ON ALTERNATING DIRECTION METHODS FOR MONOTROPIC SEMIDEFINITE PROGRAMMING

ZHANG SU

NATIONAL UNIVERSITY OF SINGAPORE

2009

ON ALTERNATING DIRECTION METHODS FOR MONOTROPIC SEMIDEFINITE PROGRAMMING

ZHANG SU

(*B.Sci., M.Sci., Nanjing University, China*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF DECISION SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2009

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere thanks to my supervisor, Professor Jie Sun, for the guidance and full support he has provided me throughout my graduate studies. In times of need or trouble, he has always been there ready to help me out.

I am grateful to my thesis committee members, Professor Gongyun Zhao and Associate Professor Mabel Chou, for their valuable suggestions and comments. I also want to thank all my colleagues and friends in National University of Singapore who have given me much support during my four-year Ph.D. studies.

I must take this opportunity to thank my parents for their unconditional love and support through all these years and my wife, Xu Wen, for accompanying me this long way with continuous love, encouragement, and care. Without you, I would be nothing.

Last but not least, I am grateful to National University of Singapore for providing me with the environment and facilities needed to complete my study.

ABSTRACT

This thesis studies a new optimization model called monotropic semidefinite programming and a type of numerical methods for solving this problem. The word "monotropic programming" was probably first popularized by Rockafellar in his seminal book, which means a linearly constrained minimization problem with convex and separable objective function. The original monotropic programming requires the decision variable to be an n-dimensional vector, while in our monotropic semidefinite programming model, the decision variable is a symmetric block-diagonal matrix. This model extends the vector monotropic programming model to the matrix space on one hand, and on the other hand it extends the linear semidefinite programming model to the convex case.

We propose certain modified alternating direction methods for solving monotropic semidefinite programming problems. The alternating direction method was originally proposed for structured variational inequality problems. We modify it to avoid solving difficult sub-variational inequality problems at each iteration, so that only metric projections onto convex sets are sufficient for the convergence. Moreover, these methods are first order algorithms (gradient-type methods) in nature, hence they are relatively easy to implement and require less computation in each iteration.

We then specialize the developed modified alternating direction methods into the algorithms for solving convex nonlinear semidefinite programming problems in which the methods are further simplified. Of particular interest to us is the convex quadratically constrained quadratic semidefinite programming problem. Compared with the well-studied linear semidefinite program, the quadratic model is so far less explored although it has important applications.

An interesting application arises from the covariance matrix estimation in financial management. In portfolio management covariance matrix is a key input to measure risk, thus correct estimation of covariance matrix is critical. The original nearest correlation matrix problem only considers linear constraints. We extend this model to include quadratic ones so as to catch the tradeoff between long-term information and short-term information. We notice that in practice the investment community often uses the multiple-factor model to explain portfolio risk. This can be also incorporated into our new model. Specifically, we adjust unreliable covariance matrix estimations of stock returns and factor returns simultaneously while requiring them to fit into the previously constructed multiple-factor model.

Another practical application of our methods is the matrix completion problem. In practice, we usually know only partial information of entries of a matrix and hope to reconstruct it according to some pre-specified properties. The most studied problems include the completion problem of distance matrix and the completion problem of low-rank matrix. Both problems can be modelled in the framework of monotropic semidefinite programming and the proposed alternating direction method provides an efficient approach for solving them.

Finally, numerical experiments are conducted to test the effectiveness of the proposed algorithms for solving monotropic semidefinite programming problems. The results are promising. In fact, the modified alternating direction method can solve a large problem with a 2000×2000 variable matrix in a moderate number of iterations and with reasonable accuracy.

CONTENTS

1. INTRODUCTION

Optimization models play a very important role in operations research and management science. Optimization models with symmetric matrix variables are often referred to as semidefinite programs. The study on these models has a relatively short history. Intensive studies on the theory, algorithms, and applications on semidefinite programs have only begun since 1990s. However, so far most of the work has been concentrated on the linear case, where, except the semidefinite cone constraint, all other constraints as well as the objective function are linear with respect to the matrix variable.

When one attempts to model some nonlinear phenomena in the above fields, linear semidefinite programming (SDP) is not enough. Therefore the research on nonlinear semidefinite programming (NLSDP) began from around 2000. Interestingly enough, some of the crucial applications of the nonlinear model arise from financial management and related business areas. For example, the nearest correlation matrix problem is introduced to adjust unqualified covariance matrix estimation. Then the objective, which is the distance between two matrices, must be nonlinear. In Chapters 5 and 6, more such applications can be raised. They motivated our project in an extent.

Much work is yet to be done to effectively solve an NLSDP. Nonlinearity could bring significant difficulty in designing algorithms. In addition, the semidefinite optimization problems easily lead to large-scale problems. For example, a 2000×2000 symmetric variable matrix has more than 2,000,000 independent variables. The situation becomes even worse if there are more than one matrix variable in the problem. Technically, we could combine all the variable matrices into a big block-diagonal matrix, but it is often not wise to do so for computational efficiency. In our research, we keep the different matrix variables and concentrate on how to take advantage of the problem structure such as separability and linearity.

1.1 Monotropic Semidefinite Programming

We study a new optimization model called monotropic semidefinite programming (MSDP) in this thesis research. "Monotropic programming", first popularized by Rockafellar in his seminal book [55], deals with a linearly constrained minimization problem with convex and separable objective function. The original monotropic programming assumes the decision variable to be an n-dimensional vector, while in our MSDP model, the decision variable is a set of symmetric matrices. In other words, we replace each variable x_i in the original model by a symmetric matrix $X_i \in \Re^{p_i \times p_i}$. As a result, the block-diagonal matrix

$$X = \operatorname{diag}(X_1, \cdots, X_n)$$

of dimension $\sum_{i=1}^{n} p_i$ could be thought of as the decision variable. Obviously, if $p_1 = \cdots = p_n = 1$, this model reduces to the n-dimensional vector case. On the other hand, if $n = 1$, this model reduces to a linearly constrained convex NLSDP problem. Since we allow additional set constraints as specified later, the later model could include nonlinear constraints and thus it is actually the convex NLSDP without loss of generality.

The MSDP has the formulation as follows.

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} f_i\left(X_i\right) \\ \text{s.t.} \quad & \sum_{i=1}^{n} \mathcal{A}_i\left(X_i\right) = b \\ & X_i \in \Omega_i \equiv \bigcap_{j=1}^{m} \Omega_{ij}, \ i = 1, \cdots, n, \end{aligned} \tag{1.1}$$

where $b \in \Re^l$, $X_i \in \Re^{p_i \times p_i}$, $f_i : \Re^{p_i \times p_i} \to \Re$ is a convex function, and Ω_{ij} is a convex set in $\Re^{p_i \times p_i}$. Furthermore, $\mathcal{A}_i$ denotes the linear operator: $\Re^{p_i \times p_i} \to \Re^l$,

$$\mathcal{A}_i\left(X_i\right) \equiv \begin{pmatrix} \langle A_{i1}, X_i \rangle \\ \vdots \\ \langle A_{il}, X_i \rangle \end{pmatrix}.$$

Usually, each Ω_{ij} is a simple convex set and we assume that it is easy to compute the projection of a point onto this set. One example is a box $\Omega_{ij} = \left\{X_i : \underline{X}_i \le X_i \le \overline{X}_i\right\}$, where the matrix inequality is understood entry-wise. Another example is a ball $\Omega_{ij} = \left\{X_i : \|X_i - C\|^2 \le \epsilon\right\}$. However, the most interesting case is when Ω_{ij} is a semidefinite cone. In this case the projection of X_i onto Ω_{ij} involves the evaluation of eigenvalues of X_i.

1.2 The Variational Inequality Formulation

Let us introduce new variables

$$Y_{ij} = \mathcal{L}_{ij}\left(X_i\right), \ i = 1, \cdots, n, \ j = 1, \cdots, m,$$

where $\mathcal{L}_{ij} : \Re^{p_i \times p_i} \rightarrow \Re^{p_i \times p_i}$ is a fixed given invertible linear operator. That is, there exists a linear operator $\mathcal{L}_{ij}^{-1} : \Re^{p_i \times p_i} \rightarrow \Re^{p_i \times p_i}$ such that

$$\mathcal{L}_{ij}^{-1}\left(\mathcal{L}_{ij}\left(X_i\right)\right) = X_i.$$

The adjoint operator $\mathcal{L}_{ij}^T : \Re^{p_i \times p_i} \rightarrow \Re^{p_i \times p_i}$ is defined by

$$\left\langle X_i', \mathcal{L}_{ij}\left(X_i\right)\right\rangle = \left\langle \mathcal{L}_{ij}^T\left(X_i'\right), X_i\right\rangle.$$

Here and below, unless otherwise specified, the inner product $\langle \cdot, \cdot \rangle$ is the Frobenius inner product defined as $\langle A, B\rangle = \mathrm{trace}(A^T B)$. Let μ_{ij} be fixed constants satisfying

$$0 \leq \mu_{ij} \leq 1 \quad \text{and} \quad \sum_{j=0}^{m} \mu_{ij} = 1, \quad \text{for } i = 0, 1, \cdots, n. \tag{1.2}$$

Then we may re-write (1.1) equivalently as

$$\begin{aligned}
\min \quad & \sum_{i=1}^{n}\left(\mu_{i0} f_i\left(X_i\right)+\sum_{j=1}^{m} \mu_{ij} f_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}\right)\right) \\
\text{s.t.} \quad & Y_{ij}=\mathcal{L}_{ij}\left(X_i\right), \; i=1,\cdots,n, \; j=1,\cdots,m \\
& \left(Y_{11},\cdots,Y_{n1}\right) \in \Omega^{'} \equiv \left\{\left(Y_{11},\cdots,Y_{n1}\right): \sum_{i=1}^{n} \mathcal{A}_i \circ \mathcal{L}_{i1}^{-1}\left(Y_{i1}\right)=b\right\} \\
& Y_{ij} \in \Omega_{ij}^{'} \equiv \left\{Y_{ij}: \mathcal{L}_{ij}^{-1}\left(Y_{ij}\right) \in \Omega_{ij}\right\}, \; i=1,\cdots,n, \; j=2,\cdots,m \\
& X_i \in \Omega_{i1}, \; i=1,\cdots,n.
\end{aligned} \tag{1.3}$$

Compared with its original form, (1.3) looks more complicated because of the addition of many new variables. In fact we do this to separate the set constraint for each X_i so that each Ω_{ij} is as simple as possible. Then it is easy to compute the projection onto it which is critical in our proposed methods. For example, consider that X_k belongs to the intersection of several balls. The set constraint $X_k \in \Omega_k$ is not simple enough. After introducing new variables Y_{kj} and letting $Y_{kj} = X_k$, each Y_{kj} is only required to be in one ball onto which there is a close-form formula for the projection. Besides, the update of Y_{ij} at each iteration can be done in parallel in our proposed methods as shown later; hence in practice there will not be much additional computational load.

The reason behind defining the matrix-to-matrix operator $\mathcal{L}_{ij}$ rather than directly defining them as matrices is that we would like to keep some specific properties of matrices, e.g., the requirement of positive semidefiniteness for symmetric matrices. The flexible choice of linear operator $\mathcal{L}_{ij}$

enables us the possibility to simplify the original problem (1.1). In Chapter 4.1 we will show that ellipsoid-type set with all positive eigenvalues can be converted to ball-type set by choosing suitable linear operators. Then the projection onto balls, rather than ellipsoids, can be calculated by using a formula instead of by using numerical algorithms such as those introduced in [14].

About the choice of $\mu_{ij},\ j = 0, \cdots, m$, the trivial way is to let $\mu_{i0} = 1$ and let the other μ_{ij}s be zero. However, the rule of (1.2) also allows other specifications of μ_{ij} based on some prior information.

The Lagrangian function of Problem (1.3) is

$$\begin{aligned} L \;\equiv\; & \sum_{i=1}^{n}\left(\mu_{i0} f_i\left(X_i\right) + \sum_{j=1}^{m} \mu_{ij} f_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}\right)\right) \\ & - \sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \lambda_{ij}, \mathcal{L}_{ij}\left(X_i\right) - Y_{ij}\right\rangle . \end{aligned} \tag{1.4}$$

Notice that the Lagrangian multipliers $\lambda_{ij},\ i = 1, \cdots, n,\ j = 1, \cdots, m$, are matrices. From now we assume each $f_i,\ i = 1, \cdots, n$, is continuously differentiable and its first order derivative is written as ∇f_i. It is well known that under mild constraint qualifications (e.g., Slater' condition), strong duality holds and hence, X_i^* is a solution of (1.3) if and only if there exists λ_{ij}^* such that $\left(X_i^*, Y_{ij}^*, \lambda_{ij}^*\right)$ satisfies

$$\begin{cases} \left\langle X_i - X_i^*, \mu_{i0} \nabla f_i\left(X_i^*\right) - \sum\limits_{j=1}^{m} \mathcal{L}_{ij}^T\left(\lambda_{ij}^*\right)\right\rangle \geq 0, \\ \forall\ X_i \in \Omega_{i1},\ i = 1, \cdots, n \\ \sum\limits_{i=1}^{n} \left\langle Y_{i1} - Y_{i1}^*, \mu_{i1}\left(\mathcal{L}_{i1}^{-1}\right)^T \circ \nabla f_i \circ \mathcal{L}_{i1}^{-1}\left(Y_{i1}^*\right) + \lambda_{i1}^*\right\rangle \geq 0, \\ \forall\ \left(Y_{11}, \cdots, Y_{n1}\right) \in \Omega' \\ \left\langle Y_{ij} - Y_{ij}^*, \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ \nabla f_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^*\right) + \lambda_{ij}^*\right\rangle \geq 0, \\ \forall\ Y_{ij} \in \Omega_{ij}',\ i = 1, \cdots, n,\ j = 2, \cdots, m \\ \mathcal{L}_{ij}\left(X_i^*\right) = Y_{ij}^*,\ i = 1, \cdots, n,\ j = 1, \cdots, m. \end{cases} \tag{1.5}$$

For convenience, we make a basic assumption to guarantee that the MSDP problem (1.3) under consideration is solvable.

Assumption 1.2.1. *The solution set $\left(X_i^*, Y_{ij}^*, \lambda_{ij}^*\right)$ of KKT system (1.5) is nonempty.*

Consequently, under this assumption, Problem (1.1) is solvable and $X_i^*, i = 1, \cdots, n$, is a solution to Problem (1.1).

1.3 Research Objectives and Results

The objectives of this thesis are:

- To study a new optimization model, namely MSDP. This model extends the monotropic programming model from vectors to matrices on one hand, and the linear SDP model to the convex case on the other hand. Then we study its optimal condition as a variational inequality problem.

- To propose some general algorithms for solving MSDP problems. The alternating direction method (ADM) appears to be an efficient first order algorithm (gradient-type method), which can take advantage of the special structure of the problem. However, the sub-variational inequality problems that appear at each iteration are not easy to solve in practice. Hence we modify the ADM so that solving the sub-variational inequalities is substituted by computing a metric projection onto a convex set. The MSDP problems with a quadratic objective function and with a general nonlinear objective are investigated, respectively. There are two respective modification procedures (the modified ADM and the prediction-correction ADM) to deal with them. For each of the modifications we present detailed convergence proof under mild conditions.

- To investigate convex NLSDP as a special case of MSDP. Particularly, we consider the convex quadratically constrained quadratic semidefinite programming (CQCQSDP) problem which generalizes the so-called convex quadratic semidefinite programming (CQSDP). We also consider the general convex nonlinear semidefinite programming (CNLSDP) problem as a special case of MSDP. These new algorithms are relatively easy to implement and require less computation at each iteration.

- To explore some important applications of MSDP in business management. The covariance matrix estimation problem is essential in financial management. We build a new optimization framework to extend the nearest correction matrix problem and the least squares covariance matrix problem. The generalized model can take into consideration the tradeoff between long-term data and short-term data. Furthermore the multiple-factor model, which is popular in investment management, can be also incorporated. Another application studied is the matrix completion problem, including the completion problem of distance matrix and the completion problem of low-rank matrix. They are very useful in practice and the proposed ADM provides another efficient solution approach for these problems.

- To perform numerical experiments on the proposed algorithms.

1.4 Structure of the Thesis

The remaining chapters of the thesis are organized as follows. In Chapter 2 we review the literature on SDP and ADM. We modify the ADM for solving MSDP problems with quadratic objective and general nonlinear objective in Chapter 3 and prove the convergence properties for two such modifications. Chapter 4 will consider the specializations on convex NLSDP, including CQC-QSDP and CNLSDP. Practical applications including the covariance matrix estimation problem and the matrix completion problem are considered respectively in Chapters 5 and 6. In Chapter 7 we present numerical results to show the efficiency of proposed algorithms. Finally, Chapter 8 concludes the thesis with a summary of results.

2. LITERATURE REVIEW

In this chapter, we briefly review the literature on SDP, focusing on NLSDP, and ADM. We also introduce our notations.

2.1 Review on Semidefinite Programming

Let $\mathbb{S}^n$ be the finite-dimensional Hilbert space of real symmetric matrices equipped with the Frobenius inner product $\langle A, B\rangle = \text{trace}(A^T B)$. Let $\mathbb{S}^n_+$ ($\mathbb{S}^n_{++}$, respectively) be the subset of $\mathbb{S}^n$ consisting of all symmetric positive semidefinite (definite, respectively) matrices. Clearly, $\mathbb{S}^n_+$ is a convex cone and is called the positive semidefinite cone. As a convention, we write $X \succeq 0$ ($X \succ 0$, respectively) to represent $X \in \mathbb{S}^n_+$ ($X \in \mathbb{S}^n_{++}$, respectively). We write $X \succeq Y$ or $Y \preceq X$ to represent $X - Y \succeq 0$, respectively. Similarly we define $X \succ Y$ and $Y \prec X$. The so-called standard form of SDP is as follows.

$$\min \langle C, X\rangle \quad \text{s.t. } X \succeq 0,\ \langle A_i, X\rangle = b_i,\ i = 1, \cdots, m,$$

where $b \in \Re^m$, $A_i \in \mathbb{S}^n$, and $C \in \mathbb{S}^n$ are given. This model has attracted researchers from diverse fields, including experts in convex optimization, linear algebra, numerical analysis, combinatorics, control theory, and statistics. The main reason is that a lot of applications lead to SDP problems [5, 7, 53]. As a consequence, there are many different approaches for solving SDP, among which the interior point method is well known for its polynomial computational property. A comprehensive survey of the early work can be found in [68].

A natural extension of SDP is NLSDP, in which either the objective function or a constraint is nonlinear in X. Certainly NLSDP model is more general and can therefore have specific applications beyond the applications of SDP. Actually NLSDP has been used in, for instance, feed back control, structural optimization, and truss design problems, etc. [4, 37].

While the mathematical formats of NLSDP may be different in various applications, it is convenient to consider the following general model.

$$\min \ f(X) \quad \text{s.t. } h(X) = 0, \ g(X) \in \mathcal{K}, \tag{2.1}$$

where $f : \mathbb{S}^n \to \Re$, $h : \mathbb{S}^n \to \Re^m$, and $g : \mathbb{S}^n \to \mathcal{Y}$ are given continuously differentiable functions, $\mathcal{Y}$ is a Hilbert space, and $\mathcal{K}$ is a symmetric (homogenous, self-dual) cone in $\mathcal{Y}$. If in addition, f is convex, h is linear, and the constraint $g(X) \in \mathcal{K}$ defines a convex set, Problem (2.1) becomes a *convex* semidefinite program.

The first order and second order optimality conditions for NLSDP have been studied in [6, 59, 60]. On the other hand, research on numerical algorithms for NLSDP is mainly in its developing stage. Comparing with linear programming, nonlinear programming is much more difficult to solve. The same happens to NLSDP.

Recently, some different methods have been proposed. Kocvara and Stingl [36] developed a code (PENNON) supporting NLSDP problems, where the augmented Lagrangian method was used. Later Sun, Sun, and Zhang [62] analyzed the convergence rate for augmented Lagrangian method in the NLSDP setting. A smoothing Newton method for NLSDP, which is a second order algorithm, is considered in Sun, Sun, and Qi [61]. A variant of the smoothing Newton methods is subsequently studied in [38]. Similar Newton-type methods [33, 34] originally proposed for SDP can also be extended to solve NLSDP. An analytic center cutting plane method is investigated by Sun, Toh, and Zhao [63, 66], which can be used for solving CNLSDPs. Another approach called successive linearization method appears in Fares, Noll, and Apkarian [20], Correa and Ramirez [12], and Kanzow et al. [35]. Noll and Apkarian [51, 52] also suggested the spectral bundle methods. In Jarre [32], Leibfitz and Mostafa [44], and Yamashita, Yabe, and Haradathe [69], interior methods are discussed. In addition, Gowda and his collaborators have extensively studied complementarity problems in general symmetric cone setting [26, 27], which are closely related to the solution of NLSDPs.

Other works focus on solving some special classes of NLSDP. Among them, the CQSDP problem, perhaps the most basic NLSDP problem in a sense, has received a lot of attention because of a number of important applications in engineering and management. In the CQSDP model, the objective is a convex quadratic function and the constraints are linear, together with the semidefinite cone constraint. For example, in order to find a positive semidefinite matrix that best approximates the solution to the matrix equation system

$$\langle A_i, X\rangle = b_i,\ i = 1, \cdots, m,$$

we need to solve the matrix least-square problem

$$\min \sum_{i=1}^{m} \|\langle A_i, X\rangle - b_i\|^2 \quad \text{s.t. } X \succeq 0, \tag{2.2}$$

which is in the form of CQSDP.

In [50], a theoretical primal-dual potential reduction algorithm was proposed for CQSDP problems by Nie and Yuan. The authors suggested to use the conjugate gradient method to compute an approximate search direction. Subsequent works include Qi and Sun [57] and Toh [64]. Qi and Sun used a Lagrangian dual approach. Toh introduced an inexact primal-dual path-following method with three classes of pre-conditioners for the augmented equation for fast convergence under suitable nondegeneracy assumptions. In two recent papers, Malick [48] and Boyd and Xiao [8], respectively applied classical quasi-Newton method (in particular, the BFGS method) and the projected gradient method to the dual problem of CQSDP. More recently, Gao and Sun [25] designed an inexact smoothing Newton method to solve a reformulated semismooth system with two level metric projection operators and demonstrated the efficiency of the proposed method in their numerical experiments.

2.2 Review on the Alternating Direction Method

The general advantage of first order algorithms is twofold. Firstly, this type of methods are relatively simple to implement, thus they are useful in finding an approximate solution of the problems, which may become the "first phase" of a hybrid first-second order algorithm. Secondly, first order methods usually require much less computation per iteration, therefore might be suitable for relatively large problems.

Among the first order approaches for solving large optimization problems, the augmented Lagrangian method is an effective one. It has desirable convergence properties. The augmented Lagrangian function of Problem (1.3) is

$$\begin{aligned} L_{\text{aug}} \equiv & \sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\mathcal{L}_{ij}\left(X_i\right)-Y_{ij}\right\|_{\beta_{ij}}^{2}-\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle\lambda_{ij},\mathcal{L}_{ij}\left(X_i\right)-Y_{ij}\right\rangle \\ & +\sum_{i=1}^{n}\left(\mu_{i0}f_i\left(X_i\right)+\sum_{j=1}^{m}\mu_{ij}f_i\circ\mathcal{L}_{ij}^{-1}\left(Y_{ij}\right)\right), \end{aligned} \tag{2.3}$$

where $\left\|\mathcal{L}_{ij}\left(X_i\right)-Y_{ij}\right\|_{\beta_{ij}}^{2} = \left\langle\mathcal{L}_{ij}\left(X_i\right)-Y_{ij},\beta_{ij}\left(\mathcal{L}_{ij}\left(X_i\right)-Y_{ij}\right)\right\rangle$ and β_{ij} is a self-adjoint positive definite linear operator. Note that a (general) quadratic penalty term has been added to the Lagrangian function (1.4). This additional term is usually not separable respective to X_i and Y_{ij}, which makes the augmented Lagrangian method more difficult to implement, therefore less attractive in practice.

To overcome this difficulty, the ADM is introduced. The ADM generally consists of three steps.

(I) Minimize the augmented Lagrangian function (2.3) with respective to X_i only.

(II) Minimize the augmented Lagrangian function (2.3) with respective to Y_{ij} only.

(III) Update the Lagrangian multipliers λ_{ij}.

Repeat (I), (II), and (III) until a stopping criterion is satisfied.

The ADM can be seen as the block Gauss-Seidel variants of the augmented Lagrangian approach. The fundamental principle involved is to use the most recent information as they are available. Furthermore, it can take advantage of block angular structure. Consequently it is very suitable for parallel computation in a data parallel environment. The ADM was probably first considered by Gabay [23] and Gabay and Mercier [24]. As shown in [46], the ADM is actually an instance of the Doulgas-Rachford splitting procedure of monotone operators [15]. It is also related to the progressive hedging algorithm of Rockafellar and Wets [56]. The ADM has been studied quite extensively in the settings of optimization and numerical analysis. Eckstein [17] and Kontogiorgis [39] gave the detailed analysis of ADMs and tested their efficiency using numerical experiments in the parallel computation environment. Some versions of the ADMs for solving different separable convex optimization problems, including monotropic optimization problems, appeared in [18, 22, 40].

The ADM is very suitable to be applied to MSDP problems in that it can take advantage of the separability structure. We are interested in the technique of decomposition − dividing a large-scale problem into many smaller ones that can be solved in parallel. The ADM just has such a nice property. When applied to Problem (1.3), the ADM becomes the following.

Algorithm 2.2.1. The ADM for MSDP

Do at each iteration until a stopping criterion is met

Step 1. $\left(X_i^k, Y_{ij}^k, \lambda_{ij}^k\right) \to \left(X_i^{k+1}, Y_{ij}^k, \lambda_{ij}^k\right)$, $i = 1, \cdots, n$, *where* X_i^{k+1} *satisfies*

$$\begin{aligned} & \left\langle X_i - X_i^{k+1}, \mu_{i0} \nabla f_i \left(X_i^{k+1}\right) - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \left(\lambda_{ij}^k - \beta_{ij} \left(\mathcal{L}_{ij} \left(X_i^{k+1}\right) - Y_{ij}^k\right)\right) \right\rangle \\ \geq \;& 0, \ \forall X_i \in \Omega_{i1} \end{aligned} \tag{2.4}$$

Step 2. $\left(X_i^{k+1}, Y_{ij}^k, \lambda_{ij}^k\right) \to \left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^k\right)$, $i = 1, \cdots, n$, $j = 1, \cdots, m$, *where* Y_{ij}^{k+1} *satisfies*

$$\begin{aligned} & \sum_{i=1}^{n} \Big\langle Y_{i1} - Y_{i1}^{k+1}, \mu_{i1} \left(\mathcal{L}_{i1}^{-1}\right)^T \circ \nabla f_i \circ \mathcal{L}_{i1}^{-1} \left(Y_{i1}^{k+1}\right) \\ & + \lambda_{i1}^k - \beta_{i1} \left(\mathcal{L}_{i1} \left(X_i^{k+1}\right) - Y_{i1}^{k+1}\right) \Big\rangle \\ \geq \;& 0, \ \forall \ \left(Y_{11}, \cdots, Y_{n1}\right) \in \Omega' \end{aligned} \tag{2.5}$$

and

$$\begin{aligned}
&\Big\langle Y_{ij} - Y_{ij}^{k+1}, \mu_{ij} \left(\mathcal{L}_{ij}^{-1}\right)^T \circ \nabla f_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) \\
&+\lambda_{ij}^{k} - \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^{k+1}\right)\Big\rangle \\
\geq\ & 0,\ \forall\ Y_{ij} \in \Omega_{ij}^{'},\ i = 1, \cdots, n,\ j = 2, \cdots, m
\end{aligned} \tag{2.6}$$

Step 3. $\left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^{k}\right) \to \left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^{k+1}\right)$, $i = 1, \cdots, n,\ j = 1, \cdots, m,$

where

$$\lambda_{ij}^{k+1} = \lambda_{ij}^{k} - \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^{k+1}\right) \tag{2.7}$$

Arbitrary X_i^0, Y_{ij}^0, and λ_{ij}^0 are chosen as the starting point. The ADM reaches optimality by taking alternating steps in the primal and dual space. The updates of variables X_i, Y_{ij}, and λ_{ij} could be done in parallel for all i and j. Primal feasibility, dual feasibility, and complementary slackness are not maintained; instead, all of them are satisfied as the algorithm finds a fixed point of the recursions.

Further studies of ADM can be found, for instance, in [11, 19, 29, 30, 41]. The inexact versions of ADM were proposed by Eckstein and Bertsekas [19] and Chen and Teboulle [11], respectively. He et al. [29] generalized the framework and proposed a new inexact ADM with flexible conditions for structured monotone variational inequalities. Recently, He et al. [30] considered alternating projection-based prediction-correction methods for structured variational inequalities. All of the work above, however, was devoted to vector optimization problems. It appears to be new to apply the idea of ADM to develop methods for solving MSDP problems.

3. MODIFIED ALTERNATING DIRECTION METHODS AND THEIR CONVERGENCE ANALYSIS

If we implement the original ADM for solving MSDP problems, we would have to solve sub-variational inequality problems on matrix spaces at each iteration. Although there are a number of methods for solving monotone variational inequalities, in many occasions it is not an easy task. As a matter of fact, there seems to be little justification on the effort of obtaining the solutions of these sub-problems at each iteration. Therefore, we modify the original ADM to make the implementation of each iteration much easier. Specifically, after the modification, the main computational load of each iteration is only the metric projections onto convex sets in the matrix space. Thus, the proposed modified ADMs are simple and easy to implement. They belong to inexact ADM in nature because we solve each iteration of the original ADM only approximately after the modification. Although generally inspired by the research of inexact ADM [11, 19, 29, 30], the procedures here are different because of special operations for matrices.

We will consider to modify ADM for monotropic quadratic semidefinite programming (MQSDP) and monotropic nonlinear semidefinite programming (MNLSDP), separately. The reason for doing so is that the quadratic case allows a more specific modification that roughly requires only half of the workload, compared to the general case. For MNLSDP problems with general nonlinear objective functions, the procedure is more complicated. In fact, it is necessary to call on a correction phase to produce the new iterate based on a predictor computed in the prediction phase. For the two different modifications, we give detailed convergence analysis under some mild conditions. It is proved that the distance between iterative point and optimal point is monotonically decreasing at each iteration.

3.1 The Modified Alternating Direction Method for Monotropic Quadratic Semidefinite Programming

In the following, we will modify the ADM into an algorithm for solving MQSDP problems. The matrix convex quadratic function has the general form

$$f(X) = \langle X, F(X) \rangle,$$

where $F : \Re^{p\times p} \to \Re^{p\times p}$ is a self-adjoint positive semidefinite linear operator. Then its first order derivative is $F(X)$. In the monotropic case the objective function is

$$\sum_{i=1}^{n} f_i(X_i) = \sum_{i=1}^{n} \langle X_i, F_i(X_i) \rangle,$$

and $\nabla f_i(X_i) = F_i(X_i)$.

At Step 1 and Step 2 of Algorithm 2.2.1, we should solve variational inequalities in matrix spaces which might be a hard job. Thus we hope to convert them to simpler projection operations through some proper modifications. We now design a modified ADM based on certain good properties of quadratic functions and prove its convergence.

Similar to the classical variational inequality [16], it is easy to see that (2.4) is equivalent to the following nonlinear equation

$$X_i^{k+1} = P_{\Omega_{i1}}\left[X_i^{k+1} - \alpha_{i0}\left(\mu_{i0}F_i\left(X_i^{k+1}\right) - \sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k - \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^k\right)\right)\right)\right], \tag{3.1}$$

where α_{i0} can be any positive number. However, it is generally impossible to select an α_{i0} so that the X_i^{k+1}s on the right hand side are cancelled. We therefore suggest to solve (3.1) approximately. Let

$$\begin{aligned} R_{i0}\left(X_i^k, X_i^{k+1}\right) \equiv{}& \mu_{i0}\left(F_i\left(X_i^{k+1}\right) - F_i\left(X_i^k\right)\right) + \sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^{k+1}\right) \\ & - \sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right) - \gamma_{i0}\left(X_i^{k+1} - X_i^k\right) \end{aligned} \tag{3.2}$$

for certain constant γ_{i0}. Given a self-adjoint linear operator $\mathcal{V}$ defined on a finite dimensional inner product space, we let $\lambda_{\max}(\mathcal{V})$ be its largest eigenvalue. Note that $|\lambda_{\max}(\mathcal{V})| = \|\mathcal{V}\| \equiv \max\{\|\mathcal{V}(M)\| : \|M\| \leq 1\}$. We choose γ_{i0} so that

$$\gamma_{i0} \geq \lambda_{\max}\left(\mu_{i0}F_i + \sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\right).$$

$R_{i0}\left(X_i^k, X_i^{k+1}\right)$ can be seen as an approximate term and will converge to 0 as X_i^k converges. With adding it in the projection of (3.1), we obtain a new formula for updating X_i as follows. It is still denoted as X_i^{k+1} for simplicity. However, remember that it is defined different from (3.1) and only solves (3.1) approximately.

$$\begin{aligned} X_i^{k+1} &= P_{\Omega_{i1}}\Bigg[X_i^{k+1} - \alpha_{i0}\Bigg(\mu_{i0}F_i\left(X_i^{k+1}\right) \\ &\quad -\sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k - \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^k\right)\right) - R_{i0}\left(X_i^k, X_i^{k+1}\right)\Bigg)\Bigg] \\ &= P_{\Omega_{i1}}\Bigg[X_i^{k+1} - \alpha_{i0}\Bigg(\gamma_{i0}X_i^{k+1} + \mu_{i0}F_i\left(X_i^k\right) - \gamma_{i0}X_i^k \\ &\quad +\sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\circ\mathcal{L}_{ij}\left(X_i^k\right) - \sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k + \beta_{ij}\left(Y_{ij}^k\right)\right)\Bigg)\Bigg]. \end{aligned} \tag{3.3}$$

Setting

$$\alpha_{i0} = \frac{1}{\gamma_{i0}},$$

and

$$D_{i0} = \mu_{i0}F_i\left(X_i^k\right) + \sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\circ\mathcal{L}_{ij}\left(X_i^k\right) - \gamma_{i0}X_i^k - \sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k + \beta_{ij}\left(Y_{ij}^k\right)\right),$$

we obtain

$$X_i^{k+1} = P_{\Omega_{i1}}\left[-\alpha_{i0}D_{i0}\right]. \tag{3.4}$$

Similarly, we can also find a solution to (2.6) by computing

$$Y_{ij}^{k+1} = P_{\Omega'_{ij}}\left[Y_{ij}^{k+1} - \alpha_{ij}\left(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) + \lambda_{ij}^k - \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^{k+1}\right)\right)\right], \tag{3.5}$$

where α_{ij} can be any positive number. Let the approximate term

$$\begin{aligned} R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right) \quad \equiv \quad & \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) + \beta_{ij}\left(Y_{ij}^{k+1}\right) - \beta_{ij}\left(Y_{ij}^k\right) \\ & -\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^k\right) - \gamma_{ij}\left(Y_{ij}^{k+1} - Y_{ij}^k\right) \end{aligned} \tag{3.6}$$

for certain constant γ_{ij} such that

$$\gamma_{ij} \geq \lambda_{\max}\left(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1} + \beta_{ij}\right).$$

Then we have the following formula for approximately solving (3.5), but still denote it as Y_{ij}^{k+1} for simplicity.

$$\begin{aligned} Y_{ij}^{k+1} \quad = \quad & P_{\Omega'_{ij}}\Big[Y_{ij}^{k+1} - \alpha_{ij}\Big(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) + \lambda_{ij}^k \\ & -\beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^{k+1}\right) - R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right)\Big)\Big] \\ = \quad & P_{\Omega'_{ij}}\Big[Y_{ij}^{k+1} - \alpha_{ij}\Big(\gamma_{ij} Y_{ij}^{k+1} + \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^k\right) \\ & +\beta_{ij}\left(Y_{ij}^k\right) - \gamma_{ij} Y_{ij}^k + \lambda_{ij}^k - \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^{k+1}\right)\Big)\Big]. \end{aligned} \tag{3.7}$$

By setting

$$\alpha_{ij} = \frac{1}{\gamma_{ij}}$$

and

$$D_{ij} = \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^k\right) + \beta_{ij}\left(Y_{ij}^k\right) - \gamma_{ij} Y_{ij}^k + \lambda_{ij}^k - \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^{k+1}\right), \tag{3.8}$$

we obtain

$$Y_{ij}^{k+1} = P_{\Omega'_{ij}}\left[-\alpha_{ij} D_{ij}\right]. \tag{3.9}$$

There is some difference for the process of (2.5). Define the following approximate term

$$\begin{aligned} R_{i1}\left(Y_{i1}^k, Y_{i1}^{k+1}\right) \equiv\ & \mu_{i1}\left(\mathcal{L}_{i1}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{i1}^{-1}\left(Y_{i1}^{k+1}\right) + \beta_{i1}\left(Y_{i1}^{k+1}\right) - \beta_{i1}\left(Y_{i1}^k\right) \\ & -\mu_{i1}\left(\mathcal{L}_{i1}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{i1}^{-1}\left(Y_{i1}^k\right) - \gamma_{i1}\left(Y_{i1}^{k+1} - Y_{i1}^k\right) \end{aligned} \tag{3.10}$$

for certain constant γ_{i1}. However, here we need an additional requirement $\gamma_{11} = \gamma_{21} = \cdots = \gamma_{n1}$. Thus the choice of γ_{11} is restricted to

$$\gamma_{11} \geq \max_{i=1,\cdots,n}\left\{\lambda_{\max}\left(\mu_{i1}\left(\mathcal{L}_{i1}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{i1}^{-1} + \beta_{i1}\right)\right\}.$$

According to this, the approximate solution of (2.5) is

$$\begin{aligned}
\left(Y_{11}^{k+1},\cdots,Y_{n1}^{k+1}\right) &= P_{\Omega'}\Bigg[\left(Y_{11}^{k+1},\cdots,Y_{n1}^{k+1}\right) \\
&\quad -\alpha_{11}\Bigg(\mu_{11}\left(\mathcal{L}_{11}^{-1}\right)^T\circ F_1\circ\mathcal{L}_{11}^{-1}\left(Y_{11}^{k+1}\right)+\lambda_{11}^k \\
&\quad -\beta_{11}\left(\mathcal{L}_{11}\left(X_1^{k+1}\right)-Y_{11}^{K+1}\right)-R_{11}\left(Y_{11}^k,Y_{11}^{k+1}\right), \\
&\quad \cdots,\mu_{n1}\left(\mathcal{L}_{n1}^{-1}\right)^T\circ F_n\circ\mathcal{L}_{n1}^{-1}\left(Y_{n1}^{k+1}\right)+\lambda_{n1}^k \\
&\quad -\beta_{n1}\left(\mathcal{L}_{n1}\left(X_n^{k+1}\right)-Y_{n1}^{K+1}\right)-R_{n1}\left(Y_{n1}^k,Y_{n1}^{k+1}\right)\Bigg)\Bigg] \\
&= P_{\Omega'}\left[\left(Y_{11}^{k+1},\cdots,Y_{n1}^{k+1}\right)\right. \\
&\quad \left.-\alpha_{11}\left(\gamma_{11}Y_{11}^{k+1}+D_{11},\cdots,\gamma_{11}Y_{n1}^{k+1}+D_{n1}\right)\right], \quad (3.11)
\end{aligned}$$

where the definition of D_{i1}, $i=1,\cdots,n$, is the same with that of (3.8). Setting $\alpha_{11}=\frac{1}{\gamma_{11}}$, we have

$$\left(Y_{11}^{k+1},\cdots,Y_{n1}^{k+1}\right)=P_{\Omega'}\left[-\alpha_{11}\left(D_{11},\cdots,D_{n1}\right)\right]. \quad (3.12)$$

In summary, the modified ADM is given as follows.

Algorithm 3.1.1. The Modified ADM for MQSDP

Do at each iteration until a stopping criterion is met

Step 1. $\left(X_i^k,Y_{ij}^k,\lambda_{ij}^k\right)\to\left(X_i^{k+1},Y_{ij}^k,\lambda_{ij}^k\right)$, $i=1,\cdots,n$, *where*

$$X_i^{k+1}=P_{\Omega_{i1}}\left[-\alpha_{i0}D_{i0}\right] \quad (3.13)$$

Step 2. $\left(X_i^{k+1}, Y_{ij}^k, \lambda_{ij}^k\right) \rightarrow \left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^k\right),\ i = 1, \cdots, n,\ j = 1, \cdots, m,$ *where*

$$\left(Y_{11}^{k+1}, \cdots, Y_{n1}^{k+1}\right) \quad = \quad P_{\Omega'}\left[-\alpha_{11}\left(D_{11}, \cdots, D_{n1}\right)\right] \tag{3.14}$$

and

$$Y_{ij}^{k+1} = P_{\Omega_{ij}'}\left[-\alpha_{ij} D_{ij}\right],\ i = 1, \cdots, n,\ j = 2, \cdots, m \tag{3.15}$$

Step 3. $\left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^k\right) \rightarrow \left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^{k+1}\right), i = 1, \cdots, n,\ j = 1, \cdots, m,$ *where*

$$\lambda_{ij}^{k+1} = \lambda_{ij}^k - \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^{k+1}\right) \tag{3.16}$$

Note that the major computation in Algorithm 3.1.1 is the metric projections onto the convex sets $\Omega_{i1},\ \Omega'$, and Ω_{ij}'. Compared with the original ADM, the computation is much simplified. In the following, we will prove a convergence result. Firstly, we prove an important proposition.

Proposition 3.1.2. *The sequence* $\left\{X_i^k, Y_{ij}^k, \lambda_{ij}^k\right\}$ *generated by the modified*

ADM for MQSDP satisfies

$$\begin{aligned} &\sum_{i=1}^{n}\Bigg(\left\langle X_i^{k+1}-X_i^*, R_{i0}\left(X_i^k, X_i^{k+1}\right)\right\rangle+\sum_{j=1}^{m}\left\langle Y_{ij}^{k+1}-Y_{ij}^*, R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right)\right\rangle \\ &+\sum_{j=1}^{m}\left\langle \lambda_{ij}^{k+1}-\lambda_{ij}^*, \beta_{ij}^{-1}\left(\lambda_{ij}^k-\lambda_{ij}^{k+1}\right)\right\rangle\Bigg) \\ \geq\ & \sum_{i=1}^{n}\left\langle X_i^{k+1}-X_i^*, \sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\left(Y_{ij}^{k+1}-Y_{ij}^k\right)\right\rangle, \end{aligned} \tag{3.17}$$

where $\left(X_i^*, Y_{ij}^*, \lambda_{ij}^*\right)$ *are defined as in (1.5).*

Proof. Note that (3.3) can be written equivalently as

$$\begin{aligned} &\left\langle X_i-X_i^{k+1}, \mu_{i0}F_i\left(X_i^{k+1}\right)-\sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^{k+1}+\beta_{ij}\left(Y_{ij}^k-Y_{ij}^{k+1}\right)\right)-R_{i0}\left(X_i^k, X_i^{k+1}\right)\right\rangle \\ \geq\ & 0,\ \forall X_i\in\Omega_{i1}. \end{aligned}$$

Setting $X_i=X_i^*$ in it, we obtain

$$\left\langle X_i^{k+1}-X_i^*, -\mu_{i0}F_i\left(X_i^{k+1}\right)+\sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^{k+1}+\beta_{ij}\left(Y_{ij}^k-Y_{ij}^{k+1}\right)\right)+R_{i0}\left(X_i^k, X_i^{k+1}\right)\right\rangle\geq 0. \tag{3.18}$$

Let $X_i=X_i^{k+1}$ in inequality (1.5). Then

$$\left\langle X_i^{k+1}-X_i^*, \mu_{i0}F_i\left(X_i^*\right)-\sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^*\right)\right\rangle\geq 0. \tag{3.19}$$

Adding (3.18) and (3.19) together, it follows that

$$\begin{aligned}
&\left\langle X_i^{k+1} - X_i^*, \sum_{j=1}^{m} \mathcal{L}_{ij}^T \left(\lambda_{ij}^{k+1} - \lambda_{ij}^*\right) + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \left(Y_{ij}^k - Y_{ij}^{k+1}\right) \right\rangle \\
&+ \left\langle X_i^{k+1} - X_i^*, R_{i0}\left(X_i^k, X_i^{k+1}\right)\right\rangle \\
\geq\ & \left\langle X_i^{k+1} - X_i^*, \mu_{i0} F_i\left(X_i^{k+1}\right) - \mu_{i0} F_i\left(X_i^*\right)\right\rangle \geq 0.
\end{aligned} \tag{3.20}$$

Note that (3.7) can be written equivalently as

$$\begin{aligned}
&\left\langle Y_{ij} - Y_{ij}^{k+1}, \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) + \lambda_{ij}^{k+1} - R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right)\right\rangle \\
\geq\ & 0,\ \forall Y_{ij} \in \Omega_{ij}^{'}.
\end{aligned}$$

Setting $Y_{ij} = Y_{ij}^*$, we obtain

$$\left\langle Y_{ij}^{k+1} - Y_{ij}^*, -\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) - \lambda_{ij}^{k+1} + R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right)\right\rangle \geq 0. \tag{3.21}$$

Let $Y_{ij} = Y_{ij}^{k+1}$ in inequality (1.5). Then

$$\left\langle Y_{ij}^{k+1} - Y_{ij}^*, \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^*\right) + \lambda_{ij}^*\right\rangle \geq 0. \tag{3.22}$$

Adding (3.21) and (3.22) together, it follows that

$$\begin{aligned} & \left\langle Y_{ij}^{k+1} - Y_{ij}^{*}, \lambda_{ij}^{*} - \lambda_{ij}^{k+1} + R_{ij}\left(Y_{ij}^{k}, Y_{ij}^{k+1}\right)\right\rangle \\ \geq\ & \left\langle Y_{ij}^{k+1} - Y_{ij}^{*}, \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^{T} \circ F_{i} \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{k+1}\right) - \mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^{T} \circ F_{i} \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^{*}\right)\right\rangle \\ \geq\ & 0. \end{aligned} \tag{3.23}$$

Similarly, there holds

$$\sum_{i=1}^{n} \left\langle Y_{i1}^{k+1} - Y_{i1}^{*}, \lambda_{i1}^{*} - \lambda_{i1}^{k+1} + R_{i1}\left(Y_{i1}^{k}, Y_{i1}^{k+1}\right)\right\rangle \geq 0. \tag{3.24}$$

It follows from (3.20), (3.23), and (3.24) that

$$\begin{aligned} & \sum_{i=1}^{n}\left\langle X_{i}^{k+1} - X_{i}^{*}, \sum_{j=1}^{m}\mathcal{L}_{ij}^{T}\left(\lambda_{ij}^{k+1} - \lambda_{ij}^{*}\right) + \sum_{j=1}^{m}\mathcal{L}_{ij}^{T}\circ\beta_{ij}\left(Y_{ij}^{k} - Y_{ij}^{k+1}\right) + R_{i0}\left(X_{i}^{k}, X_{i}^{k+1}\right)\right\rangle \\ & + \sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle Y_{ij}^{k+1} - Y_{ij}^{*}, \lambda_{ij}^{*} - \lambda_{ij}^{k+1} + R_{ij}\left(Y_{ij}^{k}, Y_{ij}^{k+1}\right)\right\rangle \\ =\ & \sum_{i=1}^{n}\left\langle X_{i}^{k+1} - X_{i}^{*}, R_{i0}\left(X_{i}^{k}, X_{i}^{k+1}\right)\right\rangle + \sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle Y_{ij}^{k+1} - Y_{ij}^{*}, R_{ij}\left(Y_{ij}^{k}, Y_{ij}^{k+1}\right)\right\rangle \\ & + \sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \lambda_{ij}^{k+1} - \lambda_{ij}^{*}, \beta_{ij}^{-1}\left(\lambda_{ij}^{k} - \lambda_{ij}^{k+1}\right)\right\rangle \\ & + \sum_{i=1}^{n}\left\langle X_{i}^{k+1} - X_{i}^{*}, \sum_{j=1}^{m}\mathcal{L}_{ij}^{T}\circ\beta_{ij}\left(Y_{ij}^{k} - Y_{ij}^{k+1}\right)\right\rangle \\ \geq\ & 0. \end{aligned}$$

The proof is complete. $\square$

We next prove the convergence theorem of the modified ADM for MQSDP.

Theorem 3.1.3. *The sequence* $\left\{X_i^k\right\}$ *generated by the modified ADM for MQSDP converges to a solution point* X_i^* *of system (1.5).*

Proof. We denote

$$W \equiv \begin{pmatrix} X_i \\ Y_{ij} \\ \lambda_{ij} \end{pmatrix}, \; G \equiv \begin{pmatrix} \overline{R}_{i0} & 0 & 0 \\ 0 & \overline{R}_{ij} & 0 \\ 0 & 0 & \beta_{ij}^{-1} \end{pmatrix}, \; G^{'} \equiv \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_{ij} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\overline{R}_{i0} = \gamma_{i0} I - \mu_{i0} F_i - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}$, $\overline{R}_{ij} = \gamma_{ij} I - \mu_{ij} \left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1} - \beta_{ij}$, and I is the identical operator with $I(M) = M$. Because of the choice of γ_{i0} and γ_{ij}, clearly G and $G^{'}$ are positive semidefinite. We define the G-inner product and $G^{'}$-inner product of W and $W^{'}$ respectively as

$$\langle W, W' \rangle_G \;\; \equiv \;\; \sum_{i=1}^{n} \left(\left\langle X_i, \overline{R}_{i0}\left(X_i^{'}\right) \right\rangle + \sum_{j=1}^{m} \left\langle Y_{ij}, \overline{R}_{ij}\left(Y_{ij}^{'}\right) \right\rangle + \sum_{j=1}^{m} \left\langle \lambda_{ij}, \beta_{ij}^{-1}\left(\lambda_{ij}^{'}\right) \right\rangle \right),$$

$$\langle W, W' \rangle_{G'} \;\; \equiv \;\; \sum_{i=1}^{n} \sum_{j=1}^{m} \left\langle Y_{ij}, \beta_{ij}\left(Y_{ij}^{'}\right) \right\rangle,$$

and the associated G-norm and $G^{'}$-norm respectively as

$$\|W\|_G \equiv \left(\sum_{i=1}^{n} \left(\|X_i\|_{\overline{R}_{i0}}^2 + \sum_{j=1}^{m} \|Y_{ij}\|_{\overline{R}_{ij}}^2 + \sum_{j=1}^{m} \|\lambda_{ij}\|_{\beta_{ij}^{-1}}^2 \right) \right)^{\frac{1}{2}},$$

$$\|W\|_{G'} \equiv \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \|Y_{ij}\|_{\beta_{ij}}^2 \right)^{\frac{1}{2}},$$

where $\|\cdot\|_{\overline{R}_{i0}}^2 \equiv \gamma_{i0} \|\cdot\|^2 - \mu_{i0} \langle \cdot, F_i(\cdot) \rangle - \left\langle \cdot, \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}(\cdot) \right\rangle$, $\|\cdot\|_{\overline{R}_{ij}}^2 \equiv \gamma_{ij} \|\cdot\|^2 - \mu_{ij} \left\langle \cdot, \left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}(\cdot) \right\rangle - \langle \cdot, \beta_{ij}(\cdot) \rangle$, $\|\cdot\|_{\beta_{ij}}^2 \equiv \langle \cdot, \beta_{ij}(\cdot) \rangle$, and $\|\cdot\|_{\beta_{ij}^{-1}}^2 \equiv \left\langle \cdot, \beta_{ij}^{-1}(\cdot) \right\rangle$. Based on these, we define the $G + G'$-inner product of W and W' as

$$\langle W, W' \rangle_{G+G'} \equiv \langle W, W' \rangle_G + \langle W, W' \rangle_{G'},$$

so that the associated $G + G'$-norm is

$$\|W\|_{G+G'} \equiv \left(\|W\|_G^2 + \|W\|_{G'}^2 \right)^{\frac{1}{2}}.$$

Note that $R_{i0}\left(X_i^k, X_i^{k+1}\right) = \overline{R}_{i0}\left(X_i^k\right) - \overline{R}_{i0}\left(X_i^{k+1}\right)$ and $R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right) = \overline{R}_{ij}\left(Y_{ij}^k\right) - \overline{R}_{ij}\left(Y_{ij}^{k+1}\right)$, then (3.17) can be written as

$$\begin{aligned} & \left\langle W^{k+1} - W^*, W^k - W^{k+1} \right\rangle_G \\ \geq \; & \sum_{i=1}^{n} \left\langle X_i^{k+1} - X_i^*, \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij}\left(Y_{ij}^{k+1} - Y_{ij}^k\right) \right\rangle. \end{aligned} \tag{3.25}$$

Observe that solving the optimal condition (1.5) is equivalent to finding a zero of the residual function

$$\|e(W)\| \equiv \left\| \begin{array}{c} X_i - P_{\Omega_{i1}}\left[X_i - \alpha_{i0}\left(\mu_{i0}F_i(X_i) - \sum_{j=1}^{m}\mathcal{L}_{ij}^T(\lambda_{ij})\right)\right] \\ (Y_{11},\cdots,Y_{n1}) - P_{\Omega'}\left[(Y_{11},\cdots,Y_{n1}) - \alpha_{11}\right. \\ \left.\left(\mu_{11}\left(\mathcal{L}_{11}^{-1}\right)^T \circ F_1 \circ \mathcal{L}_{11}^{-1}(Y_{11}) + \lambda_{11},\cdots,\mu_{n1}\left(\mathcal{L}_{n1}^{-1}\right)^T \circ F_n \circ \mathcal{L}_{n1}^{-1}(Y_{n1}) + \lambda_{n1}\right)\right] \\ Y_{ij} - P_{\Omega'_{ij}}\left[Y_{ij} - \alpha_{ij}\left(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}(Y_{ij}) + \lambda_{ij}\right)\right] \\ \beta_{ij}\left(\mathcal{L}_{ij}(X_i) - Y_{ij}\right) \end{array} \right\|.$$

Then we have from (3.3), (3.7), and (3.11) that

$$\begin{aligned} &\left\|e(W^{k+1})\right\|^2 \\ \leq\ & \left\| \begin{array}{c} \alpha_{i0}\left(\sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij}\left(Y_{ij}^k - Y_{ij}^{k+1}\right) + R_{i0}\left(X_i^k, X_i^{k+1}\right)\right) \\ \alpha_{ij}R_{ij}\left(Y_{ij}^k, Y_{ij}^{k+1}\right) \\ \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^k\right) + \beta_{ij}\left(Y_{ij}^k - Y_{ij}^{k+1}\right) \end{array} \right\|^2 \\ =\ & \left\| \begin{array}{c} \alpha_{i0}\left(\sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij}\left(Y_{ij}^k - Y_{ij}^{k+1}\right) + \overline{R}_{i0}\left(X_i^k\right) - \overline{R}_{i0}\left(X_i^{k+1}\right)\right) \\ \alpha_{ij}\left(\overline{R}_{ij}\left(Y_{ij}^k\right) - \overline{R}_{ij}\left(Y_{ij}^{k+1}\right)\right) \\ \beta_{ij}\left(\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^k\right) + \beta_{ij}\left(Y_{ij}^k - Y_{ij}^{k+1}\right) \end{array} \right\|^2 \\ \leq\ & \delta\sum_{i=1}^{n}\left(\left\|X_i^{k+1} - X_i^k\right\|_{\overline{R}_{i0}}^2 + \sum_{j=1}^{m}\left\|Y_{ij}^{k+1} - Y_{ij}^k\right\|_{\overline{R}_{ij}}^2\right. \\ & \left. + \sum_{j=1}^{m}\left\|\mathcal{L}_{ij}\left(X_i^{k+1}\right) - Y_{ij}^k\right\|_{\beta_{ij}}^2\right), \end{aligned} \tag{3.26}$$

where δ is a positive constant.

Thus,

$$
\begin{aligned}
& \left\|W^{k+1}-W^*\right\|_{G+G'}^2 \\
= \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \left\|W^{k+1}-W^k\right\|_{G+G'}^2 - 2\left\langle W^{k+1}-W^*, W^k-W^{k+1}\right\rangle_{G+G'} \\
= \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \left\|W^{k+1}-W^k\right\|_{G+G'}^2 - 2\left\langle W^{k+1}-W^*, W^k-W^{k+1}\right\rangle_{G} \\
& -2\left\langle W^{k+1}-W^*, W^k-W^{k+1}\right\rangle_{G'} \\
\leq^{(3.25)} \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - 2\sum_{i=1}^{n}\left\langle X_i^{k+1}-X_i^*, \sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\left(Y_{ij}^{k+1}-Y_{ij}^k\right)\right\rangle \\
& -\left\|W^{k+1}-W^k\right\|_{G+G'}^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle Y_{ij}^{k+1}-Y_{ij}^*, \beta_{ij}\left(Y_{ij}^k-Y_{ij}^{k+1}\right)\right\rangle \\
= \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \left\|W^{k+1}-W^k\right\|_{G+G'}^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \lambda_{ij}^{k+1}-\lambda_{ij}^k, Y_{ij}^{k+1}-Y_{ij}^k\right\rangle \\
= \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \sum_{i=1}^{n}\left(\left\|X_i^{k+1}-X_i^k\right\|_{\overline{R}_{i0}}^2 - 2\sum_{j=1}^{m}\left\langle \lambda_{ij}^{k+1}-\lambda_{ij}^k, Y_{ij}^{k+1}-Y_{ij}^k\right\rangle\right. \\
& \left.+\sum_{j=1}^{m}\left\|Y_{ij}^{k+1}-Y_{ij}^k\right\|_{\overline{R}_{ij}}^2 + \sum_{j=1}^{m}\left\|Y_{ij}^{k+1}-Y_{ij}^k\right\|_{\beta_{ij}}^2 + \sum_{j=1}^{m}\left\|\lambda_{ij}^{k+1}-\lambda_{ij}^k\right\|_{\beta_{ij}^{-1}}^2\right) \\
= \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \sum_{i=1}^{n}\left(\left\|X_i^{k+1}-X_i^k\right\|_{\overline{R}_{i0}}^2 + \sum_{j=1}^{m}\left\|Y_{ij}^{k+1}-Y_{ij}^k\right\|_{\overline{R}_{ij}}^2\right. \\
& \left.+\sum_{j=1}^{m}\left\|\lambda_{ij}^{k+1}-\lambda_{ij}^k-\beta_{ij}\left(Y_{ij}^{k+1}-Y_{ij}^k\right)\right\|_{\beta_{ij}^{-1}}^2\right) \\
= \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \sum_{i=1}^{n}\left(\left\|X_i^{k+1}-X_i^k\right\|_{\overline{R}_{i0}}^2 + \sum_{j=1}^{m}\left\|Y_{ij}^{k+1}-Y_{ij}^k\right\|_{\overline{R}_{ij}}^2\right. \\
& \left.+\sum_{j=1}^{m}\left\|\mathcal{L}_{ij}\left(X_i^{k+1}\right)-Y_{ij}^k\right\|_{\beta_{ij}}^2\right) \\
\leq^{(3.26)} \quad & \left\|W^{k}-W^*\right\|_{G+G'}^2 - \frac{1}{\delta}\left\|e\left(W^{k+1}\right)\right\|^2.
\end{aligned}
\tag{3.27}
$$

From the above inequality, we have

$$\left\|W^{k+1} - W^*\right\|_{G+G'}^2 \leq \left\|W^k - W^*\right\|_{G+G'}^2 \leq \cdots \leq \left\|W^0 - W^*\right\|_{G+G'}^2. \quad (3.28)$$

That is, the sequence $\left\{W^k\right\}$ is bounded. Thus there exists at least one cluster point of $\left\{W^k\right\}$.

It also follows from (3.27) that

$$\sum_{k=0}^{\infty} \frac{1}{\delta} \left\|e\left(W^{k+1}\right)\right\|^2 < +\infty.$$

This implies that

$$\lim_{k\to\infty} \left\|e\left(W^k\right)\right\| = 0.$$

Let $\overline{W}$ be a cluster point of $\left\{W^k\right\}$, and let $\left\{W^{k_j}\right\}$ be a corresponding subsequence converging to $\overline{W}$. Then,

$$\left\|e\left(\overline{W}\right)\right\| = \lim_{j\to\infty} \left\|e\left(W^{k_j}\right)\right\| = 0,$$

which means that $\overline{W}$ is a zero of the residual function. Therefore $\overline{W}$ satisfies (1.5). Setting $W^* = \overline{W}$ in (3.28), we have

$$\left\|W^{k+1} - \overline{W}\right\|_{G+G'}^2 \leq \left\|W^k - \overline{W}\right\|_{G+G'}^2, \ \forall k \geq 0.$$

Thus, the sequence $\{W^k\}$ has a unique cluster point and

$$\lim_{k\to\infty} W^k = \overline{W}.$$

This completes the proof. $\square$

3.2 The Prediction-Correction Alternating Direction Method for Monotropic Nonlinear Semidefinite Programming

In the following, we will modify the ADM for solving MNLSDP problems with general nonlinear objective $\sum_{i=1}^{n} f_i(X_i)$. Assume each $f_i(X_i)$, $i = 1, \cdots, n$, is continuously differentiable with the first order derivative $\nabla f_i(X_i) = F_i(X_i)$. Furthermore, we require the operator $F_i(\cdot)$, $i = 1, \cdots, n$, to be Lipschitz continuous on $\Omega_{i1} \cap \Omega_{i2} \cap \cdots \cap \Omega_{im}$ with a constant L_i, respectively.

Here the basic consideration of modifying ADM is still to remove difficult matrix variational inequalities at each iteration. Unlike the MQSDP case there are no simple formulas like (3.13)-(3.15) for computing X_i^{k+1} and Y_{ij}^{k+1}. In order to remove the implicit components in (2.4)-(2.6), we propose a more complicated prediction-correction ADM and prove its convergence.

We suggest the following approximate approaches for Step 1 and Step 2 of Algorithm 2.2.1. For $X_i,\ i=1,\cdots,n$, introduce the term

$$\begin{aligned} R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right) &\equiv X_i^k-\widetilde{X}_i^k-\alpha_{i0}^{'}\Bigg(\mu_{i0}\left(F_i\left(X_i^k\right)-F_i\left(\widetilde{X}_i^k\right)\right) \\ &\quad +\sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\circ\mathcal{L}_{ij}\left(X_i^k\right)-\sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\circ\mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)\Bigg), \end{aligned}$$

where positive scalar $\alpha_{i0}^{'}$ is chosen so that

$$\alpha_{i0}^{'}\le\frac{\eta}{\mu_{i0}L_i+\sum_{j=1}^{m}\left\|\mathcal{L}_{ij}^T\right\|\|\beta_{ij}\|\|\mathcal{L}_{ij}\|},\ i=1,\cdots,n, \tag{3.29}$$

with a fixed $0<\eta<1$. Then

$$\begin{aligned} \widetilde{X}_i^k &= P_{\Omega_{i1}}\Bigg[\widetilde{X}_i^k-\alpha_{i0}^{'}\Bigg(\mu_{i0}F_i\left(\widetilde{X}_i^k\right) \\ &\quad -\sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k-\beta_{ij}\left(\mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)-Y_{ij}^k\right)\right)\Bigg)+R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right)\Bigg] \\ &= P_{\Omega_{i1}}\Bigg[X_i^k-\alpha_{i0}^{'}\Bigg(\mu_{i0}F_i\left(X_i^k\right)+\sum_{j=1}^{m}\mathcal{L}_{ij}^T\circ\beta_{ij}\circ\mathcal{L}_{ij}\left(X_i^k\right) \\ &\quad -\sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k+\beta_{ij}\left(Y_{ij}^k\right)\right)\Bigg)\Bigg]. \end{aligned} \tag{3.30}$$

Similarly for $Y_{ij},\ i=1,\cdots,n,\ j=2,\cdots,m$, we introduce

$$\begin{aligned} R_{ij}^{'}(Y_{ij}^k,\widetilde{Y}_{ij}^k) &\equiv Y_{ij}^k-\widetilde{Y}_{ij}^k-\alpha_{ij}^{'}\left(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T\circ F_i\circ\mathcal{L}_{ij}^{-1}\left(Y_{ij}^k\right)\right. \\ &\quad \left.-\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T\circ F_i\circ\mathcal{L}_{ij}^{-1}\left(\widetilde{Y}_{ij}^k\right)+\beta_{ij}\left(Y_{ij}^k\right)-\beta_{ij}\left(\widetilde{Y}_{ij}^k\right)\right), \end{aligned}$$

and choose positive scalar $\alpha_{ij}^{'}$ so that

$$\alpha_{ij}^{'} \leq \frac{1}{\lambda_{\max}\left(\beta_{ij}\right)}, \tag{3.31}$$

where $\lambda_{\max}\left(\beta_{ij}\right)$ is the largest eigenvalue of β_{ij} and

$$\alpha_{ij}^{'} \leq \frac{\eta}{\mu_{ij} L_i \left\| \left(\mathcal{L}_{ij}^{-1}\right)^T \right\| \left\| \mathcal{L}_{ij}^{-1} \right\| + \left\| \beta_{ij} \right\|}, \;\; i = 1, \cdots, n, \;\; j = 1, \cdots, m, \tag{3.32}$$

where $0 < \eta < 1$. Adding this approximate term,

$$\begin{aligned} \widetilde{Y}_{ij}^k &= P_{\Omega_{ij}^{'}} \Big[\widetilde{Y}_{ij}^k - \alpha_{ij}^{'} \Big(\mu_{ij} \left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1} \left(\widetilde{Y}_{ij}^k \right) + \lambda_{ij}^k \\ &\quad - \beta_{ij} \left(\mathcal{L}_{ij} \left(\widetilde{X}_i^k \right) - \widetilde{Y}_{ij}^k \right) \Big) + R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) \Big] \\ &= P_{\Omega_{ij}^{'}} \Big[Y_{ij}^k - \alpha_{ij}^{'} \Big(\mu_{ij} \left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1} \left(Y_{ij}^k \right) + \lambda_{ij}^k \\ &\quad - \beta_{ij} \left(\mathcal{L}_{ij} \left(\widetilde{X}_i^k \right) - Y_{ij}^k \right) \Big) \Big]. \end{aligned} \tag{3.33}$$

Lastly, with the additional requirement $\alpha_{11}^{'} = \alpha_{21}^{'} = \cdots = \alpha_{n1}^{'}$ for the positive scalars $\alpha_{i1}^{'}$, define the following approximate term for (2.5).

$$\begin{aligned} R_{i1}^{'} \left(Y_{i1}^k, \widetilde{Y}_{i1}^k \right) &\equiv Y_{i1}^k - \widetilde{Y}_{i1}^k - \alpha_{11}^{'} \Big(\mu_{i1} \left(\mathcal{L}_{i1}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{i1}^{-1} \left(Y_{i1}^k \right) \\ &\quad - \mu_{i1} \left(\mathcal{L}_{i1}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{i1}^{-1} \left(\widetilde{Y}_{i1}^k \right) + \beta_{i1} \left(Y_{i1}^k \right) - \beta_{i1} \left(\widetilde{Y}_{i1}^k \right) \Big). \end{aligned}$$

Then the approximate solution of (2.5) is equivalent to

$$
\begin{aligned}
& \left(\widetilde{Y}_{11}^k, \cdots, \widetilde{Y}_{n1}^k\right) \\
= \ & P_{\Omega'}\Big[\left(\widetilde{Y}_{11}^k, \cdots, \widetilde{Y}_{n1}^k\right) + \left(R_{11}^{'}\left(Y_{11}^k, \widetilde{Y}_{11}^k\right), \cdots, R_{n1}^{'}\left(Y_{n1}^k, \widetilde{Y}_{n1}^k\right)\right) - \alpha_{11}^{'} \\
& \Big(\mu_{11}\left(\mathcal{L}_{11}^{-1}\right)^T \circ F_1 \circ \mathcal{L}_{11}^{-1}\left(\widetilde{Y}_{11}^k\right) + \lambda_{11}^k - \beta_{11}\left(\mathcal{L}_{11}\left(\widetilde{X}_1^k\right) - \widetilde{Y}_{11}^k\right), \cdots, \\
& \mu_{n1}\left(\mathcal{L}_{n1}^{-1}\right)^T \circ F_n \circ \mathcal{L}_{n1}^{-1}\left(\widetilde{Y}_{n1}^k\right) + \lambda_{n1}^k - \beta_{n1}\left(\mathcal{L}_{n1}\left(\widetilde{X}_n^k\right) - \widetilde{Y}_{n1}^k\right)\Big)\Big] \\
= \ & P_{\Omega'}\Big[\left(Y_{11}^k, \cdots, Y_{n1}^k\right) - \alpha_{11}^{'} \\
& \Big(\mu_{11}\left(\mathcal{L}_{11}^{-1}\right)^T \circ F_1 \circ \mathcal{L}_{11}^{-1}\left(Y_{11}^k\right) + \beta_{11}\left(Y_{11}^k\right) + \lambda_{11}^k - \beta_{11} \circ \mathcal{L}_{11}\left(\widetilde{X}_1^k\right), \cdots, \\
& \mu_{n1}\left(\mathcal{L}_{n1}^{-1}\right)^T \circ F_n \circ \mathcal{L}_{n1}^{-1}\left(Y_{n1}^k\right) + \beta_{n1}\left(Y_{n1}^k\right) + \lambda_{n1}^k - \beta_{n1} \circ \mathcal{L}_{n1}\left(\widetilde{X}_n^k\right)\Big)\Big].
\end{aligned}
\tag{3.34}
$$

Till now all implicit parts within the projections have been successfully cancelled. However, we cannot prove the convergence by just doing so. Instead we use these as the predictor and will correct them in the correction phase.

Algorithm 3.2.1. The Prediction-Correction ADM for MNLSDP

Do at each iteration until a stopping criterion is met

The Prediction Phase:

Step 1. $\left(X_i^k, Y_{ij}^k, \lambda_{ij}^k\right) \to \left(\widetilde{X}_i^k, Y_{ij}^k, \lambda_{ij}^k\right)$, $i = 1, \cdots, n$, *where*

$$\begin{aligned}\widetilde{X}_i^k &= P_{\Omega_{i1}}\left[X_i^k - \alpha_{i0}^{'}\left(\mu_{i0}F_i\left(X_i^k\right) + \sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right)\right.\right.\\ &\left.\left.\qquad - \sum_{j=1}^{m}\mathcal{L}_{ij}^T\left(\lambda_{ij}^k + \beta_{ij}\left(Y_{ij}^k\right)\right)\right)\right] \qquad (3.35)\end{aligned}$$

Step 2. $\left(\widetilde{X}_i^k, Y_{ij}^k, \lambda_{ij}^k\right) \to \left(\widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \lambda_{ij}^k\right)$, $i = 1, \cdots, n$, $j = 1, \cdots, m$, *where*

$$\begin{aligned}&\left(\widetilde{Y}_{11}^k, \cdots, \widetilde{Y}_{n1}^k\right) = P_{\Omega'}\left[\left(Y_{11}^k, \cdots, Y_{n1}^k\right) - \alpha_{11}^{'}\right.\\ &\left(\mu_{11}\left(\mathcal{L}_{11}^{-1}\right)^T \circ F_1 \circ \mathcal{L}_{11}^{-1}\left(Y_{11}^k\right) + \beta_{11}\left(Y_{11}^k\right) + \lambda_{11}^k - \beta_{11} \circ \mathcal{L}_{11}\left(\widetilde{X}_1^k\right), \cdots,\right.\\ &\left.\left.\mu_{n1}\left(\mathcal{L}_{n1}^{-1}\right)^T \circ F_n \circ \mathcal{L}_{n1}^{-1}\left(Y_{n1}^k\right) + \beta_{n1}\left(Y_{n1}^k\right) + \lambda_{n1}^k - \beta_{n1} \circ \mathcal{L}_{n1}\left(\widetilde{X}_n^k\right)\right)\right]\end{aligned} \qquad (3.36)$$

and

$$\begin{aligned}\widetilde{Y}_{ij}^k &= P_{\Omega_{ij}^{'}}\left[Y_{ij}^k - \alpha_{ij}^{'}\left(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1}\left(Y_{ij}^k\right) + \lambda_{ij}^k\right.\right.\\ &\left.\left.\qquad -\beta_{ij}\left(\mathcal{L}_{ij}\left(\widetilde{X}_i^k\right) - Y_{ij}^k\right)\right)\right], \ i = 1, \cdots, n, \ j = 2, \cdots, m \qquad (3.37)\end{aligned}$$

Step 3. $\left(\widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \lambda_{ij}^k\right) \to \left(\widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right)$, $i = 1, \cdots, n$, $j = 1, \cdots, m$, *where*

$$\widetilde{\lambda}_{ij}^k = \lambda_{ij}^k - \beta_{ij}\left(\mathcal{L}_{ij}\left(\widetilde{X}_i^k\right) - \widetilde{Y}_{ij}^k\right) \qquad (3.38)$$

The Correction Phase:

Step 4. $\left(\widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right) \to \left(X_i^{k+1}, Y_{ij}^{k+1}, \lambda_{ij}^{k+1}\right)$, $i = 1, \cdots, n$, $j = 1, \cdots, m$, *where*

$$
\begin{aligned}
X_i^{k+1} &= P_{\Omega_{i1}}\left[X_i^k - \gamma^k R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right)\right] && (3.39)\\
\left(Y_{11}^{k+1}, \cdots, Y_{n1}^{k+1}\right) &= P_{\Omega'}\Big[\left(Y_{11}^k, \cdots, Y_{n1}^k\right) \\
&\quad -\gamma^k \alpha_{11}^{'}\left(\beta_{11}\left(Y_{11}^k - \widetilde{Y}_{11}^k\right), \cdots, \beta_{n1}\left(Y_{n1}^k - \widetilde{Y}_{n1}^k\right)\right) \\
&\quad -\gamma^k\left(R_{11}^{'}\left(Y_{11}^k, \widetilde{Y}_{11}^k\right), \cdots, R_{n1}^{'}\left(Y_{n1}^k, \widetilde{Y}_{n1}^k\right)\right)\Big] && (3.40)\\
Y_{ij}^{k+1} &= P_{\Omega_{ij}^{'}}\left[Y_{ij}^k - \gamma^k R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) - \gamma^k \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right)\right], \\
&\quad i = 1, \cdots, n, \ j = 2, \cdots, m && (3.41)\\
\lambda_{ij}^{k+1} &= \lambda_{ij}^k - \gamma^k\left(\lambda_{ij}^k - \widetilde{\lambda}_{ij}^k\right) && (3.42)
\end{aligned}
$$

The positive scalar $\gamma^k < \frac{1}{2}$ *is the step-length and its optimal choice will be given later.*

In order to solve MNLSDP problems by the prediction-correction ADM, we only need to compute the metric projections onto the convex sets Ω_{i1}, Ω', and $\Omega_{ij}^{'}$. Without the special structure of quadratic objective function in MQSDP, twice as many projections are necessary. However, compared with the original ADM the computation is simplified. In the following, we will prove a convergence result. Similar to Proposition 3.1.2, with the added approximate terms, there is an important proposition for the prediction phase.

Proposition 3.2.2. *The sequence* $\left\{X_i^k, Y_{ij}^k, \lambda_{ij}^k, \widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right\}$ *generated by the prediction-correction ADM for MNLSDP satisfies*

$$\begin{aligned}
&\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i0}^{'}}\left\langle \widetilde{X}_i^k - X_i^*, R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right)\right\rangle + \sum_{j=1}^{m}\frac{1}{\alpha_{ij}^{'}}\left\langle \widetilde{Y}_{ij}^k - Y_{ij}^*, R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right)\right\rangle\right. \\
&\left.+\sum_{j=1}^{m}\left\langle \widetilde{\lambda}_{ij}^k - \lambda_{ij}^*, \beta_{ij}^{-1}\left(\lambda_{ij}^k - \widetilde{\lambda}_{ij}^k\right)\right\rangle\right) \\
\geq\ &\sum_{i=1}^{n}\left\langle \widetilde{X}_i^k - X_i^*, \sum_{j=1}^{m}\mathcal{L}_{ij}^T \circ \beta_{ij}\left(\widetilde{Y}_{ij}^k - Y_{ij}^k\right)\right\rangle, \qquad (3.43)
\end{aligned}$$

where $\left(X_i^*, Y_{ij}^*, \lambda_{ij}^*\right)$ *are defined as in (1.5).*

Proof. Similar to the proof of Proposition 3.1.2. □

Corollary 3.2.3. *The sequence* $\left\{X_i^k, Y_{ij}^k, \lambda_{ij}^k, \widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right\}$ *generated by the prediction-correction ADM for MNLSDP satisfies*

$$\begin{aligned}
&\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i0}^{'}}\left\langle \widetilde{X}_i^k - X_i^*, R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right)\right\rangle + \sum_{j=1}^{m}\left\langle \widetilde{\lambda}_{ij}^k - \lambda_{ij}^*, \beta_{ij}^{-1}\left(\lambda_{ij}^k - \widetilde{\lambda}_{ij}^k\right)\right\rangle\right. \\
&\left.+\sum_{j=1}^{m}\frac{1}{\alpha_{ij}^{'}}\left\langle \widetilde{Y}_{ij}^k - Y_{ij}^*, R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right)\right\rangle\right) \\
\geq\ &\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k\right\rangle, \qquad (3.44)
\end{aligned}$$

where $\left(X_i^*, Y_{ij}^*, \lambda_{ij}^*\right)$ *are defined as in (1.5).*

Proof. Add $\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \widetilde{Y}_{ij}^k - Y_{ij}^*, \beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right)\right\rangle$ to both sides of (3.43). □

We denote

$$W \equiv \begin{pmatrix} X_i \\ Y_{ij} \\ \lambda_{ij} \end{pmatrix} \quad \text{and} \quad G \equiv \begin{pmatrix} \left(\alpha_{i0}^{'}\right)^{-1} & 0 & 0 \\ 0 & \left(\alpha_{ij}^{'}\right)^{-1} & 0 \\ 0 & 0 & \beta_{ij}^{-1} \end{pmatrix}.$$

Clearly, G is positive definite. We define the G-inner product of W and $W^{'}$ as

$$\left\langle W, W' \right\rangle_G \equiv \sum_{i=1}^{n} \left(\frac{1}{\alpha_{i0}^{'}} \left\langle X_i, X_i^{'} \right\rangle + \sum_{j=1}^{m} \frac{1}{\alpha_{ij}^{'}} \left\langle Y_{ij}, Y_{ij}^{'} \right\rangle + \sum_{j=1}^{m} \left\langle \lambda_{ij}, \beta_{ij}^{-1} \left(\lambda_{ij}^{'} \right) \right\rangle \right),$$

and the associated G-norm as

$$\|W\|_G \equiv \left(\sum_{i=1}^{n} \left(\frac{1}{\alpha_{i0}^{'}} \|X_i\|^2 + \sum_{j=1}^{m} \frac{1}{\alpha_{ij}^{'}} \|Y_{ij}\|^2 + \sum_{j=1}^{m} \|\lambda_{ij}\|^2_{\beta_{ij}^{-1}} \right) \right)^{\frac{1}{2}},$$

where $\|\cdot\|^2_{\beta_{ij}^{-1}} \equiv \left\langle \cdot, \beta_{ij}^{-1} (\cdot) \right\rangle$.

Then (3.44) can be written as

$$\begin{aligned} &\left\langle \begin{pmatrix} \widetilde{X}_i^k - X_i^* \\ \widetilde{Y}_{ij}^k - Y_{ij}^* \\ \widetilde{\lambda}_{ij}^k - \lambda_{ij}^* \end{pmatrix}, \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) + \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G \\ \geq\ & \sum_{i=1}^{n} \sum_{j=1}^{m} \left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle. \end{aligned} \tag{3.45}$$

According to the choice criterion of $\alpha_{i0}^{'}$ and $\alpha_{ij}^{'}$, we have the following lemma.

Lemma 3.2.4. *There hold*

$$
\begin{aligned}
& \alpha_{i0}^{'} \left\| \mu_{i0} \left(F_i \left(X_i^k \right) - F_i \left(\widetilde{X}_i^k \right) \right) + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij} \left(X_i^k \right) \right. \\
& \left. - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij} \left(\widetilde{X}_i^k \right) \right\| \\
\leq \quad & \eta \left\| \widetilde{X}_i^k - X_i^k \right\|,
\end{aligned}
\tag{3.46}
$$

$$
\begin{aligned}
& \alpha_{ij}^{'} \left\| \mu_{ij} \left(\mathcal{L}_{ij}^{-1} \right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1} \left(Y_{ij}^k \right) - \mu_{ij} \left(\mathcal{L}_{ij}^{-1} \right)^T \circ F_i \circ \mathcal{L}_{ij}^{-1} \left(\widetilde{Y}_{ij}^k \right) \right. \\
& \left. + \beta_{ij} \left(Y_{ij}^k \right) - \beta_{ij} \left(\widetilde{Y}_{ij}^k \right) \right\| \\
\leq \quad & \eta \left\| \widetilde{Y}_{ij}^k - Y_{ij}^k \right\|,
\end{aligned}
\tag{3.47}
$$

and

$$
\left\| \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G \geq (1 - \eta) \left\| W^k - \widetilde{W}^k \right\|_G . \tag{3.48}
$$

Proof. (3.46) and (3.47) can be immediately derived from the conditions (3.29) and (3.32).

It follows from (3.46) that

$$\begin{aligned}
& \left\| R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right)\right\| \\
= & \left\| X_i^k - \widetilde{X}_i^k - \alpha_{i0}^{'}\Bigg(\mu_{i0}\left(F_i\left(X_i^k\right) - F_i\left(\widetilde{X}_i^k\right)\right)\right. \\
& \left. + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right) - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)\Bigg)\right\| \\
\geq & \left\| X_i^k - \widetilde{X}_i^k\right\| - \left\| \alpha_{i0}^{'}\Bigg(\mu_{i0}\left(F_i\left(X_i^k\right) - F_i\left(\widetilde{X}_i^k\right)\right)\right. \\
& \left. + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right) - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)\Bigg)\right\| \\
\geq & (1-\eta)\left\| X_i^k - \widetilde{X}_i^k\right\|.
\end{aligned} \tag{3.49}$$

Similarly, from (3.47) there is

$$\left\| R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right)\right\| \geq (1-\eta)\left\| Y_{ij}^k - \widetilde{Y}_{ij}^k\right\|. \tag{3.50}$$

Then we have (3.48) directly from (3.49) and (3.50). □

Before the main convergence result, we need another lemma based on the choice criterion of $\alpha_{i0}^{'}$ and $\alpha_{ij}^{'}$.

Lemma 3.2.5. *Define*

$$\begin{aligned}
&\Psi\left(X_i^k, Y_{ij}^k, \lambda_{ij}^k, \widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right)\\
\equiv\ &\left\langle \begin{pmatrix} X_i^k - \widetilde{X}_i^k \\ Y_{ij}^k - \widetilde{Y}_{ij}^k \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}, \begin{pmatrix} R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}'\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G\\
&+\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle,
\end{aligned}$$

then there are two lower bounds for it.

$$\begin{aligned}
&2\Psi\left(X_i^k, Y_{ij}^k, \lambda_{ij}^k, \widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right)\\
\geq\ &\left\| \begin{pmatrix} R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2 + \left\| \begin{pmatrix} 0 \\ \alpha_{ij}'\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ 0 \end{pmatrix} \right\|_G^2,
\end{aligned} \tag{3.51}$$

and

$$\begin{aligned}
&4\Psi\left(X_i^k, Y_{ij}^k, \lambda_{ij}^k, \widetilde{X}_i^k, \widetilde{Y}_{ij}^k, \widetilde{\lambda}_{ij}^k\right)\\
\geq\ &\left\| \begin{pmatrix} R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}'\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2 .
\end{aligned} \tag{3.52}$$

Proof. It follows from (3.46) that

$$
\begin{aligned}
& 2\left\langle X_i^k - \widetilde{X}_i^k, R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right)\right\rangle \\
= \ & 2\Bigg\langle X_i^k - \widetilde{X}_i^k, X_i^k - \widetilde{X}_i^k - \alpha_{i0}'\Bigg(\mu_{i0}\left(F_i\left(X_i^k\right) - F_i\left(\widetilde{X}_i^k\right)\right) \\
& + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right) - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)\Bigg)\Bigg\rangle \\
\geq \ & \left\|X_i^k - \widetilde{X}_i^k\right\|^2 - 2\Bigg\langle X_i^k - \widetilde{X}_i^k, \alpha_{i0}'\Bigg(\mu_{i0}\left(F_i\left(X_i^k\right) - F_i\left(\widetilde{X}_i^k\right)\right) \\
& + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right) - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)\Bigg)\Bigg\rangle \\
& + \Bigg\|\alpha_{i0}'\Bigg(\mu_{i0}\left(F_i\left(X_i^k\right) - \left(\widetilde{X}_i^k\right)\right) + \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(X_i^k\right) \\
& - \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \circ \mathcal{L}_{ij}\left(\widetilde{X}_i^k\right)\Bigg)\Bigg\|^2 \\
= \ & \left\|R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right)\right\|^2. \qquad (3.53)
\end{aligned}
$$

Similarly, from (3.47) there is

$$
2\left\langle Y_{ij}^k - \widetilde{Y}_{ij}^k, R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right)\right\rangle \geq \left\|R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right)\right\|^2. \qquad (3.54)
$$

For two self-adjoint linear operators $\mathcal{S}$ and $\mathcal{V}$, the notation $\mathcal{S} \preceq \mathcal{V}$ means that $\langle M, \mathcal{S}(M)\rangle \leq \langle M, \mathcal{V}(M)\rangle$ for all M. Because of (3.31), we have

$$
\begin{aligned}
& \alpha_{ij}^{'}\lambda_{\max}\left(\beta_{ij}\right) \leq 1 \\
\Longrightarrow\ & \alpha_{ij}^{'}\beta_{ij}^{2} \preceq \beta_{ij} \\
\Longrightarrow\ & \alpha_{ij}^{'}\left\|Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right\|_{\beta_{ij}^{2}}^{2} \leq \left\|Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right\|_{\beta_{ij}}^{2} \\
\Longrightarrow\ & \frac{1}{\alpha_{ij}^{'}}\left\|\alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right)\right\|^{2} \leq \left\|Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right\|_{\beta_{ij}}^{2} \\
\Longrightarrow\ & \left\|\begin{pmatrix} 0 \\ \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right) \\ 0 \end{pmatrix}\right\|_{G}^{2} \leq \sum_{i=1}^{n}\sum_{j=1}^{m}\left\|Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right\|_{\beta_{ij}}^{2}. \qquad (3.55)
\end{aligned}
$$

Thus

$$
\begin{aligned}
& 4\Psi\left(X_{i}^{k}, Y_{ij}^{k}, \lambda_{ij}^{k}, \widetilde{X}_{i}^{k}, \widetilde{Y}_{ij}^{k}, \widetilde{\lambda}_{ij}^{k}\right) \\
= \ & 4\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i0}^{'}}\left\langle X_{i}^{k}-\widetilde{X}_{i}^{k}, R_{i0}^{'}\left(X_{i}^{k}, \widetilde{X}_{i}^{k}\right)\right\rangle + \sum_{j=1}^{m}\frac{1}{\alpha_{ij}^{'}}\left\langle Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}, R_{ij}^{'}\left(Y_{ij}^{k}, \widetilde{Y}_{ij}^{k}\right)\right\rangle\right. \\
& \left. + \sum_{j=1}^{m}\left\|Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right\|_{\beta_{ij}}^{2} + \sum_{j=1}^{m}\left\|\lambda_{ij}^{k}-\widetilde{\lambda}_{ij}^{k}\right\|_{\beta_{ij}^{-1}}^{2} + \sum_{j=1}^{m}\left\langle \lambda_{ij}^{k}-\widetilde{\lambda}_{ij}^{k}, \widetilde{Y}_{ij}^{k}-Y_{ij}^{k}\right\rangle\right) \\
= \ & 4\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i0}^{'}}\left\langle X_{i}^{k}-\widetilde{X}_{i}^{k}, R_{i0}^{'}\left(X_{i}^{k}, \widetilde{X}_{i}^{k}\right)\right\rangle + \sum_{j=1}^{m}\frac{1}{\alpha_{ij}^{'}}\left\langle Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}, R_{ij}^{'}\left(Y_{ij}^{k}, \widetilde{Y}_{ij}^{k}\right)\right\rangle\right. \\
& + \frac{1}{2}\sum_{j=1}^{m}\left\|Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right\|_{\beta_{ij}}^{2} + \frac{1}{2}\sum_{j=1}^{m}\left\|\lambda_{ij}^{k}-\widetilde{\lambda}_{ij}^{k}\right\|_{\beta_{ij}^{-1}}^{2} \\
& \left. + \frac{1}{2}\sum_{j=1}^{m}\left\|\lambda_{ij}^{k}-\widetilde{\lambda}_{ij}^{k}-\beta_{ij}\left(Y_{ij}^{k}-\widetilde{Y}_{ij}^{k}\right)\right\|_{\beta_{ij}^{-1}}^{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & 2\sum_{i=1}^{n}\left(\frac{2}{\alpha_{i0}^{'}}\left\langle X_i^k-\widetilde{X}_i^k, R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right)\right\rangle+\sum_{j=1}^{m}\frac{2}{\alpha_{ij}^{'}}\left\langle Y_{ij}^k-\widetilde{Y}_{ij}^k, R_{ij}^{'}\left(Y_{ij}^k,\widetilde{Y}_{ij}^k\right)\right\rangle\right.\\
& \left.+\sum_{j=1}^{m}\left\|Y_{ij}^k-\widetilde{Y}_{ij}^k\right\|_{\beta_{ij}}^2+\sum_{j=1}^{m}\left\|\lambda_{ij}^k-\widetilde{\lambda}_{ij}^k\right\|_{\beta_{ij}^{-1}}^2+\sum_{j=1}^{m}\left\|\mathcal{L}_{ij}\left(X_i^{k+1}\right)-Y_{ij}^k\right\|_{\beta_{ij}}^2\right)\\
\geq_{(3.54)}^{(3.53)} \quad & 2\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i0}^{'}}\left\|R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right)\right\|^2+\sum_{j=1}^{m}\frac{1}{\alpha_{ij}^{'}}\left\|R_{ij}^{'}\left(Y_{ij}^k,\widetilde{Y}_{ij}^k\right)\right\|^2\right.\\
& \left.+\sum_{j=1}^{m}\left\|Y_{ij}^k-\widetilde{Y}_{ij}^k\right\|_{\beta_{ij}}^2+\sum_{j=1}^{m}\left\|\lambda_{ij}^{'}-\widetilde{\lambda}_{ij}^k\right\|_{\beta_{ij}^{-1}}^2+\sum_{j=1}^{m}\left\|\mathcal{L}_{ij}\left(X_i^{k+1}\right)-Y_{ij}^k\right\|_{\beta_{ij}}^2\right)\\
\geq^{(3.55)} \quad & 2\left(\left\|\begin{pmatrix} R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right)\\ R_{ij}^{'}\left(Y_{ij}^k,\widetilde{Y}_{ij}^k\right)\\ \lambda_{ij}^k-\widetilde{\lambda}_{ij}^k\end{pmatrix}\right\|_G^2+\left\|\begin{pmatrix}0\\ \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k-\widetilde{Y}_{ij}^k\right)\\ 0\end{pmatrix}\right\|_G^2\right)\\
& +2\sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\mathcal{L}_{ij}\left(X_i^{k+1}\right)-Y_{ij}^k\right\|_{\beta_{ij}}^2\\
\geq \quad & 2\left(\left\|\begin{pmatrix} R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right)\\ R_{ij}^{'}\left(Y_{ij}^k,\widetilde{Y}_{ij}^k\right)\\ \lambda_{ij}^k-\widetilde{\lambda}_{ij}^k\end{pmatrix}\right\|_G^2+\left\|\begin{pmatrix}0\\ \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k-\widetilde{Y}_{ij}^k\right)\\ 0\end{pmatrix}\right\|_G^2\right)\\
\geq \quad & \left\|\begin{pmatrix} R_{i0}^{'}\left(X_i^k,\widetilde{X}_i^k\right)\\ R_{ij}^{'}\left(Y_{ij}^k,\widetilde{Y}_{ij}^k\right)+\alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k-\widetilde{Y}_{ij}^k\right)\\ \lambda_{ij}^k-\widetilde{\lambda}_{ij}^k\end{pmatrix}\right\|_G^2.
\end{aligned}
$$

□

The following is the main convergence theorem.

Theorem 3.2.6. *The sequence* $\left\{X_i^k\right\}$ *generated by the prediction-correction ADM for MNLSDP converges to a solution point* X_i^* *of system (1.5).*

Proof. Observe that solving the optimal condition (1.5) is equivalent to finding a zero of the residual function

$$\|e(W)\|_G \equiv \left\| \begin{array}{c} X_i - P_{\Omega_{i1}}\left[X_i - \alpha'_{i0}\left(\mu_{i0}F_i(X_i) - \sum_{j=1}^{m}\mathcal{L}_{ij}^T(\lambda_{ij})\right)\right] \\ (Y_{11},\cdots,Y_{n1}) - P_{\Omega'}\Big[(Y_{11},\cdots,Y_{n1}) - \alpha'_{11} \\ \left(\mu_{11}\left(\mathcal{L}_{11}^{-1}\right)^T\circ F_1\circ\mathcal{L}_{11}^{-1}(Y_{11}) + \lambda_{11},\cdots,\mu_{n1}\left(\mathcal{L}_{n1}^{-1}\right)^T\circ F_n\circ\mathcal{L}_{n1}^{-1}(Y_{n1}) + \lambda_{n1}\right)\Big] \\ Y_{ij} - P_{\Omega'_{ij}}\left[Y_{ij} - \alpha'_{ij}\left(\mu_{ij}\left(\mathcal{L}_{ij}^{-1}\right)^T\circ F_i\circ\mathcal{L}_{ij}^{-1}(Y_{ij}) + \lambda_{ij}\right)\right] \\ \beta_{ij}\left(\mathcal{L}_{ij}(X_i) - Y_{ij}\right) \end{array} \right\|_G .$$

Then we have from (3.30), (3.33), and (3.34) that

$$
\begin{aligned}
&\left\| e(\widetilde{W}^k) \right\|_G \\
\leq\ & \left\| \begin{array}{c} \alpha_{i0}^{'} \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) + R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{array} \right\|_G \\
\leq\ & \left\| \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G + \left\| \begin{pmatrix} \alpha_{i0}^{'} \sum_{j=1}^{m} \mathcal{L}_{ij}^T \circ \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ 0 \\ 0 \end{pmatrix} \right\|_G \\
\leq\ & \delta \left(\left\| \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G + \left\| \begin{pmatrix} 0 \\ \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ 0 \end{pmatrix} \right\|_G \right),
\end{aligned}
\tag{3.56}
$$

where δ is a positive constant.

Thus

$$
\begin{aligned}
& \left\| W^{k+1} - W^* \right\|_G^2 \\
= \quad & \left\| \begin{pmatrix} X_i^{k+1} - X_i^* \\ Y_{ij}^{k+1} - Y_{ij}^* \\ \lambda_{ij}^{k+1} - \lambda_{ij}^* \end{pmatrix} \right\|_G^2 \\
\leq \quad & \left\| \begin{pmatrix} X_i^k - \gamma^k R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) - X_i^* \\ Y_{ij}^k - \gamma^k R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) - \gamma^k \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) - Y_{ij}^* \\ \lambda_{ij}^k - \gamma^k \left(\lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \right) - \lambda_{ij}^* \end{pmatrix} \right\|_G^2 \\
= \quad & \left\| W^k - W^* \right\|_G^2 + \left(\gamma^k \right)^2 \left\| \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) + \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2 \\
& -2\gamma^k \left\langle \begin{pmatrix} X_i^k - X_i^* \\ Y_{ij}^k - Y_{ij}^* \\ \lambda_{ij}^k - \lambda_{ij}^* \end{pmatrix}, \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) + \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G \\
\leq^{(3.45)} \quad & \left\| W^k - W^* \right\|_G^2 + \left(\gamma^k \right)^2 \left\| \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) + \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2 \\
& -2\gamma^k \left\langle \begin{pmatrix} X_i^k - \widetilde{X}_i^k \\ Y_{ij}^k - \widetilde{Y}_{ij}^k \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}, \begin{pmatrix} R_{i0}^{'} \left(X_i^k, \widetilde{X}_i^k \right) \\ R_{ij}^{'} \left(Y_{ij}^k, \widetilde{Y}_{ij}^k \right) + \alpha_{ij}^{'} \beta_{ij} \left(Y_{ij}^k - \widetilde{Y}_{ij}^k \right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G \\
& -2\gamma^k \sum_{i=1}^{n} \sum_{j=1}^{m} \left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle
\end{aligned}
$$

$$\begin{aligned}
&\leq^{(3.52)} \left\|W^k - W^*\right\|_G^2 - \left(2\gamma^k - 4(\gamma^k)^2\right) \sum_{i=1}^{n} \sum_{j=1}^{m} \left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle \\
&\quad - \left(2\gamma^k - 4(\gamma^k)^2\right) \\
&\quad \left\langle \begin{pmatrix} X_i^k - \widetilde{X}_i^k \\ Y_{ij}^k - \widetilde{Y}_{ij}^k \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}, \begin{pmatrix} R'_{i0}\left(X_i^k, \widetilde{X}_i^k\right) \\ R'_{ij}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha'_{ij}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G \qquad (3.57)
\end{aligned}$$

$$\begin{aligned}
&\leq^{(3.51)} \left\|W^k - W^*\right\|_G^2 - \left(\gamma^k - 2(\gamma^k)^2\right) \\
&\quad \left(\left\| \begin{pmatrix} R'_{i0}\left(X_i^k, \widetilde{X}_i^k\right) \\ R'_{ij}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2 + \left\| \begin{pmatrix} 0 \\ \alpha'_{ij}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ 0 \end{pmatrix} \right\|_G^2 \right) \qquad (3.58)
\end{aligned}$$

$$\begin{aligned}
&\leq \left\|W^k - W^*\right\|_G^2 - \frac{\gamma^k - 2(\gamma^k)^2}{2} \\
&\quad \left(\left\| \begin{pmatrix} R'_{i0}\left(X_i^k, \widetilde{X}_i^k\right) \\ R'_{ij}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G + \left\| \begin{pmatrix} 0 \\ \alpha'_{ij}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ 0 \end{pmatrix} \right\|_G \right)^2 \\
&\leq^{(3.56)} \left\|W^k - W^*\right\|_G^2 - \frac{\gamma^k - 2(\gamma^k)^2}{2\delta^2} \left\|e\left(\widetilde{W}^k\right)\right\|_G^2. \qquad (3.59)
\end{aligned}$$

From the above inequality, we have

$$\left\|W^{k+1} - W^*\right\|_G^2 \leq \left\|W^k - W^*\right\|_G^2 \leq \cdots \leq \left\|W^0 - W^*\right\|_G^2. \qquad (3.60)$$

That is, the sequence $\{W^k\}$ is bounded. It follows from (3.58) that

$$\sum_{k=0}^{\infty}\left(\gamma^k - 2(\gamma^k)^2\right)\left(\left\|\begin{pmatrix} R'_{i0}\left(X_i^k, \widetilde{X}_i^k\right) \\ R'_{ij}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}\right\|_G^2 + \left\|\begin{pmatrix} 0 \\ \alpha'_{ij}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ 0 \end{pmatrix}\right\|_G^2\right) < +\infty.$$

This implies that

$$\lim_{k\to\infty}\left\|\begin{pmatrix} R'_{i0}\left(X_i^k, \widetilde{X}_i^k\right) \\ R'_{ij}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}\right\|_G = 0 \Longrightarrow^{(3.48)} \lim_{k\to\infty}\left\|W^k - \widetilde{W}^k\right\|_G = 0.$$

Thus the sequence $\left\{\widetilde{W}^k\right\}$ is also bounded. Then there exists at least one cluster point of $\left\{\widetilde{W}^k\right\}$.

It also follows from (3.59) that

$$\sum_{k=0}^{\infty}\frac{\gamma^k - 2(\gamma^k)^2}{2\delta^2}\left\|e\left(\widetilde{W}^k\right)\right\|_G^2 < +\infty.$$

This implies that

$$\lim_{k\to\infty}\left\|e\left(\widetilde{W}^k\right)\right\|_G = 0.$$

Let $\overline{W}$ be a cluster point of $\left\{\widetilde{W}^k\right\}$, and let $\left\{\widetilde{W}^{k_j}\right\}$ be a corresponding subsequence converging to $\overline{W}$. Note that

$$\left\|e\left(\overline{W}\right)\right\|_G = \lim_{j\to\infty}\left\|e\left(\widetilde{W}^{k_j}\right)\right\|_G = 0,$$

which means that $\overline{W}$ is a zero of the residual function. Therefore $\overline{W}$ satisfies (1.5). Setting $W^* = \overline{W}$ in (3.60), we have

$$\left\|W^{k+1} - \overline{W}\right\|_G^2 \leq \left\|W^k - \overline{W}\right\|_G^2, \ \forall k \geq 0. \tag{3.61}$$

Since $\lim_{j\to\infty} \left\|\widetilde{W}^{k_j} - \overline{W}\right\|_G = 0$ and $\lim_{k\to\infty} \left\|W^k - \widetilde{W}^k\right\|_G = 0$, for any given $\epsilon > 0$, there exists an integer $l > 0$ such that

$$\left\|\widetilde{W}^{k_l} - \overline{W}\right\|_G < \frac{1}{2}\epsilon \quad \text{and} \quad \left\|W^{k_l} - \widetilde{W}^{k_l}\right\|_G < \frac{1}{2}\epsilon. \tag{3.62}$$

Therefore, for any $k > k_l$, it follows from (3.61) and (3.62) that

$$\left\|W^k - \overline{W}\right\|_G \leq \left\|W^{k_l} - \overline{W}\right\|_G \leq \left\|W^{k_l} - \widetilde{W}^{k_l}\right\|_G + \left\|\widetilde{W}^{k_l} - \overline{W}\right\|_G \leq \epsilon.$$

Thus, the sequence $\left\{W^k\right\}$ has a unique cluster point and

$$\lim_{k\to\infty} W^k = \overline{W}.$$

This completes the proof. $\square$

Remark: We can set optimal step-length γ^k as follows

$$
\begin{aligned}
& \gamma^k \\
\equiv\ & \nu\gamma_*^k \\
=\ & \nu \frac{\left\langle \begin{pmatrix} X_i^k - \widetilde{X}_i^k \\ Y_{ij}^k - \widetilde{Y}_{ij}^k \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}, \begin{pmatrix} R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G}{\left\| \begin{pmatrix} R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2} \\
& + \nu \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle}{\left\| \begin{pmatrix} R_{i0}^{'}\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}^{'}\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}^{'}\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2},
\end{aligned}
$$

where $\nu \in (0,2)$ is a relaxation factor. This choice of γ_*^k is to maximize the function

$$2\gamma^k \left\langle \begin{pmatrix} X_i^k - \widetilde{X}_i^k \\ Y_{ij}^k - \widetilde{Y}_{ij}^k \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}, \begin{pmatrix} R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}'\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G$$

$$+2\gamma^k \sum_{i=1}^{n}\sum_{j=1}^{m} \left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle - \left(\gamma^k\right)^2 \left\| \begin{pmatrix} R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}'\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\|_G^2,$$

which is a lower bound for the measure of improvement

$$\left\|W^k - W^*\right\|_G^2 - \left\|W^{k+1} - W^*\right\|_G^2.$$

Furthermore, γ_*^k will not be too small even when $(X_i^k, Y_{ij}^k, \lambda_{ij}^k)$ is close to the solution. Actually from (3.52), we can see $\gamma_*^k \geq \frac{1}{4}$. At this case, all parts of the convergence proof keep the same, except that (3.57) is changed to

$$\begin{aligned} & \left\|W^{k+1} - W^*\right\|_G^2 \\ \leq\ & \left\|W^k - W^*\right\|_G^2 - \nu(2-\nu)\gamma_*^k \Bigg(\sum_{i=1}^{n}\sum_{j=1}^{m} \left\langle \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k, \widetilde{Y}_{ij}^k - Y_{ij}^k \right\rangle + \\ & \left\langle \begin{pmatrix} X_i^k - \widetilde{X}_i^k \\ Y_{ij}^k - \widetilde{Y}_{ij}^k \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix}, \begin{pmatrix} R_{i0}'\left(X_i^k, \widetilde{X}_i^k\right) \\ R_{ij}'\left(Y_{ij}^k, \widetilde{Y}_{ij}^k\right) + \alpha_{ij}'\beta_{ij}\left(Y_{ij}^k - \widetilde{Y}_{ij}^k\right) \\ \lambda_{ij}^k - \widetilde{\lambda}_{ij}^k \end{pmatrix} \right\rangle_G \Bigg). \end{aligned} \tag{3.63}$$

4. SPECIALIZATION: CONVEX NONLINEAR SEMIDEFINITE PROGRAMMING

Convex NLSDP, in which all the matrix functions and constraints are convex, receives more and more interests now because a number of important applications in management and engineering lead to it. However, as mentioned in Chapter 2, the research on it is basically at the developing stage. This chapter is devoted to convex NLSDP. The modified ADM developed in the last chapter is first specialized for solving CQCQSDP problems. Thereafter, the prediction-correction ADM is specialized for solving general CNLSDP problems. In each of the specializations, we pay attention to the special structure of the problems, including the simplicities of the functions and the sets. Thus, the specialized methods are simpler and more efficient.

4.1 Convex Quadratically Constrained Quadratic Semidefinite Programming

We are concerned with the following CQCQSDP problem.

$$
\begin{aligned}
\min \quad & q_0(X) \equiv \frac{1}{2}\langle X, Q_0(X)\rangle + \langle B_0, X\rangle \\
\text{s.t.} \quad & q_i(X) \equiv \frac{1}{2}\langle X, Q_i(X)\rangle + \langle B_i, X\rangle + c_i \le 0, \quad i = 1, \cdots, m \\
& X \succeq 0,
\end{aligned}
\tag{4.1}
$$

where $Q_i : \mathbb{S}^n \to \mathbb{S}^n,\ i = 0, 1, \cdots, m$, is a self-adjoint positive semidefinite linear operator; $B_i \in \mathbb{S}^n$ and $c_i \in \mathbb{R}$ is a scalar. Basic examples of $Q(X)$ include the symmetrized Kronecker product $U \otimes U(X) = (UXU^T + UX^TU^T)/2$ for a given $U \in \mathbb{S}^n_+$ and the Hadamard product $H \circ X$ defined as $(H \circ X)_{ij} = H_{ij}X_{ij}$ for some $H \in \mathbb{S}^n \cap \Re^{n\times n}_+$, etc.

Problem (4.1) is a convex optimization problem in the space $\mathbb{S}^n$ and generalizes CQSDP model by allowing quadratic constraints. Then our proposed algorithm for CQCQSDP can be also used to solve CQSDP. In contrast, current methods [8, 25, 48, 50, 57, 64] designed for solving CQSDP heavily depend on the linearity of the constraints, thus they cannot be readily extended to solve CQCQSDP problems.

We also notice that in [3] Beck studied quadratic matrix programming of order r which may not be convex. He constructed a special semidefinite relaxation and its dual and showed that under some mild conditions strong duality holds for the relaxed problem with at most r constraints. However, Beck's model does not include the semidefinite cone constraint. Therefore, it is essentially a vector optimization model, rather than a semidefinite optimization problem like (4.1).

Recall that $q_i(X) \equiv \frac{1}{2}\langle X, Q_i(X)\rangle + \langle B_i, X\rangle + c_i \leq 0$. By introducing artificial constraints

$$Y_i = X \quad \text{and} \quad \Omega_i = \{Y_i : q_i(Y_i) \leq 0\}, \ i = 1, \cdots, m, \tag{4.2}$$

we may re-write (4.1) equivalently as

$$\min \ q_0(X) \quad \text{s.t. } X \succeq 0, \ X = Y_i, \ Y_i \in \Omega_i, \ i = 1, \cdots, m. \tag{4.3}$$

After this transformation, we can see that Problem (4.3) is exactly a special case of MQSDP. Thus the modified ADM for MQSDP Algorithm 3.1.1 can be applied to it. Our general Assumption 1.2.1 is specialized to

Assumption 4.1.1. *The solution set $(X^*, Y_i^*, \lambda_i^*)$ of KKT system of Problem (4.3) is nonempty.*

A sufficient condition that guarantees the assumption to be valid is that the CQCQSDP is feasible and at least one of $Q_0, ..., Q_m$ is positive definite.

Let

$$R\left(X^k, X^{k+1}\right) \equiv Q_0\left(X^{k+1}\right) - Q_0\left(X^k\right) - \gamma\left(X^{k+1} - X^k\right)$$

for certain constant γ such that $\gamma \geq \lambda_{\max}(Q_0)$, where $\lambda_{\max}(Q_0)$ is the largest eigenvalue of Q_0. Set

$$\alpha = \left(\sum_{i=1}^{m} \beta_i + \gamma \right)^{-1} \quad \text{and} \quad D = B_0 - \sum_{i=1}^{m} \left(\lambda_i^k + \beta_i Y_i^k \right) - \gamma X^k + Q_0 \left(X^k \right),$$

where $\beta_i,\ i = 1, \cdots, m$, is certain positive scalar. The modified ADM is given as follows.

Algorithm 4.1.2. The Modified ADM for CQCQSDP

Do at each iteration until a stopping criterion is met

Step 1. $\left(X^k, Y_i^k, \lambda_i^k\right) \to \left(X^{k+1}, Y_i^k, \lambda_i^k\right)$, *where*

$$X^{k+1} = P_{\mathbb{S}_+^n} \left[-\alpha D \right] \tag{4.4}$$

Step 2. $\left(X^{k+1}, Y_i^k, \lambda_i^k\right) \to \left(X^{k+1}, Y_i^{k+1}, \lambda_i^k\right)$, $i = 1, \cdots, m$, *where*

$$Y_i^{k+1} = P_{\Omega_i} \left[X^{k+1} - \frac{1}{\beta_i} \lambda_i^k \right] \tag{4.5}$$

Step 3. $\left(X^{k+1}, Y_i^{k+1}, \lambda_i^k\right) \to \left(X^{k+1}, Y_i^{k+1}, \lambda_i^{k+1}\right)$, $i = 1, \cdots, m$, *where*

$$\lambda_i^{k+1} = \lambda_i^k - \beta_i \left(X^{k+1} - Y_i^{k+1} \right) \tag{4.6}$$

In order to solve (4.3) by the modified ADM, we only need to compute the metric projections of a matrix onto Ω_i and $\mathbb{S}^n_+$. The projection onto Ω_i can be computed in a similar way as computing the Euclidean projection of a vector onto an ellipsoid in the real vector space. Therefore, the computation of this projection can be very fast, see, for example, [14] for the corresponding algorithms.

Remark: If some $Q_i(\cdot)$ is positive definite, we can introduce a specially designed constraint $Y_i = \mathcal{L}_i(X)$ with the invertible linear operator $\mathcal{L}_i(\cdot) = Q_i^{\frac{1}{2}}(\cdot)$. Then

$$\langle X, Q_i(X)\rangle = \left\langle Q_i^{\frac{1}{2}}(X), Q_i^{\frac{1}{2}}(X)\right\rangle = \langle Y_i, Y_i\rangle .$$

Thus the original ellipsoid-type convex set Ω_i becomes

$$\Omega_i' = \left\{Y_i : \frac{1}{2}\langle Y_i, Y_i\rangle + \left\langle \left(\mathcal{L}_i^{-1}\right)^T (B_i), Y_i\right\rangle + c_i \le 0\right\},$$

which is a ball. This choice of $\mathcal{L}_i$ can make the projection easy.

Let **vec** be an isometry identifying $\mathbb{S}^n$ with $\mathbb{R}^{n\times n}$ so that $\langle B, X\rangle = \mathbf{vec}\,(B)^T\,\mathbf{vec}\,(X)$. Let the matrix representation of operator Q under this isometry be $\bar{Q}$. Then for any X, we have $\mathbf{vec}\,(Q(X)) = \bar{Q}\,\mathbf{vec}\,(X)$. Since Q is self-adjoint and positive semidefinite, $\bar{Q}$ is a symmetric positive semidefinite matrix.

By using the **vec** function, we can convert the convex set Ω_i' to one with variable $\mathbf{vec}\,(Y_i)$ as follows.

$$\Omega_i'' = \left\{ \mathbf{vec}\ (Y_i) : \frac{1}{2}\mathbf{vec}\ (Y_i)^T\ \mathbf{vec}\ (Y_i) + \mathbf{vec}\ \left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right)^T \mathbf{vec}\ (Y_i) + c_i \leq 0 \right\}.$$

There is a close-form formula to compute the projection onto Ω_i'', namely

$$P_{\Omega_i''}\left(\mathbf{vec}\ (Y)\right) = \begin{cases} \mathbf{vec}\ (Y),\ \text{if}\ \left\|\mathbf{vec}\ (Y) + \mathbf{vec}\ \left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right)\right\|^2 \\ \leq \mathbf{vec}\ \left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right)^T \mathbf{vec}\ \left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right) - 2c_i; \\ \frac{\sqrt{\mathbf{vec}\left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right)^T \mathbf{vec}\left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right) - 2c_i}\left(\mathbf{vec}\ (Y) + \mathbf{vec}\left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right)\right)}{\left\|\mathbf{vec}\ (Y) + \mathbf{vec}\left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right)\right\|} \\ -\mathbf{vec}\ \left(\left(\mathcal{L}_i^{-1}\right)^T (B_i)\right),\ \text{otherwise} \end{cases}$$

4.2 General Convex Nonlinear Semidefinite Programming

Another special form of MSDP problems is the CNLSDP problem defined in $\mathbb{S}^n$ as follows.

$$\min\ c_0(X) \quad \text{s.t. } X \succeq 0,\ c_i(X) \leq 0,\ i = 1, \cdots, m, \tag{4.7}$$

where $c_i : \mathbb{S}^n \to \Re,\ i = 0, 1, \cdots, m$, is a convex continuously differentiable function. Let $C_i(X),\ i = 0, 1, \cdots, m$, denote the first order derivative of $c_i(X)$. Furthermore, we require the operator $C_0(\cdot)$ to be Lipschitz continuous with a constant L.

Notice that most current algorithms for solving NLSDP focus on solving the following alternative form:

$$\min \; f(x) \quad \text{s.t. } h(x) = 0, \; G(x) \preceq 0, \tag{4.8}$$

where $f : \Re^n \to \Re$, $h : \Re^n \to \Re^m$, and $G : \Re^n \to \mathbb{S}^p$ are given. However, in some applications the variable is naturally in the space of $\mathbb{S}^n_+$. In this case it seems more straightforward to consider (4.7).

By introducing

$$Y_i = X \quad \text{and} \quad \Omega_i = \{Y_i : c_i(Y_i) \le 0\}, \; i = 1, \cdots, m, \tag{4.9}$$

we re-write (4.7) equivalently as

$$\min \; c_0(X) \quad \text{s.t. } X \succeq 0, \; X = Y_i, \; Y_i \in \Omega_i, \; i = 1, \cdots, m. \tag{4.10}$$

After this transformation, we can see Problem (4.10) is exactly a special case of the MNLSDP. Thus the prediction-correction ADM for MNLSDP Algorithm 3.2.1 can be applied to it.

For convenience, we state the basic assumption to guarantee that Problem (4.10) under consideration is solvable.

Assumption 4.2.1. *The solution set* $(X^*, Y_i^*, \lambda_i^*)$ *of KKT system of Problem*

(4.10) is nonempty.

A sufficient condition for this assumption to be valid is the Slater condition, which says the interior of the feasible set of Problem (4.10) is nonempty.

Let

$$R^{'}\left(X^k,\widetilde{X}^k\right) \equiv \left(1-\alpha^{'}\sum_{i=1}^{m}\beta_i\right)\left(X^k-\widetilde{X}^k\right)-\alpha^{'}\left(C_0\left(X^k\right)-C_0\left(\widetilde{X}^k\right)\right),$$

where $\beta_i,\ i=1,\cdots,m$, is certain positive scalar and we choose positive scalar $\alpha^{'}$ so that $\alpha^{'} \leq \frac{\eta}{L+\eta\sum\limits_{i=1}^{m}\beta_i}$ with $0<\eta<1$. The prediction-correction ADM is reduced to the following form.

Algorithm 4.2.2. The Prediction-Correction ADM for CNLSDP

Do at each iteration until a stopping criterion is met

The Prediction Phase:

Step 1. $\left(X^k,Y_i^k,\lambda_i^k\right)\rightarrow\left(\widetilde{X}^k,Y_i^k,\lambda_i^k\right)$, *where*

$$\begin{aligned}\widetilde{X}^k &= P_{\mathbb{S}_+^n}\left[\widetilde{X}^k-\alpha^{'}\left(C_0\left(\widetilde{X}^k\right)-\sum_{i=1}^{m}\left(\lambda_i^k-\beta_i\left(\widetilde{X}^k-Y_i^k\right)\right)\right)\right.\\ &\qquad\left.+R^{'}\left(X^k,\widetilde{X}^k\right)\right]\\ &= P_{\mathbb{S}_+^n}\left[X^k-\alpha^{'}\left(C_0\left(X^k\right)-\sum_{i=1}^{m}\left(\lambda_i^k-\beta_i\left(X^k-Y_i^k\right)\right)\right)\right] \quad (4.11)\end{aligned}$$

Step 2. $\left(\widetilde{X}^k, Y_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, \widetilde{Y}_i^k, \lambda_i^k\right)$, $i = 1, \cdots, m$, *where*

$$\widetilde{Y}_i^k = P_{\Omega_i}\left[\widetilde{X}^k - \frac{1}{\beta_i}\lambda_i^k\right] \tag{4.12}$$

Step 3. $\left(\widetilde{X}^k, \widetilde{Y}_i^k, \lambda_i^k\right) \to \left(\widetilde{X}^k, \widetilde{Y}_i^k, \widetilde{\lambda}_i^k\right)$, $i = 1, \cdots, m$, *where*

$$\widetilde{\lambda}_i^k = \lambda_i^k - \beta_i\left(\widetilde{X}^k - \widetilde{Y}_i^k\right) \tag{4.13}$$

The Correction Phase:

Step 4. $\left(\widetilde{X}^k, \widetilde{Y}_i^k, \widetilde{\lambda}_i^k\right) \to \left(X^{k+1}, Y_i^{k+1}, \lambda_i^{k+1}\right)$, *where*

$$X^{k+1} = P_{\mathbb{S}_+^n}\left[X^k - \gamma^k R'\left(X^k, \widetilde{X}^k\right)\right] \tag{4.14}$$

$$Y_i^{k+1} = P_{\Omega_i}\left[Y_i^k - \gamma^k\left(Y_i^k - \widetilde{Y}_i^k\right)\right], \ i = 1, \cdots, m \tag{4.15}$$

$$\lambda_i^{k+1} = \lambda_i^k - \gamma^k\left(\lambda_i^k - \widetilde{\lambda}_i^k\right), \ i = 1, \cdots, m \tag{4.16}$$

The positive scalar $\gamma^k < 1$ *is the step-length.*

In order to solve (4.10) by the prediction-correction ADM, we only need to compute the metric projections of a matrix onto Ω_i and $\mathbb{S}_+^n$. The metric projections on these sets can be readily computed. Actually the projection onto convex set Ω_i can be computed by solving special convex nonlinear programming on the vector space. The interested readers can refer to [45] for the algorithms of computing the vector projection onto general convex set.

As pointed out in the Remark of Theorem 3.2.6, we can set optimal γ_*^k to maximize a lower bound of the improvement function as follows.

$$\gamma_*^k \equiv \frac{\alpha' \sum_{i=1}^{m} \frac{1}{\beta_i} \left\| \lambda_i^k - \widetilde{\lambda}_i^k \right\|^2 + \alpha' \sum_{i=1}^{m} \beta_i \left\| Y_i^k - \widetilde{Y}_i^k \right\|^2 + \left\langle X^k - \widetilde{X}^k, R' \left(X^k, \widetilde{X}^k \right) \right\rangle}{\alpha' \sum_{i=1}^{m} \frac{1}{\beta_i} \left\| \lambda_i^k - \widetilde{\lambda}_i^k \right\|^2 + \alpha' \sum_{i=1}^{m} \beta_i \left\| Y_i^k - \widetilde{Y}_i^k \right\|^2 + \left\| R' \left(X^k, \widetilde{X}^k \right) \right\|^2}.$$

This optimal step-length will not be too small (actually we can prove $\gamma_*^k \geq \frac{1}{2}$) even when the iterate point is close to the solution.

5. APPLICATION: THE COVARIANCE MATRIX ESTIMATION PROBLEM

For a random vector $x = (x_1, \cdots, x_n)^T$, the covariance matrix is defined as

$$\Sigma \equiv \mathrm{E}\left((x - \mathrm{E}(x)) \, (x - \mathrm{E}(x))^T \right),$$

where $\mathrm{E}(\cdot)$ stands for the expected value. By definition, any covariance matrix must be positive semidefinite.

The covariance matrix estimation problem is common in multivariate analysis. It occurs in many applications which involve statistical data analysis, such as in engineering design and data mining. In the field of portfolio management, the quality of covariance matrix estimation will significantly influence the measure of risk.

Markowitz [49] started modern portfolio theory in the 1950s. He stated the portfolio management problem as one of balancing expected return with risk. The famous mean-variance model is as follows

$$\min \; w^T \Sigma w \quad \text{s.t. } w^T e = 1, \; w^T \mu = q, \tag{5.1}$$

where e is the vector of ones, μ and Σ are the estimated mean and estimated covariance matrix of stock returns respectively, and q is the required expectation of the portfolio's return. The essence of this model is to show the trade-off between risk, which is measured as the portfolio's variance here, and the return. Thus the reduction of risk can translate into the increase of return. To compute the risk more accurately, the correct estimation of covariance matrix is crucial.

The traditional way of computing Σ is to use the sample covariance matrix. Sample covariance matrix is estimated from historical data, often taken as the maximum likelihood matrix under normality. It is a straightforward principle to let and only let the data speak. However, it will be problematic if there are missing elements in the observed data set. One approach is to treat the estimation of each variance or pairwise covariance separately. For example, we can compute sample covariance of pairs of stocks based on days on which both stocks have valid returns. Under the assumption that the data are missing at random, this kind of covariance matrix estimation is unbiased. However, the problem is that the obtained matrix is not guaranteed to be positive semidefinite. This could lead to negative risk for some portfolios in Markowitz's mean-variance model. Actually in some other situations of finance and statistics, the estimations of covariance matrices are also probably found to be inconsistent, i.e. $\Sigma \not\succeq 0$. To circumvent this obstacle, some modification work ought to be done for the unqualified estimators. In this chapter, we will review some previous models based on CQSDP for solving this problem and generalize them in the framework of MSDP, so that the modified ADMs can be applied.

5.1 The Nearest Correlation Matrix Problem and Its Extensions

Higham [31] introduced the following nearest correlation matrix problem. For arbitrary symmetric matrix C, one solves the optimization problem

$$\min \ \frac{1}{2}\|X - C\|^2 \quad \text{s.t. } X \in \mathbb{S}^n_+, \ X_{ii} = 1, \ i = 1, \cdots, n. \tag{5.2}$$

In [31], Higham used the modified alternating projections method to compute the solution for (5.2). Later, Qi and Sun [57] proposed a quadratically convergent Newton method for solving it.

Recently, Gao and Sun [25] extended this model to more general one.

$$\begin{aligned} \min \quad & \frac{1}{2}\|X - C\|^2 \\ \text{s.t.} \quad & \langle A_i, X\rangle = b_i, \ i = 1, \cdots, p \\ & \langle A_i, X\rangle \geq b_i, \ i = p+1, \cdots, m \\ & X \in \mathbb{S}^n_+, \end{aligned} \tag{5.3}$$

where A_i, $i = 1, \cdots, m$, are given symmetric matrices and $b \in \Re^m$ is also given. It is called the least squares covariance matrix problem. Compared with (5.2), it allows the presence of linear inequality constraints. For example, if we hope to restrict specific components of variable matrix X within some range based on parts of prior information, this kind of constraints will arise. Gao and Sun designed an inexact smoothing Newton method to solve the reformulated semidefinite system with two level metric projection operators.

For solving this linearly constrained quadratic SDP, Nie and Yuan [50], Malick [48], Boyd and Xiao [8], and Toh [64] suggested the conjugate gradient method, BFGS method, the projected gradient method, and an inexact primal-dual path-following method with pre-conditioners, respectively.

We can further extend (5.3) by adding some quadratic items to the constraint. When we want to compute a sample covariance matrix, one natural question comes out: How many historical data should we gather? In fact, we face the tradeoff between long-term data and short-term data. By using long-term data the obtained sample covariance matrix might be more stable, but less updated information has been caught; by using short-term data we can focus on current situation, but the sample covariance matrix could contain a lot of errors because of a smaller data size. It thus makes sense to combine two kinds of approaches to achieve better estimation for covariance matrix. We propose a new model for robust estimation of covariance matrix as follows.

$$
\begin{aligned}
\min \quad & \frac{1}{2}\|X - C\|^2 \\
\text{s.t.} \quad & \frac{1}{2}\left\|X - C^{'}\right\|^2 \le \epsilon \\
& \langle A_i, X\rangle = b_i, \;\; i = 1, \cdots, p \\
& \langle A_i, X\rangle \ge b_i, \;\; i = p+1, \cdots, m \\
& X \in \mathbb{S}^n_+,
\end{aligned}
\tag{5.4}
$$

where C, $C^{'}$ are the sample covariance matrices from short-term data and long-term data respectively and ϵ is a positive constant to control the size of trust region from the long-term stable estimation. Note that C can be just a given symmetric matrix if some observations are missing or wrong like Problems (5.2) and (5.3). Basically, Problem (5.4) is to find the nearest covariance matrix from short-term sample estimation within the trust region from the long-term sample estimation. Furthermore, additional linear equality and inequality constraints can be included. The optimal solution of (5.4) can be desirable because it will not be too far away from the long-term stable estimation while at the same time it can contain current information as much as possible through minimizing the distance with short-term estimation. By doing so, the estimation error can be also systematically reduced.

The next key question is how to solve the model (5.4). The problem (5.4), including (5.2) and (5.3), is exactly the special case of CQCQSDP (4.1), thus it can be solved by the modified ADM specified in Algorithm

4.1.2. Indeed, by introducing artificial constraints

$$
\begin{aligned}
Y_i &= X,\ i = 1, \cdots, m+1,\ \text{and} \\
\Omega_i &= \{Y_i : \langle A_i, Y_i \rangle = b_i\},\ i = 1, \cdots, p, \\
\Omega_i &= \{Y_i : \langle A_i, Y_i \rangle \geq b_i\},\ i = p+1, \cdots, m, \\
\Omega_{m+1} &= \left\{ Y_{m+1} : \frac{1}{2} \left\| Y_{m+1} - C' \right\|^2 \leq \epsilon \right\},
\end{aligned}
$$

we consider the equivalent problem of (5.4)

$$
\min \frac{1}{2} \|X - C\|^2 \quad \text{s.t. } X \succeq 0,\ X = Y_i,\ Y_i \in \Omega_i,\ i = 1, \cdots, m+1. \tag{5.5}
$$

Notice that except the convex set Ω_{m+1} containing quadratic item, other convex sets only involve linear equality or linear inequality. Thus it is easy to get the projections onto them comparing with the computation of the projection onto one ellipsoid which needs fast algorithm in [14]. As pointed in Remark of Algorithm 4.1.2, if the weighted Frobenius norm used in Ω_{m+1} is positive definite, we can change the ellipsoid set to a ball set through suitable chosen linear operator $\mathcal{L}_{m+1}$ and the equation $Y_{m+1} = \mathcal{L}_{m+1}(X)$. Then there is a close-form solution for the projection onto this set.

5.2 Covariance Matrix Estimation in Multiple-factor Model

In the 1960s and 1970s, single-factor model and further multiple-factor model were proposed and developed to explain expected return by many researchers.

The same structure used in the search for expected return can also be applied to explain portfolio risk, see [28]. Multiple-factor model has been popular in the investment community for many years because of good performance in making use of incisive, intuitive, and important factors to predict risk and understand return. The model can be used to analyze current portfolio risk, as well as to construct a portfolio that optimally trades off risk with expected returns. Thus it helps portfolio managers to control risk in an effective way.

The multiple-factor model has the following structure.

$$r_n = \sum_{k=1}^{K} V_{n,k} f_k + u_n, \tag{5.6}$$

where

- r_n = the excess return (return above the risk-free return) of stock n,
- $V_{n,k}$ = the exposure of asset n to factor k,
- f_k = the factor return of factor k,
- u_n = stock n's specific return. This is the return that cannot be explained by the factors.

Assume that the specific returns are uncorrelated with the factor returns and the specific returns are not correlated with each other. With these assumptions, we can express the risk structure as follows.

$$X_{n,m} = \sum_{k1,k2=1}^{K} V_{n,k1} F_{k1,k2} V_{m,k2} + \triangle_{n,m}, \tag{5.7}$$

where

- $X_{n,m}$ = the covariance of asset n with asset m (if $n = m$, this gives the variance of asset n),
- $V_{n,k1}$ = the exposure of asset n to factor $k1$,
- $F_{k1,k2}$ = the covariance of factor $k1$ with factor $k2$ (if $k1 = k2$, this gives the variance of factor $k1$),
- $\triangle_{n,m}$ = the specific covariance of asset n with asset m. By assumption, all specific risk correlations are zero, so this term is zero unless $n = m$. In that case, this term gives the specific variance of asset n.

The art of building a multiple-factor model is to choose appropriate factors. However, there is one key constraint: All factors must be a priori factors. That is, even though the factor returns are unknown, the factor exposures must be certain at the beginning of period. With this constraint, a wide variety of factors are possible. Among them, those chosen should satisfy three criteria: incisive, institutive, and interesting. According to [28], the factors can be typically divided into two broad categories: industries and risk indexes. Industry factors measure the different behavior of stocks in different industries. Industry exposures are usually 1/0 variables although the industry factors of large corporations with business in several industries must account for multiple industry memberships. Risk indexes measure the different behavior of stocks across non-industry dimensions such as volatility, momentum, liquidity, growth, value, earnings volatility, and financial leverage. Because various kinds of risk indexes involve different units and ranges, all raw exposure data must be rescaled:

$$V_{\text{normalized}} = \frac{V_{\text{raw}} - \mathrm{E}(V_{\text{raw}})}{\mathrm{SD}(V_{\text{raw}})},$$

where $\mathrm{E}(V_{\text{raw}})$ is the mean of raw exposure value and $\mathrm{SD}(V_{\text{raw}})$ is the standard deviation of raw exposure.

In a more compacted format, the multiple-factor model (5.6) and (5.7) can be written as

$$r = Vf + u \quad \text{and} \quad X = VFV^T + \triangle, \tag{5.8}$$

where r is an N vector of a stock's excess returns, V is an N by K matrix of stock factor exposures, f is a K vector of factor returns, u is an N vector of specific returns, X is the N by N covariance matrix of stock returns, F is the K by K covariance matrix of the factor returns, and $\triangle$ is the N by N diagonal matrix of specific variance.

In a multiple-factor model, the matrix of stock factor exposures V is preliminarily determined through some economical insights or statistical regression. Then as the covariance matrix estimation between factor returns and the estimation of each asset's specific variance are input, we can compute a covariance matrix estimation between asset returns through (5.8). However, these inputs may be of some errors which could transfer to the result. For example, if the covariance matrix estimation between factor returns is itself not positive semidefinite, the calculated covariance matrix estimation between asset returns might be not positive semidefinite either. Thus we need to modify them. In addition, the information of sample covariance matrix between asset returns directly from trading data can be also included. Then the new model is shown as follows:

$$\begin{aligned} \min_{X,F,\triangle} \quad & \left\|X-\overline{X}\right\|^2+\left\|F-\overline{F}\right\|^2+\left\|\triangle-\overline{\triangle}\right\|^2 \\ \text{s.t.} \quad & X = VFV^T+\triangle \\ & X\in\mathbb{S}^N_+,\ F\in\mathbb{S}^K_+,\ \triangle_{ii}\geq 0,\ i=1,\cdots,N,\ \text{and}\ \triangle_{ij}=0,\ i\neq j, \end{aligned} \tag{5.9}$$

where $\overline{X}$, $\overline{F}$, and $\overline{\triangle}$ are known matrices from pre-estimations. At the same time, more advanced models with additional linear and/or quadratic constraints are also possible. Through adding some structure in the modelling, the errors will be reduced systemically. Problem (5.9) is of the form of MQSDP problems, therefore the modified ADM Algorithm 3.1.1 can be used to solve it.

6. APPLICATION: THE MATRIX COMPLETION PROBLEM

In many applications of interest, one hopes to recover a matrix from an incomplete set of its entries. A motivating example is to infer answers in a partially filled survey. In practice, it is usual to only know very limited information. In general, it is difficult to complete the matrix and recover the entries that we have not seen. Actually we need to take advantage of the special structure of the matrix we wish to complete. If the incomplete set avoids any column or row of matrix, it is hopeless to reconstruct this unknown matrix. Thus throughout this chapter, we assume that we know at least one observation per row and one observation per column.

In the following we will consider two matrix completion problems. One is the completion problem of distance matrix and the other is the completion problem of low-rank matrix. They belong to the most studied matrix completion problems. In the first problem, we allow approximate completion which makes sense with the existence of errors. However, we require to exactly fit the data in the second problem. It is easy to switch this requirement in the models, based on the actual quality of data. For the completion problem of low-rank matrix the objective is nonconvex, therefore we only consider to solve its convex relaxation. Both problems can be seen as special cases of MSDP problems, then the ADM is applicable and it is believed that the algorithm would be efficient in solving these two problems.

6.1 The Completion Problem of Distance Matrix

An $n \times n$ symmetric matrix $D = (D_{ij})$ is called an Euclidean distance matrix (abbreviated as distance matrix) if there exist vectors $v_1, \cdots, v_n \in \Re^r$ for some $r \geq 1$ such that

$$D_{ij} = \|v_i - v_j\|^2, \; i, j = 1, \cdots, n.$$

The smallest value of r is called the embedding dimension of D. Note that $r \leq n - 1$ always.

The distance matrix is closely related to positive semidefinite matrix. The following basic connection was established by Schoenberg [58].

Proposition 6.1.1. *Given an $n \times n$ symmetric matrix $D = (D_{ij})$ with zero diagonal entries, consider the $(n-1) \times (n-1)$ symmetric matrix $X = (X_{ij})$ defined by*

$$X_{ij} \equiv \frac{1}{2}(D_{in} + D_{jn} - D_{ij}), \; i, j = 1, \cdots, n-1. \tag{6.1}$$

Then, D is a distance matrix if and only if X is a positive semidefinite *matrix.*

The applications of the distance matrix completion problem come from many areas such as multidimensional scaling in statistics [43] and molecular conformation problems in chemistry [13]. Some of these applications require a low embedding dimension. About many useful theoretical properties of distance matrix completion problem, interested readers can refer to the survey article by Laurent [42].

As pointed out in [67], one cannot provide an efficient rule to decide whether a distance matrix completion exists or not. Thus it seems more reasonable to allow approximate completions. In [2], Alfakih et al. introduced the following weighted closest Euclidean distance matrix problem.

$$\begin{aligned} \min \quad & \|H \circ (A - D)\|^2 \\ \text{s.t.} \quad & D \in \Upsilon, \end{aligned} \tag{6.2}$$

where A is a real symmetric partial matrix with zero diagonal entries, Υ denotes the cone of distance matrix, H is an $n \times n$ symmetric matrix with nonnegative elements, and $\circ$ denotes Hadamard product. For notational purposes, we assume that the free elements of A are set to 0 if they are not specified. Note that $H_{ij} = 0$ means that D_{ij} is free, while $H_{ij} > 0$ puts a weight to force the component $D_{ij} \approx A_{ij}$, i.e., D_{ij} is approximately fixed. We can add other linear equality constraints to force some components of D to exactly equal the corresponding components of A.

Then Alfakih et al. reformulated (6.2) as an equivalent SDP problem with quadratic objective

$$
\begin{aligned}
\min_{X,D} \quad & \|H \circ (A - D)\|^2 \\
\text{s.t.} \quad & X_{ij} = \frac{1}{2}(D_{in} + D_{jn} - D_{ij}),\ i, j = 1, \cdots, n-1 \\
& X \succeq 0,
\end{aligned}
\tag{6.3}
$$

and used a primal-dual interior point algorithm to solve it.

We notice Problem (6.3) is also of the form of MQSDP, therefore the modified ADM Algorithm 3.1.1 applies.

6.2 The Completion Problem of Low-rank Matrix

In many fields of engineering and science, a low-rank matrix need to be completed from small portion of entries observed. A good example is the well known Netflix problem [1]. This US large online DVD renting company needs to provide recommendations to users based on their submitted ratings on some films. That means one would like to infer their preference for unrated items. This problem seems very hard in that we should fill in the missing entries of the matrix from only small samples. However, the matrix of all user-ratings to recover has low rank because there is only a few factors to explain an individual's preference for films. Then it can be modelled as follows.

$$\begin{aligned} \min \quad & \operatorname{rank}(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij},\ (i,j) \in \Omega \\ & X \in \Re^{m\times n}, \end{aligned} \tag{6.4}$$

where M is the unknown matrix and Ω is a set of pairs of indices for known entries.

To generalize, the affine rank minimization problem is introduced.

$$\begin{aligned} \min \quad & \operatorname{rank}(X) \\ \text{s.t.} \quad & \mathcal{A}(X) = b \\ & X \in \Re^{m\times n}, \end{aligned} \tag{6.5}$$

where $\mathcal{A} : \Re^{m\times n} \to \Re^p$ is a linear operator and $b \in \Re^p$. This slight generalization appears useful in many areas such as machine learning, control, and Euclidean embedding.

Notice that the affine rank minimization problem (6.5) is an NP-hard nonconvex optimization problem. A convex relaxation is given in [21].

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \mathcal{A}(X) = b \\ & X \in \Re^{m\times n}, \end{aligned} \tag{6.6}$$

where $\|X\|_*$ is the nuclear norm of X. The nuclear norm of X is defined as

$$\|X\|_* = \sum_{i=1}^{q} \sigma_i(X),$$

where $q = \min\{m, n\}$ and $\sigma_i(X),\ i = 1, \cdots, q$, is the singular value of X. Actually the nuclear norm is the best convex approximation of the rank function over the unit ball of matrices. Candes and Recht [10] proved that a random low-rank matrix can be recovered exactly with high probability from a rather small portion of entries by solving (6.6).

The problem (6.6) can be reformulated as a SDP problem [54].

$$\begin{aligned} \min \quad & \frac{1}{2}\left(\langle W_1, I_m\rangle + \langle W_2, I_n\rangle\right) \\ \text{s.t.} \quad & \mathcal{A}(X) = b \\ & \begin{pmatrix} W_1 & X \\ X^T & W_2 \end{pmatrix} \succeq 0. \end{aligned} \tag{6.7}$$

In [10] SDPT3, one of the most advanced SDP solvers based on interior point methods, has been used to solve (6.7). However, the computational cost grows very fast as m and n increase.

The first order methods may therefore provide a promising alternative to the interior point method due to their low sensitivity to problem sizes. Ma et al. [47] proposed a Bregman iterative algorithm for solving (6.6). Recently, Cai et al. [9] proposed a singular value thresholding algorithm for solving the following Tikhonov regularized version of (6.6).

$$\begin{aligned} \min \quad & \|X\|_* + \frac{1}{2\beta}\|X\|^2 \\ \text{s.t.} \quad & \mathcal{A}(X) = b \\ & X \in \Re^{m\times n}, \end{aligned} \tag{6.8}$$

where $\beta > 0$ is a given parameter. They also showed that if β goes to ∞, the sequence of optimal solution X_β^* for (6.8) converges to the optimal solution of (6.6) with minimum Frobenius norm. Hence this algorithm approximately solves (6.6) for sufficiently large β.

Another possible model for the rank minimization problem is the nuclear norm regularized least squares problem.

$$\min \ \frac{1}{2}\|\mathcal{A}(X) - b\|^2 + \mu\|X\|_*, \tag{6.9}$$

where $\mu > 0$ is a given parameter. Here $\mathcal{A}(X) = b$ might not be feasible because of the existence of noise. Problem (6.9) is an unconstrained nonsmooth convex optimization problem. In [65], Toh and Yun proposed an accelerated proximal gradient algorithm, which terminates in $O(\frac{1}{\sqrt{\epsilon}})$ iterations with an ϵ-optimal solution, to solve it.

We point out that the ADM is also applicable here. Besides, it only needs mild condition for convergence. We may re-write (6.6) equivalently as

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & Y = X \\ & Y \in \Omega \equiv \{Y : \mathcal{A}(Y) = b\} \\ & X \in \Re^{m\times n}. \end{aligned} \tag{6.10}$$

When applied to Problem (6.10), the detail of ADM is shown as follows, where β is certain positive scalar.

Algorithm 6.2.1. The ADM for Problem (6.10)

Do at each iteration until a stopping criterion is met

Step 1. $\left(X^k, Y^k, \lambda^k\right) \rightarrow \left(X^{k+1}, Y^k, \lambda^k\right)$, *where* X^{k+1} *solves*

$$\min_{X\in\Re^{m\times n}} \|X\|_* - \left\langle \lambda^k, X\right\rangle + \frac{\beta}{2}\left\|X - Y^k\right\|^2 \tag{6.11}$$

Step 2. $\left(X^{k+1}, Y^k, \lambda^k\right) \rightarrow \left(X^{k+1}, Y^{k+1}, \lambda^k\right)$, *where* Y^{k+1} *solves*

$$\min_{Y\in\Omega} \left\langle \lambda^k, Y\right\rangle + \frac{\beta}{2}\left\|X^{k+1} - Y\right\|^2 \tag{6.12}$$

Step 3. $\left(X^{k+1}, Y^{k+1}, \lambda^k\right) \to \left(X^{k+1}, Y^{k+1}, \lambda^{k+1}\right)$, *where*

$$\lambda^{k+1} = \lambda^k - \beta\left(X^{k+1} - Y^{k+1}\right) \tag{6.13}$$

At Step 1 and Step 2 we need to solve two sub-optimization problems. (6.11) is an unconstrained optimization problem and can be reformulated to

$$\min_{X \in \Re^{m \times n}} \ \|X\|_* + \frac{\beta}{2}\left\|X - \left(Y^k + \frac{1}{\beta}\lambda^k\right)\right\|^2 \tag{6.14}$$

after ignoring some constant term. Fortunately we can solve (6.14) analytically. There is an important lemma about its solution. For the proof, see Theorem 2.1 of [9] or Theorem 3 of [47].

Lemma 6.2.2. *The solution of the minimization problem*

$$\min_{X \in \Re^{m \times n}} \ \frac{\tau}{2}\|X - G\|^2 + \mu\|X\|_*$$

for $\tau, \mu > 0$ *is given in a closed form by*

$$S_\tau(G) = U Diag\left(\left(\sigma - \frac{\mu}{\tau}\right)_+\right) V^T, \tag{6.15}$$

where $G = U\Sigma V^T$ *and* $\Sigma = Diag(\sigma)$ *are from the singular value decomposition (SVD) of* G.

Let $G^k = Y^k + \frac{1}{\beta}\lambda^k$, then for solving (6.14) we only need to compute

$$X^{k+1} = S_\beta\left(G^k\right),$$

where $S_\beta\left(G^k\right)$ has the form in (6.15) but with $\mu = 1$.

It is easy to see that (6.12) is equivalent to the following nonlinear equation

$$Y^{k+1} = P_\Omega\left[Y^{k+1} - \alpha\left(\lambda^k - \beta\left(X^{k+1} - Y^{k+1}\right)\right)\right],$$

where α can be any positive number. Thus by choosing $\alpha = \frac{1}{\beta}$, the right hand side item Y^{k+1} is cancelled. That is, in order to solve (6.12) we only have to compute

$$Y^{k+1} = P_\Omega\left[X^{k+1} - \frac{1}{\beta}\lambda^k\right]. \tag{6.16}$$

In summary, the refined ADM for solving (6.10) is given as follows.

Algorithm 6.2.3. The Refined ADM for Problem (6.10)

Do at each iteration until a stopping criterion is met

Step 1. $\left(X^k, Y^k, \lambda^k\right) \to \left(X^{k+1}, Y^k, \lambda^k\right)$, *where*

$$X^{k+1} = S_\beta\left(G^k\right) \tag{6.17}$$

Step 2. $\left(X^{k+1}, Y^k, \lambda^k\right) \rightarrow \left(X^{k+1}, Y^{k+1}, \lambda^k\right)$, *where*

$$Y^{k+1} = P_\Omega \left[X^{k+1} - \frac{1}{\beta}\lambda^k\right] \tag{6.18}$$

Step 3. $\left(X^{k+1}, Y^{k+1}, \lambda^k\right) \rightarrow \left(X^{k+1}, Y^{k+1}, \lambda^{k+1}\right)$, *where*

$$\lambda^{k+1} = \lambda^k - \beta\left(X^{k+1} - Y^{k+1}\right) \tag{6.19}$$

The main computational cost at each iteration of the refined ADM lies on computing the SVD of G^k. However, it suffices to know those singular values greater than the parameter $\frac{1}{\beta}$ and corresponding singular vectors. Therefore if this parameter is larger, the singular values to be evaluated is smaller. This motivates us to choose small β to make the decomposition of large-scale matrix possible.

7. NUMERICAL EXPERIMENTS

In this chapter, we present primary numerical results for the modified ADMs in solving MSDP problems. We should emphasize that our purpose here is not to conduct extensive computational tests but to demonstrate that the algorithms proposed are correct and can be potentially efficient. These algorithms may be taken as prototypes of those sophisticated and tailor-made algorithms for solving different classes of problems.

The codes were written in MATLAB (version 6.5) and the computations were performed on a 1.86 GHz Intel Core 2 PC with 3GB of RAM.

7.1 The Covariance Matrix Estimation Problem

We consider the following testing examples.

Example 1. QSDPs arising from the nearest correlation matrix problem (5.2). The matrix C is generated from the MATLAB segment: $x = 10\hat{}[-4 : 4/(n-1) : 0]$; C=gallary($'$randcorr$'$, $n * x/\mathrm{sum}(x)$). For the test purpose, we perturb C to

$$C = C + 10^{-3} * E, \text{ or } C = C + 10^{-2} * E, \text{ or } C = C + 10^{-1} * E,$$

where E is a randomly generated symmetric matrix with entries in $[-1, 1]$. The MATLAB code for generating E is: $E = \text{rand}(n)$; $E = (E + E')/2$; for $i = 1 : n$; $E(i, i) = 1$; end. Note that we make the perturbation larger than $10^{-4} * E$ considered in [31]. To consider the robustness of our algorithm, we use three sets of starting point:

- a) $(X^0, Y^0, \lambda^0) = (C, C, 0)$;
- b) $(X^0, Y^0, \lambda^0) = (I_n, I_n, 0)$;
- c) $X^0 = \text{rand}(n)$; $X^0 = [X^0 + (X^0)']/2$; for $i = 1 : n$; $X^0(i, i) = 1$; end; $Y^0 = X^0$; $\lambda^0 = 0$.

We test for $n = 100, 500, 1000, 2000$, respectively.

Example 2. CQCQSDPs without linear constraints arising from the extended nearest correlation problem (5.4). The matrix C is generated from the MATLAB segment: $x = 10.\verb|^|[-4 : 4/(n-1) : 0]$; C=gallary('randcorr', $n * x/\text{sum}(x)$). For the test purpose, we perturb C in the following four situations:

- $C' = C + 10^{-1} * E'$; $C = C + 10^{-1} * E$;
- $C' = C + 10^{-2} * E'$; $C = C + 10^{-2} * E$;
- $C' = C + 10^{-2} * E'$; $C = C + 10^{-1} * E$;
- $C' = C + 10^{-1} * E'$; $C = C + 10^{-2} * E$;

where E and E' are two random symmetric matrices generated as in Example 1. We use $(X^0, Y^0, \lambda^0) = (C, C, 0)$ as the starting point. We take $\epsilon =$

$r * \mathrm{norm}(C - C')$ with $r = 0 : 0.2 : 1.2$ to consider the effect of trust region size for $n = 100$. Then using $r = 0.8$, we test for $n = 100, 500, 1000, 2000$, respectively.

Example 3. The same as Example 2 but the diagonal entries of variable matrix are additionally required to be ones, i.e., it is a correlation matrix. We use $(X^0, Y_1^0, Y_2^0, \lambda_1^0, \lambda_2^0) = (C, C, C, 0, 0)$ as the starting point and set $\epsilon = 0.8 * \mathrm{norm}(C - C')$.

It is worth to mention that we simply use the parameters $\gamma = 1$ and $\beta = 1$ in the modified ADMs although we can possibly adjust them to further reduce the iteration numbers. The convergence was checked at the end of each iteration using the condition,

$$\frac{\max\{\|X^k - X^{k-1}\|_\infty, \|Y^k - Y^{k-1}\|_\infty, \|\lambda^k - \lambda^{k-1}\|_\infty\}}{\max\{\|X^1 - X^0\|_\infty, \|Y^1 - Y^0\|_\infty, \|\lambda^1 - \lambda^0\|_\infty\}} \leq 10^{-6}.$$

We also set the maximum number of iterations to 500.

The main computational cost at each iteration is matrix eigenvalue decomposition. The performance results of our modified ADMs are reported in Tables 1-4. The columns corresponding to "No. It" give the iteration numbers and the columns corresponding to "CPU Sec." give the CPU time in seconds. "$*$" means that the algorithm reaches the set maximum number of iterations before the accuracy is achieved.

From the numerical results reported in Table 1, we can see for problem's size $n = 100, 500, 1000, 2000$, the algorithm obtains the solutions mostly in less than 30 iterations and with reasonable accuracy 10^{-6}. These results

Table 1: Numerical results for Example 1

Example 1		$C+10^{-3}*E$		$C+10^{-2}*E$		$C+10^{-1}*E$	
n=	case	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
100	a)	7	0.4	14	0.6	24	1.0
	b)	20	0.9	21	0.9	28	1.2
	c)	21	1.0	20	0.9	24	1.1
500	a)	10	47.3	14	60.4	23	90.7
	b)	20	95.0	21	92.1	27	111.5
	c)	23	105.1	23	98.5	25	109.1
1000	a)	10	370.7	15	537.9	24	777.3
	b)	20	701.2	22	730.2	29	957.5
	c)	23	809.7	23	791.2	26	843.2
2000	a)	11	2972	14	3843	25	6321
	b)	20	5485	23	6377	31	7956
	c)	24	6362	24	6408	27	6823

Table 2: Numerical results for Example 2 with different trust region sizes

Example 2 n=100	$C=C+10^{-1}*E$ $C'=C+10^{-1}*E'$	$C=C+10^{-2}*E$ $C'=C+10^{-2}*E'$	$C=C+10^{-1}*E$ $C'=C+10^{-2}*E'$	$C=C+10^{-2}*E$ $C'=C+10^{-1}*E'$
r=	No. It	No. It	No. It	No. It
0	*	*	*	*
0.2	*	*	47	58
0.4	*	26	22	26
0.6	23	17	15	15
0.8	17	17	14	15
1.0	1	9	2	1
1.2	1	1	1	1

Table 3: Numerical results for Example 2 with $r = 0.8$

Example 2 r=0.8	$C=C+10^{-1}*E$ $C'=C+10^{-1}*E'$		$C=C+10^{-2}*E$ $C'=C+10^{-2}*E'$		$C=C+10^{-1}*E$ $C'=C+10^{-2}*E'$		$C=C+10^{-2}*E$ $C'=C+10^{-1}*E'$	
n=	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
100	17	0.9	16	0.9	15	0.9	16	0.9
500	16	84.2	16	80.8	13	65.2	17	84.7
1000	16	622.2	16	648.3	13	516.3	17	693.7
2000	15	4638	15	5047	13	4054	18	5726

Table 4: Numerical results for Example 3

Example 3 r=0.8	C=C+10^{-1}*E C′=C+10^{-1}*E′		C=C+10^{-2}*E C′=C+10^{-2}*E′		C=C+10^{-1}*E C′=C+10^{-2}*E′		C=C+10^{-2}*E C′=C+10^{-1}*E′	
n=	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
100	53	2.5	33	1.6	36	1.8	36	1.7
500	38	161.2	33	147.0	35	151.6	35	157.3
1000	37	1248	33	1169	36	1257	35	1254
2000	36	9571	33	9207	36	9665	36	10103

are comparative with those in [25, 64]. Actually the usage of CPU time by our proposed algorithm is between the result reported in [25] and the result reported in [64] for solving same scale problems. Furthermore, the modified ADM is quite robust for solving the nearest correlation problem (5.2) because it is little affected by the choices of starting point.

For cases with quadratic constraint for which the algorithms in [25, 64] cannot apply, the numerical results reported in Tables 2-4 are also promising. Too small r results in empty feasible set while too large r results in the uselessness of the trust region constraint. These are all verified by the numerical results reported in Table 2. It seems $r = 0.8$ is a suitable parameter regardless of different choices of C and C'. Using this r for problem's size $n = 100, 500, 1000, 2000$, the numerical results reported in Tables 3 and 4 show that our algorithm is effective to solve CQCQSDPs both without linear constraints and with linear constraints.

7.2 The Matrix Completion Problem

The random matrix completion problems considered in our numerical experiments are as follows.

Example 4. Convex relaxation problem (6.10) of low-rank matrix completion problem. For each (n, r, p) triple, where n (we set $m = n$) is the matrix dimension, r is the predetermined rank, and p is the number of entries to sample, we generate $M = M_L M_R^T$ as in [10, 65], where M_L and M_R are $n \times r$ matrices with i.i.d. standard Gaussian entries. Then a subset Ω of p elements uniformly at random from $\{(i,j) : i = 1, \cdots, n, j = 1, \cdots, n\}$ is selected. Therefore, the linear map $\mathcal{A}$ is given by

$$\mathcal{A}(X) = X_\Omega,$$

where $X_\Omega \in \Re^p$ obtained from X by selecting those elements whose indices are in Ω. We take $\beta = 0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.5, 1$ to consider the effect of parameter for $n/r = 100/10$. Then using $\beta = 0.1$, we test for $n/r = 200/10, 200/20, 500/10, 500/20, 500/50$, respectively.

We choose the initial iterate to be $X^0 = Y^0 = \text{rand}(n)$ and $\lambda^0 = 0$. The stopping criterion we use is:

$$\frac{\left\|X^k - X^{k-1}\right\|_F}{\max\left\{\left\|X^k\right\|_F, 1\right\}} < 10^{-4}.$$

The accuracy of the computed solution X_{sol} by our algorithm can be mea-

Table 5: Numerical results for Example 4 with different β

Example 4	Unknown M			ADM		
β=	n/r	p	p/d_r	iter	#sv	error
0.01	100/10	5666	3	135	19	1.4e-02
0.02	100/10	5666	3	83	18	5.6e-03
0.05	100/10	5666	3	53	13	5.3e-03
0.08	100/10	5666	3	63	11	7.0e-04
0.1	100/10	5666	3	71	10	3.5e-04
0.2	100/10	5666	3	106	10	1.2e-03
0.5	100/10	5666	3	202	11	3.7e-03
1	100/10	5666	3	351	12	8.2e-03

sured by the relative error defined as follows:

$$\text{error} \equiv \frac{\left\|X_{\text{sol}} - M\right\|_F}{\|M\|_F},$$

where M is the original matrix.

For each case, we repeat the procedure 5 times and report the performance results of the refined ADM Algorithm 6.2.3 in Tables 5 and 6. The columns corresponding to "iter", "#sv", and "error" give the average number of iterations, the average number of nonzero singular values of the computed solution matrix, and the average relative error, respectively. As indicated in [9], an $n \times n$ matrix of rank r has $d_r \equiv r(2n - r)$ degrees of freedom. Then the ratio p/d_r is also shown in the tables.

In order to free ourselves from the distraction of having to consider the storage of too large matrices in MATLAB, we only use examples with moderate dimensions. Furthermore, we compute the full SVD of G^k to obtain $S_\beta\left(G^k\right)$ at each iteration k. From Table 5, it seems $\beta = 0.1$ is a suitable

Table 6: Numerical results for Example 4 with $\beta = 0.1$

Example 4	Unknown M			ADM		
β=	n/r	p	p/d_r	iter	#sv	error
0.1	200/10	15665	4	95	10	3.7e-04
0.1	200/20	22800	3	99	20	3.5e-04
0.1	500/10	49471	5	158	10	4.3e-04
0.1	500/20	78400	4	146	20	3.8e-04
0.1	500/50	142500	3	152	50	4.1e-04

parameter. Then using this β, the numerical results reported in Table 6 are competitive with those obtained by using the fixed point continuation algorithm and the accelerated proximal gradient algorithm in [65], which are proposed to solve easier unconstrained counterpart (6.9) instead.

8. CONCLUSIONS

We study several modified ADMs for solving MSDP problems. These methods only need first order information. They may be able to deal with large-scale problems when second order information is time-consuming or even impossible to obtain.

In order to avoid solving difficult sub-variational inequality problems on matrix space at each iteration, we establish a set of projection-based algorithms. We discussed ADMs in different ways to deal with quadratic objective and general nonlinear objective. These algorithms appear to be the most efficient when they are specialized to solve convex quadratic problems, either with linear or quadratic constraints, such as CQSDP and CQCQSDP. When they are specialized to solve CNLSDP a prediction phase and a correction phase should be used, which only double the work of computing projections.

A practical application comes from the covariance matrix estimation problem. We proposed two new models, the extended nearest correlation matrix problem and the covariance matrix estimation in multiple-factor model, which are special cases of MSDP problems. Another practical application is from the matrix completion problem. We considered the completion problem of distance matrix and the completion problem of low-rank matrix. Both of them can be modelled as convex matrix optimization problems and certain modified ADM applies.

We also conducted numerical tests for problems arising from the aforementioned applications. Although the numerical results are preliminary, we are encouraged by the simplicity of the program codes, and the ability of the codes to handle medium to large sized problems. We conclude that the ADM is a promising method for MSDP.

A potential disadvantage of the first order methods, including the proposed modified ADMs, is that they cannot obtain highly accurate optimal solutions, compared with the second order methods such as the Newton method. However, we think it may not be a concern for many practical applications such as the covariance matrix estimation problem and the matrix completion problem. Moreover, for large-scale problems, it is usually very hard to get a solution even if the solution is not accurate. In this regard, we believe that the first order methods still have a room to develop.

BIBLIOGRAPHY

[1] ACM SICKDD and Netflix, Proceedings of KDD Cup and Workshop, 2007. Avaiable online at http://www.cs.uic.edu/liub/KDD-cup-2007/proceedings.html.

[2] A.Y. Alfakih, A. Khandani, and H. Wolkowicz, Solving Euclidean distance matrix completion problems via semideinite programming. *Comput. Optim. Appl.* 12, 13-30 (1999).

[3] A. Beck, Quadratic matrix programming. *SIAM J. Optim.* 17, 1224-1238 (2007).

[4] A. Ben-Tal, F. Jarre, M. Kocvara, A. Nemirovski, and J. Zowe, Optimal design of trusses under a nonconvex global buckling constraint. *Optim. Eng.* 1, 189-213 (2000).

[5] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization.* MPS-SIAM, Series on Optimization, SIAM Philadelphia, 2001.

[6] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems.* Springer-Verlag, New York, 2000.

[7] S. Boyd, L.E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics, Vol. 15, SIAM Philadelphia, 1994.

[8] S. Boyd and L. Xiao, Least-squares covariance matrix adjustment. *SIAM J. Matrix Anal. Appl.* 27, 532-546 (2005).

[9] J.F. Cai, E.J. Candes, and Z. Shen, A singular value thresholding algorithm for matrix completion. *Working Paper*, National University of Singapore, September 2008.

[10] E.J. Candes and B. Recht, Exact matrix completion via convex optimization. *Found. Comput. Math.* 9, 717-772 (2009).

[11] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems. *Math. Program.* 64, 81-101 (1994).

[12] R. Correa and C.H. Ramirez, A global algorithm for nonlinear semidefinite programming. *SIAM J. Optim.* 15, 303-318 (2004).

[13] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*. Wiley, New York, 1988.

[14] Y.H. Dai, Fast algorithms for projection on an ellipsoid. *SIAM J. Optim.* 16, 986-1006 (2006).

[15] J. Douglas and H. Rachford, On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.* 82, 421-439 (1956).

[16] B.C. Eaves, On the basic theorem of complementarity. *Math. Program.* 1, 68-75 (1971).

[17] J. Eckstein, Splitting methods for monotone operators with applications to parallel optimization. *PhD Thesis*, Massachusetts Institute of Technology, 1989.

[18] J. Eckstein, Parallel alternating direction multiplier decomposition of convex programs. *J. Optim. Theory Appl.* 80, 39-62 (1994).

[19] J. Eckstein and D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* 55, 293-318 (1992).

[20] B. Fares, D. Noll, and P. Apkarian, Robust control via sequential semidefinite programming. *SIAM J. Control Optim.* 40, 1791-1820 (2002).

[21] M. Fazel, Matrix rank minimization with applications. *PhD Thesis*, Standford University, 2002.

[22] M. Fukushima, Application of the alternating directions method of multipliers to separable convex programming problems. *Comput. Optim. Appl.* 2, 93-111 (1992).

[23] D. Gabay, Applications of the method of multipliers to variational inequalities. In: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problem*, pp. 299-331. North-Holland, Amsterdam, The Netherlands, 1983.

[24] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Comput. Math. Appl.* 2, 17-40 (1976).

[25] Y. Gao and D. Sun, Calibrating least squares covariance matrix problems with equality and inequality constraints. To appear in *SIAM J. Matrix Anal. Appl.*

[26] M.S. Gowda, Y. Song, and G. Ravindran, On some interconnections between strict monotonicity, globally uniquely solvable, and P-properties in semidefinite linear complementarity problems. *Linear Algebra Appl.* 370, 355-368 (2003).

[27] M.S. Gowda and R. Sznajder, Automorphism invariance of P- and GUS-properties of linear transformations on Euclidean Jordan algebras. *Math. Oper. Res.* 31, 109-123 (2006).

[28] R. Grinold and R.N. Kahn, Multiple-factor models for portfolio risk. *Report*, BARRA.

[29] B.S. He, L.Z. Liao, D.R. Han, and H. Yang, A new inexact alternating directions method for monotone variational inequalities. *Math. Program.* 92, 103-118 (2002).

[30] B.S. He, L.Z. Liao, and M.J. Qian, Alternating projection based prediction-correction methods for structured variational inequalities. *J. Comput. Math.* 24, 683-710 (2006).

[31] N.J. Higham, Computing the nearest correlation matrix - a problem from finance. *IMA J. Numer. Anal.* 22, 329-343 (2002).

[32] F. Jarre, An interior method for nonconvex semidefinite programs. *Optim. Eng.* 1, 347-372 (2000).

[33] C. Kanzow and C. Nagel, Some structural properties of a Newton-type method for semidefinite programs. *J. Optim. Theory Appl.* 122, 219-226 (2004).

[34] C. Kanzow and C. Nagel, Quadratic convergence of a Newton-type method for semidefinite programs without strict complementarity. *SIAM J. Optim.* 15, 654-672 (2005).

[35] C. Kanzow, C. Nagel, H. Kato, and M. Fukushima, Successive linearization methods for nonlinear semidefinite programs. *Comput. Optim. Appl.* 31, 251-273 (2005).

[36] M. Kocvara and M. Stingl, PENNON: a code for convex nonlinear and semidefinite programmings. *Optim. Methods Softw.* 18, 317-333 (2003).

[37] M. Kocvara and M. Stingl, Solving nonconvex SDP problems of structural optimization with stability control. *Optim. Methods Softw.* 19, 595-609 (2004).

[38] L. Kong, J. Sun, and N. Xiu, A regularized smoothing Newton method for symmetric cone complementarity problems. *SIAM J. Optim.* 19, 1028-1047 (2008).

[39] S. Kontogiorgis, Alternating directions methods for the parallel solution of large-scale block-structured optimization problems. *PhD Thesis*, University of Wisconsin-Madison, 1994.

[40] S. Kontogiorgis, R.D. Leone, and R.R. Meyer, Alternating direction splittings for block angular parallel optimization. *J. Optim. Theory Appl.* 90, 1-29 (1996).

[41] S. Kontogiorgis and R.R. Meyer, A variable-penalty alternating directions method for convex optimization. *Math. Program.* 83, 29-53 (1998).

[42] M. Laurent, A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. In: *Topics in Semidefinite and Interior-point Methods*, the Fields institute for research in mathematical sciences, communications series, vol. 18, American Mathematical Society: Providence, RI, 1998.

[43] J.D. Leeuw and W. Heiser, Theory of multidimensional scaling. In: P.R. Krishnaiah and L.N. Knanl, eds., *Handbook of Statistics*, vol. 2, pp. 285-316, North-Holland, 1982.

[44] F. Leibfitz and M.E. Mostafa, An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems. *SIAM J. Optim.* 12, 1048-1074 (2002).

[45] A. Lin, A class of methods for projection on a convex set. *Adv. Model. Optim.* 5, 211-221 (2003).

[46] P. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* 16, 964-979 (1979).

[47] S. Ma, D. Goldfarb, and L. Chen, Fixed point and bregman iterative methods for matrix rank minimization. To appear in *Math. Program.*

[48] J. Malick, A dual approach to semidefinite least-squares problems. *SIAM J. Matrix Anal. Appl.* 26, 272-284 (2004).

[49] H. Markowitz, Portfolio selection. *J. Finance* 7, 77-91 (1952).

[50] J.W. Nie and Y.X. Yuan, A predictor- corrector algorithm for QSDP combining Dikin-type and Newton centering steps. *Ann. Oper. Res.* 103, 115-133 (2001).

[51] D. Noll and P. Apkarian, Spectral bundle methods for non-convex maximum eigenvalue functions: first-order methods. *Math. Program.* 104, 701-727 (2005).

[52] D. Noll and P. Apkarian, Spectral bundle methods for non-convex maximum eigenvalue functions: second-order methods. *Math. Program.* 104, 729-747 (2005).

[53] M. Overton and H. Wolkowicz (eds.), Semidefinite programming, dedicated to the memory of svatopluk ployak. *Math. Program.* 77, 97-320 (1997).

[54] B. Recht, M. Fazel, and P. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *Working Paper*, California Institute of Technology, June 2007.

[55] R.T. Rockafellar, *Network Flows and Monotropic Optimization.* Wiley, New York, 1984.

[56] R.T. Rockafellar and R.J.B. Wets, Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.* 16, 119-147 (1991).

[57] H. Qi and D. Sun, A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM J. Matrix Anal. Appl.* 28, 360-385 (2006).

[58] I.J. Schoenberg, Remarks to Maurice Frechet's article: Sur la definition axiomatique d'une classe d'espaces vectoriels distancies applicables vectoriellement sur l'espace de Hilbert. *Ann. Math.* 36, 724-732 (1935).

[59] A. Shapiro, First and second order analysis of nonlinear semidefinite programs. *Math. Program.* 77, 301-320 (1997).

[60] D. Sun, The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Math. Oper. Res.* 31, 761-776 (2006).

[61] J. Sun, D. Sun, and L. Qi, A squared smoothing Newton method for nonsmooth matrix equations and its applications in semidefinite optimization problems. *SIAM J. Optim.* 14, 783-806 (2004).

[62] D. Sun, J. Sun, and L. Zhang, Rates of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math. Program.* 114, 349-391 (2008).

[63] J. Sun, K.C. Toh, and G. Zhao, An analytic cutting plane method for semidefinite feasibility problems. *Math. Oper. Res.* 27, 332-346 (2002).

[64] K.C. Toh, An inexact primal-dual path following algorithm for convex quadratic SDP. *Math. Program.* 112, 221-254 (2007).

[65] K.C. Toh and S. Yun, An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. To appear in *Pac. J. Optim.*

[66] K.C. Toh, G. Zhao, and J. Sun, A multiple-cut analytic cutting plane method for semidefinite feasibility problems. *SIAM J. Optim.* 12, 1126-1146 (2002).

[67] M.W. Trosset, Distance matrix completion by numerical optimization. *Comput. Optim. Appl.* 17, 11-22 (2000).

[68] L. Vandenberghe and S. Boyd, Semidefinite programming. *SIAM Rev.* 38, 49-95 (1996).

[69] H. Yamashita, H. Yabe, and K. Harada, A primal-dual interior point method for nonlinear semidefinite programming. *Technical Report*, Tokyo University of Science, 2007.