An Augmented Lagrangian Dual Approach for the H-Weighted Nearest Correlation Matrix Problem

Houduo Qi* and Defeng Sun†

March 3, 2008

Abstract

In [15], Higham considered two types of nearest correlation matrix problem, namely the W-weighted case and the H-weighted case. While the W-weighted case has since then been well studied to make several Lagrangian dual based efficient numerical methods available, the H-weighted case remains numerically challenging. The difficulty of extending those methods from the W-weighted case to the H-weighted case lies in the fact that an analytic formula for the metric projection onto the positive semidefinite cone under the H-weight, unlike the case under the W-weight, is not available. In this paper, we introduce an augmented Lagrangian dual based approach, which avoids the explicit computation of the metric projection under the H-weight. This method solves a sequence of unconstrained strongly convex optimization problems, each of which can be efficiently solved by a semismooth Newton method combined with the conjugate gradient method. Numerical experiments demonstrate that the augmented Lagrangian dual approach is not only fast but also robust.

AMS subject classifications. 49M45, 90C25, 90C33

1 Introduction

In [15], Higham considered two types of *nearest correlation matrix* problem. One is under the W-weighting:

$$\begin{array}{ll} \min & \frac{1}{2}\| W^{1/2}(X-G)W^{1/2} \|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X \in \mathcal{S}^n_+\,, \end{array} \tag{1}$$

where $\mathcal{S}^n$ and $\mathcal{S}^n_+$ are respectively the space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$, $\|\cdot\|$ is the Frobenius norm induced by the standard trace

*School of Mathematics, The University of Southampton, Highfield, Southampton SO17 1BJ, UK. This author's research was partially supported by EPSRC Grant EP/D502535/1. E-mail: hdqi@soton.ac.uk

†Department of Mathematics and Risk Management Institute, 2 Science Drive 2, National University of Singapore, Republic of Singapore. The author's research was supported by the Academic Research Fund under Grant R-146-000-104-112 and the Risk Management Institute under Grants R-703-000-004-720 and R-703-000-004-646, National University of Singapore. Email: matsundf@nus.edu.sg.

inner product in $\mathcal{S}^n$, and the matrix $G \in \mathcal{S}^n$ is given. The positive definite matrix $W \in \mathcal{S}^n$ is known as the W-weight to the problem and $W^{1/2}$ is the positive square root of W.

The constraints in (1), collectively known as the *correlation constraints*, specify that any feasible matrix is a correlation matrix. Solving the W-weighted problem (1) is equivalent to solving a problem of the type (cf. [22, Sec. 4.1]):

$$\begin{array}{ll} \min & \frac{1}{2}\| X - G \|^2 \\ \text{s.t.} & \mathcal{A}(X) = e \,, \\ & X \in \mathcal{S}^n_+ \,, \end{array} \tag{2}$$

where the linear operator $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^n$ is given by $\mathcal{A}(X) = \mathrm{diag}(W^{-1/2} X W^{-1/2})$ and $e \in \mathbb{R}^n$ is the vector of all ones. We often use $X \succeq 0$ to denote $X \in \mathcal{S}^n_+$.

The other nearest correlation matrix problem that is considered by Higham is under the H-weighting:

$$\begin{array}{ll} \min & \frac{1}{2}\| H \circ (X - G) \|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n \,, \\ & X \in \mathcal{S}^n_+ \,, \end{array} \tag{3}$$

where the weighting is now in the sense of Hardamard: $(A \circ B)_{ij} = A_{ij} B_{ij}$. Here the matrix H is symmetric and each of its entries is positive, i.e., $H_{ij} > 0$ for all $i, j = 1, \ldots, n$. We refer the reader to [3] for a concrete example in finance to see how H was constructed. We note that in the special case that $H = E$, the matrix of all ones, (3) turns out to be (2) with $W = I$, the identity matrix.

The W-weighted problem (1) has been well studied since Higham [15] and now there are several good methods for it including the alternating projection method [15], the gradient and quasi-Newton methods [18, 8], the semismooth Newton method combined with the conjugate gradient solver [22] and its modified version with several (preconditioned) iterative solvers [5], and the inexact interior-point methods with iterative solvers [39, 38]. All of these methods except the inexact interior-point methods crucially rely on the fact that the projection of a given matrix $X \in \mathcal{S}^n$ onto $\mathcal{S}^n_+$ under the W-weight, denoted by $\Pi^W_{\mathcal{S}^n_+}(X)$, which is the optimal solution of the following problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|W^{1/2}(Y - X)W^{1/2}\|^2 \\ \text{s.t.} & Y \in \mathcal{S}^n_+ \,, \end{array}$$

is given by the formula (see [15, Thm. 3.2])

$$\Pi^W_{\mathcal{S}^n_+}(X) = W^{1/2}\Big(W^{1/2} X W^{1/2}\Big)_+ W^{1/2},$$

where for any $A \in \mathcal{S}^n$,

$$A_+ := \Pi^I_{\mathcal{S}^n_+}(A) \,.$$

It has long been known by statisticians that for any $A \in \mathcal{S}^n$, $\Pi_{\mathcal{S}^n_+}(A) \equiv \Pi^I_{\mathcal{S}^n_+}(A)$ admits an explicit formula [32]. This means that for any $X \in \mathcal{S}^n$, $\Pi^W_{\mathcal{S}^n_+}(X)$ can be computed explicitly.

To simplify the subsequent discussions, we assume, without loss of generality, that $W \equiv I$ (for reasons, see [22, Sec. 4.1]). We note that $(\cdot)_+ = \Pi_{\mathcal{S}^n_+}(\cdot)$. To see how the metric projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ is tangled in the derivation of these methods mentioned above and also to motivate our method for the H-weighted case, let us consider the Lagrangian function of problem (2),

$$l(X,y) := \frac{1}{2}\|X - G\|^2 + y^T(b - \mathcal{A}(X)), \quad (X,y) \in \mathcal{S}^n_+ \times \mathbb{R}^n\,, \tag{4}$$

where the linear operator $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^n$ is the diagonal operator, i.e, $\mathcal{A}(X) = \mathrm{diag}(X)$ for any $X \in \mathcal{S}^n$ and $b := e$. Since problem (2) automatically satisfies the generalized Slater constraint qualification, from the duality theory developed by Rockafellar [28] we know that problem (2) can be equivalently solved by its Lagrangian dual problem

$$\max_{y \in \mathbb{R}^n} \left\{ \min_{X \in \mathcal{S}^n_+} \; l(X,y) \right\}\,, \tag{5}$$

which, via the metric projector $(\cdot)_+$, can be equivalently reformulated as the following unconstrained optimization problem (see [28, 18, 8] for details)

$$\min_{y \in \mathbb{R}^n} \theta(y) := \frac{1}{2}\|(G + \mathcal{A}^*(y))_+\|^2 - b^T y - \frac{1}{2}\|G\|^2 \tag{6}$$

in the sense that if $\bar{y}$ is an optimal solution to (6), then $\overline{X} := (G + \mathcal{A}^*(\bar{y}))_+$ solves (2). Here $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$.

The objective function $\theta(\cdot)$ in (6) is known to be once continuously differentiable and convex [28], despite the fact that the projection operator $(\cdot)_+$ is usually not differentiable. Therefore, the gradient method and quasi-Newton methods can be developed to solve (6) directly. Malick remarked in [18] that the alternating projection method is actually the gradient method for (6) with a constant steplength one. These methods converge at best linearly. Because $\theta(\cdot)$ is convex and coercive [28], solving (6) is equivalent to finding a point $\bar{y}$ satisfying its optimality condition

$$\nabla\theta(y) = \mathcal{A}(G + \mathcal{A}^*(y))_+ - b = 0.$$

Define

$$F(y) := \mathcal{A}(G + \mathcal{A}^*(y))_+, \quad y \in \mathbb{R}^n\,.$$

Then $F(\cdot)$ is Lipschitz continuous and whence the generalized Jacobian $\partial F(y)$ in the sense of Clarke [10] is well-defined. For any $y \in \mathbb{R}^n$, let $\partial^2\theta(y) := \partial F(y)$. The generalized Newton method takes the following form:

$$y^{k+1} = y^k - V_k^{-1}(\nabla\theta(y^k)), \qquad V_k \in \partial^2\theta(y^k), \; k = 0, 1, \ldots. \tag{7}$$

A formula for calculating $V \in \partial^2\theta(y)$ can be found in [22, p.378]. Clarke's Jacobian based generalized Newton method (7) was thoroughly analyzed by Qi and Sun [22] and is proven to be quadratically convergent. Numerical experiments conducted in [22, 5] seem to confirm that it is the most effective method available so far.

When it comes to the H-weighted problem (3), it is unfortunate that all of those Lagrangian dual based methods become infeasible mainly due to the lack of a computable formula for the projection of $X \in \mathcal{S}^n$ onto $\mathcal{S}^n_+$ under the H-weight. That is, the optimal solution, denoted $\Pi^{H\circ}_{\mathcal{S}^n_+}(X)$, to the following problem

$$\min \frac{1}{2}\|H \circ (Y - X)\|^2, \quad \text{s.t. } \; Y \in \mathcal{S}^n_+$$

is not known to have an explicit formula[1]. For this reason, it is not known if the Lagrangian dual problem for the H-weighted case can be reduced to an explicitly defined unconstrained optimization problem. Consequently, comparing to the original problem, not much benefit would be gained through considering the Lagrangian dual problem. This implicates, in particular, that the Newton method for the W-weighted case cannot be straightforwardly extended to the H-weighted case.

A natural question then arises: can we still expect an efficient dual approach for the H-weighted case? This paper will provide an affirmative answer to this question by exploiting the augmented Lagrangian dual approach – the augmented Lagrangian method, thanks to Rockafellar [29, 30] for his pioneering work on convex optimization problems. Let $c > 0$ be a parameter. The augmented Lagrangian function for the H-weighted problem (3) is given by (e.g., see [31, Sec. 11K])

$$\begin{aligned} L_c(X, y, Z) \;\; := \;\; & \frac{1}{2}\|H \circ (X - G)\|^2 + \; y^T(b - \mathcal{A}(X)) + \frac{c}{2}\|b - \mathcal{A}(X)\|^2 \\ & + \frac{1}{2c}\left(\|(Z - cX)_+\|^2 - \|Z\|^2\right), \end{aligned} \tag{8}$$

where $(X, y, Z) \in \mathcal{S}^n \times {\rm I\!R}^m \times \mathcal{S}^n$, $\mathcal{A} = \text{diag}$, and $b = e$. The augmented Lagrangian dual problem takes the following form

$$\min_{y \in {\rm I\!R}^n,\, Z \in \mathcal{S}^n} \left\{ \nu_c(y, Z) := - \min_{X \in \mathcal{S}^n} L_c(X, y, Z) \right\} . \tag{9}$$

The major computational task in the augmented Lagrangian dual approach, as outlined in (22)-(24), at each step for a given $(y, Z) \in {\rm I\!R}^n \times \mathcal{S}^n$, is to solve the following unconstrained optimization problem:

$$\min_{X \in \mathcal{S}^n} \; L_c(X, y, Z). \tag{10}$$

Note that for any $(y, Z) \in {\rm I\!R}^n \times \mathcal{S}^n$, $L(\cdot, y, Z)$ is a strongly convex and continuously differentiable function. Therefore, the gradient method and quasi-Newton methods can be developed in theory for (10). However, our numerical experiments show that the gradient method is extremely slow and is hence disregarded. The size of the variable X in (10) is $\bar{n} := n(n + 1)/2$. Maintaining an

---

[1] It was stated in [16, Cor. 2.2] that when H is positive definite, $\Pi^{H\circ}_{\mathcal{S}^n_+}(X)$ is uniquely determined by the equation

$$H \circ \Pi^{H\circ}_{\mathcal{S}^n_+}(X) = (H \circ X)_+ .$$

This does not seem to be true even for this special case. A counterexample is

$$H = \begin{bmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad 0 < \varepsilon \le 1/2.$$

$\bar{n} \times \bar{n}$ positive definite matrix is extremely expensive due to memory problems even when n is small, say $n = 100$. This rules out the quasi-Newton methods where an $\bar{n} \times \bar{n}$ positive definite matrix is maintained and updated at each iteration (limited-memory quasi-Newton methods may still be exploited, but their convergence analysis is hardly satisfactory).

The main purpose of this paper is to show that Newton's method is an efficient method for (10). The Newton method that we are going to use is quite similar to (7) with the difference that the number of the unknowns in the Newton equation here is $\bar{n}$, which is the order of $O(n^2)$, instead of n. These equations, even when n is relatively large, say $n = 1,000$, do not create too much difficulty when we apply the conjugate gradient method to solve them. The major reason behind this is that the H-weighted problem (3) has two mathematical properties, namely the automatic fulfillment of the *constraint nondegeneracy* and *the strong second-order sufficient condition* (see Sec. 2). These two properties not only ensure that the Newton equations encountered in the Newton method are well conditioned but also guarantee that the augmented Lagrangian method possesses a fast linear convergence, a property established by Rockafellar [29, 30] for general convex optimization problems. We will make all of those results solid in the main body of the paper.

A similar approach is also conducted in [23], where the problem considered is the type of the W-weighted with a background in correlation stress testing, which requires a large number of correlations to be fixed beforehand. Theoretically, being an augmented Lagrangian dual based method, the approach in [23] can be extended to the H-weighted case considered here. Indeed, it was [23], together with [22], that inspired us to further investigate the effectiveness of the augmented Lagrangian dual approach for the H-weighted problem (3).

The type of interior-point methods (IPMs) was deliberately left out of the above discussions because it deserves its own space for comments. As early as [16], Johnson et. al. have already started to use IPMs to solve the H-weighted matrix optimization problems of various types. The H-weighted nearest correlation matrix problem (3) can be reformulated as a linear optimization problem with mixed semidefinite and second order cone constraints [15, 39]. Consequently, publicly available IPMs based software like SeDuMi [34] and SDPT3 [41] can be applied to solve these problems directly. However, since at each iteration these solvers require to formulate and solve a dense Schur complement matrix (cf. [4]), which for the problem (3) amounts to a linear system of dimension $(n + \bar{n}) \times (n + \bar{n})$, the size of the H-weighted problem that can be solved on a Pentium IV PC (the computing machine that we are using) is limited to a small number, say $n = 80$ or 100 at most. Serious and competitive implementation of inexact IPMs was carried out by Toh et. al [39] for solving a special class of convex quadratic semidefinite programming (QSDP) including the W-weighted problem (1) and Toh [38] for a general convex QSDP with the H-weighted problem (1) being targeted and tested in particular. The search direction used in [38] was obtained by solving the *augmented equation* via the *preconditioned symmetric quasi-minimal residual* (PSQMR) iterative solver. It is this QSDP-solver that we are going to compare with. Our numerical tests show that the augmented Lagrangian dual approach for the H-weighted nearest correlation problem (3) is not only faster but also more robust.

The paper is organized as follows. In Section 2, we study some mathematical properties of the H-weighted problem (3), mainly on the constraint nondegeneracy and the strong second-order sufficient condition. Section 3 is on the augmented Lagrangian method. We first outline an abstract form of the method and then state its convergence results. In Section 3.2, we present two practical algorithms. One is the semismooth Newton-CG method for solving subproblems of the type (10) encountered in the augmented Lagrangian method, which is detailed in the second algorithm. Convergence analysis for the two algorithms is included in Section 3.3. We report numerical results in Section 4 and conclude the paper in Section 5.

2 Mathematical Properties of the H-Weighted Case

This section gives a brief account of the two mathematical properties of the H-weighted problem (3) mentioned in the introduction. The two properties will justify the use of the augmented Lagrangian method to be introduced in the next section. Although it is not our intention in this paper to address degenerate problems, the two properties will also provide some room to relax the positivity requirement on the elements of H. For example, some of the off-diagonal weights H_{ij} are allowed to be *zeros* without damaging the two properties. The *zero* weight means no restriction on the corresponding correlation.

2.1 The Constraint Nondegeneracy Property

Let us cast the problem (3) into the following convex QSDP:

$$\begin{array}{ll} \min & \frac{1}{2}\langle X,\ \mathcal{Q}(X)\rangle - \langle C,\ X\rangle + \frac{1}{2}\|H\circ G\|^2 \\ \text{s.t.} & \mathcal{A}(X) = b, \\ & X \in \mathcal{S}^n_+, \end{array} \tag{11}$$

where $\mathcal{A} = \mathrm{diag}$, $\mathcal{Q} = H\circ H\circ$, $C = H\circ H\circ G$, and $b = e$.

For any $X \in \mathcal{S}^n_+$, let $T_{\mathcal{S}^n_+}(X)$ be the tangent cone of $\mathcal{S}^n_+$ at X and $\mathrm{lin}\big(T_{\mathcal{S}^n_+}(X)\big)$ be the largest linear space contained in $T_{\mathcal{S}^n_+}(X)$, respectively. We say that the *constraint nondegeneracy* holds at a point X satisfying the constraints in (11) if

$$\mathcal{A}\big(\mathrm{lin}\ T_{\mathcal{S}^n_+}(X)\big) = \mathbb{R}^n. \tag{12}$$

For the origin of the constraint nondegeneracy, its various forms, and its role in general optimization, see [6, 7, 25, 26, 27, 33].

The constraint nondegeneracy can be easily verified for the correlation constraints. Let $X \in \mathcal{S}^n$. Suppose that X has the spectral decomposition

$$X = P\mathrm{diag}(\lambda_1, \ldots, \lambda_n)P^T, \tag{13}$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of X and P is a corresponding orthogonal matrix of orthonormal eigenvectors. Then, from [32, 14, 40] we know that

$$X_+ = P\mathrm{diag}(\max(0,\lambda_1), \ldots, \max(0,\lambda_n))P^T\,. \tag{14}$$

Define

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

Write $P = [P_\alpha\ P_\beta\ P_\gamma]$, where P_α contains columns in P indexed by α; and P_β and P_γ are defined similarly. The tangent cone $T_{\mathcal{S}^n_+}(X_+)$ was first characterized by Arnold [2] as follows

$$T_{\mathcal{S}^n_+}(X_+) = \{B \in \mathcal{S}^n \,\big|\, [\,P_\beta\ P_\gamma\,]^T B\,[\,P_\beta\ P_\gamma\,] \succeq 0\}\,.$$

Consequently,

$$\mathrm{lin}\,\big(T_{\mathcal{S}^n_+}(X_+)\big) = \big\{B \in \mathcal{S}^n \;\big|\; P_\beta^T B P_\beta = 0,\ P_\beta^T B P_\gamma = 0,\ P_\gamma^T B P_\gamma = 0\big\}\,.$$

Equivalently,

$$\mathrm{lin}\,\big(T_{\mathcal{S}^n_+}(X_+)\big) = \left\{ PBP^T \Big|\; B = \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\beta} & B_{\alpha\gamma} \\ B_{\alpha\beta}^T & 0 & 0 \\ B_{\alpha\gamma}^T & 0 & 0 \end{bmatrix},\ \begin{array}{l} B_{\alpha\alpha} \in \mathcal{S}^{|\alpha|} \\ B_{\alpha\beta} \in \mathbb{R}^{|\alpha|\times|\beta|} \\ B_{\alpha\gamma} \in \mathbb{R}^{|\alpha|\times|\gamma|} \end{array} \right\}. \tag{15}$$

The following result says that any point satisfying the correlation constraints is constraint nondegenerate. It can be proved similarly as [39, Prop. 4.2], where it used a characterization of the constraint nondegeneracy in [1] and [22, Lem. 3.3] and the result is stated only for optimal solutions. We provide here a proof for the general case.

Proposition 2.1 *Any point satisfying the correlation constraints* $\{\mathrm{diag}(X) = e,\ X \in \mathcal{S}^n_+\}$ *is constraint nondegenerate.*

Proof. Let $X \in \mathcal{S}^n$ satisfy the correlation constraints. Suppose that X has the spectral decomposition (13). Because X is positive semidefinite, $\gamma = \emptyset$. Also because $\mathrm{diag}(X) = e$, $\alpha \neq \emptyset$. Moreover, this diagonal constraint also implies (see [22, Lem. 3.3])

$$\sum_{\ell\in\alpha} P_{i\ell}^2 > 0, \qquad i = 1, \ldots, n. \tag{16}$$

To show that condition (12) holds at X, it suffices to prove

$$\Big(\mathrm{diag}\big(\mathrm{lin}\ T_{\mathcal{S}^n_+}(X)\big)\Big)^\perp = \{0\}\,.$$

Let $v \in \mathbb{R}^n$ be an arbitrary element of the left-hand side set of the above equation. We shall prove $v = 0$. It follows that for any $PBP^T \in \mathrm{lin}\,\big(T_{\mathcal{S}^n_+}(X)\big)$, we have

$$0 = \langle v,\ \mathrm{diag}(PBP^T)\rangle = \langle \mathrm{diag}(v),\ PBP^T\rangle = \langle P^T\mathrm{diag}(v)P,\ B\rangle, \tag{17}$$

where B is from (15). The structure of B implies

$$P^T\mathrm{diag}(v)P_\alpha = 0,$$

which in turn implies

$$0 = \mathrm{diag}(v)P_\alpha = \mathrm{diag}(v)\big(P_\alpha \circ P_\alpha\big).$$

Summing up each row of the above matrix equation yields

$$0 = v_i \sum_{\ell\in\alpha} P_{i\ell}^2, \quad i = 1, \ldots, n.$$

The property (16) ensures $v_i = 0$ for each $i = 1, \ldots, n$. This completes our proof. $\square$

2.2 The Strong Second-Order Sufficient Condition

Now let us consider the Karush-Kuhn-Tucker system of the QSDP (11)

$$\left.\begin{array}{rl} \mathcal{Q}(X) - \mathcal{A}^*(y) - Z & = C \\ \mathcal{A}(X) & = b \\ X \succeq 0,\ Z \succeq 0,\ \langle X,\ Z\rangle & = 0 \end{array}\right\}. \tag{18}$$

Any triple $(\overline{X}, \bar{y}, \overline{Z}) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ satisfying (18) is called a KKT point of (11). By using the fact that $\mathcal{S}^n_+$ is a self-dual cone, we know from Eaves [12] that $(\overline{X}, \bar{y}, \overline{Z}) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ satisfies the KKT conditions (18) if and only if it satisfies the following system of nonsmooth equations

$$F(X, y, Z) := \left[\begin{array}{c} \mathcal{Q}(X) - C - \mathcal{A}^*(y) - Z \\ b - \mathcal{A}(X) \\ Z - [Z - X]_+ \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right], \quad (X, y, Z) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n. \tag{19}$$

Apparently, F is globally Lipschitz continuous everywhere as $(\cdot)_+$ is so.

Let $(\overline{X}, \bar{y}, \overline{Z}) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ be a KKT point of problem (11). Denote $X := \overline{X} - \overline{Z}$. Suppose that X has the spectral decomposition (13). Define

$$\mathrm{app}\,(\bar{y}, \overline{Z}) := \left\{ B \in \mathcal{S}^n \;\middle|\; \mathcal{A}(B) = 0,\ P_\beta^T B P_\gamma = 0,\ P_\gamma^T B P_\gamma = 0 \right\}. \tag{20}$$

Note that $\mathrm{app}\,(\bar{y}, \overline{Z})$ is independent of the choice of P in (13) (see [35, Eq. (38) and Eq. (39)]). We also define

$$\mathcal{M}(\overline{X}) := \left\{ (\bar{y}, \overline{Z}) \;\middle|\; (\overline{X}, \bar{y}, \overline{Z}) \text{ is a KKT point of (11)} \right\}.$$

The set $\mathcal{M}(\overline{X})$ is known to be the set of Lagrangian multipliers at $\overline{X}$. For the H-weighted problem (3), $\mathcal{M}(\overline{X})$ contains a unique point $(\bar{y}, \overline{Z})$ because the constraint nondegeneracy holds at $\overline{X}$ by Prop. 2.1. For a proof on this, see [7, Thm. 5.85]. We say that the *strong second-order sufficient condition* (SSOSC) holds at $\overline{X}$ if

$$\langle B, H \circ H \circ B\rangle + \Upsilon_{\overline{X}}(\overline{Z}, B) > 0 \quad \forall\, 0 \neq B \in \mathrm{app}\,(\bar{y}, \overline{Z}), \tag{21}$$

where the term $\Upsilon_{\overline{X}}(\overline{Z}, B)$ is defined by

$$\Upsilon_{\overline{X}}(\overline{Z}, B) = \langle \overline{Z},\ B\overline{X}^\dagger B\rangle$$

and $\overline{X}^\dagger$ is the Moore-Penrose pseudo-inverse of $\overline{X}$. Note that $\Upsilon_{\overline{X}}(\overline{Z}, B)$ is quadratic in B and is always nonnegative because $\overline{Z} \succeq 0$ and $\overline{X} \succeq 0$. Note also that in the left hand side of (21), the first term $\langle B, H \circ H \circ B\rangle > 0$ for any $B \neq 0$ due to the assumption that $H_{ij} > 0$, $i, j, = 1, \ldots, n$. Therefore, we have the following statement.

Proposition 2.2 *Assume that $H_{ij} > 0$, $i, j = 1, \ldots, n$. Let $(\overline{X}, \bar{y}, \overline{Z})$ be the unique KKT point of the H-weighted nearest correlation matrix problem (3). Then the SSOSC (21) holds at $\overline{X}$.*

We make several remarks below.

(i) The SSOSC was first proposed by Sun [35] in the study of the strong regularity of nonlinear semidefinite programming (NSDP). The original definition runs over the set $\mathcal{M}(\overline{X})$. As this set is a singleton in our case, (21) is just a specialization of the original one given in [35, Def. 3.2].

(ii) In some practical cases [3], the diagonal weights H_{ii} are assigned *zero* values (i.e., $H_{ii} = 0$, $i = 1, \ldots, n$). This does not have any effect on (21) because for any $B \in \text{app}\,(\bar{y}, \overline{Z})$, we must have $B_{ii} = 0$, $i = 1, \ldots, n$ (see the definition (20)). Therefore, the diagonal weights in H have no contribution to the value $\langle B,\ H \circ H \circ B \rangle$. Consequently, our assumption in Prop. 2.2 can be replaced by $H_{ij} > 0$, for all $i \neq j$, $i, j = 1, \ldots, n$.

(iii) Furthermore, for the SSOSC (21) to hold at $\overline{X}$, one does not have to assume that all of the off-diagonal weights to be positive. In fact, as the following example shows, some of them are allowed to be *zeros* without damaging the SSOSC. This example also shows that too many zero off-diagonal weights do destroy the SSOSC (21).

Example 2.3 *Consider the H-weighted problem (3) in $\mathcal{S}^4$ with data given by*

$$H = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \textit{and} \quad G = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 0.5 \\ -1 & 1 & 0.5 & 1 \end{bmatrix}.$$

Such a matrix G is known as a pseudo-correlation matrix because $-1 \leq G_{ij} \leq 1$, $G_{ii} = 1$ for all $i, j = 1, \ldots, 4$, and $\lambda_{\min}(G) = -0.8860 < 0$. After running our augmented Lagrangian dual method–Algorithm 3.4, with some help of analytical cross validation, we found a KKT point $(\overline{X}, \bar{y}, \overline{Z})$ with

$$\overline{X} = \begin{bmatrix} 1 & -1 & \tau_1 & -\tau_1 \\ -1 & 1 & -\tau_1 & \tau_1 \\ \tau_1 & -\tau_1 & 1 & \tau_2 \\ -\tau_1 & \tau_1 & \tau_2 & 1 \end{bmatrix}, \quad \overline{Z} = \begin{bmatrix} 0 & 0 & \tau_1 - 1 & 1 - \tau_1 \\ 0 & 0 & 1 - \tau_1 & \tau_1 - 1 \\ \tau_1 - 1 & 1 - \tau_1 & 0 & \tau_2 - 0.5 \\ 1 - \tau_1 & \tau_1 - 1 & \tau_2 - 0.5 & 0 \end{bmatrix} + \text{diag}(\bar{y}),$$

and $\bar{y}_1 = \bar{y}_2 = 2\tau_1(1 - \tau_1)$, $\bar{y}_3 = \bar{y}_4 = \bar{y}_1 - \tau_2(\tau_2 - 0.5)$, and

$$\tau_1 = \left((1 + \sqrt{109/108})/4\right)^{1/3} - \left((-1 + \sqrt{109/108})/4\right)^{1/3}, \quad \textit{and} \quad \tau_2 = 1 - \tau_1^2.$$

Therefore, $\overline{X}$ is an optimal solution (but we cannot assess at this moment if it is the unique solution). The matrix $X := \overline{X} - \overline{Z}$ has the spectral decomposition (13) with

$$P = \begin{bmatrix} -0.5822 & -0.0000 & 0.7071 & 0.4013 \\ 0.5822 & 0.0000 & 0.7071 & -0.4013 \\ -0.4013 & 0.7071 & 0.0000 & -0.5822 \\ -0.4013 & 0.7071 & -0.0000 & 0.5822 \end{bmatrix} \quad \textit{and} \quad \lambda = \begin{bmatrix} 2.9505 \\ 1.0495 \\ -0.4283 \\ -1.3293 \end{bmatrix}.$$

Hence, $\beta = \emptyset$, $\gamma = \{3, 4\}$, implying that

$$\text{app}(\bar{y}, \overline{Z}) = \left\{ B \in \mathcal{S}^4 \;\middle|\; \text{diag}(B) = 0, \quad P_\gamma^T B P_\gamma = 0 \right\}.$$

It is only down to some elementary calculations to verify that for any $B \in$ app $(\bar{y}, \overline{Z})$, we have $B_{12} = 0$. In other words, if $0 \neq B \in$ app $(\bar{y}, \overline{Z})$, then there must exist an off-diagonal element $B_{ij} \neq 0$, $(i,j) \notin \{(1,2),(2,1)\}$. Consequently, for such B we must have

$$\langle B, H \circ H \circ B \rangle \geq H_{ij}^2 B_{ij}^2 > 0.$$

That is, the strong second-order sufficient condition (21) holds even some off-diagonal weights in H are zeros. *Because of the fulfilment of the SSOSC, we can now claim that $\overline{X}$ is indeed the unique optimal solution. We also note that the strict complementarity condition holds for this example.*

However, if H contains more zero off-diagonal weights, the SSOSC (21) may not hold any more. For example, if H becomes

$$H = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

and G remains unchanged, an optimal solution found by Algorithm 3.4 has $\bar{y} = 0$ and $\overline{Z} = 0$ as its Lagrangian multipliers. This implies $\gamma = \emptyset$ and hence

$$\mathrm{app}(\bar{y}, \overline{Z}) = \left\{ B \in \mathcal{S}^4 \,\middle|\, \mathrm{diag}(B) = 0 \right\}.$$

There exists $0 \neq B \in$ app $(\bar{y}, \overline{Z})$ such that $\langle B, H \circ H \circ B \rangle = 0$. We also note that the term $\Upsilon_{\overline{X}}(\overline{Z}, B)$ always equals 0 because $\overline{Z} = 0$. Therefore, the SSOSC (21) fails to hold.

One may be curious why we went all the way to use Algorithm 3.4 to give the seemingly nontrivial Example 2.3 in $\mathcal{S}^4$. Is it possible to have an example in $\mathcal{S}^3$? The answer is surprisingly no as long as G is a pseudo-correlation matrix. We give a brief proof to this result next.

Suppose that $H \in \mathcal{S}^3$ has only one zero off-diagonal weight, namely $H_{12} = 0$ and $H_{ij} > 0$ for all $(i,j) \notin \{(1,2),(2,1)\}$. Let

$$G = \begin{bmatrix} 1 & \tau_1 & \tau_2 \\ \tau_1 & 1 & \tau_3 \\ \tau_2 & \tau_3 & 1 \end{bmatrix}, \quad -1 \leq \tau_i \leq 1, \; i = 1,2,3.$$

The following fact can be easily verified

Fact 2.4 *For arbitrary chosen τ_2, $\tau_3 \in [-1, 1]$, the following matrix*

$$\overline{X} = \begin{bmatrix} 1 & \tau_2\tau_3 & \tau_2 \\ \tau_2\tau_3 & 1 & \tau_3 \\ \tau_2 & \tau_3 & 1 \end{bmatrix}$$

is the nearest correlation matrix to G under the H-weight (there may be more than one nearest correlation matrix). If $\tau_2 = \pm 1$ and $\tau_3 = \pm 1$, then $\overline{X}$ is the unique nearest correlation matrix.

Because of this fact and $H_{12} = 0$, the corresponding Lagrangian multipliers for $\overline{X}$ are $\bar{y} = 0 \in \mathbb{R}^3$ and $\overline{Z} = 0 \in \mathcal{S}^3$. This implies

$$\text{app}\,(\bar{y}, \overline{Z}) = \left\{ B \in \mathcal{S}^3 \;\middle|\; \text{diag}(B) = 0 \right\}.$$

Let $B \in \mathcal{S}^3$ be such that $B_{12} \neq 0$ and $B_{ij} = 0$ for $(i,j) \notin \{(1,2),(2,1)\}$. It follows that $B \in \text{app}\,(\bar{y}, \overline{Z})$ and $\langle B, H \circ H \circ B \rangle = 0$. Thus, the SSOSC (21) fails to hold because $\Upsilon_{\overline{X}}(\overline{Z}, B)$ is always *zero*. The argument certainly extends to H containing more *zero* off-diagonal weights. Hence, the SSOSC is never satisfied in $\mathcal{S}^3$ when H contains zero off-diagonal weights. The prerequisite of G being a pseudo-correlation matrix is crucial in the above argument. When G is not restricted to be a pseudo-correlation matrix, it is indeed possible to construct an example in $\mathcal{S}^3$ showing that the SSOSC may still hold even H contains some *zero* off-diagonal weights (see [21, Example 3.9]).

Let the mapping F be defined by (19). The following results states the local invertibility of F near the KKT point $(\overline{X}, \bar{y}, \overline{Z})$, which is important for the convergence analysis of the augmented Lagrangian method for solving the H-weighted problem (3).

Proposition 2.5 *There exist a neighborhood $\mathcal{N}$ of $(\overline{X}, \bar{y}, \overline{Z})$ in $\mathcal{S}^n \times \mathbb{R}^n \times \mathcal{S}^n$ and a constant $\zeta > 0$ such that*

$$\|F(X,y,Z) - F(\widetilde{X}, \widetilde{y}, \widetilde{Z})\| \geq \zeta^{-1} \|(X,y,Z) - (\widetilde{X}, \widetilde{y}, \widetilde{Z})\| \quad \forall\ (X,y,Z) \ \text{and}\ (\widetilde{X}, \widetilde{y}, \widetilde{Z}) \in \mathcal{N}.$$

Proof. This follows directly from Proposition 2.1, Proposition 2.2, and [35, Thm. 4.1]. $\square$

3 The Augmented Lagrangian Method

As we discussed in the introduction, the Lagrangian dual approach is not applicable to the H-weighted problem (3) because the metric projection onto $\mathcal{S}^n_+$ under the H-weight does not have an explicitly computable formula. The consequence is that its corresponding Lagrangian dual problem does not reduce to an explicitly defined unconstrained optimization problem. Compared with the original problem (3), not much benefit would be gained through considering the Lagrangian dual problem.

In this section, we will demonstrate that the augmented Lagrangian dual approach works well in theory for the H-weighted case. The two mathematical properties in the preceding section justify the use of the method. We arrange this section as follows.

3.1 Outline of the Augmented Lagrangian Method

Let the augmented Lagrangian function be defined by (8) with $c > 0$. The augmented Lagrangian method for solving (3) can be stated as follows. Let $c_0 > 0$ be given. Let $(y^0, Z^0) \in \mathbb{R}^m \times \mathcal{S}^n_+$ be the initial estimated Lagrangian multiplier. At the kth iteration, determine

$$X^{k+1} \in \arg\min L_{c_k}(X, y^k, Z^k); \tag{22}$$

compute (y^{k+1}, Z^{k+1}) by

$$\left\{ \begin{array}{ll} y^{k+1} & := y^k + c_k(b - \mathcal{A}(X^{k+1})), \\ Z^{k+1} & := (Z^k - c_k X^{k+1})_+, \end{array} \right. \tag{23}$$

and update c_{k+1} by

$$c_{k+1} := c_k \quad \text{or} \quad c_{k+1} > c_k \tag{24}$$

according to certain rules.

As for the global convergence and the rate of convergence of the augmented Lagrangian method for the H-weighted problem (3), we can directly use the convergence theory developed by Rockafellar in [29, Thm. 2] and [30, Thm. 5] for general convex programming problems combining with Proposition 2.5 to get the following result.

Proposition 3.1 *Let $(\overline{X}, \bar{y}, \overline{Z})$ be the unique KKT point of problem (3). Let (X^k, y^k, Z^k) be the sequence generated by the augmented Lagrangian method (22)-(24) with $\lim_{k\to\infty} c_k = c_\infty \leq \infty$. Then*

$$\lim_{k\to\infty} (X^{k+1}, y^{k+1}, Z^{k+1}) = (\overline{X}, \bar{y}, \overline{Z})$$

and for all k sufficiently large,

$$\|(y^{k+1}, Z^{k+1}) - (\bar{y}, \overline{Z})\| \leq \frac{\zeta}{\sqrt{\zeta^2 + c_k^2}} \|(y^k, Z^k) - (\bar{y}, \overline{Z})\| \tag{25}$$

and

$$\|X^{k+1} - \overline{X}\| \leq \frac{\zeta}{c_k} \|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|, \tag{26}$$

where $\zeta > 0$ is the constant given in Proposition 2.5.

Recall that for any given $c_k > 0$, the convex function $\nu_{c_k}(\cdot)$ defined in (9) is continuously differentiable with

$$\nabla \nu_{c_k}(y^k, Z^k) = \left(\begin{array}{c} -(b - \mathcal{A}(X^{k+1})) \\ \dfrac{1}{c_k}\big(Z^k - (Z^k - c_k X^{k+1})_+\big) \end{array} \right). \tag{27}$$

This means that the sequence $\{(y^{k+1}, Z^{k+1})\}$ generated by the augmented Lagrangian method (22)-(24) can be regarded as a gradient descent method applied to the augmented Lagrangian dual problem (9) with a step-length c_k at the kth iteration:

$$(y^{k+1}, Z^{k+1}) = (y^k, Z^k) - c_k \nabla \nu_{c_k}(y^k, Z^k), \ k = 0, 1, ...$$

Consequently, one may expect a slow convergence inherited by the gradient method. However, (25) in Proposition 3.1 says that the sequence $\{(y^{k+1}, Z^{k+1})\}$ converges to $(\bar{y}, \overline{Z})$ at a linear rate inversely proportional to c_k for all c_k sufficiently large. This fast convergence has a recent new interpretation in the context of NSDP: locally the augmented Lagrangian method can be treated as an approximate semismooth Newton method [37] as long as c_k is sufficiently large. It is this interpretation at the first place that attracted us to attempt to apply the augmented Lagrangian method to the H-weighted problem (3).

3.2 A Semismooth Newton-CG Method

Section 3.1 provides a convergence analysis on the augmented Lagrangian method. But one critical issue has not been addressed yet: How to solve the subproblem (22)? This issue is fundamentally important because the method is not going to be useful any way if solving each subproblem is difficult. We propose to use a semismooth Newton-CG method to solve (22) and explain in this subsection why it works.

Fix $c > 0$ and $(y, Z) \in \mathbb{R}^n \times \mathcal{S}^n$. Define

$$\theta(X) := L_c(X, y, Z), \quad X \in \mathcal{S}^n .$$

Our aim is to develop Newton's method for the problem

$$\min_{X \in \mathcal{S}^n} \theta(X). \tag{28}$$

Since $\theta(\cdot)$ is a strongly convex function, solving (28) is equivalent to solving the following non-smooth equation

$$0 = \nabla\theta(X) = \mathcal{Q}(X) - \mathcal{A}^*(y + c(b - \mathcal{A}(X))) - \Pi_{\mathcal{S}^n_+}(Z - cX) - C\,. \tag{29}$$

It is proven in [36] that the projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ is strongly semismooth. See [9] for some extensions. Since all other terms in $\nabla\theta(\cdot)$ are linear, (29) is a semismooth equation, for which generalized Newton's method has been well developed (see [17, 24]). Let $\partial\Pi_{\mathcal{S}^n_+}(Z - cX)$ denote the generalized Jacobian of $\Pi_{\mathcal{S}^n_+}(\cdot)$ at $(Z - cX)$. Then the generalized Jacobian of $\nabla\theta(\cdot)$ at X, denoted by $\partial^2\theta(X)$, is given by

$$\partial^2\theta(X) = \mathcal{Q} + c\Big(\mathcal{A}^*\mathcal{A} + \partial\Pi_{\mathcal{S}^n_+}(Z - cX)\Big).$$

The Newton method for the semismooth equation (29) is then defined by

$$X^{k+1} = X^k - V_k^{-1}(\nabla\theta(X^k)), \qquad V_k \in \partial^2\theta(X),\ k = 0, 1, \ldots. \tag{30}$$

The implementation of the Newton method (30) requires the availability of $V \in \partial^2\theta(X)$ and the nonsingularity of V, both of which can be easily realized. Any $V \in \partial^2\theta(X)$ has the formula

$$V = \mathcal{Q} + c\mathcal{A}^*\mathcal{A} + cW, \qquad W \in \partial\Pi_{\mathcal{S}^n_+}(Z - cX).$$

The operator $\mathcal{A}^*\mathcal{A}$ is obviously positive semidefinite and so is any W in $\partial\Pi_{\mathcal{S}^n_+}(Z - cX)$ [19, Prop. 1]. The positive definiteness of V comes from that of $\mathcal{Q}$ because $\mathcal{Q} = H \circ H \circ$ and $H_{ij} > 0$. An explicit formula for any $W \in \partial\Pi_{\mathcal{S}^n_+}(Z - cX)$ can be found in [20, Lem. 11].

Now we are ready to describe the algorithm for solving problem (28).

Algorithm 3.2 (A Semismooth Newton-CG Method)

Step 0. *Given* $X^0 \in \mathcal{S}^n$, $\eta \in (0, 1)$, $\mu \in (0, 1)$, $\tau_1 \in (0, 1)$, $\tau_2 \in (1, \infty)$, $\tau_3 \in (1, \infty)$, *and* $\rho \in (0, 1)$. *Let* $j := 0$.

Step 1. *Select an element* $V_j \in \partial^2\theta(X^j)$, *compute* $s_j := \min\{\tau_1, \tau_2\|\nabla\theta(X^j)\|\}$, *and apply the CG method [13] starting with the zero vector as the initial search direction to*

$$\nabla\theta(X^j) + (V_j + s_j I)\Delta X = 0 \tag{31}$$

to find a search direction ΔX^j *such that*

$$\|\nabla\theta(X^j) + (V_j + s_j I)\Delta X^j\| \leq \eta_j \|\nabla\theta(X^k)\|\,, \tag{32}$$

where $\eta_j := \min\{\eta, \tau_3\|\nabla\theta(X^j)\|\}$.

Step 2. *Let* l_j *be the smallest nonnegative integer* l *such that*

$$\theta(X^j + \rho^l(\Delta X^j)) - \theta(X^j) \leq \mu\,\rho^l \left\langle \nabla\theta(X^j), \Delta X^j \right\rangle\,.$$

Set $t_j := \rho^{l_j}$ *and* $X^{j+1} := X^j + t_j(\Delta X^j)$.

Step 3. *Replace* j *by* $j+1$ *and go to Step 1.*

Note that since for each $j \geq 0$, $V_j + s_j I$ is positive definite, one can always use the CG method to find ΔX^j such that (32) is satisfied. Furthermore, since the CG method is applied with the zero vector as the initial search direction, it is not difficult to see that ΔX^j is always a descent direction for $\theta(\cdot)$ at X^j. In fact, it holds that

$$\frac{1}{\lambda_{\max}(V_j + s_j I)}\|\nabla\theta(X^j)\|^2 \leq \left\langle -\nabla\theta(X^j), \Delta X^j \right\rangle \leq \frac{1}{\lambda_{\min}(V_j + s_j I)}\|\nabla\theta(X^j)\|^2\,, \tag{33}$$

where for any matrix $A \in \mathcal{S}^n$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ represent the smallest and largest eigenvalue of A, respectively. For a proof on (33), see [42]. Therefore, Algorithm 3.2 is well defined as long as $\nabla\theta(X^j) \neq 0$ and its convergence analysis can be conducted in a similar way as in [22, Thm. 5.3]. We state these result in the next theorem, whose proof is omitted for brevity.

Theorem 3.3 *Suppose that in Algorithm 3.2,* $\nabla\theta(X^j) \neq 0$ *for all* $j \geq 0$. *Then Algorithm 3.2 is well defined and the generated iteration sequence* $\{X^j\}$ *converges to the unique solution* X^* *of problem (28) quadratically.*

In our numerical experiments, the parameters used in Algorithm 3.2 are set as follows: $\eta = 10^{-2}$, $\mu = 10^{-12}$, $\tau_1 = 10^{-2}$, $\tau_2 = 10$, $\tau_3 = 10^4$, and $\rho = 0.5$.

3.3 A Practical Augmented Lagrangian Method

Section 3.2 addresses the fundamental issue of solving problem (22). In order to use the augmented Lagrangian method (23) for solving the H-weighted problem (3), we need to know when to terminate Algorithm 3.2 without affecting the convergence results presented in Proposition 3.1 so as to make the method practical. Fortunately, Rockafellar [29, 30] has already provided a solution on this.

For each $k \geq 0$, define

$$\theta_k(X) := L_{c_k}(X, y^k, Z^k), \quad X \in \mathcal{S}^n\,.$$

Since θ_k is strongly convex, we can use the following stopping criterion considered by Rockafellar for general convex optimization problems [29, 30] but tailored to our need:

$$\left\{\begin{array}{l} \frac{1}{h_{\min}^2}\|\nabla\theta_k(X^{k+1})\|^2 \le \frac{\varepsilon_k^2}{2c_k}, \qquad \varepsilon_k > 0,\ \sum_{k=0}^{\infty}\varepsilon_k < \infty\,, \\ \frac{1}{h_{\min}^2}\|\nabla\theta_k(X^{k+1})\|^2 \le \frac{\delta_k^2}{2c_k}\|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|^2, \qquad \delta_k > 0,\ \sum_{k=0}^{\infty}\delta_k < \infty\,, \\ \|\nabla\theta_k(X^{k+1})\| \le (\delta_k'/c_k)\|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|, \quad 0 < \delta_k' \to 0\,, \end{array}\right. \tag{34}$$

where $h_{\min} := \min\{H_{ij} \,|\, i, j = 1, \ldots, n\}$ and (y^{k+1}, Z^{k+1}) is defined by (23).

Finally, a ready-to-implement version of the augmented Lagrangian method (22)-(24) can be described as follows.

Algorithm 3.4 (A Practical Augmented Lagrangian Method)

Step 0. *Given $c_0 > 0$ and $\kappa > 1$. Let $X^0 \in \mathcal{S}^n$ be arbitrary. Let $y^0 \in \mathbb{R}^n$ and $Z^0 \in \mathcal{S}_+^n$ be the initial estimated Lagrangian multipliers. Let $k := 0$.*

Step 1. *Apply* Algorithm 3.2 *to problem*

$$\min_{X\in\mathcal{S}^n} \ \theta_k(X)$$

with $\theta(\cdot) = \theta_k(\cdot)$ and the starting point X^k to obtain X^{k+1} satisfying the stopping criterion (34).

Step 2. *Compute (y^{k+1}, Z^{k+1}) by (23) and update $c_{k+1} = c_k$ or $c_{k+1} = \kappa c_k$.*

Step 3. *Replace k by $k+1$ and go to Step 1.*

As for the case of the exact augmented Lagrangian method for the H-weighted problem (3), we can also directly use [29, Thm. 2] and [30, Thm. 5] for general convex programming problems combining with Proposition 2.5 to get the following convergence theorem for Algorithm 3.4.

Theorem 3.5 *Let $(\overline{X}, \bar{y}, \overline{Z})$ be the unique KKT point of problem (3). Let $\zeta > 0$ be the constant given in Proposition 2.5. Let (X^k, y^k, Z^k) be the sequence generated by Algorithm 3.4 with* $\lim_{k\to\infty} c_k = c_\infty \le \infty$. *Then*

$$\lim_{k\to\infty}(X^{k+1}, y^{k+1}, Z^{k+1}) = (\overline{X}, \bar{y}, \overline{Z})$$

and for all k sufficiently large,

$$\|(y^{k+1}, Z^{k+1}) - (\bar{y}, \overline{Z})\| \le a_k\|(y^k, Z^k) - (\bar{y}, \overline{Z})\|$$
$$\|X^{k+1} - \overline{X}\| \le a_k'\|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|\,,$$

where

$$a_k := \big[\zeta(\zeta^2 + c_k^2)^{-1/2} + \delta_k\big](1 - \delta_k)^{-1} \to a_\infty = \zeta(\zeta^2 + c_\infty^2)^{-1/2}$$

and

$$a_k' := \zeta(1 + \delta_k')/c_k \to a_\infty' = \zeta/c_\infty.$$

4 Numerical Results

In this section, we report our numerical experiments conducted for the H-weighted nearest correlation problem (3) in `MATLAB 7.1` running on a PC Intel Pentium IV of 2.40 GHz CPU and 512 MB of RAM.

In our numerical experiments, the initial penalty parameter c_0 is set to be 10 and the constant scalar κ is set to be 1.4. The initial point (X^0, y^0) is obtained by calling the quadratically convergent Newton method presented in ([22]) for solving the equally weighted nearest correlation matrix problem

$$\begin{array}{ll} \min & \frac{1}{2}\| X - G \|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X \in \mathcal{S}^n_+ \end{array}$$

and Z^0 is set to be

$$Z^0 := X^0 - G - \operatorname{diag}(y^0).$$

The stopping criterion for terminating Algorithm 3.4 is

$$\mathrm{Tol}_k \leq \ 5.0 \times 10^{-6}\,,$$

where

$$\mathrm{Tol}_0 := \|F(x^0, y^0, Z^0)\|$$

and for each $k \geq 0$,

$$\mathrm{Tol}_{k+1} := \max\{\|\nabla\theta_k(X^{k+1})\|,\ \|b - \mathcal{A}(X^{k+1})\|,\ \|Z^k - \Pi_{\mathcal{S}^n_+}(Z^k - c_k X^k)\|/\sqrt{c_k}\,\}\,.$$

In Step 1 of Algorithm 3.4, X^{k+1} is computed to satisfy

$$\|\nabla\theta_k(X^{k+1})\| \leq \min\{0.01, 0.5 \times \mathrm{Tol}_k\},$$

which is based on (34). In Step 2, c_{k+1} is updated to $c_{k+1} = \kappa c_k$ if $\mathrm{Tol}_{k+1} > \frac{1}{4}\mathrm{Tol}_k$ and $c_{k+1} = c_k$ otherwise.

To simulate the possible realistic situations, the H-weight matrix H is generated with all entries uniformly distributed in $[0.1, 10]$ except for 2×100 entries in $[0.01, 100]$. The `MATLAB` code for generating such a matrix H is as follows:

```
W0 = sprand(n,n,0.5); W0 = triu(W0) + triu(W0,1)'; W0 = (W0+W0')/2;
W0 = 0.01*ones(n,n) + 99.99*W0;
W1 = rand(n,n); W1 = triu(W1) + triu(W1,1)'; W1 = (W1+W1')/2;
H = 0.1*ones(n,n)+9.9*W1;
s = sprand(n,1,min(10/n,1)); I = find(s>0);
d = sprand(n,1,min(10/n,1)); J = find(d>0);
if length(I) >0 & length(J)>0
H(I,J) = W0(I,J); H(J,I) = W0(J,I); end
H = (H+H')/2;
```

Our first example is a 387×387 correlation matrix case taken from the database of the RiskMetrics.

Example 4.1 *The correlation matrix G is the 387×387 1-day correlation matrix (as of June 15, 2006) from the lagged datasets of RiskMetrics* (`www.riskmetrics.com/stddownload_edu.html`). *For the test purpose, we perturb G to*

$$G := (1-\alpha)G + \alpha E,$$

where $\alpha \in (0,1)$ and E is randomly generated symmetric matrix with entries in [-1,1]. The `MATLAB` *code for generating E is:* `E = 2.0*rand(387,387) - ones(387,387); E = triu(E) + triu(E,1)'; E = (E+E')/2`. *We also set $G_{ii} = 1$, $i = 1, \ldots, n$.*

Our second example is randomly generated with $n = 100, 500, 1000$, and 1500, respectively.

Example 4.2 *A correlation matrix G is first generated by using* `MATLAB`*'s built-in function* `randcorr: x=10.^[-4:4/(n-1):0]; G=gallery('randcorr',n*x/sum(x))`*; and is then perturbed to*

$$G := (1-\alpha)G + \alpha E,$$

where $\alpha \in (0,1)$, E is randomly generated as in Example 4.1: `E = 2.0*rand(n,n) - ones(n,n); E = triu(E) + triu(E,1)'; E = (E+E')/2`*, and G_{ii} is set to be 1 for $i = 1, \ldots, n$.*

The small added term αE in the above examples makes a correlation matrix to be a pseudo-correlation matrix. Our numerical results are reported in Tables 1-2, where `Alg. 3.4` and `IP-QSQP` refer to Algorithm 3.4 and Toh's inexact interior point method with the PSQMR as the iterative solver [38], respectively. `Iter` and `LiSys` stand for the number of total iterations and the number of total linear systems solved. `Res` represents the relative residue computed at the last iterate:

$$\texttt{Res} := \max\{\|\nabla\theta_k(X^{k+1})\|/(1+\|C\|),\ \|b - \mathcal{A}(X^{k+1})\|/(1+\|b\|),\ |\langle X^{k+1}, Z^{k+1}\rangle|/(1+|\text{obj}|)\},$$

where

$$\text{obj} := \frac{1}{2}||H \circ (X^{k+1} - G)\|^2 .$$

In Table 1, $*$ means that the `PSQMR` reaches the maximum number of steps set at 1000 and in Table 2, `out of memory` means that our PC runs out of memory.

From Tables 1-2 and other similar testing results not reported here, we have observed that our algorithm is not only faster but also more robust than `IP-QSQP`, in particular for those cases that a good initial correlation matrix estimation is available, as in many real-world situations.

5 Conclusion

The convergence theory of the augmented Lagrangian method for the convex optimization problem has been well established by Rockafellar [29, 30]. The main purpose of this paper is to demonstrate that this method is not only fast but also robust for the H-weighted correlation matrix problem. Theoretically, one only needs to verify the conditions used in [29, 30]. It turns out that the constraint nondegeneracy property and the strong second-order sufficient condition

Algorithm	α	cputime	Iter	LiSys	Res
Alg. 3.4	0.1	0:04:52	13	36	3.1×10^{-9}
	0.05	0:04:12	12	29	2.7×10^{-8}
	0.01	0:04:58	12	27	1.6×10^{-9}
	0.005	0:04:16	11	21	1.7×10^{-9}
IP-QSDP	0.1	0:17:43	17	34	1.7×10^{-8}
	0.05	0:18:36	18	36	3.3×10^{-8}
	0.01	0:37:28	25	50	8.5×10^{-8}
	0.005	0:36:21	17	34	$2.6 \times 10^{-1*}$

Table 1: Numerical results of Example 4.1

are sufficient in order to apply Rockafellar's convergence results. We outlined how the two properties naturally lead to the linear convergence of the method (cf. Proposition 3.1 and Theorem 3.5).

The key element for the practical efficiency of the augmented Lagrangian dual approach is the semismooth Newton-CG algorithm introduced in this paper. We believe that the excellent numerical results reported in this paper are largely due to this semismooth Newton-CG algorithm.

Finally, we note that, in a straightforward way, we may extend this approach to deal with a more general version that allows certain elements being fixed or contained in some confidence intervals, i.e.,

$$\begin{array}{ll} \min & \frac{1}{2}\| H \circ (X - G) \|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \dots, n \,, \\ & X_{ij} \geq l_{ij}, \ (i,j) \in \mathcal{B}_l \,, \\ & X_{ij} \leq u_{ij}, \ (i,j) \in \mathcal{B}_u \,, \\ & X \in \mathcal{S}^n_+ \,, \end{array} \tag{35}$$

where $\mathcal{B}_l$ and $\mathcal{B}_u$ are two index subsets of $\{(i,j) \,|\, 1 \leq i < j < n\}$, $l_{ij} \in [-1,1]$ for all $(i,j) \in \mathcal{B}_l$, $u_{ij} \in [-1,1]$ for all $(i,j) \in \mathcal{B}_u$, and $l_{ij} \leq u_{ij}$ for any $(i,j) \in \mathcal{B}_l \cap \mathcal{B}_u$. We omit the details here as our theoretical analysis still holds and there are no other methods available to allow us to make a reasonable comparison.

Acknowledgements. The authors would like to thank our colleague Kim-Chuan Toh for sharing with us his excellent code for solving the H-weighted nearest correlation matrix problem [38]. Several helpful discussions on the implementation of the augmented Lagrangian method with Yan Gao and Xinyuan Zhao at the National University of Singapore are also acknowledged here.

Algorithm	n	α	cputime	Iter	LiSys	Res
Alg. 3.4	100	0.1	0:00:10	10	24	1.1×10^{-8}
		0.05	0:00:10	8	22	1.1×10^{-8}
		0.01	0:00:16	8	22	1.6×10^{-8}
		0.005	0:00:41	8	34	1.1×10^{-8}
IP-QSDP		0.1	0:01:27	14	28	6.6×10^{-8}
		0.05	0:02:08	16	32	9.6×10^{-9}
		0.01	0:03:36	19	38	1.8×10^{-8}
		0.005	0:06:05	18	36	2.6×10^{-8}
Alg. 3.4	500	0.1	0:06:22	10	26	4.7×10^{-9}
		0.05	0:05:53	9	23	8.4×10^{-9}
		0.01	0:08:06	10	24	1.1×10^{-9}
		0.005	0:08:49	9	24	5.1×10^{-9}
IP-QSDP		0.1	0:41:22	14	28	9.5×10^{-8}
		0.05	0:39:47	14	28	8.7×10^{-8}
		0.01	1:34:16	19	38	1.8×10^{-8}
		0.005	1:46:42	19	38	2.9×10^{-8}
Alg. 3.4	$1,000$	0.1	0:42:24	14	32	5.6×10^{-8}
		0.05	0:36:12	11	29	3.5×10^{-10}
		0.01	0:34:59	10	26	2.0×10^{-9}
		0.005	0:33:30	9	22	2.9×10^{-9}
IP-QSDP		0.1	3:13:58	14	28	1.2×10^{-8}
		0.05	4:36:47	15	30	3.6×10^{-8}
		0.01	8:00:46	21	42	2.3×10^{-8}
		0.005	6:39:58	21	42	4.7×10^{-8}
Alg. 3.4	$1,500$	0.1	2:01:48	12	31	8.3×10^{-10}
		0.05	1:54:57	11	27	1.2×10^{-9}
		0.01	1:46:43	9	25	2.6×10^{-9}
		0.005	2:06:06	9	26	1.1×10^{-9}
IP-QSDP		–	–	–	–	`out of memory`

Table 2: Numerical results of Example 4.2

References

[1] F. Alizadeh, J.-P. A. Haeberly, and M.L. Overton, *Complementarity and nondegenracy in semidefinite programming*, Math. Programming 77 (1997), pp. 111-128.

[2] V.I. Arnold, *On matrices depending on parameters*, Russian Mathematical Surveys 26 (1971), pp. 29–43.

[3] V. Bhansali and B. Wise, *Forcasting portfolio risk in normal and stressed market*, *Jouranl of Risk* 4(1), (2001), pp. 91–106.

[4] B. Borchers and J.G. Young, *Implementation of a primal-dual method for SDP on a shared memory parallel architecture*, Computational Optimization and Applications 37 (2007), pp. 355-369.

[5] R. Borsdorf, *A Newton algorithm for the nearest correlation matrix*, Master Thesis, School of Mathematics, University of Manchester, 2007.

[6] J.F. Bonnans and A. Shapiro, *Nondegeneracy and quantitative satbility of parameterized optimization problems with multiple solutions*, SIAM Journal on Optimization 8 (1998), pp. 940–946.

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

[8] S. Boyd and L. Xiao, *Least-squares covariance matrix adjustment*, SIAM Journal on Matrix Analysis and Applications 27 (2005), pp. 532-546.

[9] X. Chen, H.-D. Qi, and P. Tseng, *Analysis of nonsmooth symmetric matrix valued functions with applications to semidefinite complementarity problems*, SIAM Journal on Optim. 13 (2003), pp. 960–985.

[10] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[11] R.L. Dykstra, *An algorithm for restricted least squares regression*, J. Amer. Stat. Assoc. 78 (1983), pp. 837–842.

[12] B.C. Eaves, *On the basic theorem for complemenarity*, Math. Programming 1 (1971), pp. 68–75.

[13] M.R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Stand. 49 (1952), pp. 409–436.

[14] N.J. Higham, *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra Appl. 103 (1988), pp. 103–118.

[15] N.J. Higham, *Computing the nearest correlation matrix – a problem from finance*, IMA J. Numer. Analysis 22 (2002), pp. 329–343.

[16] C.R. Johnson, B. Kroschel, and H. Wolkowicz, *An interior-point method for approximate positive semidefinite completions*, Computational Optimization and Applications 9 (1998), pp. 175–190.

[17] B. Kummer, *Newton's method for nondifferentiable functions*, Advances in Mathematical Optimization, 114–125, Math. Res., 45, Akademie-Verlag, Berlin, 1988.

[18] J. Malick, *A dual approach to semidefinite least-squares problems*, SIAM Journal on Matrix Analysis and Applications 26 (2004), pp. 272–284.

[19] F. Meng, D.F. Sun, and G.Y. Zhao, *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Math. Programming 104 (2005), pp. 561–581.

[20] J.S. Pang, D.F. Sun, and J. Sun, *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Mathematics of Operations Research 28 (2003), pp. 39–63.

[21] H.-D. Qi, *Local duality of nonlinear semidefinite programming*, to appear in Mathematics of Operations Research.

[22] H.-D. Qi and D.F. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM Journal on Matrix Analysis and Applications 28 (2006), pp. 360–385.

[23] H.-D. Qi and D.F. Sun, *Correlation stress testing for value-at-risk: an unconstrained convex optimization approach*, Technical Report, Department of Mathematics, National University of Singapore, 2007.

[24] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming 58 (1993), pp. 353–367.

[25] S.M. Robinson, *Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy*, Math. Programming Study 22 (1984), pp. 217–230.

[26] S.M. Robinson, *Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity*, Math. Programming Study 30 (1987), pp. 45–66.

[27] S.M. Robinson, *Constriant nondegeneracy in variational analysis*, Mathematics of Operations Research 28 (2003), pp. 201–232.

[28] R.T. Rockafellar, *Conjugate Duality and Optimization*, SIAM, Philadelphia, 1974.

[29] R.T. Rockafellar, *Augmented Lagrangains and applications of the proximal point algorithm in convex programming*, Mathematics of Operation Research 1 (1976), pp. 97-116.

[30] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14 (1976), pp. 877-898.

[31] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[32] N.C. Schwertman and D.M. Allen, *Smoothing an indefinite variance-covariance matrix*, *Journal of Statistical Computation and Simulation* 9 (1979), pp. 183–194.

[33] A. Shapiro and M.K.H. Fan, *On eigenvalue optimization*, SIAM Journal on Optimization 5 (1995), pp. 552–569.

[34] J.F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optimization Methods and Software 11 & 12 (1999), pp. 625–653.

[35] D.F. Sun, *The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications*, Mathematics of Operations Research 31 (2006), pp. 761–776.

[36] D.F. Sun and J. Sun, *Semismooth matrix valued functions*, Mathematics of Operations Research 27 (2002), pp. 150–169.

[37] D.F. Sun, J. Sun and L.W. Zhang, *The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming*, Math. Programming (2008).

[38] K.C. Toh, *An inexact path-following algorithm for convex quadratic SDP*, Math. Programming, 112 (2008), 221–254.

[39] K.C. Toh, R.H. Tütüncü, and M.J. Todd, *Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems*, Pacific J. Optimization, 3 (2007), pp. 135–164.

[40] P. Tseng, *Merit functions for semi-definite complementarity problems*, Math. Programming 83 (1998), pp. 159–185.

[41] R.H. Tütüncü, K.C. Toh, and M.J. Todd, *Solving semidefinite-quadratic-linear programs using SDPT3*, Math. Programming 95 (2003), pp. 189–217.

[42] X.Y. Zhao, D.F. Sun, and K.C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, Manuscript, Department of Mathematics, National University of Singapore, 2008.