IMA Journal of Numerical Analysis Page 1 of 21
doi:10.1093/imanum/drp031

An augmented Lagrangian dual approach for the H-weighted nearest correlation matrix problem

HOUDUO QI†
School of Mathematics, The University of Southampton, Highfield, Southampton SO17 1BJ, UK

AND

DEFENG SUN‡
Department of Mathematics and Risk Management Institute, National University of Singapore, 2 Science Drive 2, Singapore 117543

[Received on 3 March 2008; revised on 17 July 2009]

Higham (2002, *IMA J. Numer. Anal.*, **22**, 329–343) considered two types of nearest correlation matrix problems, namely the W-weighted case and the H-weighted case. While the W-weighted case has since 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 that avoids the explicit computation of the metric projection under the H-weight. This method solves a sequence of unconstrained convex optimization problems, each of which can be efficiently solved by an inexact 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.

Keywords: augmented Lagrangian; semismooth Newton method; conjugate gradient method; nearest correlation matrix.

1. Introduction

In Higham (2002) the author considered two types of *nearest correlation matrix* problems. One is under the W-weighting:

$$\begin{array}{ll} \min & \frac{1}{2}\|W^{1/2}(X-G)W^{1/2}\|^2 \\ \text{such that} & X_{ii}=1, \quad i=1,\ldots,n, \\ & X\in\mathcal{S}^n_+, \end{array} \tag{1.1}$$

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

†Corresponding author. Email: hdqi@soton.ac.uk
‡Email: matsundf@nus.edu.sg

The constraints in (1.1), collectively known as the *correlation constraints*, specify that any feasible matrix to (1.1) is a correlation matrix. Solving the W-weighted problem (1.1) is equivalent to solving a problem of the following type (cf. Qi & Sun, 2006, Section 4.1):

$$\begin{array}{ll} \min & \frac{1}{2}\|X - G\|^2 \\ \text{such that} & \operatorname{diag}(W^{-1/2} X W^{-1/2}) = e, \\ & X \in \mathcal{S}^n_+, \end{array} \tag{1.2}$$

where $e \in \mathbb{R}^n$ is the vector of all ones. We often use $X \succeq 0$ to denote $X \in \mathcal{S}^n_+$. The metric projection of $X \in \mathcal{S}^n$ onto $\mathcal{S}^n_+$ under the Frobenius norm $\|\cdot\|$ is often denoted by X_+ and sometimes by $\Pi_{\mathcal{S}^n_+}(X)$ to highlight its dependence on $\mathcal{S}^n_+$.

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

$$\begin{array}{ll} \min & \frac{1}{2}\|H \circ (X - G)\|^2 \\ \text{such that} & X_{ii} = 1, \quad i = 1, \ldots, n, \\ & X \in \mathcal{S}^n_+, \end{array} \tag{1.3}$$

where the weighting is now in the sense of Hadamard, i.e., $(A \circ B)_{ij} = A_{ij} B_{ij}$. Here the matrix H is symmetric and each of its entries is non-negative, i.e., $H_{ij} \geqslant 0$ for all $i, j = 1, \ldots, n$ (in Higham (2002), each H_{ij} was assumed to be positive.) We refer the reader to Bhansali & Wise (2001) for concrete examples in finance to see how H was constructed, where H is known as a confidence matrix. We note that, in the special case that $H = E$, the matrix of all ones, (1.3) turns out to be (1.2) with $W = I$, the identity matrix.

The W-weighted problem (1.1) has been well studied since Higham (2002) and now there are several good methods for it, including the alternating projection method (Higham, 2002), the gradient and quasi-Newton methods (Malick, 2004; Boyd & Xiao, 2005), the inexact semismooth Newton method combined with the conjugate gradient (CG) solver (Qi & Sun, 2006) and its modified version with several (preconditioned) iterative solvers (Borsdorf, 2007; Borsdorf & Higham, 2009) and the inexact interior-point methods (IPMs) with iterative solvers (Toh *et al.*, 2007; Toh, 2008). All of these methods, except the inexact IPMs, crucially rely on the fact that the projection of a given matrix $X \in \mathcal{S}^n$ onto $\mathcal{S}^n_+$ under the W-weighting, denoted by $\Pi^W_{\mathcal{S}^n_+}(X)$, which is the optimal solution of the following problem:

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

is given by the formula (see Higham, 2002, Theorem 3.2)

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

It has long been known by statisticians that, for any $A \in \mathcal{S}^n$, its metric projection A_+ over $\mathcal{S}^n_+$, which also equals $\Pi^I_{\mathcal{S}^n_+}(A)$, admits an explicit formula (Schwertman & Allen, 1979). This means that, for any $X \in \mathcal{S}^n$, one can compute $\Pi^W_{\mathcal{S}^n_+}(X)$ explicitly.

To simplify the subsequent discussions we assume, without loss of generality, that $W \equiv I$ (for reasons for this see Qi & Sun, 2006, Section 4.1). We note that $(\cdot)_+ = \Pi_{\mathcal{S}_+^n}(\cdot)$. To see how the metric projection operator $\Pi_{\mathcal{S}_+^n}(\cdot)$ is involved 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 (1.2),

$$l(X, y) := \frac{1}{2}\|X - G\|^2 + y^{\mathrm{T}}(b - \mathrm{diag}(X)), \quad (X, y) \in \mathcal{S}_+^n \times \mathbb{R}^n, \tag{1.4}$$

where $b := e$. Since problem (1.2) automatically satisfies the generalized Slater constraint qualification, from the duality theory developed by Rockafellar (1974), we know that problem (1.2) can be equivalently solved by its Lagrangian dual problem

$$\max_{y \in \mathbb{R}^n} \left\{ \min_{X \in \mathcal{S}_+^n} l(X, y) \right\}, \tag{1.5}$$

which, via the metric projector $(\cdot)_+$, can be equivalently reformulated as the following unconstrained optimization problem (see Rockafellar (1974), Malick (2004), and Boyd & Xiao (2005) for details):

$$\min_{y \in \mathbb{R}^n} \theta(y) := \frac{1}{2}\|(G + \mathrm{diag}(y))_+\|^2 - b^T y - \frac{1}{2}\|G\|^2, \tag{1.6}$$

in the sense that, if $\bar{y} \in \mathbb{R}^n$ is an optimal solution to (1.6), then $\overline{X} := (G + \mathrm{diag}(\bar{y}))_+$ solves (1.2). Here $\mathrm{diag}(y)$ is the diagonal matrix whose diagonal entries are the components of y.

The objective function $\theta(\cdot)$ in (1.6) is known to be once continuously differentiable and convex (Rockafellar, 1974), despite the fact that the projection operator $(\cdot)_+$ is not differentiable everywhere. Therefore the gradient method and quasi-Newton methods can be developed to solve (1.6) directly. Malick (2004) remarked that the alternating projection method is actually the gradient method for (1.6) with a constant step length of 1. These methods converge, at best, linearly. Because $\theta(\cdot)$ is convex and coercive (Rockafellar, 1974), solving (1.6) is equivalent to finding a point $\bar{y} \in \mathbb{R}^n$ satisfying its optimality condition

$$\nabla\theta(y) = \mathrm{diag}(G + \mathrm{diag}(y))_+ - b = 0.$$

We define

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

The function $F(\cdot)$ is Lipschitz continuous and thus the generalized Jacobian $\partial F(\cdot)$ in the sense of Clarke (1983) 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:

$$y^{k+1} = y^k - V_k^{-1}(\nabla\theta(y^k)), \quad V_k \in \partial^2\theta(y^k), \quad k = 0, 1, \ldots. \tag{1.7}$$

A formula for calculating $V \in \partial^2\theta(y)$ can be found in Qi & Sun (2006, p. 378). Clarke's Jacobian-based generalized Newton method (1.7) was thoroughly analysed by Qi & Sun (2006) and was proven to be quadratically convergent. Numerical experiments conducted in Qi & Sun (2006), Borsdorf (2007) and Borsdorf & Higham (2009) seem to confirm that the generalized Newton method (1.7), combined with iterative solvers, is the most effective one available so far.

For the H-weighted problem (1.3) all of those Lagrangian dual-based methods mentioned above are not applicable, 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-weighting, that is, the optimal solution, denoted $\Pi^{H\circ}_{\mathcal{S}^n_+}(X)$, to the following problem:

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

is not known to have an explicit formula.[1] For this reason the Lagrangian dual problem, which takes the following form for the H-weighted case:

$$\max_{y\in\mathbb{R}^n} \left\{ \min_{X\in\mathcal{S}^n_+} l(X, y) := \frac{1}{2}\|H \circ (X - G)\|^2 + y^{\mathrm{T}}(b - \mathrm{diag}(X)) \right\}, \tag{1.8}$$

cannot be reduced to an unconstrained smooth optimization problem in the dual space similar to (1.6) for the W-weighted case. Consequently, compared to the original problem, the Lagrangian dual problem (1.8) does not provide us with a better choice in terms of algorithmic design. This implies, 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, developed by Rockafellar (1976a,b) in his pioneering work on convex optimization problems. Let $c > 0$ be a parameter. By using the fact that $\mathcal{S}^n_+$ is a self-dual convex cone, we know from Wierzbicki & Kurcyusz (1977) that the augmented Lagrangian function for the H-weighted problem (1.3) takes the following form:

$$\begin{aligned} L_c(X, y, Z) := &\frac{1}{2}\|H \circ (X - G)\|^2 + y^{\mathrm{T}}(b - \mathrm{diag}(X)) + \frac{c}{2}\|b - \mathrm{diag}(X)\|^2 \\ &+ \frac{1}{2c}(\|(Z - cX)_+\|^2 - \|Z\|^2), \end{aligned} \tag{1.9}$$

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

$$\min_{y\in\mathbb{R}^n, Z\in\mathcal{S}^n} \left\{ \nu_c(y, Z) := -\min_{X\in\mathcal{S}^n} L_c(X, y, Z) \right\}. \tag{1.10}$$

The major computational task in the augmented Lagrangian dual approach, as outlined in (3.1)–(3.3), at each step for a given $(y, Z) \in \mathbb{R}^n \times \mathcal{S}^n$, is to solve the following unconstrained optimization problem:

$$\min_{X\in\mathcal{S}^n} L_c(X, y, Z). \tag{1.11}$$

[1] It was stated in Johnson *et al.* (1998, Corollary 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

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

Note that, for any $(y, Z) \in \mathbb{R}^n \times \mathcal{S}^n$, we have that $L(\cdot, y, Z)$ is a convex and continuously differentiable function. Therefore the gradient method and quasi-Newton methods can be developed in theory for solving (1.11). However, our numerical experiments show that the gradient method is extremely slow and hence disregarded. The size of the variable X in (1.11) is $\bar{n} := n(n+1)/2$. Maintaining an $\bar{n} \times \bar{n}$ 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 in which 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 solving (1.11). The Newton method that we are going to use is quite similar to (1.7) with the difference that the number of the unknowns in the Newton equation here is $\bar{n}$, which is $\mathcal{O}(n^2)$, instead of n. These equations, even when n is relatively large, say $n = 1{,}000$, do not create too much difficulty when the CG method is employed to solve them. The major reason behind this is that the H-weighted problem (1.3) may satisfy two important mathematical properties, namely, *constraint nondegeneracy* and *the strong second-order sufficient* condition (SSOSC) (see Section 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 (1976a,b) for general convex optimization problems. We will confirm all of those results in the main body of the paper.

Before commenting on other approaches, we would like to emphasize why a semismooth Newton method is possible for the augmented Lagrangian dual problem (1.10), while it is not for the Lagrangian dual problem (1.8). The major reason is that the inner optimization problem in (1.10) is unconstrained and convex. Solving this inner optimization problem is equivalent to solving the semismooth equation (3.8). Moreover, the second-order information (i.e., the generalized Hessian $\partial^2\theta(X)$) can be completely calculated by (3.9). Therefore Newton's method can be developed. However, such calculations are not available for the inner optimization problem of the Lagrangian dual problem (1.8).

We note that a similar approach was also conducted in Qi & Sun (2009), where the problem considered was of the W-weighted type with a background in correlation stress testing, which required a large number of correlations to be fixed beforehand. Theoretically, being an augmented Lagrangian dual-based method, the approach in Qi & Sun (2009) can be extended to the H-weighted case considered here. Indeed, it was Qi & Sun (2009), together with Qi & Sun (2006), that inspired us to further investigate the effectiveness of the augmented Lagrangian dual approach for the H-weighted problem (1.3).

The type of IPMs was deliberately left out of the above discussions because it deserves its own particular comments. As early as Johnson *et al.* (1998), the authors started to use IPMs to solve H-weighted matrix optimization problems of various types. The H-weighted nearest correlation matrix problem (1.3) can be reformulated as a linear optimization problem with mixed semidefinite and second-order cone constraints (Higham, 2002; Toh *et al.*, 2007). Consequently, publicly available IPMs-based software such as SeDuMi (Sturm, 1999) or SDPT3 (Tütüncü *et al.*, 2003) can be applied to solve these problems directly. However, since at each iteration these solvers require us to formulate and solve a dense Schur complement matrix (cf. Borchers & Young, 2007), which for problem (1.3) 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. The serious and competitive implementation of inexact IPMs was carried out by Toh *et al.* (2007), for solving a special class of convex quadratic semidefinite programming (QSDP) including the W-weighted problem (1.1), and Toh (2008), for a general convex QSDP with the H-weighted problem

(1.1) being targeted and tested in particular. The search direction used in Toh (2008) 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 (1.3) 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 (1.3), mainly on constraint nondegeneracy and the SSOSC. Section 3 is on the augmented Lagrangian method. We first outline an abstract form of the method in Section 3.1. In Section 3.2 we present two practical algorithms. One is the semismooth Newton-CG method for solving subproblems of the type (1.11) 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 (1.3) mentioned in Section 1. The two properties will justify the use of the augmented Lagrangian method, which is to be introduced in the next section.

2.1 *The constraint nondegeneracy property*

Let us cast the problem (1.3) into the following convex QSDP:

$$\begin{array}{ll} \min & \frac{1}{2}\langle X, \mathcal{Q}(X)\rangle - \langle C, X\rangle + \frac{1}{2}\|H \circ G\|^2 \\ \text{such that} & \operatorname{diag}(X) = b, \\ & X \in \mathcal{S}^n_+, \end{array} \tag{2.1}$$

where $\mathcal{Q} = H \circ H\circ$, $C = H \circ H \circ G$ 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 $\operatorname{lin}(T_{\mathcal{S}^n_+}(X))$ be the largest linear space contained in $T_{\mathcal{S}^n_+}(X)$. We say that *constraint nondegeneracy* holds at a point X satisfying the constraints in (2.1) if

$$\operatorname{diag}(\operatorname{lin} T_{\mathcal{S}^n_+}(X)) = \mathbb{R}^n. \tag{2.2}$$

For the origin of constraint nondegeneracy, its various forms and its role in general optimization, see Bonnans & Shapiro (1998, 2000), Robinson (1984, 1987, 2003) and Shapiro & Fan (1995).

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

$$X = P \operatorname{diag}(\lambda_1, \ldots, \lambda_n) P^{\mathrm{T}}, \tag{2.3}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ are the eigenvalues of X and P is a corresponding orthogonal matrix of orthonormal eigenvectors. Then, from Schwertman & Allen (1979), Higham (1988) and Tseng (1998), we know that

$$X_+ = P \operatorname{diag}(\max(0, \lambda_1), \ldots, \max(0, \lambda_n)) P^{\mathrm{T}}. \tag{2.4}$$

We define

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

We 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 (1971) as follows:

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

Consequently, we have

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

Equivalently, we have

$$\text{lin}\,(T_{\mathcal{S}^n_+}(X_+)) = \left\{ PBP^{\mathrm{T}} | B = \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\beta} & B_{\alpha\gamma} \\ B_{\alpha\beta}^{\mathrm{T}} & 0 & 0 \\ B_{\alpha\gamma}^{\mathrm{T}} & 0 & 0 \end{bmatrix} \begin{array}{l} B_{\alpha\alpha} \in \mathcal{S}^{|\alpha|}, \\ B_{\alpha\beta} \in \mathbb{R}^{|\alpha|\times|\beta|}, \\ B_{\alpha\gamma} \in \mathbb{R}^{|\alpha|\times|\gamma|} \end{array} \right\}. \tag{2.5}$$

The following result says that any point satisfying the correlation constraints is constraint nondegenerate. It can be proved similarly to Proposition 4.2 in Toh *et al.* (2007), where the proof used a characterization of constraint nondegeneracy in Alizadeh *et al.* (1997) and Qi & Sun (2006, Lemma 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 $\{\text{diag}(X) = e,\ X \in \mathcal{S}^n_+\}$ is constraint nondegenerate.

Proof. Let $X \in \mathcal{S}^n$ satisfy the correlation constraints. Suppose that X has the spectral decomposition (2.3). Because X is positive semidefinite, $\gamma = \emptyset$. Also, because $\text{diag}(X) = e$, we have that $\alpha \neq \emptyset$. Moreover, this diagonal constraint also implies (see Qi & Sun, 2006, Lemma 3.3)

$$\sum_{\ell \in \alpha} P_{i\ell}^2 > 0, \quad i = 1, \ldots, n. \tag{2.6}$$

To show that condition (2.2) holds at X, it suffices to prove that

$$(\text{diag}(\text{lin}\ T_{\mathcal{S}^n_+}(X)))^{\perp} = \{0\}.$$

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

$$0 = \langle v, \text{diag}(PBP^{\mathrm{T}})\rangle = \langle \text{diag}(v), PBP^{\mathrm{T}}\rangle = \langle P^{\mathrm{T}}\text{diag}(v)P, B\rangle, \tag{2.7}$$

where B is from (2.5). The structure of B implies that

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

which in turn implies that

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

Summing 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 (2.6) ensures that $v_i = 0$ for each $i = 1, \ldots, n$. This completes our proof. □

2.2 *The strong second-order sufficient condition*

Now let us consider the Karush–Kuhn–Tucker (KKT) system of the QSDP (2.1), namely,

$$\begin{aligned} \mathcal{Q}(X) - \operatorname{diag}(y) - Z &= C, \\ \operatorname{diag}(X) &= b, \\ X \succeq 0,\ Z \succeq 0,\ \langle X, Z\rangle &= 0. \end{aligned} \tag{2.8}$$

Any triple $(\overline{X}, \bar{y}, \overline{Z}) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ satisfying (2.8) is called a KKT point of (2.1). By using the fact that $\mathcal{S}^n_+$ is a self-dual cone, we know from Eaves (1971) that $(\overline{X}, \bar{y}, \overline{Z}) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ satisfies the KKT conditions (2.8) if and only if it satisfies the following system of nonsmooth equations:

$$F(X, y, Z) := \begin{bmatrix} \mathcal{Q}(X) - C - \operatorname{diag}(y) - Z \\ b - \operatorname{diag}(X) \\ Z - [Z - X]_+ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad (X, y, Z) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n. \tag{2.9}$$

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

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 (2.1). We define $X := \overline{X} - \overline{Z}$. Suppose that X has the spectral decomposition (2.3). We define

$$\text{app}\ (\bar{y}, \overline{Z}) := \{B \in \mathcal{S}^n | \operatorname{diag}(B) = 0,\ P_\beta^{\mathrm{T}} B P_\gamma = 0,\ P_\gamma^{\mathrm{T}} B P_\gamma = 0\}. \tag{2.10}$$

Note that app $(\bar{y}, \overline{Z})$ is independent of the choice of P in (2.3) (see Sun, 2006, equations (38) and (39)). We also define

$$\mathcal{M}(\overline{X}) := \{(\bar{y}, \overline{Z}) | (\overline{X}, \bar{y}, \overline{Z}) \text{ is a KKT point of (2.1)}\}.$$

The set $\mathcal{M}(\overline{X})$ is known to be the set of Lagrange multipliers at $\overline{X}$. For the H-weighted problem (1.3) $\mathcal{M}(\overline{X})$ contains a unique point $(\bar{y}, \overline{Z})$ because constraint nondegeneracy holds at $\overline{X}$ by Proposition 2.1. For a proof of this see Bonnans & Shapiro (2000, Theorem 5.85). We say that the SSOSC holds at $\overline{X}$ if

$$\langle B, H \circ H \circ B\rangle + \Upsilon_{\overline{X}}(\overline{Z}, B) > 0 \quad \forall\, 0 \neq B \in \text{app}\ (\bar{y}, \overline{Z}), \tag{2.11}$$

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 pseudoinverse of $\overline{X}$. Note that $\Upsilon_{\overline{X}}(\overline{Z}, B)$ is quadratic in B and is always non-negative because $\overline{Z} \succeq 0$ and $\overline{X} \succeq 0$. Note also that, in the left-hand side of (2.11), the first term $\langle B, H \circ H \circ B\rangle > 0$ if there exist $i, j \in \{1, \ldots, n\}$ such that $H_{ij} B_{ij} \neq 0$. Therefore we have the following statement.

PROPOSITION 2.2 Let $(\overline{X}, \bar{y}, \overline{Z})$ be the unique KKT point of the H-weighted nearest correlation matrix problem (1.3). If for any $0 \neq B \in$ app $(\bar{y}, \overline{Z})$ there exist $i, j \in \{1, \ldots, n\}$ such that $B_{ij} \neq 0$ and $H_{ij} > 0$, then the SSOSC (2.11) holds at $\overline{X}$.

We now make several remarks about the SSOSC (2.11).

(i) The SSOSC was first proposed by Sun (2006) in a 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, (2.11) is just a specialization of the original one given in Sun (2006, Definition 3.2).

(ii) In some practical cases (see Bhansali & Wise, 2001) the diagonal weights H_{ii} are assigned *zero* values (i.e., $H_{ii} = 0$, where $i = 1, \ldots, n$). This does not have any effect on (2.11) because, for any $B \in \text{app}(\bar{y}, \overline{Z})$, we must have $B_{ii} = 0$, where $i = 1, \ldots, n$ (see the definition (2.10)). Therefore the diagonal weights in H have no contribution to the value $\langle B, H \circ H \circ B\rangle$.

(iii) Furthermore, for the SSOSC (2.11) to hold at $\overline{X}$ one does not have to assume that all of the off-diagonal weights are positive. In fact, as the following example shows, some of them are allowed to be *zero* without damaging the SSOSC. This example also shows that too many zero off-diagonal weights do destroy the SSOSC (2.11).

EXAMPLE 2.3 Consider the H-weighted problem (1.3) in $\mathcal{S}^4$ with data given by

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

Such a matrix G is known as a pseudocorrelation matrix because $-1 \leqslant G_{ij} \leqslant 1$, $G_{ii} = 1$ for all $i, j = 1, \ldots, 4$, and $\lambda_{\min}(G) = -0.8860 < 0$. After running our augmented Lagrangian dual method, namely, Algorithm 3.3, 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

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

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 (2.3) with

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

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

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

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

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

that is, the SSOSC (2.11) holds even if some off-diagonal weights in H are *zero*. 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 then the SSOSC (2.11) may no longer hold. 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, then an optimal solution found by Algorithm 3.3 has $\bar{y} = 0$ and $\overline{Z} = 0$ as a pair of the Lagrange multipliers. This implies that $\gamma = \emptyset$ and hence

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

There exists $0 \neq B \in \text{app}(\bar{y}, \overline{Z})$ 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 (2.11) fails to hold.

One may wonder why we used Algorithm 3.3 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 pseudocorrelation matrix. We give a brief proof of this result below.

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

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

The following fact can easily be verified.

FACT 2.4 For arbitrary chosen $\tau_2, \tau_3 \in [-1, 1]$ the 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 a 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 Lagrange multipliers for $\overline{X}$ are $\bar{y} = 0 \in \mathbb{R}^3$ and $\overline{Z} = 0 \in \mathcal{S}^3$. This implies that

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

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 (2.11) 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 pseudocorrelation matrix is crucial in the above argument. When G is not restricted to be a pseudocorrelation matrix, it is indeed possible to construct an example in $\mathcal{S}^3$ showing that the SSOSC may still hold even when H contains some *zero* off-diagonal weights (see Qi, 2009, Example 3.9).

Let the mapping F be defined by (2.9). The following result states the local invertibility of F near the KKT point $(\overline{X}, \bar{y}, \overline{Z})$ that is important for the convergence analysis of the augmented Lagrangian method for solving the H-weighted problem (1.3).

PROPOSITION 2.5 If the assumption made in Proposition 2.2 is satisfied, in particular, if $H_{ij} > 0$ for all $i, j = 1, \ldots, n$, then there exist a neighbourhood $\mathcal{N}$ 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

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

The proof of this proposition follows directly from Propositions 2.1 and 2.2 and Sun (2006, Theorem 4.1).

3. The augmented Lagrangian method

As we discussed in Section 1, the Lagrangian dual approach is not applicable to the H-weighted problem (1.3) because the metric projection onto $\mathcal{S}^n_+$ under the H-weighting does not have an explicitly computable formula. The consequence is that its corresponding Lagrangian dual problem does not reduce to an explicitly defined unconstrained smooth optimization problem. Compared with the original problem (1.3), 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. For simplicity, for the remainder of this paper we shall assume $H_{ij} > 0$, which implies the SSOSC at the solution.

3.1 *Outline of the augmented Lagrangian method*

Let the augmented Lagrangian function be defined by (1.9) with $c > 0$. The augmented Lagrangian method for solving (1.3) 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 Lagrange multiplier. At the kth iteration, determine

$$X^{k+1} \in \arg\min L_{c_k}(X, y^k, Z^k). \tag{3.1}$$

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

$$\begin{cases} y^{k+1} := y^k + c_k(b - \text{diag}(X^{k+1})), \\ Z^{k+1} := (Z^k - c_k X^{k+1})_+ \end{cases} \tag{3.2}$$

and update c_{k+1} by

$$c_{k+1} := c_k \quad \text{or} \quad c_{k+1} > c_k \tag{3.3}$$

according to certain rules.

As for the global convergence and the rate of convergence of the augmented Lagrangian method for the H-weighted problem (1.3), we can directly use the convergence theory developed by Rockafellar (1976a, Theorem 2, 1976b, Theorem 5) for general convex programming problems, supported by Pro position 2.5. We will state such a result (Theorem 3.4) for the practical Algorithm 3.3.

Recall that, for any given $c_k > 0$, the convex function $\nu_{c_k}(\cdot)$ defined in (1.10) is continuously differentiable with

$$\nabla \nu_{c_k}(y^k, Z^k) = \begin{pmatrix} -(b - \mathrm{diag}(X^{k+1})) \\ \frac{1}{c_k}(Z^k - (Z^k - c_k X^{k+1})_+) \end{pmatrix}. \tag{3.4}$$

This means that the sequence $\{(y^{k+1}, Z^{k+1})\}$ generated by the augmented Lagrangian method (3.1)–(3.3) can be regarded as a gradient descent method applied to the augmented Lagrangian dual problem (1.10) with a step length c_k at the kth iteration:

$$(y^{k+1}, Z^{k+1}) = (y^k, Z^k) - c_k \nabla \nu_{c_k}(y^k, Z^k), \quad k = 0, 1, \ldots. \tag{3.5}$$

Consequently, one may expect a slow convergence inherited by the gradient method. Interestingly, Theorem 3.4 implies that the sequence $\{(y^{k+1}, Z^{k+1})\}$ converges to $(\bar{y}, \overline{Z})$ at a linear rate that is 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 (see Sun *et al.*, 2008) for the equation

$$\nu_{c_k}(y, Z) = 0, \tag{3.6}$$

as long as c_k is sufficiently large. In fact, it was proven in Sun *et al.* (2008) (in the NSDP setting) that, for any c large enough, $\nabla \nu_c$ is semismooth at $(\bar{y}, \overline{Z})$ and one has the following estimate:

$$V^{-1} = c\mathcal{I} + O(c^{-1})$$

for all $V \in \partial(\nabla \nu_c)(\bar{y}, \overline{Z})$. Thus, by using the fact that $\partial(\nabla \nu_c)(\cdot)$ is upper semicontinuous, for all $c_k \equiv c$ sufficiently large, the term $-c_k \nabla \nu_{c_k}(y^k, Z^k)$ in (3.5) can be regarded as a good approximation to the semismooth Newton direction $-V_k^{-1} \nabla \nu_{c_k}(y^k, Z^k)$, where $V_k \in \partial(\nabla \nu_{c_k})(y^k, Z^k)$ for the semismooth equation (3.6). It is this interpretation that attracted us to attempt to apply the augmented Lagrangian method to the H-weighted problem (1.3) in the first place.

3.2 *A semismooth Newton-CG method*

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

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

$$\min_{X \in \mathcal{S}^n} \theta(X). \tag{3.7}$$

Since $\theta(\cdot)$ is a convex function, solving (3.7) is equivalent to solving the following nonsmooth equation (see the definitions of $\mathcal{Q}$ and C in (2.1)):

$$0 = \nabla\theta(X) = \mathcal{Q}(X) - \mathrm{diag}(y + c(b - \mathrm{diag}(X))) - \Pi_{\mathcal{S}^n_+}(Z - cX) - C. \tag{3.8}$$

It was proven in Sun & Sun (2002) that the projection operator $\Pi_{\mathcal{S}^n_+}(\cdot)$ is strongly semismooth (see Chen *et al.*, 2003, for some extensions). Since all other terms in $\nabla\theta(\cdot)$ are linear, (3.8) is a semismooth equation for which the generalized Newton's method has been well developed (see Kummer, 1988; Qi & Sun, 1993). 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 $\partial^2\theta(X)$ (also known as the generalized Hessian of $\theta(\cdot)$ at X), is given by

$$\partial^2\theta(X) = \mathcal{Q} + c(\mathcal{I} + \partial\Pi_{\mathcal{S}^n_+}(Z - cX)). \tag{3.9}$$

The Newton method for the semismooth equation (3.8) is then defined by

$$X^{k+1} = X^k - V_k^{-1}(\nabla\theta(X^k)), \quad V_k \in \partial^2\theta(X), \quad k = 0, 1, \ldots. \tag{3.10}$$

The implementation of the Newton method (3.10) 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

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

The identity operator $\mathcal{I}$ is obviously positive semidefinite and so is any W in $\partial\Pi_{\mathcal{S}^n_+}(Z - cX)$ (see Meng *et al.*, 2005, Proposition 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 Pang *et al.* (2003, Lemma 11).

Now we are ready to describe the algorithm for solving problem (3.7).

ALGORITHM 3.1 (A semismooth Newton-CG method)

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

Step 1: Select an element $V_j \in \partial^2\theta(X^j)$, and compute $s_j := \min\{\tau_1, \tau_2\|\nabla\theta(X^j)\|\}$. Then apply the CG method (Hestenes & Stiefel, 1952) starting with the zero vector as the initial search direction to

$$\nabla\theta(X^j) + (V_j + s_j I)\Delta X = 0 \tag{3.11}$$

to find a search direction ΔX^j such that

$$\|\nabla\theta(X^j) + (V_j + s_j I)\Delta X^j\| \leqslant \eta_j\|\nabla\theta(X^k)\|, \tag{3.12}$$

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

Step 2: Let l_j be the smallest non-negative integer l such that

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

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 $j \geqslant 0$ we have that $V_j + s_j I$ is positive definite, one can always use the CG method to find ΔX^j such that (3.12) 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, we have that

$$\frac{1}{\lambda_{\max}(V_j + s_j I)} \|\nabla\theta(X^j)\|^2 \leqslant \langle -\nabla\theta(X^j), \Delta X^j \rangle \leqslant \frac{1}{\lambda_{\min}(V_j + s_j I)} \|\nabla\theta(X^j)\|^2, \tag{3.13}$$

where for any matrix $A \in \mathcal{S}^n$ the smallest and largest eigenvalue of A are represented by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively. For a proof of (3.13) see Zhao *et al.* (2008). Therefore Algorithm 3.1 is well defined as long as $\nabla\theta(X^j) \neq 0$, and its convergence analysis can be conducted in a similar way to that in Qi & Sun (2006, Theorem 5.3). We state these results in the next theorem, whose proof is omitted for brevity.

THEOREM 3.2 Suppose that in Algorithm 3.1 we have $\nabla\theta(X^j) \neq 0$ for all $j \geqslant 0$. Then Algorithm 3.1 is well defined and the generated iteration sequence $\{X^j\}$ converges to the unique solution X^* of problem (3.7) quadratically.

In our numerical experiments the parameters used in Algorithm 3.1 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 addressed the fundamental issue of solving problem (3.1). In order to use the augmented Lagrangian method (3.2) for solving the H-weighted problem (1.3), we need to know when to terminate Algorithm 3.1 without affecting the convergence of the augmented Lagrangian method outlined in Section 3.1 so as to make the method practical. Fortunately, Rockafellar (1976a,b) has already provided a solution to this.

For each $k \geqslant 0$ we 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 criteria considered by Rockafellar for general convex optimization problems (see Rockafellar, 1976a,b) but tailored to our needs:

$$\begin{cases} \dfrac{1}{h_{\min}^2}\|\nabla\theta_k(X^{k+1})\|^2 \leqslant \dfrac{\varepsilon_k^2}{2c_k}, & \varepsilon_k > 0, \displaystyle\sum_{k=0}^{\infty} \varepsilon_k < \infty, \\ \dfrac{1}{h_{\min}^2}\|\nabla\theta_k(X^{k+1})\|^2 \leqslant \dfrac{\delta_k^2}{2c_k}\|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|^2, & \delta_k > 0, \displaystyle\sum_{k=0}^{\infty} \delta_k < \infty, \\ \|\nabla\theta_k(X^{k+1})\| \leqslant \delta_k'/c_k\|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|, & 0 < \delta_k' \longrightarrow 0, \end{cases} \tag{3.14}$$

where $h_{\min} := \min\{H_{ij} | i, j = 1, \ldots, n\}$ and (y^{k+1}, Z^{k+1}) is defined by (3.2).

Finally, a ready-to-implement version of the augmented Lagrangian method (3.1)–(3.3) can be described as follows.

ALGORITHM 3.3 (A practical augmented Lagrangian method)

Step 0: Let $c_0 > 0$ and $\kappa > 1$ be given. 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 Lagrange multipliers. Let $k := 0$.

Step 1: Apply Algorithm 3.1 to the 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 (3.14).

Step 2: Compute (y^{k+1}, Z^{k+1}) by (3.2) 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 convergence of the algorithm, we can directly use Rockafellar (1976a, Theorem 2) and (1976b, Theorem 5) for general convex programming problems combined with Proposition 2.5 to get the following convergence theorem for Algorithm 3.3.

THEOREM 3.4 Let $(\overline{X}, \bar{y}, \overline{Z})$ be the unique KKT point of problem (1.3). Let $\zeta > 0$ be the constant given in Proposition 2.5. Let (X^k, y^k, Z^k) be the sequence generated by Algorithm 3.3 with $\lim_{k\to\infty} c_k = c_\infty \leqslant \infty$. 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 we have

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

$$\|X^{k+1} - \overline{X}\| \leqslant a'_k \|(y^{k+1}, Z^{k+1}) - (y^k, Z^k)\|,$$

where

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

and

$$a'_k := \zeta(1 + \delta'_k)/c_k \quad \longrightarrow \quad a'_\infty = \zeta/c_\infty.$$

4. Numerical results

In this section we report our numerical experiments conducted for the H-weighted nearest correlation problem (1.3) 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 Qi & Sun (2006) for solving the equally weighted nearest correlation matrix problem

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|X - G\|^2 \\ \text{such that} \quad & X_{ii} = 1, \quad i = 1, \ldots, n, \\ & X \in \mathcal{S}^n_+ \end{aligned}$$

and Z^0 is set to be

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

The stopping criterion for terminating Algorithm 3.3 is

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

where

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

and for each $k \geqslant 0$ we have

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

Note that the three terms in defining Tol_{k+1} correspond to the three conditions in the KKT system (2.8). In Step 1 of Algorithm 3.3, X^{k+1} is computed to satisfy

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

which is based on (3.14). 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 15 June 2006) from the lagged data sets of RiskMetrics (`www.riskmetrics.com/stddownload_edu.html`). For testing purposes we perturb G to

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

where $\alpha \in (0, 1)$ and E is a 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 $G_{ii} = 1$, where $i = 1, \ldots, n$.

Our second example is randomly generated with $n = 100, 500, 1{,}000$ and $1{,}500$, 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))`. It is then perturbed to

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

where $\alpha \in (0,1)$ and 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 changes the correlation matrix into a pseudocorrelation matrix. Our numerical results are reported in Tables 1 and 2, where `IP-QSDP` refers to Toh's inexact IPM with the PSQMR as the iterative solver (Toh, 2008). `Iter` and `LiSys` stand for the number of total iterations and the number of total linear systems solved, respectively. `Res` represents the relative residue computed at the last iterate and is given by

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

where

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

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 and 2 and other similar testing results not reported here, we have observed that our algorithm is not only faster but also more robust with respect to the perturbed noise level α than `IP-QSDP`, particularly, for those cases in which a good initial correlation matrix estimation is available, as in many real-world situations. Taking Table 1 as an example, we can see that Algorithm 3.3 takes almost the same time for different α, while IP-QSDP is much more sensitive to α and terminates prematurely for $\alpha = 0.005$. It is also worth mentioning that the main costs of Algorithm 3.3 include two parts, namely, the full spectral decomposition of matrices for computing the function $\theta_k(\cdot)$ and the CG steps for solving the linear systems (3.12). In cases when the condition number in the linear system (3.12) is large, the latter dominates the computing time. This explains why, in Table 2, Algorithm 3.3 (and also IP-QSDP) may take more computing time even when the perturbation is smaller.

TABLE 1 *Numerical results of Example 4.1*

Algorithm	α	CPU time	`Iter`	`LiSys`	`Res`
3.3	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 2 *Numerical results of Example 4.2*

Algorithm	n	α	CPU time	Iter	LiSys	Res
3.3	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}
3.3	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}
3.3	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}
3.3	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

5. Conclusions

The convergence theory for the augmented Lagrangian method for the convex optimization problem has been well established by Rockafellar (1976a,b). 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 Rockafellar (1976a,b). It turns out that the constraint nondegeneracy property and the SSOSC 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.

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 to be fixed or contained in some confidence intervals, i.e.,

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|H \circ (X - G)\|^2 \\ \text{such that} \quad & X_{ii} = 1, \quad i = 1, \ldots, n, \\ & X_{ij} \geqslant l_{ij}, \quad (i, j) \in \mathcal{B}_l, \\ & X_{ij} \leqslant u_{ij}, \quad (i, j) \in \mathcal{B}_u, \\ & X \in \mathcal{S}_+^n, \end{aligned} \tag{5.1}$$

where $\mathcal{B}_l$ and $\mathcal{B}_u$ are two index subsets of $\{(i, j)|1 \leqslant i < j < n\}$, $l_{ij} \in [-1, 1]$ for all $(i, j) \in \mathcal{B}_l$, $u_{ij} \in [-1, 1]$ for all $(i, j) \in \mathcal{B}_u$ and $l_{ij} \leqslant u_{ij}$ 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 comparison for problems of reasonable sizes.

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 (Toh, 2008). 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. The authors are also very grateful to the editor and two referees for their constructive comments.

Funding

Engineering and Physical Sciences Research Council (EP/D502535/1 to H.Q.); Academic Research Fund (R-146-000-104-112); the Risk Management Institute (R-703-000-004-720, R-703-000-004-646 to D.S.).

REFERENCES

ALIZADEH, F., ALIZADEH, J.-P. A. & OVERTON, M. L. (1997) Complementarity and nondegeneracy in semidefinite programming. *Math. Programming*, **77**, 111–128.

ARNOLD, V. I. (1971) On matrices depending on parameters. *Russ. Math. Surv.*, **26**, 29–43.

BHANSALI, V. & WISE, B. (2001) Forecasting portfolio risk in normal and stressed market. *J. Risk*, **4**, 91–106.

BONNANS, J. F. & SHAPIRO, A. (1998) Nondegeneracy and quantitative stability of parameterized optimization problems with multiple solutions. *SIAM J. Optim.*, **8**, 940–946.

BONNANS, J. F. & SHAPIRO, A. (2000) *Perturbation Analysis of Optimization Problems*. New York: Springer.

BORCHERS, B. & YOUNG, J. G. (2007) Implementation of a primal–dual method for SDP on a shared memory parallel architecture. *Comput. Optim. Appl.*, **37**, 355–369.

BORSDORF, R. (2007) A Newton algorithm for the nearest correlation matrix. *Master's Thesis*, School of Mathematics, University of Manchester.

BORSDORF, R. & HIGHAM, N. J. (2009) A preconditioned Newton algorithm for the nearest correlation matrix. *IMA J. Numer. Anal.* (in press). doi:10.1093/imanum/drn085.

BOYD, S. & XIAO, L. (2005) Least-squares covariance matrix adjustment. *SIAM J. Matrix Anal. Appl.*, **27**, 532–546.

CHEN, X., QI, H. & TSENG, P. (2003) Analysis of nonsmooth symmetric matrix valued functions with applications to semidefinite complementarity problems. *SIAM J. Optim.*, **13**, 960–985.

CLARKE, F. H. (1983) *Optimization and Nonsmooth Analysis*. New York: Wiley. (reprinted by SIAM (Philadelphia) in 1990).

EAVES, B. C. (1971) On the basic theorem for complementarity. *Math. Programming*, **1**, 68–75.

HESTENES, M. R. & STIEFEL, E. (1952) Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, **49**, 409–436.

HIGHAM, N. J. (1988) Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.*, **103**, 103–118.

HIGHAM, N. J. (2002) Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.*, **22**, 329–343.

JOHNSON, C. R., KROSCHEL, B. & WOLKOWICZ, H. (1998) An interior-point method for approximate positive semidefinite completions. *Comput. Optim. Appl.*, **9**, 175–190.

KUMMER, B. (1988) Newton's method for nondifferentiable functions. *Adv. Math. Optim. Math. Res.*, **45**, 114–125.

MALICK, J. (2004) A dual approach to semidefinite least-squares problems. *SIAM J. Matrix Anal. Appl.*, **26**, 272–284.

MENG, F., SUN, D. & ZHAO, G. Y. (2005) Semismoothness of solutions to generalized equations and the Moreau–Yosida regularization. *Math. Programming*, **104**, 561–581.

PANG, J. S., SUN, D. & SUN, J. (2003) Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Math. Oper. Res.*, **28**, 39–63.

QI, H. (2009) Local duality of nonlinear semidefinite programming. *Math. Oper. Res.*, **34**, 124–141.

QI, H. & SUN, D. (2006) A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM J. Matrix Anal. Appl.*, **28**, 360–385.

QI, H. & SUN, D. (2009) Correlation stress testing for value-at-risk: an unconstrained convex optimization approach. *Comput. Optim. Appl.* (in press). doi:10.1007/s10589-008-9231-4.

QI, L. & SUN, J. (1993) A nonsmooth version of Newton's method. *Math. Programming*, **58**, 353–367.

ROBINSON, S. M. (1984) Local structure of feasible sets in nonlinear programming, part II: nondegeneracy. *Math. Programming Study*, **22**, 217–230.

ROBINSON, S. M. (1987) Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity. *Math. Programming Study*, **30**, 45–66.

ROBINSON, S. M. (2003) Constraint nondegeneracy in variational analysis. *Math. Oper. Res.*, **28**, 201–232.

ROCKAFELLAR, R. T. (1974) *Conjugate Duality and Optimization*. Philadelphia, PA: SIAM.

ROCKAFELLAR, R. T. (1976a) Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, **1**, 97–116.

ROCKAFELLAR, R. T. (1976b) Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, **14**, 877–898.

SCHWERTMAN, N. C. & ALLEN, D. M. (1979) Smoothing an indefinite variance–covariance matrix. *J. Stat. Comput. Simul.*, **9**, 183–194.

SHAPIRO, A. & FAN, M. K. H. (1995) On eigenvalue optimization. *SIAM J. Optim.*, **5**, 552–569.

STURM, J. F. (1999) Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, **11/12**, 625–653.

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

SUN, D. & SUN, J. (2002) Semismooth matrix valued functions. *Math. Oper. Res.*, **27**, 150–169.

SUN, D., SUN, J. & ZHANG, L. W. (2008) The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math. Programming*, **114**, 349–391.

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

TOH, K. C., TÜTÜNCÜ, R. H. & TODD, M. J. (2007) Inexact primal–dual path-following algorithms for a special class of convex quadratic SDP and related problems. *Pac. J. Optim.*, **3**, 135–164.

TSENG, P. (1998) Merit functions for semi-definite complementarity problems. *Math. Programming*, **83**, 159–185.

TÜTÜNCÜ, R. H., TOH, K. C. & TODD, M. J. (2003) Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Programming*, **95**, 189–217.

WIERZBICKI, A. P. & KURCYUSZ, S. (1977) Projection on a cone, penalty functionals and duality theory for problems with inequality constraints in Hilbert space. *SIAM J. Control Optim.*, **15**, 25–56.

ZHAO, Z. Y., SUN, D. & TOH, K. C. (2008) A Newton-CG augmented Lagrangian method for semidefinite programming. *Manuscript*. Singapore: Department of Mathematics, National University of Singapore (revised February 2009).