Correlation Stress Testing for Value-at-Risk: An Unconstrained Convex Optimization Approach

Houduo Qi
School of Mathematics, The University of Southampton, Highfield, Southampton SO17 1BJ, UK. E-mail: hdqi@soton.ac.uk.

Defeng Sun
Department of Mathematics, National University of Singapore, Singapore.
Email: matsundf@nus.edu.sg.

Abstract

Correlation stress testing is employed in several financial models for determining the value-at-risk (VaR) of a financial institution's portfolio. The possible lack of mathematical consistence in the target correlation matrix, which must be positive semidefinite, often causes breakdown of these models. The target matrix is obtained by fixing some of the correlations (often contained in blocks of submatrices) in the current correlation matrix while stressing the remaining to a certain level to reflect various stressing scenarios. The combination of fixing and stressing effects often leads to the mathematical inconsistence of the target matrix. It is then naturally to find the *nearest correlation matrix* to the target matrix with the fixed correlations unaltered. However, the number of fixed correlations could be potentially very large, posing a computational challenge to existing methods. In this paper, we propose an *unconstrained convex optimization approach* by solving one or a sequence of continuously differentiable (but not twice continuously differentiable) convex optimization problems, depending on different stress patterns. This research fully takes advantage of the recently developed theory of strongly semismooth matrix valued functions, which makes fast convergent numerical methods applicable to the underlying unconstrained optimization problem. Promising numerical results on practical data (RiskMetrics database) and randomly generated problems of larger sizes are reported.

Subject classifications: Finance: portfolio. Programming: nonlinear.
Area of review: Financial engineering.

1 Introduction

Stress testing, an important tool to "gauge how the value of an institution's portfolio of securities and derivatives will be affected by large movements in, say, stock prices or exchange rates" (Fender, Gibson, and Mosser 2001), is conducted on a regular basis by financial institutions. One common feature of the stress test in several financial models, such as the covariance VaR and the Monte Carlo VaR models (Alexander 2001, Section 9.6), is to determine the value-at-risk (VaR) of a bank's portfolio via the stressed covariance matrices.

Covariance matrix based stress testing is often known as the correlation stress testing because of the way that the stress is reflected. Suppose that V is the current estimated covariance matrix based on historical data. By decomposing V as $V = DCD^T$, where D is a diagonal matrix with positive entries representing volatilities and C is the correlation matrix[1], one can stress volatilities (perturbing the diagonal terms of D) separately from stressing correlations (Kupiec 1998). However, unlike the diagonal matrix D, alterations in the correlation matrix C may cause breakdown of the underlying VaR models because the newly formed correlation matrix, denoted by $\widehat{C}$ and is often referred to in the literature as the target correlation matrix, is no longer guaranteed to be positive semidefinite, an essential mathematical property that any covariance/correlation matrix must satisfy. Finger (1997) illustrated this, in association with the covariance VaR model in RiskMetrics, by constructing an example.

The obstacle in conducting correlation stress testing is due to the fact that there are several desirable properties including the positive semidefiniteness that the target correlation matrix $\widehat{C}$ must possess simultaneously. One such situation is the case that some of the correlations in $\widehat{C}$ should be kept unchanged. To see why this is so in reality, let us consider a portfolio consisting of n assets $\{s_1, \ldots, s_m, s_{m+1}, \ldots, s_n\}$. Then the current estimated correlation matrix C takes the form $C = \begin{pmatrix} C_1 & C_2 \\ C_2^T & C_3 \end{pmatrix}$, where $C_1 \in \mathbb{R}^{m\times m}$ is the correlation matrix corresponding to the first group of assets $\{s_1, \ldots, s_m\}$, $C_3 \in \mathbb{R}^{(n-m)\times(n-m)}$ is the correlation matrix for the second group of assets $\{s_{m+1}, \ldots, s_n\}$, and $C_2 \in \mathbb{R}^{m\times(n-m)}$ is the cross-group correlation matrix that contains correlations between every pair of s_i and s_j, with s_i in the first group and s_j in the second group.

Suppose now that we aim to conduct stress testing on the assets in the second group of our portfolio by changing correlations in C_3 to form a new symmetric matrix $\widehat{C}_3$[2]. Of course, not all correlations in C_3 are necessarily subject to change. While the correlations in C_2 may or may not be changed depending on different situations, the ones in C_1 should not be altered (see Finger 1997 for a concrete example consisting of four Asian currencies and three non-Asian currencies). The target correlation matrix $\widehat{C}$ should therefore take the form $\widehat{C} = \begin{pmatrix} \widehat{C}_1 & \widehat{C}_2 \\ \widehat{C}_2^T & \widehat{C}_3 \end{pmatrix}$ with $\widehat{C}_1 = C_1$ and/or $\widehat{C}_2 = C_2$. However, when $\widehat{C}_2$ is not properly selected (e.g., $\widehat{C}_2 = C_2$), the matrix

[1] A real symmetric matrix is called a correlation matrix if it is positive semidefinite and all its diagonal entries are ones.

[2] Each financial institution has its own views on how $\widehat{C}_3$ can be forecasted. We will not address this issue here. Rather we focus on calibrating improperly stressed correlation matrices.

$\widehat{C}$ may fail to be positive semidefinite even if $\widehat{C}_3$ is positive (semi)definite. This phenomenon will lead to physically infeasible negative volatilities of some assets if $\widehat{C}$ were used to calculate the VaR of a portfolio. So, immediately a replacement matrix X for $\widehat{C}$ has to be introduced. In order to compute a meaningful VaR under the stress test, this replacement must be positive (semi)definite and in the meantime captures the stress information in $\widehat{C}_3$. An intrinsic question then arises: what is the best replacement and in what sense? The main purpose of this paper is to answer this question.

Many authors from the finance industry and the academia (Finger 1997, Kupiec 1998, Rebonato and Jäckel 2000, Bhansali and Wise 2001, Rapisrada, Brigo, and Mercurio 2000, León et al. 2002, Turkey, Epperlein, and Christofides 2003, Dash 2004, to name only a few) have considered the correlation stress testing problem in different scenarios by adding various constraints on the replacement matrix X. The constraints in these scenarios may look rather different. But, they actually can be classified and completely captured through the following three types of constraints.

(H1) All of the diagonal elements of X must be 1,

(H2) X has to be positive semidefinite (mathematical consistence), and

(H3) $X_{ij} = \widehat{C}_{ij}$ for some indices $(i, j) \in \mathcal{B}$, where the index set $\mathcal{B}$ specifies the locations of the fixed correlations.

Sometimes, we need the positive definiteness constraint, which replaces (H2):

(H2′) X is positive definite.

Constraints (H1) and (H2) together ensure that X is a correlation matrix. Constraint (H3) simply specifies those correlations that are not allowed to change. The index set $\mathcal{B}$ may assume various forms. The positive definiteness constraint (H2′) is particularly important in methods where it is used of the inverse of a covariance matrix, e.g., RiskMetrics VaR model using conditional covariance matrix (RiskMetrics document 1996, Page 185) or the Cholesky factor of a covariance matrix, e.g., the Monte Carlo VaR model (Alexander 2001, Section 9.4).

As mentioned above, there are a few available methods known to the finance community to select a replacement matrix X satisfying some/all of the constraints. For example, ignoring the fixed element constraint (H3), the hyperspherical decomposition method of Rebonato and Jäckel (2000) and its modified variant by Bhansali and Wise (2001) and Kercheval (2006) try to find the nearest correlation matrix to $\widehat{C}$ (measured in the Frobenius norm of matrices). See also Chapters 23 and 24 in Dash (2004) for more treatments on this case. However, the resulting optimization problem is highly nonlinear and nonconvex[3]. The shrinkage method of Kupiec (1998) and the sequential single-stress method of Turkey, Epperlein, and Christofides (2003), where the case $\widehat{C}_1 = C_1$ and $\widehat{C}_2 = C_2$ is formally referred to as the *local correlation stress testing*,

[3]The hyperspherical decomposition method can be adapted to incorporate the constraint (H3), but it would lead to a nonlinearly constrained nonconvex optimization problem, which is often very difficult to solve for a global solution.

both are capable of handling the constraints (H1)-(H3), but, as commented by Rebonato and Jäckel (2000) that "there is no way of determining to what extent the resulting matrix is optimal in any easily quantifiable sense"[4]. Finger's method as well as other spectral decomposition based methods proposed in those studies also suffer similar drawbacks.

On a parallel development in matrix analysis and optimization research, there also exist a number of successful methods of finding the nearest correlation matrix to a given matrix. Those methods include the alternating projection method of Higham (2002), quasi-Newton methods of Malick (2004) and Boyd and Xiao (2005), Newton's method of Qi and Sun (2006), and the inexact primal dual path-following method of Toh, Tütüncü, and Todd (2006). Although all of those methods are conceptually applicable to handle the fixed element constraint (H3), theory and numerical experiment are mainly developed to address the constraints (H1) and (H2). The general consensus in those studies is that the *nearest correlation matrix problem* of satisfying (H1) and (H2) can be efficiently solved even when n is moderately large. However, the presence of the fixed correlation constraint (H3) may cause a great deal of difficulty to existing methods.

The first issue that has to be dealt with is the feasibility problem. If the fixed correlations are not from an existing positive correlation matrix there may not even exist a true correlation matrix satisfying the fixed correlations (e.g., the problem is not well posed). This is why we assume that the current correlation matrix C is positive definite[5]. The second issue is the extra computational complexity that the fixed correlation constraint may bring to a already very difficult problem when n is large. For some cases, this extra complexity may not cause too much concern. The commonly studied local correlation stress testing in Turkey, Epperlein, and Christofides (2003), for example, can be reduced to a nearest correlation matrix via the Schur complement decomposition technique. Therefore, the extra complexity can be removed for this case. However, for many other cases this extra complexity has to be dealt with directly or indirectly. Furthermore, the total number, which is sometimes very large and is denoted by $\kappa(\mathcal{B})$, of the fixed correlations alone is not an accurate indicator for the complexity. The structure of $\mathcal{B}$ (i.e., whereabouts of those fixed correlations) seems to be another attribute in the complexity. Therefore, in this paper we mainly address the case where the stressed correlations have *band* structure, which is equivalent to say $\mathcal{B}$ is contained in a fixed diagonal square block. The major reason for considering this case is that theory and algorithms can be developed nicely and can be readily extended to more general cases. The third issue is whether or not a large number of fixed correlations may cause loss of the quadratic convergence of Newton's method developed for the case without the fixed correlation constraint.

To put it in another way, in correlation stress testing we face a task that requires to construct a correlation matrix with a large number of pre-fixed elements. On top of this, one is to seek such a correlation matrix that is nearest to a given target matrix (measured in the Frobenius norm here and throughout the paper). Moreover, we need a fast algorithm to accomplish the task. Combination of all of those features gives rise to a very challenging task.

[4]The comment is only on the shrinkage method, but it apparently applies to the sequential single-stress method.

[5]In practice, C may not be positive definite due to various reasons including insufficiency of historical data and numerical truncations. To make sure that the fixed correlations do come from a positive definite correlation matrix, Algorithm 3.1 (with $\tau > 0$) will be used to achieve this purpose at a very low cost.

The main purpose of this paper is to solve this challenging problem via an unconstrained optimization approach, which refers to either the *Lagrangian* dual approach or the *augmented Lagrangian* dual approach, depending on different situations in the correlation stress testing. Roughly speaking, the Lagrangian dual approach is for the case when $\kappa(\mathcal{B})$ is relatively small and the augmented Lagrangian dual approach is for the case when $\kappa(\mathcal{B})$ is relatively large. Considerable part of this paper is devoted to modelling and methodology. In the modelling part, we formulate the correlation stress testing problem via convex optimization techniques, distinguishing several interesting cases from each other (Section 2). In the methodological part, we describe in details how the Lagrangian dual method (Section 3) and the augmented Lagrangian dual method (Section 4) can be developed to solve the modelled convex optimization problems. At the core of either of the approach is the quadratically convergent Newton method (Algorithm 3.1), which involves heavy calculations to reach formulae necessary for its numerical implementation and is only made possible due to recent advances on the theory of strongly semismooth matrix valued functions. We include the calculation as well as some theoretical proofs in Appendices (Appendix A, B, and C).

The purpose of this arrangement is threefold. Firstly, for readers who are familiar with the Lagrangian and augmented Lagrangian theory, the methodological part clearly shows what the best of the theory we can have when coming to the correlation stressing testing. Secondly, for readers who are not very familiar with the theory, the minimal coverage of the methodology provides sufficient material to understand it without having to referring to a vast number of related references. Last but not least, for practitioners who may be only interested in how to use the methodology proposed here, we provide concrete algorithms (Algorithm 3.1 and Algorithm 4.3) and formulae (Appendix A) to make it easier for them to adapt the companion `MATLAB` codes (available at `http://www.math.nus.edu.sg/~matsundf`) to their own correlation stress testing problems. We hope that the superb numerical evidence reported in this paper, backed by strong theoretical convergence results, may lead to further research on this important topic and relieve practitioners from spending countless hours searching for an efficient numerical method for dealing with the challenging task of conducting the correlation stress testing in their financial models.

The paper is thus organized as follows. In the next section, we formulate the correlation stress testing problem of various types as a convex optimization problem. We study its Lagrangian dual and the augmented Lagrangian dual approaches in the next two sections. Section 5 contains our numerical results for examples collected from the relevant literature and some randomly generated hard problems. We conclude the paper in Section 6. Appendices include detailed calculations necessary for implementing the Newton method and some theoretical proofs.

Notation: $\mathcal{S}^n$ and $\mathcal{S}^n_+$ are, respectively, the linear space of $n \times n$ symmetric matrices and the cone of positive semidefinite matrices in $\mathcal{S}^n$; and $\|\cdot\|$ is the Frobenius norm defined by the trace inner product $\langle A, B\rangle = \mathrm{tr}(AB)$ for $A, B \in \mathcal{S}^n$. Sometimes, we use $X \succ (\succeq) 0$ meaning X is positive (semi-) definite.

For $X \in \mathcal{S}^n$, X_+ denotes the orthogonal projection onto $\mathcal{S}^n_+$ and X_{ij} denotes the (i,j)th entry of X. We use $\circ$ to denote the Hardmard product of matrices, i.e., for any $A, B \in \mathcal{S}^n$

$$A \circ B = [A_{ij}B_{ij}]^n_{i,j=1}.$$

The matrix E denotes the matrix of all ones in $\mathcal{S}^n$. For subsets α and β of $\{1, 2, \ldots, n\}$, $B_{\alpha\beta}$ denotes the submatrix of B indexed by α and β. For any pair (i, j), we use E^{ij} to denote the matrix whose (i, j)th entry is 1 and all other entries are zeros. Let e denote the vector of all ones. For a vector x, Diag(x) is the diagonal matrix whose diagonal entries are the vector x. Conversely, for a square matrix X, diag(X) denotes the vector formed by the diagonal entries of X.

2 The Convex Optimization Formulation

This section contains a straightforward convex optimization reformulation of finding the nearest correlation matrix to $\widehat{C}$ satisfying constraints (H1), (H2) or (H2′), and (H3). We will distinguish several cases with each having its own importance. In particular, we will treat the case $\mathcal{B} \neq \emptyset$ and $\mathcal{B} = \emptyset$ separately. We will also address how to handle the positive definiteness constraint (H2′) in our formulation.

2.1 The Case $\mathcal{B} \neq \emptyset$

Recall that the index set $\mathcal{B}$ specifies the locations of fixed correlations in $\widehat{C}$. For the purpose of later development, we assume that $\mathcal{B}$ takes the following general form:

$$\mathcal{B} := \{(i_\ell, j_\ell) \mid i_\ell < j_\ell,\ \ell = 1, \ldots, \kappa\} \tag{1}$$

for some $\kappa \leq n(n-1)/2$. We often use $\kappa(\mathcal{B})$ to indicate the dependence of κ on $\mathcal{B}$. Let c be the vector in $\mathbb{R}^\kappa$ obtained by stacking up the fixed correlations $\widehat{C}_{ij}$, $(i, j) \in \mathcal{B}$ column by column from top to bottom. Then, for any $1 \leq \ell \leq \kappa$, there exists a unique index $(i, j) \in \mathcal{B}$ such that $c_\ell = C_{ij}$. Because of this, to facilitate our description, for a vector $z \in \mathbb{R}^\kappa$, without causing confusion we often write $z \equiv (z_{ij})_{(i,j)\in\mathcal{B}}$ to match the structure of $\mathcal{B}$.

The following two examples illustrate what $\mathcal{B}$ and c may look like.

Example 2.1 (Local correlation stress testing of Turkey, Epperlein, and Christofides 2003) *In this example, $\widehat{C}_1 = C_1$ and $\widehat{C}_2 = C_2$. Then*

$$\left\{ \begin{array}{rcl} \mathcal{B} & = & \{(i, j) \mid i = 1, \ldots, m,\ j = i + 1, \ldots, n\ \}, \\ c_{ij} & = & C_{ij}, \quad \textit{for } (i, j) \in \mathcal{B}. \end{array} \right. \tag{2}$$

Example 2.2 (Band correlation stress testing) *Compared to Example 2.1, $\widehat{C}_2$ in this example is allowed to change freely while $\widehat{C}_1 = C_1$ remains unchanged*

$$\left\{ \begin{array}{rcl} \mathcal{B} & = & \{(i, j) \mid 1 \leq i < j \leq m\}, \\ c_{ij} & = & C_{ij}, \quad \textit{for } (i, j) \in \mathcal{B}. \end{array} \right. \tag{3}$$

The special structure of $\mathcal{B}$ in the *local correlation stress testing* in Example 2.1 allows us to use the Schur complement decomposition to transform it to a *nearest correlation matrix problem* of reduced dimension. Consequently, this case can be handled more directly and solved efficiently by existing methods (see Subsection 3.2). As suggested already by its name, the *band*

correlation stress testing in Example 2.2 indicates that the stressed correlations (i.e., those not in $\mathcal{B}$) form a structure of band. The importance of this case is with its generality in the sense that the approach and analysis developed can be readily extended to cover more general cases including the "rectangular" $\mathcal{B}$ (see, Remark R2 in Subsection 3.3).

Once we have the location index set $\mathcal{B}$ and the fixed correlation vector c, any replacement correlation matrix X must satisfy the following conditions:

$$X \in \mathcal{C}^n \;\text{ and }\; X_{ij} = c_{ij} \;\text{ for all } (i,j) \in \mathcal{B}, \tag{4}$$

where $\mathcal{C}^n$ is the set of all $n \times n$ correlation matrices. Note that $\mathcal{B}$ only specifies the upper part of fixed elements in X. But, since X is a symmetric matrix, the lower part of fixed elements is automatically included.

Our eventual goal is to find the nearest correlation matrix to $\widehat{C}$ from all those of satisfying conditions in (4). This leads to the following least-square optimization problem:

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - \widehat{C}\|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X_{ij} = c_{ij}, \quad (i,j) \in \mathcal{B}\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{5}$$

Note that the objective function in (5) is quadratic in X and all the constraints are linear except X being in $\mathcal{S}^n_+$, which is a closed convex cone. So, (5) is a convex optimization problem.

To single out the linear equations in (5), we define two linear operators $\mathcal{A}_1 : \mathcal{S}^n \mapsto \mathbb{R}^n$ and $\mathcal{A}_2 : \mathcal{S}^n \mapsto \mathbb{R}^\kappa$, respectively, by

$$\mathcal{A}_1(X) := \mathrm{diag}\,(X) \;\text{ and }\; (\mathcal{A}_2(X))_{ij} := X_{ij} \text{ for } (i,j) \in \mathcal{B}\,. \tag{6}$$

Note that in (6), for each $X \in \mathcal{S}^n$, $\mathcal{A}_2(X)$ is a column vector in $\mathbb{R}^\kappa$ with the notation being explained earlier. Recall that e is defined to be the vector of all ones in $\mathbb{R}^n$. Problem (5) can thus be equivalently written as

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - \widehat{C}\|^2 \\ \text{s.t.} & \mathcal{A}_1(X) = e\,, \\ & \mathcal{A}_2(X) = c\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{7}$$

This is the problem we aim to solve in this paper. To alleviate the concern about the feasibility issue of this problem as well as the unconstrained dual problems to be developed in the next section, we assume that this problem is strictly feasible. Apparently, a sufficient condition for the strict feasibility is that the current correlation matrix C is positive definite, which is a practical condition required in the financial industry.

2.2 The Case $\mathcal{B} = \emptyset$

This is the simplest case of (7) and is often referred to as the *nearest correlation matrix problem* (Higham 2002):

$$\begin{array}{ll} \min & \frac{1}{2}\|X - \widehat{C}\|^2 \\ \text{s.t.} & \mathcal{A}_1(X) = e\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{8}$$

This problem is always strictly feasible, e.g., $X = I$, regardless whether C is positive definite or not. There are n linear constraints comparing to $n + \kappa(\mathcal{B})$ in (7), where $\kappa(\mathcal{B})$ could be significantly larger than n (i.e., $\kappa(\mathcal{B}) \gg n$).

The nearest correlation matrix problem (8) also distinguishes itself from the general problem (7) numerically. Problem (8) can be efficiently solved even when n is large (e.g., $n = 2,000$), say, by Newton's method of Qi and Sun (2006) or the inexact primal-dual path-following method of Toh, Tütüncü, and Todd (2006). We will make use of this fact in two aspects. First, the Lagarangian dual approach, which the quadratically convergent Newton's method for solving problem (8) (Qi and Sun 2006) is based on, is extended to the case $\mathcal{B} \neq \emptyset$. The computational effeciency consideration of handling a large number $\kappa(\mathcal{B})$ of constraints will naturally lead us to consider an *augmented Lagrangian dual* approach, which handles constraint (H3) more directly. Second, problem (8) is used to generate a good starting point for methods developed for the case $\mathcal{B} \neq \emptyset$.

2.3 Incorporating the Positive Definiteness Constraint (H3)

In a straightforward way, we can add the positive definiteness constraint (H2′) to formulation (7) as follows

$$\begin{array}{ll} \min & \frac{1}{2}\|X - \widehat{C}\|^2 \\ \text{s.t.} & \mathcal{A}_1(X) = e\,, \\ & \mathcal{A}_2(X) = c\,, \\ & X - \tau I \in \mathcal{S}^n_+\,, \end{array} \tag{9}$$

where $0 < \tau < 1$ is a user-specified parameter, usually small (e.g., $\tau = 0.5 \times 10^{-4}$), to ensure that the solution matrix is positive definite. Because of this, problem (9) will be called the regularized version of problem (7).

After simple linear transformations, problem (9) can be reformulated to the form of (7), but with different input matrix and right-hand side constant vector. Specifically, let $Y \equiv X - \tau I$. Then (9) is equivalent to

$$\begin{array}{ll} \min & \frac{1}{2}\|Y - (\widehat{C} - \tau I)\|^2 \\ \text{s.t.} & \mathcal{A}_1(Y) = (1-\tau)e\,, \\ & \mathcal{A}_2(Y) = c\,, \\ & Y \in \mathcal{S}^n_+\,. \end{array} \tag{10}$$

We see that the input matrix now becomes $(\widehat{C} - \tau I)$ (versus $\widehat{C}$ in (7)) and the right-hand side constant vector corresponding to the linear operator $\mathcal{A}_1$ now becomes $(1-\tau)e$ (versus e in (7)).

It is obvious that these two problems have the same level of complexity. The only issue that warrants attention is about the choice of τ. To ensure the strict feasibility of (9), τ has to satisfy $0 \le \tau < \lambda_{\min}(C)$, the smallest eigenvalue of C. Problem (9) and its equivalent form (10) with $0 \le \tau < \lambda_{\min}(C)$ are the main problems we intend to solve. When $\tau = 0$, (9) reduces to (7).

The above discussion also applies to the case where $\mathcal{B} = \emptyset$. We simply write it down below without further comments for easy reference later on:

$$\begin{array}{ll} \min & \frac{1}{2}\|X - \widehat{C}\|^2 \\ \text{s.t.} & \mathcal{A}_1(X) = e\,, \\ & X - \tau I \in \mathcal{S}^n_+\,, \end{array} \tag{11}$$

where $0 < \tau < 1$. This problem is always strictly feasible and always yields a positive definite matrix near to $\widehat{C}$. The corresponding equivalent form for (11) is as follows

$$\begin{array}{ll} \min & \frac{1}{2}\|Y - (\widehat{C} - \tau I)\|^2 \\ \text{s.t.} & \mathcal{A}_1(Y) = (1-\tau)e\,, \\ & Y \in \mathcal{S}^n_+\,. \end{array} \tag{12}$$

3 A Lagarangian Dual Newton Method

It has been widely recognized that the difficulty in updating a correlation matrix to a desired one is to keep it being positive semidefinite. In other words, the constraint $X \in \mathcal{S}^n_+$ is where the difficulty arises. Moreover, in correlation stress testing, another type of difficulty comes from preserving constraint (H3), especially when $\kappa(\mathcal{B})$ is not very small. In this section, we shall focus on dealing with these two types of difficulties, by extending the generalized Newton's method studied in Qi and Sun (2006) based on a Lagrangian dual approach proposed by Rockafellar (1974) for general constrained optimization problems.

3.1 General Discussions

Consider the following least-square semidefinite optimization problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|X - X^0\|^2 \\ \text{s.t.} & \mathcal{A}(X) = b\,, \\ & X \in \mathcal{S}^p_+\,, \end{array} \tag{13}$$

where $X^0 \in \mathcal{S}^p$ is given, $\mathcal{A} : \mathcal{S}^p \to \mathbb{R}^q$ is a linear operator, and $b \in \mathbb{R}^q$. Define the ordinary Lagrangian function $l : \mathcal{S}^p_+ \times \mathbb{R}^q \to \mathbb{R}$ by

$$l(X, y) := \frac{1}{2}\|X - X^0\|^2 + \langle y, b - \mathcal{A}(X)\rangle, \quad (X, y) \in \mathcal{S}^p_+ \times \mathbb{R}^q\,. \tag{14}$$

Let $\theta : \mathbb{R}^q \to \mathbb{R}$ be defined by

$$\theta(y) := -\inf_{x \in \mathcal{S}^p_+} l(X, y), \quad y \in \mathbb{R}^q\,.$$

Then the dual of problem (13) takes the following form:

$$\begin{array}{ll} \max & -\theta(y) \\ \text{s.t.} & y \in \mathbb{R}^q \,. \end{array} \tag{15}$$

The function $\theta(y)$ has a nice analytical form. To present this form, for any $X \in \mathcal{S}^p_+$, we let $\Pi_{\mathcal{S}^p_+}(X)$ denote the metric projection of X onto $\mathcal{S}^p_+$, i.e., $\Pi_{\mathcal{S}^p_+}(X)$ is the unique optimal solution to the following convex programming problem

$$\begin{array}{ll} \min & \frac{1}{2}\|Y - X\|^2 \\ \text{s.t.} & Y \in \mathcal{S}^p_+ \,. \end{array} \tag{16}$$

Then for any $y \in \mathbb{R}^q$, we have

$$\theta(y) = \frac{1}{2}\|X^0 + \mathcal{A}^* y\|^2 - \frac{1}{2}\|X^0 + \mathcal{A}^* y - \Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^* y)\|^2 - \langle b, y\rangle - \frac{1}{2}||X^0||^2, \tag{17}$$

where $\mathcal{A}^* : \mathbb{R}^q \to \mathcal{S}^n$ is the adjoint[6] of $\mathcal{A}$. Since $\mathcal{S}^p_+$ is a closed convex cone, we know from Zarantonello (1971) that

$$\theta(y) = \frac{1}{2}\|\Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^* y)\|^2 - \langle b, y\rangle - \frac{1}{2}||X^0||^2, \quad y \in \mathbb{R}^q \tag{18}$$

and that θ is a continuously differentiable convex function with its gradient at y being given by

$$\nabla\theta(y) = \mathcal{A}\,\Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^* y) - b\,, \qquad y \in \mathbb{R}^q\,. \tag{19}$$

Recall that the generalized Slater condition is said to hold for the convex optimization problem (13) if

$$\left\{ \begin{array}{l} \mathcal{A} : \mathcal{S}^p \to \mathbb{R}^q \text{ is onto,} \\ \exists\ \overline{X} \in \mathcal{S}^p_+ \text{ such that } \ \mathcal{A}(\overline{X}) = b,\ \overline{X} \in \mathrm{int}\left(\mathcal{S}^p_+\right), \end{array} \right. \tag{20}$$

where "int" denotes the topological interior of a given set. The classical duality theory for convex programming of Rockafellar (1974) says that under the generalized Slater condition (20), the following hold: For every real number η, the level set $\{y \in \mathbb{R}^q : \theta(y) \le \eta\}$ is closed, bounded, and convex; and the unique solution to the original problem (13) is given by

$$X^* = \Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^* y^*)\,, \tag{21}$$

where $y^* \in \mathbb{R}^q$ is any optimal solution to the dual problem (15). The relation (21) suggests the following approach: first solve the unconstrained convex optimization problem (15) for y^* and then obtain the unique solution to the original problem (13) by (21). This is exactly the well-known Lagrangian dual approach outlined by Rockafellar (1974). However, θ fails to be twice continuously differentiable because the metric projector $\Pi_{\mathcal{S}^p_+}$ is not continuously differentiable. This seems to imply that one may not be able to get quadratic convergence when Newton's

[6]For the linear operator $\mathcal{A} : \mathcal{S}^p \to \mathbb{R}^q$, its adjoint $\mathcal{A}^* : \mathbb{R}^q \mapsto \mathcal{S}^p$ is defined by $\langle \mathcal{A}^* y, X\rangle = \langle y, \mathcal{A}(X)\rangle$ for any $X \in \mathcal{S}^p$ and $y \in \mathbb{R}^q$.

method is applied to the unconstrained convex optimization problem (15). As a matter of fact, the classical Newton's method is invalid in this situation as the Hessian of $\theta(\cdot)$ at some points may not exist at all. Fortunately, the recent study conducted by Qi and Sun (2006) for the nearest correlation matrix problem (8) indicates that one may still expect a quadratically converging Newton's method by using the fact $\Pi_{\mathcal{S}^p_+}$ is strongly semismooth everywhere in $\mathcal{S}^p$, a key property proven by Sun and Sun (2002) and extended by Chen, Qi, and Tseng (2003) to some more general matrix valued functions.

Denote

$$F(y) := \nabla\theta(y) = \mathcal{A}\,\Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^*y) - b\,, \qquad y \in \mathbb{R}^q\,. \tag{22}$$

Since $\Pi_{\mathcal{S}^p_+}$ is globally Lipschitz continuous with modulus 1, the mapping F is Lipschitz continuous on $\mathbb{R}^q$. According to Redemacher's Theorem (Rockafellar and Wets 1998, Section 9.J), F is differentiable almost everywhere on $\mathbb{R}^q$. We let

$$D_F := \{y \in \mathbb{R}^q |\ F \text{ is differentiable at } y\}\,.$$

Let $F'(x)$ denote the Jacobian of F at $y \in D_F$. The B-subdifferential of F at $y \in \mathbb{R}^q$, a name coined by Qi (1993), is then defined by

$$\partial_B F(y) := \left\{V \in \mathbb{R}^{q\times q} \,|\, V \text{ is an accumulation point of } F'(y^k),\ y^k \to y,\ y^k \in D_F\right\}.$$

The generalized Jacobian in the sense of Clarke (1983) is the convex hull of $\partial_B F(y)$, i.e.,

$$\partial F(y) = \operatorname{conv} \partial_B F(y)\,.$$

If F is strictly differentiable at y, Clarke's generalized Jacobian of F at y reduces to the classical Jacobian of F at y, i.e., $\partial F(y) = \{F'(y)\}$. The generalized Hessian of θ at $y \in \mathbb{R}^q$ is defined as

$$\partial^2\theta(y) := \partial F(y)\,.$$

Define

$$\widehat{\partial}^2\theta(y) := \mathcal{A}\,\partial\Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^*y)\mathcal{A}^*\,, \quad y \in \mathbb{R}^q\,.$$

By Clarke (1983, Page 75), we know from (22) that for $y \in \mathbb{R}^q$ and $d \in \mathbb{R}^q$,

$$\partial^2\theta(y)d \subseteq \widehat{\partial}^2\theta(y)d\,,$$

which implies that if every element in $\widehat{\partial}^2\theta(y)$ is positive definite, then so is every element in $\partial^2\theta(y)$.

Given the above preparations, we can extend directly the generalized Newton method developed in Qi and Sun (2006) from the nearest correlation problem (8) to problem (15) with $\partial^2\theta(\cdot)$ being replaced by $\widehat{\partial}^2\theta(\cdot)$.

Algorithm 3.1 (Newton's Method)

Step 0. *Given* $y^0 \in \mathbb{R}^q$, $\eta \in (0,1)$, $\mu \in (0,1)$, *and* $\rho \in (0,1/2)$. $k := 0$.

Step 1. *Select an element* $V_k \in \widehat{\partial}^2\theta(y^k)$ *and apply the conjugate gradient* (CG) *method of* Hestenes and Stiefel (1952) *to find an approximate solution* d^k *to*

$$\nabla\theta(y^k) + V_k d = 0 \tag{23}$$

such that

$$\|\nabla\theta(y^k) + V_k d^k\| \le \eta_k \|\nabla\theta(y^k)\|\,, \tag{24}$$

where $\eta_k := \min\{\eta, \|\nabla\theta(y^k)\|\}$. *If* (24) *is not achievable or if the condition*

$$\nabla\theta(y^k)^T d^k \le -\eta_k \|d^k\|^2 \tag{25}$$

is not satisfied, let $d^k := -B_k^{-1}\nabla\theta(y^k)$, *where* B_k *is any symmetric positive definite matrix in* $\mathcal{S}^q$.

Step 2. *Let* j_k *be the smallest nonnegative integer* j *such that*

$$\theta(y^k + \rho^j d^k) - \theta(y^k) \le \mu\, \rho^j \nabla\theta(y^k)^T d^k\,.$$

Set $t_k := \rho^{j_k}$ *and* $y^{k+1} := y^k + t_k d^k$.

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

The implementation of Algorithm 3.1 heavily hinges on the availability of computing an element from the set $\widehat{\partial}^2\theta(y^k)$ for all k. Thanks to recent progress on variational analysis of the metric projector $\Pi_{\mathcal{S}^p_+}$, the set $\widehat{\partial}^2\theta(y^k)$ can be completely charaterized. See Appendix A for a detailed account. On the theoretical side of Algorithm 3.1, by Qi and Sun (2006) and Bai, Chu, and Sun (2007), we readily have the following convergence results.

Theorem 3.2 *Assume that both* $\{\|B_k\|\}$ *and* $\{\|B_k^{-1}\|\}$ *in* Algorithm 3.1 *are uniformly bounded. Then, an infinite sequence* $\{y^k\}$ *is generated by* Algorithm 3.1 *and any accumulation point* y^* *of* $\{y^k\}$ *is a solution to the unconstrained convex optimization problem* (15). *Furthermore, if every element in* $\widehat{\partial}^2\theta(y^*)$ *is positive definite at any accumulation point* y^* *of* $\{y^k\}$, *then the whole sequence* $\{y^k\}$ *converges quadratically to the unique solution* y^* *of* (15).

Note that if the generalized Slater condition (20) holds (which is true for our correlation stress testing problems), then the infinite sequence $\{y^k\}$ generated by Algorithm 3.1 is guaranteed to be bounded and thus has at least one accumulation point y^* which solves problem (15). In Theorem 3.2, the crucial condition for obtaining the quadratic convergence of $\{y^k\}$ is on the positive definiteness of every element in $\widehat{\partial}^2\theta(y^*)$. Actually, in proof of this theorem one important hidden condition has also been used: the strong semismoothness of the metric projector $\Pi_{\mathcal{S}^p_+}$. However, as mentioned earlier, this has already been settled by Sun and Sun (2002). So, in the subsequent analysis, we shall mainly focus on checking the positive definiteness of all elements in $\widehat{\partial}^2\theta(y^*)$.

3.2 Local Correlation Stress Testing

Recall that for this case $(\mathcal{B}, c)$ is given by (2). We shall make use of the special structure of $\mathcal{B}$. Let X have the conformal structure of C as

$$X = \begin{pmatrix} X_1 & X_2 \\ X_2^T & X_3 \end{pmatrix}$$

with $X_1 \in \mathcal{S}^m$. For the sake of simplicity and for this subsection only, we let $\tau = 0$. The argument below carries through for $\tau > 0$ with minor modifications.

It is easy to see that problem (9) (now assuming $\tau = 0$) has the following equivalent formation:

$$\begin{array}{lll} \min & \frac{1}{2}\|X - \widehat{C}\|^2 & \\ \text{s.t.} & (X_3)_{ii} = 1,\ i = 1, \ldots, n-m\,, & \\ & X = \begin{pmatrix} C_1 & C_2 \\ C_2^T & X_3 \end{pmatrix} \in \mathcal{S}^n_+ & \end{array} \quad \Longleftrightarrow \quad \begin{array}{lll} \min & \frac{1}{2}\|X_3 - \widehat{C}_3\|^2 & \\ \text{s.t.} & (X_3)_{ii} = 1,\ i = 1, \ldots, n-m\,, & \\ & X = \begin{pmatrix} C_1 & C_2 \\ C_2^T & X_3 \end{pmatrix} \in \mathcal{S}^n_+\,. & \end{array} \tag{26}$$

Note that C is assumed to be positive definite. Therefore, $X \in \mathcal{S}^n_+$ if and only if the Schur complement $(X_3 - C_2^T C_1^{-1} C_2)$ of C_1 in X is positive semidefinite. This leads to the following equivalent problem of (26):

$$\begin{array}{ll} \min & \frac{1}{2}\|X_3 - \widehat{C}_3\|^2 \\ \text{s.t.} & (X_3)_{ii} = 1,\ i = 1, \ldots, n-m\,, \\ & (X_3 - C_2^T C_1^{-1} C_2) \in \mathcal{S}^{n-m}_+\,. \end{array} \tag{27}$$

Let $Y := X_3 - C_2^T C_1^{-1} C_2$. Then problem (27) becomes

$$\begin{array}{ll} \min & \frac{1}{2}\|Y - (\widehat{C}_3 - C_2^T C_1^{-1} C_2)\|^2 \\ \text{s.t.} & Y_{ii} = 1 - d_i,\ i = 1, \ldots, n-m\,, \\ & Y \in \mathcal{S}^{n-m}_+, \end{array} \tag{28}$$

where $d := \mathrm{diag}(C_2^T C_1^{-1} C_2)$. We have

Proposition 3.3 *It holds that* $d_i < 1, \quad i = 1, \ldots, n-m.$

Proof. From

$$C = \begin{pmatrix} C_1 & C_2 \\ C_2^T & C_3 \end{pmatrix} \succ 0\,,$$

we know that the Schur complement

$$C_3 - C_2^T C_1^{-1} C_2 \succ 0.$$

Thus, we have for each $i \in \{1, \ldots, n-m\}$ that

$$0 < (C_3 - C_2^T C_1^{-1} C_2)_{ii} = (C_3)_{ii} - (C_2^T C_1^{-1} C_2)_{ii} = 1 - d_i\,,$$

which completes the proof. □

Problem (28) is of the type of the nearest correlation matrix problem (8) with a reduced dimension of $(n-m)$ and the diagonal entries of Y being positive numbers instead of being all 1. This type of problem can be quite efficiently solved by Algorithm 3.1 as shown by Qi and Sun (2006). Note that for problem (28), the generalized Slater condition holds and that all elements in the corresponding set $\widehat{\partial}\theta(y^*)(=\partial\theta(y^*)$ in this case) at any solution y^* to the dual problem are positive definite (Qi and Sun 2006). Therefore, we can obtain that when Algorithm 3.1 is applied to the dual of problem (28), the infinitely generated sequence $\{y^k\}$ converges quadratically.

Once the optimal solution Y^* of problem (28) is found, the optimal solution X^* of the local correlation stress testing problem (26) can be recovered by

$$X^* := \begin{pmatrix} C_1 & C_2 \\ C_2^T & Y^* + C_2^T C_1^{-1} C_2 \end{pmatrix}.$$

3.3 The Band Correlation Stress Testing

Recall that for this case $(\mathcal{B}, c)$ is given by (3) and any replacement matrix X has to satisfy the following constraint:

$$X = \begin{pmatrix} C_1 & X_2 \\ X_2^T & X_3 \end{pmatrix} \succeq 0\,.$$

Therefore, the Schur complement decomposition technique cannot be applied to this constraint as X_2 is no longer fixed.

Let $\mathcal{A}_1 : \mathcal{S}^n \mapsto \mathbb{R}^n$ and $\mathcal{A}_2 : \mathcal{S}^n \mapsto \mathbb{R}^\kappa$ be defined by (6). Then, we have

$$\mathcal{A}_1^*(x) = \mathrm{Diag}(x) \;\; \text{for} \;\; x \in \mathbb{R}^n \;\; \text{and} \;\; \mathcal{A}_2^*(z) = \frac{1}{2}\sum_{(i,j)\in\mathcal{B}} z_{ij}\left(E^{ij} + E^{ji}\right) \;\; \text{for} \;\; z \in \mathbb{R}^\kappa. \tag{29}$$

Define $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^{n+\kappa}$ by

$$\mathcal{A}(X) := \begin{bmatrix} \mathcal{A}_1(X) \\ \mathcal{A}_2(X) \end{bmatrix}, \quad X \in \mathcal{S}^n\,. \tag{30}$$

Then, obviously, $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^{(n+\kappa)}$ is surjective. The adjoint of $\mathcal{A}$ takes the following form

$$\mathcal{A}^*(y) = \mathcal{A}_1^*(x) + \mathcal{A}_2^*(z)\,, \quad y \equiv (x,z) \in \mathbb{R}^n \times \mathbb{R}^\kappa\,. \tag{31}$$

For any $\tau \in [0,1)$, let $b_\tau = ((1-\tau)e^T,\ c^T)^T$. Then problem (10) can be written as

$$\begin{array}{ll} \min & \frac{1}{2}\|Y - (\widehat{C} - \tau I)\|^2 \\ \text{s.t.} & \mathcal{A}(y) = b_\tau\,, \\ & Y \in \mathcal{S}^n_+\,, \end{array} \tag{32}$$

which is a special case of problem (13). Here, we allow $\tau = 0$ in order to include problem (7).

For any $X \in \mathcal{S}^n$, denote $X_+ \equiv \Pi_{\mathcal{S}^n_+}(X)$. Thus, by Section 3.1, we know that the unconstrained dual problem of (32) turns to be

$$\begin{array}{ll} \min & \theta_\tau(y) \\ \text{s.t.} & y \in \mathbb{R}^{n+\kappa}\,, \end{array} \tag{33}$$

where

$$\theta_\tau(y) = \frac{1}{2}\|\big((\widehat{C} - \tau I) + \mathcal{A}^*(y)\big)_+\|^2 - b_\tau^T y - \frac{1}{2}\|\widehat{C} - \tau I\|^2\,, \quad y \in \mathbb{R}^{n+\kappa}\,.$$

Let $\tau \in [0, \lambda_{\min}(C))$. By using the facts that C is positive definite and $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^{(n+\kappa)}$ is onto, we know that the generalized Slater condition (20) for problem (32) holds. Thus, Algorithm 3.1 will generate a bounded sequence $\{y^k\}$ when it is applied to problem (33). Let y^* be an accumulation point of $\{y^k\}$. We shall next establish the quadratic convergence of $\{y^k\}$ by showing that all elements in $\widehat{\partial}\theta_\tau(y^*)$ are positive definite.

Proposition 3.4 *Assume that $\tau \in [0, \lambda_{\min}(C))$. Let $y^* \in \mathbb{R}^{n+\kappa}$ be an optimal solution to problem* (33)*. Then any element in $\widehat{\partial}^2\theta_\tau(y^*)$ is symmetric and positive definite.*

Proof. See Appendix B. □

Now we are ready to present our convergence result for solving (33).

Theorem 3.5 *Assume that $\tau \in [0, \lambda_{\min}(C))$. Let the sequence $\{y^k\}$ be generated by* Algorithm 3.1 *applied to the unconstrained dual convex optimization problem* (33)*. If both $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ in* Algorithm 3.1 *are uniformly bounded, then the whole sequence $\{y^k\}$ converges to the unique solution of* (33) *quadratically.*

Proof. First, by Theorem 3.2 and the fact that the generalized Slater's condition (20) for problem (32) holds, we know that $\{y^k\}$ is bounded, which implies that $\{y^k\}$ has at least one accumulation point, say y^*. This point also must be an optimal solution to problem (33). Therefore, from Proposition 3.4 and Theorem 3.2, we conclude that the whole sequence $\{y^k\}$ converges to y^* quadratically. □

We make a few remarks regarding using Algorithm 3.1.

(R1) The matrix $\widehat{C}$ can be any symmetric matrix. It may have nothing to do with the current correlation matrix C. The resulting optimal solution is just the nearest correlation matrix to $\widehat{C}$ satisfying the constraints involved.

(R2) For the Newton method to converge quadratically, C does not have to be positive definite. What we really need in the proof of, say, Proposition 3.4, is that the leading $m \times m$ principle submatrix of C is positive definite. The matrix C being positive definite is merely a nice sufficient condition. This observation allows us to consider more general situations. For example, $\mathcal{B}$ may take the "rectangular" form

$$\mathcal{B} := \{(i,j) \mid i = 1, \ldots, m,\ j = i+1, \ldots, m_1\}$$

for any $m_1 \geq m$. Such an index set $\mathcal{B}$ simply means that we fix the correlations contained in the rectangular submatrix $C_{\alpha\beta}$ with $\alpha := \{1, \ldots, m\}$ and $\beta := \{1, \ldots, m_1\}$.

The proof of Proposition 3.4 can readily be extended to cover the rectangular case. The proof, instead of using the positive definiteness of the leading $m \times m$ principal submatrix

of C, now uses the positive definiteness of all the $(m+1)\times(m+1)$ principal submatrices of the form $C_{\tilde{\alpha}\tilde{\alpha}}$ with

$$\tilde{\alpha} := \{1, 2, \ldots, m\} \cup \{j\} \quad \text{and} \quad m \le j \le m_1.$$

Furthermore, the fixed elements may not even form a block of submatrix (refer to the 5-factor example in the numerical experiment section). We may explore it further to state sufficient conditions ensuring the quadratic convergence for this case, but it would be more involved.

(R3) If the sequence $\{y^k\}$ generated by the Newton method converges to y^* quadratically, then the corresponding matrix sequence $\{Y^k\}$ defined by

$$Y^k = \big((\widehat{C} - \tau I) + \mathcal{A}^*(y^k)\big)_+$$

converges to the solution $Y^* = \big((\widehat{C} - \tau I) + \mathcal{A}^*(y^*)\big)_+$ and satisfies

$$\begin{aligned} \|Y^{k+1} - Y^*\| &= \|\big((\widehat{C} - \tau I) + \mathcal{A}^*(y^{k+1})\big)_+ - \big((\widehat{C} - \tau I) + \mathcal{A}^*(y^*)\big)_+\| \\ &\le \|\mathcal{A}^*\big(y^{k+1} - y^*\big)\| = O(\|y^k - y^*\|^2), \end{aligned}$$

where the non-expansion property of the projection operator has been used.

4 An Augmented Lagrangian Dual Approach

Note that the dimension of the unknown vector in the Lagrangian dual function $\theta_\tau(y)$ in problem (33) is $(n + \kappa(\mathcal{B}))$. When $\kappa(\mathcal{B})$ is large, the unconstrained convex optimization problem (33) is often costly to solve as it may need a large number of CG iterations to solve an $(n+\kappa(\mathcal{B}))\times(n+\kappa(\mathcal{B}))$ linear equation at each step even it is well-posed in theory. This computational consideration prompts us to study the augmented Lagrangian method, which attempts to solve a sequence of unconstrained convex optimization problems of lower dimensions.

Recall that problem (5) and its regularized version (9) can be uniformly written as

$$\begin{array}{ll} \min & \dfrac{1}{2}\|X - \widehat{C}\|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X_{ij} = c_{ij}, \quad (i,j) \in \mathcal{B}\,, \\ & X - \tau I \in \mathcal{S}^n_+\,, \end{array} \tag{34}$$

where $\tau \in [0, 1)$. Another way to look at problem (34), different from the Lagrangian dual approach introduced in the last section, is to first eliminate all the fixed correlations from the problem and then to solve the resulted problem of reduced dimension. This is of particular interest when the number of fixed correlations is much larger than the number of ones to be stressed.

This new way of looking at problem (34) leads to the development of the augmented Lagrangian dual approach, which goes along the following line: After eliminating the fixed correlations in problem (34), we have its equivalent version (35), whose Lagrangian dual (36) is

no longer unconstrained. To get an unconstrained convex problem, we study its augmented Lagrangian dual problem (43). The augmented Lagrangian method is then applied to this problem. To address the fast convergence of the method, we relate the iterate Y^{k+1} in (46) to the solution of the proximal-type problem (47) so that Rockafellar's classical result on the proximal method can be applied. This saves us from giving a complete convergence analysis from scratch. Below is the detailed account of the augmented Lagrangian dual approach.

Let $\mathcal{N}$ denote the indices of those correlations not in $\mathcal{B}$, i.e.,

$$\mathcal{N} := \{(i,j) \,|\, 1 \le i < j \le n\} \setminus \mathcal{B}\,.$$

Let $\bar{\kappa} := \bar{\kappa}(\mathcal{N})$ be the number of indices in $\mathcal{N}$. For the convenience of subsequent discussions, we introduce a linear operator $\overline{\mathcal{A}} : \mathcal{S}^n \mapsto \mathbb{R}^{\bar{\kappa}}$ by

$$\overline{\mathcal{A}}(Y) := x \;\; \text{with } x_{ij} = Y_{ij} \text{ for } (i,j) \in \mathcal{N}\,,$$

where, as in Section 2, for any $x \in \mathbb{R}^{\bar{\kappa}}$ we write $x \equiv (x_{ij})_{(i,j)\in\mathcal{N}}$ to match the structure of $\mathcal{N}$. Let $\overline{\mathcal{A}}^* : \mathbb{R}^{\bar{\kappa}} \mapsto \mathcal{S}^n$ be the adjoint of the linear operator $\overline{\mathcal{A}}$. Then for any $x \in \mathbb{R}^{\bar{\kappa}}$, the symmetric matrix $\overline{\mathcal{A}}^*(x)$ takes the following form

$$\big[\,\overline{\mathcal{A}}^*(x)\,\big]_{ij} = \begin{cases} 0 & \text{if } i = j\,, \\ 0 & \text{if } (i,j) \in \mathcal{B}\,, \\ \frac{1}{2}x_{ij} & \text{if } (i,j) \in \mathcal{N}\,, \end{cases} \qquad 1 \le i \le j \le n\,.$$

Denote the symmetric matrix $A_0 \in \mathcal{S}^n$ by

$$[\,A_0\,]_{ij} = \begin{cases} 1 & \text{if } i = j\,, \\ c_{ij} & \text{if } (i,j) \in \mathcal{B}\,, \\ 0 & \text{if } (i,j) \;\in \mathcal{N}\,, \end{cases} \qquad 1 \le i \le j \le n$$

and the vector $\hat{c} \in \mathbb{R}^{\bar{\kappa}}$ by

$$(\hat{c})_{ij} := \widehat{C}_{ij}, \quad (i,j) \in \mathcal{N}\,.$$

Then, in consideration of the fact that $\widehat{C}_{ij} = C_{ij}$ for all $(i,j) \in \mathcal{B}$ (i.e., the fixed correlations in the target matrix), problem (34) becomes

$$\begin{array}{ll} \min & \dfrac{1}{2}\|x - \hat{c}\|^2 \\ \text{s.t.} & A_\tau + \overline{\mathcal{A}}^*(x) \in \mathcal{S}^n_+\,, \end{array} \tag{35}$$

where we add "$\frac{1}{2}$" to the objective function for the sake of convenience and for any $\tau \in [0,1)$, we write $A_\tau \equiv A_0 - \tau I$.

Let $l_0 : \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n \to \mathbb{R}$ be the ordinary Lagrangian function for problem (35), i.e.,

$$l_0(x,Y) := \frac{1}{2}\|x - \hat{c}\|^2 - \big\langle Y, A_\tau + \overline{\mathcal{A}}^*(x)\big\rangle\,, \quad (x,Y) \in \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n\,.$$

Then, the Lagrangian dual of problem (35) is

$$\begin{array}{ll} \max & -\Big(\big\langle Y, A_\tau + \overline{\mathcal{A}}^*(\hat{c})\big\rangle + \dfrac{1}{2}\|\overline{\mathcal{A}}\,(Y)\|^2\Big) \\ \text{s.t.} & Y \in \mathcal{S}^n_+\,, \end{array} \tag{36}$$

which is no longer an unconstrained problem as problem (33). The Karush-Kukn-Tucker (KKT) conditions, i.e., the first order optimality conditions for problem(35), are

$$\begin{cases} \nabla_x l_0(x,Y) = x - \hat{c} - \overline{\mathcal{A}}(Y) = 0\,, \\ \mathcal{S}^n_+ \ni \big(A_\tau + \overline{\mathcal{A}}^*(x)\big) \perp Y \in \mathcal{S}^n_+\,, \end{cases} \tag{37}$$

where "$\big(A_\tau+\overline{\mathcal{A}}^*(x)\big) \perp Y$" means that the two matrices are orthogonal, i.e., $\big\langle A_\tau+\overline{\mathcal{A}}^*(x), Y\big\rangle = 0$. Any point $(x^*, Y^*) \in \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n$ satisfying (37) is called a KKT point. By using the fact that $\mathcal{S}^n_+$ is a self-dual cone, we know from Eaves (1971) that $(x^*, Y^*) \in \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n$ satisfying (37) if and only if (x^*, Y^*) is a solution to the following system of nonsmooth equations

$$F_\tau(x,Y) \equiv \begin{bmatrix} x - \hat{c} - \overline{\mathcal{A}}(Y) \\ Y - \big[\, Y - \big(A_\tau + \overline{\mathcal{A}}^*(x)\big)\,\big]_+ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad (x,Y) \in \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n\,.$$

Next, we shall study the existence (and uniqueness) of a KKT point $(x^*, Y^*) \in \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n$ and the local Lipschitz invertibilty of F_τ near (x^*, Y^*). Denote

$$g_\tau(x) \equiv A_\tau + \overline{\mathcal{A}}^*(x)\,, \quad x \in \mathbb{R}^{\bar{\kappa}}\,.$$

Assume that $\tau \in [0, \lambda_{\min}(C))$. Then, since the objective function in problem (35) is strongly convex and the feasible set is nonempty (e.g., C itself is a feasible solution), problem (35) has a unique optimal solution x^*. Furthermore,

$$X^* := g_\tau(x^*) + \tau I$$

is the unique optimal solution to (34), and hence

$$Y^* := g_\tau(x^*) \tag{38}$$

is the optimal solution of (10).

Let $\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))$ denote the tangent cone of $\mathcal{S}^n_+$ at $g_\tau(x^*)$ in the sense of convex analysis. We use $\mathrm{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))\big)$ to denote the largest linear space contained in $\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))$. Then, we have the following useful result.

Lemma 4.1 *Assume that $\tau \in [0, \lambda_{\min}(C))$. Then the following constraint nondegenerate condition holds at x^* :*

$$\overline{\mathcal{A}}^*(\mathbb{R}^{\bar{\kappa}}) + \mathrm{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))\big) = \mathcal{S}^n\,. \tag{39}$$

Proof. See Appendix C. □

Lemma 4.1 implies that there exists a unique $Y^* \in \mathcal{S}^n_+$ such that (x^*, Y^*) is the unique KKT point satisfying (37). See Bonnans and Shapiro (2000) for a general discussion on this. Then, we can obtain the local Lipschitz invertibility of F_τ near (x^*, Y^*).

Proposition 4.2 *Assume that $\tau \in [0, \lambda_{\min}(C))$. Then there exist a neighborhood $\mathcal{O}$ of (x^*, Y^*) in $\mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n$ and a constant $\nu > 0$ such that*

$$\|F_\tau(x,Y) - F_\tau(\tilde{x},\widetilde{Y})\| \geq \nu \|(x,Y) - (\tilde{x},\widetilde{Y})\| \quad \forall\, (x,Y) \text{ and } (\tilde{x},\widetilde{Y}) \in \mathcal{O}\,.$$

Proof. This follows directly from the strong convexity of the objective function in (35), Lemma 4.1, and Theorem 4.1 in Sun (2006). □

Now, we are ready to introduce the augmented Lagrangian dual approach for solving problem (35). Let $\sigma > 0$. The augmented Lagrangian function for problem (35) is

$$L_\sigma(x, Y) := \frac{1}{2}\|x - \hat{c}\|^2 + \frac{1}{2\sigma}\big\{\|\big(Y - \sigma g_\tau(x)\big)_+\|^2 - \|Y\|^2\big\}\,, \quad (x, Y) \in \mathbb{R}^{\bar{\kappa}} \times \mathcal{S}^n\,. \tag{40}$$

Strictly speaking, the augmented Lagrangian function L_σ should also depend on the prescribed constant τ. We drop this dependence as it can be seen clearly from the context. Here, we will omit the details on deriving this augmented Lagrangian function as excellent discussions on augmented Lagrangian functions for general optimization problems can be found easily in the literature, e.g., Section 11.K in Rockafellar and Wets (1998). For any $Y \in \mathcal{S}^n$, the augmented Lagrangian function $L_\sigma(\cdot, Y)$ is strongly convex in x and continuously differentiable with

$$\nabla_x\big(L_\sigma\big)(x, Y) = x - \hat{c} - \overline{\mathcal{A}}\,(Y - \sigma g_\tau(x))_+\,, \quad x \in \mathbb{R}^{\bar{\kappa}}\,. \tag{41}$$

For any $Y \in \mathbb{R}^{\bar{\kappa}}$, let $x_\sigma(Y)$ be the unique optimal solution to

$$\begin{array}{ll} \min & L_\sigma(x, Y) \\ \text{s.t.} & x \in \mathbb{R}^{\bar{\kappa}}\,. \end{array} \tag{42}$$

The augmented Lagrangian dual problem then takes the following form

$$\begin{array}{ll} \max & -\vartheta_\sigma(Y) \\ \text{s.t.} & Y \in \mathcal{S}^n\,, \end{array} \tag{43}$$

where $\vartheta_\sigma : \mathcal{S}^n \to \mathbb{R}$ is defined by

$$\vartheta_\sigma(Y) := -L_\sigma(x_\sigma(Y), Y)\,, \quad Y \in \mathcal{S}^n\,. \tag{44}$$

The function $\vartheta_\sigma(\cdot)$ is a continuously differentiable convex function in $\mathcal{S}^n$ with

$$\nabla(\vartheta_\sigma)(Y) := \sigma^{-1}\big[Y - (\,Y - \sigma g_\tau(x_\sigma(Y))\,)_+\big]\,, \quad Y \in \mathcal{S}^n\,. \tag{45}$$

Note that both $\vartheta_\sigma(Y)$ and $\nabla(\vartheta_\sigma)(Y)$ depend implicitly on the unique optimal solution $x_\sigma(Y)$ to problem (42). While for each $Y \in \mathcal{S}^n$, the computation of $x_\sigma(Y)$ can be obtained by applying the quadratically convergent Newton's method – Algorithm 3.1 to (42) directly, it is not clear immediately if Algorithm 3.1 can be applied to solve the augmented Lagrangian dual problem (43). However, recent research conducted by Sun, Sun, and Zhang (2006) reveals that when σ is sufficiently large, the augmented Lagrangian method in the context of general nonlinear semidefinite programming problems including problem (43) can be locally treated as an approximate version of Algorithm 3.1.

The augmented Lagrangian method for solving problem (43) can be stated as follows.

Algorithm 4.3 (An Augmented Lagrangian Method)

Step 0. *Given* $\sigma_0 > 0$. *Let* $x^0 \in \mathbb{R}^{\bar{\kappa}}$ *be arbitrary and* $Y^0 \in \mathcal{S}^n_+$ *be the initial estimated Lagrangian multiplier.* $k := 0$.

Step 1. *Let* $\theta_k(\cdot) \equiv L_\sigma(\cdot, Y^k)$. *Define*

$$\widehat{\partial}^2\theta_k(x) := I + \sigma\overline{\mathcal{A}}\,\partial\Pi_{\mathcal{S}^n_+}\big(Y^k - \sigma g_\tau(x)\big)\overline{\mathcal{A}}^*, \quad x \in \mathbb{R}^{\bar{\kappa}}.$$

Compute $x^{k+1} = x_{\sigma_k}(Y^k)$ *by applying* Algorithm 3.1 *to* problem (42) *for* $Y = Y^k$, *i.e.,*

$$\begin{array}{ll} \min & \theta_k(x) \\ \text{s.t.} & x \in \mathbb{R}^{\bar{\kappa}}, \end{array}$$

with the starting point x^k.

Step 2. *Compute* Y^{k+1} *by*

$$Y^{k+1} := \big(Y^k - \sigma_k g_\tau(x^{k+1})\big)_+ \tag{46}$$

and update σ_k *to* $\sigma_{k+1} \geq \sigma_k$.

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

Comparing (46) with (45), we can see that at the kth iteration of Algorithm 4.3:

$$Y^{k+1} = Y^k - \sigma_k\nabla(\vartheta_{\sigma_k})(Y^k) = Y^k - Y^k + \big(Y^k - \sigma_k g_\tau(x^{k+1})\big)_+ = \big(Y^k - \sigma_k g_\tau(x^{k+1})\big)_+,$$

which implies that Algorithm 4.3 is a gradient descent method of steplength σ_k at the kth iteration. This suggests that Algorithm 4.3 may converge very slowly. However, as we mentioned earlier, locally Algorithm 4.3 can be treated as a kind of approximate generalized Newton's method. So good convergence may still be expectable. In fact, from Proposition 6 in Rockafellar (1976b)[7], we know that Y^{k+1} is the unique optimal solution to the following "proximal-type" problem centered at Y^k :

$$\begin{array}{ll} \max & -\Big(\big\langle Y, A_\tau + \overline{\mathcal{A}}^*(\hat{c})\big\rangle + \dfrac{1}{2}\|\overline{\mathcal{A}}(Y)\|^2 + \dfrac{1}{2\sigma_k}\|Y - Y^k\|^2\Big) \\ \text{s.t.} & Y \in \mathcal{S}^n_+, \end{array} \tag{47}$$

which implies that Algorithm 4.3 is a proximal point algorithm applied to the Lagrangian dual problem (36). This connection allows us to use the convergence theory developed by Rockafellar (1976a, 1976b) for proximal point methods for maximal monotone operators.

Theorem 4.4 [8] *Assume that* $\tau \in [0, \lambda_{\min}(C))$. *Let* $\{Y^k\}$ *be a sequence generated by* Algorithm 4.3 *with* $\lim\limits_{k\to\infty}\sigma_k = \sigma_\infty < +\infty$. *Then,* $\{Y^k\}$ *converges to* Y^*. *Furthermore, let* $\nu > 0$ *be the constant obtained in* Proposition 4.2. *Then for all* k *sufficiently large,*

$$\|Y^{k+1} - Y^*\| \leq a_k\|Y^k - Y^*\|,$$

[7] One needs to slightly modify the proof given by Rockafellar for nonlinear convex programming to include the the problem discussed here.

[8] Theoretically speaking, in order to make Algorithm 4.3 practical, one should consider the case that $x^{k+1} \approx x_{\sigma_k}(Y^k)$ instead of $x^{k+1} = x_{\sigma_k}(Y^k)$. However, from computational point of view, this consideration is not of vital importance when the quadratically convergent Algorithm 3.1 is applied to solving problem (42).

where

$$a_k := \frac{\nu}{\sqrt{\nu^2 + \sigma_k^2}} \to a_\infty = \frac{\nu}{\sqrt{\nu^2 + \sigma_\infty^2}} < 1\,. \tag{48}$$

Proof. This is a direct application of Theorem 2 in Rockafellar (1976a) (also cf. Rockafellar 1976b, Theorem 5) and Proposition 4.2. □

We conclude this section by making the following remarks about Algorithm 4.3.

(R1) The computation of x^{k+1} for $Y = Y^k$ in Step 1 of Algorithm 4.3 can be obtained very rapidly by applying the quadratically convergent Algorithm 3.1 to problem (42) with the starting point x^k. This is particularly the case when x^k is not far away from the solution x^*. Note that there is no need to do one more spectral decomposition to get Y^{k+1} in Step 2 as, from (41), it has already been computed in Step 1 when x^{k+1} is computed.

(R2) From Theorem 4.4, (37), (41), and (46), we know that $\{x^k\}$ converges to x^* with

$$\|x^{k+1} - x^*\| = \|\overline{\mathcal{A}}\,(Y^{k+1} - Y^*)\| \le a_k \|Y^k - Y^*\|\,,$$

where a_k is defined in (48).

5 Numerical Experiments

In this section, we report our numerical experiments conducted for the correlation stress testing problem carried out in `MATLAB 7.1` running on a PC Intel Pentium IV of 2.40 GHz CPU.

Our first numerical experiment is to test some examples of small scales available in the literature. These examples include: a 7-factor example in Finger (1997), a 4-factor example in Turkay, Epperlein, and Christofides (2003), a 12-factor example in Rebonato and Jäckel (2000), and a 5-factor example in Bhansali and Wise (2001). For all these examples, Newton's method - Algorithm 3.1 found better solutions. Instead of listing all these numerical results, we just use the 5-factor example in Bhansali and Wise (2001) to illustrate the performance of Algorithm 3.1.

In the 5-factor example in Bhansali and Wise (2001), the target matrix $\widehat{C}$ is given by

$$\widehat{C} = \begin{pmatrix} 1 & \mathbf{-0.50} & -0.30 & \mathbf{-0.25} & \mathbf{-0.70} \\ & 1 & \mathbf{0.90} & 0.30 & 0.70 \\ & & 1 & 0.25 & 0.20 \\ & & & 1 & 0.75 \\ & & & & 1 \end{pmatrix}.$$

The confidence matrix of equation (7) in Bhansali and Wise (2001) indicates that the managers are highly confident that correlations in the boldface font, i.e., $\widehat{C}_{12}, \widehat{C}_{14}, \widehat{C}_{15}$, and $\widehat{C}_{23}$, are accurately estimated and want to keep those correlations unchanged. This means for this example

$$\mathcal{B} = \{(1,2), (1,4), (1,5), (2,3)\}\,.$$

Algorithm 3.1 (with $\tau = 0$ as the current correlation matrix is not available) found the following nearest correlation matrix

$$C^* = \begin{pmatrix} 1.0000 & -0.5000 & -0.2830 & -0.2500 & -0.7000 \\ & 1.0000 & 0.9000 & 0.3391 & 0.6134 \\ & & 1.0000 & 0.2179 & 0.2710 \\ & & & 1.0000 & 0.7198 \\ & & & & 1.0000 \end{pmatrix}.$$

The *sum of the squared difference* between the target matrix $\widehat{C}$ and the optimally calibrated correlation matrix C^* is $\|C^* - \widehat{C}\|^2 = 0.0326$, which is much smaller than the corresponding sum of the squared difference 1.0343 between $\widehat{C}$ and the correlation matrix found in equation (7) of Bhansali and Wise (2001). This also indicates that the hyperspherical decomposition method employed in Bhansali and Wise (2001) may not be able to produce an optimal correlation matrix.

Our next experiment is to test the capabilities of our algorithms on two relatively large scale examples. The first is a 387×387 correlation matrix case taken from the database of the RiskMetrics and the second one is randomly generated with $n = 1,000$ and $1,500$, respectively.

Example 5.1 *The current matrix C 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). The publicly available 25-day and the regulatory correlation matrices were also tested. The numerical performance*[9] *of our algorithms for the two matrices was similar to the 1-day matrix and was thus not reported here.*

Example 5.2 *The current matrix C is a randomly generated correlation matrix by using* MATLAB*'s built-in function* `randcorr`*. For the numerical testing purpose, the matrix C is deliberately generated to be ill-conditioned with very large and very small eigenvalues*[10]*.*

Note that the current correlation matrix C in examples 5.1 and 5.2 needs to be preprocessed in order to ensure that the calibrated optimal correlation matrix $X^* \geq \tau I$ (we take $\tau = 0.5 \times 10^{-4}$ in our tests, but obviously allow other choices to be specified by the users). Actually, the current matrix C in Example 5.1 contains small negative eigenvalues and the current matrix C in Example 5.2 is very close to be singular. For the local stress testing, we apply Algorithm 3.1 to the nearest correlation matrix optimization problem (12) to get an updated positive definite matrix $C \geq \tau_0 I$ (we take $\tau_0 = 1.0 \times 10^{-4}$ as τ_0 should be larger than τ). For the band stress testing, we apply Algorithm 3.1 to update only the top left m by m principal submatrix of C, i.e., C_1, to ensure that the smallest eigenvalue of the updated C_1 is at least τ_0.

[9] The stress test on the three matrices bears no physical meaning as to which correlations should be justifiably stressed. Our purpose here is solely to test the capability of our algorithms handling real and large correlation matrices.

[10] The following was used to generate the correlation matrix C: `n1 = round(n/10); d1 = rand(n1,1)/n; k =min(10, n-m); d3 = n*rand(k,1); d = rand(n,1); d(1:n1) =d1; d(n-k+1:n,1)=d3; d = n*d/sum(d); C =gallery('randcorr',d);`

In our numerical experiments for examples 5.1 and 5.2, the target correlation matrix $\widehat{C}$ is stressed in the following way:

$$\widehat{C}_{ij} := \begin{cases} C_{ij} & \text{if } (i,j) \in \mathcal{B}, \\ (1-0.1) \times C_{ij} + 0.1 \times G_{ij} & \text{if } (i,j) \notin \mathcal{B}, \end{cases}$$

where G is a randomly generated real symmetric matrix satisfying $G_{ij} \in [-1,1]$ for $i \neq j$ and $G_{ii} = 1$ for all i and j.

The initial parameters used in our numerical tests are more or less quite standard. More specifically,

• For Algorithm 3.1: $\eta = 10^{-5}$, $\mu = 10^{-4}$, and $\rho = 0.5$. The stopping criterion is $\|\nabla\theta(y^k)\| \leq 10^{-6}$. The starting point is $y^0 = 0$.

• For Algorithm 4.3: $\sigma_0 = 100$ and $\sigma_i = \min\{10\sigma_{i-1}, 10^5\}$ for $i \geq 1$. The stopping criterion is $\|F_\tau(x^{k+1}, Y^k)\| \leq 10^{-6}$ with $\tau = 0.5 \times 10^{-4}$. The starting point (x^0, Y^0) is obtained by applying Algorithm 3.1 to the nearest correlation matrix optimization problem (12).

We list our numerical results for Example 5.1 and Example 5.2 in Tables 1-4, where `It.`, `Func.`, and `Res.` stand for the number of total linear equations solved, the number of function evaluations, and the residual at the final iterate ($\|\nabla\theta(y^k)\|$ or $\|F_\tau(x^{k+1}, Y^k)\|$), respectively. Moreover, in Tables 1 and 3, `Test type` indicates the stress testing type: "Local" means the local stress testing and "Band" refers to the band stress testing. The `Levels` in Tables 2 and 4 indicates the number of calls to Algorithm 3.1 at Step 1 in Algorithm 4.3.

Test type	m	cputime	Iter.	Func.	Res.
Local	5	44 s	12	13	1.0×10^{-8}
Local	50	1 m 19 s	10	11	2.9×10^{-10}
Local	100	1 m 43 s	15	18	2.4×10^{-9}
Local	200	26 s	13	17	7.1×10^{-7}
Local	300	18 s	28	52	2.7×10^{-8}
Local	385	13 s	3	4	3.2×10^{-8}
Band	5	22 s	6	7	1.4×10^{-7}
Band	10	34 s	7	8	1.0×10^{-10}
Band	20	53 s	7	8	1.8×10^{-7}
Band	30	5 m 33 s	12	13	8.9×10^{-7}
Band	40	11 m 25 s	14	15	6.9×10^{-9}
Band	50	14 m 07 s	13	14	3.6×10^{-7}
Band	100	1 h 08 m 31 s	14	15	3.5×10^{-9}
Band	150	3 h 01 m 31 s	20	21	5.7×10^{-8}

Table 1: Numerical results of Example 5.1 for Algorithm 3.1

We make several observations about our numerical experiments in the following.

• The preprocessing step should be an integrated part of the stress testing as the current correlation matrix may even fail to be positive semidefinite due to insufficient historical data or

m	Levels	cputime	Iter.	Func.	Res.
386	5	1 m 53 s	21	31	2.2×10^{-8}
385	4	2 m 42 s	20	29	2.6×10^{-7}
384	4	4 m 05 s	21	30	8.4×10^{-7}
380	5	8 m 40 s	24	34	1.7×10^{-7}
377	6	15 m 24 s	31	43	1.3×10^{-7}
370	6	21 m 34 s	31	44	2.6×10^{-7}
350	9	1 h 15 m 46 s	51	75	7.3×10^{-7}
300	13	3 h 11 m 18 s	73	120	9.2×10^{-7}

Table 2: Numerical results of Example 5.1 for Algorithm 4.3.

numerical truncations. The quadratically convergent Newton's method - Algorithm 3.1 makes this step easy to implement.

• The local stress testing problem can be solved quite efficiently, regardless the size of m, by Algorithm 3.1.

• The band stress testing problem is more difficult to solve than the local stress testing problem when m becomes larger. Algorithm 3.1 should be used to solve the band stress testing problem with a relatively small m and Algorithm 4.3 is more favorable when m is close to n.

• The randomly generated problem is relatively easier to solve than the problem from the market data.

• The numerical results reported in Tables 1-4 indicate that our approach is highly efficient. For examples, in Table 3 for the band stress testing with $n = 1,500$ and $m = 500$, we need to solve at each step a linear equation of $125,250$ unknowns and in Table 4 for $n = 1,500$ and $m = 1,200$, we need to solve at each step a linear equation of $405,850$ unknowns.

6 Conclusion

In this paper, we developed an unconstrained convex optimization approach for the difficult correlation stress testing problem. The key for the success of our approach strongly hinges on the efficiency of the quadratically convergent Newton method - Algorithm 3.1, which makes use of recently developed theory of strongly semismooth matrix valued functions. We believe that the research conducted here can relieve practitioners from being troubled by lack of efficient numerical algorithms and allow them to pay more attention to their financial models. On the other hand, the methodology developed here is not confined to the correlation stress testing problem. It can certainly be used to deal with similar problems where an improper covariance/correlation matrix needs to be calibrated.

Test type	n	m	cputime	Iter.	Func.	Res.
Local	1,000	5	7 m 34 s	6	7	2.4×10^{-9}
Local	1,000	100	5 m 43 s	12	13	6.4×10^{-11}
Local	1,000	500	1 m 40 s	11	14	3.0×10^{-7}
Local	1,000	998	41 s	3	4	1.2×10^{-10}
Local	1,500	5	34 m 08 s	6	7	1.4×10^{-7}
Local	1,500	200	16 m 09 s	12	13	4.7×10^{-7}
Local	1,500	750	6 m 06 s	12	13	2.1×10^{-7}
Local	1,500	1,498	2 m 28 s	3	4	3.1×10^{-10}
Band	1,000	5	2 m 55 s	5	6	8.6×10^{-9}
Band	1,000	10	3 m 33 s	5	6	8.2×10^{-7}
Band	1,000	20	3 m 40 s	5	6	5.4×10^{-7}
Band	1,000	50	5 m 20 s	6	7	5.4×10^{-11}
Band	1,000	100	5 m 55 s	6	7	5.8×10^{-11}
Band	1,000	200	9 m 04 s	6	7	1.4×10^{-7}
Band	1,000	300	12 m 55 s	7	8	2.1×10^{-10}
Band	1,500	5	9 m 12 s	5	6	7.4×10^{-8}
Band	1,500	50	18 m 02 s	6	7	6.2×10^{-9}
Band	1,500	100	21 m 35 s	6	7	1.9×10^{-7}
Band	1,500	250	36 m 04 s	7	8	1.6×10^{-10}
Band	1,500	500	1 h 10 m 02 s	8	9	1.2×10^{-7}

Table 3: Numerical results of Example 5.2 for Algorithm 3.1.

Appendix A: Characterizing the set $\widehat{\partial}^2\theta(y)$

There are two issues that have to be addressed before employing Algorithm 3.1:

(i) Calculating one element $V_y \in \widehat{\partial}^2\theta(y)$ at an arbitrary point y; and

(ii) Characterizing the whole set $\widehat{\partial}^2\theta(y^*)$ at the solution point y^*.

The first issue is related to the practical implementation of Newton's method where in each step V_k is required; and the second issue is necessary to study the convergence rate of the method by assessing the nonsingularity of every element in $\widehat{\partial}^2\theta(y^*)$ (see Theorem 3.2). We address the two issues separately below.

(i) Calculating one element $V_y \in \widehat{\partial}^2\theta(y)$. Recall that $F(\cdot) = \nabla\theta(\cdot)$ is defined by (22), i.e.,

$$F(y) := \nabla\theta(y) = \mathcal{A}\,\Pi_{\mathcal{S}^p_+}(X^0 + \mathcal{A}^*y) - b\,, \qquad y \in \mathbb{R}^q\,.$$

For any given $X \in \mathcal{S}^p_+$, let $\lambda(X)$ be the eigenvalue vector of X with its components being arranged in the nonincreasing order, i.e., $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$. Let $\mathcal{O}$ denote the set

n	m	Levels	cputime	Iter.	Func.	Res.
$1,000$	999	3	11 m 13 s	12	20	2.8×10^{-8}
$1,000$	995	3	24 m 24 s	24	34	1.3×10^{-8}
$1,000$	990	3	44 m 10 s	15	34	1.9×10^{-7}
$1,000$	900	4	1 h 51 m 26 s	20	47	1.4×10^{-8}
$1,000$	700	4	3 h 32 m 30 s	25	68	2.8×10^{-8}
$1,500$	$1,499$	3	32 m 30 s	11	22	5.8×10^{-9}
$1,500$	$1,490$	3	1 h 00 m 58 s	15	31	5.7×10^{-7}
$1,500$	$1,450$	3	2 h 05 m 59 s	17	35	9.8×10^{-7}
$1,500$	$1,400$	4	3 h 38 m 04 s	19	39	1.8×10^{-8}
$1,500$	$1,200$	4	10 h 44 m 27 s	25	60	3.3×10^{-8}

Table 4: Numerical results of Example 5.2 for Algorithm 4.3.

of all orthogonal matrices in $\mathbb{R}^{p\times p}$ and $\mathcal{O}_X$ be the set of orthornormal eigenvectors of X defined by

$$\mathcal{O}_X := \left\{P \in \mathcal{O} |\ X = P\mathrm{Diag}[\lambda(X)]P^T\right\}.$$

For simplicity, let

$$X(y) := X^0 + \mathcal{A}^*(y), \qquad \lambda(y) := \lambda(X(y)).$$

We further have the spectral decomposition

$$X(y) = P\mathrm{Diag}(\lambda(y))P^T, \quad P \in \mathcal{O}_{X(y)}.$$

Define three sets of indices associated with $\lambda(y)$ by

$$\begin{array}{rcl} \alpha(y) & := & \{i|\lambda_i(y) > 0\}\ , \\ \beta(y) & := & \{i|\lambda_i(y) = 0\}\ , \\ \gamma(y) & := & \{i|\lambda_i(y) < 0\}\ . \end{array}$$

When the dependence of those sets on y is clear from the context, y is often omitted for simplicity. Let $W_y : \mathcal{S}^p \to \mathcal{S}^p$ be defined by

$$W_y H = P(M_y \circ (P^T H P))P^T \quad \forall\, H \in \mathcal{S}^p\,, \tag{49}$$

where

$$M_y := \begin{pmatrix} E_{\alpha\alpha} & E_{\alpha\beta} & (\nu_{ij})_{\substack{i\in\alpha \\ j\in\gamma}} \\ E_{\beta\alpha} & 0 & 0 \\ (\nu_{ji})_{\substack{i\in\alpha \\ j\in\gamma}} & 0 & 0 \end{pmatrix}, \qquad \nu_{ij} := \frac{\lambda_i(y)}{\lambda_i(y) - \lambda_j(y)}, \quad i \in \alpha, j \in \gamma.$$

Since $W_y \in \partial_B \Pi_{\mathcal{S}^p_+}(X(y))$ (cf. Pang, Sun, and Sun 2003), we obtain that

$$V_y := \mathcal{A}\, M_y \mathcal{A}^* \in \widehat{\partial}^2\theta(y)\,.$$

Then, for any $h \in \mathbb{R}^q$ we have

$$V_y h = \mathcal{A}\left(P(M_y \circ (P^T \mathcal{A}^*(h) P) P^T\right). \tag{50}$$

Note that there is no need to form the matrix V_y explicitly for our numerical implementation.

(ii) Characterization of the whole set $\widehat{\partial}^2\theta(y)$. Define a set of symmetric matrices at y by

$$\mathcal{M}_y := \left\{ M \in \mathcal{S}^p \mid \; M = \begin{pmatrix} E_{\alpha\alpha} & E_{\alpha\beta} & (\nu_{ij})_{\substack{i\in\alpha \\ j\in\gamma}} \\ E_{\beta\alpha} & (\omega_{ij})_{\substack{i\in\beta \\ j\in\beta}} & 0 \\ (\nu_{ji})_{\substack{i\in\alpha \\ j\in\gamma}} & 0 & 0 \end{pmatrix}, \begin{array}{l} \omega_{ij} = \omega_{ji} \in [0,1] \\ \text{for } i,j \in \beta \\ \nu_{ij} = \lambda_i(y)/(\lambda_i(y) - \lambda_j(y)) \\ \text{for } i \in \alpha, j \in \gamma \end{array} \right\}.$$

We note that $\mathcal{M}_y$ is a compact set and $1 > \nu_{ij} > 0$ for any $M \in \mathcal{M}_y$. Then, from Chen, Qi, and Teng (2003) or Pang, Sun, and Sun (2003), we obtain that

$$\widehat{\partial}^2\theta(y)h \subseteq \text{conv}\,\{\mathcal{A}(W(\mathcal{A}^*(h)))| \;\; W \in \mathcal{W}_y\} \quad \text{for any } h \in \mathbb{R}^q, \tag{51}$$

where $\mathcal{W}_y$ consists of all $W : \mathcal{S}^p \to \mathcal{S}^p$ such that

$$W(H) = P\left(M \circ (P^T H P)\right) P^T \quad \forall\, H \in \mathcal{S}^p$$

for some $P \in \mathcal{O}_{X(y)}$ and $M \in \mathcal{M}_y$.

We now specify the two characterizations to the problems we have encountered in the Lagrangian dual approach and the augmented Lagrangian dual approach respectively.

(iii) Specialization to problem (33). In the Lagrangian dual approach, θ takes the form θ_τ in (33). The corresponding F is given by

$$F(y) = \mathcal{A}\left((\widehat{C} - \tau I) + \mathcal{A}^*(y)\right)_+ - b_\tau,$$

where

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{pmatrix} \;\text{ and }\; \mathcal{A}^*(y) = \mathcal{A}_1^*(x) + \mathcal{A}_2^*(z) \;\text{ for } y = (x,z) \in \mathbb{R}^n \times \mathbb{R}^\kappa.$$

V_y can be characterized via the formula (50) with $\lambda(y)$ being the eigenvalues of the matrix $\left((\widehat{C} - \tau I) + \mathcal{A}^*(y)\right)$ of having the spectral decomposition:

$$((\widehat{C} - \tau I) + \mathcal{A}^*(y)) = P\text{Diag}(\lambda(y))P, \quad P \in \mathcal{O}_{((\widehat{C}-\tau I)+\mathcal{A}^*(y))}.$$

In the formula, α, β, and γ are respectively the set of indices of positive, zero, and negative eigenvalues of $((\widehat{C} - \tau I) + \mathcal{A}^*(y))$. Moreover, $\widehat{\partial}^2\theta_\tau(y)$ is also characterized in using the same spectral decomposition.

(iv) Specialization to problem (42). In the augmented Lagrangian dual approach (see Algorithm 4.3), when Algorithm 3.1 was applied to problem (42), θ takes the form $L_\sigma(x, Y)$ defined in (40) with x being a variable and Y being fixed. In this case (see also (41))

$$\nabla\theta(x) = \nabla_x L_\sigma(x, Y) = x - \hat{c} - \overline{\mathcal{A}}\big(Y - \sigma g_\tau(x)\big)_+, \quad x \in \mathbb{R}^{\bar{\kappa}}.$$

As we just did for calculating V_y, we can calculate a matrix $V_x \in \widehat{\partial}^2\theta(x)$ by the following formula

$$V_x h = h + \sigma\overline{\mathcal{A}}\left(P(M_x \circ (P^T\overline{\mathcal{A}}^*(h)P)P^T\right), \quad h \in \mathbb{R}^{\bar{\kappa}} \tag{52}$$

with

$$M_x := \begin{pmatrix} E_{\alpha\alpha} & E_{\alpha\beta} & (\nu_{ij})_{\substack{i\in\alpha\\ j\in\gamma}} \\ E_{\beta\alpha} & 0 & 0 \\ (\nu_{ji})_{\substack{i\in\alpha\\ j\in\gamma}} & 0 & 0 \end{pmatrix}, \qquad \nu_{ij} := \frac{\lambda_i(x)}{\lambda_i(x) - \lambda_j(x)}, \quad i \in \alpha, j \in \gamma,$$

where $\lambda(x)$ is the eigenvalue vector of the matrix $\big(Y - \sigma g_\tau(x)\big)$ of having the spectral decomposition:

$$Y - \sigma g_\tau(x) = P\mathrm{Diag}(\lambda(x))P, \quad P \in \mathcal{O}_{(Y-\sigma g_\tau(x))}.$$

In the formula, α, β, and γ are respectively the set of indices of positive, zero, and negative eigenvalues of $(Y - \sigma g_\tau(x))$.

Notice that in this case, due to the form of $\nabla\theta(x)$, any matrix in $\widehat{\partial}^2\theta(x)$ has two parts with the first being the identity matrix I and the second being a positive semidefinite matrix (e.g., see V_x). Therefore, any matrix in $\widehat{\partial}^2\theta(x)$ is always positive definite. Consequently, Algorithm 3.1 is always quadratically convergent when applied to problem (42).

Appendix B: Proof of Proposition 3.4

We first need a technical lemma.

Lemma 6.1 *Let Y^* be the optimal solution of problem* (10) *so that $X^* = Y^* + \tau I$ is the optimal solution of problem (9). Suppose Y^* has r positive eigenvalues and has the following spectral decomposition:*

$$Y^* = PDiag(\lambda_1^*, \ldots, \lambda_r^*, 0 \ldots, 0)P, \quad P \in \mathcal{O}_{Y^*}.$$

Let $P = [P_1, P_2]$, where $P_1 \in \mathbb{R}^{n\times r}$ and $P_2 \in \mathbb{R}^{n\times(n-r)}$ respectively denote the first r columns and the last $(n-r)$ columns of P. For any matrix $Z \in \mathcal{S}^n$ of the form

$$Z = \begin{pmatrix} Z_1 & 0 \\ 0 & Diag(x_{m+1}, \ldots, x_n) \end{pmatrix}, \quad Z_1 \in \mathcal{S}^m, \ x_{m+1}, \ldots, x_n \in \mathbb{R} \tag{53}$$

satisfying the condition

$$P_1^T Z = 0,$$

we have $Z = 0$.

Proof. The explicit form of Y^* in terms of the eigenvectors in P and positive eigenvalues $\lambda_1^*, \ldots, \lambda_r^*$ is

$$Y^* = \Big(\quad Y_{ij}^* = \sum_{\ell=1}^{r} (\lambda_\ell^* P_{i\ell} P_{j\ell}) \quad \Big)_{i,j=1,\ldots,n.} \tag{54}$$

The consequences of this explicit form, together with the fact that Y^* satisfies the constraints in (10), are

$$Y_{ij}^* = \sum_{\ell=1}^{r} (\lambda_\ell^* P_{i\ell} P_{j\ell}) = C_{ij}, \quad \forall (i,j) \in \mathcal{B} \tag{55}$$

and

$$Y_{ii}^* = \sum_{\ell=1}^{r} (\lambda_\ell^* P_{i\ell}^2) = 1 - \tau, \quad \forall\ i = 1, \ldots, n. \tag{56}$$

Now suppose Z takes the form (53) and satisfies

$$\Omega := P_1^T Z = P_1^T \begin{pmatrix} Z_1 & 0 \\ 0 & \mathrm{Diag}(x_{m+1}, \ldots, x_n) \end{pmatrix} = 0.$$

We prove $Z = 0$ in the following two cases.

Case 1. $j > m$. For this case we calculate

$$0 = \Omega_{ij} = x_j P_{ji} = x_j P_{ji}^2, \quad i = 1, \ldots, r.$$

Multiplying λ_i^* with Ω_{ij} and summing over $i = 1, \ldots, r$, we have

$$\begin{aligned} 0 &= \sum_{i=1}^{r} (\lambda_i^* \Omega_{ij}) \\ &= x_j \sum_{i=1}^{r} \left(\lambda_i^* P_{ji}^2\right) \\ &= x_j (Y_{jj}^*)^2 \qquad \text{(by (56))} \\ &= (1-\tau) x_j \qquad \text{(notice } 0 \le \tau < 1\text{)}. \end{aligned}$$

This proves $x_j = 0$ for $j = m+1, \ldots, n$.

Case 2. $j \le m$. For this case

$$0 = \Omega_{ij} = \sum_{\ell=1}^{m} P_{\ell i} Z_{\ell j}, \quad i = 1, \ldots, r.$$

Therefore,

$$\lambda_i^* \Omega_{ij}^2 = \sum_{\ell=1}^{m} \lambda_i^* Z_{\ell j}^2 P_{\ell i}^2 + 2 \sum_{\ell=1}^{m-1} \sum_{k=\ell+1}^{m} \lambda_i^* P_{ki} P_{\ell i} Z_{kj} Z_{\ell j}.$$

Summing over the index $i = 1, \ldots, r$ gives

$$\begin{aligned}
0 = \sum_{i=1}^{r} \lambda_i^* \Omega_{ij}^2 &= \sum_{\ell=1}^{m} \left[Z_{\ell j}^2 \left(\sum_{i=1}^{r} \left(\lambda_i^* P_{\ell i}^2 \right) \right) \right] + 2 \sum_{\ell=1}^{m-1} \sum_{k=\ell+1}^{m} \left[Z_{kj} Z_{\ell j} \left(\sum_{i=1}^{r} \left(\lambda_i^* P_{ki} P_{\ell i} \right) \right) \right] \\
&= \sum_{\ell=1}^{m} \left[Z_{\ell j}^2 Y_{\ell\ell}^* \right] + 2 \sum_{\ell=1}^{m-1} \sum_{k=\ell+1}^{m} \left[Z_{kj} Z_{\ell j} Y_{k\ell}^* \right] \qquad \text{(by (54))} \\
&= \sum_{\ell=1}^{m} (1-\tau) Z_{\ell j}^2 + 2 \sum_{\ell=1}^{m-1} \sum_{k=\ell+1}^{m} \left[Z_{kj} Z_{\ell j} (C_1)_{k\ell} \right] \qquad \text{(by (3), (55) and (56))} \\
&= Z_{\cdot j}^T (C_1 - \tau I) Z_{\cdot j}, \qquad \text{(by } (C_1)_{\ell\ell} = 1 - \tau \text{ for } \ell = 1, \ldots, m)
\end{aligned}$$

where $Z_{\cdot j}$ denotes the jth column of Z_1. Because $0 \leq \tau \lambda_{\min}(C)$, we have $\tau < \lambda_{\min}(C_1)$, which implies $(C_1 - \tau I)$ is positive definite. Therefore, $Z_{\cdot j} = 0$ for any $j \leq m$. This proves $Z_1 = 0$.

Putting Case 1 and Case 2 together we have proved $Z = 0$. This finishes the proof. □

Proof of Proposition 3.4. Recall y^* is an optimal solution of (33). Denote

$$C^* = \left(\widehat{C} - \tau I \right) + \mathcal{A}^*(y^*).$$

Suppose C^* has the spectral decomposition

$$C^* = P \mathrm{Diag}(\lambda_1^*, \ldots, \lambda_r^*, \lambda_{r+1}^*, \ldots, \lambda_n^*) P^T, \quad P \in \mathcal{O}_{C^*}$$

and suppose it has r positive eigenvalues $\lambda_1^*, \ldots, \lambda_r^*$. Define

$$\begin{aligned}
\alpha^* &:= \left\{ i \;\middle|\; \lambda_i^* > 0, \quad i = 1, \ldots, n \right\} = \{1, \ldots, r\}, \\
\beta^* &:= \left\{ i \;\middle|\; \lambda_i^* = 0, \quad i = 1, \ldots, n \right\}, \\
\gamma^* &:= \left\{ i \;\middle|\; \lambda_i^* < 0, \quad i = 1, \ldots, n \right\}.
\end{aligned}$$

It follows from the general formula (21) that the optimal solution Y^* of problem (10) is given by

$$Y^* = P \mathrm{Diag}(\lambda_1^*, \ldots, \lambda_r^*, 0 \ldots, 0) P^T, \quad P \in \mathcal{O}_{Y^*}.$$

Recall that $(\mathcal{B}, c)$ is given by (3). For a given point $y = (x, z) \in \mathbb{R}^n \times \mathbb{R}^\kappa$, it follows (29) and (31) of calculating $\mathcal{A}_1^*$, $\mathcal{A}_2^*$ and $\mathcal{A}^*$ that

$$\mathcal{A}^*(y) = \mathcal{A}_1^*(x) + \mathcal{A}_2^*(z) := Z = \begin{pmatrix} Z_1 & 0 \\ 0 & \mathrm{Diag}(x_{m+1}, \ldots, x_n) \end{pmatrix}, \tag{57}$$

where Z_1 is defined by

$$(Z_1)_{ij} := \begin{cases} 0.5 z_{ij} & \text{if } j > i \\ 0.5 z_{ji} & \text{if } j < i \\ x_i & \text{if } j = i \end{cases} \qquad i, j = 1, \ldots, m.$$

Suppose that V is an arbitrary element in $\widehat{\partial}^2\theta_\tau(y^*))$. We need to prove V is nonsingular. Suppose there exists $y = (x, z) \in \mathbb{R}^{n+\kappa}$ such that

$$V(y) = 0. \tag{58}$$

We need to show $y = 0$ in order to establish the nonsingularity of V. Now we suppose that (58) holds. It then follows from the inclusion relation (51) (see (iii) in Appendix A for the application of (51) to function θ_τ in (33)) that there exists a matrix $M \in \mathcal{M}_{y^*}$ such that

$$V(y) = \mathcal{A}\left(P(M \circ (P^T H P))P^T\right) \quad \text{with} \quad H := \mathcal{A}_1^*(x) + \mathcal{A}_2^*(z), \tag{59}$$

and M is given by

$$M = \begin{pmatrix} E_{\alpha^*\alpha^*} & E_{\alpha^*\beta^*} & (\nu_{ij})_{\substack{i\in\alpha^* \\ j\in\gamma^*}} \\ E_{\beta^*\alpha^*} & (\omega_{ij})_{\substack{i\in\beta^* \\ j\in\beta^*}} & 0 \\ (\nu_{ji})_{\substack{i\in\alpha^* \\ j\in\gamma^*}} & 0 & 0 \end{pmatrix}, \quad \begin{array}{l} \text{for some } \omega_{ij} = \omega_{ji} \in [0,1], \quad \text{for } i,j \in \beta \\ \nu_{ij} = \lambda_i^*/(\lambda_i^* - \lambda_j^*) \\ \text{for } i \in \alpha^*, j \in \gamma^*. \end{array}$$

Note that the matrix P in the above equation is the one that yields the spectral decomposition of $(\widehat{C} - \tau I) + \mathcal{A}^*(y^*)$. We then have

$$\begin{aligned} \langle y, V(y)\rangle &= \langle y, \mathcal{A}\left(P(M \circ (P^T H P))P^T\right)\rangle \\ &= \langle \mathcal{A}^*(y), P(M \circ (P^T H P))P^T\rangle \qquad \text{(by definition of the adjoint)} \\ &= \langle P^T(\mathcal{A}^*(y))P, M \circ (P^T H P)\rangle \qquad \text{(by the property of trace)} \\ &= \langle P^T H P, M \circ (P^T H P)\rangle \qquad \text{(by } \mathcal{A}^*(y) = \mathcal{A}_1^*(x) + \mathcal{A}_2^*(z)). \end{aligned}$$

Let $\tilde{H} := P^T H P$. Then by the Hardmard product we have

$$\begin{aligned} \langle y, V(y)\rangle &= \langle \tilde{H}, M \circ \tilde{H}\rangle \\ &\geq \sum_{i\in\alpha^*}\Big(\sum_{j\in\alpha^*\cup\beta^*} \tilde{H}_{ij}^2 + \sum_{j\in\gamma^*} \nu_{ij}\tilde{H}_{ij}^2\Big) \\ &\geq \nu \sum_{i=1}^{r}\sum_{j=1}^{n} \tilde{H}_{ij}^2, \end{aligned}$$

where $\nu := \min_{i\in\alpha^*, j\in\gamma^*} \nu_{ij} > 0$.

By assumption (58) we have

$$\tilde{H}_{ij} = 0 \qquad \forall i = 1, \ldots, r \text{ and } j = 1, \ldots, n.$$

This means that the first r rows of $\tilde{H}$ are zero, which, by recalling $P = [P_1, P_2]$, implies

$$P_1^T H P = 0. \tag{60}$$

Since P is nonsingular, this equation, together with (57) and (59), means

$$\Omega := P_1^T H = P_1^T(\mathcal{A}_1^*(x) + \mathcal{A}_2^*(z)) = P_1^T \begin{pmatrix} Z_1 & 0 \\ 0 & \mathrm{Diag}(x_{m+1}, \ldots, x_n) \end{pmatrix} = 0. \tag{61}$$

Now the result in Lemma 6.1 implies that $Z_1 = 0$ and $x_j = 0$ for all $j = m+1, \ldots, n$. By the way Z is defined, we know that $x = (y, z) = 0$. Hence V is nonsingular. The proof is completed. □

Appendix C: Proof of Lemma 4.1

Proof. Recall that $g_\tau(x^*)$ is positive semidefinite. We assume that it has r positive eigenvalues and it has the following spectral decomposition:

$$g_\tau(x^*) = P\mathrm{Diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)P^T, \quad P \in \mathcal{O}_{g_\tau(x^*)}.$$

It is also known from (38) that $Y^* = g_\tau(x^*)$ is the optimal solution of (10). Therefore, Lemma 6.1 applies to the eigenvector matrix P in the above decomposition.

The largest linear space $\mathrm{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))\big)$ contained in the tangent cone of $\mathcal{S}^n_+$ at $g_\tau(x^*)$ is given by (cf. Arnold 1971):

$$\mathrm{lin}\ \Big(\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))\Big) := \left\{ P \begin{pmatrix} U & V \\ V^T & 0 \end{pmatrix} P^T : \ \ U \in \mathcal{S}^r,\ V \in \mathbb{R}^{r\times(n-r)} \right\}.$$

Our first observation is that the set $\overline{\mathcal{A}}^*(\mathbb{R}^{\bar\kappa})$ has the following characterization:

$$\overline{\mathcal{A}}^*(\mathbb{R}^{\bar\kappa}) = \left\{ X \in \mathcal{S}^n \ \Big|\ \begin{array}{l} \langle E^{ij} + E^{ji}, X\rangle = 0, \ \ \text{for all } (i,j) \in \mathcal{B} \\ \langle E^{ii}, X\rangle = 0, \ \ i = 1, \ldots, n \end{array} \right\}.$$

That is, $\overline{\mathcal{A}}^*(\mathbb{R}^{\bar\kappa})$ is the intersection of the null spaces of E^{ij}, $((i,j) \in \mathcal{B})$ and E^{ii}, $i = 1, \ldots, n$ under the standard trace inner product.

Using this observation and the structure of $\mathrm{lin}\big(\mathcal{T}_{\mathcal{S}^n_+}(g_\tau(x^*))\big)$, it follows from a result (Theorem 1, Alizadeh, Haeberly, and Overton 1997) that the constraint nondegeneracy (i.e., the primal nondegeneracy in Alizadeh, Haeberly, and Overton 1997) holds if and only if the matrices

$$B^{ij} := \begin{pmatrix} P_1^T A^{ij} P_1 & P_1^T A^{ij} P_2 \\ P_2^T A^{ij} P_1 & 0 \end{pmatrix}, \ \begin{array}{l} A^{ij} = 0.5(E^{ij} + E^{ji}), \\ \text{for } (i,j) \in \mathcal{B} \text{ and } i = j = 1, \ldots, n. \end{array}$$

are linearly independent.

We now prove the linear independence of those B^{ij}. Suppose there exist $y \in \mathbb{R}^n$ and $z \in \mathbb{R}^\kappa$ (recall $\kappa = \kappa(\mathcal{B})$) such that

$$\sum_{i=1}^n y_i B^{ii} + \sum_{(i,j)\in\mathcal{B}} z_{ij} B^{ij} = 0. \tag{62}$$

We want to prove $y = 0$ and $z = 0$. Taking into account of the special structure of E^{ij}'s, it is easy to see equation (62) is equivalent to

$$\begin{pmatrix} P_1^T \mathrm{Diag}(y) P_1 & P_1^T \mathrm{Diag}(y) P_2 \\ P_2^T \mathrm{Diag}(y) P_1 & 0 \end{pmatrix} + \begin{pmatrix} P_1^T \Xi P_1 & P_1^T \Xi P_2 \\ P_2^T \Xi P_1 & 0 \end{pmatrix} = 0,$$

where

$$\Xi := 0.5 \sum_{(i,j)\in\mathcal{B}} (z_{ij} E^{ij} + z_{ij} E^{ji}).$$

Due to the symmetry, the above equation is equivalent to the first r-rows being zeros, i.e.,

$$P_1^T \mathrm{Diag}(y) P + P_1^T \Xi P = P_1^T (\mathrm{Diag}(y) + \Xi) = 0.$$

Note that the matrix $(\mathrm{Diag}(y) + \Xi)$ has the exact structure of the matrix Z defined in (53) when $(\mathcal{B}, c)$ is given by (3). Then Lemma 6.1 implies $y = 0$ and $\Xi = 0$, which in turn implies $z = 0$. This proves that the constraint nondegeneracy holds. □

As discussed in Remark R2 in Section 3, the constraint nondegeneracy property can be extended to other general cases including the rectangular $\mathcal{B}$. Consequently, the augmented Lagrangian method applies to stress testing in those general cases.

References

[1] C. Alexander, *Market Models: A Guide to Financial Data Analysis*. John Wiley & Sons, Ltd (2001).

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

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

[4] Z.-J. Bai, D. Chu, and D. Sun, *A dual optimization approach to inverse quadratic eigenvalue problems with partial eigenstructure*, Accepted by *SIAM J. Scientific Computing* in 2007.

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

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

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

[8] C. Brooks, B. Scott-Quinn, and J. Whalmsey, *Adjusting VaR model for impact of the Euro*, Working paper of ISMA Center, 1998.

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

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

[11] J.W. Dash, *Quantitative Finance and Risk Management: A Physicist's Approach*, World Scientific, 2004.

[12] B.C. Eaves, *On the basic theorem of complementarity*, Mathematical Programming 1 (1971), pp. 68–75.

[13] I. Fender, M.S. Gibson, and P.C. Mosser, *An international survey of stress tests*, Federal Reserve Bank of New York, Current Issues in Economics and Finance, Vol. 7, No 10 (2001).

[14] C. Finger, *A methodoloy for stress correlation*, In Risk Metrics Monitor, Fourth Quarter 1997.

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

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

[17] A.N. Kercheval, *A note on Rebonato and Jäckel's parametrization method for finding nearest correlation matrices*, Department of Mathematics, Florida Stata University, 2006.

[18] P.H. Kupiec, *Stress testing in a Value-at-Risk framework*, Journal of Derivatives 6(1), (1998), pp. 7–24.

[19] A. León, J.E. Peris, J. Silva, and B. Subiza, *A note on adjusting correlation matrices*, Applied Mathematics Finance 9 (2002), pp. 61–67.

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

[21] J.P. Morgan/Reuters, RiskMetrics$^{\text{TM}}$ – Techical Document. Forth Edition 1996, New York.

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

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

[24] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research 18 (1993), pp. 227–244.

[25] F. Rapisarda, D. Brigo, and F. Mercurio, *Parameterization correlations: a geometric interpretation*, Working paper, Product and Business Developement Group, Banca IMI, 2002.

[26] R. Rebonato and P. Jäckel, *The most general methodology for creating a valid correlation matrix for risk management and option pricing purpose*, Journal of Risk 2(2) (2000), pp. 17-27.

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

[28] R.T. Rockafellar, *Monotone Operators and the proximal point algorithm*, SIAM J. Control and Optimization 14 (1976), pp. 877–898.

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

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

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

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

[33] D. Sun, J. Sun, and L. Zhang, *The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming*, Technical Report, Department of Mathematics, National University of Singapore, Revised in October 2006.

[34] 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*, Accepted by Pacific J. Optimization in 2006.

[35] S. Turkay, E. Epperlein, and N. Christofides, *Correlation stress testing for value-at-risk*, Journal of Risk 5(4) (2003), pp. 75–89.

[36] E.H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory I and II.* In E.H. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, Academic Press (New York, 1971) pp. 237–424.