Math. Program., Ser. B 97: 495–515 (2003)

Digital Object Identifier (DOI) 10.1007/s10107-003-0425-3

D. Goldfarb · G. Iyengar

Robust convex quadratically constrained programs

Received: July 3, 2002 / Accepted: April 7, 2003
Published online: May 7, 2003 – © Springer-Verlag 2003

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]. In contrast to [4], where it is shown that such robust problems can be formulated as semidefinite programs, our focus in this paper is to identify uncertainty sets that allow this class of problems to be formulated as second-order cone programs (SOCP). We propose three classes of uncertainty sets for which the robust problem can be reformulated as an explicit SOCP and present examples where these classes of uncertainty sets are natural.

1. Problem formulation

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

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

where the vector of decision variables $\mathbf{x} \in \mathbf{R}^n$, and the parameters $\mathbf{c} \in \mathbf{R}^n$, $\gamma_i \in \mathbf{R}$, $\mathbf{q}_i \in \mathbf{R}^n$ and $\mathbf{Q}_i \in \mathbf{R}^{n\times n}$, $\mathbf{Q}_i \succeq \mathbf{0}$ (i.e. $\mathbf{Q}_i$ is positive semidefinite), for all $i = 1, \ldots, p$. Note that there is no loss of generality in assuming that the objective is linear. Since each $\mathbf{Q}_i \succeq \mathbf{0}$, $i = 1, \ldots, p$, it is easily verified that the convex QCP (1) is equivalent to the following second-order cone program (SOCP):

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

where $\mathbf{Q}_i = \mathbf{V}_i^T\mathbf{V}_i$, for some $\mathbf{V}_i \in \mathbf{R}^{m_i\times n}$, $i = 1, \ldots, p$, and $\|v\|$ denotes the $\mathcal{L}_n$ norm of a vector v. For a discussion of SOCPs and their applications see [2, 17, 21].

Formulations (1) and (2) 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 parameters are estimated from data, and are, therefore, subject to measurement and statistical errors [13]. Since the solutions to optimization problems are typically sensitive to parameter perturbations, errors in the input parameters tend to get amplified in the decision vector, often resulting in far from optimal solutions [3, 10].

D. Goldfarb: IEOR Department, Columbia University, e-mail: gold@ieor.columbia.edu.

Research partially supported by DOE grant GE-FG01-92ER-25126, NSF grants DMS-94-14438, CDA-97-26385, DMS-01-04282 and ONR grant N000140310514.

G. Iyengar: IEOR Department, Columbia University, e-mail: garud@ieor.columbia.edu.

Research partially supported by NSF grants CCR-00-09972, DMS-01-04282 and ONR grant N000140310514.

The problem of choosing an optimal decision vector in the presence of parameter perturbations was formalized by Ben-Tal and Nemirovski [4, 5] as the following *robust optimization problem*:

$$\begin{array}{ll} \text{minimize } \mathbf{c}^T\mathbf{x} & \\ \text{subject to } F(\mathbf{x}, \boldsymbol{\xi}) \in \mathcal{K} \subset \mathbf{R}^m, & \forall \boldsymbol{\xi} \in \mathcal{U}, \end{array} \tag{3}$$

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]. Ben-Tal and Nemirovski established that, for certain classes of uncertainty sets $\mathcal{U}$, robust counterparts of linear programs, quadratic programs, QCPs, and semidefinite programs (SDPs) are themselves tractable optimization problems. Robustness as applied to least squares problems and SDPs was independently studied by El Ghaoui and his collaborators [11, 12].

In keeping with the above formulation, a generic robust convex QCP is given by

$$\begin{array}{l} \text{minimize } \mathbf{c}^T\mathbf{x} \\ \text{subject to } \mathbf{x}^T\mathbf{Q}_i\mathbf{x} + 2\mathbf{q}_i^T\mathbf{x} + \gamma_i \leq 0, \text{ for all } (\mathbf{Q}_i, \mathbf{q}_i, \gamma_i) \in \mathcal{S}_i, i = 1, \dots, p. \end{array} \tag{4}$$

Ben-Tal and Nemirovski [4] showed that a version of (4) in which the uncertainty structures $\mathcal{S}_i$ are generalized ellipsoids can be reduced to an SDP [1, 21, 25]. In this paper we explore uncertainty structures for which (4) can be reformulated as an SOCP. We note that both the worst case and practical computational effort required to solve an SOCP is at least an order of magnitude less than that needed to solve an SDP of comparable size [2].

In Section 2 we describe three classes of uncertainty sets for which (4) can be reduced to an explicit 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 (4) can be reformulated as an SOCP.

2.1. Discrete and polytopic uncertainty sets

The simplest type of uncertainty sets is a discrete set defined as follows:

$$\mathcal{S}_a = \left\{(\mathbf{Q}, \mathbf{q}, \gamma) : (\mathbf{Q}, \mathbf{q}, \gamma) = (\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \mathbf{Q}_j \succeq \mathbf{0}, j = 1, \dots, k\right\}. \tag{5}$$

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

$$\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j \leq 0, \quad \forall j = 1, \dots, k, \tag{6}$$

or equivalently, k second-order cone (SOC) constraints.

The discrete uncertainty set (5) typically arises when one wants to be robust against several scenarios – each $(\mathbf{Q}_j, \mathbf{q}_j, \gamma_j)$ corresponds to a particular scenario (e.g., see [16]). The convex hull of the discrete uncertainty set $\mathcal{S}_a$ is the uncertainty set:

$$\mathcal{S}_{a'} = \left\{(\mathbf{Q}, \mathbf{q}, \gamma) : (\mathbf{Q}, \mathbf{q}, \gamma) = \sum_{j=1}^{k} \lambda_j(\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \mathbf{Q}_j \succeq \mathbf{0}, \lambda_j \geq 0, \ \forall j, \sum_{j=1}^{k} \lambda_j = 1\right\}. \tag{7}$$

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 $\sum_{j=1}^{k} \lambda_j(\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j) \leq 0$, for all $\lambda_j \geq 0$, $\sum_{j=1}^{k} \lambda_j = 1$. Since this is, in turn, equivalent to the set of constraints (6), it follows that a robust quadratic constraint with respect $\mathcal{S}_{a'}$ reduces to k SOC constraints.

The uncertainty sets (5) and (7) can be further extended to the following polytopic uncertainty set:

$$\mathcal{S}_b = \left\{(\mathbf{Q}, \mathbf{q}, \gamma) : \begin{array}{l} (\mathbf{Q}, \mathbf{q}, \gamma) = \sum_{j=1}^{k} \lambda_j(\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \mathbf{Q}_j \succeq \mathbf{0}, \ j = 1, \ldots, k, \\ \mathbf{A}\boldsymbol{\lambda} = \mathbf{b}, \boldsymbol{\lambda} \geq \mathbf{0}. \end{array}\right\}, \tag{8}$$

where $\{\lambda \in R^k : A\lambda = b, \lambda \geq 0\} \neq \emptyset$.

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* $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_b$, *where* $\mathcal{S}_b$ *is defined in (8), if and only if there exist* $\boldsymbol{\mu} \in \mathbf{R}^k$ *satisfying*

$$\begin{array}{c} \mathbf{b}^T\boldsymbol{\mu} \leq 0, \\ \left\| \begin{bmatrix} 2\mathbf{V}_j\mathbf{x} \\ 1 + \gamma_j + 2\mathbf{q}_j^T\mathbf{x} - \mathbf{A}_j^T\boldsymbol{\mu} \end{bmatrix} \right\| \leq 1 - \gamma_j - 2\mathbf{q}_j^T\mathbf{x} + \mathbf{A}_j^T\boldsymbol{\mu}, \ j = 1, \ldots, k, \end{array} \tag{9}$$

where $\mathbf{A}_j$ *is the* j*-th column of* $\mathbf{A}$ *and* $\mathbf{Q}_j = \mathbf{V}_j^T\mathbf{V}_j$, $j = 1, \ldots, k$.

Proof. Fix $\mathbf{x}$ and define $c_j = \mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j$, $j = 1, \ldots, k$. Then the constraint $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \leq 0$, for all $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_b$ is equivalent to

$$\mathbf{c}^T\boldsymbol{\lambda} \leq 0, \ \forall \boldsymbol{\lambda} \geq \mathbf{0} \text{ such that } \mathbf{A}\boldsymbol{\lambda} = \mathbf{b}. \tag{10}$$

By linear programming duality (10) is equivalent to

$$\exists \ \boldsymbol{\mu} \text{ such that } \mathbf{A}^T\boldsymbol{\mu} \geq \mathbf{c}, \ \mathbf{b}^T\boldsymbol{\mu} \leq 0. \tag{11}$$

The result now follows by expressing (11) as a collection of SOC constraints, using the fact (see Section 6.2.3) in [21]) that for $\mathbf{z} \in \mathbf{R}^n$, $x \in \mathbf{R}$, and $y \in \mathbf{R}$, $x, y \geq 0$,

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

□

2.2. Norm-constrained uncertainty sets

Next, we describe two closely related norm-constrained uncertainty sets that are restricted versions of the generalized ellipsoidal uncertainty sets introduced in [4]. In the first uncertainty set $\mathcal{S}_c$, all of the parameters $(\mathbf{Q}, \mathbf{q}, \gamma)$ are determined by the same set of perturbation parameters $\mathbf{u}$, i.e.

$$\mathcal{S}_c = \left\{ (\mathbf{Q}, \mathbf{q}, \gamma) : \begin{array}{l} (\mathbf{Q}, \mathbf{q}, \gamma) = (\mathbf{Q}_0, \mathbf{q}_0, \gamma_0) + \sum_{j=1}^k u_j (\mathbf{Q}_j, \mathbf{q}_j, \gamma_j), \\ \mathbf{Q}_j \succeq \mathbf{0},\ j = 0, \ldots, k, \mathbf{u} \geq \mathbf{0},\ \|\mathbf{u}\|_p \leq 1 \end{array} \right\}, \tag{12}$$

where $\|\mathbf{u}\|_p$ is the $\mathcal{L}_p$ norm for some rational $p \geq 1$. Note that for all rational $p \geq 1$ the constraint $\|\mathbf{u}\|_p \leq t$ can be reformulated as a collection of SOC constraints [2, 6].

Remark 1. The robust problem (4) with respect to $\mathcal{S}_c$ is NP-hard if the sign constraint on $\mathbf{u}$ is relaxed or if any of the $\mathbf{Q}_j$'s are indefinite [4].

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 $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_c$, where $\mathcal{S}_c$ is defined in (12), if and only if there exist $\mathbf{f} \in \mathbf{R}^k_+$ and $\nu \geq 0$ satisfying*

$$\begin{aligned} \left\| \begin{bmatrix} 2\mathbf{V}_j\mathbf{x} \\ 1 - f_j + \gamma_j + 2\mathbf{q}_j^T\mathbf{x} \end{bmatrix} \right\| &\leq 1 + f_j - \gamma_j - 2\mathbf{q}_j^T\mathbf{x},\ j = 1, \ldots, k, \\ \left\| \begin{bmatrix} 2\mathbf{V}_0\mathbf{x} \\ 1 - \nu \end{bmatrix} \right\| &\leq 1 + \nu, \\ \|\mathbf{f}\|_q &\leq -\nu - 2\mathbf{q}_0^T\mathbf{x} - \gamma_0, \end{aligned} \tag{13}$$

where $\frac{1}{p} + \frac{1}{q} = 1$, and $\mathbf{Q}_j = \mathbf{V}_j^T\mathbf{V}_j$, $j = 0, \ldots, k$.

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

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \max_{\{\mathbf{u}:\mathbf{u}\geq\mathbf{0}, \|\mathbf{u}\|_p\leq 1\}} \left\{ \sum_{j=1}^k u_j(\mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j) \right\} \leq 0. \tag{14}$$

Define $y_j = \mathbf{x}^T\mathbf{Q}_j\mathbf{x} + 2\mathbf{q}_j^T\mathbf{x} + \gamma_j$ and $z_j = \max\{y_j, 0\}$, $j = 1, \ldots, k$.

Suppose $p > 1$. Then the optimal solution $\mathbf{u}^*$ of the convex program $\max_{\{\mathbf{u}:\mathbf{u}\geq\mathbf{0}, \|\mathbf{u}\|_p\leq 1\}} \{\mathbf{u}^T\mathbf{y}\}$ is

$$\mathbf{u}_j^* = \begin{cases} \dfrac{(z_j)^{\frac{1}{p-1}}}{(\|\mathbf{z}\|_q)^{\frac{1}{p-1}}}, & \mathbf{z} \neq \mathbf{0}, \\ 0, & \text{otherwise}, \end{cases}$$

for $j = 1, \ldots, k$. Thus, (14) is equivalent to

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \|\mathbf{z}\|_q \leq 0. \tag{15}$$

For $p = 1$, the optimal value of the convex program $\max_{\{\mathbf{u}:\mathbf{u}\geq\mathbf{0}, \|\mathbf{u}\|_1\leq 1\}} \{\mathbf{u}^T\mathbf{y}\}$ is $\max_{1\leq j\leq k}\{z_j\}$. Thus, (14) is once again equivalent to

$$\mathbf{x}^T\mathbf{Q}_0\mathbf{x} + 2\mathbf{q}_0^T\mathbf{x} + \gamma_0 + \|\mathbf{z}\|_q \leq 0. \tag{16}$$

Moreover, since $\mathbf{z} \geq \mathbf{0}$ implies $\|\mathbf{f}\|_q \geq \|\mathbf{z}\|_q$ for all $f_j \geq z_j$ (i.e. $f_j \geq 0$ and $f_j \geq y_j$), $j = 1, \ldots, k$, (14) holds if and only if there exists $\mathbf{f} \geq \mathbf{0}$ and $\mathbf{f} \geq \mathbf{y}$ such that

$$\mathbf{x}^T \mathbf{Q}_0 \mathbf{x} + 2\mathbf{q}_0^T \mathbf{x} + \gamma_0 + \|\mathbf{f}\|_q \leq 0. \tag{17}$$

The result follows by expressing (17) and $f_j \geq y_j$, $j = 1, \ldots, k$, as SOC constraints. □

In $\mathcal{S}_c$ the perturbations in the quadratic term $\mathbf{Q}$ and the affine term $(\mathbf{q}, \gamma)$ are determined by the same parameter $\mathbf{u}$. However, in many applications the uncertainty in the quadratic and affine terms are independent [13]. We model this by the following uncertainty structure,

$$\mathcal{S}_d = \left\{ (\mathbf{Q}, \mathbf{q}, \gamma) : \begin{array}{l} \mathbf{Q} = \mathbf{Q}_0 + \sum_{j=1}^k u_j \mathbf{Q}_j,\ \mathbf{Q}_j \succeq \mathbf{0},\ j = 0, \ldots, k,\ \|\mathbf{u}\|_p \leq 1, \\ (\mathbf{q}, \gamma) = (\mathbf{q}_0, \gamma_0) + \sum_{j=1}^k v_j (\mathbf{q}_j, \gamma_j),\ \|\mathbf{v}\|_r \leq 1 \end{array} \right\}, \tag{18}$$

where $p, r \geq 1$ and rational.

Remark 2. Although $\mathbf{u}$ is unrestricted in sign, the constraints $\mathbf{Q}_j \succeq \mathbf{0}$ ensure that the worst case perturbation $\mathbf{u}^* \geq \mathbf{0}$. As in Remark 1, indefinite $\mathbf{Q}_j$ result in an NP-hard optimization problem [4].

Lemma 3. *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 $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_d$, where $\mathcal{S}_d$ is defined in (18), if and only if there exist $\mathbf{f}, \mathbf{g} \in \mathbf{R}^k$ and $\nu \geq 0$ such that*

$$\begin{aligned} g_j &= 2\mathbf{q}_j^T \mathbf{x} + \gamma_j, && j = 1, \ldots, k, \\ \left\| \begin{bmatrix} 2\mathbf{V}_j \mathbf{x} \\ 1 - f_j \end{bmatrix} \right\| &\leq 1 + f_j, && j = 1, \ldots, k, \\ \left\| \begin{bmatrix} 2\mathbf{V}_0 \mathbf{x} \\ 1 - \nu \end{bmatrix} \right\| &\leq 1 + \nu, \\ \|\mathbf{f}\|_q + \|\mathbf{g}\|_s &\leq -\nu - 2\mathbf{q}_0^T \mathbf{x} - \gamma_0, \end{aligned} \tag{19}$$

where $\frac{1}{p} + \frac{1}{q} = 1$, $\frac{1}{r} + \frac{1}{s} = 1$, and $\mathbf{Q}_j = \mathbf{V}_j^T \mathbf{V}_j$, $j = 0, \ldots, k$. $\mathbf{Q}_j = \mathbf{V}_j^T \mathbf{V}_j$, $j = 0, \ldots, k$.

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

$$\begin{aligned} \mathbf{x}^T \mathbf{Q}_0 \mathbf{x} + 2\mathbf{q}_0^T \mathbf{x} + \gamma_0 + \max_{\{\mathbf{u} : \|\mathbf{u}\|_p \leq 1\}} \left\{ \sum_{j=1}^k u_j (\mathbf{x}^T \mathbf{Q}_j \mathbf{x}) \right\} \\ + \max_{\{\mathbf{v} : \|\mathbf{v}\|_r \leq 1\}} \left\{ \sum_{j=1}^k v_j (2\mathbf{q}_j^T \mathbf{x} + \gamma_j) \right\} \leq 0. \end{aligned} \tag{20}$$

The Cauchy-Schwartz inequality implies that

$$\max_{\{\mathbf{u} : \|\mathbf{u}\|_p \leq 1\}} \left\{ \sum_{j=1}^k u_j (\mathbf{x}^T \mathbf{Q}_j \mathbf{x}) \right\} = \|\mathbf{f}\|_q, \tag{21}$$

where

$$f_j = \mathbf{x}^T \mathbf{Q}_j \mathbf{x}, \quad j = 1, \ldots, k, \tag{22}$$

and

$$\max_{\{\mathbf{v}:\|\mathbf{v}\|_r \le 1\}} \left\{ \sum_{j=1}^{k} v_j (2\mathbf{q}_j^T \mathbf{x} + \gamma_j) \right\} = \|\mathbf{g}\|_s ,$$

where $g_j = 2\mathbf{q}_j^T \mathbf{x} + \gamma_j$, $j = 1, \ldots, k$. Thus, (20) is equivalent to

$$\mathbf{x}^T \mathbf{Q}_0 \mathbf{x} + 2\mathbf{q}_0^T \mathbf{x} + \gamma_0 + \|\mathbf{f}\|_q + \|\mathbf{g}\|_s \le 0. \tag{23}$$

Since $\mathbf{Q}_j \succeq \mathbf{0}$, $j = 1, \ldots, k$, it is easy to verify that the constraints (22) can be relaxed to

$$f_j \ge \mathbf{x}^T \mathbf{Q}_j \mathbf{x}, \quad j = 1, \ldots, k, \tag{24}$$

without affecting the conclusion (23). The result follows from rewriting (23) and (24) as a collection of linear and SOC constraints. □

2.3. Factorized uncertainty sets

The next class of uncertainty sets is defined as follows.

$$\mathcal{S}_e = \left\{ (\mathbf{Q}, \mathbf{q}, \gamma_0) : \begin{array}{l} \mathbf{Q} = \mathbf{V}^T \mathbf{F} \mathbf{V}, \mathbf{F} \in \mathbf{R}^{m \times m}, \mathbf{V} \in \mathbf{R}^{m \times n}, \\ \mathbf{F} = \mathbf{F}_0 + \boldsymbol{\Delta} \succeq \mathbf{0}, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T, \|\mathbf{N}^{-\frac{1}{2}} \boldsymbol{\Delta} \mathbf{N}^{-\frac{1}{2}}\| \le \eta,\ \mathbf{F}_0 \succeq \mathbf{0},\ \mathbf{N} \succ \mathbf{0}, \\ \mathbf{V} = \mathbf{V}_0 + \mathbf{W},\ \|\mathbf{W}_i\|_g = \sqrt{\mathbf{W}_i^T \mathbf{G} \mathbf{W}_i} \le \rho_i, \forall i,\ \mathbf{G} \succ \mathbf{0}, \\ \mathbf{q} = \mathbf{q}_0 + \boldsymbol{\zeta} \in \mathbf{R}^n,\ \|\boldsymbol{\zeta}\|_s = \sqrt{\boldsymbol{\zeta}^T \mathbf{S} \boldsymbol{\zeta}} \le \delta,\ \mathbf{S} \succ \mathbf{0}. \end{array} \right\}, \tag{25}$$

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 $\mathbf{A} \in \mathbf{R}^{m \times m}$ is either given by the $\mathcal{L}_2$-norm, i.e. $\|\mathbf{A}\| = \max_{1 \le i \le m} \left\{ |\lambda_i(\mathbf{A})| \right\}$, or the Frobenius norm, i.e. $\|\mathbf{A}\| = \sqrt{\sum_{i=1}^m \lambda_i^2(\mathbf{A})}$, where $\left\{\lambda_i(\mathbf{A}), i = 1, \ldots, m\right\}$ are the eigenvalues of $\mathbf{A}$. The uncertainty structure $\mathcal{S}_e$ in (25) 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.

This class of uncertainty sets captures the structure of the confidence regions around the maximum likelihood estimates of the parameters. See [13] for a detailed discussion of the structure of this uncertainty set and its parametrization.

Lemma 4. *The decision vector* $\mathbf{x} \in \mathbf{R}^n$ *satisfies the robust convex quadratic constraint* $\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + \gamma \le 0$ *for all* $(\mathbf{Q}, \mathbf{q}, \gamma) \in \mathcal{S}_e$*, where* $\mathcal{S}_e$ *is defined in (25), if and only if there exist* $\tau, \nu, \sigma, r \in \mathbf{R}$*,* $\mathbf{u} \in \mathbf{R}^n$*,* $\mathbf{w} \in \mathbf{R}^m$ *and* $\mathbf{t} \in \mathbf{R}^m_+$ *such that*

$$\begin{aligned}
\tau &\ge 0,\\
\nu &\ge \tau + \mathbf{1}^T\mathbf{t}\\
\sigma &\le \frac{1}{\lambda_{\max}(\mathbf{H})},\\
r &\ge \textstyle\sum_{i=1}^n \rho_i u_i,\\
u_j &\ge x_j, \qquad j = 1, \ldots, n,\\
u_j &\ge -x_j, \qquad j = 1, \ldots, n,\\
\left\| \begin{bmatrix} 2r \\ \sigma - \tau \end{bmatrix} \right\| &\le \sigma + \tau,\\
\left\| \begin{bmatrix} 2w_i \\ (\lambda_i - \sigma - t_i) \end{bmatrix} \right\| &\le (\lambda_i - \sigma + \tau_i), \qquad i = 1, \ldots, m,\\
2\delta \left\| \mathbf{S}^{-\frac{1}{2}}\mathbf{x} \right\| &\le -\nu - 2\mathbf{q}_0^T\mathbf{x} - \gamma_0,
\end{aligned}$$

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* $\mathbf{H}$*,* $\mathbf{\Lambda} = \mathbf{diag}(\boldsymbol{\lambda})$*,* $\lambda_{\max}(\mathbf{H}) = \max_{1\le i\le m}\{\lambda_i\}$*, and* $\mathbf{w} = \mathbf{Q}^T\mathbf{H}^{\frac{1}{2}}\mathbf{G}^{\frac{1}{2}}\mathbf{V}_0\mathbf{x}$.

Proof. First fix $\mathbf{W}$, or equivalently fix $\mathbf{V}$. Thus, only $\mathbf{\Delta}$ and $\boldsymbol{\zeta}$ are variable. Define $\tilde{\mathbf{\Delta}} = \mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}$, $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ and
$\mathcal{S}_1 = \left\{\mathbf{F} : \mathbf{F} = \mathbf{F}_0 + \mathbf{\Delta} \succeq \mathbf{0}, \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}\| \le \eta\right\}$. Then

$$\begin{aligned}
&\max\left\{\mathbf{x}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\mathbf{x} : \mathbf{F} \in \mathcal{S}_1\right\}\\
&\quad= \max\left\{\mathbf{y}^T(\mathbf{F}_0 + \mathbf{\Delta})\mathbf{y} : \mathbf{\Delta} = \mathbf{\Delta}^T, \|\mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}\| \le \eta, \mathbf{F}_0 + \mathbf{\Delta} \succeq \mathbf{0}\right\}\\
&\quad= \max\left\{\mathbf{y}^T\mathbf{F}_0\mathbf{y} + (\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T\tilde{\mathbf{\Delta}}(\mathbf{N}^{\frac{1}{2}}\mathbf{y}) : \|\tilde{\mathbf{\Delta}}\| \le \eta, \mathbf{F}_0 + \mathbf{N}^{\frac{1}{2}}\tilde{\mathbf{\Delta}}\mathbf{N}^{\frac{1}{2}} \succeq \mathbf{0}\right\},\\
&\quad\le \max\left\{\mathbf{y}^T\mathbf{F}_0\mathbf{y} + (\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T\tilde{\mathbf{\Delta}}(\mathbf{N}^{\frac{1}{2}}\mathbf{y}) : \|\tilde{\mathbf{\Delta}}\| \le \eta\right\}, &(26)\\
&\quad\le \mathbf{y}^T\mathbf{F}_0\mathbf{y} + \eta(\mathbf{N}^{\frac{1}{2}}\mathbf{y})^T(\mathbf{N}^{\frac{1}{2}}\mathbf{y}), &(27)
\end{aligned}$$

where (26) 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 (27) follows from the properties of the matrix norm.

Since $\|\tilde{\mathbf{\Delta}}\| = \max_{1\le i\le m}\left\{|\lambda_i(\tilde{\mathbf{\Delta}})|\right\}$ or $\sqrt{\sum_{i=1}^m \lambda_i^2(\tilde{\mathbf{\Delta}})}$ and $\mathbf{N} \succ \mathbf{0}$, the bound (27) 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 (26) 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 (26) is, in fact, an equality, i.e.

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

Therefore, using the fact that $\boldsymbol{\zeta}^T \mathbf{x} \leq \boldsymbol{\zeta}^T \mathbf{S}^{\frac{1}{2}} \mathbf{S}^{-\frac{1}{2}} \mathbf{x} \leq \|\mathbf{S}^{\frac{1}{2}} \boldsymbol{\zeta}\| \|\mathbf{S}^{-\frac{1}{2}} \mathbf{x}\| \leq \delta \|\mathbf{S}^{-\frac{1}{2}} \mathbf{x}\|$,

$$\max_{\{(\mathbf{Q},\mathbf{q},\gamma)\in S_e\}} \left\{\mathbf{x}^T \mathbf{Q}\mathbf{x} + 2\mathbf{q}^T \mathbf{x} + \gamma\right\}$$
$$= \gamma_0 + 2\mathbf{q}_0^T \mathbf{x} + 2\delta \|\mathbf{S}^{-\frac{1}{2}} \mathbf{x}\| + \max_{\mathbf{V}\in S_v} \left\{\mathbf{x}^T \mathbf{V}^T (\mathbf{F}_0 + \eta \mathbf{N}) \mathbf{V}\mathbf{x}\right\}, \tag{28}$$

where $S_v = \{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\|_g \leq \rho_i, i = 1, \ldots, n\}$. From the definition of S_v, it follows that

$$\max_{\{\mathbf{V}\in S_v\}} \left\{\mathbf{x}^T \mathbf{V}^T (\mathbf{F}_0 + \eta \mathbf{N}) \mathbf{V}\mathbf{x}\right\}$$
$$= \max_{\left\{\mathbf{W}:\|\mathbf{W}_i\|_g \leq \rho_i, i=1,\ldots,n\right\}} \left\{(\mathbf{V}_0 \mathbf{x} + \mathbf{W}\mathbf{x})^T (\mathbf{F}_0 + \eta \mathbf{N})(\mathbf{V}_0 \mathbf{x} + \mathbf{W}\mathbf{x})\right\}. \tag{29}$$

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

$$\|\mathbf{W}\mathbf{x}\|_g = \left\|\sum_{i=1}^{n} x_i \mathbf{W}_i\right\|_g \leq \sum_{i=1}^{n} |x_i| \, \|\mathbf{W}_i\|_g \leq \sum_{i=1}^{n} \rho_i \, |x_i| \,, \tag{30}$$

the optimization problem,

$$\begin{array}{ll} \text{maximize} & (\mathbf{V}_0 \mathbf{x} + \mathbf{v})^T (\mathbf{F}_0 + \eta \mathbf{N})(\mathbf{V}_0 \mathbf{x} + \mathbf{v}) \\ \text{subject to} & \|\mathbf{v}\|_g \leq \sum_{i=1}^{n} \rho_i \, |x_i| \,, \end{array} \tag{31}$$

is a relaxation of (29). The objective function in (31) is convex in $\mathbf{v}$; therefore, the optimal solution $\mathbf{v}^*$ lies on the boundary of the feasible set, i.e. $\|\mathbf{v}^*\|_g = \boldsymbol{\rho}^T |\mathbf{x}|$, where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^T$.

For $i = 1, \ldots, n$, define

$$\mathbf{W}_i = \begin{cases} \rho_i \frac{|x_i|}{x_i} \frac{\mathbf{v}^*}{\|\mathbf{v}^*\|_g}, & x_i \neq 0, \\ \rho_i \frac{\mathbf{v}^*}{\|\mathbf{v}^*\|_g}, & \text{otherwise} \end{cases} \tag{32}$$

Clearly, the collection $\{\mathbf{W}_i : i = 1, \ldots, n\}$ is feasible for (29). Moreover,

$$\mathbf{W}\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{W}_i = \left(\boldsymbol{\rho}^T |\mathbf{x}|\right) \frac{\mathbf{v}^*}{\|\mathbf{v}^*\|_g} = \mathbf{v}^*,$$

i.e. the objective in (29) evaluated at $\mathbf{W}$ defined in (32) is equal to the objective in (31) evaluated at $\mathbf{v}^*$. Thus, (29) and (31) 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

$$\left(\mathbf{V}_0 \mathbf{x} + (\boldsymbol{\rho}^T |\mathbf{x}|)\mathbf{v}\right)^T \left(\mathbf{F}_0 + \eta \mathbf{N}\right)\left(\mathbf{V}_0 \mathbf{x} + (\boldsymbol{\rho}^T |\mathbf{x}|)\mathbf{v}\right) \leq \nu \tag{33}$$

for all $\|\mathbf{v}\|_g \leq 1$, i.e. $1 - \mathbf{v}^T \mathbf{G}\mathbf{v} \geq 0$. Define $\mathbf{y}_0 = \mathbf{V}_0 \mathbf{x}$. Then it is easy to establish (see [13]) that (33) holds for all $\mathbf{v}$ with $\|\mathbf{v}\|_g \leq 1$ if, and only if, there exists $r \geq \boldsymbol{\rho}^T |\mathbf{x}|$ such that

$$\nu - \mathbf{y}_0^T (\mathbf{F}_0 + \eta \mathbf{N})\mathbf{y}_0 - 2r\mathbf{y}_0^T (\mathbf{F}_0 + \eta \mathbf{N})\mathbf{v} - r^2 \mathbf{v}^T (\mathbf{F}_0 + \eta \mathbf{N})\mathbf{v} \geq 0, \tag{34}$$

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

Lemma 5 ($\mathcal{S}$-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 $\mathcal{S}$-procedure implies that (34) 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

$$\mathbf{M} = \begin{bmatrix} \nu - \tau - \mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}_0 & -r\mathbf{y}_0^T(\mathbf{F}_0 + \eta\mathbf{N}) \\ -r(\mathbf{F}_0 + \eta\mathbf{N})\mathbf{y}_0 & \tau\mathbf{G} - r^2(\mathbf{F}_0 + \eta\mathbf{N}) \end{bmatrix} \succeq 0. \tag{35}$$

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 (33) 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,

$$\begin{aligned} \nu &\geq \tau + \mathbf{1}^T\mathbf{t}, \\ r^2 &= \sigma\tau, \\ w_i^2 &= (1 - \sigma\lambda_i)t_i, \quad i = 1, \ldots, m, \\ \sigma &\leq \tfrac{1}{\lambda_{\max}(\mathbf{H})}. \end{aligned} \tag{36}$$

It is easy to establish that there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy (36) if and only if there exist $\tau, \sigma \geq 0$, and $\mathbf{t} \in \mathbf{R}_+^m$ that satisfy (36) with the equalities replaced by inequalities.

Note that the constraint $r^2 \leq \sigma\tau$ and $\tau \geq 0$ imply that $\sigma \geq 0$. Therefore, replacing the equalities in (36) by inequalities and reformulating them as SOC constraints, we have $\mathbf{x}^T\mathbf{V}^T(\mathbf{F}_0+\eta\mathbf{N})\mathbf{V}\mathbf{x} \le \nu$ for all $\mathbf{V} \in \mathcal{S}_v$ if and only if the following system of linear and second-order cone constraints holds,

$$\begin{aligned} \tau &\ge 0, \\ \nu &\ge \tau + \mathbf{1}^T\mathbf{t} \\ \sigma &\le \frac{1}{\lambda_{\max}(\mathbf{H})}, \\ r &\ge \textstyle\sum_{i=1}^n \rho_i\,|x_i|\,, \\ \left\|\begin{bmatrix} 2r \\ \sigma-\tau \end{bmatrix}\right\| &\le \sigma+\tau, \\ \left\|\begin{bmatrix} 2w_i \\ (1-\sigma\lambda_i-t_i) \end{bmatrix}\right\| &\le (1-\sigma\lambda_i+t_i),\ i=1,\dots,m. \end{aligned} \tag{37}$$

The constraint $r \ge \sum_{i=1}^n \rho_i\,|x_i|$ is not linear but it can be linearized by introducing a new variable $\mathbf{u}$ such that $\mathbf{u} \ge |\mathbf{x}|$, i.e. $u_j \ge x_j$ and $u_j \ge -x_j$, $i = j, \dots, n$. The result now follows by replacing $\max_{\{\mathbf{V}\in\mathcal{S}_v\}}\left\{\mathbf{x}^T\mathbf{V}^T(\mathbf{F}_0+\eta\mathbf{N})\mathbf{V}\mathbf{x}\right\}$ in (28) 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 $\mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succ \mathbf{0}$, with $\|\mathbf{F}_0^{\frac{1}{2}}\boldsymbol{\Delta}\mathbf{F}_0^{\frac{1}{2}}\| \le \eta$, $\eta < 1$. For details of these alternative formulations and their relation to probabilistic guarantees on the performance of the optimal solution see [13].

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 Section 2.

3.1. Robust mean-variance portfolio selection

Suppose the returns on n assets in a discrete time market are 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 loading matrix and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$ is the residual returns vector. Here $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes that $\mathbf{x}$ is a multivariate Normal random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. 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 $\mathbf{F} \succeq \mathbf{0}$ and the covariance matrix $\mathbf{D} = \mathbf{diag}(\mathbf{d}) \succeq \mathbf{0}$, i.e. $d_i \ge 0$, $i = 1, \dots, n$. Thus, the vector of asset returns $\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V}^T\mathbf{F}\mathbf{V} + \mathbf{D})$.

An investor's position in this 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, $j = 1, \dots, n$. 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 [19, 20] 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 have a return of at least α, i.e. $\boldsymbol{\phi}^*$ is the optimal solution of the convex quadratic optimization problem

$$\begin{array}{rl} \text{minimize} & \mathbf{Var}\big[\mathbf{r}_\phi\big] \\ \text{subject to} & \mathbf{E}[r_\phi] \geq \alpha, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1. \end{array} \tag{38}$$

The optimization problem (38) is called the minimum variance portfolio selection problem. Note that the Markowitz model implicitly assumes that the mean return vector $\mathbf{E}[\mathbf{r}]$ and the covariance matrix $\mathbf{Var}\big[\mathbf{r}\big]$ are known with certainty. Other variants include the maximum return problem and the maximum Sharpe ratio problem.

This mean-variance model has had a profound impact on the economic modeling of financial markets and the pricing of assets, earning Markowitz and Sharpe the 1990 Nobel Prize in Economics for their work. In spite of this, practitioners have shied away from this model, the primary reason being that the optimal portfolio $\boldsymbol{\phi}^*$ is extremely sensitive to the market parameters $(\mathbf{E}[\mathbf{r}], \mathbf{Var}\big[\mathbf{r}\big])$. Since these parameters are estimated from noisy data, $\boldsymbol{\phi}^*$ often amplifies noise.

One approach to mitigate the sensitivity of $\boldsymbol{\phi}^*$ to the errors and uncertainty in the problem data is to consider a robust version of (38). 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

$$S_f = \Big\{\mathbf{F} : \mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \boldsymbol{\Delta} \succeq \mathbf{0},\, \boldsymbol{\Delta} = \boldsymbol{\Delta}^T,\, \|\mathbf{F}_0^{\frac{1}{2}}\boldsymbol{\Delta}\mathbf{F}_0^{\frac{1}{2}}\| \leq \eta\Big\}; \tag{39}$$

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

$$S_d = \Big\{\mathbf{D} : \mathbf{D} = \mathbf{diag}(\mathbf{d}),\, d_i \in [\underline{d}_i, \bar{d}_i],\, i = 1, \ldots, n\Big\}; \tag{40}$$

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

$$S_v = \big\{\mathbf{V} : \mathbf{V} = \mathbf{V}_0 + \mathbf{W},\, \|\mathbf{W}_i\|_g \leq \rho_i,\, i = 1, \ldots, n\big\}, \tag{41}$$

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

$$S_m = \big\{\boldsymbol{\mu} : \boldsymbol{\mu} = \boldsymbol{\mu}_0 + \boldsymbol{\xi},\, |\xi_i| \leq \gamma_i,\, i = 1, \ldots, n\big\}. \tag{42}$$

The uncertainty sets (S_f, S_v, S_d, S_m) capture 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 $\mathbf{G}$, ρ_i, γ_i, $\bar{d}_i$, $\underline{d}_i$, $i = 1, \ldots, n$, and η are discussed in [13].

The robust analog of (38) is given by

$$\begin{array}{ll} \text{minimize} & \max_{\{\mathbf{V}\in S_v,\mathbf{D}\in S_d,\mathbf{F}\in S_f\}} \mathbf{Var}\big[\mathbf{r}_{\phi}\big] \\ \text{subject to} & \min_{\{\boldsymbol{\mu}\in S_m\}} \mathbf{E}[r_{\phi}] \geq \alpha, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1. \end{array} \tag{43}$$

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 the non-robust problem (38).

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 (43) as

$$\begin{array}{ll} \text{minimize} & \max_{\{\mathbf{V}\in S_v,\mathbf{F}\in S_f\}} \big\{\boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi}\big\} + \max_{\{\mathbf{D}\in S_d\}} \big\{\boldsymbol{\phi}^T\mathbf{D}\boldsymbol{\phi}\big\} \\ \text{subject to} & \min_{\{\boldsymbol{\mu}\in S_m\}} \boldsymbol{\mu}^T\boldsymbol{\phi} \geq \alpha, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1, \end{array} \tag{44}$$

which in turn is equivalent to the following robust QCP,

$$\begin{array}{ll} \text{minimize} & \lambda + \delta, \\ \text{subject to} & \boldsymbol{\phi}^T\mathbf{V}^T\mathbf{F}\mathbf{V}\boldsymbol{\phi} \leq \lambda, \ \forall \mathbf{V} \in S_v, \mathbf{F} \in S_f \\ & \boldsymbol{\phi}^T\mathbf{D}\boldsymbol{\phi} \leq \delta, \ \forall \mathbf{D} \in S_d, \\ & \boldsymbol{\mu}^T\boldsymbol{\phi} \geq \alpha, \ \forall \boldsymbol{\mu} \in S_m, \\ & \mathbf{1}^T\boldsymbol{\phi} = 1. \end{array} \tag{45}$$

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 (25), (45) can be reduced to an SOCP. For details on robust portfolio selection problems and the performance on real market data see [13].

3.2. Robust hyperplane separation

Let $\mathcal{L} = \big\{(\mathbf{x}_i, y_i), i = 1, \ldots, l\big\}$, $y_i \in \{+1, -1\}$, $\mathbf{x}_i \in \mathbf{R}^d$, $\forall i$, 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" $\mathbf{x}_i$, i.e. $\mathbf{x}_i$ with $y_i = -1$, from the "positive" $\mathbf{x}_i$, i.e. $\mathbf{x}_i$ with $y_i = +1$. Given such a 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".

In a typical application of linear discrimination, the hyperplane $(\mathbf{w}, b)$ is chosen by solving the quadratic program [8, 18, 27]:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\mathbf{w}\|^2 + C\mathbf{1}^T\boldsymbol{\xi}, \\ \text{subject to} & y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \ i = 1, \ldots, l, \\ & \boldsymbol{\xi} \geq \mathbf{0}, \end{array} \tag{46}$$

where C is the penalty for misclassification; or equivalently, the dual program,

$$\begin{array}{ll} \text{maximize} & \mathbf{1}^T\boldsymbol{\alpha} - \frac{1}{2}\big\| \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i \big\|^2, \\ \text{subject to} & \mathbf{y}^T\boldsymbol{\alpha} = 0, \\ & \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{1}. \end{array} \tag{47}$$

The optimal vector $\mathbf{w}^* = \sum_{i=1}^{l} \alpha_i^* \mathbf{x}_i$, where $\boldsymbol{\alpha}^*$ is the optimal solution of (47). The optimal intercept b^* is set by the complementary slack conditions (for a detailed discussion see [8]). The complexity of classifying a new point $\mathbf{x}$ is given by $n\ |\boldsymbol{\alpha}^*|$, where $|\boldsymbol{\alpha}^*|$ denotes the number of nonzero terms in the vector $\boldsymbol{\alpha}^*$. Thus, one would like the optimal $\boldsymbol{\alpha}^*$ to be sparse.

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

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

where $\bar{\mathbf{x}}_i$ is the true value of the training data and $\mathbf{u}_i$, with $\|\mathbf{u}_i\| \leq \rho_i$, is the measurement noise. If we assume that for the i-th data point all points in the ball $\mathcal{B}_i = \{\mathbf{z} : \|\mathbf{z} - \mathbf{x}_i\| \leq \rho_i\}$ are equally likely, we can increase the "margin of the classifier" (see [8]) by replacing the dual objective function $f(\boldsymbol{\alpha})$ in (47) by

$$\begin{aligned} f(\boldsymbol{\alpha}) &= \max_{\|\mathbf{u}_i\| \leq \rho_i} \left\{ \mathbf{1}^T \boldsymbol{\alpha} - \tfrac{1}{2} \boldsymbol{\alpha}^T (\mathbf{V}_0 + \mathbf{U})^T (\mathbf{V}_0 + \mathbf{U}) \boldsymbol{\alpha} \right\}, \\ &= \mathbf{1}^T \boldsymbol{\alpha} - \tfrac{1}{2} \min_{\|\mathbf{u}_i\| \leq \rho_i} \left\{ \boldsymbol{\alpha}^T (\mathbf{V}_0 + \mathbf{U})^T (\mathbf{V}_0 + \mathbf{U}) \boldsymbol{\alpha} \right\}, \end{aligned}$$

where $\mathbf{V}_0 = [y_1 \mathbf{x}_1, y_2 \mathbf{x}_2, \ldots, y_l \mathbf{x}_l]$ and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_l]$. The optimal $\boldsymbol{\alpha}^*$ is now given by the solution of the robust QCP,

$$\begin{array}{ll} \text{maximize} & \tau, \\ \text{subject to} & \mathbf{1}^T \boldsymbol{\alpha} - \tfrac{1}{2} \boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} \geq \tau, \quad \forall \mathbf{Q} \in \mathcal{S}, \\ & \mathbf{y}^T \boldsymbol{\alpha} = 0, \\ & \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{1}, \end{array} \tag{48}$$

where the uncertainty set

$$\mathcal{S} = \left\{ \mathbf{Q} : \mathbf{Q} = \mathbf{V}^T \mathbf{V}, \mathbf{V} = \mathbf{V}_0 + \mathbf{U}, \|\mathbf{U}_i\| \leq \rho, \mathbf{V}_0 = [\mathbf{x}_1, \ldots, \mathbf{x}_l]\, \mathbf{diag}(\mathbf{y}) \right\} \tag{49}$$

belongs to class of factorized uncertainty structures defined in (25). Thus, (48) can be reformulated as an SOCP. This technique can be extended to general support vector machines [27] as well.

In the above discussion we assumed that there is a confidence ball around each sample and one is indifferent to shifts within this ball. Next, we exploited this property to choose a hyperplane that maximally separates the points in the most optimistic scenario.

Suppose on the other hand one is interested in choosing a hyperplane that minimizes the misclassification in the worst case, i.e. one that minimizes the maximum misclassification when the samples are allowed to move within their corresponding confidence balls. In this case, the primal problem is given by the following robust optimization problem

$$\begin{array}{lll} \text{minimize} & \tfrac{1}{2}\|\mathbf{w}\|^2 + C\mathbf{1}^T \boldsymbol{\xi}, & \\ \text{subject to} & y_i(\mathbf{w}^T(\mathbf{x}_i + \rho_i \mathbf{u}_i) + b) \geq 1 - \xi_i,\ \|\mathbf{u}_i\| \leq 1, & i = 1, \ldots, l, \\ & \xi_i \geq 0, & i = 1, \ldots, l, \end{array}$$

or equivalently

$$\begin{aligned} \text{minimize } & \tfrac{1}{2}t^2 + C\mathbf{1}^T\boldsymbol{\xi}, \\ \text{subject to } & y_i(\mathbf{w}^T\mathbf{x}_i + b) - \rho_i t \geq 1 - \xi_i,\ i = 1, \ldots, l, \\ & \boldsymbol{\xi} \geq \mathbf{0}, \\ & \|\mathbf{w}\| \leq t. \end{aligned} \tag{50}$$

Problem (50) can be reformulated as an SOCP by introducing new scalar variables u and v, replacing t^2 in the objective by $u - v$ and requiring that u and v satisfy the linear and SOC constraints $u + v = 1$ and $\sqrt{t^2 + v^2} \leq u$, since the latter imply that $t^2 \leq u - v$.

The dual of the reformulated problem is the SOCP:

$$\begin{aligned} \text{maximize } & \beta + \mathbf{1}^T\boldsymbol{\alpha}, \\ \text{subject to } & \gamma \leq \rho^T\alpha - \|\textstyle\sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i\|, \\ & \beta + z_u = \tfrac{1}{2}, \\ & \beta + z_v = -\tfrac{1}{2}, \\ & \mathbf{y}^T\boldsymbol{\alpha} = 0, \\ & \mathbf{0} \leq \alpha \leq C\mathbf{1}, \\ & \sqrt{\gamma^2 + z_v^2} \leq z_u, \end{aligned} \tag{51}$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_l)^T$. It is easy to see from the above that $z_u - z_v = 1$ and $\beta = -\frac{1}{2}(z_u + z_v) = \frac{1}{2}(z_u^2 - z_v^2)$. Since at an optimal solution of the primal SOCP $t^2 = u - v = u^2 - v^2$, it follows from complementary slackness for SOCPs that $\gamma^2 = z_u^2 - z_v^2$. Hence, we can replace β in the objective of (51) by γ^2 to obtain the equivalent problem:

$$\begin{aligned} \text{maximize } & \mathbf{1}^T\boldsymbol{\alpha} - \tfrac{1}{2}\gamma^2, \\ \text{subject to } & \gamma \leq \boldsymbol{\alpha}^T\boldsymbol{\rho} - \|\textstyle\sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i\|, \\ & \mathbf{y}^T\boldsymbol{\alpha} = 0, \\ & \mathbf{0} \leq \alpha \leq C\mathbf{1}, \end{aligned} \tag{52}$$

The similarity of (52) and (47) is apparent – all that the uncertainty does is to reduce the norm in the objective. This analysis can be extended to general support vector machines by replacing $\| \sum_{i=1}^{l} \alpha_i^0 y_i \mathbf{x}_i \|$ with $\sqrt{\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha}}$, where $\mathbf{K}$ is the kernel matrix.

Note that in the worst case analysis discussed above we reformulated a robust quadratic program as an SOCP without using any of the general results developed in Section 2.

3.3. Linear least squares problem with deterministic and stochastic uncertainty

Consider the following linear least squares problem,

$$\min_{\mathbf{x} \in \mathbf{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \tag{53}$$

where $\mathbf{A} = [\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T]^T \in \mathbf{R}^{m \times n}$ and $\mathbf{b} \in \mathbf{R}^m$. If $m \geq n$ and the matrix $\mathbf{A}$ has full column rank, the solution of this optimization problem is given by $\mathbf{x}^* = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$ [14]. 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 (53) 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 errors. 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 (53)

$$\min_{\mathbf{x}} \max_{\{[\mathbf{A},\mathbf{b}]:\|[\mathbf{A},\mathbf{b}]-[\mathbf{A}_0,\mathbf{b}_0]\|\le\rho\}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2, \tag{54}$$

where $\|\cdot\|$ is the Frobenius norm, and showed that (54) can be reformulated as an SOCP [11]. However this uncertainty set does not appear natural, since it applies to $[\mathbf{A}\ \mathbf{b}]$ all at once.

We propose the following uncertainty structure for the rows $\mathbf{a}_i \in \mathbf{R}^n$, $i = 1, \dots, m$, of $\mathbf{A}$:

$$\mathcal{S} = \left\{\mathbf{a} : \mathbf{a} = \mathbf{a}^0 + \sum_{j=1}^{k} v^j \mathbf{a}^j + \sum_{j=1}^{l} u^j \boldsymbol{\xi}^j\right\}, \tag{55}$$

where, without loss of generality, $\|\mathbf{v}\| \le 1$, $\|\mathbf{u}\| \le 1$, and $\boldsymbol{\xi}^j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}^j)$, $j = 1, \dots, l$. Without the stochastic term, the uncertainty set (55) has the affine structure considered in [4–6]. The term $\sum_{j=1}^{l} u^j \boldsymbol{\xi}^j$ models the imperfect knowledge of the stochastic perturbations in $\mathbf{a}$ – the total variance and modes $\boldsymbol{\Omega}^j$ are known but the variance of each of the individual modes is unknown. In typical applications, the matrix $\boldsymbol{\Omega}^j = \boldsymbol{\omega}^j(\boldsymbol{\omega}^j)^T$, or equivalently $\boldsymbol{\xi}^j = (\boldsymbol{\omega}^j)^T\mathbf{Z}$ for $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where the vector $\boldsymbol{\omega}^j$ is determined by the estimation algorithm or the signal network.

The robust least squares problem corresponding to (55) is given by

$$\min_{\mathbf{x}} \left\{\sum_{i=1}^{m} \max_{\mathbf{a}_i \in \mathcal{S}_i} \left\{\mathbf{E}[(\mathbf{a}_i^T\mathbf{x} - b_i)^2]\right\}\right\}, \tag{56}$$

where each $\mathcal{S}_i$ is of the form (55) for appropriately chosen $\{\mathbf{a}_i^j : j = 0, \dots, k_i\}$ and $\{\boldsymbol{\Omega}_i^j : j = 1, \dots, l_i\}$, $i = 1, \dots, m$.

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

$$\mathbf{E}[(\mathbf{a}^T\mathbf{x} - b)^2] = \left((\mathbf{a}^0)^T\mathbf{x} + \sum_{j=1}^{k} v^j(\mathbf{a}^j)^T\mathbf{x} - b\right)^2 + \sum_{j=1}^{l}(u^j)^2\mathbf{x}^T\boldsymbol{\Omega}^j\mathbf{x}. \tag{57}$$

Therefore, $\mathbf{E}[(\mathbf{a}^T\mathbf{x} - b)^2] \le \delta$, for all $\mathbf{a} \in \mathcal{S}$, if, and only if, there exists $t \ge 0$ such that

$$\begin{aligned}&|\ \mathbf{a}^T\mathbf{x} - b\ | \le t,\ \ \forall \mathbf{a} \in \mathcal{S}_1 = \{\mathbf{a} : \mathbf{a}^0 + \textstyle\sum_{j=1}^{k} v^j\mathbf{a}^j,\ \|\mathbf{v}\| \le 1\},\\ &t^2 + \mathbf{x}^T\mathbf{Q}\mathbf{x} \le \delta,\ \forall \mathbf{Q} \in \mathcal{S}_2 = \{\mathbf{Q} : \mathbf{Q} = \textstyle\sum_{j=1}^{l} \alpha^j\boldsymbol{\Omega}^j,\ \textstyle\sum_{j=1}^{l}\alpha^j \le 1, \alpha^j \ge 0, \forall j\}.\end{aligned} \tag{58}$$

From (57) and (58), it follows that (56) is equivalent to

$$\begin{array}{lll}\text{minimize} & \sum_{i=1}^{m} \delta_i, & \\ \text{subject to} & t_i^2 + \mathbf{x}^T\mathbf{Q}_i\mathbf{x} \le \delta_i, & \mathbf{Q}_i \in \mathcal{S}_2^i,\ i = 1, \dots, m,\\ & \mathbf{a}_i^T\mathbf{x} - b_i \le t_i, & \mathbf{a}_i \in \mathcal{S}_1^i,\ i = 1, \dots, m,\\ & \mathbf{a}_i^T\mathbf{x} - b_i \ge -t_i, & \mathbf{a}_i \in \mathcal{S}_1^i,\ i = 1, \dots, m.\end{array} \tag{59}$$

The sets $\mathcal{S}_1^i$, $i = 1, \dots, m$, are a special case of the norm-constrained uncertainty set defined in (18) and $\mathcal{S}_2^i$, $i = 1, \dots, m$, are polytopic uncertainty sets of the form in (8). Thus, (59) is equivalent to an SOCP.

3.4. Equalizing uncertain channels

By sampling the input and output signals, a linear time-invariant communication channel can be described as follows (see [24] for details):

$$\begin{aligned} y_k &= \sum_{i=0}^{m-1} h_i x_{k-i} + s_k, \\ &= h_0 x_k + \sum_{i=1}^{m-1} h_i x_{k-i} + s_k, \end{aligned} \tag{60}$$

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, $\mathbf{h} = \{h_i : i = 0, \dots, m-1\}$ is the impulse response of the channel, and $\{s_k : k \geq 0\}$ are the samples of the channel noise. The channel impulse response is assumed to be finite, i.e. $m < \infty$. From (60), it is clear that in order to recover the sample x_k at time k the effects of the noise s_k and the interference $\sum_{i=1}^{m-1} h_i x_{k-i}$ from the past samples must be removed.

Since the input samples are decoded in a sequential manner, if the channel response $\mathbf{h}$ is known, interference does not cause a problem – one simply subtracts the effect of the past samples. Thus, one technique for removing interference is to "filter" the output, i.e. convolve the output $\{y_k : k \geq 0\}$ with a known vector $\mathbf{g} = \{g_i : i = 0, \dots, n-1\}$ such that the effective channel response is equal to a known vector $\mathbf{d} = \{d_k : k = 0, \dots, l-1\}$. This is called *channel shaping*.

Channel shaping using finite impulse response filters is possible only if the output is sampled at a faster rate than the input [23]. Sampling the output at a rate p times faster than the input results in p parallel channels that all see the same input samples. Let $\mathbf{h}_j = \{h_i^j : i = 0, \dots, m-1\}$, $j = 1, \dots, p$, denote the responses of the p parallel channels. Then, the samples y_k^j, $k \geq 0$, of the j-th channel are given by

$$y_k^j = \sum_{i=0}^{m-1} h_i^j x_{k-i} + s_k^j. \tag{61}$$

Suppose the outputs from the j-th channel are convolved with $\mathbf{g}_j = \{g_i^j, i = 0, \dots, n-1\}$, $n < \infty$, and the resulting signals added together. Then, the effective input-output relation is given by

$$y_k = \sum_{i=0}^{m+n-2} w_i x_{k-i} + \tilde{s}_k, \tag{62}$$

where $w_i = \sum_{j=1}^{p} \big(\sum_{r=0}^{n-1} h_i^j g_{i-r}^j \big)$, $i = 0, \dots, m+n-2$, and $\tilde{s}_k = \sum_{j=1}^{p} \big(\sum_{r=0}^{n-1} g_l^j s_{k-r}^j \big)$, $k \geq 0$.

Therefore, the effective channel is equal to $\mathbf{d}$ only if $\mathbf{g}_j$, $j = 1, \ldots, p$ satisfy

$$\sum_{j=1}^{p} \mathbf{T}_{\mathbf{h}_j} \mathbf{g}_j = \mathbf{d}, \tag{63}$$

where $\mathbf{T}_{\mathbf{h}_j}$ is the following Toeplitz matrix formed from $\mathbf{h}_j$,

$$\mathbf{T}_{\mathbf{h}_j} = \underbrace{\begin{bmatrix} h_{j0} & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_{j,m-1} & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & h_{j0} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & h_{j,m-1} \end{bmatrix}}_{(n+m-2)\times n}.$$

For $p \geq 2$ the system of equations (63) has a solution under fairly general conditions [23, 22].

In the development so far it has been implicitly assumed that the responses $\mathbf{h}_j$, $j = 1, \ldots, p$, are known. In practice, the responses $\mathbf{h}_j$ are estimated by transmitting a known finite length training sequence. Consequently, the estimates are subject to statistical errors. These errors in the estimates can be modeled as follows:

$$\bar{\mathbf{h}}_j = \mathbf{h}_j + u_j \boldsymbol{\xi}_j, \quad j = 1, \ldots, p \tag{64}$$

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, $\boldsymbol{\xi}_j \sim \mathcal{N}(0, \boldsymbol{\Omega}_j)$, $\mathbf{Tr}(\boldsymbol{\Omega}_j) = 1$, $j = 1, \ldots, p$, $\mathbf{E}[\boldsymbol{\xi}_j \boldsymbol{\xi}_k^T] = \mathbf{0}$, $j \neq k$, and $\|\mathbf{u}\| \leq \sigma^2$. The uncertainty structure (64) 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.

In a robust approach, $\{\mathbf{g}_j, j = 1, \ldots, p\}$ are chosen by solving the problem:

$$\min_{\{\mathbf{g}_j : j=1,\ldots,p\}} \max_{\{\mathbf{u}: \|\mathbf{u}\| \leq \sigma^2\}} \mathbf{E}\left[\left\| \sum_{j=1}^{p} \mathbf{T}_{\bar{\mathbf{h}}_j} \mathbf{g}_j - \mathbf{d} \right\|^2 \right] \tag{65}$$

After substituting for $\bar{\mathbf{h}}_j$, $j = 1, \ldots, p$, from (64), we have that

$$\begin{aligned} &\max_{\{\mathbf{u}: \|\mathbf{u}\| \leq \sigma^2\}} \mathbf{E}\left[\left\| \sum_{j=1}^{p} \mathbf{T}_{\mathbf{h}_j} \mathbf{g}_j - \mathbf{d} + \sum_{j=1}^{p} u_j \mathbf{T}_{\boldsymbol{\xi}_j} \mathbf{g}_j \right\|^2 \right] \\ &= \max_{\{\mathbf{u}: \|\mathbf{u}\| \leq \sigma^2\}} \mathbf{E}\left[\left\| \sum_{j=1}^{p} \mathbf{T}_{\mathbf{h}_j} \mathbf{g}_j - \mathbf{d} + \sum_{j=1}^{p} u_j \mathbf{T}_{\mathbf{g}_j} \boldsymbol{\xi}_j \right\|^2 \right], \end{aligned} \tag{66}$$

$$= \Big\| \sum_{j=1}^{p} \mathbf{T}_{\mathbf{h}_j}\mathbf{g}_j - \mathbf{d} \Big\|^2 + \max_{\{\mathbf{u}:\|\mathbf{u}\|\leq\sigma^2\}} \Big\{ \sum_{j=1}^{p} u_j^2\, \mathbf{Tr}\big(\mathbf{T}_{\mathbf{g}_j}^T \mathbf{T}_{\mathbf{g}_j} \Omega_j\big) \Big\},$$

$$= \Big\| \sum_{j=1}^{p} \mathbf{T}_{\mathbf{h}_j}\mathbf{g}_j - \mathbf{d} \Big\|^2 + \max_{\{\mathbf{u}:\|\mathbf{u}\|\leq\sigma^2\}} \Big\{ \sum_{j=1}^{p} \mathbf{g}_j (u_j^2 \mathbf{\Lambda}_j) \mathbf{g}_j \Big\} \tag{67}$$

where (66) follows from the fact that $\mathbf{T}_{\boldsymbol{\xi}_j}\mathbf{g}_j = \mathbf{T}_{\mathbf{g}_j}\boldsymbol{\xi}_j$ and $\mathbf{\Lambda}_j$ in (67) is set by the identity $\mathbf{g}_j\mathbf{\Lambda}_j\mathbf{g}_j = \mathbf{Tr}\big(\mathbf{T}_{\mathbf{g}_j}^T\mathbf{T}_{\mathbf{g}_j}\Omega_j\big)$. Thus, (65) is equivalent to the robust convex QCP

$$\begin{array}{ll} \text{minimize } \delta + \nu, & \\ \text{subject to } \big\| \mathbf{T}\mathbf{g} - \mathbf{d} \big\|^2 \leq \delta, & \\ \qquad \mathbf{g}^T\mathbf{Q}\mathbf{g} \leq \nu, & \forall \mathbf{Q} \in \mathcal{S}, \end{array} \tag{68}$$

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

$$\mathcal{S} = \Big\{ \mathbf{Q} : \mathbf{Q} = \mathbf{diag}(\mathbf{Q}_1, \ldots, \mathbf{Q}_p), \mathbf{Q}_j = \alpha_j \mathbf{\Lambda}_j, \sum_{j=1}^{p} \alpha_j \leq 1, \alpha_j \geq 0, j = 1, \ldots, p \Big\}, \tag{69}$$

belongs to the class of polytopic uncertainty sets described in (8). Consequently, (68) 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 $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ with an unknown mean $\boldsymbol{\mu}$ and covariance

$$\mathbf{\Sigma} \in \mathcal{S}_1 = \Big\{ \mathbf{\Sigma} : \mathbf{\Sigma}^{-1} = \mathbf{\Sigma}_0^{-1} + \mathbf{\Delta} \succeq \mathbf{0}, \mathbf{\Delta} = \mathbf{\Delta}^T, \big\| \mathbf{\Sigma}_0^{\frac{1}{2}} \mathbf{\Delta} \mathbf{\Sigma}_0^{\frac{1}{2}} \big\| \leq \eta \Big\}. \tag{70}$$

We will assume that $\eta < 1$. The set (70) is precisely the confidence region associated with the maximum likelihood estimate of the covariance $\mathbf{\Sigma}$ of $\mathbf{x}$. See [13] for details.

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

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{d}, \tag{71}$$

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

$$\mathbf{D} \in \mathcal{S}_2 = \left\{ \mathbf{D} : \begin{array}{l} \mathbf{D} = \mathbf{V}^T\mathbf{F}\mathbf{V}, \mathbf{F} = \mathbf{F}_0 + \mathbf{\Delta} \succeq \mathbf{0}, \|\mathbf{N}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{N}^{-\frac{1}{2}}\| \leq \eta, \\ \mathbf{V} = \mathbf{V}_0 + \mathbf{W}, \|\mathbf{W}_i\| \leq \rho_i, i = 1, \ldots, m, \|\mathbf{\Delta}\| \leq \eta \end{array} \right\}. \tag{72}$$

The uncertainty set (72) 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 observations $\mathbf{y}$ and an *a priori* unbiased estimate $\bar{\boldsymbol{\mu}}$ of $\boldsymbol{\mu}$, we consider a linear unbiased estimator of the form

$$\hat{\boldsymbol{\mu}} = (\mathbf{I} - \mathbf{KC})\bar{\boldsymbol{\mu}} + \mathbf{Ky}, \tag{73}$$

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

$$\mathbf{P} \equiv \mathbf{E}[(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T] = (\mathbf{I} - \mathbf{KC})^T \boldsymbol{\Sigma}(\mathbf{I} - \mathbf{KC}) + \mathbf{K}^T \mathbf{DK}. \tag{74}$$

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 [15]. The robust measurement model developed here is a variant of the model proposed by Calafiore and El Ghaoui [9], in which the *a priori* covariance $\boldsymbol{\Sigma}$ is assumed to be known exactly and the noise covariance

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

In [9] it is shown 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

$$\min_{\mathbf{K}} \max_{\{\boldsymbol{\Sigma} \in \mathcal{S}_1, \mathbf{D} \in \mathcal{S}_2\}} \max_{\{1 \leq j \leq k\}} \left\{\mathbf{v}_j^T (\mathbf{I} - \mathbf{KC})^T \boldsymbol{\Sigma}(\mathbf{I} - \mathbf{KC})\mathbf{v}_j + \mathbf{v}_j^T \mathbf{K}^T \mathbf{DK}\mathbf{v}_j\right\}, \tag{75}$$

or equivalently, the robust QCP,

$$\begin{array}{lll} \text{minimize } \nu, & & \\ \text{subject to } \mathbf{v}_j^T (\mathbf{I} - \mathbf{KC})^T \boldsymbol{\Sigma}(\mathbf{I} - \mathbf{KC})\mathbf{v}_j \leq \delta_j, & \forall\, \boldsymbol{\Sigma} \in \mathcal{S}_1, & j = 1, \ldots, k, \\ \qquad\qquad \mathbf{v}_j^T \mathbf{K}^T \mathbf{DK}\mathbf{v}_j \leq \nu - \delta_j, & \forall\, \mathbf{D} \in \mathcal{S}_2, & j = 1, \ldots, k. \end{array} \tag{76}$$

For fixed $\mathbf{K}$, Lemma 3 in [13] implies that

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

Thus, $\eta < 1$ implies that the first constraint in (76) 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 $\mathcal{S}_2$ belongs to the class of factorized uncertainty sets defined in (25), Lemma 4 implies that that the robust quadratic constraint $\mathbf{v}_j^T \mathbf{K}^T \mathbf{DK}\mathbf{v}_j = \mathbf{y}_j^T \mathbf{D}\mathbf{y}_j \leq \nu - \delta_j$, for all $\mathbf{D} \in \mathcal{S}_2$, can be reformulated as a collection of linear and SOC constraints. Hence, (76) can be transformed into an SOCP.

4. Conclusion

Ben-Tal and Nemirovski initiated the study of robust convex QCPs and showed that for generalized ellipsoidal uncertainty sets these problems can be reformulated as SDPs [4] (see also [6]). In Section 2 we described three general classes of uncertainty sets that enable robust QCPs to be reformulated as SOCPs.

Adding robustness reduces the sensitivity of the optimal decision to perturbations in a model's parameters, often resulting in significant improvement in performance [3, 13, 26]. Typically, the complexity of the deterministic reformulation of the robust counterpart of a problem is higher than it is for the original problem. However, since the worst case complexity of SOCPs is comparable to that of convex QCPs, our results show that one can add robustness to convex QCPs with a relatively modest increase in computational effort. Moreover, the examples presented in Section 3 show that the natural uncertainty sets for optimization problems arising in a wide variety of application areas belong to the classes introduced in Section 2.

An important issue in robust optimization is how to choose 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 (25), the parametrization is not obvious; in [13] it is shown that the factorized uncertainty set is parametrized by the confidence regions corresponding to the statistical techniques used to estimate the parameters of the original non-robust problem.

References

[1] Alizadeh, F.: Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. Optim. **5**(1), 13–51 (1995)
[2] Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Prog. **95**(1), 3–51 (2003)
[3] Ben-Tal, A., Nemirovski, A.: Robust truss topology design via semidefinite programming. SIAM J. Optim. **7**(4), 991–1016 (1997)
[4] Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**(4), 769–805 (1998)
[5] Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. Oper. Res. Lett. **25**(1), 1–13 (1999)
[6] Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization. SIAM, Philadelphia, PA, 2001
[7] Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: Linear matrix inequalities in system and control theory. SIAM, Philadelphia, PA, 1994
[8] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. **2**, 121–167 (1998)
[9] Calafiore, G., El Ghaoui, L.: Minimum variance estimation with uncertain statistical model. In: Proc. of 40th Conf. Dec. and Cont., 2001
[10] Chopra, V.K., Ziemba, W.T.: The effect of errors in means, variances and covariances on optimal portfolio choice. J. Portfolio Manag., Winter, 1993, pp. 6–11
[11] El Ghaoui, L., Lebret, H.: Robust solutions to least-squares problems with uncertain data. SIAM J. Matrix Anal. Appl. **18**(4), 1035–1064 (1997)
[12] El Ghaoui, L., Oustry, F., Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM J. Optim. **9**(1), 33–52 (1998)
[13] Goldfarb, D., Iyengar, G.: Robust portfolio selection problems. Math. Oper. Res. **28**(1), (2003)
[14] Golub, G.H., Van Loan, C.: Matrix Computations. John Hopkins Press, 3rd edition, 1996
[15] Huber, P.J.: Robust Statistics. John Wiley & Sons, 1981
[16] Lobo, M.S., Fazel, M., Boyd, S.: Portfolio optimization with linear and fixed transaction costs and bounds on risk. Submitted to Operations Research, June 2000
[17] Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. Linear Algebra Appl. **284**(1-3), 193–228 (1998)
[18] Mangasarian, O.L., Musicant, D.R.: Robust linear and support vector regression. IEEE Trans. PAMI, **22**, 9 (2000)
[19] Markowitz, H.M.: Portfolio selection. J. Finance **7**, 77–91 (1952)
[20] Markowitz, H.M.: Portfolio Selection. Wiley, New York, 1959
[21] Nesterov, Y., Nemirovskii, A.: Interior-point polynomial algorithms in convex programming. SIAM, Philadelphia, 1993

[22] Pal, D., Iyengar, G.N., Cioffi, J.M.: A new method for channel shortening with applications to discrete multi-tone – part i : Theory. Submitted to IEEE Transactions on Communications
[23] Pal, D., Iyengar, G.N., 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, **2**, 763–768 (1998)
[24] Proakis, J.G.: Digital Communication. McGraw-Hill, 4th edition, 2000
[25] Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Reveiw **38**, 49–95 (1996)
[26] Vandenberghe, L., Boyd, S., Nouralishahi, M.: Robust linear programming and optimal control. In: Proceedings of the 15th IFAC World Congress on Automatic Control, 2002
[27] Vapnik, V.: Statistical learning theory. John Wiley and Sons, Inc., New York, 1998