^n$ by

$$Z^* := \sum_{i=1}^{r} |\hat{\lambda}(Y)|_i U_i V_i^T = \sum_{i=1}^{r} |\hat{\lambda}_i(Y)| U_i (v_i U_i^T) = \sum_{i=1}^{r} \hat{\lambda}_i(Y) U_i U_i^T\,. \tag{2.32}$$

Thus, by using the fact that $Z^* \in \mathcal{S}^n(r)$, we have

$$Z^* \in \Pi_{\mathcal{S}^n(r)}(Y) \quad \text{and} \quad \Psi_r^s(Y) = \frac{1}{2}\sum_{i=r+1}^{n} \hat{\lambda}_i^2(Y)\,. \tag{2.33}$$

Then we can fully characterize all the solutions to problem (2.31) in the following lemma.

Lemma 2.9. *Let $Y \in \mathcal{S}^n$ have the spectral decomposition as in (2.30). Then the solution set $\Pi_{\mathcal{S}^n(r)}(Y)$ to problem (2.31) can be characterized as follows*

$$\begin{array}{rcl} \Pi_{\mathcal{S}^n(r)}(Y) & = & \Bigg\{ [\,U_{\hat{\alpha}} \; U_{\hat{\beta}} Q_{\hat{\beta}} \; U_{\hat{\gamma}}\,]\mathrm{diag}(v)\,[\,U_{\hat{\alpha}} \; U_{\hat{\beta}} Q_{\hat{\beta}} \; U_{\hat{\gamma}}\,]^T \Big| \\ & & v \in \mathcal{V}\,, Q_{\hat{\beta}} = \begin{bmatrix} Q_{\hat{\beta}^+} & 0 \\ 0 & Q_{\hat{\beta}^-} \end{bmatrix}, Q_{\hat{\beta}^+} \in \mathcal{O}^{|\hat{\beta}^+|}, Q_{\hat{\beta}^-} \in \mathcal{O}^{|\hat{\beta}^-|} \Bigg\}, \end{array} \tag{2.34}$$

where

$$\mathcal{V} := \big\{ v \in \Re^n \,|\, \ v_i = \hat{\lambda}_i(Y) \ \textit{for} \ \ i \in \hat{\alpha} \cup \hat{\beta}_1, \ v_i = 0 \ \textit{for} \ i \in (\hat{\beta} \setminus \hat{\beta}_1) \cup \hat{\gamma} \,, \\ \textit{where} \ \hat{\beta}_1 \subseteq \hat{\beta} \ \textit{and} \ |\hat{\beta}_1| = r - |\hat{\alpha}| \ \big\} \,. \tag{2.35}$$

Proof. By employing Ky Fan's inequality (2.12), i.e.,

$$\|Z - Y\| \geq \|\lambda(Z) - \lambda(Y)\|, \quad Z \in \mathcal{S}^n \,,$$

where the equality holds if and only if Y and Z admit a simultaneous ordered spectral decomposition, together with (2.33), we have for any $\overline{Z} \in \Pi_{\mathcal{S}^n(r)}(Y)$, that

$$\sum_{i=r+1}^{n} \hat{\lambda}_i^2(Y) = \|\overline{Z} - Y\|^2 \geq \|\lambda(\overline{Z}) - \lambda(Y)\|^2 \geq \sum_{i=r+1}^{n} \hat{\lambda}_i^2(Y) \,, \tag{2.36}$$

which implies that there exists $P \in \mathcal{O}^n$ such that Y and $\overline{Z}$ admit the spectral decompositions as in (2.8) with the same orthogonal matrix P.

$$Y = P\Lambda(Y)P^T \quad \text{and} \quad \overline{Z} = P\Lambda(\overline{Z})P^T. \tag{2.37}$$

Note that there exists a permutation matrix $S \in \Re^{n\times n}$ such that $\hat{\lambda}(Y) = S\lambda(Y)$. Under this permutation matrix, $v = S\lambda(\overline{Z})$ for some $v \in \mathcal{V}$ defined in (2.35) and

$$\widehat{\Lambda}(Y) = S\Lambda(Y)S^T \quad \text{and} \quad \mathrm{diag}(v) = S\Lambda(\overline{Z})S^T. \tag{2.38}$$

Noting that $Y = U\widehat{\Lambda}(Y)U^T$, one has

$$Y = P\Lambda(Y)P^T = PS^T\widehat{\Lambda}(Y)SP^T = U\widehat{\Lambda}(Y)U^T \,.$$

Then, by applying Proposition 2.2, we obtain that

$$\overline{Z} = P\Lambda(\overline{Z})P^T = PS^T\mathrm{diag}(v)SP^T = (UQ)\mathrm{diag}(v)(UQ)^T,$$

where $Q \in \mathcal{O}^n$ takes the form as in (2.10). Notice that the two matrices $\sum_{i\in\hat{\alpha}} v_i U_i U_i^T$ and $\sum_{i\in\hat{\gamma}} v_i U_i U_i^T$ are independent of the choices of $Q \in \mathcal{O}^n$ satisfying (2.10), thus we can easily derive the conclusion (2.34) and complete the proof. □

Since $\Psi_r^s(Y)$ takes the same value as in (2.33) for any element in $\Pi_{\mathcal{S}^n(r)}(Y)$, for notational convenience, with no ambiguity, we use $\frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(Y) - Y\|^2$ to represent $\Psi_r^s(Y)$.

Define $\Xi_r^s : \mathcal{S}^n \to \Re$ by

$$\Xi_r^s(Z) = -\frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(Z) - Z\|^2 + \frac{1}{2}\|Z\|^2\,,\ Z \in \mathcal{S}^n\,. \tag{2.39}$$

Then we have

$$\Xi_r^s(Y) = \frac{1}{2}\sum_{i=1}^{r}\hat{\lambda}_i^2(Y) = \frac{1}{2}\|\Pi_{\mathcal{S}^n(r)}(Y)\|^2\,,$$

where $\|\Pi_{\mathcal{S}^n(r)}(Y)\|$ is interpreted as $\|\overline{Z}\|$ for any $\overline{Z} \in \Pi_{\mathcal{S}^n(r)}(Y)$, e.g., the matrix Z^* defined by (2.32). By noting that for any $Z \in \mathcal{S}^n$, $\Xi_r^s(Z)$ can be reformulated as

$$\begin{aligned}\Xi_r(Z) &= \max_{X\in\mathcal{S}^n(r)}\left\{\frac{1}{2}\|Z\|^2 - \frac{1}{2}\|X - Z\|^2\right\}\\ &= \max_{X\in\mathcal{S}^n(r)}\left\{\langle X, Z\rangle - \frac{1}{2}\|X\|^2\right\},\end{aligned} \tag{2.40}$$

we know that $\Xi_r^s(\cdot)$ is a convex function as it is the maximum of infinitely many affine functions.

Proposition 2.10. *Let $Y \in \mathcal{S}^n$ have the spectral decomposition as in (2.30). Then*

$$\partial\Xi_r^s(Y) = \operatorname{conv}\Pi_{\mathcal{S}^n(r)}(Y)\,. \tag{2.41}$$

Proof. For any $z \in \Re^n$, define

$$\xi_r(z) = \max_{x\in\mathcal{F}_r}\left\{\frac{1}{2}\|z\|^2 - \frac{1}{2}\|x - z\|^2\right\} = \max_{x\in\mathcal{F}_r}\left\{\langle x, z\rangle - \frac{1}{2}\|x\|^2\right\}, \tag{2.42}$$

where $\mathcal{F}_r := \{x \in \Re^n \,|\, \|x\|_0 \le r\}$. Then $\xi_r(\cdot)$ is a convex function and its subdifferential is well defined. Let $y = \hat{\lambda}(Y)$. Thus

$$\xi_r(y) = \max_{x\in\mathcal{F}_r}\left\{\frac{1}{2}\|y\|^2 - \frac{1}{2}\|x - y\|^2\right\} = \frac{1}{2}\|\hat{\lambda}(Y)\|^2 - \min_{x\in\mathcal{F}_r}\frac{1}{2}\|x - y\|^2\,. \tag{2.43}$$

Denote the solution set of (2.43) by $\mathcal{F}_r^*$. Define $\alpha := \{i \,|\, x_i \neq 0\}$ and $\beta := \{i \,|\, x_i = 0\}$ for any given $x \in \mathcal{F}_r$. It then follows that

$$\sum_{i=1}^{n}(x_i - y_i)^2 = \sum_{i\in\alpha}(x_i - y_i)^2 + \sum_{i\in\beta}(x_i - y_i)^2 \geq \sum_{i\in\beta} y_i^2 \geq \sum_{i=r+1}^{n} y_i^2\,, \tag{2.44}$$

where the last inequality is from the facts that $|\beta| \geq n - r$ and the non-increasing order of y in terms of the absolute value. Therefore, we know that

$$\xi_r(y) = \frac{1}{2}\sum_{i=1}^{r} y_i^2 \quad \text{and} \quad \mathcal{F}_r^* = \mathcal{V}\,, \tag{2.45}$$

where $\mathcal{V}$ is defined in (2.35). From convex analysis [105], we can easily derive that

$$\partial\xi_r(y) = \operatorname{conv}\mathcal{V}$$

and that $\xi_r(\cdot)$ is differentiable at y if and only if $|\hat{\lambda}(Y)|_r > |\hat{\lambda}(Y)|_{r+1}$. In the latter case,

$$\partial\xi_r(y) = \{\nabla\xi_r(y)\} = \big\{v \in \Re^n \,|\, v_i = \hat{\lambda}_i(Y) \text{ for } 1 \leq i \leq r \text{ and } v_i = 0 \text{ for } r{+}1 \leq i \leq n\big\}.$$

Since the convex function $\xi_r(\cdot)$ is symmetric, i.e., $\xi_r(z) = \xi_r(Sz)$ for any $z \in \Re^n$ and any permutation matrix S, for $Z \in \mathcal{S}^n$ we can rewrite $\Xi_r^s(Z)$ as

$$\Xi_r^s(Z) = \xi_r(\hat{\lambda}(Z)) = \xi_r(\lambda(Z))\,,$$

where $\hat{\lambda}(Z) = (\hat{\lambda}_1(z), \ldots, \hat{\lambda}_n(Z))^T$ is the column vector containing all the eigenvalues of Z being arranged in the non-increasing order in terms of their absolute values. By [64, Theorem 1.4], we know that $\Xi_r^s(\cdot)$ is differentiable at $Y \in \mathcal{S}^n$ if and only if $\xi_r(\cdot)$ is differentiable at $\hat{\lambda}(Y)$ and

$$\partial\Xi_r^s(Y) = \{U\operatorname{diag}(v)U^T \,|\, v \in \partial\xi_r(\hat{\lambda}(Y))\,,\, U \in \mathcal{O}^n,\, U\operatorname{diag}(\hat{\lambda}(Y))U^T = Y\,\}.$$

Thus $\Xi_r^s(\cdot)$ is differentiable at Y if and only if $|\hat{\lambda}(Y)|_r > |\hat{\lambda}(Y)|_{r+1}$. In the latter case,

$$\partial\Xi_r^s(Y) = \{(\Xi_r^s)'(Y)\} = \big\{U\operatorname{diag}(v)U^T \,|\, v_i = \hat{\lambda}_i(Y) \text{ for } 1 \leq i \leq r \text{ and } v_i = 0 \text{ for } r{+}1 \leq i \leq n\big\}.$$

Let the B-subdifferential of $\Xi_r^s(\cdot)$ at Y be defined by

$$\partial_B \Xi_r^s(Y) = \big\{ \lim_{Y^k \to Y} (\Xi_r^s)'(Y^k)\,,\ \Xi_r^s(\cdot) \text{ is differentiable at } Y^k \big\}.$$

Then we can easily check that

$$\partial_B \Xi_r^s(Y) = \Pi_{\mathcal{S}^n(r)}(Y)\,, \tag{2.46}$$

where we used the fact that the two matrices $\sum_{i\in\hat{\alpha}} \hat{\lambda}_i(Y) U_i U_i^T$ and $\sum_{i\in\hat{\gamma}} \hat{\lambda}_i(Y) U_i U_i^T$ are independent of the choices of $U \in \mathcal{O}^n$ satisfying (2.30). Thus, by Theorem 2.5.1 in [18], one has

$$\partial \Xi_r^s(Y) = \operatorname{conv} \partial_B \Xi_r^s(Y) = \operatorname{conv} \Pi_{\mathcal{S}^n(r)}(Y)\,.$$

The proof is completed. □

Remark 2.11. *Proposition 2.10 implies that when* $|\hat{\lambda}_r(Y)| > |\hat{\lambda}_{r+1}(Y)|$, $\Xi_r^s(\cdot)$ *is continuously differentiable near* Y *and* $(\Xi_r^s)'(Y) = \Pi_{\mathcal{S}^n(r)}(Y) = \{Z^*\}$, *where* Z^* *is defined in (2.32).*

Remark 2.12. *Since, for a given symmetric positive definite matrix* $W \in \mathcal{S}^n$, *the following* W*-weighted problem*

$$\begin{array}{ll} \min & \dfrac{1}{2}\|W^{1/2}(Z-Y)W^{1/2}\|^2 \\ \text{s.t.} & Z \in \mathcal{S}^n(r) \end{array} \tag{2.47}$$

admits the solution set as $W^{-\frac{1}{2}} \Pi_{\mathcal{S}^n(r)}(W^{\frac{1}{2}} Y W^{\frac{1}{2}}) W^{-\frac{1}{2}}$, *there is no difficulty to work out the corresponding results presented in Lemma 2.9 and Proposition 2.10 for this more general case.*

2.4.3 Generalized projection onto the nonconvex set $\Re_r^{n_1\times n_2}$

Let $Y \in \Re^{n_1\times n_2}(n_1 \le n_2)$ admit the singular value decomposition

$$Y = U[\Sigma(Y) \ \ \mathbf{0}]V^T = U[\Sigma(Y) \ \ \mathbf{0}][V_1 \ \ V_2]^T, \tag{2.48}$$

where $U \in \mathcal{O}^{n_1}$ and $V \in \mathcal{O}^{n_2}$ are orthogonal matrices, and $\Sigma(Y) := \mathrm{diag}(\sigma(Y))$, $\sigma_1(Y) \geq \cdots \geq \sigma_{n_1}(Y) \geq 0$ are the singular values of Y being arranged in the non-increasing order. Decompose $V \in \mathcal{O}^{n_2}$ into the form $V = [V_1 \;\; V_2]$, where $V_1 \in \Re^{n_2 \times n_1}$ and $V_2 \in \Re^{n_2 \times (n_2 - n_1)}$. The set of such matrix pairs (U, V) in the singular value decomposition (2.48) is denoted by $\mathcal{O}^{n_1,n_2}(Y)$. Denote the nuclear norm of Y by $\|Y\|_* = \sum_{i=1}^{n_1} \sigma_i(Y)$.

Define the index sets of positive and zero singular values of Y, by

$$\tau := \{i \,|\, \sigma_i(Y) > 0\} \quad \text{and} \quad \tau_0 := \{i \,|\, \sigma_i(Y) = 0\}\,.$$

Let $\nu_1 > \nu_2 > \ldots > \nu_t > 0$ be the nonzero distinct singular values of Y. Let $\{\tau_k\}_{k=1}^t$ be a partition of τ, which is given by

$$\tau_k := \{i \,|\, \sigma_i(Y) = \nu_k\}, \quad k = 1, \ldots, t\,.$$

Proposition 2.13. *For any given $Y \in \Re^{n_1 \times n_2}$, denote $\Sigma := \Sigma(Y)$. Let $P \in \mathcal{O}^{n_1}$ and $W \in \mathcal{O}^{n_2}$ satisfy*

$$P[\Sigma \;\; \mathbf{0}] = [\Sigma \;\; \mathbf{0}]W\,. \tag{2.49}$$

Then, there exist $Q \in \mathcal{O}^{|\tau|}$, $Q' \in \mathcal{O}^{n_1 - |\tau|}$ and $Q'' \in \mathcal{O}^{n_2 - |\tau|}$ such that

$$P = \begin{bmatrix} Q & \mathbf{0} \\ \mathbf{0} & Q' \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} Q & \mathbf{0} \\ \mathbf{0} & Q'' \end{bmatrix}.$$

Moreover, the orthogonal matrix Q is a block diagonal matrix which takes the form as follows:

$$\begin{aligned} &Q_{\tau_k \tau_l} = 0_{\tau_k \tau_l}, && k, l = 1, \cdots, t,\ k \neq l, \\ &Q_{\tau_k \tau_k} Q^T_{\tau_k \tau_k} = Q^T_{\tau_k \tau_k} Q_{\tau_k \tau_k} = I_{|\tau_k|}, && k = 1, \cdots, t. \end{aligned} \tag{2.50}$$

For the proof of this proposition, see [29].

Now we discuss the generalized metric projection over the set $\Re_r^{n_1\times n_2}$ in the non-symmetric case. Denote

$$\begin{aligned}\Psi_{\rho,r}(Y) := \quad & \min \quad \frac{1}{2}\|Z-Y\|^2+\rho\|Z\|_* \\ & \text{s.t.} \quad Z\in\Re_r^{n_1\times n_2}\,,\end{aligned} \tag{2.51}$$

where $\rho\geq 0$ is a given parameter. Denote the set of optimal solutions to (2.51) by $\mathcal{P}_{\rho,r}(Y)$, which is called the generalized metric projection of Y over $\Re_r^{n_1\times n_2}$.

In order to characterize the solution set $\mathcal{P}_{\rho,r}(Y)$, we need the von Neumann's trace inequality first proved by von Neumann [83]. For the condition when the equality holds, see [66].

Lemma 2.14. *Any matrices X and Y in $\Re^{n_1\times n_2}$ satisfy $trX^TY\leq\sigma(X)^T\sigma(Y)$, where X and Y have the singular value decomposition as in (2.48). The equality holds if and only if X and Y have a simultaneous ordered singular value decomposition.*

Define $Z^*\in\Re^{n_1\times n_2}$ by

$$Z^* := \sum_{i=1}^{r}\big(\sigma_i(Y)-\rho\big)_+U_iV_i^T\,. \tag{2.52}$$

By noting that von Neumann's trace inequality implies

$$\|Z-Y\|\geq\|\sigma(Z)-\sigma(Y)\|,\quad\forall\,Z\in\Re^{n_1\times n_2}\,, \tag{2.53}$$

we obtain that for any $Z \in \Re_r^{n_1 \times n_2}$,

$$\begin{aligned}
& \frac{1}{2}\|Z-Y\|^2 + \rho\|Z\|_* \\
\geq\ & \frac{1}{2}\|\sigma(Z)-\sigma(Y)\|^2 + \rho\|Z\|_* \\
=\ & \frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Z)-\sigma_i(Y)\big)^2 + \frac{1}{2}\sum_{i=r+1}^{n_1}\sigma_i^2(Y) + \rho\sum_{i=1}^{r}\sigma_i(Z) \\
=\ & \frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Z)-(\sigma_i(Y)-\rho)\big)^2 + \frac{1}{2}\sum_{i=1}^{n_1}\sigma_i^2(Y) - \frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Y)-\rho\big)^2 \\
\geq\ & \frac{1}{2}\sum_{i=1}^{r}\big((\sigma_i(Y)-\rho)_+ - (\sigma_i(Y)-\rho)\big)^2 + \frac{1}{2}\sum_{i=1}^{n_1}\sigma_i^2(Y) - \frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Y)-\rho\big)^2 \\
=\ & -\frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Y)-\rho\big)_+^2 + \frac{1}{2}\sum_{i=1}^{n_1}\sigma_i^2(Y) \\
=\ & \frac{1}{2}\|Z^*-Y\|^2 + \rho\|Z^*\|_*\,,
\end{aligned} \tag{2.54}$$

which, together with the fact that $Z^* \in \Re_r^{n_1 \times n_2}$, implies that $Z^* \in \mathcal{P}_{\rho,r}(Y)$ and

$$\Psi_{\rho,r}(Y) = -\frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Y)-\rho\big)_+^2 + \frac{1}{2}\sum_{i=1}^{n_1}\sigma_i^2(Y)\,. \tag{2.55}$$

Lemma 2.15. *Let $Y \in \Re^{n_1 \times n_2}$ have the singular value decomposition as in (2.48). Define the index sets by $\bar{\alpha} := \{i \mid \sigma_i(Y) > \sigma_r(Y)\}$, $\bar{\beta} := \{i \mid \sigma_i(Y) = \sigma_r(Y)\}$, and $\bar{\gamma} := \{i \mid \sigma_i(Y) < \sigma_r(Y)\}$. Then the solution set $\mathcal{P}_{\rho,r}(Y)$ to problem (2.51) can be characterized as follows*

$$\mathcal{P}_{\rho,r}(Y) = \big\{\,[\,U_{\bar{\alpha}}\ U_{\bar{\beta}}Q_{\bar{\beta}}\ U_{\bar{\gamma}}\,][\,\mathrm{diag}(v)\ \mathbf{0}\,][\,V_{1\bar{\alpha}}\ V_{1\bar{\beta}}Q_{\bar{\beta}}\ V_{1\bar{\gamma}}\ V_2\,]^T \mid Q_{\bar{\beta}} \in \mathcal{O}^{|\bar{\beta}|}\big\}, \tag{2.56}$$

where $v = \big((\sigma_1(Y)-\rho)_+, \cdots, (\sigma_r(Y)-\rho)_+, 0, \cdots, 0\big)^T \in \Re^{n_1}$.

Proof. By (2.55), we have that for any $\overline{Z} \in \mathcal{P}_{\rho,r}(Y)$,

$$\frac{1}{2}\|\overline{Z}-Y\|^2 + \rho\|\overline{Z}\|_* = -\frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Y)-\rho\big)_+^2 + \frac{1}{2}\sum_{i=1}^{n_1}\sigma_i^2(Y)\,,$$

which implies that the inequalities in (2.54) are both equalities if Z is replaced by $\overline{Z}$. Therefore, from the first inequality in (2.54), we can see that there exist

$\overline{U} \in \mathcal{O}^{n_1}$ and $\overline{V} \in \mathcal{O}^{n_2}$ such that Y and $\overline{Z}$ admit the singular value decompositions as in (2.48) with the same orthogonal matrices $\overline{U}$ and $\overline{V}$, i.e.,

$$Y = \overline{U}[\Sigma(Y) \;\; \mathbf{0}]\overline{V}^T \quad \text{and} \quad \overline{Z} = \overline{U}[\Sigma(\overline{Z}) \;\; \mathbf{0}]\overline{V}^T. \tag{2.57}$$

Moreover, by the second inequality in (2.54), together with the fact that $\sigma(Y)$ is arranged in the non-increasing order, we obtain that

$$\sigma(\overline{Z}) = v := \big((\sigma_1(Y) - \rho)_+, \cdots, (\sigma_r(Y) - \rho)_+, 0, \cdots, 0\big)^T. \tag{2.58}$$

Then from Proposition 2.13 and $Y = U[\Sigma(Y) \;\; \mathbf{0}]V^T$, we know that

$$\overline{Z} = \overline{U}[\Sigma(\overline{Z}) \;\; \mathbf{0}]\overline{V}^T = (UW_1)[\mathrm{diag}(v) \;\; \mathbf{0}](VW_2)^T,$$

with $W_1 \in \mathcal{O}^{n_1}$ and $W_2 \in \mathcal{O}^{n_2}$ taking the form

$$W_1 = \begin{bmatrix} Q & \mathbf{0} \\ \mathbf{0} & Q' \end{bmatrix} \quad \text{and} \quad W_2 = \begin{bmatrix} Q & \mathbf{0} \\ \mathbf{0} & Q'' \end{bmatrix}, \tag{2.59}$$

where $\tau := \{i \,|\, \sigma_i(Y) > 0\}$, $Q' \in \mathcal{O}^{n_1 - |\tau|}$, $Q'' \in \mathcal{O}^{n_2 - |\tau|}$ and $Q \in \mathcal{O}^{|\tau|}$ is a block diagonal matrix taking the same form as in (2.50). Notice that the two matrices $\sum_{i \in \bar{\alpha}} v_i U_i V_i^T$ and $\sum_{i \in \bar{\gamma}} v_i U_i V_i^T$ are independent of the choices of $W_1 \in \mathcal{O}^{n_1}$ and $W_2 \in \mathcal{O}^{n_2}$ satisfying (2.59), then the conclusion (2.51) holds. □

Since $\Psi_{\rho,r}(Y)$ takes the same value as in (2.14) for any element in $\mathcal{P}_{\rho,r}(Y)$, for notational convenience, with no ambiguity, we use $\frac{1}{2}\|\mathcal{P}_{\rho,r}(Y) - Y\|^2 + \rho\|Y\|_*$ to represent $\Psi_{\rho,r}(Y)$. Define $\Xi_{\rho,r} : \Re^{n_1 \times n_2} \to \Re$ by

$$\Xi_{\rho,r}(Z) = -\frac{1}{2}\|\mathcal{P}_{\rho,r}(Z) - Z\|^2 - \rho\|Z\|_* + \frac{1}{2}\|Z\|^2, \quad Z \in \Re^{n_1 \times n_2}. \tag{2.60}$$

Then we have

$$\Xi_{\rho,r}(Y) = \frac{1}{2}\sum_{i=1}^{r}\big(\sigma_i(Y) - \rho\big)_+^2 = \frac{1}{2}\|\mathcal{P}_{\rho,r}(Y)\|^2,$$

where $\|\mathcal{P}_{\rho,r}(Y)\|$ is interpreted as $\|\overline{Z}\|$ for any $\overline{Z} \in \mathcal{P}_{\rho,r}(Y)$, e.g., the matrix Z^* defined by (2.52). By noting that for any $Z \in \Re^{n_1 \times n_2}$, $\Xi_{\rho,r}(Z)$ can be reformulated as

$$\begin{aligned} \Xi_{\rho,r}(Z) &= \max_{X \in \Re_r^{n_1 \times n_2}} \left\{ \frac{1}{2}\|Z\|^2 - \frac{1}{2}\|X - Z\|^2 - \rho\|X\|_* \right\} \\ &= \max_{X \in \Re_r^{n_1 \times n_2}} \left\{ \langle X, Z \rangle - \frac{1}{2}\|X\|^2 - \rho\|X\|_* \right\}, \end{aligned} \tag{2.61}$$

we know that $\Xi_{\rho,r}(\cdot)$ is a convex function as it is the maximum of infinitely many affine functions.

Proposition 2.16. *Let $Y \in \Re^{n_1 \times n_2}$ have the spectral decomposition as in (2.48). Then*

$$\partial \Xi_{\rho,r}(Y) = \operatorname{conv} \mathcal{P}_{\rho,r}(Y)\,. \tag{2.62}$$

Proof. For any $z \in \Re^{n_1}$, define

$$\xi_{\rho,r}(z) = \max_{x \in \mathcal{F}_r} \left\{ \frac{1}{2}\|z\|^2 - \frac{1}{2}\|x - z\|^2 - \rho\|x\|_1 \right\} = \max_{x \in \mathcal{F}_r} \left\{ \langle x, z \rangle - \frac{1}{2}\|x\|^2 - \rho\|x\|_1 \right\}, \tag{2.63}$$

where $\mathcal{F}_r := \{x \in \Re^{n_1} \,|\, \|x\|_0 \le r\}$. Then $\xi_{\rho,r}(\cdot)$ is a convex function and its sub-differential is well defined. Let $y = \sigma(Y)$. Thus

$$\xi_{\rho,r}(y) = \max_{x \in \mathcal{F}_r} \left\{ \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x-y\|^2 - \rho\|x\|_1 \right\} = \frac{1}{2}\|\sigma(Y)\|^2 - \min_{x \in \mathcal{F}_r} \left\{ \frac{1}{2}\|x-y\|^2 + \rho\|x\|_1 \right\}. \tag{2.64}$$

Denote the solution set of (2.64) by $\mathcal{F}_r^*$. Thus, from the non-increasing order of $\sigma(Y)$, one can easily show that

$$\xi_{\rho,r}(y) = \frac{1}{2}\sum_{i=1}^{r} (y_i - \rho)_+^2 \quad \text{and} \quad \mathcal{F}_r^* = \mathcal{V}, \tag{2.65}$$

where

$$\begin{aligned} \mathcal{V} := \; \big\{ v \in \Re^{n_1} \,|\; & v_i = \big(\sigma_i(Y) - \rho\big)_+ \text{ for } i \in \bar{\alpha} \cup \bar{\beta}_1,\; v_i = 0 \text{ for } i \in (\bar{\beta} \setminus \bar{\beta}_1) \cup \bar{\gamma}\,, \\ & \text{where } \bar{\beta}_1 \subseteq \bar{\beta} \text{ and } |\bar{\beta}_1| = r - |\bar{\alpha}| \;\big\}. \end{aligned} \tag{2.66}$$

From convex analysis [105], we can easily derive that

$$\partial\xi_{\rho,r}(y) = \operatorname{conv}\mathcal{V}$$

and that $\xi_{\rho,r}(\cdot)$ is differentiable at y if and only if $\sigma_r(Y) > \sigma_{r+1}(Y) > 0$ or $\sigma_{r+1}(Y) = 0$. In the latter case,

$$\partial\xi_{\rho,r}(y) = \{\nabla\xi_{\rho,r}(y)\} = \{v\},$$

where v is defined in (2.58), i.e., $v = \big((\sigma_1(Y)-\rho)_+, \ldots, (\sigma_r(Y)-\rho)_+, 0, \ldots, 0\big)^T \in \Re^{n_1}$.

Since the convex function $\xi_{\rho,r}(\cdot)$ is absolutely symmetric, i.e., $\xi_{\rho,r}(z) = \xi_{\rho,r}(Sz)$ for any $z \in \Re^n$ and any generalized permutation matrix S which has exactly one nonzero entry in each row and column, that entry being ± 1. From [63, Theorem 3.1 & Corollary 2.5], we know that $\Xi_{\rho,r}(\cdot)$ is differentiable at $Y \in \Re^{n_1\times n_2}$ if and only if $\xi_{\rho,r}(\cdot)$ is differentiable at $\sigma(Y)$ and

$$\partial\Xi_{\rho,r}(Y) = \big\{U[\operatorname{diag}(v)\ \mathbf{0}]V^T \,|\, v \in \partial\xi_{\rho,r}(\sigma(Y))\,,\ (U,V) \in \mathcal{O}^{n_1,n_2}(Y)\big\}.$$

Thus $\Xi_{\rho,r}(\cdot)$ is differentiable at Y if and only if $\sigma_r(Y) > \sigma_{r+1}(Y) > 0$ or $\sigma_{r+1}(Y) = 0$. In the latter case,

$$\partial\Xi_{\rho,r}(Y) = \{\,\Xi'_{\rho,r}(Y)\,\} = \{\,U[\operatorname{diag}(v)\ \mathbf{0}]V^T\,\}.$$

Let the B-subdifferential of $\Xi_{\rho,r}(\cdot)$ at Y be defined by

$$\partial_B\Xi_{\rho,r}(Y) = \big\{\lim_{Y^k\to Y}\Xi'_{\rho,r}(Y^k)\,,\ \Xi_{\rho,r}(\cdot) \text{ is differentiable at } Y^k\big\}.$$

Then we can easily check that

$$\partial_B\Xi_{\rho,r}(Y) = \mathcal{P}_{\rho,r}(Y)\,, \tag{2.67}$$

where we used the fact that the two matrices $\displaystyle\sum_{i\in\bar{\alpha}}\sigma_i(Y)U_iV_i^T$ and $\displaystyle\sum_{i\in\bar{\gamma}}\sigma_i(Y)U_iV_i^T$ are independent of the choices of $(U,V) \in \mathcal{O}^{n_1,n_2}(Y)$ satisfying (2.48). Thus, by

Theorem 2.5.1 in [18], one has

$$\partial \Xi_{\rho,r}(Y) = \operatorname{conv} \partial_B \Xi_{\rho,r}(Y) = \operatorname{conv} \mathcal{P}_{\rho,r}(Y)\,.$$

The proof is completed. □

Remark 2.17. *Proposition 2.16 implies that when $\sigma_r(Y) > \sigma_{r+1}(Y) > 0$ or $\sigma_{r+1}(Y) = 0$, $\Xi_{\rho,r}(\cdot)$ is continuously differentiable near Y and $\Xi'_{\rho,r}(Y) = \mathcal{P}_{\rho,r}(Y) = \{Z^*\}$, where Z^* is defined in (2.52).*

Remark 2.18. *Since, for given matrices $W_1 \in \Re^{n_1 \times n_2}$ and $W_2 \in \Re^{n_1 \times n_2}$, the following weighted problem*

$$\begin{array}{ll} \min & \dfrac{1}{2}\|W_1(Z-Y)W_2\|^2 + \rho\|W_1 Z W_2\|_* \\ \text{s.t.} & Z \in \Re_r^{n_1 \times n_2}\,, \end{array} \tag{2.68}$$

admits the solution set as ${W_1}^{-1}\mathcal{P}_{\rho,r}(W_1 Y W_2){W_2}^{-1}$, there is no difficulty to work out the corresponding results presented in Lemma 2.15 and Proposition 2.16 for this more general case.

Remark 2.19. *When $r = n_1$, $\mathcal{P}_{\rho,r}(\cdot)$ reduces to soft thresholding operator $\mathcal{P}_\rho(\cdot)$ [12]; when $\rho = 0$, the generalized metric projection $\mathcal{P}_{\rho,r}(\cdot)$ reduces to the metric projection $\Pi_{\Re_r^{n_1 \times n_2}}(\cdot)$.*

The equations (2.21) , (2.41) and (2.62) are particularly useful in developing a technique for global optimality checking in Chapter 4.

2.5 The smoothing functions

In this section, we shall introduce the smoothing functions for the real-valued nonsmooth function $t_+ := \max(0, t)$, which is not differentiable at $t = 0$.

Let $\phi_H : \Re \times \Re \to \Re$ be defined by the following Huber function for t_+

$$\phi_H(\varepsilon, t) = \begin{cases} t & \text{if } t \geq \dfrac{|\varepsilon|}{2}\,, \\ \dfrac{1}{2|\varepsilon|}(t + \dfrac{|\varepsilon|}{2})^2 & \text{if } -\dfrac{|\varepsilon|}{2} < t < \dfrac{|\varepsilon|}{2},\ (\varepsilon, t) \in \Re \times \Re\,. \\ 0 & \text{if } t \leq -\dfrac{|\varepsilon|}{2}, \end{cases} \tag{2.69}$$

and the Smale smoothing function $\phi_S \colon \Re \times \Re \to \Re$

$$\phi_S(\varepsilon, t) = [\, t + \sqrt{\varepsilon^2 + t^2}\,]/2\,, \quad (\varepsilon, t) \in \Re \times \Re\,. \tag{2.70}$$

Discussions on the properties of the smoothing functions can be found in [93, 124]. It has been known that both ϕ_H and ϕ_S are globally Lipschitz continuous, continuously differentiable around (ε, t) whenever $\varepsilon \neq 0$, and are strongly semismooth at $(0, t)$ (see [124] and references therein for details). Since ϕ_H and ϕ_S share similar differential properties, in the following, unless we specify we will use ϕ to denote the smoothing function either ϕ_H or ϕ_S.

2.6 The Slater condition

We consider the following problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \mathcal{A}x = b\,, \\ & x \in \mathcal{K}\,, \end{array} \tag{2.71}$$

where $\mathcal{A} : \mathcal{X} \to \Re^m$ is a linear mapping, $b \in \Re^m$ and $\mathcal{K}$ is a closed convex cone with nonempty interior. We always assume that $b \in \mathcal{A}\mathcal{X}$. The Slater condition for problem (2.71) is as follows

$$\mathcal{A} \text{ is onto and there exists } \ x^0 \in \mathrm{int}(\mathcal{K}) \text{ such that } \mathcal{A}x^0 = b\,. \tag{2.72}$$

Proposition 2.20. *If the Slater condition (2.72) holds for problem (2.71), then*

$$\langle b, y\rangle < 0 \text{ for any } 0 \neq y \in \Re^m \text{ satisfying } \mathcal{A}^* y \in -\mathcal{K}^* \,. \tag{2.73}$$

Conversely, if the condition (2.73) holds, then there exists $x \in \mathrm{int}(\mathcal{K})$ *such that* $\mathcal{A}x = b$.

Proof. Suppose that the Slater condition (2.72) holds. Then there exists $\bar{x} \in \mathrm{int}(\mathcal{K})$ such that $\mathcal{A}\bar{x} = b$. Let $0 \neq \bar{y} \in \Re^m$ such that $\mathcal{A}^*\bar{y} \in -\mathcal{K}^*$. Thus, $\mathcal{A}^*\bar{y} \neq 0$ from the fact that $\mathcal{A}$ is onto. Furthermore, $\bar{x} \in \mathrm{int}(\mathcal{K})$ implies that there exists $\delta > 0$ such that $\bar{x} + \delta\mathcal{A}^*\bar{y} \in \mathcal{K}$. It then follows that

$$\langle \bar{x},\, \mathcal{A}^*\bar{y}\rangle = \langle \bar{x} + \delta\mathcal{A}^*\bar{y},\, \mathcal{A}^*\bar{y}\rangle - \delta\langle \mathcal{A}^*\bar{y},\, \mathcal{A}^*\bar{y}\rangle \leq -\delta\langle \mathcal{A}^*\bar{y},\, \mathcal{A}^*\bar{y}\rangle < 0\,.$$

Therefore, $\langle b, \bar{y}\rangle = \langle \mathcal{A}\bar{x}\,,\, \bar{y}\,\rangle = \langle\, \bar{x},\, \mathcal{A}^*\bar{y}\,\rangle < 0$, which proves the first part of this proposition.

Next we prove the remaining part by contradiction. Define $\mathcal{S} := \{\, x \in \mathcal{X} \,|\, \mathcal{A}x = b\,\}$. Suppose that there does not exist $x \in \mathrm{int}(\mathcal{K})$ such that $\mathcal{A}x = b$, i.e.,

$$\{\, x \in \mathcal{X} \,|\, x \in \mathrm{int}(\mathcal{K})\,\} \cap \mathcal{S} = \emptyset\,.$$

Then, from Separation Theorem [105], there exists $0 \neq p \in \mathcal{X}$ such that

$$\langle p\,,\, y\rangle \leq \langle p\,,\, x\rangle\,, \qquad \forall\, y \in \mathrm{int}(\mathcal{K}) \text{ and } x \in \mathcal{S}\,. \tag{2.74}$$

As $\mathcal{K}$ is a closed convex cone, for any $\kappa > 0$

$$\kappa\langle p\,,\, y\rangle \leq \langle p\,,\, x\rangle\,, \qquad \forall\, y \in \mathrm{int}(\mathcal{K}) \text{ and } x \in \mathcal{S}\,.$$

It follows that

$$\langle p\,,\, y\rangle \leq \lim_{\kappa\to+\infty} \frac{\langle p\,,\, x\rangle}{\kappa} = 0\,, \qquad \forall\, y \in \mathrm{int}(\mathcal{K})\,.$$

Thus, $\langle p\,,\, y\rangle \leq 0$ for all $y \in \mathcal{K}$. That is, $p \in -\mathcal{K}^*$. Similarly, we can show that

$$\langle p\,,\, x\rangle \geq 0\,, \quad \forall\, x \in \mathcal{S}. \tag{2.75}$$

Let $\bar{x} \in \mathcal{S}$, i.e., $\mathcal{A}\bar{x} = b$. Then $\mathcal{S} = \bar{x} + \mathrm{Ker}\mathcal{A}$, where $\mathrm{Ker}\mathcal{A} = \{x \in \mathcal{X} \,|\, \mathcal{A}x = 0\}$ is the kernel of $\mathcal{A}$. Note that $\mathrm{Ker}\mathcal{A}$ is a subspace, together with (2.75), one can easily show that

$$\langle p\,,\, x^0\rangle = 0,\ \forall\, x^0 \in \mathrm{Ker}\mathcal{A} \quad \text{and} \quad \langle p\,,\, \bar{x}\rangle \geq 0\,.$$

Therefore, $0 \neq p \in \mathrm{Im}\mathcal{A}^*$. That is, there exists $0 \neq \bar{y} \in \Re^m$ such that $p = \mathcal{A}^*\bar{y} \in -\mathcal{K}^*$. It then follows that

$$\langle p\,,\, \bar{x}\rangle = \langle \mathcal{A}^*\bar{y},\, \bar{x}\rangle = \langle \bar{y},\, \mathcal{A}\bar{x}\rangle = \langle \bar{y},\, b\rangle \geq 0\,,$$

which is contradictory to the condition (2.73). Thus we complete the proof. □

Remark 2.21. *Consider the following problem*

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & \mathcal{A}^p x = b^p\,, \\ & \mathcal{A}^q x \in b^q + \mathcal{Q}\,, \\ & x \in \mathcal{K}\,, \end{array} \tag{2.76}$$

where $b^p \in \Re^p$, $b^q \in \Re^q$ and $\mathcal{Q}$, $\mathcal{K}$ are two closed convex cones with nonempty interior. By adding a variable, we can rewrite it as

$$\begin{array}{ll} \min & \hat{f}(x_s) \\ \text{s.t.} & \mathcal{A}x_s = \begin{pmatrix} \mathcal{A}^p x \\ \mathcal{A}^q x - s \end{pmatrix} = \begin{pmatrix} b^p \\ b^q \end{pmatrix} = b\,, \\ & x_s \in \widehat{\mathcal{K}}, \end{array} \tag{2.77}$$

where

$$\hat{f}(x_s) = f(x)\,,\ x_s = \begin{pmatrix} x \\ s \end{pmatrix} \ \textit{and}\ \ \widehat{\mathcal{K}} = \begin{pmatrix} \mathcal{K} \\ \mathcal{Q} \end{pmatrix}\,.$$

Again, we assume that $b \in \mathrm{Im}\mathcal{A}$. Then Proposition 2.20 can be directly applied to problem (2.77).

2.7 $P_0(P)$-matrix and quasi $P_0(P)$-matrix

A matrix $M \in \Re^{n\times n}$ is a called a P_0-matrix (P-matrix) if all of its principal minors are nonnegative (positive). Here, we will introduce some generalizations of P_0-matrix and P-matrix in order to exploit the properties of the generalized Jacobians.

Definition 2. A matrix $M \in \Re^{n\times n}$ is called a quasi P_0-matrix (P-matrix) if there exists an orthogonal matrix $U \in \Re^{n\times n}$ such that UMU^T is a P_0-matrix (P-matrix).

It is obvious that any P_0-matrix (P-matrix) is a quasi P_0-matrix (P-matrix). Any quasi P-matrix is a quasi P_0-matrix and any quasi P-matrix is nonsingular. If A is a quasi P_0-matrix, then for any $\varepsilon > 0$, $B := A + \varepsilon I$ is a quasi P-matrix, where I is the identity matrix. We will see later that the concepts of quasi P_0-matrix and quasi P-matrix are useful in the analysis of nonsingularity of generalized Jacobians.

Next we shall introduce the concept of a block quasi P_0-function. Suppose that the set $\mathcal{K}$ is the Cartesian product of m (with $m \geq 1$) lower dimensional sets:

$$\mathcal{K} := \prod_{j=1}^{m} \mathcal{K}^j,$$

with each $\mathcal{K}^j$ being a nonempty closed convex subset of $\Re^{n_j}$ and $\sum_{j=1}^{m} n_j = n$. Correspondingly, we partition both the variable x and the function F in the following way:

$$x = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^m \end{pmatrix} \quad \text{and} \quad F(x) = \begin{pmatrix} F^1(x) \\ F^2(x) \\ \vdots \\ F^m(x) \end{pmatrix},$$

where for every j, both x^j and $F^j(x)$ belong to $\Re^{n_j}$. Let $L(\mathcal{K})$ denote all the sets in $\Re^n$ which have the same partitioned structure as $\mathcal{K}$, i.e., $\mathcal{D} \in L(\mathcal{K})$ if and only

if $\mathcal{D}$ can be expressed as

$$\mathcal{D} = \prod_{j=1}^{m} \mathcal{D}^j,$$

where $\mathcal{D}^j \in \Re^{n_j}$ for $j = 1, \ldots, m$.

Definition 3. F is called a block quasi P_0-function on $\mathcal{D} \in L(\mathcal{K})$ if for every pair $x, y \in \mathcal{D}$ with $x \neq y$, there exist a block diagonal orthogonal matrix $Q \in \mathcal{O}^m$ which takes the following form

$$Q := \begin{bmatrix} Q^1 & 0 & \ldots & 0 \\ 0 & Q^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & Q^m \end{bmatrix},$$

where for $j = 1, \ldots, m$, $Q^j \in \mathcal{O}^{n_j}$, such that

$$\max_{\substack{1 \leq i \leq m \\ \hat{x}^i \neq \hat{y}^i}} \langle \hat{x}^i - \hat{y}^i, \, \widehat{F}^i_x - \widehat{F}^i_y \rangle \geq 0 \,,$$

where $\hat{x} := Qx$, $\hat{y} := Qy$, $\widehat{F}_x := QF(x)$ and $\widehat{F}_y := QF(y)$.

Definition 4. Let $\mathcal{X}$ be a finite dimensional space. We shall say that $f : \mathcal{X} \to \Re^n$ is weakly univalent if it is continuous and there exists a sequence of univalent (i.e., one-to-one and continuous) functions f_k from $\mathcal{X}$ to $\Re^n$ such that f_k converges to f uniformly on bounded subsets of $\mathcal{X}$.

Note that univalent functions, affine functions, monotone, and more generally P_0-functions on $\Re^n$ are all weakly univalent.

Chapter 3

A Framework of Proximal Subgradient Method

Let $\mathcal{X}$ be a a finite-dimensional real Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. Let $h : \mathcal{X} \to \Re$ be a smooth function (i.e., continuously differentiable), and $g : \mathcal{X} \to \Re \cup \{\pm\infty\}$ and $p : \mathcal{X} \to \Re \cup \{\pm\infty\}$ be two convex functions. A type of nonconvex nonsmooth optimization problem we will consider in this chapter takes the following form:

$$\min_{x \in \mathcal{X}} f(x) := h(x) + g(x) - p(x)\,. \tag{3.1}$$

In next chapter, one will clearly see how this kind of problems arises from the low rank matrix optimization problems we are dealing with in this thesis.

Remark 3.1. *Suppose that $\Omega \subseteq \mathcal{X}$ is a closed convex set. The constraint $x \in \Omega$ in problem (3.1) can be absorbed into the convex function $g(\cdot)$ via an indicator function $I_\Omega(x) : \mathcal{X} \to [-\infty, +\infty]$*

$$I_\Omega(x) := \begin{cases} 0\,, & \text{if } x \in \Omega\,, \\ +\infty\,, & \text{otherwise}\,. \end{cases}$$

Now we introduce a proximal subgradient method for solving problem (3.1).

Algorithm 3.2. (A proximal subgradient method)

Step 0. *Choose* $x^0 \in \mathcal{X}$. *Set* $k := 0$.

Step 1. *Choose* $M^k \succ 0$ *and* $W^k \in \partial_B p(x^k)$.

Step 2. *Solve*

$$\begin{aligned} \min \quad & \hat{f}^k(d) := \langle \nabla h(x^k), d\rangle + \tfrac{1}{2}\langle d, M^k d\rangle + g(x^k + d) - g(x^k) - \langle W^k, d\rangle \\ \text{s.t.} \quad & x^k + d \in \mathcal{X} \end{aligned} \tag{3.2}$$

to get d^k.

Step 3. Armijo Line Search.

Choose $\alpha^k_{init} > 0$. *Let* l_k *be the smallest nonnegative integer* l *satisfying*

$$f(x^k + \alpha^k_{init}\rho^l d^k) \le f(x^k) + \sigma\alpha^k_{init}\rho^l \Delta^k, \tag{3.3}$$

where $0 < \rho < 1$, $0 < \sigma < 1$, *and*

$$\Delta^k := \langle \nabla h(x^k), d^k\rangle + g(x^k + d^k) - g(x^k) - \langle W^k, d^k\rangle. \tag{3.4}$$

Set $\alpha^k := \alpha^k_{init}\rho^{l_k}$ *and* $x^{k+1} := x^k + \alpha^k d^k$.

Step 4. *If* $x^{k+1} = x^k$, *stop; otherwise, set* $k := k + 1$ *and go to* ***Step 1***.

Remark 3.3. *When* $p \equiv 0$, *the proximal subgradient method reduces to the proximal gradient method which was studied in [40, 77], see also [116] and reference therein. Recently, there are intensive studies in accelerated proximal gradient methods for large-scale convex-concave optimization by Nesterov [82], Nemirovski [81] and others. How to extend these accelerated versions to problem (3.1), however, is still unknown and we leave it to further study.*

Note that various line search rules for smooth optimization can be adapted to our nonsmooth setting to choose α^k. In Algorithm 3.2, we adapt the Armijo rule, which is simple and effective. We will show the well-definedness of the Armijo rule in the following two lemmas.

Lemma 3.4. *Let $\{x^k\}$ and $\{d^k\}$ be two sequences generated by Algorithm 3.2. Then for any $\alpha \in (0,1]$ and $k \geq 0$, we have*

$$f(x^k+\alpha d^k) \leq f(x^k)+\alpha\big(\langle \nabla h(x^k), d^k\rangle + g(x^k+d^k) - g(x^k) - \langle W^k, d^k\rangle\big) + o(\alpha) \quad (3.5)$$

and

$$\langle \nabla h(x^k), d^k\rangle + g(x^k+d^k) - g(x^k) - \langle W^k, d^k\rangle \leq -\langle d^k, M^k d^k\rangle . \quad (3.6)$$

Proof. For any $\alpha \in (0,1]$, from the convexity of g and p, we obtain

$$\begin{aligned}
& f(x^k+\alpha d^k) - f(x^k) \\
= \;& h(x^k+\alpha d^k) + g(x^k+\alpha d^k) - p(x^k+\alpha d^k) - \big(h(x^k)+g(x^k)-p(x^k)\big) \\
\leq \;& h(x^k+\alpha d^k) - h(x^k) + \alpha g(x^k+d^k) + (1-\alpha)g(x^k) - g(x^k) \\
& -\big(p(x^k)+\alpha\langle W^k, d^k\rangle - p(x^k)\big) \\
= \;& \alpha\big(\langle \nabla h(x^k), d^k\rangle + g(x^k+d^k) - g(x^k) - \langle W^k, d^k\rangle\big) + o(\alpha),
\end{aligned}$$

which proves (3.5). Moreover, by using the convexity of g and the fact that $d^k \in \arg\min_{(x^k+d)\in\mathcal{X}} \hat{f}^k(d)$, we know that for any $\alpha \in (0,1)$

$$\begin{aligned}
& \langle \nabla h(x^k), d^k\rangle + \tfrac{1}{2}\langle d^k, M^k d^k\rangle + g(x^k+d^k) - g(x^k) - \langle W^k, d^k\rangle \\
\leq \;& \langle \nabla h(x^k), \alpha d^k\rangle + \tfrac{1}{2}\langle \alpha d^k, M^k(\alpha d^k)\rangle + g(x^k+\alpha d^k) - g(x^k) - \langle W^k, \alpha d^k\rangle \\
\leq \;& \alpha\langle \nabla h(x^k), d^k\rangle + \tfrac{\alpha^2}{2}\langle d^k, M^k d^k\rangle + \alpha g(x^k+d^k) + (1-\alpha)g(x^k) - g(x^k) - \alpha\langle W^k, d^k\rangle \\
= \;& \alpha\langle \nabla h(x^k), d^k\rangle + \tfrac{\alpha^2}{2}\langle d^k, M^k d^k\rangle + \alpha(g(x^k+d^k) - g(x^k)) - \alpha\langle W^k, d^k\rangle .
\end{aligned}$$

Rearranging the terms yields

$$\langle \nabla h(x^k), d^k\rangle + g(x^k+d^k) - g(x^k) - \langle W^k, d^k\rangle \leq -\frac{1+\alpha}{2}\langle d^k, M^k d^k\rangle ,$$

then taking $\alpha \uparrow 1$ proves (3.6). □

Lemma 3.5. *Let $\{x^k\}$ and $\{d^k\}$ be two sequences generated by Algorithm 3.2. Assume that for all $k \geq 0$, $0 < \underline{\nu}\|d\|^2 \leq \langle d, M^k d\rangle$ for any $d \in \mathcal{X}$. If h satisfies*

$$\|\nabla h(y) - \nabla h(z)\| \leq L\|y - z\|, \quad \forall\, y, z \in \mathcal{X}\,, \tag{3.7}$$

for some $L \geq 0$, then, for each integer $k \geq 0$, the descent condition

$$f(x^k + \alpha d^k) \leq f(x^k) + \sigma\alpha\Delta^k \tag{3.8}$$

is satisfied for any $\sigma \in (0,1)$ whenever $0 \leq \alpha \leq \min\{1, 2\underline{\nu}(1-\sigma)/L\}$.

Proof. Without any ambiguity, we drop the superscript k for simplicity. For any $\alpha \in (0,1]$, we obtain

$$\begin{aligned}
& f(x+\alpha d) - f(x) \\
= \;& h(x+\alpha d) - h(x) + g(x+\alpha d) - g(x) - \big(p(x+\alpha d) - p(x)\big) \\
= \;& \alpha\langle \nabla h(x), d\rangle + g(x+\alpha d) - g(x) - \big(p(x+\alpha d) - p(x)\big) \\
& \quad + \textstyle\int_0^1 \langle \nabla h(x+t\alpha d) - \nabla h(x), \alpha d\rangle dt \\
\leq \;& \alpha\big(\langle \nabla h(x), d\rangle + g(x+d) - g(x) - \langle W, d\rangle\big) + \alpha \textstyle\int_0^1 \|\nabla h(x+t\alpha d) - \nabla h(x)\| \cdot \|d\| dt \\
\leq \;& \alpha\big(\langle \nabla h(x), d\rangle + g(x+d) - g(x) - \langle W, d\rangle\big) + \tfrac{L}{2}\alpha^2\|d\|^2\,,
\end{aligned}$$

where $W \in \partial p(x)$. If $\alpha \leq 2\underline{\nu}(1-\sigma)/L$, then

$$\frac{L}{2}\alpha\|d\|^2 \leq (1-\sigma)\langle d, Md\rangle \leq -(1-\sigma)\big(\langle \nabla h(x), d\rangle + g(x+d) - g(x) - \langle W, d\rangle\big)\,.$$

Therefore, when $0 \leq \alpha \leq \min\{1, 2\underline{\nu}(1-\sigma)/L\}$, the inequality (3.8) holds for any $\sigma \in (0,1)$. □

Definition 5. A point $x \in \mathcal{X}$ is said to be a *stationary* point of problem (3.1) if

$$\partial(h(x) + g(x)) \cap (\partial p(x)) = (\nabla h(x) + \partial g(x)) \cap (\partial p(x)) \neq \emptyset \tag{3.9}$$

and a *B-stationary* point of problem (3.1) if

$$\partial(h(x) + g(x)) \cap (\partial_B p(x)) = (\nabla h(x) + \partial g(x)) \cap (\partial_B p(x)) \neq \emptyset\,. \tag{3.10}$$

Assumption 3.6. *For all $k \geq 0$, $\underline{\nu}\|d\|^2 \leq \langle d, M^k d\rangle \leq \bar{\nu}\|d\|^2$ for any $d \in \mathcal{X}$, where $0 < \underline{\nu} \leq \bar{\nu} < +\infty$.*

Theorem 3.7. *Let $\{x^k\}$ and $\{d^k\}$ be two sequences generated by Algorithm 3.2 under Assumption 3.6. Then $\{f(x^k)\}$ is a monotonically decreasing sequence. If $x^{k+1} = x^k$ for some integer $k \geq 0$, then x^k is a B-stationary point of problem (3.1). Otherwise, suppose that $\inf \alpha_{init}^k > 0$, the following results hold:*

(a). For each integer $k \geq 0$, Δ^k satisfies

$$\Delta^k \leq -\langle d^k, M^k d^k\rangle \leq -\underline{\nu}\|d^k\|^2\,,$$

$$f(x^{k+1}) - f(x^k) \leq \sigma\alpha^k\Delta^k \leq 0\,.$$

(b). If $\{x^{k_j}\}$ is a converging subsequence of $\{x^k\}$, then $\lim_{j\to+\infty} d^{k_j} = 0$.

(c). Any accumulation point of $\{x^k\}$ is a B-stationary point of problem (3.1).

Proof. The monotone decreasing property of $\{f(x^k)\}$ follows easily from the line search condition (3.3) in Algorithm 1.

We first consider the case that $x^{k+1} = x^k$ for some integer $k \geq 0$. It is clear that $d^k = 0$ is the optimal solution to problem (3.2). Then one has

$$0 \in \nabla h(x^k) + \partial g(x^k) - W^k\,,$$

which implies that x^k is a B-stationary point of problem (3.1) from the definition (3.10).

Next we assume that $x^{k+1} \neq x^k$ for all $k \geq 0$. Then an infinite sequence $\{x^k\}$ is generated. Suppose that $\{x^{k_j}\}$ is a converging subsequence of $\{x^k\}$. Let $\bar{x} := \lim_{j\to+\infty} x^{k_j}$. Since h is continuous and g, p are continuous in the relative interiors of $\mathrm{dom}(g)$ and $\mathrm{dom}(p)$, $\lim_{j\to+\infty} f(x^{k_j}) = f(\bar{x})$. Note that $\{f(x^k)\}$ is a decreasing sequence, this implies that $\lim_{k\to+\infty} f(x^k) = f(\bar{x})$. Hence

$$\lim_{k\to+\infty} \alpha^k\Delta^k = 0 \tag{3.11}$$

follows directly from the fact that $f(x^k) - f(x^{k+1}) \to 0$ as $k \to +\infty$.

Now we prove $\lim\limits_{j\to+\infty} d^{k_j} = 0$. By contradiction, suppose that $d^{k_j} \nrightarrow 0$ when $j \to +\infty$. By passing to a subsequence if necessary, we can assume that, for some $\delta > 0$, $\|d^{k_j}\| \geq \delta$ for all $j \geq 0$. Thus, $\Delta^k \leq -\underline{\nu}\|d^k\|^2 \leq -\underline{\nu}\delta$. By noting (3.11), one has $\lim\limits_{j\to+\infty} \alpha^{k_j} = 0$. Recall that $\alpha^{k_j} = \alpha_{init}^{k_j}\rho^{l_{k_j}}$ and $\inf \alpha_{init}^k > 0$. Then there exists some index $\bar{k} > 0$ such that $\alpha^{k_j} < \alpha_{init}^{k_j}$ and $\alpha^{k_j} \leq \rho$ for all $k_j > \bar{k}$. Furthermore, α^k is chosen by the Armijo rule , it implies that

$$f\big(x^{k_j} + (\alpha^{k_j}/\rho)d^{k_j}\big) - f(x^{k_j}) > \sigma(\alpha^{k_j}/\rho)\Delta^{k_j}\,, \forall\ k_j \geq \bar{k}\,.$$

Thus,

$$\begin{aligned}
\sigma\Delta^{k_j} &= \sigma\big(\langle\nabla h(x^{k_j}), d^{k_j}\rangle + g(x^{k_j} + d^{k_j}) - g(x^{k_j}) - \langle W^{k_j}, d^{k_j}\rangle\big) \\
&< \frac{h(x^{k_j}+(\alpha^{k_j}/\rho)d^{k_j}) - h(x^{k_j}) + g(x^{k_j}+(\alpha^{k_j}/\rho)d^{k_j}) - g(x^{k_j}) - \big(p(x^{k_j}+(\alpha^{k_j}/\rho)d^{k_j}) - p(x^{k_j})\big)}{\alpha^{k_j}/\rho} \\
&\leq \frac{h(x^{k_j}+(\alpha^{k_j}/\rho)d^{k_j}) - h(x^{k_j})}{\alpha^{k_j}/\rho} + g(x^{k_j} + d^{k_j}) - g(x^{k_j}) - \langle W^{k_j}, d^{k_j}\rangle
\end{aligned}$$

It follows that

$$\frac{h(x^{k_j} + (\alpha^{k_j}/\rho)d^{k_j}) - h(x^{k_j})}{\alpha^{k_j}/\rho} - \langle\nabla h(x^{k_j}), d^{k_j}\rangle \geq -(1-\delta)\Delta^{k_j} \geq (1-\delta)\underline{\nu}\|d^k\|^2\,,$$

$$\text{and}\quad \frac{h\big(x^{k_j} + \hat{\alpha}^{k_j}\frac{d^{k_j}}{\|d^{k_j}\|}\big) - h(x^{k_j})}{\hat{\alpha}^{k_j}} - \langle\nabla h(x^{k_j}), \frac{d^{k_j}}{\|d^{k_j}\|}\rangle \geq (1-\delta)\underline{\nu}\delta\,,$$

where $\hat{\alpha}^{k_j} := \frac{\alpha^{k_j}\|d^{k_j}\|}{\rho}$. Then $-\alpha^{k_j}\Delta^k \geq \underline{\nu}\alpha^{k_j}\|d^{k_j}\|^2 \geq \delta\underline{\nu}\alpha^{k_j}\|d^{k_j}\| \geq 0$, thus $\alpha^{k_j}\|d^{k_j}\| \to 0$ as $j \to +\infty$, which implies that $\hat{\alpha}^{k_j} \to 0$.

Note that there exists $\bar{d}$, by further passing to a subsequence if necessary, such that $\lim\limits_{j\to+\infty} \frac{d^{k_j}}{\|d^{k_j}\|} = \bar{d}$. It then follows that

$$0 = \langle\nabla h(\bar{x}), \bar{d}\rangle - \langle\nabla h(\bar{x}), \bar{d}\rangle \geq (1-\delta)\underline{\nu}\delta > 0\,,$$

and thus this contradiction shows that $d^{k_j} \to 0$ as $j \to +\infty$.

To complete the proof, we still need to show that $\bar{x}$ is a B-stationary point of problem (3.1). Noting that $d^{k_j} = \arg\min_{d\in\mathcal{X}} \hat{f}^{k_j}(d)$, there exists $V^{k_j} \in \partial g(x^{k_j} + d^{k_j})$ such that

$$\nabla h(x^{k_j}) + M^{k_j} d^{k_j} + V^{k_j} - W^{k_j} = 0\,.$$

Since both $\{x^{k_j}\}$ and $\{x^{k_j} + d^{k_j}\}$ are bounded, from convex analysis [105, Chap 24, Theorem 24.7], we know that $\{V^{k_j}\}$ and $\{W^{k_j}\}$ are also bounded. By taking subsequences respectively, if necessary, we assume that there exist $\overline{V} \in \partial g(\bar{x})$ and $\overline{W} \in \partial_B p(\bar{x})$ such that $\lim_{j\to+\infty} V^{k_j} = \overline{V}$ and $\lim_{j\to+\infty} W^{k_j} = \overline{W}$, respectively. Hence,

$$\nabla h(\bar{x}) + \overline{V} - \overline{W} = 0\,,$$

which implies that $\bar{x}$ is a B-stationary point of problem (3.1), i.e.,

$$(\nabla h(\bar{x}) + \partial g(\bar{x})) \cap \partial_B p(\bar{x}) \neq \emptyset\,.$$

□

By considering some special choices of M^k in Algorithm 3.2, we have the following lemma concerning the stepsize satisfying the Armijo descent condition (3.3).

Lemma 3.8. *Let $\{x^k\}$ and $\{d^k\}$ be two sequences generated by Algorithm 3.2. If for each integer $k \geq 0$, one can choose $M^k \succ 0$ such that*

$$h(y) \leq h(x^k) + \langle \nabla h(x^k), y - x^k\rangle + \frac{1}{2}\langle y - x^k,\, M^k(y - x^k)\rangle\,, \quad \forall\ y \in \mathcal{X}\,, \tag{3.12}$$

then the descent condition

$$f(x^k + \alpha d^k) \leq f(x^k) + \sigma\alpha\Delta^k \tag{3.13}$$

is satisfied for any $\sigma \in (0,1)$ whenever $0 \leq \alpha \leq \min\{1, 2(1-\sigma)\}$.

Proof. Without any ambiguity, we drop the superscript k for simplicity. For any $\alpha \in (0, 1]$, we obtain

$$\begin{aligned} & f(x+\alpha d) - f(x) \\ = \; & h(x+\alpha d) - h(x) + g(x+\alpha d) - g(x) - \big(p(x+\alpha d) - p(x)\big) \\ \leq \; & \alpha\langle \nabla h(x), d\rangle + \tfrac{\alpha^2}{2}\langle d, Md\rangle + g(x+\alpha d) - g(x) - \big(p(x+\alpha d) - p(x)\big) \\ \leq \; & \alpha\big(\langle \nabla h(x), d\rangle + g(x+d) - g(x) - \langle W, d\rangle\big) + \tfrac{\alpha^2}{2}\langle d, Md\rangle\,, \end{aligned}$$

where $W \in \partial p(x)$. If $\alpha \leq 2(1-\sigma)$, then

$$\frac{\alpha}{2}\langle d, Md\rangle \leq (1-\sigma)\langle d, Md\rangle \leq -(1-\sigma)\big(\langle \nabla h(x), d\rangle + g(x+d) - g(x) - \langle W, d\rangle\big)\,.$$

Therefore, when $0 \leq \alpha \leq \min\{1, 2(1-\sigma)\}$, the inequality (3.13) holds for any $\sigma \in (0, 1)$. □

One important implication of Lemma 3.8 is that if $0 < \sigma \leq \frac{1}{2}$, then for any $k \geq 0$, we can take $\alpha^k \equiv 1$, i.e., the unit stepsize is attainable. Using this observation, we have the following proximal subgradient algorithm with no line search.

Algorithm 3.9. (A proximal subgradient method with no line search)

Step 0. *Choose* $x^0 \in \mathcal{X}$. *Set* $k := 0$.

Step 1. *Choose* $M^k \succ 0$ *such that for any* $x \in \mathcal{X}$

$$h(x) \le \hat{h}^k(x) := h(x^k) + \langle \nabla h(x^k),\, x - x^k \rangle + \frac{1}{2}\langle x - x^k,\, M^k(x - x^k) \rangle \,. \quad (3.14)$$

Choose $W^k \in \partial_B p(x^k)$ *and define* $\hat{p}^k(x) : \mathcal{X} \to \Re$ *by*

$$\hat{p}^k(x) := p(x^k) + \langle W^k, x - x^k \rangle \,. \quad (3.15)$$

Step 2. *Solve*

$$\min_{x \in \mathcal{X}} \hat{f}^k(x) := \hat{h}^k(x) + g(x) - \hat{p}^k(x)$$

to get x^{k+1}.

Step 3. *If* $x^{k+1} = x^k$, *stop; otherwise, set* $k := k + 1$ *and go to* ***Step 1***.

From Theorem 3.7 and Lemma 3.8, we can easily derive the following corollary.

Corollary 3.10. *Let* $\{x^k\}$ *be the sequences generated by Algorithm 3.9. Then* $\{f(x^k)\}$ *is a monotonically decreasing sequence. If* $x^{k+1} = x^k$ *for some integer* $k \ge 0$, *then* x^k *is a B-stationary point of problem (3.1). Otherwise, the infinite sequence* $\{f(x^k)\}$ *satisfies*

$$\frac{1}{2}\langle x^{k+1} - x^k,\, M^k(x^{k+1} - x^k) \rangle \le f(x^k) - f(x^{k+1}) \,, \quad k = 0, 1, \ldots$$

Moreover, any accumulation point of $\{x^k\}$ *is a B-stationary point of problem (3.1) provided that Assumption 3.6 holds.*

Remark 3.11. *If* ∇h *satisfies the condition (3.7), i.e.,* ∇h *is Lipschitz continuous with the Lipschitz constant* L, *we can simply choose* $M^k \succ 0$ *for all* $k \ge 0$ *such that* $\langle d, M^k d \rangle \le L\|d\|^2$ *for any* $d \in \mathcal{X}$.

Remark 3.12. *Let $\Omega \subset \mathcal{X}$ be a closed (bounded) set, which is not necessarily convex. For a given continuous function $f : \mathcal{X} \to \Re$, the principle of a majorization method for minimizing $f(x)$ over Ω is to start with an initial point $x^0 \in \Omega$ and for each $k \geq 0$, to minimize $\hat{f}^k(x)$ over Ω to get x^{k+1}, where $\hat{f}^k(\cdot)$ is a majorization function of f at x^k, i.e., $\hat{f}^k(\cdot)$ satisfies*

$$\hat{f}^k(x^k) = f(x^k) \quad \text{and} \quad \hat{f}^k(x) \geq f(x), \quad \forall\, x \in \Omega\,.$$

The monotone decreasing property of the generated sequence $\{f(x^k)\}$ comes from the so-called sandwich inequality [25] for the majorization method, i.e.,

$$f(x^{k+1}) \leq \hat{f}^k(x^{k+1}) \leq \hat{f}^k(x^k) = f(x^k), \quad k = 0, 1, \ldots. \tag{3.16}$$

The efficiency of the above majorization method hinges on two key issues: i) the majorization functions should be simpler than the original function f so that the resulting minimization problems are easier to solve, and ii) they should not deviate too much from f in order to get fast convergence. These two often conflicting issues need to be addressed on a case by case basis to achieve best possible overall performance.

The idea of using a majorization function in optimization appeared as early as in Ortega and Rheinboldt [84, Section 8.3] for the purpose of doing line searches to decide a step length along a descent direction. This technique was quickly replaced by more effective inexact line search models such as the back tracking. The very first majorization method was introduced by de Leeuw[23, 24] and de Leeuw and Heiser [28] to solve multidimensional scaling problems. Since then much progress has been made on using majorization methods to solve various optimization problems [26, 27, 49, 50, 57, 58], to name only a few.

In Algorithm 3.9, one may notice that $\hat{h}^k(\cdot)$ and $\hat{p}^k(\cdot)$ defined in (3.14) and (3.15) are actually a special kind of the majorization functions of $h(\cdot)$ and $p(\cdot)$ at x^k, respectively. In this sense, the proximal subgradient method with no line search can

be treated as a majorization method, thus, it can also handle the nonconvex constraints like the majorization method. However, the proximal subgradient method is designed for the cases where Ω is convex and the majorization functions are not easy to compute.

Chapter 4

A Penalty Approach

Let $C \in \Re^{n_1 \times n_2}$ be a given matrix and $H \in \Re^{n_1 \times n_2}$ a given weight matrix whose entries are nonnegative. Let $\widehat{\mathcal{Q}} \in \Re^q$ be a closed convex cone with nonempty interior and define $\mathcal{Q} := \{0\}^p \times \widehat{\mathcal{Q}}$. Denote $\mathcal{Q}^*$ as the dual cone of $\mathcal{Q}$ under the natural inner product of $\Re^p \times \Re^q$. Let $m := p + q$ and $\rho \geq 0$ be a given number. Then we consider the following structured low rank matrix, not necessarily symmetric, approximation problem

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 + \rho\|X\|_* \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & \text{rank}(X) \leq r\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.1}$$

where "$\circ$" denotes the Hadamard product, i.e., $(A \circ B)_{ij} = A_{ij}B_{ij}$ for all i, j, $\mathcal{A} : \Re^{n_1 \times n_2} \to \Re^m$ is a linear operator and $r \in \{1, \ldots, n_1\}$ is a given integer.

4.1 A penalty approach for the rank constraint

In this subsection, we shall introduce a penalty technique to deal with the non-convex rank constraint in (4.1). Given the fact that for any $X \in \Re^{n_1 \times n_2}$, rank$(X) \leq$

r if and only if $\sigma_{r+1}(X) + \ldots + \sigma_{n_1}(X) = 0$, we can equivalently rewrite (4.1) as follows

$$\begin{array}{rll} \bar{\theta} := \min & \theta(X) = h(X) + \rho\|X\|_* & \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, & \\ & \sigma_{r+1}(X) + \ldots + \sigma_{n_1}(X) = 0\,, & \\ & X \in \Re^{n_1 \times n_2}\,, & \end{array} \tag{4.2}$$

where $h(X) := \dfrac{1}{2}\|H \circ (X - C)\|^2$. Now we consider the following penalized problem by taking a trade-off between the rank constraint and the weighted least squares distance:

$$\begin{array}{ll} \min & \theta(X) + c(\sigma_{r+1}(X) + \ldots + \sigma_{n_1}(X)) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.3}$$

where $c > 0$ is a given penalty parameter that decides the allocated weight to the rank constraint in the objective function. By noting that for any $X \in \Re^{n_1 \times n_2}$,

$$\sum_{i=r+1}^{n_1} \sigma_i(X) = \sum_{i=1}^{n_1} \sigma_i(X) - \sum_{i=1}^{r} \sigma_i(X) = \|X\|_* - \sum_{i=1}^{r} \sigma_i(X)\,, \tag{4.4}$$

we can equivalently write problem (4.3) as

$$\begin{array}{ll} \min & f_c(X) := \theta(X) - cp(X) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.5}$$

where for any $X \in \Re^{n_1 \times n_2}$,

$$p(X) := \sum_{i=1}^{r} \sigma_i(X) - \|X\|_* \leq 0\,, \tag{4.6}$$

which is the difference of two convex functions. Note that the penalized problem (4.5) is not equivalent to the original problem (4.1). Then the question is how much we can say about the solutions to (4.1) by solving the penalized problem (4.5). We will address this question in the following two propositions.

Since the objective function $f_c(\cdot)$ of problem (4.5) is coercive in $\Re^{n_1\times n_2}$, we know that the problem (4.5) exists at least one global solution, say X_c^*.

Proposition 4.1. *If the rank of X_c^* is not larger than r, then X_c^* is a global optimal solution to problem (4.1).*

Proof. Assume that the rank of X_c^* is not larger than r. Then X_c^* is a feasible solution to (4.1) and $p(X_c^*) = 0$. Let $X_r \in \Re^{n_1\times n_2}$ be any feasible point to (4.1). Thus, by noting that $p(X_r) = 0$, we have

$$\theta(X_c^*) = \theta(X_c^*) - cp(X_c^*) \leq \theta(X_r) - cp(X_r) = \theta(X_r)\,.$$

This shows that the conclusion of this proposition holds. □

Proposition 4.1 says in the ideal situation when the rank of X_c^* is not larger than r, X_c^* actually solves the original problem (4.1). Though this ideal situation is always observed in our numerical experiments for a properly chosen penalty parameter $c > 0$, there is no theoretical guarantee that this is the case. However, when the penalty parameter c is large enough, $|p(X_c^*)|$ can be proven to be very small. To see this, let X^* be an optimal solution to the following least squares convex optimization problem

$$\begin{array}{ll} \min & \theta(X) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1\times n_2}\,. \end{array} \tag{4.7}$$

Proposition 4.2. *Let $\varepsilon > 0$ be a given positive number and $X_r \in \Re^{n_1\times n_2}$ a feasible solution to problem (4.1). Assume that $c > 0$ is chosen such that $\big(\theta(X_r) - \theta(X^*)\big)/c \leq \varepsilon$. Then we have*

$$|\,p(X_c^*)\,| \leq \varepsilon \quad \text{and} \quad \theta(X_c^*) \leq \bar{\theta} - c|p(X_c^*)| \leq \bar{\theta}. \tag{4.8}$$

Proof. By noting that X_r is feasible to the penalized problem (4.5) and $p(X_r) = 0$, we have

$$\theta(X_r) = \theta(X_r) - cp(X_r) = f_c(X_r) \geq f_c(X_c^*) = \theta(X_c^*) - cp(X_c^*) \geq \theta(X^*) - cp(X_c^*)\,,$$

which implies

$$|\, p(X_c^*)\,| = -p(X_c^*) \leq \big(\theta(X_r) - \theta(X^*)\big)/c \leq \varepsilon\,.$$

Let $\overline{X}$ be a global optimal solution to problem (4.1). Then from

$$\theta(\overline{X}) - cp(\overline{X}) = f_c(\overline{X}) \geq f_c(X_c^*) = \theta(X_c^*) - cp(X_c^*)$$

and the fact that $p(\overline{X}) = 0$, we obtain that $\theta(X_c^*) \leq \theta(\overline{X}) - c|p(X_c^*)| = \bar{\theta} - c|p(X_c^*)|$. The proof is completed. □

Proposition 4.2 says that an ε-optimal solution to the original problem (4.1) in the sense of (4.8) is guaranteed by solving the penalized problem (4.5) as long as the penalty parameter c is above some ε-dependent number. This provides the rationale to replace the rank constraint in problem (4.1) by the penalty function $-cp(\cdot)$ in problem (4.5).

Remark 4.3. *In Proposition 4.2, we need to choose a feasible point X_r to problem (4.1). That is equivalently to say that we need to find a global solution to*

$$\begin{array}{ll} \min & \sigma_{r+1}(X) + \ldots + \sigma_{n_1}(X) = -p(X) \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1 \times n_2}\,. \end{array} \tag{4.9}$$

To solve problem (4.9), one may use the majorization method to be introduced in next subsection. This corresponds to the case that $H = 0$. However, this is not needed in many situations when a feasible point to problem (4.1) is readily available. For example, the truncated singular value decomposition (TSVD) of X^* is such a choice if there are no constraints.

Remark 4.4. *There are different choices to penalize the rank constraint. For example, one may use* $\sigma_{r+1}(X)$ *or* $\sum_{i=r+1}^{n_1} \sigma_i^2(X)$ *instead. However, the performance needs to be tested further.*

Recall that in the symmetric counterpart of problem (4.1), we consider the following problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} \quad & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \operatorname{rank}(X) \le r\,, \end{aligned} \tag{4.10}$$

where $C \in \mathcal{S}^n$ is given and $H \in \mathcal{S}^n$ is a given weight matrix whose entries are nonnegative.

Given the fact that for any $X \in \mathcal{S}^n_+$, $\operatorname{rank}(X) \le r$ if and only if $\lambda_{r+1}(X) + \ldots + \lambda_n(X) = 0$, and that

$$\sum_{i=r+1}^{n} \lambda_i(X) = \sum_{i=1}^{n} \lambda_i(X) - \sum_{i=1}^{r} \lambda_i(X) = \langle I, X\rangle - \sum_{i=1}^{r} \lambda_i(X)\,,$$

the penalized problem for (4.10) takes the following form

$$\begin{aligned} \min \quad & f_c(X) := \theta(X) - cp(X) \\ \text{s.t.} \quad & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \succeq 0\,, \end{aligned} \tag{4.11}$$

where $\theta(X) := \frac{1}{2}\|H \circ (X - C)\|^2$ and for any $X \in \mathcal{S}^n$,

$$p(X) := \sum_{i=1}^{r} \lambda_i(X) - \langle I, X\rangle\,, \tag{4.12}$$

which is a convex function and simpler than (4.6). Note that problem (4.11) is similar to the penalized problem (4.5) in the nonsymmetric setting, thus Proposition 4.1 and 4.2 still hold for the symmetric counterpart problem (4.11).

4.2 The proximal subgradient method for the penalized problem

In this section, we shall study the penalized problem (4.5), which can be rewritten as follows

$$\begin{aligned} \min \quad & f_c(X) = h(X) + g(X) - cp(X) \\ \text{s.t.} \quad & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{aligned} \tag{4.13}$$

where $h(X) = \frac{1}{2}\|H \circ (X - C)\|^2$, $g(X) := (\rho + c)\|X\|_*$ and $p(X) := \sum_{i=1}^{r} \sigma_i(X)$. Let Ω denote the feasible set of problem (4.13), i.e.,

$$\Omega := \{X \in \Re^{n_1 \times n_2} \,|\, \mathcal{A}X \in b + \mathcal{Q}\}\,.$$

For any $X \in \Omega$, denote the normal cone of Ω at the point X by

$$\mathcal{N}_\Omega(X) := \{Z \in \Re^{n_1 \times n_2} \,|\, \langle Z, Y - X\rangle \leq 0 \;\forall\, Y \in \Omega\}.$$

A point $X \in \Omega$ is said to be a *stationary* point of problem (4.13) if

$$(\nabla h(X) + \partial g(X) + N_\Omega(X)) \cap (c\partial p(X)) \neq \emptyset\,,$$

and a *B-stationary* point of problem (4.13) if

$$(\nabla h(X) + \partial g(X) + N_\Omega(X)) \cap (c\partial_B p(X)) \neq \emptyset\,.$$

A B-stationary point of problem (4.13) is always a stationary point of the problem itself and the converse is not necessarily true.

From Remark 3.1, we know that the penalized problem (4.13) can directly be solved by the proximal subgradient method introduced in Chapter 3 via the following problem

$$\min_{X \in \Re^{n_1 \times n_2}} \hat{f}_c(X) := h(X) + \hat{g}(X) - cp(X)\,, \tag{4.14}$$

where $\hat{g}(X) := g(X) + I_\Omega(X)$ is still a convex function as Ω is a closed convex set. By noting that h is a twice differentiable quadratic function and for any $Y \in \Re^{n_1 \times n_2}$

$$h(X) \;=\; h(Y) + \langle \nabla h(Y),\, X - Y \rangle + \frac{1}{2}\|H \circ (X - Y)\|^2 \,,$$

when applying Algorithm 3.2 to problem (4.14), for each integer $k \geq 0$, we only need to choose a componentwise nonnegative matrix $\widehat{H}^k \geq 0$ in $\Re^{n_1 \times n_2}$ such that $\langle Y, M^k Y \rangle = \|\widehat{H}^k \circ Y\|^2$ for any $Y \in \Re^{n_1 \times n_2}$. Then the following corollary comes directly from Theorem 3.7.

Assumption 4.5. *For all $k \geq 0$,*

$$\kappa_1 \leq \min_{\substack{i=1,\ldots,n_1 \\ j=1,\ldots,n_2}} \widehat{H}^k_{ij} \leq \max_{\substack{i=1,\ldots,n_1 \\ j=1,\ldots,n_2}} \widehat{H}^k_{ij} \leq \kappa_2 \,,$$

where $0 < \kappa_1 \leq \kappa_2 < +\infty$.

Corollary 4.6. *Let $\{X^k\}$, $\{d^k\}$ be two sequences generated by Algorithm 3.2 under Assumption 4.5. Then $\{\hat{f}_c(X^k)\}$ is a monotonically decreasing sequence. If $X^{k+1} = X^k$ for some integer $k \geq 0$, then X^k is a B-stationary point of problem (4.13). Otherwise, suppose that $\inf \alpha^k_{init} > 0$, the following results hold:*

(a). For each integer $k \geq 0$, Δ^k satisfies

$$\Delta^k \leq -\|\widehat{H}^k \circ d^k\|^2 \leq -\underline{\nu}\|d^k\|^2 \,,$$

$$\hat{f}_c(X^{k+1}) - \hat{f}_c(X^k) \leq \sigma \alpha^k \Delta^k \leq 0 \,.$$

(b). If $\{X^{k_j}\}$ is a converging subsequence of $\{X^k\}$, then $\lim_{j \to +\infty} d^{k_j} = 0$.

(c). Any accumulation point of $\{X^k\}$ is a B-stationary point of problem (4.13).

Furthermore, one may notice that at each iteration k, it is not difficult to find $\Re^{n_1 \times n_2} \ni \widehat{H}^k \geq 0$ satisfying

$$\|H \circ (X - X^k)\|^2 \leq \|\widehat{H}^k \circ (X - X^k)\|^2, \quad \forall\, X \in \mathcal{X} \,,$$

thus, for any $X \in \Re^{n_1 \times n_2}$,

$$h(X) \leq \hat{h}^k(X) := h(X^k) + \langle \nabla h(X^k),\, X - X^k \rangle + \frac{1}{2}\|\widehat{H}^k \circ (X - X^k)\|^2 .$$

This implies that one may also apply the proximal subgradient method without line search, i.e, Algorithm 3.9, to problem (4.13).

Corollary 4.7. *Let $\{X^k\}$ be the sequence generated by Algorithm 3.9. Then $\{\hat{f}_c(X^k)\}$ is a monotonically decreasing sequence. If $X^{k+1} = X^k$ for some integer $k \geq 0$, then X^{k+1} is a B-stationary point of problem (4.13). Otherwise, the infinite sequence $\{\hat{f}_c(X^k)\}$ satisfies*

$$\frac{1}{2}\|\widehat{H}^k \circ (X^{k+1} - X^k)\|^2 \leq \hat{f}_c(X^k) - \hat{f}_c(X^{k+1}) , \quad k = 0, 1, \ldots \tag{4.15}$$

Moreover, any accumulation point of the bounded sequence $\{X^k\}$ is a B-stationary point of problem (4.13) provided that Assumption 4.5 holds.

Similarly, in the symmetric case, for the penalized problem (4.11), we can also define the *stationary* (*B-stationary*) point. Let Ω denote the feasible set of problem (4.11), i.e., $\Omega = \{X \in \mathcal{S}^n \,|\, \mathcal{A}X \in b + \mathcal{Q}\}$. A point $X \in \Omega$ is said to be a *stationary* point of problem (4.11) if

$$(\nabla\theta(X) + N_\Omega(X)) \cap (c\partial p(X)) \neq \emptyset ,$$

and a *B-stationary* point of problem (4.11) if

$$(\nabla\theta(X) + N_\Omega(X)) \cap (c\partial_B p(X)) \neq \emptyset .$$

Hence, one can easily show that both Corollary 4.7 and 4.6 still hold for the penalized problem (4.11).

4.2.1 Implementation issues

In this subsection, we discuss several implementation issues when applying the proximal subgradient method to penalized problem (4.13) and (4.11). Due to

the nice properties of the first term $h(\cdot)$ (smooth and quadratic) in the objective function of problem (4.13) in the nonsymmetric setting, we simply apply Algorithm 3.9 to problem (4.13) and eventually need to solve a sequence of problems taking the form of

$$\begin{array}{ll} \min & \hat{f}_c^k(X) = \dfrac{1}{2}\|\widehat{H}^k \circ (X - X^k)\|^2 + \langle X, H \circ H \circ (X^k - C) - cW^k \rangle + g(X) + q_c^k \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.16}$$

where $q_c^k := h(X^k) - \langle \nabla h(X^k),\, X^k \rangle - cp_\sigma(X^k) + c\langle W^k,\, X^k \rangle$. Here we assume that $0 \le H_{ij} \le 1$ for $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$ (see Remark 4.8 if it fails to hold). It then follows that for all $k \ge 0$, $\widehat{H}^k$ can simply be chosen as E whose entries are all ones. Then the objective function $\hat{f}_c^k(\cdot)$ in (4.16) can be equivalently written as

$$\begin{array}{rcl} \hat{f}_c^k(X) & = & \frac{1}{2}\|X - X^k\|^2 + \left\langle X, H \circ H \circ (X^k - C) - cW^k \right\rangle + g(X) + q_c^k \\ & = & \frac{1}{2}\|X - (X^k + C^k)\|^2 + g(X) + f_c(X^k) - \frac{1}{2}\|C^k\|^2\,, \end{array}$$

where $C^k := cW^k - H \circ H \circ (X^k - C)$. By dropping the constant terms in $\hat{f}_c^k(X)$ and noting that $g(X) = (\rho + c)\|X\|_*$, we can equivalently write problem (4.16) as the following well-studied least squares nuclear norm problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - (X^k + C^k)\|^2 + (\rho + c)\|X\|_* \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.17}$$

which can be efficiently solved by the well developed smoothing Newton-BiCGStab method.

Remark 4.8. *If not all the components of the given weight matrix H are in $[0, 1]$, one can do the preprocessing as follows. Define two vectors $d_1 \in \Re^{n_1}$ and $d_2 \in \Re^{n_2}$ by*

$$(d_1)_i = \max\{\delta,\, \max\{\, H_{ij} \,|\, j = 1, \ldots, n_2\}\}\,, \quad i = 1, \ldots, n_1\,,$$

and

$$(d_2)_j = \max\{\delta,\, \max\{\, H_{ij} \,|\, i = 1, \ldots, n_1\}\}\,, \quad j = 1, \ldots, n_2\,,$$

where $\delta > 0$ *is a small positive number. Let* $D_1 = diag(d_1)$ *and* $D_2 = diag(d_2)$. *Then we consider the following problem instead of the original problem (4.1)*

$$\begin{array}{ll} \min & \dfrac{1}{2}\|H \circ (X - C)\|^2 + \rho\|D_1^{\frac{1}{2}} X D_2^{\frac{1}{2}}\|_* \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & \mathrm{rank}(X) \le r\,, \\ & X \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.18}$$

which can be equivalently written as

$$\begin{array}{ll} \min & \dfrac{1}{2}\|\widetilde{H} \circ (\widetilde{X} - \widetilde{C})\|^2 + \rho\|\widetilde{X}\|_* \\ \text{s.t.} & \widetilde{\mathcal{A}}\widetilde{X} := \mathcal{A}X \in b + Q\,, \\ & \mathrm{rank}(\widetilde{X}) \le r\,, \\ & \widetilde{X} \in \Re^{n_1 \times n_2}\,, \end{array} \tag{4.19}$$

where $\widetilde{H} = D_1^{-1/2} H D_2^{-1/2}$, $\widetilde{X} = D_1^{1/2} X D_2^{1/2}$ *and* $\widetilde{C} = D_1^{1/2} C D_2^{1/2}$.

Note that problem (4.19) now takes the same form as problem (4.1). Moreover, the components of the weight matrix $\widetilde{H}$ *are all in* $[0, 1]$.

Remark 4.9. *Alternatively, one may also apply Algorithm 3.2 to problem (4.13), which again leads to a sequence of least squares problems. We omit the details here.*

Now we turn our attention to the penalized problem (4.11) in the symmetric setting. Similarly, we eventually need to solve a sequence of problems in the following form

$$\begin{array}{ll} \min & \hat{f}_c^k(X) = \dfrac{1}{2}\|\widehat{H}^k \circ (X - X^k)\|^2 + \left\langle X,\, H \circ H \circ (X^k - C) - cW^k \right\rangle + g_c^k \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}_+^n\,, \end{array} \tag{4.20}$$

where $g_c^k := \theta(X^k) - \langle \nabla\theta(X^k),\, X^k\rangle - cp(X^k) + c\langle W^k,\, X^k\rangle$. For the sake of easy computations, in our implementation, we always choose a positive vector $d \in \Re^n$ such that $H_{ij} \le \widehat{H}_{ij}^k = \sqrt{d_i d_j}$ for all $i, j \in \{1, \ldots, n\}$. Let $D = \mathrm{diag}(d)$. Then the objective function $\hat{f}_c^k(\cdot)$ in (4.20) can be equivalently written as

$$\begin{aligned} \hat{f}_c^k(X) &= \tfrac{1}{2}\|D^{1/2}(X - X^k)D^{1/2}\|^2 + \big\langle X, H \circ H \circ (X^k - C) - cW^k\big\rangle + g_c^k \\ &= \tfrac{1}{2}\|D^{1/2}\big(X - (X^k + C^k)\big)D^{1/2}\|^2 + f_c(X^k) - \tfrac{1}{2}\|D^{1/2}C^k D^{1/2}\|^2\,, \end{aligned}$$

where $C^k := D^{-1}\big(cW^k - H \circ H \circ (X^k - C)\big)D^{-1}$. By dropping the constant terms in $\hat{f}_c^k(X)$, we can equivalently write problem (4.20) as the following well-studied diagonally weighted least squares positive semidefinite problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|D^{1/2}\big(X - (X^k + C^k)\big)D^{1/2}\|^2 \\ \text{s.t.} \quad & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}_+^n\,, \end{aligned} \tag{4.21}$$

which can be solved efficiently by the recently developed smoothing Newton-BiCGStab method [42].

For the choice of $d \in \Re^n$, one can simply take

$$d_1 = \ldots = d_n = \max\big\{\delta,\, \max\{\, H_{ij} \,|\, i, j = 1, \ldots, n\}\big\}\,, \tag{4.22}$$

where $\delta > 0$ is a small positive number. However, a better way is to choose $d \in \Re^n$ as follows

$$d_i = \max\big\{\delta,\, \max\{\, H_{ij} \,|\, j = 1, \ldots, n\,\}\big\}\,, \quad i = 1, \ldots, n\,. \tag{4.23}$$

Remark 4.10. *The choice of d in (4.22) is simpler and will lead to an unweighted least squares problem. The disadvantage of this choice is that the resulting problem generally takes more iterations to converge than the one obtained from the choice of (4.23) due to the fact that the error $\|H - dd^T\|$ is larger for the choice of (4.22). If H takes the form of hh^T for some column vector $\Re^n \ni h > 0$, we can just take $\widehat{H}^k \equiv H$ for all $k \ge 1$. In this case, the majorization function of $\theta(\cdot)$ is itself.*

4.2.2 Some rationale for the penalty approach

Now we consider the following simplest case of problem (4.1)

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X-C\|^2+\rho\|X\|_* \\ \text{s.t.} & \text{rank}(X)\le r\,, \\ & X\in\Re^{n_1\times n_2}\,, \end{array} \tag{4.24}$$

i.e., there is no weight matrix H and no linear constraints for X. Suppose that the given matrix C has the singular value decomposition as in (2.48), i.e.,

$$C=U[\Sigma(C)\ \ \mathbf{0}]V^T\,, \tag{4.25}$$

where $U\in\mathcal{O}^{n_1}$, $V\in\mathcal{O}^{n_2}$ and $\Sigma(C)=\text{diag}(\sigma(C))=(\sigma_1(C),\ldots,\sigma_{n_1}(C))^T$ with $\sigma_1(C)\ge\cdots\ge\sigma_{n_1}(C)\ge 0$. Write

$$U=[\,U_1,\ldots,U_{n_1}\,]\quad\text{and}\quad V=[\,V_1,\ldots,V_{n_2}\,].$$

Recall that problem (4.24) is exactly the problem (2.51) we studied in Chapter 2 and one of its global optimal solution is given by

$$X^*:=\sum_{i=1}^{r}\big(\sigma_i(C)-\rho\big)_+U_iV_i^T.$$

Now we claim that this global optimal solution to problem (4.24) can be obtained in two iterations by our majorized penalty approach provided that the penalty parameter $c\ge\sigma_{r+1}(C)-\rho$.

To prove this claim, let the initial point $X^0=0$. Then $W^0=0$. Noting that $X^{k+1}=\mathcal{P}_{\rho+c}(C+cW^k)$, we obtain that

$$X^1=\mathcal{P}_{\rho+c}(C+cW^0)=\sum_{i=1}^{n_1}\big(\sigma_i(C)-\rho-c\big)_+U_iV_i^T\,,$$

and thus $W^1\in\partial_B p_\sigma(X^1)$ can be simply chosen as $W^1=\sum_{i=1}^{r}U_iV_i^T$. It then

follows that

$$\begin{aligned} X^2 &= \mathcal{P}_{\rho+c}(C + cW^1) \\ &= \sum_{i=1}^{r} \big(\sigma_i(C) - \rho\big)_+ U_i V_i^T + \sum_{i=r+1}^{n_1} \big(\sigma_i(C) - \rho - c\big)_+ U_i V_i^T \\ &= \sum_{i=1}^{r} \big(\sigma_i(C) - \rho\big)_+ U_i V_i^T = X^* \,, \end{aligned}$$

which implies that we can recover the original optimal solution to problem **??** by solving its penalized problem.

This interesting result provides us the justification for using the penalty approach to deal with the rank constraint.

4.3 The Lagrangian dual reformulation

In this section, we shall study the Lagrangian dual problems in both nonsymmetric case and symmetric case in order to check the optimality of the solutions obtained by applying the proximal subgradient method to the penalized problems.

4.3.1 The Lagrangian dual problem for the nonsymmetric problem

We first study the Lagrangian dual of problem (4.1), which takes the form as follows

$$\max_{y \in \mathcal{Q}^*} V(y) := \inf_{X \in \Re_r^{n_1 \times n_2}} L(X, y) \,, \tag{4.26}$$

where $L(X, y)$ is the Lagrangian function of (4.1)

$$L(X, y) := \frac{1}{2}\|H \circ (X - C)\|^2 + \rho\|X\|_* + \langle b - \mathcal{A}X,\, y\rangle \,, \quad (X, y) \in \Re^{n_1 \times n_2} \times \Re^m \,.$$

Suppose that $\bar{y} \in \mathcal{Q}^*$ is an optimal solution to (4.26). Then for any feasible solution $\widehat{X}$ to (4.1), one has

$$\begin{aligned} & \frac{1}{2}\|H \circ (\widehat{X} - C)\|^2 + \rho\|\widehat{X}\|_* \\ \geq \; & \frac{1}{2}\|H \circ (\widehat{X} - C)\|^2 + \rho\|\widehat{X}\|_* + \langle b - \mathcal{A}\widehat{X},\, \bar{y}\rangle \\ \geq \; & V(\bar{y})\,, \end{aligned} \tag{4.27}$$

which implies that the optimal dual value $V(\bar{y})$ provides a valid lower bound for checking the optimality of the primal solution. When H is the matrix with all the entries equal to 1, we can further simplify the expression for $V(y)$ and write it explicitly as

$$\begin{aligned} & V(y) \\ = \; & \inf_{X \in \Re^{n_1 \times n_2}_r} \left\{ \frac{1}{2}\|X - C\|^2 + \rho\|X\|_* + \langle b - \mathcal{A}X,\, y\rangle \right\} \\ = \; & \inf_{X \in \Re^{n_1 \times n_2}_r} \left\{ \frac{1}{2}\|X - (C + \mathcal{A}^* y)\|^2 + \rho\|X\|_* + \langle b, y\rangle - \frac{1}{2}\|C + \mathcal{A}^* y\|^2 + \frac{1}{2}\|C\|^2 \right\} \\ = \; & -\frac{1}{2}\|\mathcal{P}_{\rho,r}(C + \mathcal{A}^* y)\|^2 + \langle b, y\rangle + \frac{1}{2}\|C\|^2\,, \end{aligned}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. Define $\Phi(y) := -V(y) + \frac{1}{2}\|C\|^2$ for any $y \in \mathcal{Q}^*$. Now we can rewrite the dual problem as follows

$$\begin{array}{ll} \min & \Phi(y) = \dfrac{1}{2}\|\mathcal{P}_{\rho,r}(C + \mathcal{A}^* y)\|^2 - \langle b, y\rangle \\ \text{s.t.} & y \in \mathcal{Q}^* = \Re^p \times \widehat{\mathcal{Q}}^*\,. \end{array} \tag{4.28}$$

In order to facilitate subsequent analysis, we first rewrite $\mathcal{A}$ and b as

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}^p \\ \mathcal{A}^q \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b^p \\ b^q \end{bmatrix},$$

where $\mathcal{A}^p : \Re^{n_1 \times n_2} \to \Re^p$, $\mathcal{A}^q : \Re^{n_1 \times n_2} \to \Re^q$, $b^p \in \Re^p$ and $b^q \in \Re^q$.

Now we discuss the existence of the optimal solutions to (4.28). For this purpose, we need the following Slater condition:

$$\begin{cases} \mathcal{A}^p \text{ is onto, and} \\ \exists\, X^0 \in \Re^{n_1 \times n_2} \text{ such that } \mathcal{A}^p X^0 = b^p \text{ and } \mathcal{A}^q X^0 - b^q \in \mathrm{int}(\widehat{\mathcal{Q}})\,. \end{cases} \tag{4.29}$$

Using Proposition 2.20 and Remark 2.21, we have the following corollary.

Corollary 4.11. *Assume that the Slater condition (4.29) holds. Then* $\langle b, \bar{y}\rangle < 0$ *for any* $0 \neq \bar{y} \in \mathcal{Q}^*$ *satisfying* $\mathcal{A}^*\bar{y} = 0$ *.*

Proposition 4.12. *Assume that the Slater condition (4.29) holds. Then, for any constant* $\nu \in \Re$*, the level set* $L_\nu := \big\{y \in \mathcal{Q}^* \mid \Phi(y) \leq \nu\big\}$ *is bounded.*

Proof. We prove the conclusion of this proposition by contradiction. Suppose that on the contrary that there exists a constant $\nu \in \Re$ such that L_ν is unbounded. Then there exists a sequence $\{y^k\} \in \mathcal{Q}^*$ such that $\Phi(y^k) \leq \nu$ for all $k \geq 1$ and $\limsup_{k\to+\infty} \|y^k\| = +\infty$. Without loss of generality, we may assume that $y^k \neq 0$ for each $k \geq 1$ and $||y^k|| \to \infty$ as $k \to \infty$. We assume, by taking a subsequence if necessary, that there exists $\bar{y} \neq 0$ such that

$$\lim_{k\to+\infty} \frac{y^k}{\|y^k\|} = \bar{y}\,.$$

Next we consider the following two subcases:

1). $\mathcal{A}^*\bar{y} \neq 0$. Let $D^k := C + \mathcal{A}^*y^k$ and its singular value decomposition (SVD) be

$$D^k = U^k[\Sigma^k \ \mathbf{0}](V^k)^T,$$

where $U^k \in \Re^{n_1\times n_1}$ and $V^k \in \Re^{n_2\times n_2}$ are two orthogonal matrices, $\Sigma^k := \mathrm{diag}(\sigma_1^k, \ldots, \sigma_{n_1}^k)$, and $\sigma_1^k \geq \cdots \geq \sigma_{n_1}^k \geq 0$ are singular values of D^k. Let $B^k := D^k/\|y^k\|$. Then $B^k = U^k[\frac{\Sigma^k}{\|y^k\|} \ 0](V^k)^T \to \mathcal{A}^*\bar{y}$. It follows that there exists a positive number $\delta > 0$ such that $\frac{\sigma_1^k}{\|y^k\|} \geq 2\delta > 0$ and $\delta\|y^k\| > \rho$ for k sufficiently large. Hence, we have

$$\|\mathcal{P}_{\rho,r}(C + \mathcal{A}^*y^k)\| = \|\mathcal{P}_{\rho,r}(D^k)\| \geq \max(\sigma_1^k - \rho, 0) \geq 2\delta\|y^k\| - \rho \geq \delta\|y^k\|\,,$$

and thus,

$$\begin{aligned}\liminf_{k\to+\infty} \Phi(y^k) &= \liminf_{k\to+\infty} \Big(\frac{1}{2}\|\mathcal{P}_{\rho,r}(C + \mathcal{A}^*y^k)\|^2 - \langle b, y^k\rangle\Big)\\ &\geq \liminf_{k\to+\infty} \|y^k\|\Big(\frac{\delta^2}{2}\|y^k\| - \|b\|\Big) = +\infty\,.\end{aligned}$$

2). $\mathcal{A}^*\bar{y} = 0$. Then $\langle b, \bar{y}\rangle < 0$ follows immediately from Corollary 4.11. Therefore,

$$\liminf_{k\to+\infty} \Phi(y^k) \geq \liminf_{k\to+\infty} \|y^k\|\big(-\langle b, y^k/\|y^k\|\rangle\big) \geq -\langle b, \bar{y}\rangle \liminf_{k\to+\infty} \|y^k\|/2 = +\infty\,.$$

In summary, we have shown that $\Phi(y^k) \to +\infty$ as $k \to \infty$, which is a contradiction to our assumption that $\Phi(y^k) \leq \nu$ for all $k \geq 1$. This contradiction shows that the conclusion of this proposition holds. □

Proposition 4.12 says that if the Slater condition (4.29) holds, the dual problem (4.28) always has optimal solutions. Let $\bar{y} \in \mathcal{Q}^*$ be an optimal solution to (4.28). Then we have

$$0 \in \partial\Phi(\bar{y}) + \mathcal{N}_{\mathcal{Q}^*}(\bar{y})\,. \tag{4.30}$$

Theorem 4.13. *The optimal solution $\bar{y} \in \mathcal{Q}^*$ to the dual problem (4.28) satisfies*

$$\emptyset \neq \big(b - \mathcal{A}\mathrm{conv}\,\mathcal{P}_{\rho,r}(C + \mathcal{A}^*\bar{y})\big) \cap \mathcal{N}_{\mathcal{Q}^*}(\bar{y})\,. \tag{4.31}$$

Furthermore, if there exists a matrix $\overline{X} \in \mathcal{P}_{\rho,r}(C + \mathcal{A}^*\bar{y})$ such that $b - \mathcal{A}\overline{X} \in \mathcal{N}_{\mathcal{Q}^*}(\bar{y})$, then $\overline{X}$ and $\bar{y}$ globally solve the primal problem (4.1) with $H = E$ and the corresponding dual problem (4.28), respectively and there is no duality gap between the primal and dual problems.

Proof. Recall that for $y \in \mathcal{Q}^*$, $\Phi(y) = \dfrac{1}{2}\|\mathcal{P}_{\rho,r}(C+\mathcal{A}^*y)\|^2 - \langle b, y\rangle$. From Proposition 2.16, we know that the sub-differential of $\Phi(\cdot)$ at the optimal solution point $\bar{y}$ can be written as

$$\partial\Phi(\bar{y}) = \mathcal{A}\mathrm{conv}\,\mathcal{P}_{\rho,r}(C + \mathcal{A}^*\bar{y}) - b\,. \tag{4.32}$$

Then (4.31) now follows directly from (4.30). If there exists a matrix $\overline{X} \in \mathcal{P}_{\rho,r}(C + \mathcal{A}^*y)$ such that $b - \mathcal{A}\overline{X} \in \mathcal{N}_{\mathcal{Q}^*}(\bar{y})$, we have that

$$\mathcal{A}\overline{X} \in b + \mathcal{Q} \;\text{ and }\; \big\langle b - \mathcal{A}\overline{X},\, \bar{y}\big\rangle = 0\,.$$

Then $\overline{X}$ is feasible to the primal problem (4.1) and that

$$V(\bar{y}) = \frac{1}{2}\|\overline{X} - C\|^2 + \rho\|\overline{X}\|_* + \langle b - \mathcal{A}\overline{X},\, \bar{y}\rangle = \frac{1}{2}\|\overline{X} - C\|^2 + \rho\|\overline{X}\|_*\,,$$

which, together with the fact that $\bar{y} \in \mathcal{Q}^*$ is feasible to the dual problem (4.28), completes the proof of the remaining part of the theorem.

□

Corollary 4.14. *Let $\bar{y}$ be an optimal solution of (4.28). If $\sigma_r(C+\mathcal{A}^*\bar{y}) > \sigma_{r+1}(C+\mathcal{A}^*\bar{y}) > 0$ or $\sigma_{r+1}(C+\mathcal{A}^*\bar{y}) = 0$, then $\overline{X} = \mathcal{P}_{\rho,r}(C+\mathcal{A}^*\bar{y})$ globally solves problem (4.1).*

Proof. It follows directly from Remark 2.17. □

4.3.2 The Lagrangian dual problem for the symmetric problem

In this subsection, we shall study the Lagrangian dual of problem (4.10), i.e,

$$\begin{array}{ll} \min & \dfrac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \mathrm{rank}(X) \le r\,. \end{array} \tag{4.33}$$

The Lagrangian function of (4.33) is

$$L(X, y) = \frac{1}{2}\|H \circ (X - C)\|^2 + \langle b - \mathcal{A}X,\, y\rangle\,, \qquad (X, y) \in \mathcal{S}^n \times \Re^m.$$

Then the Lagrangian dual problem of (4.33) takes the form of

$$\max_{y \in \mathcal{Q}^*} V(y)\,, \tag{4.34}$$

where $\mathcal{Q}^*$ is the dual cone of $\mathcal{Q}$ and $V(y)$ is defined by

$$V(y) := \inf_{X \in \mathcal{S}^n_+} L(X, y) = \inf_{X \in \mathcal{S}^n_+(r)} \left\{ \frac{1}{2}\|H \circ (X - C)\|^2 + \langle b - \mathcal{A}X,\, y\rangle \right\}. \quad (4.35)$$

Suppose that $\bar{y} \in \mathcal{Q}^*$ is an optimal solution to (4.34). Then for any feasible $\widehat{X}$ to (4.33), one has

$$\begin{array}{rcl} \frac{1}{2}\|H \circ (\widehat{X} - C)\|^2 & \geq & \frac{1}{2}\|H \circ (\widehat{X} - C)\|^2 + \langle b - \mathcal{A}\widehat{X},\, \bar{y}\rangle \\ & \geq & V(\bar{y})\,, \end{array} \quad (4.36)$$

which implies that the dual solution $\bar{y}$ provides a valid lower bound for checking the optimality of the primal solution. When H is the matrix with all the entries equal to 1, we can further simplify (4.35) and write $V(y)$ explicitly as

$$\begin{array}{rcl} V(y) & = & \inf_{X \in \mathcal{S}^n_+(r)} \left\{ \frac{1}{2}\|X - C\|^2 + \langle b - \mathcal{A}X,\, y\rangle \right\} \\ & = & \inf_{X \in \mathcal{S}^n_+(r)} \left\{ \frac{1}{2}\|X - (C + \mathcal{A}^*y)\|^2 - \frac{1}{2}\|C + \mathcal{A}^*y\|^2 + \langle b, y\rangle + \frac{1}{2}\|C\|^2 \right\} \\ & = & \frac{1}{2}\|\Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*y) - (C + \mathcal{A}^*y)\|^2 - \frac{1}{2}\|C + \mathcal{A}^*y\|^2 + \langle b, y\rangle + \frac{1}{2}\|C\|^2 \\ & = & -\frac{1}{2}\|\Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*y)\|^2 + \langle b, y\rangle + \frac{1}{2}\|C\|^2 \end{array}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. For any $y \in \mathcal{Q}^*$, let $\Phi^s(y) := -V(y) + \frac{1}{2}\|C\|^2$. Now we can rewrite the dual problem as follows

$$\begin{array}{ll} \min & \Phi^s(y) = \frac{1}{2}\|\Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*y)\|^2 - \langle b, y\rangle \\ \text{s.t.} & y \in \mathcal{Q}^* = \Re^p \times \Re^q_+\,. \end{array} \quad (4.37)$$

Remark 4.15. *When H takes the form of $H = hh^T$ for some column vector $h > 0$ in $\Re^n$, we can also derive a similar explicit expression for $V(y)$ as follows*

$$V(y) = -\frac{1}{2}\|\Pi_{\mathcal{S}^n_+(r)}\big(D^{\frac{1}{2}}(C + D^{-1}\mathcal{A}^*yD^{-1})D^{\frac{1}{2}}\big)\|^2 + \langle b, y\rangle + \frac{1}{2}\|D^{\frac{1}{2}}CD^{\frac{1}{2}}\|^2\,,$$

where $D^{\frac{1}{2}} = \mathrm{diag}(h)$. For the general weight matrix H, we cannot reformulate (4.35) explicitly. However, we can still apply the majorized penalty method introduced earlier in this paper to compute $V(y)$.

Next we discuss the existence of the optimal solution to (4.37). For this purpose, we need the following Slater condition:

$$\begin{cases} \{\mathcal{A}_i\}_{i=1}^p \text{ are linearly independent,} \\ \text{there exists } X^0 \succ 0 \text{ such that } \mathcal{A}_j X^0 = b_j \text{ for } j = 1, \ldots, p\,, \\ \text{and } \mathcal{A}_j X^0 > b_j \text{ for } j = p+1, \ldots, m\,. \end{cases} \tag{4.38}$$

Using Proposition 2.20 and Remark 2.21, we have the following corollary.

Corollary 4.16. *Assume that the Slater condition (4.38) holds. Then* $\langle b, \bar{y}\rangle < 0$ *for any* $0 \neq \bar{y} \in \mathcal{Q}^*$ *satisfying* $\mathcal{A}^*\bar{y} \preceq 0$ *.*

Proposition 4.17. *Assume that the Slater condition (4.38) holds. Then, for any constant* $\nu \in \Re$*, the level set* $L_\nu := \left\{y \in \mathcal{Q}^* \mid \Phi^s(y) \leq \nu\right\}$ *is bounded.*

Proof. We prove the conclusion of this proposition by contradiction. Suppose that on the contrary that there exists a constant $\nu \in \Re$ such that L_ν is unbounded. Then there exists a sequence $\{y^k\} \in \mathcal{Q}^*$ such that $\Phi^s(y^k) \leq \nu$ for all $k \geq 1$ and $\limsup_{k\to+\infty} \|y^k\| = +\infty$. Without loss of generality, we may assume that $\|y^k\| \neq 0$ for each $k \geq 1$ and $||y^k|| \to \infty$ as $k \to \infty$. For $k \geq 1$, let $B^k := (C + \mathcal{A}^*y^k)/\|y^k\|$. We assume, by taking a subsequence if necessary, that there exists $\bar{y} \neq 0$ such that

$$\lim_{k\to+\infty} \frac{y^k}{\|y^k\|} = \bar{y}\,.$$

Next we consider the following two cases:

1). $\mathcal{A}^*\bar{y} \not\succeq 0$, i.e., $\mathcal{A}^*\bar{y}$ has at least one positive eigenvalue. It then follows that there exists a positive number $\delta > 0$ such that

$$\liminf_{k\to+\infty} \|\Pi_{\mathcal{S}^n_+(r)}(B^k)\|^2 = \liminf_{k\to+\infty} \|\Pi_{\mathcal{S}^n_+(r)}(\mathcal{A}^*\bar{y})\|^2 \geq \delta > 0\,.$$

Hence, we have

$$\begin{aligned} \liminf_{k\to+\infty} \Phi^s(y^k) &= \liminf_{k\to+\infty} \Big(\frac{1}{2}\|\Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*y^k)\|^2 - \langle b, y^k\rangle\Big) \\ &\geq \liminf_{k\to+\infty} \|y^k\|\Big(\frac{1}{2}\|y^k\|\|\Pi_{\mathcal{S}^n_+(r)}(B^k)\|^2 - \|b\|\Big) = +\infty\,. \end{aligned}$$

2). $\mathcal{A}^*\bar{y} \preceq 0$. Then $\langle b, \bar{y}\rangle < 0$ follows immediately from Corollary 4.16. Therefore,

$$\liminf_{k\to+\infty} \Phi^s(y^k) \geq \liminf_{k\to+\infty} \|y^k\|\big(-\langle b, y^k/\|y^k\|\rangle\big) \geq -\langle b, \bar{y}\rangle \liminf_{k\to+\infty} \|y^k\|/2 = +\infty\,.$$

In summary, we have shown that $\Phi^s(y^k) \to +\infty$ as $k \to \infty$, which is a contradiction to our assumption that $\Phi^s(y^k) \leq \nu$ for all $k \geq 1$. This contradiction shows that the conclusion of this proposition holds. □

Proposition 4.17 says that if the Slater condition (4.38) holds, the dual problem (4.37) always has optimal solutions. Let $\bar{y} \in \mathcal{Q}^*$ be an optimal solution to (4.37). Then we have

$$0 \in \partial\Phi^s(\bar{y}) + \mathcal{N}_{\mathcal{Q}^*}(\bar{y})\,. \tag{4.39}$$

Theorem 4.18. *The optimal solution $\bar{y} \in \mathcal{Q}^*$ to the dual problem (4.37) satisfies*

$$\emptyset \neq \big(b - \mathcal{A}\mathrm{conv}\,\Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*\bar{y})\big) \cap \mathcal{N}_{\mathcal{Q}^*}(\bar{y})\,. \tag{4.40}$$

Furthermore, if there exists a matrix $\overline{X} \in \Pi_{\mathcal{S}^n_+(r)}\big(C + \mathcal{A}^*\bar{y}\big)$ such that $b - \mathcal{A}\overline{X} \in \mathcal{N}_{\mathcal{Q}^*}(\bar{y})$, then $\overline{X}$ and $\bar{y}$ globally solve the primal problem (4.33) with $H = E$ and the corresponding dual problem (4.37), respectively and there is no duality gap between the primal and dual problems.

Proof. From Proposition 2.5, we know that the sub-differential of $\Phi^s(\cdot)$ at the optimal solution point $\bar{y}$ can be written as

$$\partial\Phi^s(\bar{y}) = \mathcal{A}\mathrm{conv}\,\Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*\bar{y}) - b\,. \tag{4.41}$$

Then (4.40) now follows directly from (4.39). If there exists a matrix $\overline{X} \in \Pi_{\mathcal{S}^n_+(r)}(C + \mathcal{A}^*\bar{y})$ such that $b - \mathcal{A}\overline{X} \in \mathcal{N}_{\mathcal{Q}^*}(\bar{y})$, we know that

$$\mathcal{A}\overline{X} \in b + \mathcal{Q} \;\text{ and }\; \big\langle b - \mathcal{A}\overline{X},\, \bar{y}\big\rangle = 0\,.$$

Then $\overline{X}$ is feasible to the primal problem (4.33) and

$$V(\bar{y}) = \frac{1}{2}\|\overline{X} - C\|^2 + \langle b - \mathcal{A}\overline{X}, \bar{y}\rangle = \frac{1}{2}\|\overline{X} - C\|^2 ,$$

which, together with the fact that $\bar{y} \in \mathcal{Q}^*$ is feasible to the dual problem (4.37), completes the proof of the remaining part of the theorem. □

Corollary 4.19. *Let $\bar{y}$ be an optimal solution of (4.37). If $\lambda_r(C+\mathcal{A}^*\bar{y}) > \lambda_{r+1}(C+\mathcal{A}^*\bar{y}) > 0$ or $\lambda_{r+1}(C+\mathcal{A}^*\bar{y}) \le 0$, then $\overline{X} = \Pi_{\mathcal{S}^n_+(r)}(C+\mathcal{A}^*\bar{y})$ globally solves problem (4.33).*

Proof. It follows directly from Remark 2.6. □

Remark 4.20. *Theorem 4.18 also holds for the following W-weighted problem*

$$\begin{array}{ll} \min & \frac{1}{2}\|W^{1/2}(X-C)W^{1/2}\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q} , \\ & X \in \mathcal{S}^n_+ , \\ & \mathrm{rank}(X) \le r , \end{array} \tag{4.42}$$

where W is a symmetric positive definite matrix.

Remark 4.21. *Theorem 4.18 can be regarded as an extension of the globalization checking results of Zhang and Wu [123, Theorem 4.5] which only holds for a special kind of correlation matrix calibration problems. However, the technique introduced in Theorem 4.18 allows us to deal with more general cases in several aspects:*

(E1). The matrix C is no longer required to be a valid correlation matrix.

(E2). The problem may have more general constraints including the simple lower and upper bound constraints.

(E3). The assumption $|\lambda_r\big(C + \mathrm{diag}(\bar{y})\big)| > |\lambda_{r+1}\big(C + \mathrm{diag}(\bar{y})\big)|$ is much weaker to include more general situations.

Chapter 5

A Smoothing Newton-BiCGStab Method

5.1 The algorithm

The purpose of this section is to introduce an inexact smoothing Newton method for solving the general nonsmooth equation

$$F(y) = 0, \quad y \in \Re^m ,$$

where $F : \Re^m \to \Re^m$ is a locally Lipschitz continuous function. This inexact smoothing Newton method is largely modified from the exact smoothing Newton method constructed in [94] for solving complementarity and variational inequality problems. The motivation to introduce an inexact version is completely from the computational point of view because the costs of the exact smoothing Newton method for solving problems such as the LSSDP problem (5.16) are prohibitive.

Let $G : \Re \times \Re^m \to \Re^m$ be a locally Lipschitz continuous function satisfying

$$G(\varepsilon, y') \to F(y) \quad \text{as} \quad (\varepsilon, y') \to (0, y) .$$

Furthermore, G is required to be continuously differentiable around any (ε, y) unless $\varepsilon = 0$. The existence of such a function G can be easily proven via convolution.

Define $E : \Re \times \Re^m \to \Re \times \Re^m$ by

$$E(\varepsilon, y) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, y) \end{bmatrix}, \quad (\varepsilon, y) \in \Re \times \Re^m .$$

Then solving the nonsmooth equation $F(y) = 0$ is equivalent to solving the following smoothing-nonsmooth equation

$$E(\varepsilon, y) = 0 .$$

Our inexact smoothing Newton method is specifically designed for solving the latter one.

Define the merit function $\varphi : \Re \times \Re^m \to \Re_+$ by

$$\varphi(\varepsilon, y) := \|E(\varepsilon, y)\|^2 , \quad (\varepsilon, y) \in \Re \times \Re^m .$$

Choose $r \in (0, 1)$. Let

$$\zeta(\varepsilon, y) := r \min\{1, \varphi(\varepsilon, y)\} , \quad (\varepsilon, y) \in \Re \times \Re^m .$$

Then the inexact smoothing Newton method can be described as follows.

Algorithm 5.1. (An inexact smoothing Newton method)

Step 0. *Let $\hat{\varepsilon} \in (0, \infty)$ and $\eta \in (0, 1)$ be such that*

$$\delta := \sqrt{2} \max\{r\hat{\varepsilon}, \eta\} < 1 .$$

Select constants $\rho \in (0, 1)$, $\sigma \in (0, 1/2)$, $\tau \in (0, 1)$, and $\hat{\tau} \in [1, \infty)$. Let $\varepsilon^0 := \hat{\varepsilon}$ and $y^0 \in \Re^m$ be an arbitrary point. $k := 0$.

Step 1. *If $E(\varepsilon^k, y^k) = 0$, then stop. Otherwise, compute*

$$\zeta_k := r \min\{1, \varphi(\varepsilon^k, y^k)\} \quad \text{and} \quad \eta_k := \min\{\tau, \hat{\tau}\|E(\varepsilon^k, y^k)\|\} .$$

Step 2. *Solve the following equation*

$$E(\varepsilon^k, y^k) + E'(\varepsilon^k, y^k) \begin{bmatrix} \Delta\varepsilon^k \\ \Delta y^k \end{bmatrix} = \begin{bmatrix} \zeta_k \hat{\varepsilon} \\ 0 \end{bmatrix} \tag{5.1}$$

approximately such that

$$\|R_k\| \leq \min\{\eta_k \|G(\varepsilon^k, y^k) + G'_\varepsilon(\varepsilon^k, y^k)\Delta\varepsilon^k\|,\, \eta\|E(\varepsilon^k, y^k)\|\}\,, \tag{5.2}$$

where

$$\Delta\varepsilon^k := -\varepsilon^k + \zeta_k\hat{\varepsilon}$$

and

$$R_k := G(\varepsilon^k, y^k) + G'(\varepsilon^k, y^k) \begin{bmatrix} \Delta\varepsilon^k \\ \Delta y^k \end{bmatrix}.$$

Step 3. *Let l_k be the smallest nonnegative integer l satisfying*

$$\varphi(\varepsilon^k + \rho^l\Delta\varepsilon^k, y^k + \rho^l\Delta y^k) \leq [1 - 2\sigma(1-\delta)\rho^l]\varphi(\varepsilon^k, y^k)\,. \tag{5.3}$$

Define:

$$(\varepsilon^{k+1}, y^{k+1}) := (\varepsilon^k + \rho^{l_k}\Delta\varepsilon^k, y^k + \rho^{l_k}\Delta y^k)\,.$$

Step 4. *Replace k by $k+1$ and go to* **Step 1**.

Lemma 5.2. *If for some $(\tilde{\varepsilon}, \tilde{y}) \in \Re_{++} \times \Re^m$, $E'(\tilde{\varepsilon}, \tilde{y})$ is nonsingular, then there exist an open neighborhood $\mathcal{O}$ of $(\tilde{\varepsilon}, \tilde{y})$ and a positive number $\bar{\alpha} \in (0, 1]$ such that for any $(\varepsilon, y) \in \mathcal{O}$ and $\alpha \in [0, \bar{\alpha}]$, $\varepsilon \in \Re_{++}$, $E'(\varepsilon, y)$ is nonsingular, and*

$$\varphi(\varepsilon + \alpha\Delta\varepsilon, y + \alpha\Delta y) \leq [1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon, y)\,, \tag{5.4}$$

where $(\Delta\varepsilon, \Delta y) \in \Re \times \Re^m$ satisfies

$$\Delta\varepsilon = -\varepsilon + \zeta(\varepsilon, y)\hat{\varepsilon}$$

and

$$\left\| G(\varepsilon, y) + G'(\varepsilon, y) \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\| \leq \eta\|E(\varepsilon, y)\|\,.$$

Proof. Since $\tilde{\varepsilon} \in \Re_{++}$ and $E'(\tilde{\varepsilon}, \tilde{y})$ is nonsingular, there exists an open neighborhood $\mathcal{O}$ of $(\tilde{\varepsilon}, \tilde{y})$ such that for any $(\varepsilon, y) \in \mathcal{O}$, $\varepsilon \in \Re_{++}$ and $E'(\varepsilon, y)$ is nonsingular. For any $(\varepsilon, y) \in \mathcal{O}$, denote

$$R(\varepsilon, y) := G(\varepsilon, y) + G'(\varepsilon, y) \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix}.$$

Then $(\Delta\varepsilon, \Delta y)$ is the unique solution of the following equation

$$E(\varepsilon, y) \; + \; E'(\varepsilon, y) \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \; = \; \begin{bmatrix} \zeta(\varepsilon, y)\hat{\varepsilon} \\ R(\varepsilon, y) \end{bmatrix}.$$

Thus,

$$\begin{aligned}
& \left\langle \nabla\varphi(\varepsilon, y), \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\rangle = \left\langle 2\nabla E(\varepsilon, y) E(\varepsilon, y), \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\rangle \\
= \; & \left\langle 2E(\varepsilon, y), \begin{bmatrix} \zeta(\varepsilon, y)\hat{\varepsilon} \\ R(\varepsilon, y) \end{bmatrix} - E(\varepsilon, y) \right\rangle \\
= \; & -2\varphi(\varepsilon, y) + 2\varepsilon\zeta(\varepsilon, y)\hat{\varepsilon} + 2\langle R(\varepsilon, y), G(\varepsilon, y)\rangle \\
\leq \; & -2\varphi(\varepsilon, y) + 2\varepsilon(r\hat{\varepsilon}) \min\{1, \varphi(\varepsilon, y)\} + 2\eta\varphi(\varepsilon, y)^{1/2} \|G(\varepsilon, y)\|,
\end{aligned}$$

which, implies that if $\varphi(\varepsilon, y) > 1$ we have

$$\begin{aligned}
& \left\langle \nabla\varphi(\varepsilon, y), \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\rangle \\
\leq \; & -2\varphi(\varepsilon, y) + 2\varepsilon(r\hat{\varepsilon}) + 2\eta\varphi(\varepsilon, y)^{1/2} \|G(\varepsilon, y)\| \\
\leq \; & -2\varphi(\varepsilon, y) + 2\max\{r\hat{\varepsilon}, \eta\}\big(\varepsilon + \varphi(\varepsilon, y)^{1/2}\sqrt{\varphi(\varepsilon, y) - \varepsilon^2}\big) \\
\leq \; & -2\varphi(\varepsilon, y) + 2\sqrt{2}\max\{r\hat{\varepsilon}, \eta\}\varphi(\varepsilon, y) \\
= \; & 2\big(\sqrt{2}\max\{r\hat{\varepsilon}, \eta\} - 1\big)\varphi(\varepsilon, y) \qquad (5.5)
\end{aligned}$$

and if $\varphi(\varepsilon, y) < 1$ we have

$$\begin{aligned}
&\left\langle \nabla\varphi(\varepsilon, y), \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\rangle \\
\leq\ & -2\varphi(\varepsilon, y) + 2\varepsilon(r\hat{\varepsilon})\varphi(\varepsilon, y) + 2\eta\varphi(\varepsilon, y)^{1/2}\|G(\varepsilon, y)\| \\
\leq\ & -2\varphi(\varepsilon, y) + 2\max\{r\hat{\varepsilon}, \eta\}\varphi(\varepsilon, y)^{1/2}\big(\varepsilon\varphi(\varepsilon, y)^{1/2} + \sqrt{\varphi(\varepsilon, y) - \varepsilon^2}\big) \\
\leq\ & -2\varphi(\varepsilon, y) + 2\sqrt{2}\max\{r\hat{\varepsilon}, \eta\}\varphi(\varepsilon, y) \\
=\ & 2\big(\sqrt{2}\max\{r\hat{\varepsilon}, \eta\} - 1\big)\varphi(\varepsilon, y)\,. \qquad (5.6)
\end{aligned}$$

Therefore, by inequalities (5.5) and (5.6), we have

$$\left\langle \nabla\varphi(\varepsilon, y), \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\rangle \leq -2\,(1-\delta)\,\varphi(\varepsilon, y)\,. \qquad (5.7)$$

By using the fact that $\nabla\varphi(\cdot, \cdot)$ is uniformly continuous on $\mathcal{O}$, we obtain from the Taylor expansion that

$$\varphi(\varepsilon + \alpha\Delta\varepsilon, y + \alpha\Delta y) = \varphi(\varepsilon, y) + \alpha\left\langle \nabla\varphi(\varepsilon, y), \begin{bmatrix} \Delta\varepsilon \\ \Delta y \end{bmatrix} \right\rangle + o(\alpha) \quad \forall\,(\varepsilon, y) \in \mathcal{O}\,,$$

which, together with (5.7), implies that there exists a positive number $\bar{\alpha} \in (0, 1]$ such that for all $\alpha \in [0, \bar{\alpha}]$, (5.4) holds. □

Let

$$\mathcal{N} := \{(\varepsilon, y)\,|\,\varepsilon \geq \zeta(\varepsilon, y)\hat{\varepsilon}\}\,. \qquad (5.8)$$

Proposition 5.3. *For each fixed $k \geq 0$, if $\varepsilon^k \in \Re_{++}$, $(\varepsilon^k, y^k) \in \mathcal{N}$, and $E'(\varepsilon^k, y^k)$ is nonsingular, then for any $\alpha \in [0, 1]$ such that*

$$\varphi(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) \leq [\,1 - 2\sigma(1-\delta)\alpha]\varphi(\varepsilon^k, y^k) \qquad (5.9)$$

it holds that $(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k) \in \mathcal{N}$.

Proof. Note that $(\varepsilon^k, y^k) \in \mathcal{N}$, i.e., $\varepsilon^k \geq \zeta_k \hat{\varepsilon}$, so $\Delta\varepsilon^k = -\varepsilon^k + \zeta_k\hat{\varepsilon} \leq 0$. Thus, by the definition of ζ, together with (5.9), we have

$$\begin{aligned} & \varepsilon^k + \alpha\Delta\varepsilon^k - \zeta(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k)\hat{\varepsilon} \\ \geq\ & \varepsilon^k + \Delta\varepsilon^k - \zeta(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k)\hat{\varepsilon} \\ =\ & \zeta_k\hat{\varepsilon} - \zeta(\varepsilon^k + \alpha\Delta\varepsilon^k,\, y^k + \alpha\Delta y^k)\hat{\varepsilon} \\ \geq\ & 0\,. \end{aligned} \tag{5.10}$$

This completes our proof. □

In order to discuss the global convergence of Algorithm 5.1 we need the following assumption.

Assumption 5.4. *For any $(\varepsilon, y) \in \Re_{++} \times \Re^n$, $E'(\varepsilon, y)$ is nonsingular.*

Theorem 5.5. *Suppose that Assumptions 5.4 is satisfied. Then Algorithm 5.1 is well defined and generates an infinite sequence $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ with the property that any accumulation point $(\bar{\varepsilon}, \bar{y})$ of $\{(\varepsilon^k, y^k)\}$ is a solution of $E(\varepsilon, y) = 0$.*

Proof. It follows from Lemma 5.2, Proposition 5.3, and Assumption 5.4 that Algorithm 5.1 is well defined and generates an infinite sequence $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$.

From the design of Algorithm 5.1, $\varphi(\varepsilon^{k+1}, y^{k+1}) < \varphi(\varepsilon^k, y^k)$ for all $k \geq 0$. Hence, the two sequences $\{\varphi(\varepsilon^k, y^k)\}$ and $\{\zeta(\varepsilon^k, y^k)\}$ are monotonically decreasing. Since both $\varphi(\varepsilon^k, y^k)$ and $\zeta(\varepsilon^k, y^k)$ are nonnegative for $k \geq 0$, there exist $\bar{\psi} \geq 0$ and $\bar{\zeta} \geq 0$ such that $\varphi(\varepsilon^k, y^k) \to \bar{\varphi}$ and $\zeta(\varepsilon^k, y^k) \to \bar{\zeta}$ as $k \to \infty$.

Let $(\bar{\varepsilon}, \bar{y})$ be any accumulation point (if it exists) of $\{(\varepsilon^k, y^k)\}$. By taking a subsequence if necessary, we may assume that $\{(\varepsilon^k, y^k)\}$ converges to $(\bar{\varepsilon}, \bar{y})$. Then $\bar{\varphi} = \varphi(\bar{\varepsilon}, \bar{y})$, $\bar{\zeta} = \zeta(\bar{\varepsilon}, \bar{y})$, and $(\bar{\varepsilon}, \bar{y}) \in \mathcal{N}$.

Suppose that $\bar{\varphi} > 0$. Then, from $\zeta(\bar{\varepsilon}, \bar{y}) = r\min\{1, \varphi(\bar{\varepsilon}, \bar{y})\}$ and $(\bar{\varepsilon}, \bar{y}) \in \mathcal{N}$, we see that $\bar{\varepsilon} \in \Re_{++}$. Thus, from Assumption 5.4, $E'(\bar{\varepsilon}, \bar{y})$ exists and is invertible. Hence, from Lemma 5.2, there exist an open neighborhood $\mathcal{O}$ of $(\bar{\varepsilon}, \bar{y})$ and a positive

number $\bar{\alpha} \in (0,1]$ such that for any $(\varepsilon, y) \in \mathcal{O}$ and all $\alpha \in [0, \bar{\alpha}]$, $\varepsilon \in \Re_{++}$, $E'(\varepsilon, y)$ is invertible, and (5.4) holds. Therefore, there exists a nonnegative integer l such that $\rho^l \in (0, \bar{\alpha}]$ and $\rho^{l_k} \geq \rho^l$ for all k sufficiently large. Thus

$$\varphi(\varepsilon^{k+1}, y^{k+1}) \leq [1 - 2\sigma(1-\delta)\rho^{l_k}]\varphi(\varepsilon^k, y^k) \leq [1 - 2\sigma(1-\delta)\rho^{l}]\varphi(\varepsilon^k, y^k)$$

for all sufficiently large k. This contradicts the fact that the sequence $\{\varphi(\varepsilon^k, y^k)\}$ converges to $\bar{\varphi} > 0$. This contradiction shows that $\varphi(\bar{\varepsilon}, \bar{y}) = \bar{\varphi} = 0$. i.e., $E(\bar{\varepsilon}, \bar{y}) = 0$. The proof is completed. □

Theorem 5.6. *Suppose that Assumptions 5.4 is satisfied and that $(\bar{\varepsilon}, \bar{y})$ is an accumulation point of the infinite sequence $\{(\varepsilon^k, y^k)\}$ generated by Algorithm 5.1. Suppose that E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and that all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular. Then the whole sequence $\{(\varepsilon^k, y^k)\}$ converges to $(\bar{\varepsilon}, \bar{y})$ quadratically, i.e.,*

$$\left\|(\varepsilon^{k+1} - \bar{\varepsilon}, y^{k+1} - \bar{y})\right\| = O\big(\|(\varepsilon^k - \bar{\varepsilon}, y^k - \bar{y})\|^2\big)\,. \tag{5.11}$$

Proof. First, from Theorem 5.5, $(\bar{\varepsilon}, \bar{y})$ is a solution of $E(\varepsilon, y) = 0$. Then, since all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular, from [92], for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$,

$$\|E'(\varepsilon^k, y^k)^{-1}\| = O(1)$$

and

$$\begin{aligned}
& \left\| \begin{pmatrix} \varepsilon^k \\ y^k \end{pmatrix} + \begin{pmatrix} \Delta\varepsilon^k \\ \Delta y^k \end{pmatrix} - \begin{pmatrix} \bar{\varepsilon} \\ \bar{y} \end{pmatrix} \right\| \\
= & \left\| \begin{pmatrix} \varepsilon^k \\ y^k \end{pmatrix} + E'(\varepsilon^k, y^k)^{-1} \left[\begin{pmatrix} r\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ R_k \end{pmatrix} - E(\varepsilon^k, y^k) \right] - \begin{pmatrix} \bar{\varepsilon} \\ \bar{y} \end{pmatrix} \right\| \\
= & \left\| -E'(\varepsilon^k, y^k)^{-1} \left[E(\varepsilon^k, y^k) - E'(\varepsilon^k, y^k) \begin{pmatrix} \varepsilon^k - \bar{\varepsilon} \\ y^k - \bar{y} \end{pmatrix} - \begin{pmatrix} r\varphi(\varepsilon^k, y^k)\hat{\varepsilon} \\ R_k \end{pmatrix} \right] \right\| \\
= & O\left(\left\| E(\varepsilon^k, y^k) - E(\bar{\varepsilon}, \bar{y}) - E'(\varepsilon^k, y^k) \begin{pmatrix} \varepsilon^k - \bar{\varepsilon} \\ y^k - \bar{y} \end{pmatrix} \right\| \right) + O(\varphi(\varepsilon^k, y^k)) + O(\|R_k\|)\,.
\end{aligned} \tag{5.12}$$

Since E is locally Lipschitz continuous near $(\bar{\varepsilon}, \bar{y})$, for all (ε^k, y^k) close to $(\bar{\varepsilon}, \bar{y})$ we have

$$\varphi(\varepsilon^k, y^k) = \|E(\varepsilon^k, y^k) - E(\bar{\varepsilon}, \bar{y})\|^2 = O\big(\|(\varepsilon^k - \bar{\varepsilon}, y^k - \bar{y})\|^2\big) \tag{5.13}$$

and

$$\begin{aligned} \|R_k\| &\leq \eta_k \|G(\varepsilon^k, y^k) + G'_\varepsilon(\varepsilon^k, y^k)\Delta\varepsilon^k\| \\ &\leq O(\|E(\varepsilon^k, y^k)\|)\big(\|G(\varepsilon^k, y^k)\| + O(|\Delta\varepsilon^k|)\big) \\ &\leq O(\|E(\varepsilon^k, y^k) - E(\bar{\varepsilon}, \bar{y})\|^2)\,. \end{aligned} \tag{5.14}$$

Therefore, by using the assumption that E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and the relations (5.12), (5.13), and (5.14), we have for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$ that

$$\|(\varepsilon^k, y^k) + (\Delta\varepsilon^k, \Delta y^k) - (\bar{\varepsilon}, \bar{y})\| = O\big(\|(\varepsilon^k, y^k) - (\bar{\varepsilon}, \bar{y})\|^2\big)\,. \tag{5.15}$$

Finally, since E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and that all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular, we have for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$ that

$$\|(\varepsilon^k, y^k) - (\bar{\varepsilon}, \bar{y})\| \leq O(\|E(\varepsilon^k, y^k)\|)\,,$$

which, together with (5.15) and the Lipschitz continuity of E, implies that

$$\varphi(\varepsilon^k + \Delta\varepsilon^k, y^k + \Delta y^k) = O(\varphi^2(\varepsilon^k, y^k))\,.$$

This shows that for all (ε^k, y^k) sufficiently close to $(\bar{\varepsilon}, \bar{y})$,

$$(\varepsilon^{k+1}, y^{k+1}) = (\varepsilon^k, y^k) + (\Delta\varepsilon^k, \Delta y^k)\,.$$

Thus, by using (5.15) we know that (5.11) holds. □

5.2 Least squares semidefinite programming

In this section, we apply the general inexact smoothing Newton method developed in the last section to the following least squares semidefinite programming (LSSDP)

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X-C\|^2 \\ \text{s.t.} & \langle A_i, X\rangle = b_i, \quad i=1,\ldots,p\,, \\ & \langle A_i, X\rangle \geq b_i, \quad i=p+1,\ldots,m\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{5.16}$$

where $\mathcal{S}^n$ and $\mathcal{S}^n_+$ are, respectively, the space of $n\times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$, $\|\cdot\|$ is the Frobenius norm induced by the standard trace inner product $\langle\cdot,\cdot\rangle$ in $\mathcal{S}^n$, C and A_i, $i=1,\ldots,m$ are given matrices in $\mathcal{S}^n$, and $b\in\Re^m$. Mathematically, the LSSDP problem (5.16) can be equivalently written as

$$\begin{array}{ll} \min & t \\ \text{s.t.} & \langle A_i, X\rangle = b_i, \quad i=1,\ldots,p\,, \\ & \langle A_i, X\rangle \geq b_i, \quad i=p+1,\ldots,m\,, \\ & t \geq \|X-C\|\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{5.17}$$

Problem (5.17) is a linear optimization problem with linear equality/inequality, the second order cone, and the positive semidefinite cone constraints. This suggests that one may then use well developed and publicly available softwares, based on interior point methods (IPMs), such as SeDuMi [112], SDPT3 [117], and a few others to solve (5.17), and so the LSSDP problem (5.16), directly. This is indeed feasible on a Pentium IV PC (the computing machine that we will use in our numerical experiments) as long as n is small (say 80 at most) and m is not too large (say $5,000$). The reason is that at each iteration these solvers require to

formulate and solve a linear system with a dense Schur complement matrix (for example, see [5]) of the size $(m+1+\bar{n}) \times (m+1+\bar{n})$, where $\bar{n} := \frac{1}{2}n(n+1)$.

Realizing the difficulties in using IPMs to solve the LSSDP problem, in two recent papers, Malick [74] and Boyd and Xiao [7] proposed, respectively, to apply classical quasi-Newton methods (in particular, the BFGS method) and the projected gradient method to the Lagrangian dual of problem (5.16) as the objective function in the corresponding Lagrangian dual (dual in short) problem is continuously differentiable. Unlike the IPMs, these two dual based approaches are relatively inexpensive at each iteration as the dual problem is of dimension m only. The overall numerical performance of these two approaches vary from problem to problem. They may take dozens of iterations for some testing examples and several hundreds or thousands for some others.

For subsequent discussions, in this section we introduce some basic properties of matrix valued functions related to the LSSDP problem (5.16) and its dual.

Let $\mathcal{F}$ denote the feasible set of of problem (5.16). Assume that $\mathcal{F} \neq \emptyset$. Then problem (5.16) has a unique optimal solution $\overline{X}$. Let $q = m-p$ and $\mathcal{Q} = \{0\}^p \times \Re^q_+$. Denote $\mathcal{A} : \mathcal{S}^n \to \Re^m$ by

$$\mathcal{A}(X) := \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{bmatrix}, \quad X \in \mathcal{S}^n .$$

For any symmetric $X \in \mathcal{S}^n$, we write $X \succeq 0$ and $X \succ 0$ to represent that X is positive semidefinite and positive definite, respectively. Then

$$\mathcal{F} = \{X \in \mathcal{S}^n \,|\, \mathcal{A}(X) \in b + \mathcal{Q},\ X \succeq 0\}$$

and the dual problem of (5.16) takes the form

$$\begin{array}{ll} \min & \theta(y) := \dfrac{1}{2}\|\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^* y)\|^2 - \langle b, y\rangle - \dfrac{1}{2}\|C\|^2 \\ \text{s.t.} & y \in \mathcal{Q}^* = \Re^p \times \Re^q_+ . \end{array} \tag{5.18}$$

The objective function $\theta(\cdot)$ in (5.18) is a continuously differentiable convex function with

$$\nabla\theta(y) = \mathcal{A}\Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*y) - b, \quad y \in \Re^m\,,$$

where the adjoint $\mathcal{A}^* : \Re^m \to \mathcal{S}^n$ takes the form

$$\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i, \quad y \in \Re^m\,. \tag{5.19}$$

One classical dual approach described by Rockafellar in [106, Page 4], when specialized to problem (5.16), is to first find an optimal solution $\bar{y}$, if it exists, to the dual problem (5.18), and then to obtain the unique optimal solution $\overline{X}$ to problem (5.16) via $\overline{X} = \Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*\bar{y})$. See Malick [74] and Boyd and Xiao [7] for the worked out details.

In order to apply a dual based optimization method to solve problem (5.16), we need the following Slater condition to hold:

$$\begin{cases} \{A_i\}_{i=1}^p \text{ are linearly independent,} \\ \exists\ X^0 \in \mathcal{F} \text{ such that } \ \langle A_i, X^0\rangle > b_i,\ i = p+1, \ldots, m \text{ and } X^0 \succ 0\,. \end{cases} \tag{5.20}$$

The next proposition is a straightforward application of [106, Theorems 17 & 18].

Proposition 5.7. *Under the Slater condition (5.20), the following hold:*

(i) *There exists at least one $\bar{y} \in \mathcal{Q}^*$ that solves the dual problem (5.18). The unique solution to problem (5.16) is given by*

$$\overline{X} = \Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*\bar{y}). \tag{5.21}$$

(ii) *For every real number τ, the constrained level set $\{y \in \mathcal{Q}^* \,|\, \theta(y) \le \tau\}$ is closed, bounded, and convex.*

Proposition 5.7 says that one should be able to use any gradient based optimization method to find an optimal solution to the convex problem (5.18), and thus solves

problem (5.16), as long as the Slater condition (5.20) holds. Note that for any given $y \in \Re^m$, both $\theta(y)$ and $\nabla\theta(y)$ can be computed explicitly as the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$ has long been known by statisticians to admit an analytic formula [109]. Since $\theta(\cdot)$ is a convex function, $\bar{y} \in \mathcal{Q}^*$ solves problem (5.18) if and only if it satisfies the following variational inequality

$$\langle y - \bar{y}, \nabla\theta(\bar{y})\rangle \geq 0 \quad \forall\, y \in \mathcal{Q}^*\,. \tag{5.22}$$

Define $F : \Re^m \to \Re^m$ by

$$F(y) := y - \Pi_{\mathcal{Q}^*}(y - \nabla\theta(y)), \quad y \in \Re^m\,. \tag{5.23}$$

Then one can easily check that $\bar{y} \in \mathcal{Q}^*$ solves (5.22) if and only if $F(\bar{y}) = 0$ [30]. Thus, solving the dual problem (5.18) is equivalent to solving the following equation

$$F(y) = 0\,, \quad y \in \Re^m\,. \tag{5.24}$$

Since both $\Pi_{\mathcal{Q}^*}(\cdot)$ and $\Pi_{\mathcal{S}^n_+}(\cdot)$ are globally Lipschitz continuous, F is globally Lipschitz continuous. This means that though one cannot use classical Newton method to solve (5.24), one can still use Clarke's generalized Jacobian based Newton methods [61, 92, 95]. Unlike the case with equality constraints only, however, $F(\cdot)$ is no longer the gradient mapping of any real valued function. This means that we cannot use the techniques in [89] to globalize these Clarke's generalized Jacobian based Newton methods. In this paper, we shall introduce an inexact smoothing Newton method to overcome this difficulty. For this purpose, we need smoothing functions for $F(\cdot)$.

Next, we shall first discuss smoothing functions for the metric projector $\Pi_{\mathcal{S}^n_+}(\cdot)$. Let $X \in \mathcal{S}^n$. Suppose that X has the spectral decomposition

$$X = P\Lambda P^T = P\mathrm{diag}(\lambda_1, \ldots, \lambda_n)P^T, \tag{5.25}$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of X and P is a corresponding orthogonal matrix of orthonormal eigenvectors of X. Then, from [109],

$$\Pi_{\mathcal{S}^n_+}(X) = P\mathrm{diag}(\max(0,\lambda_1),\ldots,\max(0,\lambda_n))P^T\,. \tag{5.26}$$

Define

$$\alpha := \{i \mid \lambda_i > 0\},\quad \beta := \{i \mid \lambda_i = 0\},\ \text{and}\ \gamma := \{i \mid \lambda_i < 0\}.$$

Write $P = [P_\alpha\ P_\beta\ P_\gamma]$ with P_α, P_β, and P_γ containing the columns in P indexed by α, β, and γ, respectively. Let $\phi : \Re \times \Re \to \Re$ be defined by the following Huber smoothing function

$$\phi(\varepsilon,t) = \begin{cases} t & \text{if } t \geq \frac{|\varepsilon|}{2}\,, \\ \dfrac{1}{2|\varepsilon|}(t+\dfrac{|\varepsilon|}{2})^2 & \text{if } -\frac{|\varepsilon|}{2} < t < \frac{|\varepsilon|}{2}, \quad (\varepsilon,t) \in \Re\times\Re\,. \\ 0 & \text{if } t \leq -\frac{|\varepsilon|}{2}\,, \end{cases} \tag{5.27}$$

For any $\varepsilon \in \Re$, let

$$\Phi(\varepsilon,X) := P \begin{bmatrix} \phi(\varepsilon,\lambda_1) & & \\ & \ddots & \\ & & \phi(\varepsilon,\lambda_n) \end{bmatrix} P^T\,. \tag{5.28}$$

Note that when $\varepsilon = 0$, $\Phi(0,X) = \Pi_{\mathcal{S}^n_+}(X)$. From Proposition 2.1, we know that when $\varepsilon \neq 0$ or $\beta = \emptyset$,

$$\Phi'_X(\varepsilon,X)(H) = P[\Omega(\varepsilon,\lambda)\circ(P^THP)]P^T \quad \forall\, H \in \mathcal{S}^n\,, \tag{5.29}$$

where "$\circ$" denotes the Hadamard product, $\lambda = (\lambda_1,\ldots,\lambda_n)^T$, and the symmetric matrix $\Omega(\varepsilon,\lambda)$ is given by

$$\big[\,\Omega(\varepsilon,\lambda)\,\big]_{ij} = \begin{cases} \dfrac{\phi(\varepsilon,\lambda_i)-\phi(\varepsilon,\lambda_j)}{\lambda_i-\lambda_j} \in [0,1] & \text{if } \lambda_i \neq \lambda_j\,, \\ \phi'_{\lambda_i}(\varepsilon,\lambda_i) \in [0,1] & \text{if } \lambda_i = \lambda_j\,, \end{cases} \quad i,j = 1,\ldots,n\,. \tag{5.30}$$

When $\varepsilon \neq 0$ or $\beta = \emptyset$, the partial derivative of $\Phi(\cdot,\cdot)$ with respect to ε can be computed by

$$\Phi'_\varepsilon(\varepsilon, X) = P\mathrm{diag}(\phi'_\varepsilon(\varepsilon, \lambda_1), \cdots, \phi'_\varepsilon(\varepsilon, \lambda_n))P^T .$$

Thus, $\Phi(\cdot,\cdot)$ is continuously differentiable around $(\varepsilon, X) \in \Re \times \mathcal{S}^n$ if $\varepsilon \neq 0$ or $\beta = \emptyset$. Furthermore, $\Phi(\cdot,\cdot)$ is globally Lipschitz continuous and strongly semismooth at any $(0, X) \in \Re \times \mathcal{S}^n$ [124]. In particular, for any $\varepsilon \downarrow 0$ and $\mathcal{S}^n \ni H \to 0$, it holds that

$$\Phi(\varepsilon, X + H) - \Phi(0, X) - \Phi'(\varepsilon, X + H)(\varepsilon, H) = O(\|(\varepsilon, H)\|^2) . \tag{5.31}$$

Recall that for a locally Lipschitz continuous function Γ from a finite dimensional real Hilbert space $\mathcal{X}$ to $\Re^n$, the B-subdifferential of Γ at $x \in \mathcal{X}$ in the sense of Qi [92] is defined by

$$\partial_B\Gamma(x) := \{V \,|\, V = \lim_{k\to\infty} \Gamma'(x^k),\ x^k \to x,\ x_k \in \mathcal{D}_\Gamma\} ,$$

where $\mathcal{D}_\Gamma$ is the set of points where Γ is Fréchet differentiable. The generalized Jacobian $\partial\Gamma(x)$ of Γ at x in the sense of Clarke [18] is just the convex hull of $\partial_B\Gamma(x)$.

Define $\Phi_{|\beta|} : \Re \times \mathcal{S}^{|\beta|} \to \mathcal{S}^{|\beta|}$ by replacing the dimension n in the definition of $\Phi : \Re \times \mathcal{S}^n \to \mathcal{S}^n$ with $|\beta|$. As the case for $\Phi(\cdot,\cdot)$, the mapping $\Phi_{|\beta|}(\cdot,\cdot)$ is also Lipschitz continuous. Then the B-subdifferentials $\partial_B\Phi(0, X)$ of Φ at $(0, X)$ and $\partial_B\Phi_{|\beta|}(0, Z)$ of $\Phi_{|\beta|}$ at $(0, Z) \in \Re \times \mathcal{S}^{|\beta|}$ in the sense of Qi [92] are both well defined. The following result can be proven similarly as in [15, Proposition 5].

Proposition 5.8. *Suppose that $X \in \mathcal{S}^n$ has the spectral decomposition as in* (5.25)*. Then $V \in \partial_B\Phi(0, X)$ if and only if there exists $V_{|\beta|} \in \partial_B\Phi_{|\beta|}(0, 0)$ such that for all*

$(\varepsilon, H) \in \Re \times \mathcal{S}^n$,

$$V(\varepsilon, H) = P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma} \circ (P_\alpha^T H P_\gamma) \\ (P_\alpha^T H P_\beta)^T & V_{|\beta|}(\varepsilon, P_\beta^T H P_\beta) & 0 \\ (P_\alpha^T H P_\gamma)^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T, \quad (5.32)$$

where $U \in \mathcal{S}^n$ *is defined by*

$$U_{ij} := \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\,\lambda_i\,| + |\,\lambda_j\,|}, \quad i, j \,=\, 1\,, \ldots, n, \qquad (5.33)$$

where $0/0$ *is defined to be 1.*

In order to define smoothing functions for $F(\cdot)$, we need to define smoothing functions for $\Pi_{\mathcal{Q}^*}(\cdot)$. This, however, can be done in many different ways. For simplicity, we shall only use the function ϕ given by (5.27) to define a smoothing function for $\Pi_{\mathcal{Q}^*}(\cdot)$. Let $\psi : \Re \times \Re^m \to \Re^m$ be defined by

$$\psi_i(\varepsilon, z) = \begin{cases} z_i & \text{if } i = 1, \ldots, p, \\ \phi(\varepsilon, z_i) & \text{if } i = p+1, \ldots, m, \end{cases} \qquad (\varepsilon, z) \in \Re \times \Re^m\,. \qquad (5.34)$$

The function ψ is obviously continuously differentiable around any $(\varepsilon, z) \in \Re \times \Re^m$ as long as $\varepsilon \neq 0$ and is strongly semismooth everywhere.

Now, we are ready to define a smoothing function for $F(\cdot)$ itself. Let

$$\Upsilon(\varepsilon, y) := y - \psi\left(\varepsilon, y - (\mathcal{A}\Phi(\varepsilon, C + \mathcal{A}^* y) - b)\right), \quad (\varepsilon, y) \in \Re \times \Re^m\,. \qquad (5.35)$$

By the definitions of Υ, ψ, and Φ, we know that for any $y \in \Re^m$, $F(y) = \Upsilon(0, y)$. We summarize several useful properties of Υ in the next proposition.

Proposition 5.9. *Let* $\Upsilon : \Re \times \Re^m$ *be defined by (5.35). Let* $y \in \Re^m$*. Then it holds that*

(i) Υ *is globally Lipschitz continuous on* $\Re \times \Re^m$*.*

(ii) Υ *is continuously differentiable around* (ε, y) *when* $\varepsilon \neq 0$. *For any fixed* $\varepsilon \in \Re$, $\Upsilon(\varepsilon, \cdot)$ *is a* P_0*-function, i.e., for any* $(y, h) \in \Re^m \times \Re^m$ *with* $y \neq h$,

$$\max_{y_i \neq h_i} (y_i - h_i)(\Upsilon_i(\varepsilon, y) - \Upsilon_i(\varepsilon, h)) \geq 0\,, \tag{5.36}$$

and thus for any fixed $\varepsilon \neq 0$, $\Upsilon'_y(\varepsilon, y)$ *is a* P_0*-matrix.*

(iii) Υ *is strongly semismooth at* $(0, y)$. *In particular, for any* $\varepsilon \downarrow 0$ *and* $\Re^m \ni h \to 0$ *we have*

$$\Upsilon(\varepsilon, y + h) - \Upsilon(0, y) - \Upsilon'(\varepsilon, y + h) \begin{pmatrix} \varepsilon \\ h \end{pmatrix} = O(\|(\varepsilon, h)\|^2)\,.$$

(iv) *For any* $h \in \Re^m$,

$$\partial_B \Upsilon(0, y)(0, h) \subseteq h - \partial_B \psi(0, y - \nabla\theta(y))(0, h - \mathcal{A}\partial_B \Phi(0, C + \mathcal{A}^* y)(0, \mathcal{A}^* h))\,.$$

Proof. (i) Since both ψ and Φ are globally Lipschitz continuous, Υ is also globally Lipschitz continuous.

(ii) From the definitions of ψ and Φ we know that Υ is continuously differentiable around $(\varepsilon, y) \in \Re \times \Re^m$ when $\varepsilon \neq 0$.

Since, by part (i), Υ is continuous on $\Re \times \Re^m$, we only need to show that for any $0 \neq \varepsilon \in \Re$, $\Upsilon(\varepsilon, \cdot)$ is a P_0-function.

Fix $\varepsilon \neq 0$. Define $g_\varepsilon : \Re^m \to \Re^m$ by

$$g_\varepsilon(y) = \mathcal{A}\Phi(\varepsilon, C + \mathcal{A}^* y) - b, \quad y \in \Re^m\,.$$

Then g_ε is continuously differentiable on $\Re^m$. From (5.29) and (5.30), we have

$$\langle h, (g_\varepsilon)'(y)h\rangle = \langle h, \mathcal{A}\Phi'_X(\varepsilon, X)(\mathcal{A}^* h)\rangle = \langle \mathcal{A}^* h, \Phi'_X(\varepsilon, X)(\mathcal{A}^* h)\rangle \geq 0 \quad \forall\, h \in \Re^m\,,$$

which implies that g_ε is a P_0-function on $\Re^m$. Let $(y, h) \in \Re^m \times \Re^m$ with $y \neq h$. Then there exists $i \in \{1, \ldots, m\}$ with $y_i \neq h_i$ such that

$$(y_i - h_i)((g_\varepsilon)_i(y) - (g_\varepsilon)_i(h)) \geq 0\,.$$

Furthermore, by noting that for any $z \in \Re^m$,

$$\phi'_{z_i}(\varepsilon, z_i) \in [0,1], \quad i = 1, \ldots, m\,,$$

we obtain that

$$(y_i - h_i)(\Upsilon_i(\varepsilon, y) - \Upsilon_i(\varepsilon, h)) \geq 0\,.$$

This shows that (5.36) holds. Thus, $\Upsilon'_y(\varepsilon, y)$ is P_0-matrix for any fixed $\varepsilon \neq 0$.

(iii) Since it can be checked directly that the composite of strongly semismooth functions is still strongly semismooth [37], Υ is strongly semismooth at $(0, y)$.

(iv) Since both ψ and Φ are directionally differentiable, for any $(\varepsilon, y') \in \Re \times \Re^m$ such that Υ is Fréchet differentiable at (ε, y'),

$$\Upsilon'(\varepsilon, y')(0, h) = h - \psi'\left((\varepsilon, z'); \left(0, h - \mathcal{A}\Phi'((\varepsilon, C + \mathcal{A}^* y'); (0, \mathcal{A}^* h))\right)\right)\,,$$

which, together with the semismoothness of ψ and Φ, implies

$$\Upsilon'(\varepsilon, y')(0, h) \in h - \partial_B \psi(\varepsilon, z')\big(0, h - \mathcal{A}\partial_B \Phi(\varepsilon, C + \mathcal{A}^* y')(0, \mathcal{A}^* h)\big)\,,$$

where $z' := y' - (\mathcal{A}\Phi(\varepsilon, C + \mathcal{A}^* y') - b)$. By taking $(\varepsilon, y') \to (0, y)$ in the above inclusion, we complete the proof. □

5.2.1 Global and local convergence analysis

In this section, we apply the general inexact smoothing Newton method developed in the last section to the least squares semidefinite programming (5.16).

Let $F : \Re^m \to \Re^m$ be defined by (5.23). Let $\kappa \in (0, \infty)$ be a constant. Define $G : \Re \times \Re^m \to \Re^m$ by

$$G(\varepsilon, y) := \Upsilon(\varepsilon, y) + \kappa|\varepsilon| y\,, \quad (\varepsilon, y) \in \Re \times \Re^m\,, \tag{5.37}$$

where $\Upsilon : \Re \times \Re^m \to \Re^m$ is defined by (5.35). The reason for defining G by (5.37) is that for any $(\varepsilon, y) \in \Re \times \Re^m$ with $\varepsilon \neq 0$, $G'_y(\varepsilon, y)$ is a P-matrix (i.e., all its principal

minors are positive), thus nonsingular while by part (ii) of Proposition 5.9 $\Upsilon'_y(\varepsilon, y)$ is only a P_0-matrix (i.e., all its principal minors are nonnegative), which may be singular.

Let $E : \Re \times \Re^m \to \Re \times \Re^m$ be defined by

$$E(\varepsilon, y) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, y) \end{bmatrix} = \begin{bmatrix} \varepsilon \\ \Upsilon(\varepsilon, y) + \kappa|\varepsilon|y \end{bmatrix}, \quad (\varepsilon, y) \in \Re \times \Re^m . \tag{5.38}$$

Let $\mathcal{N}$ be defined by (5.8). Next, we discuss convergent properties of Algorithm 5.1 when it is applied to solve $E(\varepsilon, y) = 0$.

Theorem 5.10. *Algorithm 5.1 is well defined and generates an infinite sequence* $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ *with the properties that any accumulation point* $(\bar{\varepsilon}, \bar{y})$ *of* $\{(\varepsilon^k, y^k)\}$ *is a solution of* $E(\varepsilon, y) = 0$ *and* $\lim_{k\to\infty} \varphi(\varepsilon^k, y^k) = 0$. *Additionally, if the Slater condition (5.20) holds, then* $\{(\varepsilon^k, y^k)\}$ *is bounded.*

Proof. From part (ii) of Proposition 5.9 and the definitions of G and E we know that for any $(\varepsilon, y) \in \Re_{++} \times \Re^m$, $G'_y(\varepsilon, y)$, and so $E'(\varepsilon, y)$, is a P-matrix. Then from Theorem 5.5 we know that Algorithm 5.1 is well defined and generates an infinite sequence $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ with the property that any accumulation point $(\bar{\varepsilon}, \bar{y})$ of $\{(\varepsilon^k, y^k)\}$ is a solution of $E(\varepsilon, y) = 0$.

Since $\varphi(\varepsilon^k, y^k)$ is a decreasing sequence, $\lim_{k\to\infty} \varphi(\varepsilon^k, y^k)$ exists. Let

$$\bar{\varphi} := \lim_{k\to\infty} \varphi(\varepsilon^k, y^k) \geq 0 .$$

If $\bar{\varphi} > 0$, then there exists an $\varepsilon' > 0$ such that $\varepsilon^k \geq \varepsilon'$ for all $k \geq 0$. For any $\upsilon \geq 0$, let

$$L_\upsilon := \{y \in \Re^m \mid \|\Upsilon(\nu, y) + \kappa\nu y\| \leq \upsilon,\ \nu \in [\varepsilon', \hat{\varepsilon}]\} .$$

Then it is not difficult to prove that for any $\upsilon \geq 0$, L_υ is bounded. In fact, suppose that for some $\upsilon \geq 0$, L_υ is unbounded. Then there exist two sequences $\{z^l\}$ and $\{\nu^l\}$ such that $\lim_{l\to\infty} ||z^l|| = \infty$ and for all $l \geq 1$, $\varepsilon' \leq \nu^l \leq \hat{\varepsilon}$ and

$\|\Upsilon(\nu^l, z^l) + \kappa\nu^l z^l\| \leq \upsilon$. By taking subsequences if necessary, we may assume that $\lim_{l\to\infty} \nu^l = \bar{\nu} \in [\varepsilon', \hat{\varepsilon}]$ and

$$i \in I^{\infty} \cup I^{-\infty} \cup I^{\upsilon} \quad \forall\, i \in \{1, \ldots, m\}\,,$$

where

$$\begin{aligned} I^{\infty} &:= \{i \,|\, \lim_{l\to\infty} z_i^l = \infty,\ i = 1, \ldots, m\}\,, \\ I^{-\infty} &:= \{i \,|\, \lim_{l\to\infty} z_i^l = -\infty,\ i = 1, \ldots, m\}, \quad \text{and} \\ I^{\upsilon} &:= \{i \,|\, \{z_i^l\} \text{ is uniformly bounded},\ i = 1, \ldots, m\}\,. \end{aligned}$$

Then, we have

$$\Upsilon_i(\nu^l, z^l) \to -\infty \quad \forall\, i \in I^{\infty}, \tag{5.39}$$

and

$$\Upsilon_i(\nu^l, z^l) \to \infty \quad \forall\, i \in I^{-\infty}\,. \tag{5.40}$$

For each $l \geq 1$, define $h^l \in \Re^m$ as follows

$$h_i^l = \begin{cases} 0 & \text{if } i \in I^{\infty} \cup I^{-\infty}\,, \\ z_i^l & \text{if } i \in I^{\upsilon}\,, \end{cases} \qquad i = 1, \ldots, m\,.$$

Since, by part (ii) of Proposition 5.9, for any $l \geq 1$, $\Upsilon(\nu^l, \cdot)$ is a P_0-function, by further taking subsequences if necessary, we know that there exists $i \in I^{\infty} \cup I^{-\infty}$ (note that $h_j^l = z_j^l$ for all $j \in I^{\upsilon}$ and $l \geq 1$) such that

$$(z_i^l - h_i^l)(\Upsilon_i(\nu^l, z^l) - \Upsilon_i(\nu^l, h^l)) \geq 0 \quad \forall\, l \geq 1\,,$$

which is impossible in view of (5.39), (5.40), and the fact that $\{\Upsilon(\nu^l, h^l)\}$ is bounded (note that Υ is globally Lipschitz continuous). This shows that for any $\upsilon \geq 0$, L_{υ} is bounded, i.e.,

$$\{y \in \Re^m \,|\, \|G(\varepsilon, y)\| \leq \upsilon,\ \varepsilon \in [\varepsilon', \hat{\varepsilon}]\}$$

is bounded. This implies that $\{(\varepsilon^k, y^k)\}$ is bounded. Thus, $\{(\varepsilon^k, y^k)\}$ has at least one accumulation point, which is a solution of $E(\varepsilon, y) = 0$, contradicting $\bar{\varphi} > 0$. Therefore, $\bar{\varphi} = 0$.

Suppose that the Slater condition (5.20) holds. Then from Proposition 5.7 we know that the solution set of the dual problem is nonempty and compact. Thus, $E(\varepsilon, y) = 0$ also has a nonempty and compact solution set. Since part (ii) of Proposition 5.9 implies that E is a P_0-function, the boundedness of $\{(\varepsilon^k, y^k)\}$ follows directly from [97, Theorem 2.5]. □

Assume that the Slater condition (5.20) holds. Let $(\bar{\varepsilon}, \bar{y})$ be an accumulation point of the infinite sequence $\{(\varepsilon^k, y^k)\}$ generated by Algorithm 5.1. Then, by Theorem 5.10, we know that $\bar{\varepsilon} = 0$ and $F(\bar{y}) = 0$, i.e., $\bar{y} \in \mathcal{Q}^* = \Re^p \times \Re^q_+$ is an optimal solution to the dual problem (5.18). Let $\overline{X} := \Pi_{\mathcal{S}^n_+}(C + \mathcal{A}^*\bar{y})$. By Proposition 5.7 we know that $\overline{X} \in \mathcal{S}^n_+$ is the unique optimal solution to problem (5.16).

For quadratic convergence of Algorithm 5.1, we need the concept of constraint nondegeneracy initiated by Robinson [104] and extensively developed by Bonnans and Shapiro [4]. This concept is a generalization of the well-known linear independence constraint qualification (or LICQ) used in nonlinear programming. For a given closed $K \in \mathcal{X}$, a finite dimensional real Hilbert space, as in convex analysis [105] we use $T_K(x)$ to denote the tangent cone of K at $x \in K$. The largest linear space contained in $T_K(x)$ is denoted by $\mathrm{lin}\big(T_K(x)\big)$. Let $\mathcal{I}$ be the identity mapping from $\mathcal{S}^n$ to $\mathcal{S}^n$. Then the constraint nondegeneracy is said to hold at $\overline{X}$ for the problem (5.16) if

$$\begin{pmatrix} \mathcal{A} \\ \mathcal{I} \end{pmatrix} \mathcal{S}^n + \begin{pmatrix} \mathrm{lin}\big(T_{\mathcal{Q}}(\mathcal{A}(\overline{X}) - b)\big) \\ \mathrm{lin}\big(T_{\mathcal{S}^n_+}(\overline{X})\big) \end{pmatrix} = \begin{pmatrix} \Re^m \\ \mathcal{S}^n \end{pmatrix}, \tag{5.41}$$

where $\mathcal{Q} = \{0\}^p \times \Re^q_+$. Note that the constraint nondegenerate condition (5.41) is called the primal nondegeneracy in [1].

Let $\mathrm{Ind}(\overline{X})$ denote the index set of active constraints at $\overline{X}$:

$$\mathrm{Ind}(\overline{X}) := \{i \,|\, \langle A_i, \overline{X}\rangle = b_i,\ i = p+1, \ldots, m\},$$

and s be the number of elements in $\mathrm{Ind}(\overline{X})$. Without loss of generality, we assume

that

$$\mathrm{Ind}(\overline{X}) = \{p+1, \ldots, p+s\}\,.$$

Define $\widehat{\mathcal{A}} : \mathcal{S}^n \to \Re^{p+s}$ by

$$\widehat{\mathcal{A}}(X) := \begin{bmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_{p+s}, X\rangle \end{bmatrix}, \quad X \in \mathcal{S}^n\,, \tag{5.42}$$

and the adjoint of $\widehat{\mathcal{A}}$ is denoted by $\widehat{\mathcal{A}}^*$.

Lemma 5.11. *Let $X := C + \mathcal{A}^*\bar{y}$ have the spectral decomposition as in (5.25). Then the constraint nondegenerate condition (5.41) holds at $\overline{X}$ if and only if for any $h \in \Re^{p+s}$,*

$$P_\alpha^T \widehat{\mathcal{A}}^* h = 0 \Longleftrightarrow h = 0\,. \tag{5.43}$$

Proof. Since the linearity space $\mathrm{lin}\big(T_{\mathcal{Q}}(\mathcal{A}(\overline{X}) - b)\big)$ in (5.41) can be computed directly as follows

$$\mathrm{lin}\big(T_{\mathcal{Q}}(\mathcal{A}(\overline{X}) - b)\big) = \{h \in \Re^m \,|\, h_i = 0,\ i = 1, \ldots, p,\ i \in \mathrm{Ind}(\overline{X})\}\,, \tag{5.44}$$

we can see that (5.41) is reduced to

$$\begin{pmatrix} \widehat{\mathcal{A}} \\ \mathcal{I} \end{pmatrix} \mathcal{S}^n + \begin{pmatrix} \{0\}^{p+s} \\ \mathrm{lin}\big(T_{\mathcal{S}^n_+}(\overline{X})\big) \end{pmatrix} = \begin{pmatrix} \Re^{p+s} \\ \mathcal{S}^n \end{pmatrix},$$

which is equivalent to

$$\widehat{\mathcal{A}}\big(\mathrm{lin} T_{\mathcal{S}^n_+}(\overline{X})\big) = \Re^{p+s}\,. \tag{5.45}$$

Note that

$$\overline{X} = \Pi_{\mathcal{S}^n_+}(X) = P\mathrm{diag}(\max(0,\lambda_1), \ldots, \max(0,\lambda_n))P^T\,,$$

the tangent cone $T_{\mathcal{S}^n_+}(\overline{X})$, which was first characterized by Arnold [2], takes the form

$$T_{\mathcal{S}^n_+}(\overline{X}) = \{B \in \mathcal{S}^n \,\big|\, [\,P_\beta\ P_\gamma\,]^T B\,[\,P_\beta\ P_\gamma\,] \succeq 0\}\,.$$

Consequently,

$$\mathrm{lin}\big(T_{\mathcal{S}^n_+}(\overline{X})\big) = \big\{B \in \mathcal{S}^n \;\big|\; P_\beta^T B P_\beta = 0,\ P_\beta^T B P_\gamma = 0,\ P_\gamma^T B P_\gamma = 0\big\}\,. \tag{5.46}$$

Thus, from (5.45), the constraint nondegeneracy condition (5.41) holds if and only if (5.43) holds. □

Lemma 5.12. *Let $\Phi : \Re \times \mathcal{S}^n \to \mathcal{S}^n$ be defined by (5.28). Assume that the constraint nondegeneracy (5.41) holds at $\overline{X}$. Then for any $V \in \partial_B \Phi(0, \overline{X})$ we have*

$$\big\langle h, \widehat{\mathcal{A}} V(0, \widehat{\mathcal{A}}^* h)\big\rangle > 0 \quad \forall\, 0 \neq h \in \Re^{p+s}\,. \tag{5.47}$$

Proof. Let $V \in \partial_B \Phi(0, \overline{X})$. Suppose that there exists $0 \neq h \in \Re^{p+s}$ such that (5.47) fails to hold, i.e.,

$$\big\langle h, \widehat{\mathcal{A}} V(0, \widehat{\mathcal{A}}^* h)\big\rangle \leq 0\,.$$

Denote $H := \widehat{\mathcal{A}}^* h$. Then, by Proposition 5.8, there exists $V_{|\beta|} \in \partial_B \Phi_{|\beta|}(0,0)$ such that

$$V(0,H) = P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma} \circ (P_\alpha^T H P_\gamma) \\ (P_\alpha^T H P_\beta)^T & V_{|\beta|}(0, P_\beta^T H P_\beta) & 0 \\ (P_\alpha^T H P_\gamma)^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T\,,$$

where $U \in \mathcal{S}^n$ is defined by (5.33). Since $\langle P_\beta^T H P_\beta, V_{|\beta|}(0, P_\beta^T H P_\beta)\rangle \geq 0$ and $\big\langle h, \widehat{\mathcal{A}} V(0, \widehat{\mathcal{A}}^* h)\big\rangle \leq 0$, we obtain from $\big\langle h, \widehat{\mathcal{A}} V(0, \widehat{\mathcal{A}}^* h)\big\rangle = \langle H, V(0,H)\rangle$ that

$$P_\alpha^T H P_\alpha = 0,\ P_\alpha^T H P_\beta = 0, \text{ and } P_\alpha^T H P_\gamma = 0\,,$$

i.e.,

$$P_\alpha^T H = P_\alpha^T \widehat{\mathcal{A}}^* h = 0\,.$$

On the other hand, since the constraint nondegeneracy (5.41) holds at $\overline{X}$, from (5.43) we know that $h = 0$. This contradiction shows that for any $V \in \partial_B \Phi(0, X)$, (5.47) holds. □

Proposition 5.13. *Let $\Upsilon : \Re \times \Re^m \to \Re^m$ be defined by (5.35). Assume that the constraint nondegeneracy (5.41) holds at $\overline{X}$. Then for any $W \in \partial_B \Upsilon(0, \bar{y})$ we have*

$$\max_i h_i(W(0,h))_i > 0 \quad \forall\, 0 \neq h \in \Re^m\,. \tag{5.48}$$

Proof. Let $W \in \partial_B \Upsilon(0, \bar{y})$. Suppose that there exists $0 \neq h \in \Re^m$ such that (5.48) does not hold, i.e.,

$$\max_i h_i(W(0,h))_i \leq 0\,. \tag{5.49}$$

Then from part (iv) of Proposition 5.9 we know that there exist $D \in \partial_B \psi(0, \bar{z})$ and $V \in \partial_B \Phi(0, \overline{X})$ such that

$$W(0,h) = h - D(0, h - \mathcal{A}V(0, \mathcal{A}^* h)) = h - D(0,h) + D(0, \mathcal{A}V(0, \mathcal{A}^* h))\,, \tag{5.50}$$

where $\bar{z} := \bar{y} - \nabla\theta(\bar{y}) = \bar{y} - (\mathcal{A}\Phi(0, \overline{X}) - b)$. By simple calculations, we can see that there exists a nonnegative vector $d \in \Re^m$ satisfying

$$d_i = \begin{cases} 1 & \text{if } 1 \leq i \leq p\,, \\ \in [0,1] & \text{if } p+1 \leq i \leq p+s\,, \\ 0 & \text{if } p+s+1 \leq i \leq m \end{cases}$$

such that for any $y \in \Re^m$,

$$(D(0,y))_i = d_i y_i,\ i = 1, \ldots, m\,.$$

Thus, we obtain from (5.50) and (5.49) that

$$\begin{cases} h_i(\mathcal{A}V(0, \mathcal{A}^* h))_i \leq 0 & \text{if } 1 \leq i \leq p\,, \\ h_i(\mathcal{A}V(0, \mathcal{A}^* h))_i \leq 0 \text{ or } h_i = 0 & \text{if } p+1 \leq i \leq p+s\,, \\ h_i = 0 & \text{if } p+s+1 \leq i \leq m\,, \end{cases}$$

which, implies

$$\langle h, \mathcal{A}V(0, \mathcal{A}^* h)\rangle = \langle \hat{h}, \widehat{\mathcal{A}}V(0, \widehat{\mathcal{A}}^* \hat{h})\rangle \leq 0\,,$$

where $0 \neq \hat{h} \in \Re^{p+s}$ is defined by $\hat{h}_i = h_i$, $i = 1, \ldots, p+s$. This, however, contradicts (5.47) in Lemma 5.12. This contradiction shows that (5.48) holds. □

Theorem 5.14. *Let $(\bar{\varepsilon}, \bar{y})$ be an accumulation point of the infinite sequence $\{(\varepsilon^k, y^k)\}$ generated by Algorithm 5.1. Assume that the constraint nondegeneracy (5.41) holds at $\overline{X}$. Then the whole sequence $\{(\varepsilon^k, y^k)\}$ converges to $(\bar{\varepsilon}, \bar{y})$ quadratically, i.e.,*

$$\left\|(\varepsilon^{k+1} - \bar{\varepsilon}, y^{k+1} - \bar{y})\right\| = O\big(\|(\varepsilon^k - \bar{\varepsilon}, y^k - \bar{y})\|^2\big)\,. \tag{5.51}$$

Proof. In order to apply Theorem 5.6 to obtain the quadratic convergence of $\{(\varepsilon^k, y^k)\}$, we only need to check that E is strongly semismooth at $(\bar{\varepsilon}, \bar{y})$ and that all $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ are nonsingular.

The strong semismoothness of E at $(\bar{\varepsilon}, \bar{y})$ follows directly from part (iii) of Proposition 5.9 and the fact that the modulus function $|\cdot|$ is strongly semismooth everywhere on $\Re$. The nonsingularity of all matrices in $\partial_B E(\bar{\varepsilon}, \bar{y})$ can be proved as follows.

Let $V \in \partial_B E(\bar{\varepsilon}, \bar{y})$ be arbitrarily chosen. From Proposition 5.13 and the definition of E, we know that for any $0 \neq d \in \Re^{m+1}$,

$$\max_i d_i (Vd)_i > 0\,,$$

which, by [19, Theorem 3.3.4], implies that V is a P-matrix, and so nonsingular. Then the proof is completed. □

Theorem 5.14 says that Algorithm 5.1 can achieve quadratic convergence under the assumption that the constraint nondegenerate condition (5.41) holds at $\overline{X}$. Next, we shall discuss about this assumption by considering the following special least squares semidefinite programming

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - C\|^2 \\ \text{s.t.} & X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \geq l_{ij},\ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \leq u_{ij},\ (i,j) \in \mathcal{B}_u\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{5.52}$$

where $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ are three index subsets of $\{(i,j) \,|\, 1 \le i \le j \le n\}$ satisfying $\mathcal{B}_e \cap \mathcal{B}_l = \emptyset$, $\mathcal{B}_e \cap \mathcal{B}_u = \emptyset$, and $l_{ij} < u_{ij}$ for any $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$. Denote the cardinalities of $\mathcal{B}_e$, $\mathcal{B}_l$, and $\mathcal{B}_u$ by p, q_l, and q_u, respectively. Let $m := p + q_l + q_u$. For any $(i,j) \in \{1,\ldots,n\} \times \{1,\ldots,n\}$, define $\mathcal{E}^{ij} \in \Re^{n\times n}$ by

$$(\mathcal{E}^{ij})_{st} := \begin{cases} 1 & \text{if } (s,t) = (i,j)\,, \\ 0 & \text{otherwise}\,, \end{cases} \qquad s,t = 1,\ldots,n\,.$$

Thus, problem (5.52) can be written as a special case of (5.16) with

$$\mathcal{A}(X) := \begin{bmatrix} \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_e} \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_l} \\ -\{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_u} \end{bmatrix}, \quad X \in \mathcal{S}^n \tag{5.53}$$

and

$$b := \begin{pmatrix} \{e_{ij}\}_{(i,j)\in\mathcal{B}_e} \\ \{l_{ij}\}_{(i,j)\in\mathcal{B}_l} \\ -\{u_{ij}\}_{(i,j)\in\mathcal{B}_u} \end{pmatrix},$$

where $A^{ij} := \frac{1}{2}(\mathcal{E}^{ij} + \mathcal{E}^{ji})$. Then, its dual problem takes the same form as (5.18) with $q := q_l + q_u$. The index set $\mathrm{Ind}(\overline{X})$ of active constraints at $\overline{X}$ now becomes

$$\mathrm{Ind}(\overline{X}) = \widehat{\mathcal{B}}_l \cup \widehat{\mathcal{B}}_u\,,$$

where

$$\widehat{\mathcal{B}}_l := \{\,(i,j) \in \mathcal{B}_l \,|\, \langle\, A^{ij}, \overline{X}\,\rangle = l_{ij}\} \quad \text{and} \quad \widehat{\mathcal{B}}_u := \{\,(i,j) \in \mathcal{B}_u \,|\, \langle\, A^{ij}, \overline{X}\,\rangle = u_{ij}\}\,.$$

Let s be the cardinality of $\mathrm{Ind}(\overline{X})$. Then the mapping $\widehat{\mathcal{A}} : \mathcal{S}^n \to \Re^{p+s}$ defined by (5.42) takes the form

$$\widehat{\mathcal{A}}(X) := \begin{bmatrix} \{\langle A^{ij}, X\rangle\}_{(i,j)\in\mathcal{B}_e} \\ \{\langle A^{ij}, X\rangle\}_{(i,j)\in\widehat{\mathcal{B}}_l} \\ -\{\langle A^{ij}, X\rangle\}_{(i,j)\in\widehat{\mathcal{B}}_u} \end{bmatrix}.$$

Recall that the constraint nondegenerate condition (5.41) holds at $\overline{X}$ if and only if for any $h \in \Re^{p+s}$, (5.43) holds. A particular case for (5.43) to hold is when $\mathcal{B}_e = \{(i,i) \,|\, i = 1,\ldots,n\}$, $\mathcal{B}_l \cup \mathcal{B}_u = \emptyset$, and $b > 0$ [89, 91]. Furthermore, if $\mathcal{B}_e$ has a band structure, (5.43) also holds as long as the corresponding band of the given matrix C is positive definite [91]. In general, the equivalent constraint nondegenerate condition (5.43) may fail to hold for problem (5.52). In [88], Qi establishes an interesting connection between the constraint nondegeneracy and the positive semidefinite matrix completions on chordal graphs.

5.3 Least squares matrix nuclear norm problems

In this section, we shall introduce the least squares matrix nuclear norm programming (LSNNP) and then still apply the general inexact smoothing Newton method to solve it.

Let $\mathcal{A}^e : \Re^{n_1 \times n_2} \to \Re^{m_e}$, $\mathcal{A}^l : \Re^{n_1 \times n_2} \to \Re^{m_l}$ and $\mathcal{A}^q : \Re^{n_1 \times n_2} \to \Re^{m_q}$ be the linear operators defined by

$$\begin{aligned}
\mathcal{A}^e(X) &= [\langle A^e_1, X\rangle; \cdots ; \langle A^e_{m_e}, X\rangle], \\
\mathcal{A}^l(X) &= [\langle A^l_1, X\rangle; \cdots ; \langle A^l_{m_l}, X\rangle], \\
\mathcal{A}^q(X) &= [\langle A^q_1, X\rangle; \cdots ; \langle A^q_{m_q-1}, X\rangle, 0].
\end{aligned}$$

Denote a second order cone by

$$\mathcal{K}^{m_q} := \{\, y \in \Re^{m_q} \,|\, \|y^t\|_2 \le y_{m_q} \,\},$$

where $y = [y_1;\, y_2;\, \cdots ;\, y_{m_q-1};\, y_{m_q}] = [y^t;\, y_{m_q}]$.

Let $\rho \ge 0$ and $\lambda > 0$ be two given numbers and $C \in \Re^{n_1 \times n_2}$ be a given matrix. The least squares matrix nuclear norm problem (LSNNP) then takes the following

form

$$\begin{array}{lll} \min & \dfrac{\lambda}{2}\|X-C\|^2+\rho\|X\|_* & \\ \text{s.t.} & \mathcal{A}^e(X)-b^e = 0, & b^e\in\Re^{m_e}, \\ & \mathcal{A}^l(X)-b^l \geq 0, & b^l\in\Re^{m_l}, \\ & \mathcal{A}^q(X)-b^q \in \mathcal{K}^{m_q}, & b^q\in\Re^{m_q}, \\ & X\in\Re^{n_1\times n_2}. & \end{array} \tag{5.54}$$

Denote $b := [b^e;\, b^l;\, b^q]$ and $Q := \{0\}^{m_e}\times\Re_+^{m_l}\times\mathcal{K}^{m_q}$. Let $m := m_e+m_l+m_q$. Define $\mathcal{A}:\Re^{n_1\times n_2}\to\Re^m$ by $\mathcal{A}=[\mathcal{A}^e;\,\mathcal{A}^l;\,\mathcal{A}^q]$. Then problem (5.54) can be rewritten in the following compact form

$$\begin{array}{ll} \min & f_{\lambda,\rho}(X) := \dfrac{\lambda}{2}\|X-C\|^2+\rho\|X\|_* \\ \text{s.t.} & \mathcal{A}(X)\in b+\mathcal{Q}\,, \\ & X\in\Re^{n_1\times n_2}\,. \end{array} \tag{5.55}$$

5.3.1 The Lagrangian dual problem and optimality conditions

The Lagrangian function $L(X,y):\Re^{n_1\times n_2}\times\Re^m\to\Re$ for problem (5.55) is defined by

$$L(X,y) := f_{\lambda,\rho}(X)-\langle\mathcal{A}(X)-b,\,y\rangle = \frac{\lambda}{2}\|X-C\|^2+\rho\|X\|_*+\langle b-\mathcal{A}(X),y\rangle\,. \tag{5.56}$$

The dual objective function $g(y)$ can be derived from the Lagrangian function (5.56) by

$$\begin{aligned} g(y) &= \inf_{X\in\Re^{n_1\times n_2}} L(X,y) \\ &= \inf_{X\in\Re^{n_1\times n_2}} \Big\{\frac{\lambda}{2}\|X-C\|^2+\rho\|X\|_*+\langle b-\mathcal{A}(X),\, y\rangle\Big\} \\ &= \inf_{X\in\Re^{n_1\times n_2}} \Big\{\frac{\lambda}{2}\big(\|X\|^2-2\langle C+\frac{1}{\lambda}\mathcal{A}^*y, X\rangle+\|C+\frac{1}{\lambda}\mathcal{A}^*y\|^2\big)+\rho\|X\|_* \\ &\qquad -\frac{\lambda}{2}\|C+\frac{1}{\lambda}\mathcal{A}^*y\|^2+\frac{\lambda}{2}\|C\|^2+\langle b,y\rangle\Big\} \\ &= \inf_{X\in\Re^{n_1\times n_2}} \Big\{\frac{\lambda}{2}\|X-(C+\frac{1}{\lambda}\mathcal{A}^*y)\|^2+\rho\|X\|_*-\frac{\lambda}{2}\|C+\frac{1}{\lambda}\mathcal{A}^*y\|^2+\frac{\lambda}{2}\|C\|^2+\langle b,y\rangle\Big\}. \end{aligned}$$

where $\mathcal{A}^* = [(\mathcal{A}^e)^*\ (\mathcal{A}^l)^*\ (\mathcal{A}^q)^*]$ is the adjoint operator of $\mathcal{A}$.

In order to get the infimum of $\frac{\lambda}{2}\|X-C-\frac{1}{\lambda}\mathcal{A}^*y\|^2+\rho\|X\|_*$ in $g(y)$, we need to introduce the singular value thresholding operator $\mathcal{P}_\tau(\cdot)$ for any $\tau>0$. Let $Y\in\Re^{n_1\times n_2}$ have the singular value decomposition (SVD) as in (2.48)

$$Y = U[\Sigma(Y)\ \ \mathbf{0}][V_1\ \ V_2]^T,\quad \Sigma(Y)=\mathrm{diag}(\sigma(Y)),$$

where $\sigma(Y) := \big(\sigma_1(Y),\ldots,\sigma_{n_1}(Y)\big)^T$ are singular values of Y. For any $\tau\geq 0$, $\mathcal{P}_\tau(Y)$ is defined by:

$$\mathcal{P}_\tau(Y) := U[\Sigma_\tau(Y)\ \ \mathbf{0}][V_1\ \ V_2]^T = U\Sigma_\tau(Y)V_1^T,$$

where $\Sigma_\tau(Y)=\mathrm{diag}\big((\sigma_1(Y)-\tau)_+,\ldots,(\sigma_{n_1}(Y)-\tau)_+\big)^T$. The singular value thresholding operator is a proximity operator associated with nuclear norm. Details of proximity operator can be found in [52].

The following proposition[1] allows us to obtain the result of $\inf_X\{\frac{\lambda}{2}\|X-(C+\frac{1}{\lambda}\mathcal{A}^*y)\|^2+\rho\|X\|_*\}$. Its proof can be found in [12, 73].

[1]Donald Goldfard first reported the formula (5.57) at the "Foundations of Computational Mathematics Conference'08" held at the City University of Hong Kong, June 2008

Proposition 5.15. *For each $\tau \geq 0$ and $Y \in \Re^{n_1 \times n_2}$, the singular value thresholding operator obeys*

$$\mathcal{P}_\tau(Y) = \arg\min_X \{\frac{1}{2}\|X - Y\|_F^2 + \tau\|X\|_*\}. \tag{5.57}$$

Proposition 5.15 implies that

$$\begin{aligned} g(y) &= \frac{\lambda}{2}\|\mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*y) - (C + \frac{1}{\lambda}\mathcal{A}^*y)\|^2 + \rho\|\mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*y)\|_* + \\ &\quad -\frac{\lambda}{2}\|C + \frac{1}{\lambda}\mathcal{A}^*y\|^2 + \frac{\lambda}{2}\|C\|^2 + \langle b, y\rangle \\ &= -\frac{\lambda}{2}\|\mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*y)\|^2 + \frac{\lambda}{2}\|C\|^2 + \langle b, y\rangle. \end{aligned}$$

Let

$$\theta(y) := -g(y) + \frac{\lambda}{2}\|C\|^2 = \frac{\lambda}{2}\|\mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*y)\|^2 - \langle b, y\rangle.$$

Then we obtain the dual problem for problem (5.55) is

$$\begin{aligned} \min \quad & \theta(y) \\ \text{s.t} \quad & y \in \mathcal{Q}^* \,. \end{aligned} \tag{5.58}$$

The objective function $\theta(\cdot)$ in the dual problem (5.58) is a continuously differentiable convex function [52]. However it is not twice continuously differentiable. Its gradient is given by

$$\nabla\theta(y) = \mathcal{A}\mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*y) - b\,, \tag{5.59}$$

The dual problem (5.58) of problem (5.55) is a convex constrained vector-valued problem, in contrast to the matrix-valued problem (5.55). When it is easier to apply optimization algorithms to solve for the dual problem (5.58) than for the primal problem (5.55), one can use Rockafellar's dual approach [106] to find an optimal solution $\bar{y}$ for (5.58) first. An optimal solution $\overline{X}$ for (5.55) can then be obtained by

$$\overline{X} = \arg\inf_X L(X, \bar{y}) = \mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*\bar{y})\,.$$

Before introducing optimality conditions, we assume that the Slater condition holds for the primal problem (5.55):

$$\begin{cases} \{A_i\}_{i=1}^{m_e} \text{ are linearly independent,} \\ \exists\, X^0 \text{ such that } \mathcal{A}^l(X^0) > b^l \text{ and } \mathcal{A}^q(X^0) - b^q \in ri(\mathcal{K}^{m_q})\,. \end{cases} \tag{5.60}$$

where $ri(\mathcal{K}^{m_q})$ denotes the relative interior of $\mathcal{K}^{m_q}$. When the Slater condition is satisfied, the following proposition, which is a straightforward application of Rockafellar's results in [106, Theorems 17 & 18], holds.

Proposition 5.16. *Under the Slater condition (5.60), the following results hold:*

(i) There exists at least one $\bar{y} \in \mathcal{Q}^*$ that solves the dual problem (5.58). The unique solution to the primal problem (P) *is given by*

$$\overline{X} = \mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^*\bar{y}). \tag{5.61}$$

(ii) For every real number τ, the constrained level set $\{y \in \mathcal{Q}^* |\, \theta(y) \leq \tau\}$ is closed, bounded and convex.

The convexity in the second part of Proposition 5.16 allows us to apply any gradient based optimization method to obtain an optimal solution for the dual problem (5.58). When a solution is found for (5.58), one can always use (5.61) to obtain a unique optimal solution to the primal problem .

Define $F : \Re^m \to \Re^m$ by

$$F(y) := y - \Pi_{\mathcal{Q}^*}\big(y - \nabla\theta(y)\big), \quad y \in \Re^m\,. \tag{5.62}$$

Then, one can easily check that solving the dual problem (5.58) is equivalent to solving the following equation:

$$F(y) = 0\,, \qquad y \in \Re^m\,. \tag{5.63}$$

It is known that F is globally Lipschitz continuous but not everywhere continuously differentiable. Similarly, we can apply the smoothing Newton-BiCGStab method to solve (5.62).

Recall that

$$F(y) = y - \Pi_{\mathcal{Q}^*}\big(y - (\mathcal{A}\mathcal{P}_{\frac{\rho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^* y) - b)\big)\,.$$

Now we introduce the smoothing functions for the $\Pi_{\mathcal{Q}^*}(\cdot)$ and $\mathcal{P}_{\frac{\rho}{\lambda}}(\cdot)$, respectively. F contains a composition of two nonsmooth functions. In the outer layer, $\Pi_{Q^*}(\cdot)$ is a metric projection operator from $\Re^m$ to Q^*. Recall that $\mathcal{Q}^* = \Re^{m_e} \times \Re^{m_l}_+ \times \mathcal{K}^{m_q}$, then $\Pi_{Q^*}(\cdot)$ is given by

$$\Pi_{\mathcal{Q}^*}(z) = [\, z^e;\ \Pi_{\Re^{m_l}_+}(z^l);\ \Pi_{\mathcal{K}^{m_q}}(z^q)\,]\,, \tag{5.64}$$

where $z = [z^e;\ z^l;\ z^q]$ and $\Pi_{\mathcal{K}^{m_q}}(z)$ denotes the projection of z onto the second-order cone $\mathcal{K}^{m_q}$. The properties of second order cone have been well studied. The following well known proposition gives an analytical solution to $\Pi_{\mathcal{K}^n}(\cdot)$, the metric projection onto a second order cone $\mathcal{K}^n$ of dimension n. See [39, 85] and references therein for more discussions on $\Pi_{\mathcal{K}^n}(\cdot)$.

Proposition 5.17. *For any $z \in \Re^n$, let $z = [z^t;\ z_n]$ where $z^t \in \Re^{n-1}$ and $z_n \in \Re$. Then z has the following spectral decomposition*

$$z = \lambda_1(z)c_1(z) + \lambda_2(z)c_2(z), \tag{5.65}$$

where for $i = 1, 2$,

$$\begin{array}{rcl} \lambda_i(z) & = & z_n + (-1)^i \|z^t\|_2\,, \\ c_i(z) & = & \left\{ \begin{array}{ll} \frac{1}{2}((-1)^i \frac{z^t}{\|z^t\|_2},\, 1)^T & \text{if } z^t \neq 0\,, \\ \frac{1}{2}((-1)^i w,\, 1)^T & \text{if } z^t = 0\,, \end{array} \right. \end{array}$$

where $w \in \Re^{n-1}$ satisfies $\|w\|_2 = 1$. Then $\Pi_{\mathcal{K}^n}(z)$ is given by

$$\Pi_{\mathcal{K}^n}(z) \;=\; (\lambda_1(z))_+ c_1(z) + (\lambda_2(z))_+ c_2(z)\,. \tag{5.66}$$

Now we are ready to introduce a smoothing function $\psi^{soc} : \Re \times \Re^{m_q} \to \Re^{m_q}$ for $\Pi_{\mathcal{K}^{m_q}}(\cdot)$. For any $z^q \in \Re^{m_q}$ which has the spectral decomposition as in (5.65), we define $\psi^{soc} : \Re \times \Re^{m_q} \to \Re^{m_q}$ by

$$\psi^{soc}(\varepsilon, z^q) \;\; = \;\; \phi(\varepsilon, \lambda_1(z^q))c_1(z^q) + \phi(\varepsilon, \lambda_2(z^q))c_2(z^q)\,, \tag{5.67}$$

where $\phi(\cdot,\cdot)$ is the Huber or Smale smoothing function defined as in (2.69) or (2.70). It has been shown in [124, Theorem 5.1] that $\psi^{soc}(\cdot,\cdot)$ is globally Lipschitz continuous and strongly semismooth on $\Re_+ \times \Re^{m_q}$ if the smoothing function ϕ is globally Lipschitz continuous and strongly semismooth on $\Re_+ \times \Re$.

Next we consider the smoothing for $\Pi_{\Re^n_+}(\cdot)$. Define $\psi^{nno} : \Re \times \Re^{m_l} \to \Re^{m_l}$ by

$$\psi^{nno}_i(\varepsilon, z^l) = \phi(\varepsilon, z^l_i) \qquad \text{for } i = 1, \ldots, m_l\,, \quad (\varepsilon, z^l) \in \Re \times \Re^{m_l}\,. \tag{5.68}$$

In order to define smoothing function for $F(\cdot)$, we need to define smoothing function for $\Pi_{\mathcal{Q}^*}(\cdot)$. Let $\psi : \Re \times \Re^m \to \Re^m$ be defined by

$$\psi(\varepsilon, z) = \begin{pmatrix} z^e \\ \psi^{nno}(\varepsilon, z^l) \\ \psi^{soc}(\varepsilon, z^q) \end{pmatrix}. \tag{5.69}$$

It is obvious that ψ is a globally Lipschitz continuous, and strongly semismooth function on $\Re \times \Re^m$. Furthermore, it can be easily checked that for any fixed $\varepsilon \neq 0$, any $t, s \in \Re$ and $t \neq s$,

$$\phi'_t(\varepsilon, t) \in [0, 1] \quad \text{ and } \quad \frac{\phi(\varepsilon, t) - \phi(\varepsilon, s)}{t - s} \in [0, 1], \tag{5.70}$$

thus, together with the result of Korányi [60, Page 74], we know that for any $z \in \Re^m$,

$$\begin{array}{l} \psi'_z(\varepsilon, z) = (\psi'_z(\varepsilon, z))^T, \\ \mathbf{0} \preceq \psi'_z(\varepsilon, z) \preceq I. \end{array} \tag{5.71}$$

Next we will construct a smoothing function for the inner layer on the nonsymmetric matrix operator $\mathcal{P}_{\frac{\rho}{\lambda}}(\cdot)$.

Let $Y \in \Re^{n_1 \times n_2}$ $(n_1 \leq n_2)$. Suppose that Y has the following singular value decomposition as in (2.48), i.e.,

$$Y = U[\Sigma(Y) \;\; \mathbf{0}]V^T = U[\Sigma(Y) \;\; \mathbf{0}][V_1 \;\; V_2]^T. \tag{5.72}$$

Let the orthogonal matrix $Q \in \mathcal{O}^{n_1+n_2}$ be defined by

$$Q := \frac{1}{\sqrt{2}} \begin{pmatrix} U & U & \mathbf{0} \\ V_1 & -V_1 & \sqrt{2}V_2 \end{pmatrix}. \tag{5.73}$$

In order to properly define the smoothing function for nonsymmetric matrix-valued functions, we will transform a nonsymmetric matrix into a symmetric matrix and make use of the known properties of the symmetric matrix-valued functions. Define $\Xi : \Re^{n_1 \times n_2} \to S^{n_1+n_2}$ by

$$\Xi(Y) := \begin{pmatrix} \mathbf{0} & Y \\ Y^T & \mathbf{0} \end{pmatrix}, \qquad Y \in \Re^{n_1 \times n_2}.$$

Then, from [44, Section 8.6], $\Xi(Y)$ has the following spectral decomposition:

$$\Xi(Y) = Q \begin{pmatrix} \Sigma & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} Q^T, \tag{5.74}$$

i.e., the eigenvalues of $\Xi(Y)$ are $\pm\sigma_i(Y)$, $i = 1, \ldots, n_1$, and 0 of multiplicity $n_2 - n_1$. For some $\tau > 0$, we define a real-valued function $g_\tau : \Re \to \Re$ by

$$g_\tau(t) := (t-\tau)_+ - (-t-\tau)_+ = \begin{cases} t-\tau & \text{if } t > \tau \\ 0 & \text{if } -\tau \leq t \leq \tau \\ t+\tau & \text{if } t < -\tau \end{cases}, \quad t \in \mathbb{R}. \tag{5.75}$$

For any $W = Q\mathrm{diag}(\lambda_1, \ldots, \lambda_{n_1+n_2})Q^T \in \mathcal{S}^{n_1+n_2}$, define

$$\begin{aligned} G_\tau(W) &:= Q\mathrm{diag}\big(g_\tau(\lambda_1), \ldots, g_\tau(\lambda_{n_1+n_2})\big)Q^T \\ &= \Pi_{\mathcal{S}^n_+}(W - \tau I) - \Pi_{\mathcal{S}^n_+}(-W - \tau I). \end{aligned} \tag{5.76}$$

By direct calculation, we have

$$G_\tau(\Xi(Y)) = \begin{pmatrix} \mathbf{0} & \mathcal{P}_\tau(Y) \\ \mathcal{P}_\tau(Y)^T & \mathbf{0} \end{pmatrix}. \tag{5.77}$$

The smoothing functions for $g_\tau(\cdot)$ in (5.75) and $G_\tau(\cdot)$ in (5.76) can be defined, respectively, by

$$\phi_{g_\tau}(\varepsilon, t) \;:=\; \phi(\varepsilon, t-\tau) - \phi(\varepsilon, -t-\tau), \tag{5.78}$$

where $\phi(\cdot,\cdot)$ is the Huber or Smale smoothing function defined as in (2.69) or (2.70), and

$$\Phi_{G_\tau}(\varepsilon, \Xi(Y)) \;:=\; Q \begin{bmatrix} \Sigma_{\phi_{g_\tau}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\Sigma_{\phi_{g_\tau}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} Q^T, \tag{5.79}$$

where $\Sigma_{\phi_{g_\tau}} := \mathrm{diag}\big(\phi_{g_\tau}(\varepsilon, \sigma_1), \ldots, \phi_{g_\tau}(\varepsilon, \sigma_{n_1})\big)$.

From (5.77), One can easily derive that Φ_{G_τ} has the following form

$$\Phi_{G_\tau}(\varepsilon, \Xi(Y)) = \begin{pmatrix} \mathbf{0} & \Phi_{\mathcal{P}_\tau}(\varepsilon, Y) \\ (\Phi_{\mathcal{P}_\tau}(\varepsilon, Y))^T & \mathbf{0} \end{pmatrix}, \tag{5.80}$$

where $\Phi_{\mathcal{P}_\tau} : \Re \times \Re^{n_1 \times n_2} \to \Re^{n_1 \times n_2}$ is defined by

$$\Phi_{\mathcal{P}_\tau}(\varepsilon, Y) := U[\Sigma_{\phi_{g_\tau}} \;\; \mathbf{0}]V^T, \tag{5.81}$$

which is the smoothing function for the soft thresholding operator $\mathcal{P}_\tau(\cdot)$. Note that when $\varepsilon = 0$, $\Phi_{G_\tau}(0, \Xi(Y)) = G_\tau(\Xi(Y))$ and $\Phi_{\mathcal{P}_\tau}(0, Y) = \mathcal{P}_\tau(Y)$.

We have known that the smoothing function (5.69) for the outer layer of F in (5.64) is strongly semismooth at $(0, y)$. Next we will show the strong semismoothness of $\Phi_{\mathcal{P}_\tau}$, which is a smoothing function for the inner layer projection of F.

Applying Proposition 2.1, we obtain that when $\varepsilon \neq 0$ or $\sigma_i \neq \tau$, $i = 1, \ldots, n_1$, for any $H \in \Re^{n_1 \times n_2}$,

$$(\Phi_{G_\tau})'_{\Xi(Y)}(\varepsilon, \Xi(Y))\, \Xi(H) = Q\big[\Omega\big(\varepsilon, \lambda(\Xi(Y))\big) \circ \big(Q^T \Xi(H) Q\big)\big] Q^T, \qquad (5.82)$$

where $\Omega(\varepsilon, \lambda(\Xi(Y)))$ is the first divided difference matrix of Φ_{G_τ} at $\lambda(\Xi(Y))$ and

$$\lambda(\Xi(Y)) = (\sigma_1, \ldots, \sigma_{n_1}, -\sigma_1, \ldots, -\sigma_{n_1}, 0, \ldots, 0)^T \in \Re^{n_1+n_2}.$$

One can easily check that $\Omega\big(\varepsilon, \lambda(\Xi(Y))\big) \in \mathcal{S}^{n_1+n_2}$ takes the following form

$$\Omega\big(\varepsilon, \lambda(\Xi(Y))\big) = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{12}^T & \Omega_{22} & \Omega_{23} \\ \Omega_{13}^T & \Omega_{23}^T & \Omega_{33} \end{bmatrix},$$

where

$$[\Omega_{11}]_{ij} = [\Omega(\varepsilon, \lambda)]_{ij} = \begin{cases} \dfrac{\phi_{g_\tau}(\varepsilon, \sigma_i) - \phi_{g_\tau}(\varepsilon, \sigma_j)}{\sigma_i - \sigma_j} & \text{if } \sigma_i \neq \sigma_j \\ (\phi_{g_\tau})'_{\sigma_i}(\varepsilon, \sigma_i) & \text{if } \sigma_i = \sigma_j \end{cases}, \text{ for } i, j = 1, \ldots, n_1,$$

$$[\Omega_{12}]_{ij} = [\Omega(\varepsilon, \lambda)]_{i(j+n_1)} = \begin{cases} \dfrac{\phi_{g_\tau}(\varepsilon, \sigma_i) + \phi_{g_\tau}(\varepsilon, \sigma_j)}{\sigma_i + \sigma_j} & \text{if } \sigma_i \neq 0 \text{ or } \sigma_j \neq 0 \\ (\phi_{g_\tau})'_{\sigma_i}(\varepsilon, \sigma_i) & \text{if } \sigma_i = \sigma_j = 0 \end{cases} \text{ for } i, j = 1, \ldots, n_1,$$

$$[\Omega_{13}]_{ij} = [\Omega(\varepsilon, \lambda)]_{i(j+2n_1)} = \begin{cases} \dfrac{\phi_{g_\tau}(\varepsilon, \sigma_i)}{\sigma_i} & \text{if } \sigma_i \neq 0 \\ (\phi_{g_\tau})'_{\sigma_i}(\varepsilon, \sigma_i) & \text{if } \sigma_i = 0 \end{cases}, \text{ for } i = 1, \ldots, n_1,\ j = 1, \ldots, n_2 - n_1,$$

$$[\Omega_{33}]_{ij} = [\Omega(\varepsilon, \lambda)]_{(i+2n_1)(j+2n_1)} = (\phi_{g_\tau})'_t(\varepsilon, 0), \text{ for } i, j = 1, \ldots, n_2 - n_1,$$

and

$$\Omega_{22} = \Omega_{11}, \quad \Omega_{23} = \Omega_{13}.$$

Note that $\Omega_{11} = \Omega_{11}^T, \Omega_{12} = \Omega_{12}^T$ and $[\Omega(\varepsilon, \lambda)]_{ij} \in [0, 1]$ for all $i, j = 1, \ldots, n_1 + n_2$.

By direct calculation, we can easily show that

$$\begin{aligned}
&(\Phi_{G_\tau})'_{\Xi(Y)}(\varepsilon, \Xi(Y))\,\Xi(H) \\
= \;& Q\Big(\Omega(\varepsilon, \lambda(\Xi(Y))) \circ \frac{1}{2}\begin{pmatrix} U^T & V_1^T \\ U^T & -V_1^T \\ \mathbf{0} & \sqrt{2}V_2^T \end{pmatrix}\begin{pmatrix} \mathbf{0} & H \\ H^T & \mathbf{0} \end{pmatrix}\begin{pmatrix} U & U & \mathbf{0} \\ V_1 & -V_1 & \sqrt{2}V_2 \end{pmatrix}\Big)Q^T \\
= \;& \frac{1}{2}Q\begin{pmatrix} (A^T + A) \circ \Omega_{11} & (A^T - A) \circ \Omega_{12} & \sqrt{2}B \circ \Omega_{13} \\ (A - A^T) \circ \Omega_{21} & -(A^T + A) \circ \Omega_{22} & \sqrt{2}B \circ \Omega_{23} \\ \sqrt{2}B^T \circ \Omega_{31} & \sqrt{2}B^T \circ \Omega_{32} & \mathbf{0} \end{pmatrix}Q^T \\
= \;& \begin{pmatrix} \mathbf{0} & P_{12} \\ P_{12}^T & \mathbf{0} \end{pmatrix},
\end{aligned}$$

where $A = U^T H V_1$, $B = U^T H V_2$ and

$$P_{12} \;=\; \frac{1}{2}U\big((A + A^T) \circ \Omega_{11} + (A - A^T) \circ \Omega_{12}\big)V_1^T + U(B \circ \Omega_{13})V_2^T.$$

When $\varepsilon \neq 0$ or $\sigma_i \neq \tau$, $i = 1, \ldots, n_1$, the partial derivative of $\Phi_{G_\tau}(\cdot, \cdot)$ with respect to ε can be computed by

$$(\Phi_{G_\tau})'_\varepsilon(\varepsilon, Y) = Q\begin{bmatrix} D(\varepsilon, \Sigma) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -D(\varepsilon, \Sigma) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}Q^T, \tag{5.83}$$

where

$$D(\varepsilon, \Sigma) = \operatorname{diag}\big((\phi_{g_\tau})'_\varepsilon(\varepsilon, \sigma_1), \ldots, (\phi_{g_\tau})'_\varepsilon(\varepsilon, \sigma_{n_1})\big)\,. \tag{5.84}$$

Note that

$$(\Phi_{G_\tau})'(\varepsilon, \Xi(Y))(\nu, \Xi(H)) = \begin{bmatrix} \mathbf{0} & (\Phi_{\mathcal{P}_\tau})'(\varepsilon, Y)(\nu, H) \\ \big((\Phi_{\mathcal{P}_\tau})'(\varepsilon, Y)(\nu, H)\big)^T & \mathbf{0} \end{bmatrix},$$

then $\Phi_{\mathcal{P}_\tau}(\cdot, \cdot)$ is continuously differentiable around $(\varepsilon, Y) \in \Re^{n_1 \times n_2}$ if $\varepsilon \neq 0$ or $\sigma_i \neq \tau$, $i = 1, \ldots, n_1$, and its derivative is given by

$$\begin{aligned}
&(\Phi_{\mathcal{P}_\tau})'(\varepsilon, Y)(\nu, H) \\
= \;& U\Big(\Omega_{11} \circ \frac{A + A^T}{2} + \Omega_{12} \circ \frac{A - A^T}{2} + \nu D(\varepsilon, \Sigma)\Big)V_1^T + U(\Omega_{13} \circ B)V_2^T.
\end{aligned} \tag{5.85}$$

Furthermore, $\Phi_{\mathcal{P}_\tau}(\cdot,\cdot)$ is globally Lipschitz continuous and strongly semismooth at any $(0,Y)\in\Re\times\Re^{n_1\times n_2}$ [124]. In particular, for any $\varepsilon\to 0$ and $\Re^{n_1\times n_2}\ni\Delta Y\to\mathbf{0}$ and, it holds that

$$\Phi_{\mathcal{P}_\tau}(\varepsilon,Y+\Delta Y)-\Phi_{\mathcal{P}_\tau}(0,Y)-(\Phi_{\mathcal{P}_\tau})'(\varepsilon,Y+\Delta Y)(\varepsilon,\Delta Y)=\mathcal{O}(\|(\varepsilon,\Delta Y)\|^2). \quad (5.86)$$

Now we are ready to introduce a smoothing function $\Upsilon:\Re\times\Re^m\to\Re^m$ for F defined in (5.62) with (5.69) and (5.81),

$$\Upsilon(\varepsilon,y):=y-\psi\Big(\varepsilon,\ y-\big(\mathcal{A}\Phi_{\mathcal{P}_{\frac{\rho}{\lambda}}}(\varepsilon,C+\frac{1}{\lambda}\mathcal{A}^*y)-b\big)\Big). \quad (5.87)$$

By the definitions of Υ, ψ and $\Phi_{\mathcal{P}_\tau}$, we know that for any $y\in\Re^m$, $F(y)=\Upsilon(0,y)$.

In order to study the properties of Υ, we need the following notations. Let $m_1:=m_e+m_l$. Define

$$\mathcal{K}:=\underbrace{\Re\times\cdots\times\Re}_{m_1}\times\Re^{m_q},$$

and

$$\mathcal{D}:=\underbrace{\Re\times\cdots\times\Re}_{m_1}\times\mathcal{K}^{m_q},$$

where $\mathcal{K}^{m_q}$ denotes the second order cone with dimension m_q as usual. Then $\mathcal{D}\in L(\mathcal{K})$.

Proposition 5.18. *Let $\Upsilon:\Re\times\Re^m$ be defined by (5.87). Let $y\in\Re^m$. Then it holds that*

(i) Υ is globally Lipschitz continuous on $\Re\times\Re^m$.

(ii) Υ is continuously differentiable around (ε,y) where $\varepsilon\neq 0$. For any fixed $\varepsilon\in\Re$, $\Upsilon(\varepsilon,\cdot)$ is a block quasi P_0-function on $\mathcal{D}\in L(\mathcal{K})$, i.e., for any y, h in $\Re^m$ with $y\neq h$, there exists an orthogonal matrix $Q\in\mathcal{O}^m$ taking the form of

$$Q=\begin{bmatrix} I & \mathbf{0}\\ \mathbf{0} & Q^q\end{bmatrix}, \quad (5.88)$$

where $Q^q \in \mathcal{O}^{m_q}$ *such that*

$$\max_{\substack{1\le i\le m_1+1\\ \hat{y}^i\neq\hat{h}^i}} \langle \hat{y}^i - \hat{h}^i,\, \widehat{\Upsilon}^i_y - \widehat{\Upsilon}^i_h\rangle \ge 0\,, \tag{5.89}$$

where

$$\hat{y}^i := \begin{cases} y^i, & i = 1,\ldots,m_1,\\ Q^q y^i, & i = m_1+1,\end{cases}, \quad \hat{h}^i := \begin{cases} h^i, & i = 1,\ldots,m_1,\\ Q^q h^i, & i = m_1+1,\end{cases},$$

$$\widehat{\Upsilon}^i_y := \begin{cases} \Upsilon^i(\varepsilon,y), & i = 1,\ldots,m_1,\\ Q^q\Upsilon^i(\varepsilon,y), & i = m_1+1,\end{cases}, \quad \widehat{\Upsilon}^i_h := \begin{cases} \Upsilon^i(\varepsilon,h), & i = 1,\ldots,m_1,\\ Q^q\Upsilon^i(\varepsilon,h), & i = m_1+1.\end{cases}$$

Furthermore, for any fixed $\varepsilon \neq 0$, $\Upsilon'_y(\varepsilon,y)$ *is a quasi* P_0*-matrix.*

In particular, if $m^q = 0$, *then for any fixed* $\varepsilon \in \Re$, $\Upsilon(\varepsilon,\cdot)$ *is a* P_0*-function, i.e., for any* y, h *in* $\Re^m$ *with* $y \neq h$,

$$\max_{\substack{i=1,\ldots,m\\ y_i\neq h_i}} (y_i - h_i)\big(\Upsilon_i(\varepsilon,y) - \Upsilon_i(\varepsilon,h)\big) \ge 0\,, \tag{5.90}$$

and thus for any fixed $\varepsilon \neq 0$, $\Upsilon'_y(\varepsilon,y)$ *is a* P_0*-matrix.*

(iii) Υ *is strongly semismooth at (0, y). In particular, for any* $\varepsilon \downarrow 0$ *and* $\Re^m \ni h \to 0$ *we have*

$$\Upsilon(\varepsilon, y+h) - \Upsilon(0,y) - \Upsilon'(\varepsilon, y+h)\begin{pmatrix}\varepsilon\\ h\end{pmatrix} = O(\|(\varepsilon,h)\|^2)\,.$$

(iv) For any $h \in \Re^m$,

$$\partial_B\Upsilon(0,y)(0,h) \subseteq h - \partial_B\psi(0, y-\nabla\theta(y))\big(0, h - \frac{1}{\lambda}\mathcal{A}\partial_B\Phi_{\mathcal{P}_{\frac{\rho}{\lambda}}}(0, C+\frac{1}{\lambda}\mathcal{A}^*y)(0,\mathcal{A}^*h)\big). \tag{5.91}$$

Proof. (i) Since both ψ and $\Phi_{\mathcal{P}_{\frac{\rho}{\lambda}}}$ are globally Lipschitz continuous, Υ is also globally Lipschitz continuous.

(ii) From the definitions of ψ and $\Phi_{\mathcal{P}_{\frac{\rho}{\lambda}}}$, we know that Υ is continuously differentiable for any $(\varepsilon, y) \in \Re \times \Re^m$ when $\varepsilon \neq 0$.

We first show that for any $0 \neq \varepsilon \in \Re$, $\Upsilon(\varepsilon, \cdot)$ is a block quasi P_0-function on $\mathcal{D} \in L(\mathcal{K})$. Fix $\varepsilon \neq 0$ and define $g_\varepsilon : \Re^m \to \Re^m$ by

$$g_\varepsilon(y) = \mathcal{A}\Phi_{\mathcal{P}_{\frac{\rho}{\lambda}}}(\varepsilon,\, C + \frac{1}{\lambda}\mathcal{A}^* y) - b, \quad y \in \Re^m.$$

Then g_ε is continuously differentiable and monotone on $\Re^m$ [54].

Applying the classical mean value theorem to $\psi(\varepsilon, \cdot)$, together with the structure of ψ and (5.71), one has that there exists $\mathbf{0} \preceq S \preceq I_m$ taking the form

$$S = \begin{bmatrix} D^{m_1} & \mathbf{0} \\ \mathbf{0} & S^q \end{bmatrix}, \tag{5.92}$$

where $D^{m_1} \in \mathcal{S}^{m_1}$ is a diagonal matrix and $S^q \in \mathcal{S}^{m_q}$, such that for any y, $h \in \Re^n$ with $y \neq h$,

$$\begin{aligned} \Upsilon(\varepsilon, y) - \Upsilon(\varepsilon, h) &= \big(y - \psi(\varepsilon, y - g_\varepsilon(y))\big) - \big(h - \psi(\varepsilon, h - g_\varepsilon(h))\big) \\ &= (y - h) - S\big((y - g_\varepsilon(y)) - (h - g_\varepsilon(h))\big) \\ &= (I - S)(y - h) + S\big(g_\varepsilon(y) - g_\varepsilon(h)\big)\,. \end{aligned}$$

From the structure of S in (5.92), we know that there exists an orthogonal matrix $Q \in \mathcal{O}^m$ taking the form

$$Q = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & Q^q \end{bmatrix},$$

where $Q^q \in \mathcal{O}^{m_q}$ such that $QSQ^T = D$ and $\mathbf{0} \preceq D \preceq I$ is a diagonal matrix. Then it follows that

$$\begin{aligned} \widehat{\Upsilon}_y - \widehat{\Upsilon}_h &:= Q\Upsilon(\varepsilon, y) - Q\Upsilon(\varepsilon, h) \\ &= (I - D)(Qy - Qh) + D\big(Qg_\varepsilon(y) - Qg_\varepsilon(h)\big) \\ &= (I - D)(\hat{y} - \hat{h}) + D(\hat{g}_\varepsilon^y - \hat{g}_\varepsilon^h)\,, \end{aligned}$$

where $\hat{g}_\varepsilon^y := Qg_\varepsilon(y)$ and $\hat{g}_\varepsilon^h := Qg_\varepsilon(h)$. Note that

$$\begin{aligned}\langle \hat{y} - \hat{h},\, \hat{g}_\varepsilon^y - \hat{g}_\varepsilon^h \rangle &= \langle Qy - Qh,\, Qg_\varepsilon(y) - Qg_\varepsilon(h) \rangle \\ &= \langle y - h,\, g_\varepsilon(y) - g_\varepsilon(h) \rangle \geq 0\,,\end{aligned}$$

where the inequality comes from the monotocity of g_ε. Thus, there exists $i \in \{1, \ldots, m_1 + 1\}$ such that $\hat{y}^i \neq \hat{h}^i$ and

$$\langle \hat{y}^i - \hat{h}^i,\, \widehat{\Upsilon}_y^i - \widehat{\Upsilon}_h^i \rangle \geq 0\,,$$

which implies that (5.89) holds for any $\varepsilon^k \neq 0$.

Since Υ is continuous on $\Re \times \Re^m$, in order to show that $\Upsilon(0, \cdot)$ is also a block quasi P_0-function on $\mathcal{D} \in L(\mathcal{K})$, we choose an arbitrary positive sequence $\{\varepsilon^k\}$ such that $\lim_{k \to +\infty} \varepsilon^k = 0$. Since (5.89) holds for all $\varepsilon^k > 0$, we can easily get the conclusion by taking $k \to +\infty$ on both sides of (5.89) and noting that the index set $\{i \,|\, \hat{y}^i \neq \hat{h}^i, i = 1, \ldots, m_1\}$ is independent of k. Thus, $\Upsilon(\varepsilon, \cdot)$ is a block quasi P_0-function on $\mathcal{D} \in L(\mathcal{K})$ for any $\varepsilon \in \Re$.

Next we will show that for any fixed $\varepsilon \neq 0$, $\Upsilon_y'(\varepsilon, y)$ is a quasi P_0-matrix. Fix $\varepsilon \neq 0$. Let $z := y - g_\varepsilon(y)$, $V := \psi_z'(\varepsilon, z)$ and $A := (g_\varepsilon)'(y)$ for any $y \in \Re^m$. By using above arguments, we know that there exists $\widehat{Q} \in \mathcal{O}^m$ which takes the same form as in (5.88), i.e.,

$$\widehat{Q} = \begin{bmatrix} I_{m_1} & \mathbf{0} \\ \mathbf{0} & \widehat{Q}^q \end{bmatrix},$$

where $\widehat{Q}^q \in \mathcal{O}^{m_q}$, such that $\widehat{Q} V \widehat{Q}^T = \widehat{D}$ and $\mathbf{0} \preceq \widehat{D} \preceq I$ is a diagonal matrix. Then, one has

$$\widehat{Q}\Upsilon_y'(\varepsilon, y)\widehat{Q}^T = \widehat{Q}\big(I - V(I - A)\big)\widehat{Q}^T = I - \widehat{D} + \widehat{D}(\widehat{Q}A\widehat{Q}^T)\,.$$

Note that A is a P_0-matrix, so is $\widehat{Q}A\widehat{Q}^T$. Thus, $\widehat{Q}\Upsilon_y'(\varepsilon, y)\widehat{Q}^T$ is also P_0-matrix, which implies that $\Upsilon_y'(\varepsilon, y)$ is a quasi P_0-matrix for any fixed $\varepsilon \neq 0$.

In particular, if there is no second order cone constraint, i.e., $m_q = 0$, note that for any $z \in \Re^m$,

$$\psi'_{z_i}(\varepsilon, z_i) \in [0, 1], \quad i = 1, \ldots, m.$$

Let $y, h \in \Re^m$ with $y \neq h$. Then there exists $i \in \{1, \ldots, m\}$ with $y_i \neq h_i$ such that

$$(y_i - h_i)((g_\varepsilon)_i(y) - (g_\varepsilon)_i(h)) \geq 0.$$

Then we obtain that

$$(y_i - h_i)(\Upsilon_i(\varepsilon, y) - \Upsilon_i(\varepsilon, h)) \geq 0.$$

Thus Υ is a P_0-function and (5.90) holds for any $y, h \in \Re^m$ such that $y \neq h$.

(iii) From the fact that the composite of strongly semismooth functions is still strongly semismooth [37] and that both ψ defined in (5.69) and $\Phi_{\mathcal{P}_{\frac{\varrho}{\lambda}}}$ defined in (5.81) are strongly semismooth at any $(0, y)$, we conclude that Υ is strongly semismooth at $(0, y)$.

(iv) Both ψ and $\Phi_{\mathcal{P}_{\frac{\varrho}{\lambda}}}$ are directionally differentiable. For any $(\varepsilon, y') \in \Re \times \Re^m$ such that Υ is Fréchet differentiable at (ε, y'), we know that

$$\Upsilon'(\varepsilon, y')(0, h) = h - \psi'\big((\varepsilon, z'); (0, h - \frac{1}{\lambda}\mathcal{A}\Phi'_{\mathcal{P}_{\frac{\varrho}{\lambda}}}((\varepsilon, C + \frac{1}{\lambda}\mathcal{A}^* y'); (0, \mathcal{A}^* y)))\big),$$

which, together with semismoothness of ψ and $\Phi_{\mathcal{P}_{\frac{\varrho}{\lambda}}}$, implies

$$\Upsilon'((\varepsilon, y'); (0, h)) \in h - \partial_B \psi((\varepsilon, z')\big(0, h - \frac{1}{\lambda}\mathcal{A}\partial_B \Phi_{\mathcal{P}_{\frac{\varrho}{\lambda}}}(\varepsilon, C + \frac{1}{\lambda}\mathcal{A}^* y')(0, \mathcal{A}^* y)\big),$$

where $z' = y' - \big(\mathcal{A}\Phi_{\mathcal{P}_{\frac{\varrho}{\lambda}}}(\varepsilon, C + \frac{1}{\lambda}\mathcal{A}^* y') - b\big)$. By taking $(\varepsilon, y') \to (0, y)$ in the above inclusion, we complete the proof.

□

5.3.2 Global convergence analysis

Let $F : \Re^m \to \Re^m$ be defined by (5.62). Let $\kappa \in (0, \infty)$ be a constant. Define $G : \Re \times \Re^m \to \Re^m$ by

$$G(\varepsilon, y) := \Upsilon(\varepsilon, y) + \kappa|\varepsilon|y\,, \quad (\varepsilon, y) \in \Re \times \Re^m\,, \tag{5.93}$$

where $\Upsilon : \Re \times \Re^m \to \Re^m$ is defined by (5.87). The reason for defining G by (5.93) is that for any $(\varepsilon, y) \in \Re \times \Re^m$ with $\varepsilon \neq 0$, $G'_y(\varepsilon, y)$ is a quasi P-matrix, thus nonsingular while by part (ii) of Proposition 5.18, $\Upsilon'_y(\varepsilon, y)$ is only a quasi P_0-matrix, which may be singular.

Let $E : \Re \times \Re^m \to \Re \times \Re^m$ be defined by

$$E(\varepsilon, y) := \begin{bmatrix} \varepsilon \\ G(\varepsilon, y) \end{bmatrix} = \begin{bmatrix} \varepsilon \\ \Upsilon(\varepsilon, y) + \kappa|\varepsilon|y \end{bmatrix}, \quad (\varepsilon, y) \in \Re \times \Re^m\,. \tag{5.94}$$

Let $\mathcal{N}$ be defined by (5.8). Next, we discuss convergent properties of Algorithm 5.1 when it is applied to solve $E(\varepsilon, y) = 0$.

Lemma 5.19. *The mapping E defined in (5.94) is weakly univalent.*

Proof. For every positive integer $k \geq 1$, consider the mapping

$$E^k(\varepsilon, y) := E(\varepsilon, y) + \begin{bmatrix} 0 \\ y/k \end{bmatrix} = \begin{bmatrix} \varepsilon \\ G^k(\varepsilon, y) \end{bmatrix}, \quad (\varepsilon, y) \in \Re \times \Re^m\,,$$

where $G^k(\varepsilon, y) := G(\varepsilon, y) + y/k = \Upsilon(\varepsilon, y) + (\kappa\varepsilon + 1/k)y$. It is obvious that E^k is continuous for every k and the sequence $\{E^k\}$ converges to E uniformly on bounded subsets. So, to proof the Lemma, we only need to show that for each k, E^k is one-to-one. Let (ε, y) and $(\hat{\varepsilon}, h)$ be two vectors in $\Re \times \Re^m$ such that $E^k(\varepsilon, y) = E^k(\hat{\varepsilon}, h)$. Thus, $\varepsilon = \hat{\varepsilon}$ and $G^k(\varepsilon, y) = G^k(\varepsilon, h)$. Suppose that $y \neq h$. Since, by part (ii) of Proposition 5.18, $\Upsilon(\varepsilon, \cdot)$ is a block quasi P_0-function on $\mathcal{D} \in L(\mathcal{K})$ for any $\varepsilon \in \Re$, we

obtain that for any $k \geq 1$, there exists $i \in \{1, \ldots, m_1 + 1\}$ and a block orthogonal matrix $Q^k \in \mathcal{O}^m$ taking the form as in (5.88) such that $\hat{y}^i \neq \hat{h}^i$ and

$$0 = \langle \hat{y}^i - \hat{h}^i, (\widehat{G}_y^k)^i - (\widehat{G}_h^k)^i \rangle \geq (\kappa\varepsilon + 1/k)\|\hat{y}^i - \hat{h}^i\|^2 > 0, \quad \forall\, k \geq 1\,,$$

where $\hat{y} := Q^k y$, $\hat{h} := Q^k h$, $\widehat{G}_y^k := Q^k G^k(\varepsilon, y)$ and $\widehat{G}_h^k := Q^k G^k(\varepsilon, h)$. Thus we complete the proof. □

Theorem 5.20. *Algorithm 5.1 is well defined and generates an infinite sequence* $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ *with the properties that any accumulation point* $(\bar{\varepsilon}, \bar{y})$ *of* $\{(\varepsilon^k, y^k)\}$ *is a solution of* $E(\varepsilon, y) = 0$ *and* $\lim_{k\to\infty} \varphi(\varepsilon^k, y^k) = 0$. *Additionally, if the Slater condition (5.60) holds, then* $\{(\varepsilon^k, y^k)\}$ *is bounded.*

Proof. From part (ii) of Proposition 5.18 and the definitions of G and E we know that for any $(\varepsilon, y) \in \Re_{++} \times \Re^m$, $G'_y(\varepsilon, y)$, and so $E'(\varepsilon, y)$, is a quasi P-matrix. Then from Theorem 5.5, we know that Algorithm 5.1 is well defined and generates an infinite sequence $\{(\varepsilon^k, y^k)\} \in \mathcal{N}$ with the property that any accumulation point $(\bar{\varepsilon}, \bar{y})$ of $\{(\varepsilon^k, y^k)\}$ is a solution of $E(\varepsilon, y) = 0$.

Since $\varphi(\varepsilon^k, y^k)$ is a decreasing sequence, $\lim_{k\to\infty} \varphi(\varepsilon^k, y^k)$ exists. Let

$$\bar{\varphi} := \lim_{k\to\infty} \varphi(\varepsilon^k, y^k) \geq 0\,.$$

If $\bar{\varphi} > 0$, then there exists an $\varepsilon' > 0$ such that $\varepsilon^k \geq \varepsilon'$ for all $k \geq 0$. For any $\upsilon \geq 0$, let

$$L_\upsilon := \{y \in \Re^m \,|\, \|\Upsilon(\nu, y) + \kappa\nu y\| \leq \upsilon,\ \nu \in [\varepsilon', \hat{\varepsilon}]\}\,.$$

Then it is not difficult to prove that for any $\upsilon \geq 0$, L_υ is bounded. In fact, suppose that for some $\upsilon \geq 0$, L_υ is unbounded. Then there exist two sequences $\{z^l\}$ and $\{\nu^l\}$ such that $\lim_{l\to\infty} ||z^l|| = \infty$ and for all $l \geq 1$, $\varepsilon' \leq \nu^l \leq \hat{\varepsilon}$ and $\|\Upsilon(\nu^l, z^l) + \kappa\nu^l z^l\| \leq \upsilon$. By taking subsequences if necessary, we may assume that $\lim_{l\to\infty} \nu^l = \bar{\nu} \in [\varepsilon', \hat{\varepsilon}]$ and define an index set by

$$I^\infty := \{i \,|\, \lim_{l\to\infty} \|(z^l)^i\| = \infty,\ i = 1, \ldots, m_1 + 1\}\,.$$

For each $l \geq 1$, define $h^l \in \Re^m$ as follows

$$(h^l)^i = \begin{cases} 0 & \text{if } i \in I^\infty, \\ (z^l)^i & \text{if } i \in \{1, \ldots, m_1 + 1\} \setminus I^\infty. \end{cases} \tag{5.95}$$

Since, by part (ii) of Proposition 5.18, for any $l \geq 1$, $\Upsilon(\nu^l, \cdot)$ is a block quasi P_0-function on $\mathcal{D} \in L(\mathcal{K})$, i.e., there exists $i \in \{1, \ldots, m_1 + 1\}$ and an orthogonal matrix $Q^l \in \mathcal{O}^m$ which takes the form as in (5.88) such that $(\hat{z}^l)^i \neq (\hat{h}^l)^i$ and

$$\left\langle (\hat{z}^l)^i - (\hat{h}^l)^i, (\widehat{G}^l_{z_l})^i - (\widehat{G}^l_{h_l})^i \right\rangle \geq \kappa \nu^l \|(\hat{z}^l)^i - (\hat{h}^l)^i\|^2, \quad \forall\, l \geq 1\,,$$

where $\hat{z}^l := Q^l z^l$, $\hat{h}^l := Q^l h^l$, $\widehat{G}^l_{z_l} := Q^l G^l(\nu^l, z^l)$ and $\widehat{G}^l_{h_l} := Q^l G^l(\nu^l, h^l)$, which fails to hold for all l sufficiently large since $\{\Upsilon(\nu^l, z^l) + \kappa \nu^l z^l\}$ and $\{\Upsilon(\nu^l, h^l) + \kappa \nu^l h^l\}$ are bounded. Thus, for any $\upsilon \geq 0$, L_υ is bounded, i.e.,

$$\{y \in \Re^m \mid \|G(\varepsilon, y)\| \leq \upsilon,\ \varepsilon \in [\varepsilon', \hat{\varepsilon}]\}$$

is bounded. This implies that $\{(\varepsilon^k, y^k)\}$ is bounded. Thus, $\{(\varepsilon^k, y^k)\}$ has at least one accumulation point, which is a solution of $E(\varepsilon, y) = 0$, contradicting $\bar{\varphi} > 0$. Therefore, $\bar{\varphi} = 0$.

Suppose that the Slater condition (5.60) holds. Then from Proposition 5.16 we know that the solution set of the dual problem is nonempty and compact. Thus, $E(\varepsilon, y) = 0$ also has a nonempty and compact solution set.

Since E is weakly univalent from Lemma 5.19, the boundedness of $\{(\varepsilon^k, y^k)\}$ follows directly from [97, Theorem 2.5]. □

Assume that the Slater condition (5.60) holds. Let $(\bar{\varepsilon}, \bar{y})$ be an accumulation point of the infinite sequence $\{(\varepsilon^k, y^k)\}$ generated by Algorithm 5.1. Then, by Theorem 5.20, we know that $\bar{\varepsilon} = 0$ and $F(\bar{y}) = 0$, i.e., $\bar{y} \in \mathcal{Q}^*$ is an optimal solution to the dual problem (5.58). Let $\overline{X} := \mathcal{P}_{\frac{\varrho}{\lambda}}(C + \frac{1}{\lambda}\mathcal{A}^* y)$. By Proposition 5.16, we know that $\overline{X} \in \Re^{n_1 \times n_2}$ is the unique optimal solution to problem (5.55).

5.3.3 Local convergence analysis

Define $h : \Re^{n_1\times n_2} \to \Re$ by $h(X) = \|X\|_*$. Let $\mathcal{K}^{n_1\times n_2}$ be the epigraph of h, i.e.,

$$\mathcal{K}^{n_1\times n_2} \;:=\; \mathrm{epi}\, h = \{\, (X,t) \in \Re^{n_1\times n_2} \times \Re \mid h(X) \le t \,\},$$

which is a close convex cone. Let $\widehat{\mathcal{A}} = (\mathcal{A}\ 0)$. Then problem (5.54) can be reformulated as

$$\begin{array}{ll} \min & \dfrac{\lambda}{2}\|X - C\|^2 + \rho t \\ \text{s.t.} & \widehat{\mathcal{A}}(X,t) \in b + \mathcal{Q}\,, \\ & (X,t) \in \mathcal{K}^{n_1\times n_2}\,. \end{array} \tag{5.96}$$

It is easy to see that if $\overline{X}$ is an optimal solution to problem (5.55) if and only if $(\overline{X}, \bar{t})$ is an optimal solution to (5.96) and $\bar{t} = \|\overline{X}\|_*$.

For quadratic convergence analysis, we need the concept of constraint nondegeneracy. Let $\mathcal{I}$ be an identity mapping from $\Re^{n_1\times n_2} \times \Re$ to $\Re^{n_1\times n_2} \times \Re$. Then the constraint nondegeneracy is said to hold at $(\overline{X}, \bar{t})$ if

$$\begin{pmatrix} \widehat{\mathcal{A}} \\ \mathcal{I} \end{pmatrix} \big(\Re^{n_1\times n_2} \times \Re \big) + \begin{pmatrix} \mathrm{lin}\Big(T_{\mathcal{Q}}\big(\widehat{\mathcal{A}}(\overline{X}, \bar{t}) - b\big)\Big) \\ \mathrm{lin}\big(T_{\mathcal{K}^{n_1\times n_2}}(\overline{X}, \bar{t})\big) \end{pmatrix} = \begin{pmatrix} \Re^m \\ \Re^{n_1\times n_2} \times \Re \end{pmatrix} \tag{5.97}$$

Now we try to characterize $T_{\mathcal{K}^{n_1\times n_2}}(\overline{X}, \bar{t})$ which involves the epigraph of h. Let $X \in \Re^{n_1\times n_2}$ have the singular value decomposition

$$X = U[\Sigma(X)\ \ \mathbf{0}][V_1\ \ V_2]^T,$$

where $\Sigma(X) = \mathrm{diag}(\sigma_1(X), \ldots, \sigma_{n_1}(X))$. Suppose that X is of rank r. Write $U = [U_1\ U_2]$ where $U_1 \in \Re^{n_1\times r}$ consists of the first r columns in U and $U_2 \in \Re^{n_1\times(n_1-r)}$ denotes the remaining part in U. Similarly, V_1 can be partitioned into $V_1 = [V_{11}\ V_{12}]$. For any $H \in \Re^{n_1\times n_2}$, define $g(H) := h'(X; H)$. Noting that $h(X) = \sum_{i=1}^{n_1} \sigma_i(X)$, by the result of Watson [118] about the directional derivative

of the singular values, we obtain that

$$g(H) = \begin{cases} \|H\|_*, & \text{if } \sigma(X) = 0, \\ \langle UV_1^T, H\rangle, & \text{if } \sigma_{\min}(X) > 0, \\ \langle U_1V_{11}^T, H\rangle + \|U_2^T H[V_{12}\ V_2]\|_*, & \text{if } \sigma_{\min}(X) = 0 \text{ and } \sigma_{\max}(X) > 0. \end{cases} \tag{5.98}$$

By using [18, Proposition 2.3.6 & Theorem 2.4.7], we have that

$$T_{\operatorname{epi} h}\big(X, h(X)\big) = \operatorname{epi} h'(X;\cdot)\,.$$

It follows that

$$T_{\mathcal{K}^{n_1\times n_2}}(\overline{X}, h(\overline{X})) = \big\{(H,s) \in \Re^{n_1\times n_2}\times\Re \,|\, \langle U_1V_{11}^T, H\rangle + \|U_2^T H[V_{12}\ V_2]\|_* \le s\big\}\,.$$

Thus, its linearity space is as follows

$$\operatorname{lin}\big(T_{\mathcal{K}^{n_1\times n_2}}(\overline{X}, h(\overline{X}))\big) = \big\{(H,s) \in \Re^{n_1\times n_2}\times\Re \,|\, \langle U_1V_{11}^T, H\rangle = s,\ U_2^T H[V_{12}\,V_2] = 0\big\}\,. \tag{5.99}$$

Under the constraint nondegeneracy condition (5.97), it is possible to prove that all $V \in \partial_B E(0,\bar{y})$ are nonsingular, which implies that the sequence generated by Algorithm 5.1 will converge quadratically to $(0,\bar{y})$ according to Theorem 5.6. Actually, when there is no second order cone constraint, i.e., $m_q = 0$, this has already been proven in [54]. Note that in this case, the constraint nondegeneracy condition (5.97) can be further simplified as follows. Let $\operatorname{Ind}(\overline{X})$ denote the index set of active constraints at $\overline{X}$

$$\operatorname{Ind}(\overline{X}) := \{\, i \,|\, \langle A_i^l, \overline{X}\rangle = b_i^l,\ i = m_e+1,\ldots,m\}\,,$$

and $s = |\operatorname{Ind}(\overline{X})|$. Without loss of generality, we assume that

$$\operatorname{Ind}(\overline{X}) = \{m_e+1,\ldots,m_e+s\}\,.$$

Define $\widetilde{\mathcal{A}} : \Re^{n_1\times n_2} \to \Re^{m_e+s}$ by

$$\widetilde{\mathcal{A}}(X) := \big[\langle A_1^e, X\rangle, \ldots, \langle A_{m_e}^e, X\rangle, \langle A_{m_e+1}^l, X\rangle, \ldots, \langle A_{m_e+s}^l, X\rangle\big]^T, \quad X \in \Re^{n_1\times n_2}\,. \tag{5.100}$$

Let $\overline{\mathcal{A}} = (\widetilde{\mathcal{A}} \;\; 0)$. Then (5.97) can be reduced to

$$\begin{pmatrix} \overline{\mathcal{A}} \\ \mathcal{I} \end{pmatrix} \big(\Re^{n_1 \times n_2} \times \Re \big) + \begin{pmatrix} \{0\}^{m_e+s} \\ \operatorname{lin}\big(T_{\mathcal{K}^{n_1 \times n_2}}(\overline{X}, \bar{t})\big) \end{pmatrix} = \begin{pmatrix} \Re^{m_e+s} \\ \Re^{n_1 \times n_2} \times \Re \end{pmatrix} \tag{5.101}$$

which is equivalent to

$$\overline{\mathcal{A}}\big(\operatorname{lin}\big(T_{\mathcal{K}^{n_1 \times n_2}}(\overline{X}, \bar{t})\big)\big) = \Re^{m_e+s}\,. \tag{5.102}$$

When $m_q \neq 0$, the proof for the nonsingularity of all $V \in \partial_B E(0, \bar{y})$ under the constraint nondegeneracy 5.97 can be done similarly, but its analysis is much more involved. To save some space, we omit the details in this thesis.

Chapter 6

Numerical Results

In this chapter, we conduct some numerical experiments on the SLR-MOPs and report our numerical results for the symmetric SLR-MOPs and the nonsymmetric SLR-MOPs, respectively, in the following two sections.

6.1 Numerical results for the symmetric SLR-MOPs

For the symmetric SLR-MOPs, we consider problem (1.5) introduced in Chapter 1, i.e.,

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} \quad & X_{ii} = 1,\ i = 1, \ldots, n\,, \\ & X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \geq l_{ij},\ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \leq u_{ij},\ (i,j) \in \mathcal{B}_u\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \text{rank}(X) \leq r\,. \end{aligned} \tag{6.1}$$

We have shown that problem (6.1) has many applications among a variety of fields. Here we shall first discuss some existing methods for solving this problem. For this purpose, we start from a simple version of problem (6.1). The so-called rank constrained nearest correlation matrix problem (rank-NCM)

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} \quad & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \operatorname{rank}(X) \leq r \end{aligned} \tag{6.2}$$

has been investigated by many researchers. In [111], Simon gave a comprehensive literature review and summarized thirteen methods for solving the rank-NCM problem (6.2) and its many different variations. Here we will only briefly discuss several methods which are most relevant to our approach to be introduced in this thesis.

We start with mentioning the method of "principal component analysis" (PCA). This method truncates the spectral decomposition of the symmetric matrix C to a positive semidefinite matrix by taking the first r largest eigenvalues of C. Its modified version (mPCA), perhaps firstly introduced by Flurry [38], is to take account of the unit diagonal constraints via a normalization procedure. The mPCA method is very popular in the financial industry due to its simplicity and has been widely implemented by many financial institutions for obtaining a correlation matrix with the required rank. The major drawback of the mPCA approach is that it only produces a non-optimal feasible solution to problem (6.2). Nevertheless, it can be used as a good initial feasible point for other methods of solving the rank-NCM problem. In terms of finding an optimal solution, Zhang and Wu [123] and Wu [121] took an important step by using a Lagrange dual method to solve the rank-NCM problem (6.2) with equal weights, i.e., $H = E$, where E is a symmetric matrix whose entries are all ones. Under the assumptions that the given matrix C

is a valid correlation matrix and the rth and $(r+1)$th eigenvalues (arranged in the non-increasing order in terms of their absolute values) of $C+\text{diag}(\bar{y})$ have different absolute values, where $\bar{y}$ is an optimal solution to the Lagrange dual problem of (6.2) and $\text{diag}(\bar{y})$ is a diagonal matrix whose diagonal is $\bar{y}$, Zhang and Wu [123] provided a way to get a global solution of problem (6.2). This global optimality checking is very rare in non-convex optimization. The Lagrange dual method is effective when the required rank r is large. The next major progress is achieved by Pietersz and Groenen [87] who proposed an innovative row by row alternating majorization method. This method can be applied to problem (6.2) with an arbitrary symmetric nonnegative weight matrix H and is particularly efficient when r is small as its computational cost at each iteration is of the order $O(r^2n^2)$. In [47], Grubisic and Pietersz introduced a geometric programming approach for solving problem (6.2). This approach is applicable to any weight matrix H too, but its numerical performance is not so efficient as the majorization method of Pietersz and Groenen as far as we know. Another well studied method for solving problem (6.2) is the trigonometric parametrization method of Rebonato [98, 99, 100, 101], Brigo [8], Brigo and Mercurio [10] and Rapisarda et al. [96]. In this method, they first decompose $X = RR^T$ with $R \in \Re^{n\times r}$ and then parameterize each row vector of R by trigonometric functions through spherical coordinates. The resulting problem is unconstrained, but highly nonlinear and non-convex. It is not clear to us if the problem can be efficiently solved in practice. The trigonometric parametrization method has been considered earlier for the cases without the rank constraint [72, 101]. A class of alternating direction methods, which are easy to implement, are also well studied by many researchers for solving the rank-NCM problem. For example, Morini and Webber [79] suggested an iterative algorithm called eigenvalue zeroing by iteration (EZI). This algorithm generally does not converge to a stationary point of the rank-NCM problem and cannot be extended to the case

with a general weight matrix H. Very recently, Li and Qi [68] proposed a sequential semismooth Newton method for solving problem (6.2) with $H = E$. They formulate the problem as a bi-affine semidefinite programming and then use an augmented Lagrange method to solve a sequence of least squares problems. This approach can be effective when the required rank r is relatively large.

So far we have seen that unless $r \leq O(\sqrt{n})$ in which case the majorization method of Pietersz and Groenen [87] is an excellent choice, there still lacks an efficient method. Note that problem (6.1) is a generalization of problem (6.2) and for problem (6.1) to have a feasible solution, the required rank r cannot be arbitrarily chosen as in problem (6.2) when m is large. From numerical algorithmic point of view, however, there is no much progress in extending approaches from problem (6.2) to deal with the more challenging problem (6.1). Only recently, Simon and Abell [111] extended the majorization method of Pietersz and Groenen [87] by incorporating some equality constraints of the kind $X_{ij} = 0$. But unlike the case for the simpler problem (6.2), this extension can easily fail even the number of such constraints is not large. The main reason is that the desired monotone decreasing property of the objective function is no longer valid whenever the off-diagonal bounds exist. Under this situation, our proposed approach seems to be the only choice so far.

Next, we address several practical issues in the implementation of the proximal subgradient method to the penalized problem of (6.1).

1. *The choice of the initial point* $X^0 \in \Omega$. Compute d as in (4.23). Let $D = \mathrm{diag}(d)$. We then apply the majorization method alternatively (first fix Z

and then X) to approximately solve

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 + \frac{1}{2}\|H \circ (Z - C)\|^2 + \frac{\rho}{2}\|D^{1/2}(X - Z)D^{1/2}\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,, \\ & \text{rank}(Z) \le r \end{array} \tag{6.3}$$

to obtain a feasible solution, say $(\widetilde{X}, \widetilde{Z})$, where $\rho > 0$ is initially set as 100 and is increased by 10 times at each step. The maximum number of steps is set as 10. Then we set $X^0 := \widetilde{X} \in \Omega$.

2. *The choice of the penalty parameter c.* Let X^* be an optimal solution to the following problem

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} & \mathcal{A}X \in b + \mathcal{Q}\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{6.4}$$

We choose the initial penalty parameter c to be

$$c := \min\big\{1,\, 0.25\big(\theta(X^0) - \theta(X^*)\big)/\max\{1,\, p(X^0) - p(X^*)\}\big\}\,.$$

Thereafter, c is updated as follows: when $|p(X^k)|/\max\{1, r\} > 0.1$, c is increased by 4 times; otherwise, c is increased by 1.4 times. The penalty parameter c will be kept unchanged if $|p(X^k)| \le 10^{-8}$.

3. *The choice of the algorithm for solving the subproblems (4.21).* The success of our approach heavily relies on our ability in solving a sequence of the subproblems of the form (4.21). For this purpose, we use the well tested smoothing Newton-BiCGStab method developed in [42].

4. *The stopping criterion.* We terminate our algorithm if

$$|p(X^k)| \le 10^{-8} \quad \text{and} \quad \frac{|\sqrt{f_c(X^k)} - \sqrt{f_c(X^{k-1})}|}{\max\big(100, \sqrt{f_c(X^{k-1})}\big)} \le 10^{-5}\,.$$

We did our numerical experiments in `MATLAB 7.8.0 (R2009a)` running on a PC Intel (R) Core (TM) 2 of 3.16 GHz CPU each and 2.96 GB of RAM. The testing examples to be reported are given below.

Example 6.1. *Let* $n = 500$ *and the weight matrix* $H = E$*. For* $i, j = 1, \ldots, n$*,* $C_{ij} = 0.5 + 0.5e^{-0.05|i-j|}$*. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*. This matrix* C *is a valid correlation matrix and has been used by a number of authors [8, 68].*

Example 6.2. *Let* $n = 500$ *and the weight matrix* $H = E$*. The matrix* C *is extracted from the correlation matrix which is based on a* $10,000$ *gene micro-array data set obtained from* 256 *drugs treated rat livers; see Natsoulis et al. [80] for details. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*.*

Example 6.3. *Let* $n = 500$*. The matrix* C *is the same as in Example 6.1, i.e.,* $C = 0.5 + 0.5e^{-0.05|i-j|}$ *for* $i, j = 1, \ldots, n$*. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*. The weight matrix* H *is generated in the same way as in [91] such that all its entries are uniformly distributed in* $[0.1, 10]$ *except for* 2×100 *entries in* $[0.01, 100]$*.*

Example 6.4. *Let* $n = 500$*. The matrix* C *is the same as in Example 6.2. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$ *. The weight matrix* H *is generated in the same way as in Example 6.3.*

Example 6.5. *The matrix* C *is an estimated* 943×943 *correlation matrix based on* $100,000$ *ratings for* 1682 *movies by* 943 *users. Due to missing data, the generated matrix* G *is not positive semi-definite [41]. This rating data set can be downloaded from* `http://www.grouplens.org/node/73`*. The index sets are* $\mathcal{B}_e = \mathcal{B}_l = \mathcal{B}_u = \emptyset$*. The weight matrix* H *is provided by T. Fushiki at Institute of Statistical Mathematics, Japan.*

Example 6.6. *The matrix* C *is obtained from the gene data sets with dimension* $n = 1,000$ *as in Example 6.2. The weight matrix* H *is the same as in Example*

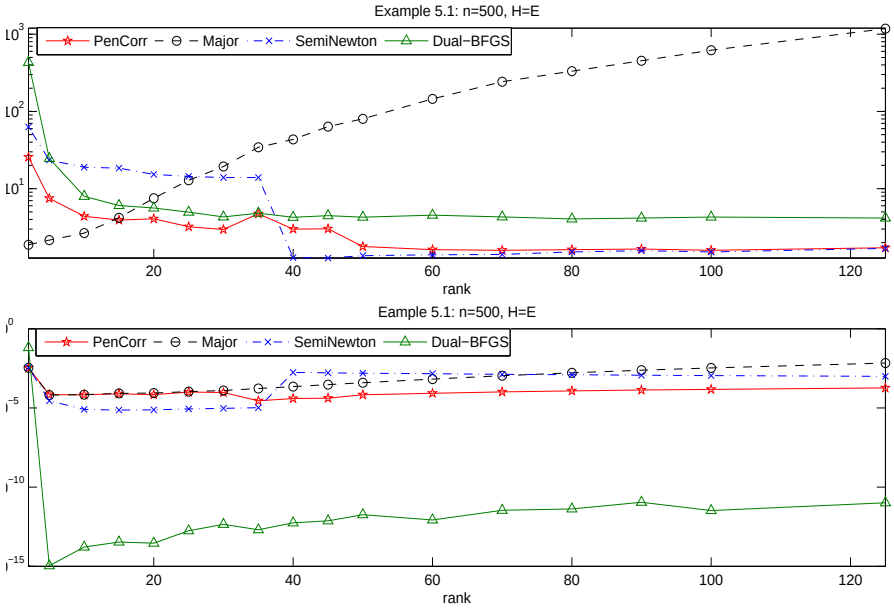

Figure 6.1: Example 6.1

6.3. The index sets $\mathcal{B}_e$, $\mathcal{B}_l$, *and* $\mathcal{B}_u \subset \{(i,j) \,|\, 1 \le i < j \le n\}$ *consist of the indices of* $\min(\hat{n}_r, n-i)$ *randomly generated elements at the* i*th row of* X, $i = 1, \ldots, n$ *with* $\hat{n}_r = 5$ *for* $\mathcal{B}_e$ *and* $\hat{n}_r = 10$ *for* $\mathcal{B}_l$ *and* $\mathcal{B}_u$. *We take* $e_{ij} = 0$ *for* $(i,j) \in \mathcal{B}_e$, $l_{ij} = -0.1$ *for* $(i,j) \in \mathcal{B}_l$ *and* $u_{ij} = 0.1$ *for* $(i,j) \in \mathcal{B}_u$.

Our numerical results are reported in Tables 6.1-6.5, where "*time*" and "*residue*" stand for the total computing time used (in seconds) and the residue $\sqrt{2\theta(X^k)}$ at the final iterate X^k of each algorithm, respectively. For the simplest rank-NCM problem (6.2) of equal weights (i.e., $H = E$), there are many algorithms to choose from. For the purpose of comparison, we only selected three most efficient ones from the literure: the dual approach of Zhang and Wu [123] and Wu [121] (C is required to be a valid correlation matrix), the majorization approach of Pietersz and Groenen [87], and the augmented Lagrangian approach of Li and Qi [68]. For the majorization approach and the augmented Lagrangian approach, we used the

Example 6.1	Major	SemiNewton	Dual-BFGS	PenCorr
rank	*time residue relgap*	*time residue relgap*	*time residue relgap*	*time residue relgap*
2	1.9 1.564e2 3.4e-3	63.0 1.564e2 3.5e-3	432.0 1.660e2 6.5e-2	25.7 1.564e2 3.4e-3
5	2.2 7.883e1 6.5e-5	23.5 7.883e1 2.8e-5	24.6 7.883e1 1.1e-15	7.5 7.883e1 7.0e-5
10	2.7 3.869e1 6.9e-5	19.0 3.868e1 8.0e-6	8.0 3.868e1 1.7e-14	4.4 3.869e1 6.7e-5
15	4.2 2.325e1 8.3e-5	18.5 2.324e1 7.3e-6	6.0 2.324e1 3.4e-14	3.9 2.325e1 7.9e-5
20	7.5 1.571e1 8.8e-5	15.3 1.571e1 7.6e-6	5.6 1.571e1 2.9e-14	4.1 1.571e1 6.9e-5
25	12.8 1.145e1 1.1e-4	14.4 1.145e1 8.6e-6	5.0 1.145e1 1.8e-13	3.2 1.145e1 1.0e-4
30	19.4 8.797e0 1.3e-4	14.0 8.796e0 9.5e-6	4.3 8.795e0 4.4e-13	3.0 8.796e0 9.4e-5
35	34.4 7.020e0 1.7e-4	14.0 7.019e0 1.0e-5	4.8 7.019e0 2.0e-13	4.7 7.019e0 2.8e-5
40	43.4 5.766e0 2.2e-4	1.3 5.774e0 1.7e-3	4.3 5.764e0 5.6e-13	3.0 5.765e0 3.9e-5
45	63.6 4.843e0 3.0e-4	1.3 4.849e0 1.6e-3	4.5 4.841e0 7.4e-13	3.0 4.841e0 4.2e-5
50	80.1 4.141e0 4.0e-4	1.4 4.146e0 1.6e-3	4.3 4.139e0 1.8e-12	1.8 4.139e0 6.8e-5
60	145.0 3.156e0 6.7e-4	1.4 3.158e0 1.4e-3	4.5 3.153e0 8.4e-13	1.6 3.154e0 8.4e-5
70	243.0 2.507e0 1.1e-3	1.4 2.507e0 1.3e-3	4.3 2.504e0 3.4e-12	1.6 2.504e0 1.0e-4
80	333.0 2.053e0 1.6e-3	1.5 2.052e0 1.2e-3	4.1 2.050e0 4.2e-12	1.6 2.050e0 1.2e-4
90	452.0 1.722e0 2.4e-3	1.6 1.720e0 1.2e-3	4.2 1.718e0 1.1e-11	1.7 1.718e0 1.4e-4
100	620.0 1.471e0 3.3e-3	1.5 1.468e0 1.1e-3	4.3 1.467e0 3.3e-12	1.6 1.467e0 1.5e-4
125	1180.0 1.055e0 6.8e-3	1.7 1.049e0 9.9e-4	4.2 1.048e0 1.0e-11	1.7 1.048e0 1.8e-4

Table 6.1: Numerical results for Example 6.1 with $C \in \mathcal{S}^{500}$

Example 6.2	Major			SemiNewton			Dual-BFGS			PenCorr		
rank	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*	*time*	*residue*	*relgap*
2	0.6	2.858e2	6.5e-4	54.4	2.860e2	1.5e-3	304.5	2.862e2	2.1e-3	37.2	2.859e2	8.2e-4
5	6.0	1.350e2	2.0e-3	38.2	1.358e2	8.1e-3	78.8	1.367e2	1.5e-2	99.2	1.351e2	2.4e-3
10	9.3	6.716e1	4.4e-4	32.7	6.735e1	3.2e-3	58.3	6.802e1	1.3e-2	32.1	6.719e1	9.7e-4
15	8.8	4.097e1	3.4e-4	26.8	4.100e1	1.0e-3	44.6	4.096e1	1.0e-4	18.4	4.099e1	7.5e-4
20	13.0	2.842e1	7.3e-4	18.8	2.844e1	1.4e-3	40.4	2.842e1	8.9e-4	16.6	2.843e1	1.1e-3
25	34.9	2.149e1	1.2e-3	18.0	2.152e1	2.6e-3	26.6	2.149e1	1.2e-3	16.4	2.151e1	2.2e-3
30	33.7	1.693e1	4.3e-4	17.3	1.695e1	1.7e-3	23.0	1.694e1	7.8e-4	14.5	1.694e1	1.2e-3
35	71.8	1.379e1	1.3e-3	18.1	1.381e1	2.6e-3	19.7	1.378e1	7.1e-4	11.9	1.379e1	1.6e-3
40	50.0	1.151e1	1.5e-3	12.5	1.152e1	2.1e-3	34.7	1.145e1	3.2e-4	7.7	1.151e1	1.6e-3
45	43.3	9.733e0	9.6e-4	10.6	9.736e0	1.3e-3	23.1	9.733e0	9.2e-4	6.3	9.733e0	1.0e-3
50	44.5	8.318e0	4.1e-4	10.7	8.319e0	4.8e-4	19.7	8.315e0	5.1e-6	5.7	8.318e0	4.5e-4
60	66.5	6.214e0	8.1e-4	10.9	6.214e0	7.4e-4	6.1	6.209e0	1.4e-13	6.9	6.213e0	5.9e-4
70	91.2	4.733e0	1.1e-3	11.0	4.731e0	8.2e-4	23.1	4.728e0	1.9e-4	4.6	4.731e0	7.2e-4
80	93.0	3.663e0	8.7e-4	2.2	3.800e0	3.8e-2	5.2	3.660e0	4.0e-13	2.9	3.662e0	4.5e-4
90	125.0	2.865e0	1.2e-3	2.0	2.962e0	3.5e-2	5.0	2.862e0	5.1e-13	3.0	2.864e0	7.0e-4
100	150.0	2.255e0	1.4e-3	1.7	2.323e0	3.2e-2	15.1	2.254e0	7.8e-4	2.9	2.254e0	8.3e-4
125	288.6	1.269e0	2.4e-3	1.4	1.304e0	3.0e-2	17.1	1.266e0	1.6e-4	2.7	1.268e0	1.4e-3

Table 6.2: Numerical results for Example 6.2 with $C \in \mathcal{S}^{500}$

	Example 6.3				Example 6.4			
	Majorw		PenCorr		Majorw		PenCorr	
rank	*time*	*residue*	*time*	*residue*	*time*	*residue*	*time*	*residue*
2	8.8	1.805e2	81.2	1.804e2	2.9	3.274e2	141.6	3.277e2
5	27.0	8.984e1	70.0	8.986e1	34.4	1.523e2	245.0	1.522e2
10	38.7	4.382e1	48.7	4.383e1	48.5	7.423e1	98.7	7.428e1
15	55.5	2.616e1	43.7	2.618e1	70.5	4.442e1	79.9	4.446e1
20	84.4	1.751e1	39.1	1.753e1	101.4	2.985e1	67.0	2.987e1
25	117.0	1.265e1	38.2	1.266e1	289.6	2.197e1	69.8	2.204e1
30	171.8	9.657e0	36.5	9.657e0	335.6	1.694e1	65.8	1.699e1
35	250.6	7.639e0	39.8	7.632e0	436.7	1.345e1	71.0	1.343e1
40	324.7	6.213e0	38.8	6.203e0	470.7	1.098e1	50.5	1.098e1
45	408.4	5.169e0	38.4	5.148e0	498.7	9.104e0	47.7	9.094e0
50	502.2	4.391e0	37.5	4.355e0	639.5	7.625e0	48.0	7.623e0
60	654.1	3.290e0	35.6	3.219e0	837.6	5.552e0	44.0	5.523e0
70	972.5	2.579e0	38.2	2.481e0	987.5	4.135e0	44.9	4.084e0
80	1274.9	2.090e0	42.6	1.959e0	1212.0	3.127e0	38.0	3.082e0
90	1526.9	1.740e0	44.0	1.588e0	1417.0	2.393e0	35.6	2.345e0
100	1713.7	1.478e0	40.9	1.310e0	1612.0	1.865e0	32.7	1.814e0
125	2438.1	1.052e0	44.6	8.591e-1	1873.0	1.030e0	27.7	9.748e-1

Table 6.3: Numerical results for Examples 6.3 and 6.4 with $C \in \mathcal{S}^{500}$

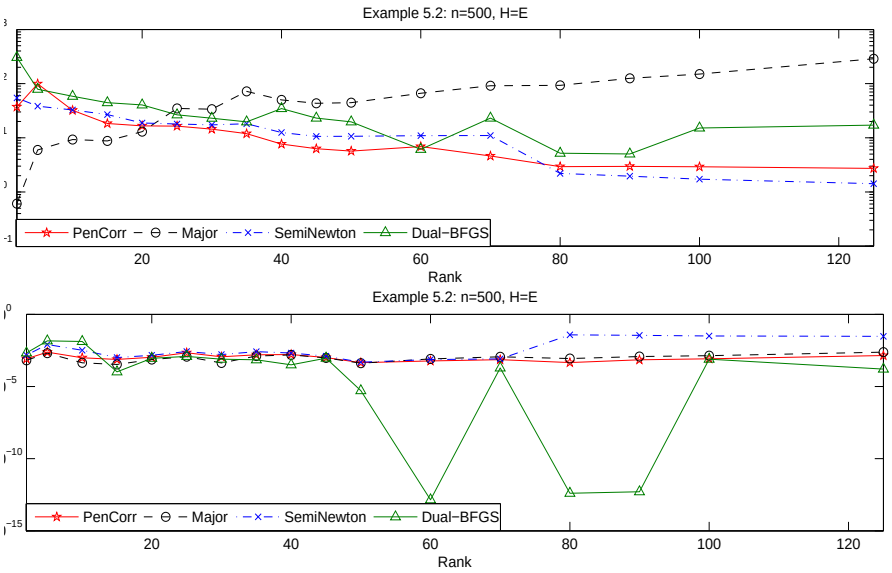

Figure 6.2: Example 6.2

Example 6.5	Majorw		PenCorr	
rank	*time*	*residue*	*time*	*residue*
5	233.4	5.242e2	1534.9	5.273e2
10	706.5	3.485e2	1634.6	3.509e2
20	926.7	2.389e2	1430.2	2.398e2
50	2020.1	1.706e2	829.9	1.709e2
100	3174.3	1.609e2	537.5	1.611e2
150	3890.6	1.608e2	687.1	1.610e2
250	7622.5	1.608e2	694.2	1.610e2

Table 6.4: Numerical results for Example 6.5 with $C \in \mathcal{S}^{943}$

Example 6.6	PenCorr	
rank	*time*	*residue*
20	11640.0	1.872e2
50	1570.0	1.011e2
100	899.0	8.068e1
250	318.3	7.574e1
500	326.3	7.574e1

Table 6.5: Numerical results for Example 6.6 with $C \in \mathcal{S}^{1000}$

codes developed by the authors of [87] and [68]. They are referred to as `Major`[1] and `SemiNewton`, respectively, in Examples 6.1 and 6.2. For the dual approach of [123, 121], we used the BFGS implementation of Lewis and Overton [65] to solve the Lagrangian dual problem. This is denoted by `Dual-BFGS`. The `Dual-BFGS` solves the Lagrangian dual problem to get an approximate optimal dual solution y^k. This approximate optimal dual solution may not always be able to generate an optimal solution to the primal problem as the rth and $(r+1)$th eigenvalues (arranged in the non-increasing order in terms of their absolute values) of $C + \mathrm{diag}(y^k)$ may be of the same absolute values, but it does provide a valid lower bound for the optimal value of the primal problem. The final iterate of the `Dual-BFGS` is obtained by applying the modified PCA procedure to $C + \mathrm{diag}(y^k)$. Our own code is indicated by `PenCorr`. In Tables 6.1-6.2, "*relgap*" denotes the relative gap which is computed as

$$relgap := \frac{residue - \text{ lower bound}}{\max\{1, \text{ lower bound}\}},$$

where the lower bound is obtained by the `Dual-BFGS`. This "*relgap*" indicates the worst possible relative error from the global optimal value.

[1] `Majorw` is the corresponding code for solving the weighted cases.

From Tables 6.1-6.2, we can see that even for the simplest rank-NCM problem (6.2) of equal weights (i.e., $H = E$), `PenCorr` is quite competitive in terms of computing time and solution quality except for small rank cases that `Major` is a clear winner. Examples 6.3, 6.4, and 6.5 belong to the rank-NCM problem (6.2) of general weights. For these three examples, we can see clearly from Tables 6.3-6.4 that `Majorw` performs better than `PenCorr` when the ranks are not large and loses its competitiveness quickly to `PenCorr` as the rank increases. When there are constraints on the off-diagonal parts as in Example 6.6, `PenCorr` seems to be the only viable approach.

6.2 Numerical results for the nonsymmetric SLR-MOPs

To conduct the numerical experiments on the nonsymmetric SLR-MOPs, we consider the following problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|H \circ (X - C)\|^2 \\ \text{s.t.} \quad & X_{ij} = e_{ij},\ (i,j) \in \mathcal{B}_e\,, \\ & X_{ij} \geq l_{ij},\ (i,j) \in \mathcal{B}_l\,, \\ & X_{ij} \leq u_{ij},\ (i,j) \in \mathcal{B}_u\,, \\ & \mathrm{rank}(X) \leq r\,. \end{aligned} \tag{6.5}$$

Notice that problem (6.5) is a special problem of (4.1) with $\rho = 0$ [The case that $\rho > 0$ is not reported here because its performance is similar to the case that $\rho = 0$]. In our implementation, the initial point X^0, the initial penalty parameter c, and the termination criterion are chosen in the same way as in the symmetric SLR-MOPs. We did our numerical experiments in `MATLAB 7.8.0 (R2009a)` running on a PC. The testing examples to be reported are given below.

Example 6.7. *Let* $n_1 = 300$ *and* $n_2 = 500$. *The matrix* C *is a randomly generated* $n_1 \times n_2$ *matrix with entries in* $[-1, 1]$ *and the weight matrix* H *is generated in the same way as in [91] such that all its entries are uniformly distributed in* $[0.1, 10]$ *except for* 2×100 *entries in* $[0.01, 100]$. *The index sets* $\mathcal{B}_e$, $\mathcal{B}_l$, *and* $\mathcal{B}_u \subset \{(i, j) \,|\, 1 \leq i < j \leq n_1\}$ *consist of the indices of* $\min(n_r, n_1 - i)$ *randomly generated elements at the* i*th row of* X, $i = 1, \ldots, n_1$ *with* $n_r = 1$ *for* $\mathcal{B}_e$ *and* $n_r = 2$ *for* $\mathcal{B}_l$ *and* $\mathcal{B}_u$. *We take* $e_{ij} = 0$ *for* $(i, j) \in \mathcal{B}_e$, $l_{ij} = -0.1$ *for* $(i, j) \in \mathcal{B}_l$ *and* $u_{ij} = 0.1$ *for* $(i, j) \in \mathcal{B}_u$.

Example 6.8. *Let* $n_1 = 300$ *and* $n_2 = 500$. *The matrix* $C \in \Re^{n_1\times n_2}$ *and three index sets are generated in the same way as in Example 6.7. The weight matrix* H *is extracted from the matrix provided by T. Fushiki at Institute of Statistical Mathematics, Japan. We still take* $e_{ij} = 0$ *for* $(i, j) \in \mathcal{B}_e$, $l_{ij} = -0.1$ *for* $(i, j) \in \mathcal{B}_l$ *and* $u_{ij} = 0.1$ *for* $(i, j) \in \mathcal{B}_u$.

Example 6.9. *Let* $n_1 = 500$ *and* $n_2 = 1,000$. *The matrices* C *and* H *are generated in the same way as in Example 6.7. The index sets* $\mathcal{B}_e$, $\mathcal{B}_l$, *and* $\mathcal{B}_u$ *are generated in the same way as in Example 6.7 with* $n_r = 2$ *for* $\mathcal{B}_e$ *and* $n_r = 5$ *for* $\mathcal{B}_l$ *and* $\mathcal{B}_u$. *Again, we take* $e_{ij} = 0$ *for* $(i, j) \in \mathcal{B}_e$, $l_{ij} = -0.1$ *for* $(i, j) \in \mathcal{B}_l$ *and* $u_{ij} = 0.1$ *for* $(i, j) \in \mathcal{B}_u$.

Our numerical results are reported in Tables 6.6 and 6.7, where "*time*" and "*residue*" stand for the total computing time used (in seconds) and the residue $\sqrt{2\theta(X^k)}$ at the final iterate X^k, respectively. Tables 6.6 and 6.7 show that our approach also performs well for the nonsymmetric SLR-MOPs.

PenCorr	Example 6.7		Example 6.8	
rank	*time*	*residue*	*time*	*residue*
5	3370.3	1.249e3	2886.2	7.986e3
10	1241.5	1.195e3	2699.5	7.106e3
15	1130.0	1.144e3	1729.5	6.469e3
30	852.0	1.004e3	2084.8	5.015e3
50	579.3	8.390e2	2190.9	3.683e3
100	943.7	5.183e2	1615.9	1.846e3

Table 6.6: Numerical results for Examples 6.7 and 6.8 with $C \in \Re^{300\times 500}$

PenCorr	Example 6.9	
rank	*time*	*residue*
15	12451.3	2.213e3
25	7248.9	2.109e3
50	3561.0	1.867e3
100	2305.7	1.450e3

Table 6.7: Numerical results for Example 6.9 with $C \in \Re^{500\times 1000}$

Chapter 7

Conclusions

In this thesis, we studied the structured low rank matrix optimization problems (SLR-MOPs) which concern the construction of the nearest approximation to a given matrix by another matrix with a specific linear structure and a rank no more than a specified number. This approximation is needed in many important applications arising from a wide range of fields. The SLR-MOPs are known to be non-convex and NP-hard. Thus we proposed a penalty approach for solving the structured low rank matrix problems of the general form (4.1), i.e., absorbing the non-convex rank constraint into the objective function via a penalty technique by using the fact that for any $X \in \Re^{n_1 \times n_2}$, $\text{rank}(X) \leq r$ if and only if $\sigma_{r+1}(X) + \ldots + \sigma_{n_1}(X) = 0$. We further proved that an ε-optimal solution to the original problem is guaranteed by solving the penalized problem as long as the penalty parameter c is above some ε-dependent number which provides some rationale for using this penalty technique. In order to solve the related penalized problem, we presented a framework of proximal subgradient method and further proposed a smoothing Newton-BiCGStab method to solve the resulting sequence of least squares nuclear norm problems which are recently well studied. Interestingly, we also extended the globalization checking results of Zhang and Wu [123, Theorem 4.5] to deal with

more realistic problems. These results are very rare for non-convex optimization problems. Numerical results indicate that our approach is able to handle both the rank and the linear constraints effectively, in particular in the situations when the rank is not very small.

Our approach has paved a new way to deal with the structured low rank matrix optimization problems by solving a sequence of least squares nuclear norm problems. We believe that it represents a good progress for the non-convex low rank matrix approximation problems.

There are still many unanswered questions whose solutions will introduce further development on rank constrained matrix optimization problems. Here we list some of them:

Q1. Is it possible to accelerate our proximal subgradient method as for the case in the accelerated proximal gradient method for convex problems?

Q2. How to further improve the efficiency of the smoothing Newton-BiCGStab method when there are a large number of constraints in the primal problem?

Q3. How to deal with other matrix norms such as the spectral norm and the maximum norm?

Q4. Numerically, though in order to make problem (6.1) feasible, one cannot ask the rank to be very small when there are a large number of bound constraints, it is still interesting to know if one can design a more efficient method to solve problem (6.1) with a small rank and a small number of bound constraints.

Bibliography

[1] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, *Complementarity and nondegenracy in semidefinite programming*, Mathematical Programming 77 (1997), pp. 111-128.

[2] V. I. Arnold, *On matrices depending on parameters*, Russian Mathematical Surveys 26 (1971), pp. 29–43.

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

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[5] B. Borchers and J. G. Young, *Implementation of a primal-dual method for SDP on a shared memory parallel architecture*, Computational Optimization and Applications 37 (2007), pp. 355-369.

[6] I. Borg and P. Groenen, *Modern Multidimensional Scaling*, Springer, New York, 1997.

[7] S. Boyd and L. Xiao, *Least-squares covariance matrix adjustment*, SIAM Journal on Matrix Analysis and Applications 27 (2005), pp. 532-546.

[8] D. Brigo, *A note on correlation and rank reduction*, working paper, 2002. Downloadable from `http://www.damianobrigo.it`.

[9] D. Brigo and F. Mercurio, *Interest rate models: theory and practice*, Springer-Verlag, Berlin, 2001.

[10] D. Brigo and F. Mercurio, *Calibrating LIBOR*, Risk Magazine 15 (2002), pp. 117–122.

[11] J. P. Burge, D. G. Luenberger and D. L. Wenger, *Estimation of structured covariance matrices*, Proceedings of the IEEE 70 (1982), pp. 963–974.

[12] J.-F. Cai, E. J. Candès and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization 20 (2010), pp. 1956–1982.

[13] E.J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics 9 (2009), pp. 717–772.

[14] E.J. Candès and T. Tao, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Transaction on Information Theory 56 (2010), pp. 2053–2080.

[15] Z. X. Chan and D. F. Sun, *Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming*, SIAM Journal on Optimization 19 (2008), pp. 370–396.

[16] Y. D. Chen, Y. Gao, and Y.-J. Liu, *An inexact SQP Newton method for convex SC^1 minimization problems*, to appear in Journal of Optimization Theory and Applications, 2010.

[17] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, *Structured low rank approximation*, Linear Algebra and its Applications 366 (2003), pp. 157–172.

[18] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[19] R. W. Cottle, J.-S. Pang and R. E. Stone, *The Linear Complementarity Problem*, Aacdemic Press, Boston, 1992.

[20] G. Cybenko, *Moment problems and low rank Toeplitz approximatons*, Circuits, Systems, and Signal Processing 1 (1983), pp. 245–366.

[21] A. d'Aspremont, *Interest rate model calibration using semidefinite programming*, Applied Mathematical Finance 10 (2003), pp. 183-213.

[22] A. d'Aspremont, *Risk-Management method for the Libor market model using semidefinite programming*, Journal of Computational Finance 8 (2005), pp. 77-99.

[23] J. de Leeuw, *Applications of convex analysis to multidimensional scaling.* In J. R. Barra, F. Brodeau, G. Romier, and B. van Cutsem (Eds.), Recent developments in statistics, Amsterdam, The Netherlands, 1977, pp. 133–145.

[24] J. de Leeuw, *Convergence of the majorization method for multidimensional scaling*, Journal of classification 5 (1988), pp. 163–180.

[25] J. de Leeuw, *Fitting distances by least squares*, technical report, University of California, Los Angeles, 1993.

[26] J. de Leeuw, *Block relaxation algorithms in statistics*. In H. H. Bock, W. Lenski and M. M. Richter (Eds.), Information Systems and Data Analysis, Springer-Verlag., Berlin, 1994, pp. 308–325.

[27] J. de Leeuw, *A decomposition method for weighted least squares low-rank approximation of symmetric matrices*, Department of Statistics, UCLA, April 2006. Available at `http://repositories.cdlib.org/uclastat/papers/2006041602`.

[28] J. de Leeuw and W. J. Heiser, *Convergence of correction matrix algorithms for multidimensional scaling*, In J. C. Lingoes, I. Borg and E. E. C. I. Roskam (Eds.), Geometric Representations of Relational Data, Mathesis Press, 1977, pp. 735–752.

[29] C. Ding, D. F. Sun and K. -C. Toh, *An introduction to a class of matrix cone programming* technical report, National University of Singapore, 2010.

[30] B.C. Eaves, *On the basic theorem for complemenarity*, Mathematical Programming 1 (1971), pp. 68–75.

[31] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika 1 (1936), pp. 211–218.

[32] K. Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations*, Proceedings of the National Academy of Science of U.S.A. 35 (1949), pp. 652–655.

[33] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.

[34] M. Fazel, H. Hindi, and S. Boyd *A rank minimization heuristic with application to minimum order system approximation*, Proceedings of the American Control Conference 6 (2001), pp. 4734–4739.

[35] M. Fazel, H. Hindi, and S. Boyd, *Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices*, Proceedings of the American Control Conference (2003), pp. 2156–2162.

[36] M. Fazel, H. Hindi, and S. Boyd, *Rank minimization and application in system theory*, Proceedings of the American Control Conference (2004), pp. 3273–3278.

[37] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Mathematical Programming 76 (1997), pp. 513–532.

[38] B. Flury, *Common Principal Components and Related Multivariate Models*, John Wiley & Sons, New York, 1988.

[39] M. Fukushima, Z.-Q. Luo, and P. Tseng, *Smoothing functions for second-order cone complementarity problems*, SIAM Journal on Optimization 12 (2002), pp. 436–460.

[40] M. Fukushima and H. Mine, *A generalized proximal point algorithm for certain non-convex minimization problems*, International Journal of Systems Science 12 (1981), pp. 989–1000.

[41] T. Fushiki, *Estimation of positive semidefinite correlation matrices by using convex quadratic semidefinite programming*, Neural Computation 21 (2009), pp. 2028–2048.

[42] Y. Gao and D. F. Sun, *Calibrating least squares semidefinite programming with equality and inequality constraints*, SIAM Journal on Matrix Analysis and Applications 31 (2009), pp. 1432–1457.

[43] N. Gillis *Weighted Low-Rank Approximations*, talk presented in the 20th International Symposium on Mathematical Programming, August 2009.

[44] G. H. Golub and C. F. van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, USA, Third Edition, 1996.

[45] D. Gross, *Recovering low-rank matrices from few coefficients in any basis*, Preprint, 2009.

[46] I. Grubišić, *Interest Rate Theory: The BGM Model*, master thesis, Leiden University, August 2002. Available at `http://www.math.uu.nl/people/grubisic`.

[47] I. Grubišić and R. Pietersz, *Efficient rank reduction of correlation matrices*, Linear Algebra and Its Applications 422 (2007), pp. 629–653.

[48] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd edition, Cambridge University Press, 1952.

[49] W. J. Heiser, *A generalized majorization method for least squares multidimensional scaling of pseudodistance that may be negative*, Psychometrika 56 (1991), pp. 7–27.

[50] W. J. Heiser, *Convergent computation by iterative majorization: theory and applications in multidimensional data analysis*, In W. J. Krzanowski (Ed.), Recent Advances in Descriptive Multivariate Analysis, Oxford University Press, Oxford, 1995, pp. 157–189.

[51] N. J. Higham, *Computing the nearest correlation matrix – a problem from finance*, IMA Journal of Numerical Analysis 22 (2002), pp. 329–343.

[52] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, I, volume 305 of Grundlehren der Mathematischen Wissenschaften, Fundamental Principles of Mathematical Sciences. SpringerVerlag, Berlin, 1993.

[53] W. Hoge, *A subspace identification extension to the phase correlation method*, IEEE Transactions on Medical Imaging 22 (2003), pp. 277–280.

[54] K. F. Jiang, D. F. Sun and K. -C. T, *A proximal point method for matrix least squares problem with nuclear norm regularization*, technical report, National University of Singapore, 2010.

[55] A. N. Kercheval, *On Rebonato and Jäckel's parametrization method for finding nearest correlation matrices*, International Journal of Pure and Applied Mathematics 45 (2008), pp. 383–390.

[56] R. H. Keshavan, A. Montanari, and S. Oh, *Matrix completion from a few entries*, preprint, 2009.

[57] H. A. L. Kiers, *Majorization as a tool for optimizing a class of matrix functions*, Psychometrika 55 (1990), pp. 417–428.

[58] H. A. L. Kiers, *Setting up alternating least squares and iterative majorization algorithm for solving various matrix optimization problems*, Computational Statistics & Data Analysis 41 (2002), pp. 157–170.

[59] D. L. Knol and J. M. F. ten Berge, *Least-squares approximation of an improper matrix by a proper one*, Psychometrika 54 (1989), pp. 53–61.

[60] M. Korányi, *Monotone functions on formally real Jordan algebras*, Mathematische Annalem 269 (1984), pp. 73–76.

[61] B. Kummer, Newton's method for non-differentiable functions, in *Advances in Mathematical Optimization*, J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach and K. Zimmerman, eds., Akademi-Verlag, Berlin, pp. 114–125, 1988.

[62] A. B. Kurtulan, *Correlations in economic capital models for pension fund pooling*, Master Thesis, Tilburg University, December 2009.

[63] A.S. Lewis, *The convex analysis of unitarily invariant matrix functions*, Journal of Convex Analysis 2 (1995), pp. 173–183.

[64] A. S. Lewis, *Derivatives of spectral functions*, Mathematics of Operations Research 21 (1996), pp. 576–588.

[65] A. S. Lewis and M. L. Overton, *Nonsmooth optimization via BFGS*, 2008. The MATLAB software is downloadable at `http://cs.nyu.edu/overton/software/index.html`.

[66] A. S. Lewis and H. S. Sendov, *Nonsmooth Analysis of Singular Values. Part I: Theory*, Set-Valued Analysis 13 (2005), pp. 213–241.

[67] D. Li, X. L. Sun, and J. Wang, *Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection*, Mathematical Finance 16 (2006), pp. 83–101.

[68] Q. N. Li and H. D. Qi, *A sequential semismooth Newton method for the nearest low-rank correlation matrix problem*, Technical Report, University of Southampton, September 2009.

[69] F. Lillo and R. N. Mantegna, *Spectral density of the correlation matrix of factor models: A random matrix theory approach*, Physical Review E 72 (2005), pp. 016219-1–016219-10.

[70] M. Lobo, M. Fazel, and S. Boyd, *Portfolio optimization with linear and fixed transaction costs*, Annals of Operations Research 152 (2006), pp. 341–365.

[71] K. Löwner, *Über monotone matrixfunktionen*, Mathematische Zeitschrift 38 (1934), pp. 177–216.

[72] P. M. Lurie and M. S. Goldberg, *An approximate method for sampling correlated variables from partially-specified distributions*, Management Science 44 (1998), pp. 203–218.

[73] S. Q. Ma, D. Goldfarb and L. F. Chen *Fixed point and Bregman iterative methods for matrix rank minimization*, to appear in Mathematical Programming Series A, 2008.

[74] J. Malick, *A dual approach to semidefinite least-squares problems*, SIAM Journal on Matrix Analysis and Applications 26 (2004), pp. 272–284.

[75] F. W. Meng, D. F. Sun, and G. Y. Zhao *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Mathematical Programming 104 (2005), pp. 561–581.

[76] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM Journal on Control and Optimization 15 (1977), pp. 959–972.

[77] H. Mine and M. Fukushima, *A minimization method for the sum of a convex function and a continuously differentiable function*, Journal of Optimization Theory and Applications 33 (1981), pp. 9–23.

[78] S. K. Mishra, *Optimal solution of the nearest correlation matrix problem by minimization of the maximum norm*, Munich Personal RePEc Archive, August 2004. Available at `http://mpra.ub.uni-muenchen.de/1783`.

[79] M. Morini and N. Webber, *An EZI method to reduce the rank of a correlation matrix in financial modelling*, Applied Mathematical Finance 13 (2009), pp. 309–331.

[80] G. Natsoulis, C. I Pearson, J. Gollub, B. P. Eynon, J. Ferng, R. Nair, R. Idury, M. D Lee, M. R Fielden, R. J Brennan, A. H Roter and K. Jarnagin, *The liver pharmacological and xenobiotic gene response repertoire*, Molecular Systems Biology 4 (2008), pp. 1–12.

[81] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM Journal on Optimization 15 (2005), pp. 229–251.

[82] Y. Nesterov, *Smooth minimization of nonsmooth functions*, Mathematical Programming 103 (2005), pp. 127–152.

[83] J. von Neumann, *Some matrix inequalities and metrization of matric-space*, Tomsk University Review 1 (1937), pp. 286–300.

[84] J. M. Otega and W. C. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York, 1970.

[85] J. V. Outrata and D. F. Sun *On the Coderivative of the Projection Operator onto the Second-order Cone*, Set-Valued Analysis 16 (2008), pp. 999–1014.

[86] M. Overton and R. S. Womersley, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Mathematical Programming 62 (1993), pp. 321–357.

[87] R. Pietersz and P. Groenen, *Rank reduction of correlation matrices by majorization*, Quantitative Finance 4 (2004), pp. 649–662.

[88] H.-D. Qi, *Positive semidefinite matrix completions on chordal graphs and the constraint nondegeneracy in semidefinite programming*, Linear Algebra and Its Applications 430 (2009), pp. 1151–1164.

[89] H.-D. Qi and D. F. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM Journal on Matrix Analysis and Applications 28 (2006), pp. 360–385.

[90] H.-D. Qi and D. F. Sun, *An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem*, to appear in IMA Journal of Numerical Analysis, 2010.

[91] H.-D. Qi and D. F. Sun, *Correlation stress testing for value-at-risk: an unconstrained convex optimization approach*, Computational Optimization and Applications 45 (2010), pp. 427–462.

[92] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research 18 (1993), pp. 227–244.

[93] L. Qi and D. F. Sun, *Nonsmooth and smoothing methods for NCP and VI*, Encyclopedia of Optimization, C. Floudas and P. Pardalos (editors), Kluwer Academic Publisher, USA, 2001, pp. 100–104.

[94] L. Qi, D. F. Sun and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Mathematical Programming 87 (2000), pp. 1–35.

[95] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Mathematical Programming 58 (1993), pp. 353–367.

[96] F. Rapisarda, D. Brigo and F. Mercurio, *Parametrizing correlations: a geometric interpretation*, IMA Journal of Management Mathematics 18 (2007), pp. 55–73.

[97] G. Ravindran and M. S. Gowda, *Regularization of P_0-functions in box variational inequality problems*, SIAM Journal on Optimization 11 (2000), pp. 748–760.

[98] R. Rebonato, *Calibrating the BGM model*, Risk Magazine (1999), pp. 74–79.

[99] R. Rebonato, *On the simultaneous calibration of multifactor lognormal interest rate models to black volatilities and to the correlation matrix*, Journal of Computational Finance 2 (1999), pp. 5–27.

[100] R. Rebonato, *Morden pricing of interest-rate derivatives*, Princeton University Press, New Jersey, 2002.

[101] R. Rebonato and P. Jackel, *The most general methodology to create a valid correlation matrix for risk management and option pricing purposes*, The Journal of Risk 2 (1999), pp. 17–27.

[102] B. Recht, *A Simpler Approach to Matrix Completion*, to appear in Journal of Machine Learning Research, 2009.

[103] B. Recht, M. Fazel, and P.A. Parrilo, *Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization*, to appear in SIAM Review, 2007.

[104] S. M. Robinson, *Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy*, Mathematical Programming Study 22 (1984), pp. 217–230.

[105] R. T. Rockafellar, *Convex Analyis*, Princeton University Press, Princeton, 1970.

[106] R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

[107] R. T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[108] E. Schmidt, *Zur Theorie der linearen nichtlinearen Integralgleichungen*, Mathematische Annalen 63 (1907), pp. 433–476.

[109] N. C. Schwertman and D. M. Allen, *Smoothing an indefinite variance-covariance matrix*, Journal of Statistical Computation and Simulation 9 (1979), pp. 183–194.

[110] D. Simon, *Reduced order kalman filtering without model reduction*, Control and Intelligent Systems 35 (2007), pp. 169–174.

[111] D. Simon and J. Abell, *A Majorization Algorithm for Constrained Correlation Matrix Approximation*, Linear Algebra and its Applications 432 (2010), pp. 1152–1164.

[112] J.F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optimization Methods and Software 11 & 12 (1999), pp. 625–653.

[113] D.F. Sun and J. Sun, *Semismooth matrix valued functions*, Mathematics of Operations Research 27 (2002), pp. 150–169.

[114] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, Philadephia, 1997.

[115] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, submitted to SIAM Journal on Optimization, 2008.

[116] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Mathematical Programming 117 (2009), pp. 387–423.

[117] R.H. Tütüncü, K.C. Toh, and M.J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming 95 (2003), pp. 189–217.

[118] G.A. Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra and its Applications 170 (1992), pp. 33–45.

[119] G.A. Watson, *On matrix approximation problems with Ky Fan k norms*, Numerical Algorithms 5 (1993), pp. 263–272.

[120] R. Werner and K. Schöttle, *Calibration of correlation matrices - SDP or not SDP*, technical report, Munich University of Technology, 2007.

[121] L.X. Wu, *Fast at-the-money calibration of the LIBOR market model using Lagrange multipliers*, Journal of Computational Finance 6 (2003), pp. 39–77.

[122] E.H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory I and II.* In E. H. Zarantonello (Ed.), Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, pp. 237–424.

[123] Z.Y. Zhang and L.X. Wu, *Optimal low-rank approximation to a correlation matrix*, Linear Algebra and Its Applications 364 (2003), pp. 161–187.

[124] J.Y. Zhao, *The Smoothing Function of the Nonsmooth Matrix Valued Function*, Master thesis, National University of Singapore, July 2004. Downloadable from `http://www.math.nus.edu.sg/~matsundf/Zhao_July_2004.pdf`.

Name: Gao Yan

Degree: Doctor of Philosophy

Department: Mathematics

Thesis Title: STRUCTURED LOW RANK MATRIX OPTIMIZATION PROBLEMS: A PENALTY APPROACH

Abstract

In this thesis, we study a class of structured low rank matrix optimization problems (SLR-MOPs) which aim at finding an approximate matrix of certain specific structures and whose rank is no more than a prescribed number. This kind of approximation is needed in many important applications arising from a wide range of fields. The SLR-MOPs are in general non-convex and thus difficult to solve due to the presence of the rank constraint. In this thesis, we propose a penalty approach to deal with this difficulty. Some rationale to motivate this penalty technique is also addressed. We further present a general proximal subgradient method for the purpose of solving the penalized problem. Finally, we design a quadratically convergent smoothing Newton-BiCGStab method to solve the resulted sub-problems. Numerical results indicate that our approach is able to handle both the rank and the linear constraints effectively, in particular in the situations when the rank is not very small.

Keywords:
structured low rank matrix, a proximal subgradient method, a penalty approach, a smoothing Newton-BiCGStab method.

STRUCTURED LOW RANK MATRIX OPTIMIZATION PROBLEMS: A PENALTY APPROACH

GAO YAN

NATIONAL UNIVERSITY OF SINGAPORE

AUGUST 2010

STRUCTURED LOW RANK MATRIX OPTIMIZATION PROBLEMS: A PENALTY APPROACH GAO YAN 2010