Robust Least Square Semidefinite Programming with Applications

Guoyin Li [*] Alfred Ka Chun Ma [†] Ting Kei Pong [‡]

July 13, 2013

Abstract

In this paper, we consider a least square semidefinite programming problem under ellipsoidal data uncertainty. We show that the robustification of this uncertain problem can be reformulated as a semidefinite linear programming problem with an additional second-order cone constraint. We then provide an explicit quantitative sensitivity analysis on how the solution under the robustification depends on the size/shape of the ellipsoidal data uncertainty set. Next, we prove that, under suitable constraint qualifications, the reformulation has zero duality gap with its dual problem, even when the primal problem itself is infeasible. The dual problem is equivalent to minimizing a smooth objective function over the Cartesian product of second-order cones and the Euclidean space, which is easy to project onto. Thus, we propose a simple variant of the spectral projected gradient method [7] to solve the dual problem. While it is well-known that any accumulation point of the sequence generated from the algorithm is a dual optimal solution, we show in addition that the dual objective value along the sequence generated converges to a finite value if and only if the primal problem is feasible, again under suitable constraint qualifications. This latter fact leads to a simple certificate for primal infeasibility in situations when the primal feasible set lies in a known compact set. As an application, we consider *robust* correlation stress testing where data uncertainty arises due to untimely recording of portfolio holdings. In our computational experiments on this particular application, our algorithm performs reasonably well on medium-sized problems for real data when finding the optimal solution (if exists) or identifying primal infeasibility, and usually outperforms the standard interior-point solver SDPT3 in terms of CPU time.

[*]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: g.li@unsw.edu.au. This author was partially supported by a research grant from Australian Research Council.

[†]Department of Finance, The Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: alfred@baf.msmail.cuhk.edu.hk

[‡]Department of Combinatorics and Optimization, University of Waterloo, Waterloo, ON, Canada, N2L 3G1. E-mail: ptingkei@uwaterloo.ca. This author was supported by research grants from AFOSR and NSERC.

1 Introduction

It is widely recognized that solutions to optimization problems can be highly sensitive to perturbations of input data. Indeed, as discussed in [6], under data uncertainty, an "optimal" solution could even turn out to be highly infeasible, unreliable or both. See also, for example, [2], for a recent exposition. Thus, various approaches have been taken to tackle optimization with uncertainty. One popular approach is stochastic optimization, in which the decision-maker first imposes a probabilistic description on the uncertainty set and then finds a solution optimal in some probabilistic sense. We refer interested readers to several textbooks [8, 24] and references therein for a detailed discussion of this subject. Another popular approach is robust optimization, which is the subject of this paper. In this approach, data uncertainty is modeled deterministically via a set called "uncertainty set". With this uncertainty set incorporated, the feasible region of the original optimization problem is usually shrunk. The decision-maker then computes an optimal solution over the smaller feasible region as a conservative strategy. This approach has a long history. Soyster [25] was one of the first researchers to study explicit approaches to robust optimization, dating back to the early 1970s. In recent years, Ben-Tal and Nemirovski [1, 2, 3], El Ghaoui et al. [11] and Bertsimas et al. [5, 6] further developed and lifted Soyster's idea into a comprehensive theory and presented numerous applications. We refer the readers to the recent comprehensive monograph [1] and the survey paper [6] as well as the references therein for more details.

In this paper, we consider the following least square semidefinite programming problem:

$$\begin{aligned} \min_{X\in\mathcal{S}^n} \ & \frac{1}{2}\|X-G\|_F^2 \\ \text{s.t. } & \mathcal{A}(X)\le b, \\ & \mathcal{B}(X)=h, \\ & X\succeq 0, \end{aligned} \qquad \text{(LSSDP)}$$

where $\|\cdot\|_F$ denotes the usual Fröbenius norm of a matrix, $b=(b_1,\dots,b_m)^T\in\mathbb{R}^m$ and $h=(h_1,\dots,h_p)^T\in\mathbb{R}^p$, and $\mathcal{A}\in\mathcal{L}(\mathcal{S}^n:\mathbb{R}^m)$ and $\mathcal{B}\in\mathcal{L}(\mathcal{S}^n:\mathbb{R}^p)$ are two linear mappings defined by

$$\mathcal{A}(X)=(\mathrm{tr}(A_1X),\dots,\mathrm{tr}(A_mX))^T,\ \mathcal{B}(X)=(\mathrm{tr}(B_1X),\dots,\mathrm{tr}(B_pX))^T,$$

with $A_i,B_j\in\mathcal{S}^n$ (see the next section for notations). This model arises in many important applications in finance and management science. A simple example is the problem of finding the nearest correlation matrix subject to additional linear constraints, i.e., $p=n$ and $\mathrm{tr}(B_iX)=X_{ii}=1$ for all $i=1,...,n$; see, for example [12, 20]. We assume in addition that the data $(\mathcal{A},b)\in\mathcal{L}(\mathcal{S}^n:\mathbb{R}^m)\times\mathbb{R}^m$ in (LSSDP) is uncertain and they belong to the following *ellipsoidal uncertainty set*:

$$\Upsilon=\{(\mathcal{A}^{(0)},b^{(0)})+\sum_{k=1}^{l}w^k(\mathcal{A}^{(k)},b^{(k)}):\|(w^1,\dots,w^l)\|\le 1\}, \tag{1.1}$$

where $(\mathcal{A}^{(k)},b^{(k)})\in\mathcal{L}(\mathcal{S}^n:\mathbb{R}^m)\times\mathbb{R}^m$, $k=0,1,\dots,l$. We note that the ellipsoidal uncertainty is a popular tool in describing and handling uncertainty, and has natural statistical interpretation. On the other hand, we do not impose uncertainty in the equality constraints as the equality constraints usually represent some normalized or fixed element condition (for example, the constraints $X_{ii}=1$ for all i in finding the nearest correlation matrix), and is thus often free of uncertainty. Following the robust optimization approach, the robust counterpart of (LSSDP) is then formulated as follows:

$$\begin{aligned} v := \min_{X \in \mathcal{S}^n} & \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t. } & \mathcal{A}(X) \leq b, \forall (\mathcal{A}, b) \in \Upsilon, \\ & \mathcal{B}(X) = h, \\ & X \succeq 0, \end{aligned} \tag{RSDP}$$

We say $\bar{X}$ is a robust solution to (LSSDP) if it is a solution of the robust counterpart (RSDP).

Our motivation for considering (RSDP) stems from correlation stress testing, which is an important tool in risk management. Stress testing is not a clearly defined technique and is in general very subjective [4], stress testing correlation matrices can even said to be *ad hoc* [16, 17, 21, 29]. Hypothetical scenarios are usually used for stress testing purposes. In the context of correlation stress testing, we expect the correlations to increase in crisis. Since such kinds of situations may not be observed from the historical data, the correlation matrix estimated from the historical data needs to be modified for stress testing. To stress test a portfolio of correlated assets, the risk manager can specify the hypothetical scenarios and accordingly compute the loss of the portfolio. Ideally, we want to modify some entries of the correlation matrix but the resulting matrix may not be a legitimate correlation matrix. To ensure the modified matrix is a correlation matrix, we propose the robust correlation stress testing, modeled as a robust least square semidefinite programming problem with ellipsoidal data uncertainty. This robust setting particularly fits the stress testing framework in that we actually can stress test the portfolio with a class of hypothetical scenarios within one optimization routine instead of working on each scenario one at a time. A standard stress test is to find the worst-case performance of the portfolio and the robust optimization framework exactly provides this solution. In addition, when converting a 1-day Value-at-Risk to a 10-day Value-at-Risk, we usually assume the portfolio weights are unchanged. In our proposed robust stress testing framework, we can also stress test the portfolio even for changes in the portfolio weights. We present more details in Section 6.

In this paper, we study the properties of (RSDP), propose an algorithm for solving it and illustrate how (RSDP) can be applied in robust correlation stress testing. More specifically, we study the asymptotic rate of convergence of the solution of (RSDP) to the solution of (LSSDP), as the uncertainty set shrinks. We also show that (RSDP) can be formulated as a least square semidefinite linear programming problem with an additional second-order cone constraint. To develop an algorithm, as in [12, 20], we consider the dual problem of (RSDP). We present two constraint qualifications under which the duality gap is zero; while the first one is a standard Slater condition, the second condition does not require (RSDP) to be feasible. This second condition is particularly interesting since infeasible instances naturally arise when the uncertainty set is too large. In the correlation stress testing, infeasible instances also arise when there are too many pre-fixed elements so that there may not be a correlation matrix with those fixed correlations. The dual problem of (RSDP) consists of maximizing a smooth function over a simple closed convex set. Thus, various algorithms can be applied to solve this problem. In this paper, we look at a simple variant of the spectral projected gradient method [7]. While it is well-known that the sequence thus generated accumulates at a dual optimal solution, we show under suitable constraint qualifications that the dual objective value along the sequence generated converges to a finite value if and only if (RSDP) is feasible, providing a simple certificate for primal infeasibility. We remark that Newton type method described in [12] could also be suitably adapted, but it is not immediate to us how primal infeasibility can be certified directly from the algorithm. Finally, as an application of our model, we present the idea of robust correlation stress testing and discuss how it can be modeled as (RSDP).

The rest of this paper is organized as follows. In Section 2, we introduce notations and recall some basic facts that will be used throughout the paper. In Section 3, we reformulate (RSDP) as a semidefinite linear programming problem with an additional second-order cone constraint, and present sensitivity analysis of this problem. In Section 4, we establish that the reformulated problem and its Lagrangian dual have zero duality gap under suitable constraint qualifications. In Section 5, we propose a simple variant of the spectral projected gradient method to solve the dual problem. In Section 6, we discuss robust correlation stress testing and report the numerical experiment on medium-sized problems for real data on this particular application. We compare our algorithm with the standard interior-point solver SDPT3. Finally, we conclude this paper and present some future research topics in Section 7.

2 Notations and preliminaries

In this paper, the symbols $\mathbb{R}^n$ and $\mathbb{R}^n_+$ denote the n-dimensional real Euclidean space and its nonnegative orthant, respectively. For a vector $v \in \mathbb{R}^n$, v^T denotes its transpose and $\|v\|$ denotes the Euclidean norm of v. The $(n+1)$-dimensional second-order cone is denoted by SOC_n, which is defined as

$$\mathrm{SOC}_n = \{x \in \mathbb{R}^{n+1} :\ x_1 \geq \sqrt{x_2^2 + \cdots + x_{n+1}^2}\}.$$

The space of $n \times n$ real symmetric matrices is denoted by $\mathcal{S}^n$. For a matrix $A \in \mathcal{S}^n$, $\|A\|_F$ denotes the Fröbenius norm of A and $\mathrm{vec}(A)$ denotes the vector obtained by stacking the columns of A. The cone of $n \times n$ positive semidefinite matrices is denoted by $\mathcal{S}^n_+$. For matrices $A, B \in \mathcal{S}^n$, $A \succeq B$ (resp., $A \succ B$) means $A - B$ is positive semidefinite (resp., positive definite). The vector of all ones is denoted by e and the identity matrix is denoted by I, whose dimensions should be clear from the context. The space of all linear mappings from $\mathcal{S}^n$ to $\mathbb{R}^m$ (resp., $\mathbb{R}^m$ to $\mathcal{S}^n$) is denoted by $\mathcal{L}(\mathcal{S}^n : \mathbb{R}^m)$ (resp., $\mathcal{L}(\mathbb{R}^m : \mathcal{S}^n)$). Elements in $\mathcal{L}(\mathcal{S}^n : \mathbb{R}^m)$ are represented by scripted letters, for example, $\mathcal{A}$. Let $\mathcal{A} \in \mathcal{L}(\mathcal{S}^n : \mathbb{R}^m)$. The conjugate linear mapping $\mathcal{A}^* \in \mathcal{L}(\mathbb{R}^m : \mathcal{S}^n)$ of $\mathcal{A}$ is defined by

$$\mathrm{tr}(X\mathcal{A}^*(y)) = \mathcal{A}(x)^T y \text{ for all } X \in \mathcal{S}^n, y \in \mathbb{R}^m,$$

where we use tr to denote the trace throughout.

For a set $\Omega \subseteq \mathbb{R}^n$, $\mathrm{int}(\Omega)$ denotes the interior of Ω, $\overline{\Omega}$ is the closure of Ω and $\mathrm{dist}(x, \Omega) := \inf_{y \in \Omega} \|x - y\|$ is the distance from $x \in \mathbb{R}^n$ to Ω. Similarly, for a set $\Omega \subseteq \mathcal{S}^n$, $\mathrm{dist}(X, \Omega) := \inf_{Y \in \Omega} \|X - Y\|_F$ is the distance from $X \in \mathcal{S}^n$ to Ω. We say a set Ω is convex if $\mu x + (1-\mu)y \in \Omega$ for all $x, y \in \Omega$ and $\mu \in [0, 1]$. For a closed convex set $\Omega \subseteq \mathbb{R}^n$ (resp., $\mathcal{S}^n$), let $\mathrm{Pr}_\Omega(\cdot)$ be the projection mapping onto Ω, measured with respect to the Euclidean (resp., Fröbenius) norm. For a closed convex set $\Omega \subseteq \mathbb{R}^n$, the normal cone of Ω at $x \in \Omega$ is defined as

$$N_\Omega(x) := \left\{z \in \mathbb{R}^n : z^T(y - x) \leq 0 \text{ for all } y \in \Omega\right\}.$$

Finally, a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if, for all $\mu \in [0, 1]$ and $x, y \in \mathbb{R}^n$,

$$f((1-\mu)x + \mu y) \leq (1-\mu)f(x) + \mu f(y).$$

A function f is said to be concave whenever $-f$ is convex.

We close this section by stating three useful facts which concerns the estimations of the distance to a given convex set. The first two lemmas present explicit error bounds (with explicit moduli) bounding the distance to the intersection of closed convex sets by the distances to each individual set. In particular, the first lemma is a generalization of [15, Theorem 3.1] which assumes the two sets intersect with nonempty interior, and is a slight modification of the result in [18, Lemma 4.10].

For the convenience of the readers, the proof of the first lemma is included in the appendix. The second lemma is a special case of [15, Theorem 3.1] by invoking [15, Remark 2.1(c)]. Finally, the third lemma is an important error bound result derived by Robinson [22] which bounds the distance to a convex set in terms of a related residual function. Here and throughout the remainder of the paper, we use $B(X,r) \subseteq \mathcal{S}^n$ to denote the closed ball centered at $X \in \mathcal{S}^n$ with radius r.

Lemma 2.1. *Let Ω_1 and Ω_2 be closed convex sets in $\mathcal{S}^n$ and let $X_0 \in \Omega_1 \cap \Omega_2$. If $B(X_0,\delta) \subseteq \Omega_2$ and $\Omega_1 \cap \Omega_2 \subseteq B(X_0,R)$ for some $R > \delta > 0$, then we have*

$$\operatorname{dist}(X, \Omega_1 \cap \Omega_2) \leq 2\left(1 + \frac{R}{\delta}\right) \max\{\operatorname{dist}(X,\Omega_1), \operatorname{dist}(X,\Omega_2)\} \quad \text{for all } X \in \mathcal{S}^n.$$

Lemma 2.2. (cf. [15, Theorem 3.1]) *Let Ω_i, $i = 1,\ldots,m$, be closed convex sets in $\mathcal{S}^n$ and let $\Omega = \bigcap_{i=1}^m \Omega_i$. Suppose that there exist $X_0 \in \Omega$, $\delta > 0$ and $R > 0$ be such that $B(X_0,\delta) \subseteq \Omega \subseteq B(X_0,R)$. Then, we have*

$$\operatorname{dist}(X,\Omega) \leq \frac{R}{\delta} \max_{1\leq i\leq m} \operatorname{dist}(X,\Omega_i) \quad \text{for all } X \in \mathcal{S}^n.$$

Lemma 2.3. (cf. [22, Equation 4]) *Let $g : \mathcal{S}^n \to \mathbb{R}$ be a convex function. Let $\Omega = \{X : g(X) \leq 0\}$. Suppose that $D := \sup_{X,Y\in\Omega} \|X - Y\|_F < \infty$ and there exist $X_0 \in \mathcal{S}^n$ and $\delta > 0$ such that $g(X_0) \leq -\delta < 0$. Then, for each $X \in \mathcal{S}^n$, we have*

$$d(X,\Omega) \leq \frac{D}{\delta} \max\{g(X), 0\}.$$

3 Reformulation of RSDP and sensitivity analysis

In this section, we study various properties of (RSDP). Specifically, we reformulate (RSDP) as a least square linear semidefinite programming problem with an additional second-order cone constraint. We also show that, under certain conditions, the robust solution will converge to the optimal solution of (LSSDP) when the diameter of the uncertainty set shrinks to zero.

We start by deriving a reformulation of (RSDP). For notational simplicity, from now on, we denote

$$\mathcal{P}_i(X) = \begin{pmatrix} \operatorname{tr}(A_i^{(0)}X) \\ \vdots \\ \operatorname{tr}(A_i^{(l)}X) \end{pmatrix} \text{ and } q_i := \begin{pmatrix} b_i^{(0)} \\ \vdots \\ b_i^{(l)} \end{pmatrix} \quad \forall i = 1,\ldots,m. \tag{3.1}$$

Proposition 3.1. (Second-order cone reformulation of the robust problem) *Problem* (RSDP) *is equivalent to the following convex optimization problem:*

$$\begin{aligned} \min_{X\in\mathcal{S}^n} \ & \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t. } & \mathcal{P}_i(X) - q_i \in -\mathrm{SOC}_l \quad \forall i = 1,\ldots,m, \\ & \mathcal{B}(X) = h, \\ & X \succeq 0. \end{aligned} \tag{3.2}$$

Proof. First, let $(\mathcal{A}^{(k)}, b^{(k)}) \in \mathcal{L}(\mathcal{S}^n : \mathbb{R}^m) \times \mathbb{R}^m$, $k = 0,1,\ldots,l$, be defined by

$$\mathcal{A}^{(k)}(X) = (\operatorname{tr}(A_1^{(k)}X), \ldots, \operatorname{tr}(A_m^{(k)}X)) \text{ and } b^{(k)} = (b_1^{(k)}, \ldots, b_m^{(k)}).$$

Then it is easy to see that for any $X \in \mathcal{S}^n$,

$$\begin{aligned}\mathcal{A}(X) \le b \ \ \forall (\mathcal{A}, b) \in \Upsilon &\iff \mathcal{A}^{(0)}(X) + \sum_{k=1}^{l} w^k \mathcal{A}^{(k)}(X) \le b^{(0)} + \sum_{k=1}^{l} w^k b^{(k)} \ \ \forall \|w\| \le 1 \\ &\iff \sum_{k=1}^{l} w^k \left(\mathrm{tr}(A_i^{(k)} X) - b_i^{(k)}\right) \le -\left(\mathrm{tr}(A_i^{(0)} X) - b_i^{(0)}\right) \ \ \forall \|w\| \le 1, \forall i = 1, ..., m, \\ &\iff \mathcal{P}_i(X) - q_i \in -\mathrm{SOC}_l \ \ \forall i = 1, ..., m.\end{aligned}$$

This proves the proposition. □

The next interesting and important question is to study the behaviour of the optimal solution of (RSDP) as $\mathrm{diam}(\Upsilon) \to 0$, where

$$\mathrm{diam}\Upsilon := \sup_{(\mathcal{A}_1, u), (\mathcal{A}_2, v) \in \Upsilon} \{\|\mathcal{A}_1 - \mathcal{A}_2\|_2 + \|u - v\|\},$$

and $\|\cdot\|_2$ is the induced operator norm in $\mathcal{L}(\mathcal{S}^n : \mathbb{R}^m)$. Before proceeding, we first introduce some notations. For $t = 1, 2, \ldots$, consider the following optimization problem:

$$\begin{aligned}\min_{X \in \mathcal{S}^n} \ & \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t.} \ & \mathcal{A}(X) \le b, \forall (A, b) \in \Upsilon_t, \\ & \mathcal{B}(X) = h, \\ & X \succeq 0,\end{aligned} \tag{RSDP_t}$$

where $\Upsilon_t := \{(\mathcal{A}^{(0)}, b^{(0)}) + \sum_{k=1}^{l} w^k (\mathcal{A}^{(k),t}, b^{(k),t}) : \|(w^1, \ldots, w^l)\| \le 1\}$ and the problem (LSSDP) with $(\mathcal{A}^{(0)}, b^{(0)})$ in place of $(\mathcal{A}, b)$, namely,

$$\begin{aligned}\min_{X \in \mathcal{S}^n} \ & \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t.} \ & \mathcal{A}^{(0)}(X) \le b^{(0)}, \\ & \mathcal{B}(X) = h, \\ & X \succeq 0.\end{aligned} \tag{LSSDP_0}$$

We show in the next theorem that the optimal solution of (RSDP_t) converges to the optimal solution of (LSSDP_0) under a pseudo-Slater-type assumption on (LSSDP_0). Moreover, we obtain an explicit asymptotic rate of convergence in terms of $\mathrm{diam}(\Upsilon_t)$.

Theorem 3.1. (Sensitivity analysis) *Suppose there exists $X_0 \succeq 0$ with $\mathcal{B}(X_0) = h$ and $\mathcal{A}^{(0)}(X_0) < b^{(0)}$. For each $t \ge 1$, let $\bar{X}^t$ denote the unique minimizer of* (RSDP_t)*. Assume that* $\mathrm{diam}(\Upsilon_t) \to 0$*. Then $\bar{X}^t$ is well-defined for large t and converges to the unique minimizer X^* of* (LSSDP_0)*. Moreover, it holds that*

$$\|\bar{X}^t - X^*\|_F = O\left(\sqrt{\mathrm{diam}(\Upsilon_t)}\right),$$

for all sufficiently large t.

Proof. Let Ω denote the feasible set of (LSSDP_0). Then Ω is closed and convex. Moreover, it follows from the existence of X_0 in the assumption that $\Theta \ne \emptyset$ and $\Omega = \overline{\Theta}$, where

$$\Theta := \{X \succeq 0 : \ \mathcal{B}(X) = h, \mathcal{A}^{(0)}(X) < b^{(0)}\}.$$

Consider any $X \in \Theta$. Since $\operatorname{diam}(\Upsilon_t) \to 0$, for all sufficiently large t, we see that X is feasible for (RSDP$_t$). This implies that (RSDP$_t$) is feasible for all sufficiently large t and hence $\bar{X}^t$ is well-defined for all large enough t. We next show that $\lim \bar{X}^t = X^*$.

To this end, observe that from the feasibility of X for (RSDP$_t$) with large t, we have

$$\|X - G\|_F \geq \|\bar{X}^t - G\|_F \tag{3.3}$$

for all sufficiently large t. This shows in particular that $\{\bar{X}^t\}$ is a bounded sequence and thus has at least one cluster point. Let $\tilde{X}$ be one cluster point of $\{\bar{X}^t\}$. It then follows from the definition of cluster point and (3.3) that

$$\|X - G\|_F \geq \|\tilde{X} - G\|_F. \tag{3.4}$$

Note that this last relation holds for any $X \in \Theta$. Taking infimum over all such X in (3.4) and using the fact that $\Omega = \overline{\Theta}$, we obtain further that

$$\|X^* - G\|_F \geq \|\tilde{X} - G\|_F.$$

Since $\tilde{X}$ is clearly feasible for (LSSDP$_0$), we conclude from the uniqueness of optimal solution of (LSSDP$_0$) that $\tilde{X} = X^*$. Thus, X^* is the unique cluster point for the sequence $\{\bar{X}^t\}$, showing that $\lim \bar{X}^t = X^*$.

We now turn to the second part. Let $R > \|X^* - X_0\|_F$ be large enough so that the ball $B(X_0, R)$ contains the bounded sequence $\{\bar{X}^t\}$ and let $\Omega(u^t)$ denote the feasible set of (RSDP$_t$), where u^t is given by

$$u^t = \Big((\mathcal{A}^{(1),t}, b^{(1),t}), \ldots, (\mathcal{A}^{(l),t}, b^{(l),t}) \Big),$$

which converges to zero by assumption. From the definition of $\Omega(u^t)$ and Proposition 3.1, we see immediately that

$$\Omega(u^t) = \Xi(u^t) \cap \{X : \ \mathcal{B}(X) = h\} \cap \mathcal{S}^n_+,$$

where

$$\Xi(u^t) := \bigcap_{i=1}^{m} \{X : \ \mathcal{P}_i^t(X) - q_i^t \in -\mathrm{SOC}_l\},$$

and $\mathcal{P}_i^t$ and q_i^t are defined as in (3.1) for u^t. In addition, from the assumption on X_0, we deduce that there exist $\delta \in (0, R)$ and $t_0 \geq 1$ such that

$$B(X_0, \delta) \subseteq \Xi(u^t) \ \text{ whenever } t \geq t_0. \tag{3.5}$$

Then, for all $t \geq t_0$, we have from (3.5) and Lemma 2.1 that

$$\operatorname{dist}(X^*, \Omega(u^t) \cap B(X_0, R)) \leq 2\left(1 + \frac{R}{\delta}\right) \operatorname{dist}\left(X^*, \Xi(u^t) \cap B(X_0, R)\right). \tag{3.6}$$

Furthermore, it follows from (3.5) and Lemma 2.2 that for all $t \geq t_0$

$$\operatorname{dist}\left(X^*, \Xi(u^t) \cap B(X_0, R)\right) \leq \frac{R}{\delta} \max_{i=1,\ldots,m} \left\{\operatorname{dist}\left(X^*, \{X \in B(X_0, R) : \ q_i^t - \mathcal{P}_i^t(X) \in \mathrm{SOC}_l\}\right)\right\}. \tag{3.7}$$

We next derive another upper bound for the right hand side of (3.7). To this end, for a vector $y \in \mathbb{R}^{l+1}$, denote $y^\sim := (y_2, ..., y_{l+1}) \in \mathbb{R}^l$. For each $i = 1, ..., m$, consider the continuous convex function $\ell_i^t(X) := \max\{\|y^\sim\| - y_1, \|X - X_0\| - R\}$ with $y := q_i^t - \mathcal{P}_i^t(X)$. Then it is clear that

$$\{X \in B(X_0, R) : \ q_i^t - \mathcal{P}_i^t(X) \in \mathrm{SOC}_l\} = \{X : \ \ell_i^t(X) \leq 0\} =: \Delta_i.$$

Moreover, observe that the set Δ_i is compact convex and that there exist $t_1 \geq t_0$ and $\beta \in (0, R)$ such that $\ell_i^t(X_0) < -\beta < 0$ for all $t \geq t_1$, $i = 1, ..., m$. Thus, by Lemma 2.3, we have for all $t \geq t_1$ and each $i = 1, ..., m$ that

$$\mathrm{dist}(X^*, \{X \in B(X_0, R) : \; q_i^t - \mathcal{P}_i^t(X) \in \mathrm{SOC}_l\}) \leq \frac{2R}{\beta} \max\{0, \ell_i^t(X^*)\}. \tag{3.8}$$

In addition, observe that $\|y^\sim\| - y_1 \leq \sqrt{2}\mathrm{dist}(y, \mathrm{SOC}_l)$ for any $y \in \mathbb{R}^{l+1}$. Combining this observation with (3.8) and (3.7), we obtain further that, for each $i = 1, ..., m$ and $t \geq t_1$,

$$\mathrm{dist}\left(X^*, \Xi(u^t) \cap B(X_0, R)\right) \leq \frac{2\sqrt{2}R^2}{\delta\beta} \max_{i=1,\ldots,m} \left\{\mathrm{dist}\left(\mathcal{P}_i^t(X^*) - q_i^t, -\mathrm{SOC}_l\right)\right\}. \tag{3.9}$$

Combining (3.6) and (3.9), we conclude that for any $t \geq t_1$,

$$\mathrm{dist}\left(X^*, \Omega(u^t) \cap B(X_0, R)\right) = O(\|u^t\|), \tag{3.10}$$

Next, let $f(X) := \frac{1}{2}\|X - G\|_F^2$. From the optimality of X^* for (LSSDP_0), we see that

$$f(X) - f(X^*) \geq \frac{1}{2}\|X - X^*\|_F^2 \tag{3.11}$$

for all X feasible for (LSSDP_0). Also, f is locally Lipschitz and hence there exist $L > 0$ and $\eta > 0$ such that

$$|f(X) - f(X^*)| \leq L\|X - X^*\|_F \tag{3.12}$$

whenever $X \in B(X^*, \eta)$. On the other hand, from (3.10), we see that for all $t \geq t_1$, there exists $\hat{X}^t \in \Omega(u^t) \cap B(X_0, R)$ such that

$$\|X^* - \hat{X}^t\|_F = O(\|u^t\|). \tag{3.13}$$

This together with the fact that $\bar{X}^t \to X^*$ implies that there exists $t_2 \geq t_1$ such that $\bar{X}^t, \hat{X}^t \in B(X^*, \eta)$ whenever $t \geq t_2$. Using this fact, (3.11), (3.12) and (3.13), we obtain that for any $t \geq t_2$,

$$\frac{1}{2}\|\bar{X}^t - X^*\|_F^2 \leq f(\bar{X}^t) - f(X^*) \leq f(\hat{X}^t) - f(X^*) \leq L\|X^* - \hat{X}^t\|_F = O(\|u^t\|)$$

Hence, for all $t \geq t_2$,

$$\|\bar{X}^t - X^*\|_F = O\left(\sqrt{\|u^t\|}\right).$$

The conclusion now immediately follows by noting that $\mathrm{diam}(\Upsilon_t) = O(\|u^t\|)$. □

4 Duality and Constraint Qualifications

In this section, we derive the dual of (3.2), and thus of (RSDP). We present two conditions under which the duality gap is zero. Specifically, we consider the following two constraint qualifications:

(CQ I) There exists $X_0 \succ 0$ with $\mathcal{B}(X_0) = h$ and $\mathcal{A}(X_0) < b$ for all $(\mathcal{A}, b) \in \Upsilon$.

(CQ II) There exists an index set $J \subseteq \{1, \ldots, p\}$ such that the set $\{X \succeq 0 : \; (\mathcal{B}(X))_J = h_J\}$ is nonempty and compact.

Similarly as in the proof of Proposition 3.1, it can be shown that **(CQ I)** is equivalent to the following condition **(CQ I′)**:

(CQ I′) There exists $X_0 \succ 0$ with $\mathcal{B}(X_0) = h$ and $\mathcal{P}_i(X_0) - q_i \in -\mathrm{int}\; \mathrm{SOC}_l$ for all i.

Remark 4.1. (Comparison of the two constraint qualifications) *While easy to state, the condition* **(CQ I)***, or in particular, feasibility of* (RSDP)*, is in general nontrivial to certify especially when the size of the problem is large.*

On the other hand, condition **(CQ II)** *does not concern the feasibility of* (RSDP) *and can sometimes be easily verified. In particular, in the important application of finding nearest correlation matrix, the condition* **(CQ II)** *is always satisfied as* $\{X \succeq 0 : X_{ii} = 1, i = 1, \ldots, n\}$ *is compact and contains the identity matrix. However, the corresponding feasible set of* (RSDP) *in this case could be empty even though* **(CQ II)** *is satisfied.*

In the next theorem, we derive the dual problem of (RSDP) and show that the duality gap is zero under either condition **(CQ I)** (or, equivalently, **(CQ I′)**) or **(CQ II)**. Furthermore, the unique solution of (RSDP) can be obtained from a solution of the dual problem. The proof of this theorem is standard and is based on the well-known Lagrangian duality and minimax theory. Here, we provide the proof for the completeness of the paper. Before proceeding, we introduce the following concave function

$$\theta(s,\mu) = -\frac{1}{2}\left\|\Pr_{\mathcal{S}^n_+}\left(G - \sum_{i=1}^m \mathcal{P}_i^*(s_i) - \mathcal{B}^*(\mu)\right)\right\|_F^2 + \frac{1}{2}\|G\|_F^2 - \left(\sum_{i=1}^m s_i^T q_i + \mu^T h\right)$$

for any $s = (s_1^{(0)}, \ldots, s_1^{(l)}, s_2^{(0)}, \ldots, s_2^{(l)}, \ldots, s_m^{(0)}, \ldots, s_m^{(l)}) \in \mathbb{R}^{m(l+1)}$ and $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$, and the optimization problem

$$\max_{s\in \mathrm{SOC}_l^m, \mu \in \mathbb{R}^p} \theta(s,\mu). \tag{DSDP}$$

Here $\mathcal{P}_i^*$ and $\mathcal{B}^*$ are respectively given by

$$\mathcal{P}_i^*(s_i) = \sum_{k=0}^l s_i^{(k)} A_i^{(k)} \quad \text{and} \quad \mathcal{B}^*(\mu) = \sum_{j=1}^p \mu_j B_j,$$

where $s_i := (s_i^{(0)}, s_i^{(1)}, \ldots, s_i^{(l)})$, defined in accordance with q_i in (3.1).

Theorem 4.1. (Duality and realization of the primal robust solution) *The following statements are true.*

(i) *If* **(CQ I)** *holds, then*

$$\min_{X\in\mathcal{S}^n}\left\{\frac{1}{2}\|X-G\|_F^2 : \mathcal{A}(X) \le b, \forall(\mathcal{A},b)\in\Upsilon, \mathcal{B}(X) = h, X \succeq 0\right\} = \max_{s\in \mathrm{SOC}_l^m, \mu\in\mathbb{R}^p} \theta(s,\mu), \tag{4.1}$$

and the maximum in the dual problem is attained.

(ii) *If* **(CQ II)** *holds, then* (4.1) *holds, with both values possibly equal positive infinity.*

Furthermore, in either case, if $(\bar{s}, \bar{\mu}) \in \mathrm{SOC}_l^m \times \mathbb{R}^p$ *solves* (DSDP)*, then*

$$\bar{X} = \Pr_{\mathcal{S}^n_+}\left(G - \sum_{i=1}^m \mathcal{P}_i^*(\bar{s}_i) - \mathcal{B}^*(\bar{\mu})\right) \tag{4.2}$$

is a solution of (RSDP).

Proof. If **(CQ I)** holds, then we have, from Proposition 3.1, that

$$
\begin{aligned}
v = & \min_{X\in\mathcal{S}^n}\left\{\frac{1}{2}\|X-G\|_F^2 : \mathcal{A}(X)\le b, \forall(\mathcal{A},b)\in\Upsilon, \mathcal{B}(X)=h, X\succeq 0\right\}\\
= & \min_{X\succeq 0}\left\{\frac{1}{2}\|X-G\|_F^2 : \mathcal{P}_i(X)-q_i\in -\mathrm{SOC}_l, \forall i=1,...,m, \mathcal{B}(X)=h\right\}\\
= & \min_{X\succeq 0}\max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu\in\mathbb{R}^p}}\left\{\frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m s_i^T(\mathcal{P}_i(X)-q_i)+\mu^T(\mathcal{B}(X)-h)\right\} \qquad (4.3)\\
= & \max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu\in\mathbb{R}^p}}\min_{X\succeq 0}\left\{\frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m s_i^T(\mathcal{P}_i(X)-q_i)+\mu^T(\mathcal{B}(X)-h)\right\}, \qquad (4.4)
\end{aligned}
$$

where the last equality follows from **(CQ I′)** and [23, Theorem 28.2].

On the other hand, suppose that **(CQ II)** holds instead. Writing $\bar{J}=\{1,...,p\}\backslash J$, we have

$$
\begin{aligned}
v = & \min_{X\succeq 0}\left\{\frac{1}{2}\|X-G\|_F^2 : \mathcal{P}_i(X)-q_i\in -\mathrm{SOC}_l, \forall i=1,...,m, \mathcal{B}(X)=h\right\}\\
= & \min_{\substack{(\mathcal{B}(X))_J=h_J\\ X\succeq 0}}\max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu_{\bar{J}}\in\mathbb{R}^{|\bar{J}|}}}\left\{\frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m s_i^T(\mathcal{P}_i(X)-q_i)+\mu_{\bar{J}}^T(\mathcal{B}(X)-h)_{\bar{J}}\right\}\\
= & \max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu_{\bar{J}}\in\mathbb{R}^{|\bar{J}|}}}\min_{\substack{(\mathcal{B}(X))_J=h_J\\ X\succeq 0}}\left\{\frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m s_i^T(\mathcal{P}_i(X)-q_i)+\mu_{\bar{J}}^T(\mathcal{B}(X)-h)_{\bar{J}}\right\} \qquad (4.5)\\
= & \max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu\in\mathbb{R}^p}}\min_{X\succeq 0}\left\{\frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m s_i^T(\mathcal{P}_i(X)-q_i)+\mu^T(\mathcal{B}(X)-h)\right\}, \qquad (4.6)
\end{aligned}
$$

where (4.5) follows from the compactness assumption and the minimax theorem [23, Corollary 37.3.2], while (4.6) follows from the compactness assumption and [23, Theorem 30.4 (i)]; see also [23, Theorem 30.3] and [23, Pages 320-321].

Thus, under either condition, we have (4.4) (or, equivalently, (4.6)). Starting from this relation and completing the square, we see further that

$$
\begin{aligned}
v = & \max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu\in\mathbb{R}^p}}\min_{X\succeq 0}\left\{\frac{1}{2}\left\|X-G+\sum_{i=1}^m\mathcal{P}_i^*(s_i)+\mathcal{B}^*(\mu)\right\|_F^2+\left\langle G,\sum_{i=1}^m\mathcal{P}_i^*(s_i)+\mathcal{B}^*(\mu)\right\rangle\right.\\
& \left.-\frac{1}{2}\left\|\sum_{i=1}^m\mathcal{P}_i^*(s_i)+\mathcal{B}^*(\mu)\right\|_F^2-\sum_{i=1}^m s_i^Tq_i-\mu^Th\right\}\\
= & \max_{\substack{(s_1,\ldots,s_m)\in\mathrm{SOC}_l^m\\ \mu\in\mathbb{R}^p}}\min_{X\succeq 0}\left\{\frac{1}{2}\left\|X-G+\sum_{i=1}^m\mathcal{P}_i^*(s_i)+\mathcal{B}^*(\mu)\right\|_F^2-\frac{1}{2}\left\|G-\sum_{i=1}^m\mathcal{P}_i^*(s_i)-\mathcal{B}^*(\mu)\right\|_F^2\right.\\
& \left.+\frac{1}{2}\|G\|_F^2-\sum_{i=1}^m s_i^Tq_i-\mu^Th\right\}. \qquad (4.7)
\end{aligned}
$$

It is clear that the inner minimization of (4.7) is attained at

$$
X=\mathrm{Pr}_{\mathcal{S}_+^n}\left(G-\sum_{i=1}^m\mathcal{P}_i^*(s_i)-\mathcal{B}^*(\mu)\right). \qquad (4.8)
$$

Plugging this expression of X into (4.7), we obtain further that

$$v = \max_{\substack{(s_1,\ldots,s_m)\in \mathrm{SOC}_l^m \\ \mu\in \mathbb{R}^p}} \underbrace{-\frac{1}{2}\left\|\mathrm{Pr}_{\mathcal{S}_+^n}\left(G-\sum_{i=1}^m \mathcal{P}_i^*(s_i)-\mathcal{B}^*(\mu)\right)\right\|_F^2+\frac{1}{2}\|G\|_F^2-\left(\sum_{i=1}^m s_i^T q_i+\mu^T h\right)}_{\theta(s,\mu)},$$

where we have made use of the facts that, for any $M \in \mathcal{S}^n$, $M = \mathrm{Pr}_{\mathcal{S}_+^n}(M) + \mathrm{Pr}_{-\mathcal{S}_+^n}(M)$ and $\|M\|_F^2 = \|\mathrm{Pr}_{\mathcal{S}_+^n}(M)\|_F^2 + \|\mathrm{Pr}_{-\mathcal{S}_+^n}(M)\|_F^2$. This proves (4.1). The attainment of the dual maximization problem under **(CQ I)** follows from [23, Corollary 28.2.2].

Finally, suppose that either **(CQ I)** or **(CQ II)** holds. Then from (4.3), (4.4) and (4.6), we conclude that if $(\bar{s},\bar{\mu})$ solves (DSDP) and $\bar{X}$ solves (RSDP), then $(\bar{X},\bar{s},\bar{\mu})$ is a saddle point of the function

$$(X,s,\mu)\mapsto \frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m s_i^T(\mathcal{P}_i(X)-q_i)+\mu^T(\mathcal{B}(X)-h)$$

over the set $\mathcal{S}_+^n \times \mathrm{SOC}_l^m \times \mathbb{R}^p$. Equation (4.2) now follows from this fact and (4.8). □

In view of Theorem 4.1 (ii), we immediately have the following corollary concerning maximizing sequences for (DSDP).

Corollary 4.1. *Suppose that* **(CQ II)** *holds. Let* $\{(s^t,\mu^t)\}$ *be a maximizing sequence for* (DSDP). *Then* $\lim_{t\to\infty}\theta(s^t,\mu^t)=v$. *Furthermore, we have the following consequences:*

(i) *If* $v=+\infty$, *then* (RSDP) *is infeasible.*

(ii) *If* $\{(s^t,\mu^t)\}$ *is bounded (and so, has cluster points), then* v *is finite and the primal optimal solution* $\bar{X}$ *is given by* (4.2), *where* $(\bar{s},\bar{\mu})$ *is a cluster point of* $\{(s^t,\mu^t)\}$.

(iii) *If* $\|(s^t,\mu^t)\|\to+\infty$, *then any cluster point* $(\tilde{s},\tilde{\mu})$ *of the normalized sequence* $\{(\frac{s^t}{\|(s^t,\mu^t)\|},\frac{\mu^t}{\|(s^t,\mu^t)\|})\}$ *is a solution of the following system in* (s,μ):

$$\begin{cases} \displaystyle\sum_{i=1}^m \mathcal{P}_i^*(s_i)+\mathcal{B}^*(\mu)\in\mathcal{S}_+^n \\ \displaystyle\sum_{i=1}^m s_i^T q_i+\mu^T h\le 0 \\ s\in \mathrm{SOC}_l^m,\ \mu\in\mathbb{R}^p,\ \|(s,\mu)\|=1. \end{cases} \tag{4.9}$$

Proof. The fact that $\lim_{t\to\infty}\theta(s^t,\mu^t)=v$ follows from the definition of maximizing sequence and (4.1). The claims in (i) and (ii) follow immediately from the preceding theorem. To finish the proof, we only need to establish (iii). Let $\{(s^t,\mu^t)\}$ be a maximizing sequence for (DSDP) with $\|(s^t,\mu^t)\|\to+\infty$. Then, the sequence $\{(s^t,\mu^t)\}$ is in the dual feasible set $\mathcal{F}:=\mathrm{SOC}_l^m\times\mathbb{R}^p$. Suppose that $(\tilde{s},\tilde{\mu})$ is a cluster point of the normalized sequence. Then by passing to a subsequence if necessary, we may assume without loss of generality that

$$\frac{s^t}{\|(s^t,\mu^t)\|}\to\tilde{s}\in\mathrm{SOC}_l^m \text{ and } \frac{\mu^t}{\|(s^t,\mu^t)\|}\to\tilde{\mu}\in\mathbb{R}^p.$$

Since $\theta(s^t,\mu^t)\to v\in\mathbb{R}\cup\{\infty\}$, there exists $\eta_0\in\mathbb{R}$ such that $\theta(s^t,\mu^t)\ge\eta_0$ for all t. Thus, we have,

$$\eta_0\le -\frac{1}{2}\left\|\mathrm{Pr}_{\mathcal{S}_+^n}\left(G-\sum_{i=1}^m\mathcal{P}_i^*(s_i^t)-\mathcal{B}^*(\mu^t)\right)\right\|_F^2+\frac{1}{2}\|G\|_F^2-\left(\sum_{i=1}^m q_i^T s_i^t+h^T\mu^t\right). \tag{4.10}$$

Dividing $\|(s^t,\mu^t)\|^2$ on both sides of (4.10), making use of the homogeneity of $\mathrm{Pr}_{\mathcal{S}^n_+}(\cdot)$ and passing to limit, we obtain that

$$\left\|\mathrm{Pr}_{\mathcal{S}^n_+}\left(-\sum_{i=1}^m \mathcal{P}_i^*(\tilde{s})-\mathcal{B}^*(\tilde{\mu})\right)\right\|_F^2=\lim_{t\to\infty}\left\|\frac{\mathrm{Pr}_{\mathcal{S}^n_+}\left(G-\sum_{i=1}^m \mathcal{P}_i^*(s_i^t)-\mathcal{B}^*(\mu^t)\right)}{\|(s^t,\mu^t)\|}\right\|_F^2\le 0$$

This gives $\mathrm{Pr}_{\mathcal{S}^n_+}\left(-\sum_{i=1}^m \mathcal{P}_i^*(\tilde{s})-\mathcal{B}^*(\tilde{\mu})\right)=0$, which in turn implies that

$$\sum_{i=1}^m \mathcal{P}_i^*(\tilde{s})+\mathcal{B}^*(\tilde{\mu})\in\mathcal{S}^n_+. \tag{4.11}$$

Finally, dividing $\|(s^t,\mu^t)\|$ on both sides of (4.10) and passing to lower limit, we have

$$2\left(\sum_{i=1}^m q_i^T\tilde{s}_i+h^T\tilde{\mu}\right)\le 2\left(\sum_{i=1}^m q_i^T\tilde{s}_i+h^T\tilde{\mu}\right)+\liminf_{t\to\infty}\left\|\frac{\mathrm{Pr}_{\mathcal{S}^n_+}\left(G-\sum_{i=1}^m \mathcal{P}_i^*(s_i^t)-\mathcal{B}^*(\mu^t)\right)}{\|(s^t,\mu^t)\|^{\frac{1}{2}}}\right\|_F^2\le 0. \tag{4.12}$$

Thus, $(\tilde{s},\tilde{\mu})$ satisfies (4.9). This completes the proof. □

Next, we show that, under either condition **(CQ I)** or **(CQ II)**, one can bound the optimal value of (LSSDP) for any fixed $(\mathcal{A},b)\in\Upsilon$ using the optimal value of (RSDP).

Proposition 4.1. *Suppose that either condition* **(CQ I)** *or* **(CQ II)** *holds. Fix any* $(\mathcal{A},b)\in\Upsilon$ *and let* X^* *be the unique solution of the corresponding problem* (LSSDP). *Furthermore, let* $\bar{X}$ *and* $(\bar{s},\bar{\mu})$ *be the solution of* (RSDP) *and* (DSDP), *respectively. Then we have the following inequality:*

$$\frac{1}{2}\|\bar{X}-G\|_F^2-\sum_{i=1}^m \bar{s}_i^T(\mathcal{P}_i(X^*)-q_i)\le\frac{1}{2}\|X^*-G\|_F^2\le\frac{1}{2}\|\bar{X}-G\|_F^2. \tag{4.13}$$

Proof. Notice first that the second inequality in (4.13) is trivial, since $\bar{X}$ is feasible for (LSSDP) by the definition of (RSDP). We now prove the first inequality. From Theorem 4.1 and the definition of saddle point, we have

$$\frac{1}{2}\|\bar{X}-G\|_F^2=v=\min_{X\succeq 0}\left\{\frac{1}{2}\|X-G\|_F^2+\sum_{i=1}^m \bar{s}_i^T(\mathcal{P}_i(X)-q_i)+\bar{\mu}^T(\mathcal{B}(X)-h)\right\}.$$

Since $\mathcal{B}(X^*)=h$ and $X^*\succeq 0$, we see further that

$$\frac{1}{2}\|\bar{X}-G\|_F^2\le\frac{1}{2}\|X^*-G\|_F^2+\sum_{i=1}^m \bar{s}_i^T(\mathcal{P}_i(X^*)-q_i),$$

which is just the first inequality in (4.13). □

Remark 4.2. *The relation* (4.13) *bounds the optimal value of* (LSSDP) *in terms of that of* (RSDP). *Although the quantity* $\sum_{i=1}^m \bar{s}_i^T(\mathcal{P}_i(X^*)-q_i)$ *in the bound depends on* X^*, *it can actually be upper estimated independently of* X^* *under condition* **(CQ II)**. *Indeed, by complementary slackness and the optimality of* $\bar{X}$ *and* $(\bar{s},\bar{\mu})$ *for* (RSDP) *and* (DSDP), *we see that*

$$\left|\sum_{i=1}^m \bar{s}_i^T(\mathcal{P}_i(X^*)-q_i)\right|=\left|\sum_{i=1}^m \bar{s}_i^T(\mathcal{P}_i(X^*-\bar{X}))\right|\le\left\|\sum_{i=1}^m \mathcal{P}_i^*(\bar{s}_i)\right\|_F\|X^*-\bar{X}\|_F,$$

and $\|X^*-\bar{X}\|_F$ *is, in turn, bounded by the diameter of the set* $\{X\succeq 0:\ (B(X))_J=h_J\}$, *which is finite under condition* **(CQ II)**. *Thus, under condition* **(CQ II)**, *one can estimate how deviated the robust solution is in terms of the optimal value, without prior knowledge of the "true" solution.*

Recall that the function $\psi(X) = \frac{1}{2}\|\mathrm{Pr}_{\mathcal{S}_+^n}(X)\|_F^2$ is known to be continuously differentiable with $\nabla\psi(X) = \mathrm{Pr}_{\mathcal{S}_+^n}(X)$. Thus, θ is also continuously differentiable and its derivative is given by

$$\nabla\theta(s,\mu) = \left(F_1^{(0)}(s,\mu),\ldots,F_1^{(l)}(s,\mu),\ldots,F_m^{(0)}(s,\mu),\ldots,F_m^{(l)}(s,\mu),H_1(s,\mu),\ldots,H_p(s,\mu)\right)^T,$$

where, for each $i = 1,\ldots,m$, $k = 0,\ldots,l$, $j = 1,\ldots,p$,

$$F_i^{(k)}(s,\mu) = \mathrm{tr}\left(A_i^{(k)}\Xi\right) - b_i^{(k)}, \qquad H_j(s,\mu) = \mathrm{tr}\left(B_j\Xi\right) - h_j,$$

and $\Xi = \mathrm{Pr}_{\mathcal{S}_+^n}\left(G - \sum_{i=1}^m \mathcal{P}_i^*(s_i) - \mathcal{B}^*(\mu)\right)$. It can be easily seen that $\nabla\theta$ is a Lipschitz function over $\mathbb{R}^{m(l+1)+p}$. Hence (DSDP) can be solved by various algorithms including the gradient projection algorithm.

We next examine when the upper level set of the dual function is compact.

Proposition 4.2. (Compactness of the upper level set of the dual function) *Suppose the following regularity condition holds:*

$$\left.\begin{array}{c} \sum_{i=1}^m \mathcal{P}_i^*(s_i) + \mathcal{B}^*(\mu) \in \mathcal{S}_+^n \\ \sum_{i=1}^m s_i^T q_i + \mu^T h \le 0 \\ s \in \mathrm{SOC}_l^m, \mu \in \mathbb{R}^p \end{array}\right\} \Rightarrow (s,\mu) = (0,0). \tag{4.14}$$

Then, the function $-\theta$ is coercive on the dual feasible set $\mathcal{F} = \mathrm{SOC}_l^m \times \mathbb{R}^p$, that is, for all $\eta \in \mathbb{R}$ $\{(s,\mu) \in \mathcal{F} : \theta(s,\mu) \ge \eta\}$ is compact.

Proof. Suppose to the contrary that $-\theta$ is not coercive. Then there exists η_0 such that $K := \{(s,\mu) \in \mathcal{F} : \theta(s,\mu) \ge \eta_0\}$ is not compact. In particular, K contains a sequence $(s^t,\mu^t) \in \mathcal{F}$ such that $\|(s^t,\mu^t)\| \to \infty$. By passing to a subsequence if necessary, we assume that

$$\frac{s^t}{\|(s^t,\mu^t)\|} \to \tilde{s} \in \mathrm{SOC}_l^m \text{ and } \frac{\mu^t}{\|(s^t,\mu^t)\|} \to \tilde{\mu} \in \mathbb{R}^p \text{ for some } \|(\tilde{s},\tilde{\mu})\| = 1.$$

Similar to the proof of Corollary 4.1 (iii), we see that both (4.11) and (4.12) hold. Thus, our regularity assumption (4.14) gives $(\tilde{s},\tilde{\mu}) = (0,0)$, which contradicts $\|(\tilde{s},\tilde{\mu})\| = 1$. Consequently, $-\theta$ has to be coercive. This completes the proof. □

Remark 4.3. *It should be pointed out that if the mapping $\mathcal{B}^*$ is injective, i.e., the matrices $B_1,\ldots,B_p$ are linearly independent, then* **(CQ I)** *implies the regularity condition* (4.14).

We end this section by presenting a simple example showing that the function $-\theta$ may not be coercive if (4.14) is not satisfied.

Example 4.1. (Noncompactness of the upper level set without regularity condition) *Consider the problem* (RSDP) *with $l = m = 1$ and $p = 2$, where G is the two by two zero matrix,*

$$A_1^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_1^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$b_1^{(0)} = 2$, $b_1^{(1)} = 0$, *and*

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, B_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

and $h_1 = h_2 = 1$. It is easy to see that

$$\theta(s_1^{(0)}, s_1^{(1)}, \mu_1, \mu_2) = -\frac{1}{2}\left\| \mathrm{Pr}_{S_+^2}\begin{pmatrix} -s_1^{(0)} - \mu_1 & -s_1^{(1)} \\ -s_1^{(1)} & -s_1^{(0)} - \mu_2 \end{pmatrix} \right\|_F^2 - 2s_1^{(0)} - \mu_1 - \mu_2.$$

Consider a sequence $(s_1^{(0),t}, s_1^{(1),t}, \mu_1^t, \mu_2^t) = (t, 0, -t, -t) \in \mathrm{SOC}_1 \times \mathbb{R}^2$. Then $\|(s_1^{(0),t}, s_1^{(1),t}, \mu_1^t, \mu_2^t)\| \to \infty$ and

$$\theta(s_1^{(0),t}, s_1^{(1),t}, \mu_1^t, \mu_2^t) = -\frac{1}{2}\left\| \mathrm{Pr}_{S_+^2}\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\|_F^2 - 2t - (-t) - (-t) = 0.$$

Thus, $\{(s, \mu) \in \mathrm{SOC}_1 \times \mathbb{R}^2 : \theta(s, \mu) \geq 0\}$ is not compact. Moreover, it can be directly verified that $(1, 0, -1, -1)$ is a point which violates the condition (4.14).

5 Algorithm

In the section, we discuss a simple algorithm for solving (DSDP). Motivated by Corollary 4.1, we focus on algorithms that maintain *dual feasibility* at each iteration so that the sequence generated can likely provide information about primal feasibility. Since the feasible set of (DSDP) is easy to project onto (see, for example, [13, Proposition 3.3(b)]), we consider an adaptation of the spectral projected gradient method with nonmonotone line search in [7] to approximately solve (DSDP). We remark that Newton type method described in [12] could also be suitably adapted, but it is not immediate to us how primal infeasibility can be readily certified from the algorithm. The algorithm is described as follows. For notational convenience, we write $z := (s, \mu)$ where $s = (s_1, \ldots, s_m)$ with $s_i \in \mathbb{R}^{l+1}$ and $\mu \in \mathbb{R}^p$. Moreover, we use $\nabla_\mu \theta$ to denote the partial derivative with respect to μ and $\nabla_{s_i}\theta$ to denote the partial derivative with respect to s_i.

Spectral projected gradient (SPG) method:

1. **Start:** Choose parameters $0 < \beta, \sigma < 1$, $0 < \underline{\eta} < \bar{\eta}$ and integer $M \geq 1$. Let $z^0 \in \mathrm{SOC}_l^m \times \mathbb{R}^p$ be given and set $\bar{\tau}_0 = \bar{\eta}_{0,i} = 1$ for $i = 1, ..., m$.

2. **For** $t = 0, 1, \ldots$

   (a) Let $d_\mu^t = \bar{\tau}_t \nabla_\mu \theta(z^t)$ and $d_{s_i}^t = \mathrm{Pr}_{\mathrm{SOC}_l}(s_i^t + \bar{\eta}_{t,i}\nabla_{s_i}\theta(z^t)) - s_i^t$ for $i = 1, ..., m$. Set $d^t = (d_s^t, d_\mu^t)$.

   (b) Find the largest $\alpha \in \{1, \beta, \beta^2, ...\}$ such that

   $$\theta(z^t + \alpha d^t) \geq \min_{(t+1-M)_+ \leq r \leq t} \{\theta(z^r)\} + \sigma\alpha\nabla\theta(z^t)^T d^t. \tag{5.1}$$

   Set $\alpha_t \leftarrow \alpha$, $z^{t+1} \leftarrow z^t + \alpha_t d^t$ and $t \leftarrow t + 1$.

   (c) Update

   $$\bar{\tau}_{t+1} \leftarrow \min\left\{\max\left\{-\frac{\|s_\mu^t\|^2}{(s_\mu^t)^T y_\mu^t}, \underline{\eta}\right\}, \bar{\eta}\right\}, \quad \bar{\eta}_{t+1,i} \leftarrow \min\left\{\max\left\{-\frac{\|s_{s_i}^t\|^2}{(s_{s_i}^t)^T y_{s_i}^t}, \underline{\eta}\right\}, \bar{\eta}\right\},$$

   where $s_\mu^t = \mu^{t+1} - \mu^t$, $s_{s_i}^t = s_i^{t+1} - s_i^t$, $y_\mu^t = \nabla_\mu\theta(z^{t+1}) - \nabla_\mu\theta(z^t)$, $y_{s_i}^t = \nabla_{s_i}\theta(z^{t+1}) - \nabla_{s_i}\theta(z^t)$, for $i = 1, ..., m$.

   End (for)

Note that the main computational cost per iteration is to obtain the gradient of θ, which involves an eigenvalue decomposition and thus takes $O(n^3)$ flops in general. Furthermore, notice that from the definition of d^t and [28, Lemma 1], there exists $\delta > 0$ such that

$$\nabla\theta(z^t)^T d^t \geq \delta \|d^t\|^2 \tag{5.2}$$

for all $t > 0$. In particular, the step (5.1), and hence each iteration of the above algorithm, is well-defined at any nonstationary point z^t, where $d^t \neq 0$. In the next proposition, we analyze asymptotic properties of $\{z^t\}$. We first recall two important results concerning projections in the next lemma. The proof of the first property can be found in [14, Lemma 1], while the proof of second one can be found in [27, Lemma 2].

Lemma 5.1. *For any closed convex set $\Omega \subseteq \mathrm{IR}^n$, $x \in \Omega$ and $y \in \mathrm{IR}^n$,*

(i) *the mapping $\zeta \mapsto \left\| \frac{\Pr_\Omega(x+\zeta y) - x}{\zeta} \right\|$ is monotonically nonincreasing on $\zeta > 0$.*

(ii) *the mapping $\zeta \mapsto \|\Pr_\Omega(x+\zeta y) - x\|$ is monotonically nondecreasing on $\zeta \geq 0$.*

We show in the next proposition that the norm of the search direction $\{d^t\}$ decays to zero in the limit. Unlike similar analysis in [19], our function θ does not necessarily have compact upper level sets nor is it necessarily uniformly continuous on its upper level sets; hence our conclusion does not follow directly from their analysis. In our proof, we make use of uniform continuity of $\nabla\theta$ and Lemma 5.1 instead. Furthermore, we show that in the case when $\{z^t\}$ is unbounded and $\{\theta(z^t)\}$ is bounded, we can obtain a sequence from $\{z^t\}$ that is asymptotically feasible for (RSDP).

Theorem 5.1. (Convergence of the SPG method) *Let $\{z^t\}$ be generated by the SPG method and suppose that* **(CQ II)** *holds. Then, the following statements hold:*

(i) *Suppose that $\limsup_{t\to\infty} \theta(z^t) < \infty$. Then $\lim_{t\to\infty} \|d^t\| = 0$ and any cluster point $\bar{z} = (\bar{s}, \bar{\mu})$ of $\{z^t\}$ is a maximizer of* (DSDP) *and $\bar{X} := \Pr_{\mathcal{S}^n_+}\left(G - \sum_{i=1}^m \mathcal{P}_i^*(\bar{s}_i) - \mathcal{B}^*(\bar{\mu})\right)$ is a minimizer of* (RSDP)*. In addition, $\lim_{t\to\infty} \theta(z^t)$ exists and is finite. Moreover, the sequence*

$$X^t := \Pr_{\mathcal{S}^n_+}\left(G - \sum_{i=1}^m \mathcal{P}_i^*(s_i^t) - \mathcal{B}^*(\mu^t)\right) \tag{5.3}$$

is bounded. If $\{z^t\}$ does not have any cluster point, then any cluster point of $\{X^t\}$ is feasible for (RSDP).

(ii) *If $\limsup_{t\to\infty} \theta(z^t) = \infty$, then* (RSDP) *is infeasible.*

Proof. If $\limsup_{t\to\infty} \theta(z^t) = \infty$, then (RSDP) is infeasible by Theorem 4.1 (ii). We now consider the case when $\limsup_{t\to\infty} \theta(z^t) < \infty$. If $d^t = 0$ for some $t > 0$, then clearly z^t is a maximizer of (DSDP), and the conclusion immediately follows from this fact and Theorem 4.1. Thus, we assume further that $d^t \neq 0$ for all $t > 0$. Since $\nabla\theta$ is globally Lipschitz continuous, we obtain immediately from [28, Lemma 5(b)] the existence of $\underline{\alpha} > 0$ such that

$$\alpha_t \geq \underline{\alpha} \quad \text{for all } t > 0. \tag{5.4}$$

Next, for each $t > 0$, let the index $\iota(t)$ be the smallest integer in $\{(t+1-M)_+, ..., t\}$ such that

$$\theta(z^{\iota(t)}) = \min_{(t+1-M)_+ \leq r \leq t} \{\theta(z^r)\}. \tag{5.5}$$

Then we see from the definition, (5.1) and (5.2) that $\theta(z^{\iota(t)})$ is increasing. Since $\limsup_{t\to\infty} \theta(z^t) < \infty$, we conclude that $\lim_{t\to\infty} \theta(z^{\iota(t)}) = \bar{\theta}$ for some $\bar{\theta} \in \mathbb{R}$. On the other hand, using (5.5) and (5.1), we have for all $t > 0$ that

$$\begin{aligned}\theta(z^{\iota(t)}) &\geq \min_{(\iota(t)-M)_+ \leq r \leq \iota(t)-1} \{\theta(z^r)\} + \sigma\alpha_{\iota(t)-1}\nabla\theta(z^{\iota(t)-1})^T d^{\iota(t)-1} \\ &= \theta(z^{\iota(\iota(t)-1)}) + \sigma\alpha_{\iota(t)-1}\nabla\theta(z^{\iota(t)-1})^T d^{\iota(t)-1}. \end{aligned} \tag{5.6}$$

Combining (5.4), (5.6) and (5.2), we obtain that

$$\lim_{t\to\infty} \|d^{\iota(t)-1}\| = 0. \tag{5.7}$$

We claim that $\lim_{t\to\infty} \|d^t\| = 0$. To see this, notice first from Lemma 5.1, (5.7) and the definition of d^t that

$$\lim_{t\to\infty} \|\nabla_\mu\theta(z^{\iota(t)-1})\| = 0 \text{ and } \lim_{t\to\infty} \max_{1\leq i\leq m} \|\mathrm{Pr}_{\mathrm{SOC}_l}(s_i^{\iota(t)-1} + \bar{\eta}_{\iota(t)-1,i}\nabla_{s_i}\theta(z^{\iota(t)-1})) - s_i^{\iota(t)-1}\| = 0.$$

Using the Lipschitz continuity of $\nabla\theta$ and the definition that $z^{\iota(t)} = z^{\iota(t)-1} + \alpha_{\iota(t)-1}d^{\iota(t)-1}$, we conclude from the last relation that

$$\lim_{t\to\infty} \|\nabla_\mu\theta(z^{\iota(t)})\| = 0 \text{ and } \lim_{t\to\infty} \max_{1\leq i\leq m} \|\mathrm{Pr}_{\mathrm{SOC}_l}(s_i^{\iota(t)} + \bar{\eta}_{\iota(t)-1,i}\nabla_{s_i}\theta(z^{\iota(t)})) - s_i^{\iota(t)}\| = 0. \tag{5.8}$$

Now, define $h_t(\xi) = \max_{1\leq i\leq m} \|\mathrm{Pr}_{\mathrm{SOC}_l}(s_i^{\iota(t)} + \xi\nabla_{s_i}\theta(z^{\iota(t)}))) - s_i^{\iota(t)}\|$ for any $\xi \in \mathbb{R}$. Then Lemma 5.1 implies that, for any $b > a$, we have $h_t(a) \leq h_t(b) \leq (b/a)h_t(a)$ and thus

$$h_t(a) \leq h_t(b) \leq \frac{\overline{\eta}}{\underline{\eta}} h_t(a) \tag{5.9}$$

whenever $\underline{\eta} \leq a < b \leq \overline{\eta}$. This last relation together with (5.8) implies that $\lim_{t\to\infty} \|d^{\iota(t)}\| = 0$. Repeating this argument, we obtain that $\lim_{t\to\infty} \|d^{\iota(t)+l}\| = 0$ for any $1 \leq l \leq M$. Hence, we have proved that $\lim_{t\to\infty} \|d^t\| = 0$. The conclusion that any cluster point of $\{z^t\}$ is a maximizer of (DSDP) now follows from the definition of d^t and standard arguments. As **(CQ II)** holds, Theorem 4.1 gives us that $\bar{X} := \mathrm{Pr}_{\mathcal{S}_+^n}(G - \sum_{i=1}^m \mathcal{P}_i^*(\bar{s}_i) - \mathcal{B}^*(\bar{\mu}))$ is a minimizer of (RSDP).

Next, we show that $\{X^t\}$ is bounded and $\lim_{t\to\infty} \theta(z^t)$ exists. First, from **(CQ II)**, there exists an index set $J \subseteq \{1, \ldots, p\}$ such that the set $\{X \succeq 0 : (\mathcal{B}(X))_J = h_J\}$ is compact, which implies that $\{X \succeq 0 : h_J - \epsilon e \leq (\mathcal{B}(X))_J \leq h_J + \epsilon e\}$ is compact for any $\epsilon > 0$, where $e \in \mathbb{R}^{|J|}$ is the vector whose elements are all one. Since $\lim_{t\to\infty} \|d^t\| = 0$ implies that $\mathcal{B}(X^t) - h \to 0$, this implies that the sequence $\{X^t\}$ is bounded as claimed. Next, recall from the arguments preceding (5.6) that we have $\lim_{t\to\infty} \theta(z^{\iota(t)}) = \bar{\theta}$ for some $\bar{\theta} \in \mathbb{R}$. We now show further that $\lim_{t\to\infty} \theta(z^t) = \bar{\theta}$. To this end, note that $\nabla\theta$ is Lipschitz continuous. Hence, using Taylor's inequality and concavity of θ, we have

$$|\theta(z^{\iota(t)}) - \theta(z^{\iota(t)+1})| \leq \|\nabla\theta(z^{\iota(t)})\|\|z^{\iota(t)+1} - z^{\iota(t)}\| + \frac{L_\theta}{2}\|z^{\iota(t)+1} - z^{\iota(t)}\|^2, \tag{5.10}$$

where L_θ is the Lipschitz constant of $\nabla\theta$. Combining the boundedness of $\{X^t\}$ and the definition of $\nabla\theta$, we conclude that the sequence $\{\nabla\theta(z^t)\}$ is bounded. Using this fact, $\lim_{t\to\infty} \|d^t\| = 0$, $\lim_{t\to\infty} \theta(z^{\iota(t)}) = \bar{\theta}$ and (5.10), we obtain that

$$\lim_{t\to\infty} \theta(z^{\iota(t)+1}) = \lim_{t\to\infty} \theta(z^{\iota(t)}) = \bar{\theta}.$$

Similarly, one can show that $\lim_{t\to\infty} \theta(z^{\iota(t)+l}) = \bar\theta$ for any $1 \le l \le M$. Combining this with the definition of $\iota(t)$, we conclude that $\lim_{t\to\infty} \theta(z^t) = \bar\theta$.

Finally, let us consider the case when $\{z^t\}$ does not have any cluster point. We see from the previous paragraph that any cluster point $\bar X$ of $\{X^t\}$ satisfies $\mathcal{B}(\bar X) = h$. We now show further that $\mathcal{P}_i(\bar X) - q_i \in -\mathrm{SOC}_l$ for all $i = 1, ..., m$. To this end, notice from the definition of d^t and Lemma 5.1 that for each $i = 1, ..., m$,

$$c_i^t := \mathrm{Pr}_{\mathrm{SOC}_l}(s_i^t + \bar\eta \nabla_{s_i}\theta(z^t)) - s_i^t \to 0.$$

Moreover, from the definition of c_i^t, we have that

$$\begin{aligned} s_i^t + c_i^t &= \mathrm{Pr}_{\mathrm{SOC}_l}(s_i^t + \bar\eta \nabla_{s_i}\theta(z^t)) \\ &= \mathrm{Pr}_{\mathrm{SOC}_l}(s_i^t + c_i^t + \bar\eta \nabla_{s_i}\theta(z^t) - c_i^t) \\ \Rightarrow \bar\eta \nabla_{s_i}\theta(z^t) - c_i^t &\in N_{\mathrm{SOC}_l}(s_i^t + c_i^t). \end{aligned}$$

Since the normal cone $N_{\mathrm{SOC}_l}(s_i^t + c_i^t)$ is contained in $-\mathrm{SOC}_l$, taking limit and using the definition of $\bar\eta \nabla_{s_i}\theta(z^t)$, we conclude that $\mathcal{P}_i(\bar X) - q_i \in -\mathrm{SOC}_l$ for all i. This completes the proof. □

Remark 5.1. (Comments about the proof of convergence of SPG)

(1) *In Theorem 5.1* (i)*, we have shown that* $\lim_{t\to\infty} \|d^t\| = 0$ *whenever* $\limsup_{t\to\infty} \theta(z^t) < \infty$. *Indeed, a close inspection of the proof reveals that* $\lim_{t\to\infty} \|d^t\| = 0$ *if and only if* $\limsup_{t\to\infty} \theta(z^t) < \infty$ *when* **(CQ II)** *holds. To see this, suppose that* $\lim_{t\to\infty} \|d^t\| = 0$. *Then, the same proof for the boundedness of the primal iterate* $\{X^t\}$ *is valid and the same argument shows that any cluster point of* $\{X^t\}$ *is feasible for the primal problem. Thus,* $\{\theta(z^t)\}$ *has to be bounded from above by the weak duality.*

(2) *In Theorem 5.1* (i)*, we were not able to establish that any cluster point of* $\{X^t\}$ *is optimal when* $\{z^t\}$ *has no cluster points. Such a result is often related to the relationship between the Levitin-Polyak minimizing sequence and the naturally stationary sequence studied extensively in [10] and will be considered in future work.*

(3) (Sharper convergence for SPG under strict feasibility condition) *If the mapping* $\mathcal{B}^*$ *is injective, i.e., the matrices* $B_1, \ldots, B_p$ *are linearly independent and* **(CQ I)** *holds, then Remark 4.3 and Proposition 4.2 give us that, for each* $\eta \in \mathbb{R}$*, the upper level set* $\{z : \theta(z) \ge \eta\}$ *is compact. Let* $\{z^t\}$ *be generated by the SPG method. It is routine to show that* $\{z^t\}$ *is always bounded and any cluster point* $\bar z = (\bar s, \bar\mu)$ *of* $\{z^t\}$ *is a maximizer of* (DSDP) *and* $\bar X := \mathrm{Pr}_{\mathcal{S}_+^n}(G - \sum_{i=1}^m \mathcal{P}_i^*(\bar s_i) - \mathcal{B}^*(\bar\mu))$ *is a minimizer of* (RSDP).

Theorem 5.1 enables us to detect infeasibility of the robustification problem (RSDP). Indeed, under **(CQ II)**, the set $\{X \succeq 0 :\ (\mathcal{B}(X))_J = h_J\}$ is compact and nonempty. Thus, we have from Theorem 5.1 that (RSDP) is infeasible if and only if

$$\theta(z^t) > \max\left\{\frac{1}{2}\|X - G\|_F^2 :\ (\mathcal{B}(X))_J = h_J, X \succeq 0\right\} \tag{5.11}$$

for some t. We remark that this equivalence remains true if the maximum on the right hand side of (5.11) is replaced by any of its upper bounds. For infeasibility detection to be effective, an estimate of the rate of divergence of $\{\theta(z^t)\}$ is crucial. In the next proposition, we provide a rough estimate.

Proposition 5.1. *Suppose that* **(CQ II)** *holds. Then* (RSDP) *is infeasible if and only if*

$$\rho := \inf \left\{ \left\| (s,\mu) - \mathrm{Pr}_{\mathcal{F}}\left((s,\mu) + \underline{\eta} \nabla \theta(s,\mu) \right) \right\|^2 : (s,\mu) \in \mathcal{F} \right\} > 0, \tag{5.12}$$

where $\mathcal{F} := \mathrm{SOC}_l^m \times \mathbb{R}^p$ *is the dual feasible set. Moreover, let* $\{z^t\}$ *be generated by the SPG method. Then for all* $t \geq 1$*, we have*

$$\theta(z^t) \geq \theta(z^0) + \sigma \underline{\alpha} \delta \rho \left\lfloor \frac{t}{M} \right\rfloor, \tag{5.13}$$

with δ *and* $\underline{\alpha}$ *defined as in* (5.2) *and* (5.4)*, respectively, and* M *and* σ *are parameters of the SPG method.*

Proof. Suppose that the ρ defined as above is zero and let $\{(s^t, \mu^t)\}$ be a minimizing sequence for the infimum. Then it is easy to see that any cluster point of this sequence is a maximizer of (DSDP). Hence, if a cluster point exists, then by Theorem 4.1 (ii), the primal problem (RSDP) is feasible. Now, suppose that $\{(s^t, \mu^t)\}$ has no cluster point. Let $X^t := \mathrm{Pr}_{\mathcal{S}_+^n}\left(G - \sum_{i=1}^m \mathcal{P}_i^*(s_i^t) - \mathcal{B}^*(\mu^t)\right)$. As $\rho = 0$, we have $\nabla_\mu \theta(s^t, \mu^t) = \mathcal{B}(X^t) - h \to 0$. Proceeding as in the proof of Theorem 5.1 (i) shows that the sequence $\{X^t\}$ is bounded and has a cluster point that is feasible for (RSDP).

Conversely, suppose that (RSDP) is feasible and let $\{z^t\}$ be generated from the SPG method. Then by weak duality, $\limsup_{t\to\infty} \theta(z^t)$ is finite. Hence, Theorem 5.1 (i) asserts that $\lim_{t\to\infty} \|d^t\| = 0$, which together with the definition of ρ gives $\rho = 0$. This proves the first part of the theorem.

Now, for each $t \geq 1$, in view of (5.6), (5.2) and (5.4), we have

$$\theta(z^{\iota(t)}) \geq \theta(z^{\iota(\iota(t)-1)}) + \sigma \underline{\alpha} \delta \|d^{\iota(t)-1}\|^2$$

Combining this inequality with (5.9), the definition of ρ and $\|d^{\iota(t)-1}\|$, we obtain

$$\theta(z^{\iota(t)}) \geq \theta(z^{\iota(\iota(t)-1)}) + \sigma \underline{\alpha} \delta \rho,$$

from which the inequality (5.13) follows by summing both sides of the inequality along a suitable subsequence of $\{\theta(z^t)\}$ and invoking the definition of $\iota(t)$. □

It is worth noting that, in general, it is very challenging to compute the quantity ρ directly. Nevertheless, from the above proposition, we can see that the quantity ρ in (5.12) can be viewed as a measure of infeasibility which is of theoretical interest. Roughly speaking, the lower bound in (5.13) increases faster when (RSDP) is less likely feasible, which could suggest that $\{\theta(z^t)\}$ increases faster when (RSDP) is less likely feasible. This phenomenon is indeed observed in our numerical experiments in the next section.

6 Robust correlation stress testing and numerical simulations

In this section, we perform numerical experiments to demonstrate the effectiveness of the SPG method in solving instances of (RSDP). In particular, we consider problem instances arising from *robust* correlation stress testing. We will describe this application in detail in Section 6.1. Numerical results concerning real data from the CRSP data base is presented in Section 6.2.

6.1 Robust correlation stress testing

Consider a portfolio of n stocks with mean daily returns $r = (r_1, \ldots, r_n)^T$. In the context of risk management, we usually assume $r \approx (0, \ldots, 0)^T$. We assume that the variance of return for stock i is σ_i^2 with correlation matrix C. Define $D \in \mathbb{R}^{n\times n}$ to be a diagonal matrix with $d_{ii} = \sigma_i$ and denote by w the weight vector so that the portfolio on stock i is w_i. Then the $100(1-\alpha)\%$ Value-at-Risk (VaR) of the portfolio return is given by $-Z_\alpha\sqrt{w^TDCDw}$ where Z_α is the $100(1-\alpha)$th percentile of the standard normal random variable [17]. In correlation stress testing, we often need to modify values of the correlation matrix to reflect some plausible scenarios. However, after the modification, the resulting matrix may no longer be a valid correlation matrix. It is hence natural to find a correlation matrix having the specified entries, staying close to the given correlation matrix C. Among all matrices being close to the hypothetical correlation matrix needed for stress testing, we achieve the objective of stress testing by finding the largest VaR among all such matrices by solving the following optimization problem:

$$\begin{array}{ll} \min\limits_{X\in\mathcal{S}^n} & -\lambda Z_\alpha^2 w^TDXDw + \dfrac{1}{2}\|X - C\|_F^2 \\ \text{s.t.} & X_{ii} = 1,\ i = 1, ..., n, \\ & X_{ij} = h_{ij},\ (i,j) \in S, \\ & X \succeq 0, \end{array}$$

where $S \subseteq \{(i,j) :\ 1 \le i < j \le n\}$ is a set of indices, h_{ij} are the modified correlation between the ith stock and the jth stock and λ is a tuning parameter trading off the proximity to the given matrix C and stress testing portfolio.

Moreover, in reality, it is likely that the portfolio is changed in the near future and the risk manager wants to ensure that this framework can stress test over all such portfolios. More specifically, we would like to guarantee that

$$u^TDXDu \ge w^TDCDw \quad \forall u,\ e^Tu = 1,\ u \text{ "close" to } w. \tag{6.1}$$

Suppose that we allow a possible change of up to k stocks and let Q be the $k \times (k-1)$ matrix whose columns form an orthonormal basis of the nullspace of $\{y :\ e^Ty = 0\}$. Then we can quantify proximity of u to w by requiring that $u = w + \tilde{Q}v$ for some $v \in \mathbb{R}^{k-1}$ with norm less than or equal to ϵ, where $\tilde{Q}$ is the $n \times (k-1)$ matrix whose submatrix formed by the rows indexed by the k stocks equal Q, and is zero otherwise. Thus, we can rewrite (6.1) as

$$w^TDXDw + 2\epsilon w^TDXD\tilde{Q}v + \epsilon^2 v^T\tilde{Q}^TDXD\tilde{Q}v \ge w^TDCDw \quad \forall\|v\| \le 1. \tag{6.2}$$

While it is well-known that this last relation is LMI representable (see, for example, [1]), such a reformulation gives rise to an additional positive semidefinite constraint of size k by k, which is expensive from the computational point of view when $k = \Omega(n)$. Hence, we consider an alternative approach. To this end, notice that the relation (6.2) is always implied by

$$w^TDXDw + 2\epsilon w^TDXD\tilde{Q}v \ge w^TDCDw \quad \forall\|v\| \le 1. \tag{6.3}$$

Moreover, we have $w^TDXD\tilde{Q}v = \operatorname{tr}(XD\tilde{Q}vw^TD)$, and that

$$\operatorname{vec}(2\epsilon D\tilde{Q}vw^TD) = \underbrace{2\epsilon(Dw) \otimes (D\tilde{Q})}_{M} v.$$

Thus, if we define the set Υ as in (1.1) by letting $A^{(0)} = -\frac{Dww^TD}{w^TDCDw}$, $b^0 = -1$ and $b^{(i)} = 0$ for $i \ge 1$, and $A^{(i)} = -\frac{\hat{M}_i}{w^TDCDw}$, where $\hat{M}_i$ is the symmetric matrix obtained from a symmetrization of the

$n \times n$ matrix that forms the ith column of M by stacking the columns, we obtain the following optimization problem, whose first constraint is equivalent to (6.3), a condition that *implies* (6.2):

$$\begin{array}{ll} \min\limits_{X \in \mathcal{S}^n} & -\lambda Z_\alpha^2 w^T DXDw + \frac{1}{2}\|X - C\|_F^2 \\ \text{s.t.} & \mathcal{A}(X) \le b, \ \forall (\mathcal{A}, b) \in \Upsilon, \\ & X_{ii} = 1, \ i = 1, ..., n, \\ & X_{ij} = h_{ij}, \ (i,j) \in S, \\ & X \succeq 0, \end{array} \tag{6.4}$$

where S, h_{ij}, λ and Υ are defined as above. Let $G = C + \lambda Z_\alpha^2 Dww^T D$. It is easy to see that (6.4) is equivalent to

$$\begin{array}{ll} \min\limits_{X \in \mathcal{S}^n} & \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t.} & \mathcal{A}(X) \le b, \ \forall (\mathcal{A}, b) \in \Upsilon, \\ & X_{ii} = 1, \ i = 1, ..., n, \\ & X_{ij} = h_{ij}, \ (i,j) \in S, \\ & X \succeq 0, \end{array} \tag{6.5}$$

which is an instance of (RSDP). The robustness constraint in (6.4) addresses the issue of untimely recording of the portfolio holdings, which is especially useful when one needs to convert a 1-day VaR into a several-day VaR.

Remark 6.1. *As an alternative approach to handle correlation stress testing, one may consider the following weighted least square model:*

$$\begin{array}{ll} \min\limits_{X \in S^n} & \sum\limits_{(i,j) \in S} W_{ij}(X_{ij} - h_{ij})^2 - \lambda Z_\alpha^2 w^T DXDw + \frac{1}{2}\|X - C\|_F^2 \\ \text{s.t.} & X_{ii} = 1, \ i = 1, \ldots, n, \\ & X \succeq 0, \end{array}$$

where $W_{ij} > 0$, $(i,j) \in S$, are weights enforced on the stressed correlations. Note that the equality constraint $X_{ij} = h_{ij}$ in (RSDP) *has been replaced by the quadratic penalty term $W_{ij}(X_{ij} - h_{ij})^2$ in the above problem. Hence, one cannot readily specify known correlations in this approach. However, as a trade-off, the resulting optimization problem is always feasible.*

Following our discussion in Section 6.1, we have the following robust counterpart to address the issue of untimely recording of the portfolio holdings:

$$\begin{array}{ll} \min\limits_{X \in \mathcal{S}^n} & \sum\limits_{(i,j) \in S} W_{ij}(X_{ij} - h_{ij})^2 + \frac{1}{2}\|X - G\|_F^2 \\ \text{s.t.} & \mathcal{A}(X) \le b, \ \forall (\mathcal{A}, b) \in \Upsilon, \\ & X_{ii} = 1, \ i = 1, ..., n, \\ & X \succeq 0, \end{array} \tag{6.6}$$

where $G = C + \lambda Z_\alpha^2 Dww^T D$. Upon a suitable linear transformation, (6.6) *can be viewed as an instance of* (RSDP). *Further study of this alternative model would be one of our future research directions.*

6.2 Test results on data from CRSP database

In the implementation of our SPG method in this section, we set $\beta = 0.5$, $\sigma = 1e-4$, $\underline{\eta} = 1e-8$, $\overline{\eta} = 1e+5$ and $M = 200$. We initialize at the origin, and terminate the algorithm when one of the following three criteria is met:

Case 1: When $\alpha_t < 1e-15$ or the number of iterations hits 25000. In this case, we terminate early.

Case 2: When $\theta(z^t) > \frac{1}{2}(n + \|G\|_F)^2$. Notice that the latter quantity is an upper bound of the objective function of (6.5) constrained over the set of all correlation matrices. Thus, from the discussion preceding (5.11), we identify that (6.5) is likely infeasible.

Case 3: When

$$\frac{\left|\frac{1}{2}\|X^t - G\|_F^2 - \max_{s \leq t} \theta(z^s)\right|}{\max\left\{\frac{1}{2}\|X^t - G\|_F^2, 1\right\}} < 1e - 7,$$

and

$$\max\left\{\|\text{diag}(X^t) - e\|_F, \|(X^t - h)_S\|_F, \text{dist}(\mathcal{P}_i(X^t) - q_i, -\text{SOC}_l) : \ i = 1, ...m\right\} < 1e - 9,$$

where X^t is defined in (5.3). In this case, we obtain an approximate optimal solution to (6.5).

All numerical experiments are performed on an SGI XE340 system, with two 2.4 GHz quad-core Intel E5620 Xeon 64-bit CPUs and 48 GB RAM, equipped with SUSE Linux Enterprise server 11 SP1 and Matlab 7.14 (R2012a).

We collect daily stock return data from 2007 to 2011 from the CRSP database. Since survivorship bias is not our concern, we filter out all the stocks with missing data in this period. This results in a set of 4248 stocks with 1260 of daily returns. We randomly select two sets of stocks of size $n = 150$ and 250 respectively, and compute the corresponding covariance matrix, from which the diagonal matrix D and the correlation matrix C are obtained. We are interested in the VaR at 99.5% confidence level, i.e., $Z_\alpha \approx 2.575829$.

In our tests below, we consider three different weights λ corresponding to $\lambda Z_\alpha^2 = 10n$, $100n$ and $1000n$, and assume that the current portfolio w is e/n, i.e., it is equally weighted. We randomly select $k = \frac{n}{2}$ stocks, which means that we allow half of the stock holdings to be adjusted in the near future. We then randomly select the index set $S \subseteq \{(i,j) : \ 1 \leq j < i \leq n\}$ and set the corresponding h_{ij} to be 0.9. Finally, with these fixed k stocks and fixed S, we consider different λ and ϵ. To benchmark performance of our algorithm, we also solve the corresponding formulation (3.2) using the standard primal-dual interior point solver SDPT3 (version 4.0) [26, 30] with the default tolerance. We use the precompiled mex files within the package available on the software webpage.

In our first test, we set $|S| = \lceil\frac{n(n-1)}{200}\rceil$ and consider $\epsilon = 0.02,\ 0.04,\ 0.06,\ 0.08$. The computational results are reported in Tables 1 and 2, where we report the CPU time in seconds (cpu), number of iterations (iter), approximate primal optimal value (fval), primal infeasibility of the approximate solution X^* at termination (pfeas) defined as

$$\max\left\{-\lambda_{\min}(X^*), \|\text{diag}(X^*) - e\|_F, \|(X^* - h)_S\|_F, \text{dist}(\mathcal{P}_i(X^*) - q_i, -\text{SOC}_l) : \ i = 1, ...m\right\},$$

the difference $\|X^* - C\|_F$ (ndiff) and the VaR $Z_\alpha\sqrt{w^T D X^* D w}$ at termination (VaR). We observe that the ndiff generally increases with VaR for a fixed tradeoff parameter λ. Moreover, the VaR increases with λ. Finally, we observe that our algorithm is faster than SDPT3 while producing solutions of similar qualities.

In our second test, we consider a larger percentage of fixed entries, namely $|S| = \lceil\frac{n(n-1)}{40}\rceil$. In our experiments, we observe that the robustness constraint is only in effect for larger ϵ and thus we report results for $\epsilon = 0.05,\ 0.1,\ 0.5,\ 1$. The computational results are reported in Tables 3 and 4. We report similar quantities as before. In addition, when SDPT3 returns infeasibility while solving (3.2), we report also in parenthesis the time taken for SDPT3 to solve (3.2) but with 0 as objective; i.e., the time taken for SDPT3 to solve the associated feasibility problem. From the

λZ_α^2	ϵ	SPG cpu/iter/fval/pfeas/ndiff/VaR	SDPT3 cpu/fval/pfeas/ndiff/VaR
1500	0.02	22.5/1277/1.6253e+2/9.9e-10/1.80339e+1/4.82055e-2	436.3/1.6253e+2/8.7e-11/1.80339e+1/4.82055e-2
	0.04	20.3/1140/2.0073e+2/9.9e-10/2.00430e+1/5.05270e-2	429.9/2.0073e+2/3.8e-10/2.00430e+1/5.05270e-2
	0.06	27.3/1499/2.5851e+2/9.9e-10/2.27441e+1/5.10933e-2	406.0/2.5851e+2/2.6e-10/2.27441e+1/5.10933e-2
	0.08	37.0/2081/3.1371e+2/6.3e-10/2.50538e+1/5.08811e-2	395.6/3.1371e+2/7.3e-11/2.50538e+1/5.08811e-2
15000	0.02	22.5/1218/1.6185e+2/8.8e-10/1.80340e+1/4.82160e-2	445.8/1.6185e+2/1.5e-10/1.80340e+1/4.82160e-2
	0.04	23.3/1347/1.9958e+2/8.3e-10/2.00432e+1/5.05528e-2	406.5/1.9958e+2/5.9e-10/2.00432e+1/5.05528e-2
	0.06	28.7/1612/2.5724e+2/9.2e-10/2.27443e+1/5.11247e-2	400.9/2.5724e+2/2.3e-10/2.27443e+1/5.11247e-2
	0.08	35.2/2007/3.1248e+2/9.2e-10/2.50539e+1/5.09135e-2	393.2/3.1248e+2/9.9e-11/2.50539e+1/5.09135e-2
150000	0.02	19.2/1062/1.5578e+2/9.6e-10/1.80413e+1/4.83243e-2	369.3/1.5578e+2/3.0e-10/1.80413e+1/4.83243e-2
	0.04	22.1/1208/1.8865e+2/4.0e-10/2.00601e+1/5.08197e-2	381.4/1.8865e+2/7.3e-10/2.00601e+1/5.08197e-2
	0.06	28.1/1524/2.4506e+2/9.5e-10/2.27625e+1/5.14472e-2	401.6/2.4506e+2/3.1e-10/2.27625e+1/5.14472e-2
	0.08	37.4/2109/3.0074e+2/9.0e-10/2.50709e+1/5.12462e-2	462.7/3.0074e+2/3.9e-10/2.50709e+1/5.12462e-2

Table 1: Computational results for real data with 150 stocks, $|S| = \lceil \frac{n(n-1)}{200} \rceil$.

λZ_α^2	ϵ	SPG cpu/iter/fval/pfeas/ndiff/VaR	SDPT3 cpu/fval/pfeas/ndiff/VaR
2500	0.02	121.0/3129/9.4124e+2/7.9e-10/4.33930e+1/5.11321e-2	4724.9/9.4124e+2/1.4e-10/4.33930e+1/5.11321e-2
	0.04	135.9/3599/1.0189e+3/9.5e-10/4.51498e+1/5.33138e-2	5395.0/1.0189e+3/6.2e-10/4.51498e+1/5.33138e-2
	0.06	157.2/4166/1.2712e+3/5.3e-10/5.04299e+1/5.44995e-2	5859.1/1.2712e+3/2.8e-10/5.04299e+1/5.44995e-2
	0.08	231.3/5994/1.5685e+3/8.3e-10/5.60149e+1/5.47685e-2	5392.4/1.5685e+3/4.8e-10/5.60149e+1/5.47685e-2
25000	0.02	147.9/2547/9.3914e+2/1.0e-09/4.33931e+1/5.11606e-2	4733.5/9.3914e+2/2.0e-10/4.33931e+1/5.11606e-2
	0.04	206.7/3599/1.0161e+3/8.2e-10/4.51499e+1/5.33279e-2	5712.1/1.0161e+3/2.1e-10/4.51499e+1/5.33279e-2
	0.06	275.9/4802/1.2679e+3/9.9e-10/5.04300e+1/5.45143e-2	5918.2/1.2679e+3/3.8e-10/5.04300e+1/5.45143e-2
	0.08	285.1/4878/1.5651e+3/8.4e-10/5.60149e+1/5.47830e-2	5619.5/1.5651e+3/1.4e-10/5.60149e+1/5.47830e-2
250000	0.02	152.8/2673/9.1852e+2/9.9e-10/4.34071e+1/5.14461e-2	4752.3/9.1852e+2/7.5e-11/4.34071e+1/5.14461e-2
	0.04	209.6/3655/9.8800e+2/9.5e-10/4.51570e+1/5.34716e-2	5195.3/9.8800e+2/1.1e-10/4.51570e+1/5.34716e-2
	0.06	244.3/4247/1.2355e+3/8.9e-10/5.04367e+1/5.46643e-2	5868.3/1.2355e+3/3.1e-10/5.04367e+1/5.46643e-2
	0.08	358.6/6380/1.5316e+3/9.1e-10/5.60209e+1/5.49291e-2	5704.1/1.5316e+3/1.0e-10/5.60209e+1/5.49291e-2

Table 2: Computational results for real data with 250 stocks, $|S| = \lceil \frac{n(n-1)}{200} \rceil$.

tables, we still observe that the ndiff generally increases with VaR for a fixed tradeoff parameter λ and the VaR increases with λ. Furthermore, we also observe that our algorithm is faster than SDPT3 in finding an approximate solution and in identifying infeasibility except when $\epsilon = 0.1$. Moreover, note that our method is slower than SDPT3 when the latter only solves the feasibility problem, but the difference in time narrows as dimension increases for $\epsilon = 0.5, 1$. Finally, we note that our method tends to be faster in certifying infeasibility when the problem is less likely feasible.

7 Conclusion and future research

In this paper, we study a least square semidefinite programming problem under ellipsoidal uncertainty. We show that the problem can be reformulated as a least square semidefinite linear programming problem with an additional second-order cone constraint and is thus tractable. We also provide an explicit quantitative sensitivity analysis on how the solution under the robustification depends on the size/shape of the ellipsoidal data uncertainty set. We present two constraint qualifications under which the reformulation has zero duality gap with its dual, one of which does not require feasibility of the primal problem. We then adapt the spectral projected gradient method [7] to solve the dual problem. While it is folk-lore that any cluster point of the sequence generated is a dual optimal solution, we establish under suitable constraint qualifications that the

λZ_α^2	ϵ	SPG cpu/iter/fval/pfeas/ndiff/VaR	SDPT3 cpu/fval/pfeas/ndiff/VaR
1500	0.05	25.4/1388/2.7610e+3/9.6e-10/7.43212e+1/7.54233e-2	337.7/2.7610e+3/5.5e-11/7.43212e+1/7.54233e-2
	0.10	219.7/12084/2.9771e+3/9.8e-10/7.71747e+1/7.60129e-2	337.6/2.9771e+3/5.8e-11/7.71747e+1/7.60129e-2
	0.50	Infeasible: cpu/iter = 16.0/848	Infeasible: cpu = 488.0 (4.8)
	1.00	Infeasible: cpu/iter = 10.9/609	Infeasible: cpu = 486.2 (5.2)
15000	0.05	27.0/1506/2.7535e+3/7.5e-10/7.43212e+1/7.54252e-2	321.7/2.7535e+3/1.0e-09/7.43212e+1/7.54252e-2
	0.10	238.2/13047/2.9694e+3/9.4e-10/7.71747e+1/7.60146e-2	331.2/2.9694e+3/6.6e-11/7.71747e+1/7.60146e-2
	0.50	Infeasible: cpu/iter = 13.9/751	Infeasible: cpu = 490.2 (4.7)
	1.00	Infeasible: cpu/iter = 11.1/643	Infeasible: cpu = 492.5 (4.9)
150000	0.05	22.0/1242/2.6790e+3/3.6e-10/7.43217e+1/7.54441e-2	335.3/2.6790e+3/1.2e-09/7.43217e+1/7.54441e-2
	0.10	219.0/12001/2.8931e+3/8.8e-10/7.71751e+1/7.60318e-2	330.1/2.8931e+3/1.6e-11/7.71751e+1/7.60318e-2
	0.50	Infeasible: cpu/iter = 19.1/1075	Infeasible: cpu = 496.8 (5.0)
	1.00	Infeasible: cpu/iter = 10.3/539	Infeasible: cpu = 487.4 (5.0)

Table 3: Computational results for real data with 150 stocks, $|S| = \lceil \frac{n(n-1)}{40} \rceil$.

λZ_α^2	ϵ	SPG cpu/iter/fval/pfeas/ndiff/VaR	SDPT3 cpu/fval/pfeas/ndiff/VaR
2500	0.05	66.4/1642/9.1282e+3/5.7e-10/1.35129e+2/7.94189e-2	3963.1/9.1282e+3/1.2e-10/1.35129e+2/7.94189e-2
	0.10	Infeasible: cpu/iter = 818.0/21798	Infeasible: cpu = 5664.5 (26.0)
	0.50	Infeasible: cpu/iter = 40.8/689	Infeasible: cpu = 5873.4 (27.1)
	1.00	Infeasible: cpu/iter = 37.3/626	Infeasible: cpu = 5838.1 (24.3)
25000	0.05	92.1/1555/9.1136e+3/5.3e-10/1.35129e+2/7.94196e-2	3984.9/9.1136e+3/1.9e-10/1.35129e+2/7.94196e-2
	0.10	Infeasible: cpu/iter = 1169.2/20276	Infeasible: cpu = 6257.5 (27.6)
	0.50	Infeasible: cpu/iter = 29.7/490	Infeasible: cpu = 5975.2 (24.3)
	1.00	Infeasible: cpu/iter = 12.8/198	Infeasible: cpu = 5821.3 (23.4)
250000	0.05	86.3/1492/8.9683e+3/7.1e-10/1.35129e+2/7.94265e-2	4003.6/8.9683e+3/2.6e-10/1.35129e+2/7.94265e-2
	0.10	Infeasible: cpu/iter = 1248.7/21423	Infeasible: cpu = 6218.9 (26.1)
	0.50	Infeasible: cpu/iter = 51.7/898	Infeasible: cpu = 5825.8 (26.1)
	1.00	Infeasible: cpu/iter = 25.8/430	Infeasible: cpu = 5874.0 (23.9)

Table 4: Computational results for real data with 250 stocks, $|S| = \lceil \frac{n(n-1)}{40} \rceil$.

dual objective value along the sequence generated converges to a finite value if and only if the primal problem is feasible. Finally, we discuss robust correlation stress testing as an application of our model.

In discussing the mathematical model for the robust correlation stress testing, we have relaxed the problem by replacing the positive semidefinite constraint with a second-order cone constraint, and hence yielding a problem in the form of (RSDP). It is certainly interesting to study the performance and implication of the unrelaxed (exact) model in the future. Furthermore, while the rate of convergence of steepest-descent-type algorithms is relatively well-known, the rate of "divergence", which is essential in the effectiveness of certifying primal infeasibility by our algorithm, is not well understood. Although the constant ρ in (5.12) can be viewed as a measure of infeasibility, it seems to be a very rough under estimate. Thus, it would be interesting to better understand ρ and derive a tighter lower bound to $\|d^t\|^2$ for all t. This is another interesting future research direction.

A Proof of Lemma 2.1

Proof. By a translation, we may assume without loss of generality that $X_0 = 0$. For a closed convex set Ω, let σ_Ω denote the usual support function of Ω (that is, $\sigma_\Omega(X) = \sup_{Y\in\Omega} \operatorname{tr}(YX)$ for all $X \in \mathcal{S}^n$) and let epiσ_Ω be the epigraph of the support function. Now, notice that if $(Y_1, \alpha_1) \in$

epi σ_{Ω_1} and $(Y_2, \alpha_2) \in$ epi σ_{Ω_2} are such that $\alpha_1 + \alpha_2 = \sigma_{\Omega_1 \cap \Omega_2}(Y_1 + Y_2)$ and $\|Y_1 + Y_2\|_F \leq 1$, then we have $0 \leq \sigma_{\Omega_1}(Y_1) \leq \alpha_1$ and $\alpha_2 \leq \sigma_{\Omega_1 \cap \Omega_2}(Y_1 + Y_2) \leq R$. From these we obtain further that $\delta \|Y_2\|_F \leq \sigma_{\Omega_2}(Y_2) \leq \alpha_2 \leq R$ and consequently

$$\|Y_1\|_F \leq \|Y_1 + Y_2\|_F + \|Y_2\|_F \leq 1 + \frac{R}{\delta}.$$

The proof of this lemma now follows similarly as in [18, Lemma 4.10], making use of the bounds derived above. □

References

[1] A. Ben-Tal, L. El Ghaoui and A. Nemirovski. *Robust Optimization.* Princeton Series in Applied Mathematics, Princeton, NJ (2009).

[2] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Program.* 88, pp. 411–424 (2000).

[3] A. Ben-Tal and A. Nemirovski. Selected topics in robust convex optimization. *Math. Program.* 112, pp. 125–158 (2008).

[4] J. Berkowitz. A coherent framework for stress testing. *J. Risk* 2(2), pp. 5–15 (1999).

[5] D. Bertsimas and D. B. Brown. Constructing uncertainty sets for robust linear optimization. *Oper. Res.* 57, pp. 1483–1495 (2009).

[6] D. Bertsimas, D. B. Brown and C. Caramanis. Theory and applications of robust optimization. *SIAM Rev.* 53, pp. 464–501 (2011).

[7] E. G. Birgin, J. M. Martínez and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.* 10, pp. 1196–1211 (2000).

[8] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer, Berlin (1997).

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

[10] C.-C. Chou, K.-F. Ng and J.-S. Pang. Minimizing and stationary sequences of constrained optimization problems. *SIAM J. Control Optim.* 36, pp. 1908–1936 (1998).

[11] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.* 18, pp. 1035–1064 (1997).

[12] Y. Gao and D. Sun. Calibrating least squares semidefinite programming with equality and inequality constraints. *SIAM J. Matrix Anal. Appl.* 31, pp. 1432–1457 (2009).

[13] M. Fukushima, Z.-Q. Luo and P. Tseng. Smoothing functions for second-order-cone complementarity problems. *SIAM J. Optim.* 12, pp. 436–460 (2001).

[14] E. M. Gafni and D. P. Bertsekas. Two metric projection methods for constrained optimization. *SIAM J. Control Optim.* 22, pp. 936–964 (1984).

[15] H. Hu. Geometric condition measures and smoothness condition measures for closed convex sets and linear regularity of infinitely many closed convex sets. *J. Optim. Theory Appl.* 126, pp. 287–308 (2005).

[16] J. Kim and C. C. Finger. A stress test to incorporate correlation breakdown. *J. Risk* 2(3), pp. 5–19 (2000).

[17] P. H. Kupiec. Stress testing in a value at risk framework. *J. Deriv.* 6, pp. 7–24 (1998).

[18] C. Li, K. F. Ng and T. K. Pong. The SECQ, linear regularity, and the strong CHIP for an infinite system of closed convex sets in normed linear spaces. *SIAM J. Optim.* 18, pp. 643–665 (2007).

[19] Z. Lu and Y. Zhang. An augmented Lagrangian approach for sparse principal component analysis. *Math. Program.* 135, pp. 149–193 (2012).

[20] H. Qi and D. Sun. Correlation stress testing for value-at-risk: an unconstrained convex optimization approach. *Comput. Optim. Appl.* 45, pp. 427–462 (2010).

[21] R. Rebonato and P. Jäckel. The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *J. Risk* 2(2), pp. 17–27 (1999).

[22] S. M. Robinson. An application of error bounds for convex programming in a linear space. *SIAM J. Control* 13, pp. 271–273 (1975).

[23] R. T. Rockafellar. *Convex Analysis.* Princeton University Press (1970).

[24] A. Shapiro, D. Dentcheva and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory.* SIAM, Philadelphia (2009).

[25] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper. Res.* 21, pp. 1154–1157 (1973).

[26] K. C. Toh, M. J. Todd and R. H. Tütüncü. SDPT3 — a Matlab software package for semidefinite programming. *Optim. Method Softw.* 11, pp. 545–581 (1999).

[27] Ph. L. Toint. Global convergence of a class of trust-region methods for nonconvex minimization in Hilbert space. *IMA J. Numer. Anal.* 8, pp. 231–252 (1988).

[28] P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.* 117, pp. 387–423 (2007).

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

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