CORC REPORT TR-2002-04

Robust convex quadratically constrained programs*

D. Goldfarb †   G. Iyengar ‡

July 1st, 2002

Abstract

In this paper we study robust convex quadratically constrained programs, a subset of the class of robust convex programs introduced by Ben-Tal and Nemirovski [4]. Unlike [4], our focus in this paper is to identify uncertainty structures that allow the corresponding robust quadratically constrained programs to be reformulated as second-order cone programs. We propose three classes of uncertainty sets that satisfy this criterion and present examples where these classes of uncertainty sets are natural.

1 Problem formulation

A generic quadratically constrained program (QCP) is defined as follows.

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{x}^T\mathbf{Q}_i\mathbf{x} + 2\mathbf{q}_i^T\mathbf{x} + \gamma_i \leq 0, i = 1,\ldots,p, \end{array} \tag{1}$$

where the vector of decision variables $\mathbf{x} \in \mathbf{R}^n$, and the data $\mathbf{c} \in \mathbf{R}^n$, $\gamma_i \in \mathbf{R}$, $\mathbf{q}_i \in \mathbf{R}^n$ and $\mathbf{Q}_i \in \mathbf{R}^{n\times n}$, for all $i = 1,\ldots,p$. Note that without any loss of generality one may assume that the objective is linear. The QCP (1) is a convex optimization problem if and only if $\mathbf{Q}_i \succeq \mathbf{0}$ for all $i = 1,\ldots,p$, where $\mathbf{Q} \succeq \mathbf{0}$ denotes that the matrix $\mathbf{Q}$ is positive semidefinite.

Suppose $\mathbf{Q} \succeq \mathbf{0}$. Then $\mathbf{Q} = \mathbf{V}^T\mathbf{V}$ for some $\mathbf{V} \in \mathbf{R}^{m\times n}$ and the quadratic constraint $\mathbf{x}^T\mathbf{Q}\mathbf{x} \leq -(2\mathbf{q}^T\mathbf{x} + \gamma)$ is equivalent to the second-order cone (SOC) constraint [2, 18, 22]

$$\left\| \begin{bmatrix} 2\mathbf{V}\mathbf{x} \\ (1+\gamma+2\mathbf{q}^T\mathbf{x}) \end{bmatrix} \right\| \leq 1 - \gamma - 2\mathbf{q}^T\mathbf{x}. \tag{2}$$

*Submitted to *Math Programming, Series B.* Please do not circulate.

†IEOR Department, Columbia University, Email: gold@ieor.columbia.edu. Research partially supported by DOE grant GE-FG01-92ER-25126, and NSF grants DMS-94-14438, CDA-97-26385 and DMS-01-04282

‡IEOR Department, Columbia University, Email: garud@ieor.columbia.edu. Research partially supported by NSF grants CCR-00-09972 and DMS-01-04282.

Thus, the convex QCP (1) is equivalent to the following second-order cone program (SOCP)

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \left\| \begin{bmatrix} 2\mathbf{V}_i\mathbf{x} \\ (1+\gamma_i+2\mathbf{q}_i^T\mathbf{x}) \end{bmatrix} \right\| \leq 1-\gamma_i-2\mathbf{q}_i^T\mathbf{x}, \quad i=1,\ldots,p, \end{array} \tag{3}$$

where $\mathbf{Q}_i = \mathbf{V}_i^T\mathbf{V}_i$, $i=1,\ldots,p$. For a detailed discussion of SOCPs and their applications see [2, 18, 22].

The formulations (1) and (3) implicitly assume that the parameters defining the problem – $\{(\mathbf{Q}_i,\mathbf{q}_i,\gamma_i), i = 1\ldots,p\}$ – are known exactly. However, in practice these are estimated from data, and are, therefore, subject to measurement and statistical errors [16]. Since the solutions to optimization problems are typically sensitive to parameter fluctuations, the errors in the input parameters tend to get amplified in the decision vector. This, in turn, leads to a sharp deterioration in performance [3, 10].

The problem of choosing a decision vector in the presence of parameter perturbation was formalized by Ben-Tal and Nemirovski [4, 5] as follows:

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & F(\mathbf{x},\boldsymbol{\xi}) \in \mathcal{K} \subset \mathbf{R}^m, \qquad \forall \boldsymbol{\xi} \in \mathcal{U}, \end{array} \tag{4}$$

where $\boldsymbol{\xi}$ are the uncertain parameters, $\mathcal{U}$ is the uncertainty set, $\mathbf{x} \in \mathbf{R}^n$ is the decision vector, $\mathcal{K}$ is a convex cone and, for fixed $\boldsymbol{\xi} \in \mathcal{U}$, the function $F(\mathbf{x},\boldsymbol{\xi})$ is $\mathcal{K}$-concave [4, 6]. The optimization problem (4) is called a *robust optimization problem.* Ben-Tal and Nemirovski established that for certain classes of uncertainty sets $\mathcal{U}$, robust counterparts of linear programs, quadratic programs, quadratically constrained quadratic programs, and semidefinite programs are themselves tractable optimization problems. Robustness as applied to uncertain least squares problems and uncertain semidefinite programs was independently studied by El Ghaoui and his collaborators [11, 12].

In keeping with the formulation proposed by Ben-Tal and Nemirovski, a generic robust convex QCP is given by

$$\begin{array}{ll} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{x}^T\mathbf{Q}_i\mathbf{x} + 2\mathbf{q}_i^T\mathbf{x} + \gamma_i \leq 0, \text{for all } (\mathbf{Q}_i,\mathbf{q}_i,\gamma_i) \in \mathcal{S}_i, i=1,\ldots,p. \end{array} \tag{5}$$

Ben-Tal and Nemirovski [4] investigated a version of (5) in which the uncertainty structures $\mathcal{S}_i$ are generalized ellipsoids and showed that the resulting robust optimization problem can be reduced to a semidefinite program (SDP) [1, 22, 26]. In this paper we explore uncertainty structures for which the corresponding robust problems can be reformulated as SOCPs. Our interest in this class of structures stems from the fact that both the worst case and practical computational effort required to solve SOCPs is at least an order of magnitude less than that needed to solve SDPs [2].

The organization of the rest of this paper is as follows. In Section 2 we propose three classes of uncertainty sets for which a robust convex QCP can be reduced to an SOCP. In Section 3 we present several applications where the natural uncertainty structures are combinations of those presented in Section 2. Section 4 contains some concluding remarks.

2 Uncertainty structures

In this section we introduce three classes of uncertainty sets for which the robust convex QCP (5) can be reformulated as an SOCP.

2.1 Discrete and polytopic uncertainty sets

The simplest class of the uncertainty sets is a discrete set defined as follows.

$$\mathcal{S}_a = \big\{(\mathbf{Q}, \mathbf{q}, \gamma) : (\mathbf{Q}, \mathbf{q}, \gamma) = (\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \mathbf{Q}_j \succeq \mathbf{0}, j = 1, \ldots, k\big\}. \tag{6}$$

The robust constraint $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$ for all $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_a$ is equivalent to the k convex quadratic constraints

$$\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j \leq 0, \quad \forall j = 1, \ldots, k. \tag{7}$$

Thus, the resulting robust optimization problem is an SOCP.

The discrete uncertainty set (6) typically arises when one wants to be robust against several scenarios – each $(\mathbf{Q}_i, \mathbf{q}_i, \gamma_i)$ corresponds to a particular scenario [17]. A closely related related uncertainty structure is the polytopic uncertainty set defined as follows.

$$\mathcal{S}_b = \Big\{(\mathbf{Q}, \mathbf{q}, \gamma) : (\mathbf{Q}, \mathbf{q}, \gamma) = \sum_{j=1}^{k} \lambda_j(\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \mathbf{Q}_j \succeq \mathbf{0}, \lambda_j \geq 0, \forall j, \sum_{j=1}^{n} \lambda_j = 1\Big\}. \tag{8}$$

The robust constraint $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$, for all $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_b$ is equivalent to the constraint $\sum_j \lambda_j(\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j) \leq 0$, for all $\lambda_j \geq 0$, $\sum_j \lambda_j = 1$. The latter constraint is, in turn, equivalent to the set of constraints:

$$\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j \leq 0, \quad \forall j = 1, \ldots, k. \tag{9}$$

From (7) and (9), it follows that the SOCP reformulations of the robust problems corresponding to the discrete and the polytopic uncertainty sets are identical.

2.2 Affine uncertainty sets

Next, we propose two closely related affine uncertainty sets both of which are restricted versions of the generalized ellipsoidal uncertainty sets introduced by Ben-Tal and Nemirovski [4].

In the first uncertainty set $\mathcal{S}_c$, the parameter $(\mathbf{Q}, \mathbf{q}, \gamma)$ is affinely perturbed by a single set of perturbation parameters $\mathbf{u}$, i.e.

$$\mathcal{S}_c = \left\{(\mathbf{Q}, \mathbf{q}, \gamma) : \begin{array}{l} (\mathbf{Q}, \mathbf{q}, \gamma) = (\mathbf{Q}_0, \mathbf{q}_0, \gamma_0) + \sum_{j=1}^{k} u_i(\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \\ \mathbf{Q}_j \succeq \mathbf{0}, j = 0, \ldots, k, \mathbf{u} \geq \mathbf{0}, \|\mathbf{u}\| \leq 1 \end{array} \right\}. \tag{10}$$

Remark 1 *In this case, the robust problem (5) is NP-hard if the sign constraint on* $\mathbf{u}$ *is relaxed or if any of the* $\mathbf{Q}_j$*'s are indefinite [4].*

The following lemma shows that the robust convex quadratic constraint corresponding to $\mathcal{S}_c$ can be reformulated as a collection of SOC constraints.

Lemma 1 *The decision vector* $\mathbf{x} \in \mathbf{R}^n$ *satisfies the robust constraint* $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$ *for all* $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_c$*, where* $\mathcal{S}_c$ *is defined by (10), if and only if there exist* $\mathbf{f} \in \mathbf{R}^k_+$ *and* $\nu \geq 0$ *satisfying*

$$\begin{aligned} \left\| \begin{bmatrix} 2\mathbf{V}_j\mathbf{x} \\ 1 - f_j + \gamma_j + 2\mathbf{q}_j^T\mathbf{x} \end{bmatrix} \right\| &\leq 1 + f_j - \gamma_j - 2\mathbf{q}_j^T\mathbf{x}, \quad j = 1, \ldots, k, \\ \left\| \begin{bmatrix} 2\mathbf{V}_0\mathbf{x} \\ 1 - \nu \end{bmatrix} \right\| &\leq 1 + \nu, \\ \|\mathbf{f}\| &\leq -\nu - 2\mathbf{q}_0^T\mathbf{x} - \gamma_0, \end{aligned} \tag{11}$$

where $\mathbf{Q}_j = \mathbf{V}_j^T\mathbf{V}_j$*,* $j = 0, \ldots, k$*.*

Proof: The constraint $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$ for all $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_c$ is equivalent to

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \max_{\{\mathbf{u}:\mathbf{u}\geq\mathbf{0},\|\mathbf{u}\|\leq 1\}} \Big\{ \sum_{j=1}^k u_j(\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j) \Big\} \leq 0, \tag{12}$$

Let $\mathbf{f} \in \mathbf{R}^k$ with

$$f_j \geq \max\{\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j, 0\}, \quad j = 1, \ldots, k. \tag{13}$$

Then (12) holds if and only if there exists $\mathbf{f}$ satisfying (13) and

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \|\mathbf{f}\| \leq 0. \tag{14}$$

The result follows from rewriting (14) as a collection of linear and SOC constraints. ■

In the uncertainty structure $\mathcal{S}_c$ the perturbations in the quadratic term $\mathbf{Q}$ and the affine term $(\mathbf{q}, \gamma)$ is controlled by the same parameter $\mathbf{u}$. However, in many applications the uncertainty in the quadratic and affine terms are independent [16]. We model this situation by the following uncertainty structure,

$$\mathcal{S}_d = \left\{ (\mathbf{Q}, \mathbf{q}, \gamma) : \begin{array}{l} \mathbf{Q} = \mathbf{Q}_0 + \sum_{j=1}^k u_i\mathbf{Q}_j, \mathbf{Q}_j \succeq \mathbf{0}, j = 0, \ldots, k, \|\mathbf{u}\| \leq 1, \\ (\mathbf{q}, \gamma) = (\mathbf{q}_0, \gamma_0) + \sum_{j=1}^k v_i(\mathbf{q}_j, \gamma_j), \|\mathbf{v}\| \leq 1 \end{array} \right\}. \tag{15}$$

Remark 2 *Although we allow general* $\mathbf{u}$*, the constraint* $\mathbf{Q}_i \succeq \mathbf{0}$ *implies that the worst case perturbation* $\mathbf{u}^* \geq \mathbf{0}$*. As in Remark 1, allowing indefinite* $\mathbf{Q}_i$ *results in an NP-hard optimization problem [4].*

The following lemma establishes that the robust convex QCP corresponding to $\mathcal{S}_d$ can be reformulated as an SOCP.

Lemma 2 *The decision vector* $\mathbf{x} \in \mathbf{R}^n$ *satisfies the robust constraint* $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$ *for all* $(\mathbf{Q}, \mathbf{q}, \gamma) \in S_d$*, where* $\mathcal{S}_d$ *is defined by (15), if and only if there exist* $\mathbf{f}, \mathbf{g} \in \mathbf{R}^k$ *and* $\nu \geq 0$ *such that*

$$\begin{array}{rcll} g_j & = & 2\mathbf{q}_j^T\mathbf{x} + \gamma_j, & j = 1, \ldots, k, \\ \left\| \begin{bmatrix} 2\mathbf{V}_j\mathbf{x} \\ 1 - f_j \end{bmatrix} \right\| & \leq & 1 + f_j, & j = 1, \ldots, k, \\ \left\| \begin{bmatrix} 2\mathbf{V}_0\mathbf{x} \\ 1 - \nu \end{bmatrix} \right\| & \leq & 1 + \nu, & \\ \|\mathbf{f}\| + \|\mathbf{g}\| & \leq & -\nu - 2\mathbf{q}_0^T\mathbf{x} - \gamma_0, & \end{array} \tag{16}$$

where $\mathbf{Q}_j = \mathbf{V}_j^T\mathbf{V}_j$*,* $j = 0, \ldots, k$*.*

Proof: The robust convex quadratic constraint $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$ for all $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_d$ is equivalent to

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \max_{\{\mathbf{u}:\mathbf{u}\geq\mathbf{0}, \|\mathbf{u}\|\leq 1\}} \left\{ \sum_{j=1}^k u_j(\mathbf{x}^T\mathbf{Q}_j\mathbf{x}) \right\} + \max_{\{\mathbf{v}:\|\mathbf{v}\|\leq 1\}} \left\{ \sum_{j=1}^k v_j(2\mathbf{q}_j^T\mathbf{x} + \gamma_j) \right\} \leq 0. \tag{17}$$

Thus, (17) holds if and only if there exist $\mathbf{f}, \mathbf{g} \in \mathbf{R}^k$ such that $f_j \geq \mathbf{x}^T\mathbf{Q}_j\mathbf{x}$, $g_j \geq 2\mathbf{q}_j^T\mathbf{x} + \gamma_j$, $j = 1, \ldots, k$, and

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \|\mathbf{f}\| + \|\mathbf{g}\| \leq 0. \tag{18}$$

The result follows from rewriting (18) as a collection of linear and SOC constraints. ∎

2.3 Factorized uncertainty sets

The next class of uncertainty sets is defined as follows.

$$\mathcal{S}_e = \left\{ (\mathbf{Q}, \mathbf{q}, \gamma_0) : \begin{array}{l} \mathbf{Q} = \mathbf{V}^T\mathbf{F}\mathbf{V}, \mathbf{F} = \mathbf{F}_0 + \mathbf{\Delta} \succ \mathbf{0}, \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}\| \leq \eta, \mathbf{N} \succ \mathbf{0}, \\ \mathbf{V} = \mathbf{V}_0 + \mathbf{W} \in \mathbf{R}^{m\times n}, \|\mathbf{W}_i\|_g = \sqrt{\mathbf{W}_i^T\mathbf{G}\mathbf{W}_i} \leq \rho_i,\ \forall i, \mathbf{G} \succeq \mathbf{0}, \\ \mathbf{q} = \mathbf{q}_0 + \boldsymbol{\zeta}, \|\boldsymbol{\zeta}\|_s = \sqrt{\boldsymbol{\zeta}^T\mathbf{S}\boldsymbol{\zeta}} \leq \delta, \mathbf{S} \succ \mathbf{0}. \end{array} \right\}, \tag{19}$$

where $\mathbf{W}_i$, $i = 1, \ldots, n$, is the i-th column of the matrix $\mathbf{W}$ and the norm $\|\mathbf{A}\|$ of a symmetric matrix $\mathbf{A}$ is either given by the $\mathcal{L}_2$-norm, i.e. $\|\mathbf{A}\| = \max_{1\leq i\leq m}\{|\lambda_i(\mathbf{A})|\}$, or the Frobenius norm, i.e. $\|\mathbf{A}\| = \sqrt{\sum_{i=1}^m \lambda_i^2(\mathbf{A})}$, where $\{\lambda_i(\mathbf{A}), i = 1, \ldots, m\}$ are the eigenvalues of the matrix $\mathbf{A}$. The uncertainty structure $\mathcal{S}_e$ in (19) is quite general and includes as special cases: (i) fixed $\mathbf{F}$ (e.g., $\mathbf{F} = \mathbf{I}$, i.e. $\mathbf{Q} = \mathbf{V}^T\mathbf{V}$) and (ii) fixed $\mathbf{V}$, i.e. only $\mathbf{F}$ is uncertain.

Although the uncertainty structure $\mathcal{S}_e$ does not appear as natural as the discrete or the affine uncertainty sets, it captures the structure of the confidence regions around the maximum likelihood estimates of the parameters. See [16] for a detailed discussion of the structure of this uncertainty set and its parametrization.

Lemma 3 below establishes that a robust quadratic constraint corresponding to (19) can be reformulated as a collection of linear and SOC constraints.

Lemma 3 *The decision vector* $\mathbf{x} \in \mathbf{R}^n$ *satisfies the robust convex quadratic constraint* $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$ *for all* $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_e$*, where* $\mathcal{S}_e$ *is defined by (19), if and only if there exist* $\tau, \nu, \sigma, r \in \mathbf{R}$*,* $\mathbf{u} \in \mathbf{R}^n$ *and* $\mathbf{w} \in \mathbf{R}^m$ *such that*

$$
\begin{array}{rcll}
\tau & \geq & 0, & \\
\nu & \geq & \tau + \mathbf{1}^T\mathbf{t} & \\
\sigma & \leq & \lambda_{\min}(\mathbf{H}), & \\
r & \geq & \sum_{i=1}^n \rho_i u_i, & \\
u_j & \geq & x_j, & j = 1, \ldots, n, \\
u_j & \geq & -x_j, & j = 1, \ldots, n, \\
\left\| \begin{bmatrix} 2r \\ \sigma - \tau \end{bmatrix} \right\| & \leq & \sigma + \tau, & \\
\left\| \begin{bmatrix} 2w_i \\ (\lambda_i - \sigma - t_i) \end{bmatrix} \right\| & \leq & (\lambda_i - \sigma + \tau_i), & i = 1, \ldots, m, \\
2\delta \|\mathbf{S}^{-\frac{1}{2}}\mathbf{x}\| & \leq & -\nu - 2\mathbf{q}_0^T\mathbf{x} - \gamma_0, &
\end{array}
\tag{20}
$$

where $\mathbf{H} = \mathbf{G}^{-\frac{1}{2}}(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{G}^{-\frac{1}{2}}$*,* $\mathbf{H} = \mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}$ *is the spectral decomposition of* $\mathbf{H}$*,* $\mathbf{\Lambda} = \mathbf{diag}(\lambda_i)$*, and* $\mathbf{w} = \mathbf{Q}^T\mathbf{H}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}}\mathbf{V}_0\mathbf{x}$.

Proof: Fix $\mathbf{V} \in \mathcal{S}_e$. Define $\tilde{\mathbf{\Delta}} = \mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}$, $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ and $\mathcal{S}_1 = \{\mathbf{F} : \mathbf{F} = \mathbf{F}_0 + \mathbf{\Delta} \succeq \mathbf{0}, \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{N}^{\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{\frac{1}{2}}\| \leq \eta\}$. Then

$$
\begin{aligned}
\max\{\mathbf{x}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\mathbf{x} : \mathbf{F} \in \mathcal{S}_1\} &= \max\Big\{\mathbf{y}^T(\mathbf{F}_0 + \mathbf{\Delta})\mathbf{y} : \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{N}^{\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{\frac{1}{2}}\| \leq \eta, \mathbf{F}_0 + \mathbf{\Delta} \succeq \mathbf{0}\Big\} \\
&= \max\Big\{\mathbf{y}^T\mathbf{F}_0\mathbf{y} + (\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T\tilde{\mathbf{\Delta}}(\mathbf{N}^{\frac{1}{2}}\mathbf{y}) : \|\tilde{\mathbf{\Delta}}\| \leq \eta, \mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}\Big\}, \\
&\leq \max\Big\{\mathbf{y}^T\mathbf{F}_0\mathbf{y} + (\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T\tilde{\mathbf{\Delta}}(\mathbf{N}^{\frac{1}{2}}\mathbf{y}) : \|\tilde{\mathbf{\Delta}}\| \leq \eta\Big\}, \qquad (21) \\
&\leq \mathbf{y}^T\mathbf{F}_0\mathbf{y} + \eta(\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T(\mathbf{N}^{\frac{1}{2}}\mathbf{y}), \qquad (22)
\end{aligned}
$$

where (21) follows from relaxing the constraint $\mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}$ and (22) follows from the properties of the matrix norm.

Since $\|\tilde{\mathbf{\Delta}}\| = \max\{|\lambda_i(\tilde{\mathbf{\Delta}})|\}$ or $\sqrt{\sum_i \lambda_i^2(\tilde{\mathbf{\Delta}})}$ and $\mathbf{N} \succ \mathbf{0}$, the bound (22) is achieved by

$$
\tilde{\mathbf{\Delta}}^* = \eta \frac{(\mathbf{N}^{\frac{1}{2}}\mathbf{y})(\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T}{\|\mathbf{N}^{\frac{1}{2}}\mathbf{y}\|^2},
$$

unless $\mathbf{y} = \mathbf{0}$. Thus, the right hand side of (21) is given by $\mathbf{y}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}$ and is achieved by $\tilde{\mathbf{\Delta}}^* = \eta\frac{\mathbf{N}\mathbf{y}\mathbf{y}^T\mathbf{N}}{\mathbf{y}^T\mathbf{N}\mathbf{y}}$, unless $\mathbf{y} = \mathbf{0}$. Since $\mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}^*\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}$, it follows that the inequality (21) is, in fact, an equality, i.e.

$$
\max_{\mathbf{F} \in \mathcal{S}_1} \{\mathbf{y}^T\mathbf{F}\mathbf{y}\} = \mathbf{y}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}.
$$

Therefore,

$$
\max_{\{(\mathbf{Q},\mathbf{q},\gamma) \in S_e\}} \{\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma\} = \gamma_0 + 2\mathbf{q}_0^T\mathbf{x} + \delta\|\mathbf{S}^{-\frac{1}{2}}\mathbf{x}\| + \max_{\mathbf{V} \in S_v}\{\mathbf{x}^T\mathbf{V}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{V}\mathbf{x}\}, \tag{23}
$$

where $S_v = \{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \leq \rho_i, i = 1, \ldots, m\}$. The rest of the proof reformulates the robust quadratic constraint $\max_{\{\mathbf{V} \in S_v\}}\{\mathbf{x}^T\mathbf{V}^T(\mathbf{I}+\eta\mathbf{N})\mathbf{V}\mathbf{x}\} \leq \nu$ as a collection of SOC constraints.

Since the constraints $\|\mathbf{W}_i\|_g \leq \rho_i$, $i = 1, \ldots, n$ imply the bound,

$$\|\mathbf{W}\mathbf{x}\|_g = \left\|\sum_{i=1}^{n} x_i \mathbf{W}_i\right\|_g \leq \sum_{i=1}^{n} |x_i| \, \|\mathbf{W}_i\|_g \leq \sum_{i=1}^{n} \rho_i \, |x_i| \, , \tag{24}$$

the optimization problem,

$$\begin{array}{ll} \text{maximize} & \|\mathbf{V}_0\boldsymbol{\phi} + \mathbf{W}\mathbf{x}\|^2 \\ \text{subject to} & \|\mathbf{W}\mathbf{x}\|_g \leq \sum_{i=1}^{n} \rho_i \, |x_i| \, , \end{array} \tag{25}$$

is a relaxation of

$$\begin{array}{ll} \text{maximize} & \|\mathbf{V}_0\mathbf{x} + \mathbf{W}\mathbf{x}\|^2 \\ \text{subject to} & \|\mathbf{W}_i\|_g \leq \rho_i, \quad i = 1, \ldots, n. \end{array} \tag{26}$$

The objective function in (25) is convex; therefore the optimal solution $\mathbf{W}^*\boldsymbol{\phi}$ lies on the boundary of the feasible set, i.e. $\|\mathbf{W}^*\mathbf{x}\|_g = \boldsymbol{\rho}^T|\mathbf{x}|$, where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^T$. But from (24), it follows that $\mathbf{W}$ lies on the boundary only if the columns of the matrix $\mathbf{W}$ satisfy $\mathbf{W}_i = \rho_i\mathbf{v}$ for some vector $\mathbf{v}$ with $\|\mathbf{v}\|_g = 1$. But any such choice of $\mathbf{W}_i$ is feasible for (26). Therefore (26) and (25) are, in fact, equivalent.

Thus $\mathbf{x}^T\mathbf{V}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{V}\mathbf{x} \leq \nu$ for all $\mathbf{V} \in \mathcal{S}_v$ if and only if

$$\big(\mathbf{V}_0\mathbf{x} + (\boldsymbol{\rho}^T|\mathbf{x}|)\mathbf{v}\big)^T\big(\mathbf{F}_0 + \eta\mathbf{N}\big)\big(\mathbf{V}_0\mathbf{x} + (\boldsymbol{\rho}^T|\mathbf{x}|)\mathbf{v}\big) \leq \nu \tag{27}$$

for all $\|\mathbf{v}\|_g \leq 1$, i.e. $1 - \mathbf{v}^T\mathbf{G}\mathbf{v} \geq 0$. Define $\mathbf{y}_0 = \mathbf{V}_0\mathbf{x}$ and $r = \boldsymbol{\rho}^T|\mathbf{x}|$ then (27) is equivalent to

$$\nu - \mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}_0 - 2r\mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{v} - r^2\mathbf{v}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{v} \geq 0, \tag{28}$$

for all $\mathbf{v}$ such that $1 - \mathbf{v}^T\mathbf{G}\mathbf{v} \geq 0$. Before proceeding further, we need the following:

Lemma 4 ($\mathcal{S}$-procedure) *Let $F_i(\mathbf{x}) = \mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i$, $i = 0, \ldots, p$ be quadratic functions of $\mathbf{x} \in \mathbf{R}^n$. Then $F_0(\mathbf{x}) \geq 0$ for all $\mathbf{x}$ such that $F_i(\mathbf{x}) \geq 0$, $i = 1, \ldots, p$, if there exist $\tau_i \geq 0$ such that*

$$\begin{bmatrix} c_0 & \mathbf{b}_0^T \\ \mathbf{b}_0 & \mathbf{A}_0 \end{bmatrix} - \sum_{i=1}^{p} \tau_i \begin{bmatrix} c_i & \mathbf{b}_i^T \\ \mathbf{b}_i & \mathbf{A}_i \end{bmatrix} \succeq 0.$$

Moreover, if $p = 1$ then the converse holds if there exists $\mathbf{x}_0$ such that $F_1(\mathbf{x}_0) > 0$.

For a discussion of the $\mathcal{S}$-procedure and its applications, see [7].

Since $\mathbf{v} = \mathbf{0}$ is strictly feasible for $1 - \mathbf{v}^T\mathbf{G}\mathbf{v} \geq 0$, the S-procedure implies that (28) holds for all $1 - \mathbf{v}^T\mathbf{G}\mathbf{v} \geq 0$ if and only if there exists a $\tau \geq 0$ such that

$$\mathbf{M} = \begin{bmatrix} \nu - \tau - \mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}_0 & -r\mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N})^{\frac{1}{2}} \\ -r(\mathbf{F}_0 + \eta\mathbf{N})^{\frac{1}{2}}\mathbf{y}_0 & \tau\mathbf{G} - r^2\mathbf{I} \end{bmatrix} \succeq 0. \tag{29}$$

Let the spectral decomposition of $\mathbf{H} = \mathbf{G}^{-\frac{1}{2}}(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{G}^{-\frac{1}{2}}$ be $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{\Lambda} = \mathbf{diag}(\boldsymbol{\lambda})$, and define $\mathbf{w} = \mathbf{Q}^T\mathbf{H}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}}\mathbf{y}_0 = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{Q}^T\mathbf{G}^{\frac{1}{2}}\mathbf{y}_0$. Observing that $\mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}_0 = \mathbf{w}^T\mathbf{w}$, we have that the matrix $\mathbf{M} \succeq \mathbf{0}$ if and only if

$$\bar{\mathbf{M}} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{Q}^T\mathbf{G}^{\frac{1}{2}} \end{bmatrix} \mathbf{M} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{G}^{\frac{1}{2}}\mathbf{Q} \end{bmatrix} = \begin{bmatrix} \nu - \tau - \mathbf{w}^T\mathbf{w} & -r\mathbf{w}^T\mathbf{\Lambda}^{\frac{1}{2}} \\ -r\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{w} & \tau\mathbf{I} - r^2\mathbf{\Lambda} \end{bmatrix} \succeq \mathbf{0}.$$

The matrix $\bar{\mathbf{M}} \succeq \mathbf{0}$ if and only if $\tau \geq r^2\lambda_i$, for all $i = 1, \ldots, m$ (i.e. $\tau \geq r^2\lambda_{\max}(\mathbf{H})$), $w_i = 0$ for all i such that $\tau = r^2\lambda_i$, and the Schur complement of the nonzero rows and columns of $\tau\mathbf{I} - r^2\mathbf{\Lambda}$

$$\nu - \tau - \mathbf{w}^T\mathbf{w} - r^2 \left(\sum_{i:\tau \neq r^2\lambda_i} \frac{\lambda_i w_i^2}{\tau - r^2\lambda_i} \right) = \nu - \tau - \sum_{i:\sigma\lambda_i \neq 1} \frac{w_i^2}{1 - \sigma\lambda_i} \geq 0,$$

where $\sigma = \frac{r^2}{\tau}$. It follows that (27) holds for all $\mathbf{v}^T\mathbf{G}\mathbf{v} \leq 1$ if and only if there exists $\tau, \sigma \geq 0$ and $\mathbf{t} \in \mathbf{R}_+^m$ satisfying,

$$\begin{array}{rcl} \nu & \geq & \tau + \mathbf{1}^T\mathbf{t}, \\ r^2 & = & \sigma\tau, \\ w_i^2 & = & (1 - \sigma\lambda_i)t_i, \quad i = 1, \ldots, m, \\ \sigma & \leq & \frac{1}{\lambda_{\max}(\mathbf{H})}. \end{array} \tag{30}$$

It is easy to establish that there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy (30) if and only if there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy (30) with the equalities replaced by inequalities.

From [22] (see Section 6.2.3) and [18], we have that for $\mathbf{z} \in \mathbf{R}^n$, $x \in \mathbf{R}$, and $y \in \mathbf{R}$, $x, y \geq 0$,

$$\mathbf{z}^T\mathbf{z} \leq xy \Leftrightarrow \left\| \begin{bmatrix} 2\mathbf{z} \\ x - y \end{bmatrix} \right\| \leq x + y.$$

Note that the constraint $r^2 \leq \sigma\tau$ and $\tau \geq 0$ imply that $\sigma \geq 0$. Therefore, replacing the equalities in (30) and reformulating the inequalities as SOC constraints, we have that $\mathbf{x}^T\mathbf{V}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{V}\mathbf{x} \leq \nu$ for all $\mathbf{V} \in \mathcal{S}_v$ if and only if the following system of linear and second-order cone constraints holds,

$$\begin{array}{rcl} \tau & \geq & 0, \\ \nu & \geq & \tau + \mathbf{1}^T\mathbf{t} \\ \sigma & \leq & \lambda_{\min}(\mathbf{G}), \\ r & \geq & \sum_{i=1}^n \rho_i \left| x_i \right|, \\ \left\| \begin{bmatrix} 2r \\ \sigma - \tau \end{bmatrix} \right\| & \leq & \sigma + \tau, \\ \left\| \begin{bmatrix} 2w_i \\ (1 - \sigma\lambda_i - t_i) \end{bmatrix} \right\| & \leq & (1 - \sigma\lambda_i + t_i), \quad i = 1, \ldots, m. \end{array} \tag{31}$$

The constraint $r \geq \sum_{i=1}^m \rho_i |x_i|$ is not linear but it can be converted into one by introducing a new variable $\mathbf{z}$ such that $\mathbf{u} \geq |\mathbf{x}|$, i.e. $u_j \geq x_j$ and $u_j \geq -x_j$, $i = j, \ldots, n$.

The result now follows by replacing $\max_{\{\mathbf{V} \in S_c\}} \{\mathbf{x}^T\mathbf{V}^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{V}\mathbf{x}\}$ in (23) by the bound ν. ■

There are several closely related versions of the factorized uncertainty set $\mathcal{S}_e$ that also result in robust problems that can be reduced to SOCPs. These include the special case where the matrix $\mathbf{F}$ is known, i.e. $\eta = 0$; and the variant of $\mathcal{S}_e$ where $\mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \mathbf{\Delta} \succ \mathbf{0}$, with $\|\mathbf{F}_0^{\frac{1}{2}} \mathbf{\Delta} \mathbf{F}_0^{\frac{1}{2}}\| \leq \eta$, $\eta < 1$. For details of these alternative formulations and their relation to probabilistic guarantees on the performance of the optimal solution see [16].

3 Applications

In this section we present several applications of robust convex QCPs. We show that the uncertainty in these applications can be adequately modeled by the uncertainty sets introduced in the previous section.

3.1 Robust mean-variance portfolio selection

Suppose an investor wants to invest in a market with n assets. The random returns on the assets is given by the random return vector

$$\mathbf{r} = \boldsymbol{\mu} + \mathbf{V}^T \mathbf{f} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\mu} \in \mathbf{R}^n$ is the mean return vector, $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{F}) \in \mathbf{R}^m$ is the vector of returns on the factors that drive the market, $\mathbf{V} \in \mathbf{R}^{m \times n}$ is the factor loadings matrix and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$ is the residual returns vector. Here $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes that $\mathbf{x}$ is a multivariate Normal random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. In addition, we assume that the vector of residual returns $\boldsymbol{\epsilon}$ is independent of the vector of factor returns $\mathbf{f}$, the covariance matrix $\mathbf{F} \succ \mathbf{0}$ and the covariance matrix $\mathbf{D} = \mathbf{diag}(\mathbf{d}) \succ \mathbf{0}$, i.e. $d_i > 0$, $i = 1, \ldots, n$. Thus, the vector of asset returns $\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}^T \mathbf{F} \mathbf{V} + \mathbf{D})$.

The investor's position in the market is described by a portfolio vector $\boldsymbol{\phi} \in \mathbf{R}^n$, where ϕ_j denotes the fraction of the capital invested in asset j, $j = 1, \ldots, n$. The random return r_ϕ on a portfolio $\boldsymbol{\phi}$ is given by $r_\phi = \mathbf{r}^T \boldsymbol{\phi}$. The objective is to choose a portfolio that maximizes some measure of "return" on the investment subject to appropriate constraints on the associated "risk".

Markowitz [20, 21] proposed a model for portfolio selection in which the "return" is the expected value $\mathbf{E}[r_\phi]$ of the portfolio return, the "risk" is the variance $\mathbf{Var}\big[\mathbf{r}_\phi\big]$ of the return, and the optimal portfolio $\boldsymbol{\phi}^*$ is one that has the minimum variance amongst those that a return of at least α, i.e. $\boldsymbol{\phi}^*$ is the optimal solution of the convex quadratic optimization problem

$$\begin{array}{rl} \text{minimize} & \mathbf{Var}\big[\mathbf{r}_\phi\big] \\ \text{subject to} & \mathbf{E}[r_\phi] \geq \alpha, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1. \end{array} \tag{32}$$

The optimization problem (32) is called the minimum variance portfolio selection problem. Other variants include the maximum return problem and the maximum Sharpe ratio problem.

Note that the Markowitz model implicitly assumes that the mean return vector $\mathbf{E}[\mathbf{r}]$ and the covariance matrix $\mathbf{Var}[\mathbf{r}]$ are known with certainty. This mean-variance model has had a profound impact on the economic modeling of financial markets and the pricing of assets. In 1990, Sharpe and Markowitz shared the Nobel Memorial Prize in Economic Sciences for this work. In spite of the theoretical success of the mean-variance model, practitioners have shied away from it. The primary criticism leveled against the Markowitz model is that the optimal portfolio $\boldsymbol{\phi}^*$ is extremely sensitive to the market parameters $(\mathbf{E}[\mathbf{r}], \mathbf{Var}[\mathbf{r}])$: since these parameters are estimated from noisy data, $\boldsymbol{\phi}^*$ often amplifies noise.

One solution to the sensitivity of $\boldsymbol{\phi}^*$ to the perturbation of the problem data is to consider a robust version of the Markowitz problem (32). To this end, we define the uncertainty structures as follows. The covariance matrix $\mathbf{F}$ of the factor returns $\mathbf{f}$ is assumed to belong to

$$S_f = \left\{\mathbf{F} : \mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succeq \mathbf{0}, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T, \|\mathbf{F}_0^{\frac{1}{2}} \boldsymbol{\Delta} \mathbf{F}_0^{\frac{1}{2}}\| \leq \eta\right\}; \tag{33}$$

the uncertainty set S_d for the matrix $\mathbf{D}$ is given by

$$S_d = \left\{\mathbf{D} : \mathbf{D} = \mathbf{diag}(\mathbf{d}), d_i \in [\underline{d}_i, \overline{d}_i], i = 1, \ldots, n\right\}; \tag{34}$$

the factor loadings matrix $\mathbf{V}$ belongs to the elliptical uncertainty set S_v given by

$$S_v = \left\{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \leq \rho_i, i = 1, \ldots, n\right\}, \tag{35}$$

where $\mathbf{W}_i$ is the i-th column of $\mathbf{W}$ and $\|\mathbf{w}\|_g = \sqrt{\mathbf{w}^T \mathbf{G} \mathbf{w}}$; and the mean returns vector $\boldsymbol{\mu}$ lies in

$$S_m = \{\boldsymbol{\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\xi}, |\xi_i| \leq \gamma_i, i = 1, \ldots, n\}. \tag{36}$$

The uncertainty sets (S_f, S_v, S_d, S_m) mimic the structure of the confidence region around the minimum mean square estimates of $(\boldsymbol{\mu}, \mathbf{V}, \mathbf{F})$. The justification for this choice of uncertainty structures and suitable choices for the matrix $\mathbf{G}$, and the bounds ρ_i, γ_i, $\overline{d}_i$, $\underline{d}_i$, $i = 1, \ldots, n$, and η are discussed in [16].

The robust analog of the Markowitz mean-variance optimization problem (32) is given by

$$\begin{array}{ll} \text{minimize} & \max_{\{\mathbf{V} \in S_v, \mathbf{D} \in S_d\}} \mathbf{Var}[\mathbf{r}_\phi] \\ \text{subject to} & \min_{\{\boldsymbol{\mu} \in S_m\}} \mathbf{E}[r_\phi] \geq \alpha, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1. \end{array} \tag{37}$$

We expect that the sensitivity of the optimal solution of this mathematical program to parameter fluctuations will be significantly smaller than it would be for its classical counterpart (32).

Since the return $r_\phi \sim \mathcal{N}(\boldsymbol{\mu}^T \boldsymbol{\phi}, \boldsymbol{\phi}^T (\mathbf{V}^T \mathbf{F} \mathbf{V} + \mathbf{D}) \boldsymbol{\phi})$, we can write (37) as

$$\begin{array}{ll} \text{minimize} & \max_{\{\mathbf{V} \in S_v\}} \left\{\boldsymbol{\phi}^T \mathbf{V}^T \mathbf{F} \mathbf{V} \boldsymbol{\phi}\right\} + \max_{\{\mathbf{D} \in S_d\}} \left\{\boldsymbol{\phi}^T \mathbf{D} \boldsymbol{\phi}\right\} \\ \text{subject to} & \min_{\{\boldsymbol{\mu} \in S_m\}} \boldsymbol{\mu}^T \boldsymbol{\phi} \geq \alpha, \\ & \mathbf{1}^T \boldsymbol{\phi} = 1, \end{array} \tag{38}$$

which in turn is equivalent to the following robust quadratically constrained problem,

$$
\begin{array}{rll}
\text{minimize} & \lambda + \delta, & \\
\text{subject to} & \boldsymbol{\phi}^T \mathbf{V}^T \mathbf{F} \mathbf{V} \boldsymbol{\phi} \leq \lambda, & \forall \mathbf{V} \in S_v, \mathbf{F} \in S_f \\
& \boldsymbol{\phi}^T \mathbf{D} \boldsymbol{\phi} \leq \delta, & \forall \mathbf{D} \in S_d, \\
& \boldsymbol{\mu}^T \boldsymbol{\phi} \geq \alpha, & \forall \boldsymbol{\mu} \in S_m, \\
& \mathbf{1}^T \boldsymbol{\phi} = 1. &
\end{array} \tag{39}
$$

Since the uncertainty sets $S_m \times S_v \times S_f$ and S_d are special cases of the factorized uncertainty structure proposed in (19), (39) can be reduced to an SOCP. For details on robust portfolio selection problems and the performance on real market data see [16].

3.2 Robust hyperplane separation

Let $\mathcal{L} = \{(\mathbf{x}_i, y_i), i = 1, \ldots, l\}$, $y_i \in \{+1, -1\}$, $\mathbf{x}_i \in \mathbf{R}^d$, $\forall i$, be a labeled set of training data. The objective in the hyperplane separation problem is to choose a hyperplane $(\mathbf{w}, b)$, $b \in \mathbf{R}$, $\mathbf{w} \in \mathbf{R}^d$, that maximally separates the "negative examples", i.e. $\mathbf{x}_i$ with $y_i = -1$, from the "positive examples", i.e. $\mathbf{x}_i$ with $y_i = +1$. Then given this separating hyperplane $(\mathbf{w}, b)$, a new sample $\mathbf{x}$ is classified as "positive" provided $\mathbf{w}^T\mathbf{x} + b \geq 0$, otherwise it is classified as "negative". Pattern classification using hyperplanes is called linear discrimination.

In a typical application of linear discrimination, the hyperplane $(\mathbf{w}, b)$ is chosen by solving the following quadratic program [8, 19, 28].

$$
\begin{array}{rll}
\text{minimize} & \frac{1}{2}\|\mathbf{w}\|^2 + C\big(\sum_{i=1} \xi_i\big), & \\
\text{subject to} & \mathbf{w}^T\mathbf{x}_i + b \geq 1 - \xi_i, & \text{if } y_i = +1, \\
& \mathbf{w}^T\mathbf{x}_i + b \geq 1 + \xi_i, & \text{if } y_i = -1, \\
& \xi_i \geq 0, & i = 1, \ldots, l.
\end{array} \tag{40}
$$

Instead of solving (40), one typically solves its dual given by

$$
\begin{array}{rl}
\text{maximize} & \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i\alpha_j (y_i\mathbf{x}_i)^T(y_j\mathbf{x}_j), \\
\text{subject to} & \sum_{i=1}^{l} \alpha_i y_i = 0, \\
& 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, l.
\end{array} \tag{41}
$$

The optimal vector $\mathbf{w}^* = \sum_{i=1}^{l} \alpha_i^* \mathbf{x}_i$, where $\boldsymbol{\alpha}^*$ is the optimal solution of the dual (41). The optimal intercept b^* is set by the complementary slack conditions (for a detailed discussion see [8]).

In several applications of linear discrimination, the training data $\mathbf{x}_i$ is corrupted by measurement noise. A simple additive model for measurement error is given by

$$
\mathbf{x}_i = \bar{\mathbf{x}}_i + \mathbf{u}_i, \quad i = 1, \ldots, l,
$$

where $\bar{\mathbf{x}}_i$ is the true value of the training data and $\mathbf{u}_i$ is the measurement noise. Typically, one assumes that $\|\mathbf{u}_i\| \leq \rho$, $i = 1, \ldots, l$. If the measurement noise $\{\mathbf{u}_i : i = 1, \ldots, l\}$ were known, the appropriate dual problem would be

$$\begin{array}{rl} \text{maximize} & \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j(y_i(\mathbf{x}_i - \mathbf{u}_i))^T(y_j(\mathbf{x}_j - \mathbf{u}_j)), \\ \text{subject to} & \sum_{i=1}^{l}\alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, l. \end{array} \tag{42}$$

When the perturbations due to measurement noise are unknown, a conservative approach is to replace the objective function in (42) by $\mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\max_{\{\|\mathbf{u}_i\|\leq\rho\}}\Big\{\sum_{i,j=1}^{l}\alpha_i\alpha_j(y_i(\mathbf{x}_i - \mathbf{u}_i))^T(y_j(\mathbf{x}_j - \mathbf{u}_j))\Big\}$, i.e. solve the following robust quadratically constrained problem,

$$\begin{array}{rl} \text{maximize} & \tau, \\ \text{subject to} & \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{Q}\boldsymbol{\alpha} \geq \tau, \quad \forall \mathbf{Q} \in \mathcal{S}, \\ & \sum_{i=1}^{l}\alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, l, \end{array} \tag{43}$$

where the uncertainty set

$$\mathcal{S} = \Big\{\mathbf{Q} : \mathbf{Q} = \mathbf{V}^T\mathbf{V}, \mathbf{V} = \mathbf{V}_0 + \mathbf{U}, \|\mathbf{U}_i\| \leq \rho, \mathbf{V}_0 = [\mathbf{x}_1, \ldots, \mathbf{x}_l]\,\mathbf{diag}(\mathbf{y})\Big\} \tag{44}$$

belongs to class of factorized uncertainty structures defined in (19). Thus, (44) can be reformulated as an SOCP. This technique can be extended to general support vector machines [28] as well.

3.3 Linear least squares problem with deterministic and stochastic uncertainty

Consider the following linear least squares problem,

$$\min_{\{\mathbf{x}\in\mathbf{R}^n\}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \tag{45}$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]^T \in \mathbf{R}^{m\times n}$ and $\mathbf{b} \in \mathbf{R}^n$. If $m \geq n$ and the matrix $\mathbf{A}$ has full column rank, the solution of this optimization problem is given by $\mathbf{x}^* = \big(\mathbf{A}^T\mathbf{A}\big)^{-1}\mathbf{A}^T\mathbf{b}$ [13]. Even when additional linear and convex quadratic constraints are imposed on the solution $\mathbf{x}$, such as $\|\mathbf{x}\|^2 \leq M$, the linear least squares problem (45) is still a convex QCP.

In many applications of least squares problems, the problem data $(\mathbf{A}, \mathbf{b})$ is either estimated from empirical data or is the result of measurement, and therefore, subject to perturbations. In order to reduce the sensitivity of the decision $\mathbf{x}$ to perturbations in the data, El Ghaoui and Lebret formulated the following robust version of (45)

$$\min_{\mathbf{x}} \max_{\big\{[\mathbf{A},\mathbf{b}]:\|[\mathbf{A},\mathbf{b}]-[\mathbf{A}_0,\mathbf{b}_0]\|\leq\rho\big\}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \tag{46}$$

where $\|\cdot\|$ is the Frobenius norm, and showed that (46) can be reformulated as an SOCP [11]. However it is not clear that the uncertainty set that appears above is natural, since it applies to $[\mathbf{A}\ \mathbf{b}]$ at once.

In this paper, we propose the following uncertainty structure for the rows $\mathbf{a}_i \in \mathbf{R}^n$, $i = 1, \ldots, m$ of the data matrix $\mathbf{A}$:

$$\mathcal{S} = \Big\{\mathbf{a} : \mathbf{a} = \mathbf{a}^0 + \sum_{j=1}^{k} v^j \mathbf{a}^j + \sum_{j=1}^{l} u^j \boldsymbol{\xi}^j\Big\}, \tag{47}$$

where without any loss of generality $\|\mathbf{v}\| \le 1$, $\|\mathbf{u}\| \le 1$, and $\boldsymbol{\xi}^j \sim \mathcal{N}(\mathbf{0}, \Omega^j)$, $j = 1, \ldots, l$. Without the stochastic term, the uncertainty set (47) has the affine structure considered in [4, 5, 6]. The term $\sum_{j=1}^{l} u^j \boldsymbol{\xi}^j$ models the imperfect knowledge of the stochastic perturbations in $\mathbf{a}$ – the decision maker knows the total variance and modes Ω^j but does not know the variance of each of the individual modes. In typical applications, the matrix $\Omega^j = \boldsymbol{\omega}^j(\boldsymbol{\omega}^j)^T$, or equivalently $\boldsymbol{\xi}^j = (\boldsymbol{\omega}^j)^T \mathbf{Z}$ for $Z \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where the vector $\boldsymbol{\omega}^j$ is determined by the estimation algorithm or the signal network.

The robust least squares problem corresponding to (47) is given by

$$\min_{\mathbf{x}} \Big\{ \sum_{i=1}^{m} \max_{\{(\mathbf{u}_i, \mathbf{v}_i)\}} \Big\{ \mathbf{E}[(\mathbf{a}_i^T \mathbf{x} - b_i)^2] \Big\} \Big\}, \tag{48}$$

where each $\mathbf{a}_i$ belongs to a uncertainty set of the form in (47) for appropriately chosen $\{\mathbf{a}_i^j : j = 1, \ldots, k_i\}$ and $\{\Omega_i^j : j = 1, \ldots, l_i\}$, $i = 1, \ldots, m$.

For a fixed $\mathbf{a}$ in $\mathcal{S}$ and $b \in \mathbf{R}$, the expected error $\mathbf{E}[(\mathbf{a}^T\mathbf{x} - b)^2]$ is given by

$$\mathbf{E}[(\mathbf{a}^T\mathbf{x} - b)^2] = \Big((\mathbf{a}^0)^T\mathbf{x} + \sum_{j=1}^{k} v^j (\mathbf{a}^j)^T \mathbf{x} - b\Big)^2 + \sum_{j=1}^{l} (u^j)^2 \mathbf{x}^T \Omega^j \mathbf{x}. \tag{49}$$

The constraint $\mathbf{E}[(\mathbf{a}^T\mathbf{x} - b)^2] \le \delta$, for all $\mathbf{a} \in \mathcal{S}$, is equivalent to the following set of constraints,

$$\begin{array}{ll} |\ \mathbf{a}^T\mathbf{x} - b\ | \le t, & \forall \mathbf{a} \in \mathcal{S}_1 = \{\mathbf{a} : \mathbf{a}^0 + \sum_{j=1}^{k} v^j \mathbf{a}^j, \|\mathbf{v}\| \le 1\}, \\ t^2 + \mathbf{x}^T\mathbf{Q}\mathbf{x} \le \delta, & \forall \mathbf{Q} \in \mathcal{S}_2 = \{\mathbf{Q} : \mathbf{Q} = \sum_{j=1}^{l} \alpha^j \Omega^j, \sum_{j=1}^{l} \alpha^j \le 1, \alpha^j \ge 0, \forall j\}. \end{array} \tag{50}$$

From (48), (49) and (50), it follows that the robust optimization problem (48) is equivalent to the following robust convex QCP

$$\begin{array}{rll} \text{minimize} & \sum_{i=1}^{m} \delta_i, & \\ \text{subject to} & t_i^2 + \mathbf{x}^T \mathbf{Q}_i \mathbf{x} \le \delta_i, & \mathbf{Q}_i \in \mathcal{S}_2^i, i = 1, \ldots, m, \\ & \mathbf{a}_i^T\mathbf{x} - b_i \le t_i, & \mathbf{a}_i \in \mathcal{S}_1^i, i = 1, \ldots, m, \\ & \mathbf{a}_i^T\mathbf{x} - b_i \ge -t_i, & \mathbf{a}_i \in \mathcal{S}_1^i, i = 1, \ldots, m. \end{array} \tag{51}$$

The set $\mathcal{S}_1^i$, $i = 1, \ldots, m$, is a special case of the affine uncertainty set defined in (15) and $\mathcal{S}_2^i$, $i = 1, \ldots, m$, belongs to the class of polytopic uncertainty sets defined in (8). Therefore, the robust problem (51) can be reduced to an SOCP.

3.4 Equalizing uncertain channels

By suitably sampling the input and output signals, the input-output relation of any linear time-invariant communication system can be written as follows [25]:

$$\begin{aligned} y_k &= \sum_{i=0}^{m-1} h_i x_{k-i} + s_k, \\ &= h_0 x_k + \sum_{i=1}^{m-1} h_i x_{k-i} + s_k, \end{aligned} \tag{52}$$

where $\{x_k, k \geq 0\}$ are the samples of the input signal, $\{y_k, k \geq 0\}$ are the samples of the output signal, $\{h_i, i = 0, \ldots, m\}$ is the impulse response of the channel, and $\{s_k, k \geq 0\}$ are the samples of the channel noise. We assume that the channel impulse response is finite, i.e. $m < \infty$. The term $\sum_{i=1}^{m-1} h_i x_{k-i}$ is called the inter-symbol interference (ISI). In order to recover the input sequence one has to remove the ISI and the effects of the noise.

The $\mathcal{Z}$-transform $W(z)$ of any sequence $\{w_k : k \geq 0\}$ is defined by

$$W(z) = \sum_i w_i z^i, \tag{53}$$

where z is a complex number. Under fairly general conditions [25], the sequence $\{w_k : k \geq 0\}$ can be recovered from $\{W(z) : |z| = 1\}$. Therefore, we will treat the sequence and its $\mathcal{Z}$-transform as equivalent. Taking the $\mathcal{Z}$-transform of both sides of (52), we get

$$Y(z) = H(z)X(z) + S(z), \tag{54}$$

where $Y(z)$, $H(z)$, $X(z)$ and $S(z)$ are the $\mathcal{Z}$-transforms of $\{y_k\}$, $\{h_k\}$, $\{x_k\}$ and $\{s_k\}$ respectively. From (54) we have that

$$\frac{1}{H(z)} Y(z) = X(z) + \frac{S(z)}{H(z)}, \tag{55}$$

i.e. one can remove the ISI by processing the output sequence through a linear time-invariant filter with impulse response $G(z) = \frac{1}{H(z)}$ [25].

The process of removing ISI using a linear time-invariant filter is called *channel equalization*. Thus, (55) describes a technique for channel equalization. This technique, although simple, is not practical because the impulse response $G(z)$ is infinite, i.e. an infinite number of output samples are required to reconstruct one input sample.

Channel equalization using finite impulse response (FIR) filters is possible provided the output signal is sampled at a faster rate [24]. Oversampling the output signal at a rate p times faster than the input is equivalent to p parallel input-output channels that all see the same input sequence. Let $H_j(z) = \sum_{i=0}^{m-1} h_{ji} z^i$, $j = 1, \ldots, p$ denote the channel responses of the p parallel input-output channels obtained by oversampling the output. Then, the output $Y_j(z)$ of the j-th channel is given by

$$Y_j(z) = H_j(z)X(z) + S_j(z). \tag{56}$$

Suppose the output $Y_j(z)$ is passed through a filter with impulse response $G_j(z) = \sum_{i=0}^{n-1} g_{ji} z^i$, $n < \infty$, and the output signals added together. Then, the effective input-output relation is given by

$$Y(z) = \Big(\sum_{j=1}^{p} G_j(z) H_j(z)\Big) X(z) + \sum_{j=1}^{p} G_j(z) S_j(z). \tag{57}$$

Suppose we require the effective input-output channel be $D(z) = \sum_{j=0}^{l-1} d_j z^j$, for some $l \leq m+n$. Then $\{G_j(z), j = 1, \ldots, p\}$ must satisfy the polynomial equation

$$\sum_{j=1}^{p} G_j(z) H_j(z) = D(z),$$

or equivalently

$$\sum_{j=1}^{p} \mathbf{T}_{h_j} \mathbf{g}_j = \mathbf{d}, \tag{58}$$

where

$$\mathbf{d} = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_{m+n-2} \end{bmatrix}, \qquad \mathbf{T}_{h_j} = \underbrace{\begin{bmatrix} h_{j0} & 0 & 0 & \ldots & 0 \\ h_{j1} & h_{j0} & \ddots & \ddots & \vdots \\ h_{j2} & h_{j1} & h_{j0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & h_{j0} \\ h_{j,m-1} & \ddots & \ddots & \ddots & h_{j1} \\ 0 & \ddots & \ddots & \ddots & h_{j2} \\ \vdots & \ddots & h_{a_{M-1}} & \ddots & h_{j3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & 0 & h_{j,m-1} \end{bmatrix}}_{(n+m-1)\times n}, \qquad \mathbf{g}_j = \begin{bmatrix} g_{j0} \\ g_{j1} \\ \vdots \\ g_{j,n-1} \end{bmatrix}.$$

Under fairly general conditions [24, 23] the system of equations (58) has a solution provided $p \geq 2$, i.e. on can find finite impulse response filters $G_j(z)$, $j = 1, \ldots, p$ that can shorten the channel to any given target $D(z)$.

This equalization method assumes that the channel responses $\{H_j(z), j = 1, \ldots, p\}$ are known. In practice, the channel responses are estimated by transmitting a known finite length training sequence and, therefore, the estimates are subject to statistical errors. We model the uncertainty in the channel response as follows,

$$\bar{\mathbf{h}}_j = \mathbf{h}_j + u_j \boldsymbol{\xi}_j, \quad j = 1, \ldots, p \tag{59}$$

where $\bar{\mathbf{h}}_j$ is the *true* value of the j-th channel response, $\mathbf{h}_j$ is our estimate of the j-th channel response, $\boldsymbol{\xi}_j \sim \mathcal{N}(0, \Omega_i)$, $\mathbf{Tr}(\Omega_i) = 1$, $j = 1, \ldots, p$, $\mathbf{E}[\boldsymbol{\xi}_j \boldsymbol{\xi}_k^T] = \mathbf{0}$, $j \neq k$, and $\|\mathbf{u}\| \leq \sigma^2$. The uncertainty structure (59) reflects our limited knowledge of the noise in each of the p parallel channels.

The total noise variance is $\sum_{j=1}^p u_j^2 \,\mathbf{Tr}(\Omega_j) = \sigma^2$ but the noise variance of the individual channels is not known. The filters $\{\mathbf{g}_j, j = 1, \ldots, p\}$ are now chosen by solving the robust optimization problem

$$\begin{aligned}
&\min_{\{\mathbf{g}_j: j=1,\ldots,p\}} \max_{\{\mathbf{u}:\|\mathbf{u}\|\le\sigma^2\}} \mathbf{E}\left[\left\|\sum_{j=1}^p \mathbf{T}_{\bar{h}_j}\mathbf{g}_j - \mathbf{d}\right\|^2\right] \\
&\quad= \min_{\{\mathbf{g}_j: j=1,\ldots,p\}} \max_{\{\mathbf{u}:\|\mathbf{u}\|\le\sigma^2\}} \mathbf{E}\left[\left\|\sum_{j=1}^p \mathbf{T}_{h_j}\mathbf{g}_j - \mathbf{d} + \sum_{j=1}^p u_j \mathbf{T}_{\boldsymbol{\xi}_j}\mathbf{g}_j\right\|^2\right], \\
&\quad= \min_{\{\mathbf{g}_j: j=1,\ldots,p\}} \max_{\{\mathbf{u}:\|\mathbf{u}\|\le\sigma^2\}} \mathbf{E}\left[\left\|\sum_{j=1}^p \mathbf{T}_{h_j}\mathbf{g}_j - \mathbf{d} + \sum_{j=1}^p u_j \mathbf{T}_{\mathbf{g}_j}\boldsymbol{\xi}_j\right\|^2\right], \\
&\quad= \min_{\{\mathbf{g}_j: j=1,\ldots,p\}} \left\{\left\|\sum_{j=1}^p \mathbf{T}_{h_j}\mathbf{g}_j - \mathbf{d}\right\|^2 + \max_{\{\mathbf{u}:\|\mathbf{u}\|\le\sigma^2\}} \left\{\sum_{j=1}^p u_j^2 \,\mathbf{Tr}\big(\mathbf{T}_{g_j}^T \mathbf{T}_{g_j}\Omega_j\big)\right\}\right\}, \\
&\quad= \min_{\{\mathbf{g}_j: j=1,\ldots,p\}} \left\{\left\|\sum_{j=1}^p \mathbf{T}_{h_j}\mathbf{g}_j - \mathbf{d}\right\|^2 + \max_{\{\mathbf{u}:\|\mathbf{u}\|\le\sigma^2\}} \left\{\sum_{j=1}^p \mathbf{g}_j(u_j^2\boldsymbol{\Lambda}_j)\mathbf{g}_j\right\}\right\}, \qquad (60)
\end{aligned}$$

where $\boldsymbol{\Lambda}_j$ is set by the identity $\mathbf{g}_j\boldsymbol{\Lambda}_j\mathbf{g}_j = \mathbf{Tr}\big(\mathbf{T}_{g_j}^T \mathbf{T}_{g_j}\Omega_j\big)$. From (60) it follows that the robust equalization problem is equivalent to the following robust convex QCP

$$\begin{array}{ll}
\text{minimize} & \delta + \nu, \\
\text{subject to} & \left\|\mathbf{T}\mathbf{g} - \mathbf{d}\right\|^2 \le \delta, \\
& \mathbf{g}^T\mathbf{Q}\mathbf{g} \le \nu, \qquad \forall \mathbf{Q} \in \mathcal{S},
\end{array} \qquad (61)$$

where $\mathbf{g} = [\mathbf{g}_1^T, \ldots, \mathbf{g}_p^T]^T \in \mathbf{R}^{np}$, $\mathbf{T} = [\mathbf{T}_{h_1}, \ldots, \mathbf{T}_{h_p}] \in \mathbf{R}^{(n+m-1)\times(np)}$, and the uncertainty set

$$\mathcal{S} = \Big\{\mathbf{Q} : \mathbf{Q} = \mathbf{diag}(\mathbf{Q}_1, \ldots, \mathbf{Q}_p), \mathbf{Q}_j = \alpha_j\boldsymbol{\Lambda}_j, \sum_{j=1}^p \alpha_j \le 1, \alpha_j \ge 0, j = 1, \ldots, p\Big\}, \qquad (62)$$

belongs to the class of polytopic uncertainty sets described in (8). Therefore, (61) can be reformulated as an SOCP.

3.5 Robust estimation in uncertain statistical models

Suppose $\mathbf{x} \in \mathbf{R}^n$ is a Gaussian random variable with *a priori* distribution $x \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with an unknown mean $\boldsymbol{\mu}$ and covariance

$$\boldsymbol{\Sigma} \in \mathcal{S}_1 = \Big\{\boldsymbol{\Sigma} : \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Delta} \succeq \mathbf{0}, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T, \big\|\boldsymbol{\Sigma}_0^{\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\Sigma}_0^{\frac{1}{2}}\big\| \le \eta\Big\}. \qquad (63)$$

We will assume that $\eta < 1$. The structure (63) is precisely the confidence region associated with the maximum likelihood estimate of the covariance $\boldsymbol{\Sigma}$ of $\mathbf{x}$. See [16] for details.

Suppose a vector of measurements $\mathbf{y} \in \mathbf{R}^m$ is given by the linear observation model

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d}, \qquad (64)$$

where $\mathbf{C} \in \mathbf{R}^{m \times n}$ is the known regression matrix, and the disturbance vector $\mathbf{d} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$, independent of $\mathbf{x}$, with

$$\mathbf{D} \in \mathcal{S}_2 = \left\{ \mathbf{D} : \begin{array}{l} \mathbf{D} = \mathbf{V}^T \mathbf{F} \mathbf{V}, \mathbf{F} = \mathbf{F}_0 + \boldsymbol{\Delta} \succeq \mathbf{0}, \|\mathbf{N}^{-\frac{1}{2}} \boldsymbol{\Delta} \mathbf{N}^{-\frac{1}{2}}\| \leq \eta, \\ \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\| \leq \rho_i, i = 1, \ldots, m, \|\boldsymbol{\Delta}\| \leq \eta \end{array} \right\}. \tag{65}$$

The uncertainty set (19) is quite general. For example, one can control the rank of the covariance matrix $\mathbf{D}$ by appropriately setting the dimension of $\mathbf{F}_0$ and model any norm-like perturbation by suitably choosing $\mathbf{N}$.

Given the vector of observations $\mathbf{y}$ and an *a priori* unbiased estimate $\bar{\boldsymbol{\mu}}$ of the mean vector $\boldsymbol{\mu}$, we consider a linear unbiased estimator of the form

$$\widehat{\boldsymbol{\mu}} = (\mathbf{I} - \mathbf{K}\mathbf{C})\bar{\boldsymbol{\mu}} + \mathbf{K}\mathbf{y}, \tag{66}$$

where the gain matrix $\mathbf{K} \in \mathbf{R}^{n \times m}$ is to be determined. Since $\bar{\mu}$ is unbiased, the estimate $\widehat{\boldsymbol{\mu}}$ is also unbiased. The covariance $\mathbf{P}$ of the *a posteriori* estimate $\widehat{\boldsymbol{\mu}}$ is given by

$$\mathbf{P} \equiv \mathbf{E}[(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T] = (\mathbf{I} - \mathbf{K}\mathbf{C})^T \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{K}\mathbf{C}) + \mathbf{K}^T \mathbf{D} \mathbf{K}. \tag{67}$$

The non-robust version of this measurement model (i.e. $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ and $\mathbf{D} = \mathbf{D}_0$ for fixed $\boldsymbol{\Sigma}_0$ and $\mathbf{D}_0$) is the well-known Gaussian linear stochastic model [14]. The robust measurement model developed here is a variant of the model proposed by Calafiore and El Ghaoui [9] where the *a priori* covariance $\boldsymbol{\Sigma}$ was known exactly and the noise covariance

$$\mathbf{D} \in \left\{ \mathbf{D} : \mathbf{D}^{-1} = \mathbf{D}_0^{-1} + \mathbf{L}\boldsymbol{\Delta}\mathbf{R} + \mathbf{R}^T \boldsymbol{\Delta}^T \mathbf{L}^T \succeq \mathbf{0}, \|\boldsymbol{\Delta}\| \leq 1 \right\}.$$

They show that the problem of choosing the gain matrix $\mathbf{K}$ to minimize the worst-case value of $\mathbf{Tr}(\mathbf{P})$ or $\mathbf{det}(\mathbf{P})$ can be reduced to an SDP.

In this paper, we are interested in minimizing the worst-case variance along a given fixed set of vectors $\{\mathbf{v}_j : \|\mathbf{v}_j\| = 1, j = 1, \ldots, k\}$, i.e. we want to solve the following optimization problem

$$\min_{\mathbf{K}} \max_{\{\mathbf{D} \in \mathcal{S}\}} \max_{\{1 \leq j \leq k\}} \left\{ \mathbf{v}_j^T (\mathbf{I} - \mathbf{K}\mathbf{C})^T \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{K}\mathbf{C}) \mathbf{v}_j + \mathbf{v}_j^T \mathbf{K}^T \mathbf{D} \mathbf{K} \mathbf{v}_j \right\}, \tag{68}$$

or equivalently, the robust quadratically constrained problem,

$$\begin{array}{rll} \text{minimize} & \nu, & \\ \text{subject to} & \mathbf{v}_j^T (\mathbf{I} - \mathbf{K}\mathbf{C})^T \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{K}\mathbf{C}) \mathbf{v}_j \leq \delta_j, & \boldsymbol{\Sigma} \in \mathcal{S}_1, j = 1, \ldots, k, \\ & \mathbf{v}_j^T \mathbf{K}^T \mathbf{D} \mathbf{K} \mathbf{v}_j \leq \nu - \delta_j, & \forall \mathbf{D} \in \mathcal{S}_2, j = 1, \ldots, k. \end{array} \tag{69}$$

If we fix $\mathbf{K}$, from Lemma 3 in [16] it follows that

$$\max_{\boldsymbol{\Sigma} \in \mathcal{S}_1} \left\{ \mathbf{v}_j^T (\mathbf{I} - \mathbf{K}\mathbf{C})^T \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{K}\mathbf{C}) \mathbf{v}_j \right\} = \begin{cases} \infty & \eta \geq 1, \\ \frac{1}{(1-\eta)} \mathbf{v}_j^T (\mathbf{I} - \mathbf{K}\mathbf{C})^T \boldsymbol{\Sigma}_0 (\mathbf{I} - \mathbf{K}\mathbf{C}) \mathbf{v}_j, & \eta < 1. \end{cases}$$

Thus, $\eta < 1$ implies that the first constraint in (69) can be reformulated as a collection of SOC constraints. Fix an index j and let $\mathbf{y}_j = \mathbf{K}\mathbf{v}_j$. Since the uncertainty set $\mathcal{S}$ belongs to the to the class of factorized uncertainty sets defined in (19), Lemma 3 implies that that the robust quadratic constraint $\mathbf{v}_j^T \mathbf{K}^T \mathbf{D} \mathbf{K} \mathbf{v}_j = \mathbf{y}_j^T \mathbf{D} \mathbf{y}_j \leq \nu - \delta_j$, for all $\mathbf{D} \in \mathcal{S}_2$, can be reformulated as a collection of linear and SOC constraints. Thus, the robust problem (69) can be transformed into an SOCP.

4 Conclusion

In this paper we study robust convex quadratically constrained programs. Ben-Tal and Nemirovski initiated the study of these problems and showed that for generalized ellipsoidal uncertainty sets these robust problems can be reformulated as SDPs [4] (see also [6]). In this work, our focus was on identifying uncertainty structures that allow an SOCP reformulation for the corresponding robust convex QCPs. In Section 2 we proposed three different classes of uncertainty sets that meet this criterion.

Adding robustness reduces the sensitivity of the optimal decision to fluctuations in the parameters and can often result in significant improvement in performance [3, 16, 27]. Typically, the complexity of the deterministic reformulation of the robust problem is higher than the non-robust version of the problem. However, since the worst case complexity of convex QCPs is comparable to that of SOCPs, the results in this paper show that one can add robustness to convex QCPs with a relatively modest increase in the computational effort. Moreover, the examples presented in Section 3 show that the natural uncertainty sets for optimization problems arising a wide variety of application areas belong to the classes introduced in Section 2.

An important issue with regards to robustness is that of parametrization of the uncertainty structures, i.e. setting the parameters that define the uncertainty structures. In some cases, such as the polytopic uncertainty (8), the parametrization is clear – the uncertainty set is defined by scenario analysis. However, in others, such as the factorized uncertainty set (19), the parametrization is not obvious – in [16] it is shown that the factorized uncertainty set is parametrized by the confidence regions corresponding to statistical technique used to estimate the parameters of the original non-robust problem.

References

[1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5(1):13–51, 1995.

[2] F. Alizadeh and D. Goldfarb. Second-order cone programming. To appear in *Math Programming*.

[3] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semidefinite programming. *SIAM J. Optim.*, 7(4):991–1016, 1997.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[5] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, 1999.

[6] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, algorithms, and engineering applications.

[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory.* SIAM, Philadelphia, PA, 1994.

[8] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[9] G. Calafiore and L. El Ghaoui. Minimum variance estimation with uncertain statistical model. In *Proc. of 40th Conf. Dec. and Cont.*, 2001.

[10] V. K. Chopra and W. T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *J. Portfolio Manag.*, (Winter):6–11, 1993.

[11] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.*, 18(4):1035–1064, 1997.

[12] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.

[13] G. H. Golub and C. Van Loan. *Matrix Computations.* John Hopkins Press, 3rd edition, 1996.

[14] P. J. Huber. *Robust Statistics.* John Wiley & Sons, 1981.

[15] G. Iyengar and D. Goldfarb. Robust portfolio selection problems. *Submitted to Math Programming.*

[16] G. Iyengar and D. Goldfarb. Robust portfolio selection problems. Technical Report TR-2001-05, Computational Optimization Research Center (CORC), IEOR Department, Columbia University, 2001. Available at *http://www.corc.ieor.columbia.edu/reports/techreports.html.*

[17] M. S. Lobo, M. Fazel, and S. Boyd. Portfolio optimization with linear and fixed transaction costs and bounds on risk. Submitted to *Operations Research*, June 2000.

[18] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra Appl.*, 284(1-3):193–228, 1998.

[19] O. L. Mangasarian and D. R. Musicant. Robust linear and support vector regression. *IEEE Trans. PAMI*, 22, 9 2000.

[20] H. M. Markowitz. Portfolio selection. *J. Finance*, 7:77–91, 1952.

[21] H. M. Markowitz. *Portfolio Selection.* Wiley, New York, 1959.

[22] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, Philadelphia, 1993.

[23] D. Pal, G. N. Iyengar, and J. M. Cioffi. A new method for channel shortening with applications to discrete multi-tone - part i : Theory. Submitted to IEEE Transactions on Communications.

[24] D. Pal, G. N. Iyengar, and Cioffi J. M. A new method of channel shortening with applications to discrete multi-tone (dmt) systems. In *Proceeding of the IEEE International Conference on Communications*, volume 2, pages 763–768, 1998.

[25] J. G. Proakis. *Digital Communication*. McGraw-Hill, 4th edition, 2000.

[26] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Reveiw*, 38:49–95, 1996.

[27] L. Vandenberghe, S. Boyd, and M. Nouralishahi. Robust linear programming and optimal control. In *Proceedings of the 15th IFAC World Congress on Automatic Control*, 2002.

[28] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, Inc., New York, 1998.