OPERATIONS RESEARCH
Vol. 57, No. 6, November–December 2009, pp. 1483–1495
ISSN 0030-364X | EISSN 1526-5463 | 09 | 5706 | 1483

DOI 10.1287/opre.1080.0646

Constructing Uncertainty Sets for Robust Linear Optimization

Dimitris Bertsimas
Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, dbertsim@mit.edu

David B. Brown
The Fuqua School of Business, Duke University, Durham, North Carolina 27708, dbbrown@duke.edu

In this paper, we propose a methodology for constructing uncertainty sets within the framework of robust optimization for linear optimization problems with uncertain parameters. Our approach relies on decision maker risk preferences. Specifically, we utilize the theory of *coherent risk measures* initiated by Artzner et al. (1999) [Artzner, P., F. Delbaen, J. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–228.], and show that such risk measures, in conjunction with the support of the uncertain parameters, are equivalent to explicit uncertainty sets for robust optimization. We explore the structure of these sets in detail. In particular, we study a class of coherent risk measures, called *distortion risk measures*, which give rise to polyhedral uncertainty sets of a special structure that is tractable in the context of robust optimization. In the case of discrete distributions with rational probabilities, which is useful in practical settings when we are sampling from data, we show that the class of all distortion risk measures (and their corresponding polyhedral sets) are generated by a finite number of conditional value-at-risk (CVaR) measures. A subclass of the distortion risk measures corresponds to polyhedral uncertainty sets symmetric through the sample mean. We show that this subclass is also finitely generated and can be used to find inner approximations to arbitrary, polyhedral uncertainty sets.

Subject classifications: robust optimization; uncertainty sets; coherent risk measures; distortion risk measures.
Area of review: Optimization.
History: Received July 2007; revision received March 2008; accepted April 2008. Published online in *Articles in Advance* April 24, 2009.

1. Introduction

Robust optimization dates back at least 30 years to the work of Soyster (1973). With the advent of efficient algorithms for conic optimization problems, the field of robust optimization has expanded significantly over the past decade. Helping to spark this recent growth was the work of Ben-Tal and Nemirovski (1999, 2000), who show that even small perturbations of uncertain quantities can result in highly infeasible solutions. More recent work has focused on the properties of the solutions and the tractability of various robust formulations, as well as extending robustness to more general conic problems (e.g., Bertsimas et al. 2004, Bertsimas and Sim 2006, El Ghaoui and Lebret 1997, El Ghaoui et al. 1998).

In this paper, we consider the case of robust optimization in the context of linear optimization problems. In particular, for linear optimization with uncertain constraint matrix $\tilde{\mathbf{A}}$, the robust problem is without loss of generality in the form

$$\min\{\mathbf{c}'\mathbf{x}\colon \tilde{\mathbf{A}}\mathbf{x} \leqslant \mathbf{b}\ \forall\, \tilde{\mathbf{A}} \in \mathcal{U}\},$$

where $\mathbf{x} \in \mathbb{R}^n$ is a decision vector, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\tilde{\mathbf{A}} \in \mathbb{R}^{m\times n}$ is a matrix of uncertain coefficients, and $\mathcal{U}$ is an *uncertainty set* for $\tilde{\mathbf{A}}$. Note that this form is general enough to capture uncertainty in the right-hand side vector and the cost vector as well; indeed, if $\mathbf{b}$ is uncertain, we may add a variable x_{n+1} and include $\mathbf{b}$ as a column of the matrix $\tilde{\mathbf{A}}$ and constrain $x_{n+1} = 1$. If $\mathbf{c}$ is uncertain, we may add a variable t and the constraint $\mathbf{c}'\mathbf{x} \geqslant t$, and change the objective to minimizing t.

The idea of the robust optimization approach is to compute optimal solutions that retain feasibility for all possible realizations of $\mathbf{A}$ within this prescribed uncertainty set $\mathcal{U}$. The literature on robust optimization, however, is essentially silent on the the question of *constructing* these uncertainty sets. Ellipsoidal uncertainty sets, as well as other norms, are common in many treatments (e.g., Bertsimas et al. 2004). Although often rooted in some statistical considerations, these approaches are fundamentally ad hoc, with emphasis usually placed on sets that preserve computational tractability.

Here we provide a prescriptive methodology for constructing uncertainty sets within a robust optimization framework for linear optimization problems with uncertain data. We accomplish this by taking as primitive the decision maker's attitude toward risk. We show that when this attitude can be expressed in the form of a *coherent risk measure* (Artzner et al. 1999), then the optimization problem with such a risk measure is equivalent to a robust optimization problem with an explicit, convex uncertainty set.

Our approach is "data-driven," i.e., the only information on the uncertain matrix $\tilde{\mathbf{A}}$ at our disposal is a finite set of sampled matrices $\mathbf{A}_1, \ldots, \mathbf{A}_N$. In addition to avoiding complicated, distributional assumptions, this data-driven approach is well suited to practical settings, in which the realizations of the uncertain parameters are typically the only information available.

We summarize our key results as follows:

1. Given a coherent risk measure as a primitive, as well as realizations of the uncertain data in the problem, we construct a corresponding convex uncertainty set in a robust optimization framework. This is important because the uncertainty set becomes a *consequence* of the particular risk measure the decision maker selects. Thus, risk preferences that can be expressed in the form of a coherent risk measure lead to convex uncertainty sets of an explicit construction. A converse implication also holds; convex uncertainty sets within the parameter support induce a coherent risk measure.

2. We consider an important subclass of coherent risk measures that we call *distortion risk measures*, which satisfy some additional risk hedging and distribution invariance properties. In the case of a discrete distribution over rational probabilities (which may always be converted to a uniform one over a potentially larger probability space), we show that these distortion measures are generated by a *finite* number of *conditional value-at-risk* (CVaR) measures. Furthermore, we show that distortion risk measures correspond to uncertainty polytopes with a special structure that allows for efficiently solvable robust linear programs.

3. A further subclass of distortion risk measures induce uncertainty polytopes that are symmetric through the sample mean regardless of the underlying support. We show that this subclass is also finitely generated. Moreover, these symmetric uncertainty sets may be described by a norm and can be used to find inner approximations to arbitrary polyhedral sets, as we show.

We emphasize that we are attempting to make a contribution to robust optimization, *not* risk theory. For the most part, we are leveraging known results from risk theory to develop our approach. Indeed, the risk theory community is generally focused on more general probability spaces; as we have stated, our restriction to discrete spaces is motivated in part by the practical issue of sampling, as well as ease of insight into the geometry of the resulting uncertainty sets (e.g., associating extreme points with mixtures of particular "outlier" samples, etc.).[1]

A contemporaneous paper done independently by Natarajan et al. (2009) also explores the connection between risk measures and uncertainty sets. The thrust of their work, however, is to derive risk measures *from* uncertainty sets. The focus of our work here, on the other hand, is on a methodology for uncertainty set construction beginning with a risk measure as a primitive.

The outline of this paper is as follows. Section 2 introduces some necessary background from risk theory. Section 3 considers general coherent risk measures. Section 4 introduces distortion risk measures and develops the associated polyhedral uncertainty sets in detail. Section 5 concludes the paper.

Notation. Throughout the paper, $\mathbf{e}$ will denote the vector of ones and $\mathbf{e}_N \triangleq \mathbf{e}/N$. We will denote the N-dimensional probability simplex by Δ^N, i.e.,

$$\Delta^N \triangleq \{\mathbf{p} \in \mathbb{R}^N_+ : \mathbf{e}'\mathbf{p} = 1\}.$$

Also, for any two probability measures $\mathbb{Q}$ and $\mathbb{P}$ defined on the same underlying measure space $(\Omega, \mathcal{F})$, $\ll$ denotes absolute continuity; specifically,

$$\mathbb{Q} \ll \mathbb{P} \iff \mathbb{Q}(A) = 0 \quad \forall A \in \mathcal{F} \quad \text{s.t. } \mathbb{P}(A) = 0.$$

Lastly, if $\mathbb{Q} \ll \mathbb{P}$, then $d\mathbb{Q}/d\mathbb{P}$ denotes the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$.

2. Background from Risk Theory

Before describing our approach to constructing uncertainty sets, we need some background from risk theory.

2.1. Risk Measures

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{X}$ be a linear space of random variables on Ω, i.e., a set of functions $X: \Omega \to \mathbb{R}$. It is typically assumed that X is bounded; in particular, $\mathcal{X} \subseteq L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.[2] X, for the introduction, can be thought of as a reward from an uncertain position. We will use the notation $X \geqslant Y$ for $X, Y \in \mathcal{X}$ to represent state-wise dominance, i.e., $X(\omega) \geqslant Y(\omega)$ for all $\omega \in \Omega$.

We have the following definition.

Definition 2.1. A function $\mu: \mathcal{X} \to \mathbb{R}$, which satisfies, for all $X, Y \in \mathcal{X}$:

1. *Monotonicity*: If $X \geqslant Y$, then $\mu(X) \leqslant \mu(Y)$.
2. *Translation invariance*: If $c \in \mathbb{R}$, then $\mu(X + c) = \mu(X) - c$, is called a *risk measure*.

A risk measure can be interpreted as the smallest amount of capital necessary by which to augment a position X to make it "acceptable." As such, the properties above are clear; if one position X never performs worse than another position Y, then it cannot be any riskier. In addition, if we augment our position by a guaranteed amount c, then our capital requirement is reduced correspondingly by c as well. A classic example of a risk measure is the so-called *value-at-risk* defined as

$$\mathrm{VaR}_\alpha(X) \triangleq \inf\{t \in \mathbb{R}: \mathbb{P}\{t + X \geqslant 0\} \geqslant 1 - \alpha\}.$$

This risk measure can be interpreted as the smallest amount of additional capital required to ensure that a position breaks even with probability at least $1 - \alpha$. See, for example, Föllmer and Schied (2004) for more on risk measures.

2.2. Coherent Risk Measures

Although a risk measure need only satisfy translation invariance and monotonicity, we may desire additional, structural properties, such as the way risk measures should deal with diversification, etc. Artzner et al. (1999) present an axiomatic definition of risk measures satisfying some natural properties and termed such measures coherent, as we now define.

Definition 2.2. A function $\mu: \mathscr{X} \to \mathbb{R}$ that in addition to being a risk measure satisfies for all $X, Y \in \mathscr{X}$:

1. *Convexity*: If $\lambda \in [0, 1]$, then $\mu(\lambda X + (1-\lambda)Y) \leqslant \lambda\mu(X) + (1-\lambda)\mu(Y)$.
2. *Positive homogeneity*: If $\lambda \geqslant 0$, then $\mu(\lambda X) = \lambda\mu(X)$,

is called a *coherent risk measure*.

The intuition behind these axioms is fairly clear. For example, the convexity property ensures that diversification of positions can never increase risk under a coherent risk measure; this is desirable for both economic reasons (convex preferences) and computational ones (ensuring that optimization over risk measures induces convex optimization problems). The positive homogeneity axiom states that risk scales linearly with the size of a position; when this axiom is lifted, we obtain the more general class of *convex risk measures* introduced by Föllmer and Schied (2002). Note also that, when positive homogeneity holds, the convexity axiom is equivalent to the requirement of *subadditivity*, i.e.,

$$\mu(X+Y) \leqslant \mu(X) + \mu(Y) \quad \text{for all } X, Y \in \mathscr{X}.$$

One of most noteworthy coherent risk measures is the *conditional value-at-risk*, defined as

$$\mu(X) \triangleq \inf_{\nu \in \mathbb{R}} \left\{ \nu + \frac{1}{\alpha}\mathbb{E}[(-\nu - X)^+] \right\}$$

for any $\alpha \in (0, 1]$. This coherent risk measure is explored in detail by Rockafellar and Uryasev (2000). For atomless distributions, this is equivalent to $-\mathbb{E}[X \mid X \leqslant -\mathrm{VaR}_\alpha(X)]$. Delbaen (2000) shows that CVaR is the smallest upper bound to VaR among all coherent risk measures that depend only on the distribution of the underlying random variable. Acerbi and Tasche (2002) define the same risk measure, but name it *expected shortfall* (explored also by Bertsimas et al. 2004). Nemirovski and Shapiro (2006) use CVaR as a means of finding convex approximations to chance constrained optimization problems.

We will also find this risk measure to be of central importance, and it will have a variety of interesting properties which we will examine and discuss in §4.

2.3. Representation Theorem for Coherent Risk Measures

The following is the main result related to coherent risk measures. In essence, it states that we can describe any coherent risk measure equivalently in terms of expectations over a family of distributions. The result is largely a consequence of the separation theorem for convex sets. The proof actually predates the introduction of coherent risk measures (see, e.g., Chapter 10 of Huber 1981 for one version of the proof).

Theorem 2.1 (Representation of Coherent Risk Measures). *A function $\mu: \mathscr{X} \to \mathbb{R}$ is a coherent risk measure if and only if there exists a family of probability measures $\mathscr{Q}$ on $(\Omega, \mathscr{F})$ with $\mathbb{Q} \ll \mathbb{P}$ for all $\mathbb{Q} \in \mathscr{Q}$ such that*

$$\mu(X) = \sup_{\mathbb{Q} \in \mathscr{Q}} \mathbb{E}_{\mathbb{Q}}[-X] \quad \forall X \in \mathscr{X}, \tag{1}$$

where $\mathbb{E}_{\mathbb{Q}}[X]$ denotes the expectation of the random variable X under the measure $\mathbb{Q}$ (as opposed to the measure of X itself).[3]

The representation theorem says that all coherent risk measures may be represented as the worst-case expected value over a family of "generalized scenarios." For example, the generating family for CVaR is $\mathscr{Q}_\alpha = \{\mathbb{Q} \ll \mathbb{P}: d\mathbb{Q}/d\mathbb{P} \leqslant 1/\alpha\}$.

This is a duality theorem, and the connection to robustness is clear; a risk measure is coherent if and only if it can be expressed as the worst-case expected value over a family of distributions. This is a very clear *ambiguity* interpretation and is the crucial idea as we attempt to construct uncertainty sets in a robust optimization framework from a given coherent risk measure.

3. Coherent Risk Measures and Convex Uncertainty

In this section, we show how the concepts from risk theory—in particular, coherent risk measures—allow us to construct a robust counterpart to a linear optimization problem with uncertain data. We will focus on a *single* constraint of the form $\tilde{\mathbf{a}}'\mathbf{x} \geqslant b$. For multiple constraints, we can obviously apply this framework in constraintwise fashion; on the other hand, depending on how we wish to weigh the risk associated with the various constraints, this may or may not be appropriate. Chen et al. (2009) consider using the CVaR risk measure as a tractable means of addressing multiple chance constraints. The issue of vector-valued risk measures is an interesting one and very much open; see, for example, Jouini et al. (2004) for an effort at extending coherent risk measures to more general vector spaces. Because this is still unresolved, our focus is therefore on a single constraint.

We note the following issues in a practical context.

(1) We generally do not know the distribution of $\tilde{\mathbf{a}}$. In fact, we usually only have some finite number N of *observations* of the uncertain vector $\tilde{\mathbf{a}}$.

(2) Even equipped with a perfect description of the distribution of $\tilde{\mathbf{a}}$, it is not clear how we should construct an uncertainty set $\mathscr{U}$ within a robust optimization setting.

To address the first issue, we will make the following assumption.

ASSUMPTION 3.1. *The uncertain vector* $\tilde{\mathbf{a}}$ *is a random variable in* $\mathbb{R}^n$ *on the finite probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *where* $|\Omega| = N$, $\mathcal{F} = 2^\Omega$. *We denote* $\mathbf{a}_i \triangleq \tilde{\mathbf{a}}(\omega_i)$ *and the support of* $\tilde{\mathbf{a}}$ *by* $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$.

REMARK 3.1. Actually, this assumption is not really needed for the results in this section; the result will still go through for infinite probability spaces (see Natarajan et al. 2009). Because our underlying motivation in this paper is the case in which we are *data-driven*, and because we will rely heavily on the discrete space in the following sections, however, we adopt the assumption now.

REMARK 3.2. We will occasionally refer to $\mathcal{A}$ as the *data* of the problem. In some cases, it will also be convenient to use the matrix form $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$.

Thus, we assume that the sample space is confined to $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$, and $\tilde{\mathbf{a}}$ is distributed across these N values. On the one hand, this is restrictive because N could be much smaller than the size of the actual sample space, which may be astronomically large in number, if not continuous. Moreover, this approach cannot capture events outside the set of samples. On the other hand, it seems quite useful practically because the data are many times (e.g., financial applications relying on historical returns) the *only information* we have about the distribution of $\tilde{\mathbf{a}}$. In other words, although such an approach admittedly has its shortcomings, we nonetheless find it to be very useful in applications and will therefore focus on this approach for the remainder of the paper.

For the second issue, we take as primitive a coherent risk measure. The choice of this risk measure clearly depends on the preferences of the decision maker. Given a constraint based on this coherent risk measure and a distribution defined as above, we will show that there exists an equivalent robust optimization problem with a unique convex uncertainty set.

Specifically, the decision maker would like to ensure some level of conservatism for satisfying the constraint, and they impose this with a *risk aversion constraint* of the form $\mu(\tilde{\mathbf{a}}'\mathbf{x} - b) \leqslant 0$. This is appropriate in situations when simply taking the expected value of $\tilde{\mathbf{a}}$ is not a good enough guarantee; notice also that this constraint is convex in $\mathbf{x}$, which contrasts with the approach of chance constraints often proposed as a method for embedding conservatism into the optimization problem.

To put this into more concrete terms, imagine $\mathbf{x}$ as a decision vector representing allocation across n production units with uncertain production levels $\tilde{\mathbf{a}}$; we are concerned about the possibility of not meeting a particular total production level b. Meeting b in expectation is simply not a good enough guarantee. Of course, we could enforce $\mathbb{P}\{\tilde{\mathbf{a}}'\mathbf{x} \geqslant b\} \geqslant 1 - \alpha$ for some sufficiently small α, but this destroys convexity of the problem in general.

A coherent risk measure, on the other hand, is an approach that allows us to obtain some degree of conservatism without compromising convexity of the problem. For example, the constraint $\mathrm{CVaR}_\alpha(\tilde{\mathbf{a}}'\mathbf{x} - b) \leqslant 0$ says, roughly,[4] that the expected value of the total production, in the $\alpha\%$ worst cases, is no less than b. Notice also that this implies $\mathbb{P}\{\tilde{\mathbf{a}}'\mathbf{x} \geqslant b\} \geqslant 1 - \alpha$.

We have the first result, which stems in straightforward fashion from Theorem 2.1.

THEOREM 3.1. *If the risk measure* μ *is coherent and* $\tilde{\mathbf{a}}$ *is distributed as in Assumption* 3.1, *then*

$$\{\mathbf{x} \in \mathbb{R}^n\colon \mu(\tilde{\mathbf{a}}'\mathbf{x} - b) \geqslant 0\} = \{\mathbf{x} \in \mathbb{R}^n\colon \mathbf{a}'\mathbf{x} \geqslant b \ \forall\, \mathbf{a} \in \mathcal{U}\}, \quad (2)$$

where

$$\mathcal{U} = \mathrm{conv}(\{\mathbf{A}\mathbf{q}\colon \mathbf{q} \in \mathcal{Q}\}),$$

and $\mathcal{Q}$ *is the family of generating measures for* μ. *Conversely, if* $\mathcal{U} \subseteq \mathrm{conv}(\mathcal{A})$, *then* (2) *holds with the coherent risk measure generated by*

$$\mathcal{Q} = \{\mathbf{q} \in \Delta^N\colon \exists\, \mathbf{a} \in \mathcal{U} \text{ s.t. } \mathbf{A}\mathbf{q} = \mathbf{a}\}.$$

PROOF. Assume that μ is given and coherent; by Theorem 2.1 and the fact that $\tilde{\mathbf{a}}$ is distributed on $\mathcal{A}$, we have

$$\begin{aligned}\mu(\tilde{\mathbf{a}}'\mathbf{x} - b) &= \mu(\tilde{\mathbf{a}}'\mathbf{x}) + b\\ &= \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[-\tilde{\mathbf{a}}'\mathbf{x}] + b\\ &= \sup_{\mathbf{q} \in \mathcal{Q}}\{-(\mathbf{A}\mathbf{q})'\mathbf{x}\} + b\\ &= -\inf_{\mathbf{a} \in \widetilde{\mathcal{U}}}\{\mathbf{a}'\mathbf{x}\} + b\\ &= -\inf_{\mathbf{a} \in \mathcal{U}}\{\mathbf{a}'\mathbf{x}\} + b,\end{aligned}$$

where $\widetilde{\mathcal{U}} = \{\mathbf{A}\mathbf{q}\colon \mathbf{q} \in \mathcal{Q}\}$ and $\mathcal{U} = \mathrm{conv}(\widetilde{\mathcal{U}})$, and the last line follows from the simple observation that the inf of a linear function over any bounded set is equal to the inf of that function over the convex hull of that set. The converse direction follows in nearly identical fashion with the steps reversed. □

Theorem 3.1 provides a methodology for constructing robust optimization problems with uncertainty sets possessing a direct, physical meaning. The decision maker has some risk measure μ that depends on their preferences. If μ is coherent, there is an *explicit* uncertainty set that should be used in the robust optimization framework. This uncertainty set is convex and its structure depends on the generating family $\mathcal{Q}$ for μ and the data $\mathcal{A}$. We provide a few examples.

EXAMPLE 3.1 (SCENARIO-BASED SETS). Consider the coherent risk measure generated by

$$\mathcal{Q} = \mathrm{conv}(\{\mathbf{q}_1, \ldots, \mathbf{q}_m\}),$$

where $\mathbf{q}_i \in \Delta^N$. This is simply a coherent risk measure based on *scenarios* for the underlying distribution, and the connection to robustness is transparent. The uncertainty set is just

$$\mathcal{U} = \text{conv}(\{\mathbf{A}\mathbf{q}_1, \ldots, \mathbf{A}\mathbf{q}_m\}).$$

Example 3.2 (Conditional Value-at-Risk). For CVaR, we have the generating family $\mathcal{Q} = \{\mathbf{q} \in \Delta^N\colon q_i \leqslant p_i/\alpha\}$. This leads to the uncertainty set

$$\mathcal{U} = \text{conv}\left(\left\{\frac{1}{\alpha}\sum_{i\in I} p_i \mathbf{a}_i + \left(1 - \frac{1}{\alpha}\sum_{i\in I} p_i\right)\mathbf{a}_j\colon\right.\right.$$
$$\left.\left. I \subseteq \{1, \ldots, N\},\ j \in \{1, \ldots, N\}\backslash I,\ \sum_{i\in I} p_i \leqslant \alpha\right\}\right).$$

This set is a polytope. When $p_i = 1/N$ and $\alpha = j/N$ for some $j \in \mathbb{Z}_+$, this has the interpretation of the convex hull of all *j-point averages of* $\mathcal{A}$. We will explore these kinds of sets in more detail in the following section.

Example 3.3 (One-Sided Moments). For an example that uses higher-order moments, consider the risk measures

$$\mu_{r,\alpha}(X) = -\mathbb{E}[X] + \alpha\sigma_{r,-}(X),$$

where $r \geqslant 1$, $\alpha \in [0, 1]$, and

$$\sigma_{r,-}(X) \triangleq [\mathbb{E}((X - \mathbb{E}[X])^-)^r]^{1/r}.$$

These are coherent risk measures (Fischer 2001). Moreover, they are representable by the family of measures

$$\mathcal{Q}_{s,\alpha} = \{\mathbf{q} \in \Delta^N\colon q_i = p_i(1 + \alpha(\delta_i - \mathbf{p}'\boldsymbol{\delta})),\ \boldsymbol{\delta} \geqslant \mathbf{0},\ \|\delta\|_s \leqslant 1\},$$

where $s = r/(r-1)$ and $\|\mathbf{q}\|_s = (\mathbb{E}|\mathbf{q}|^s)^{1/s}$. These lead to norm-bounded uncertainty sets of the form

$$\mathcal{U} = \{\hat{\mathbf{a}} + \alpha(\mathbf{A}\mathbf{y} - (\mathbf{e}'\mathbf{y})\hat{\mathbf{a}})\colon \mathbf{y} \geqslant \mathbf{0},\ \|\mathbf{y}\|_s \leqslant 1\}.$$

The remainder of this paper focuses on classes of coherent risk measures that give rise to uncertainty sets with special structure.

4. Distortion Risk Measures and Polyhedral Uncertainty

For an arbitrary coherent risk measure μ, the uncertainty set in the corresponding robust optimization problem depends explicitly on the generator $\mathcal{Q}$ of μ given by Theorem 2.1. In general, without assuming a structural form for the coherent risk measure or imposing some additional properties on it, we cannot say anything more about the structure of the resulting uncertainty sets.

In this section, we explore a subclass of coherent risk measures that satisfy some additional properties that are often desirable in practice. We show that the resulting risk measures are equivalent to polyhedral uncertainty sets of a special structure. We begin with a bit of necessary background.

4.1. Comonotonicity, Choquet Integrals, and Law Invariance

As stated, when a risk measure satisfies positive homogeneity, the convexity property is equivalent to the property of subadditivity, i.e., for all $X, Y \in \mathcal{X}$, $\mu(X+Y) \leqslant \mu(X) + \mu(Y)$. This has the flavor of rewarding diversification; aggregating positions should never increase the risk.

Often, however, it may be that two random variables "move together" in a way that one cannot be hedged by another. In such situations, it seems reasonable that the risks should simply sum up rather than being reduced. We now define the following.

Definition 4.1. Two random variables $X, Y \in \mathcal{X}$ that satisfy, for all $(\omega, \omega') \in \Omega \times \Omega$,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0,$$

are called *comonotone*. A risk measure $\mu\colon \mathcal{X} \to \mathbb{R}$ that satisfies, for all comonotone $X, Y \in \mathcal{X}$,

$$\mu(X+Y) = \mu(X) + \mu(Y),$$

is called *comonotonic*.

As a simple example, consider a call option with strike price K and the price S of the underlying stock. The exercise value is $C = \max(0, S - K)$. Clearly, S and C are comonotone. In this case, a comonotonic risk measure would not allow us to reduce the risk of a long position in the stock with a long position in the call.

On a related note, comonotonicity is an important property when considering sums of random variables with arbitrary dependencies. Comonotone random variables have worst-case summation properties among all dependence structures, and, as such, have been used by Dhaene et al. (2002) to compute upper bounds on sums of random variables. Coherent risk measures that are also comonotonic are linked to *Choquet integrals*, which we now define with some more terminology (originally introduced by Choquet 1954 himself).

Definition 4.2. A set function $g\colon \mathcal{F} \to [0, 1]$ is called *monotone* if $g(A) \leqslant g(B)$ for all $A \subseteq B \subseteq \Omega$, and *normalized* if $g(\varnothing) = 0$ and $g(\Omega) = 1$. If, in addition, g satisfies

$$g(A \cup B) + g(A \cap B) \leqslant g(A) + g(B),$$

we say g is *submodular*.

This now allows us to define the Choquet integral, introduced in Choquet (1954).

Definition 4.3. The *Choquet integral* of a random variable $X \in \mathcal{X}$ with respect to the monotone, normalized set function $g\colon \mathcal{F} \to [0, 1]$ is defined as

$$\int X\,dg \triangleq \int_{-\infty}^{0} (g(X > x) - 1)\,dx + \int_{0}^{\infty} g(X > x)\,dx. \tag{3}$$

Choquet integrals have been used in the application of pricing insurance premia (e.g., Denneberg 1990, Wang 2000). One direction of the following result is originally due to Dellacherie (1970); Schmeidler (1986) later completed it.

THEOREM 4.1 (SCHMEIDLER 1986). *A coherent risk measure* μ: $\mathscr{X} \to \mathbb{R}$ *is comonotonic if and only if it can be written in the form* $\int(-X)\,dg$, *where* g: $\mathscr{F} \to [0,1]$ *is a monotone, normalized, and submodular set function.*

For example, CVaR can be written as the Choquet integral of the function $g(A) = \min\{\mathbb{P}\{A\}/\alpha, 1\}$, which is clearly monotone, normalized, and submodular; this alone tells us that CVaR is comonotonic and coherent.

We define one more property of risk measures; this property is intimately connected to the ability to estimate risk measures from historical data.

DEFINITION 4.4. A risk measure μ: $\mathscr{X} \to \mathbb{R}$ that satisfies $\mu(X) = \mu(Y)$ for all $X, Y \in \mathscr{X}$ such that X and Y have the same distribution under $\mathbb{P}$ is called *law invariant*.

Although law invariance is intuitive, it could be violated when there is some notion of ordering to the events of the underlying probability space. As a simple example, consider $|\Omega| = 2$ and let $X = (1,0)$ and $Y = (0,1)$ be two random variables on Ω, with $\mathbb{P} = (1/2, 1/2)$. Clearly, X and Y have the same distribution under $\mathbb{P}$. On the other hand, consider the coherent risk measure defined by the singleton family $\mathscr{Q} = \{(1/3, 2/3)\}$; then $\mu(X) = -1/3 \neq \mu(Y) = -2/3$.

On the other hand, law invariance is an eminently reasonable property when we are dealing with the situation when we are estimating risk measures from data, which is our motivation. In fact, if a risk measure is not law invariant, then in general it is impossible to consistently estimate the risk measure from data. To see this, consider X and Y with the same distribution and let μ be a risk measure such that $\mu(X) \neq \mu(Y)$. Let $\{X_i\}$ and $\{Y_i\}$ be a sequence of N i.i.d. samples of X and Y, and let $\hat{\mu}$: $\mathbb{R}^N \to \mathbb{R}$ be any estimator of μ that depends only on the sample distribution. As $N \to \infty$, the sequences are asymptotically identical, so $\lim_{N\to\infty} \hat{\mu}(X_1, \ldots, X_N) = \lim_{N\to\infty} \hat{\mu}(Y_1, \ldots, Y_N)$. On the other hand, $\mu(X) \neq \mu(Y)$, so at least one of the estimators must be asymptotically wrong. Therefore, law invariance is a necessary condition to be able to consistently estimate the risk measure.

Summarizing the properties in this section, we introduce the following nomenclature.

DEFINITION 4.5. A coherent risk measure that is also comonotonic and law invariant is called a *distortion risk measure*.[5]

4.2. Representation of Distortion Risk Measures

We now show a representation theorem for distortion risk measures in terms of conditional value-at-risk. The main result is already known in the case of atomless distributions (Kusuoka 2001, Föllmer and Schied 2004). Because our focus is on the data-driven case, i.e., discrete distributions, we cannot directly apply these previously known results. Moreover, in contrast to the known results in the atomless case, our result shows generation by a *finite* number of conditional value-at-risk measures. Throughout the remainder of the paper, we will operate under the following assumption.

ASSUMPTION 4.1. *The distribution* $\mathbb{P}$ *satisfies* $\mathbb{P}\{\omega_i\} = 1/N$ *for all* $i \in \{1, \ldots, N\}$.

REMARK 4.1. Although this is obviously a very special distribution, it is not a very limiting assumption given that we have already restricted ourselves to discrete spaces. One of the central motivations of this paper is to explore the connection between risk measures and uncertainty sets in the setting, very common in practice, in which we are obtaining distributional information from samples. As such, this is obviously far and away the most relevant discrete distribution; of course, it also encompasses any distribution $\mathbb{P}\{\omega_i\} = k_i/N$, where $k_i \in \mathbb{Z}_+$ and $k_1 + \cdots + k_N = N$ simply by replicating samples (or discarding them if $k_i = 0$). Furthermore, given an arbitrary, discrete distribution representable with rational numbers, we may always convert it to such a form for some N. The price we will pay for such a conversion to a larger N will be that of increased complexity in terms of the size of the corresponding robust problem.

We now work toward a representation theorem, which tells us that a risk measure in this setting is a distortion risk measure if and only if it is a mixture of CVaR measures. We start with the following lemma.

LEMMA 4.1. *A risk measure* μ *is a distortion risk measure if and only if there exists a function* ν: $[0,1] \to [0,1]$, *satisfying* $\int_{\alpha=0}^{1} \nu(d\alpha) = 1$, *such that*

$$\mu(X) = \int_{\alpha=0}^{1} \mathrm{CVaR}_\alpha(X)\,\nu(d\alpha).$$

PROOF. Note that the result is known (Kusuoka 2001) in the case of atomless distributions. Leitner (2005) proves a result for second-order stochastic dominance preserving (a stronger condition than law invariance) coherent risk measures on general probability spaces, but does not consider comonotonicity. As far as we know, our claim in this special case has not been shown; nonetheless, our proof closely follows the proof of the atomless case from Föllmer and Schied (2004).

One direction is clear because it is well known that CVaR is a distortion risk measure, and such risk measures are closed under convex combinations. For the other direction, let μ be a distortion risk measure; because it is comonotonic and coherent, Theorem 4.1 tells us there exists a normalized, monotone, and submodular function g such that

$\mu(X) = \int(-X)\,dg$. Furthermore, because μ is law invariant, g must be a function of the probability alone, i.e., there exists a function θ: $[0,1] \to [0,1]$ that is nondecreasing and satisfies $\theta(0) = 0$ and $\theta(1) = 1$ such that $\theta(\mathbb{P}\{A\}) := g(A)$ for all $A \in \mathcal{F}$. Note that this only defines θ at the points i/N, $i \in \{0, \ldots, N\}$; we simply take the piecewise-linear function through these points. If we can show that this piecewise-linear function θ is concave, the result follows by Föllmer and Schied (2004, Theorem 4.64).

To this end, we need to show that $\theta(i/N) - \theta((i-1)/N) \geqslant \theta((i+1)/N) - \theta(i/N)$ for all $i \in \{1, \ldots, N-1\}$. Choosing $A_i = \bigcup_{k=1}^{i} \omega_k$, this is equivalent to $g(A_i) - g(A_i \backslash \omega_i) \geqslant g(A_i \cup \omega_{i+1}) - g(A_i)$. However, a function g is submodular if and only if it satisfies this very property of nonincreasing second differences (e.g., Queyranne 2002), so we are done. □

Remark 4.2. Note that Assumption 4.1 is critical for the reverse direction of the proof above. Indeed, consider a probability space $\mathbb{P}\{\omega_1\} = 1/3$, $\mathbb{P}\{\omega_2\} = 2/3$ on $\Omega = \{\omega_1, \omega_2\}$. The function $g(\varnothing) = 0$, $g(\{\omega_1\}) = 0$, $g(\{\omega_2\}) = 1$, and $g(\Omega) = 1$ is normalized, monotone, submodular, and a function of the probability alone, but the resulting θ above is not concave. If the probabilities are equal, however, law invariance requires $g(\{\omega_1\}) = g(\{\omega_2\})$, which eliminates the possibility of such counterexamples.

We can strengthen this representation to one of *finite* generation, as we now show. We start with a definition.

Definition 4.6. The *restricted simplex in N-dimensions* is denoted by $\hat{\Delta}^N$ and defined as

$$\hat{\Delta}^N \triangleq \{\mathbf{q} \in \Delta^N\colon q_1 \geqslant \cdots \geqslant q_N\}.$$

Theorem 4.2. *A risk measure μ is a distortion risk measure if and only if there exists a $\mathbf{q} \in \hat{\Delta}^N$ such that*

$$\mu(X) = -\sum_{i=1}^{N} q_i x_{(i)}, \tag{4}$$

where $x_{(i)}$ are the increasing order statistics of X, i.e., $x_{(1)} \leqslant \cdots \leqslant x_{(N)}$. Moreover, every such $\mathbf{q} \in \hat{\Delta}^N$ may be written in the form

$$\mathbf{q} = \sum_{i=1}^{N} \lambda_j \hat{\mathbf{q}}^j, \tag{5}$$

where $\boldsymbol{\lambda} \geqslant \mathbf{0}$, $\sum_{j=1}^{N} \lambda_j = 1$, and $\hat{\mathbf{q}}^j \in \hat{\Delta}^N$ corresponds to the distortion risk measure $CVaR_{j/N}$.

Proof. First, consider a risk measure given as in (4). Such a risk measure is coherent because it is generated by the set $\mathcal{Q} = \{\mathbf{q}' \in \Delta^N\colon q'_i = q_{\sigma(i)}, \sigma \in S(N)\}$, where $S(N)$ is all permutations of N elements. It also obviously comonotonic. Law invariance follows because if X and Y have the same distribution under Assumption 4.1, they must have $x_{(i)} = y_{(i)}$ for all $i \in \{1, \ldots, N\}$.

For the other direction of the first part, note that we have, for $\alpha \geqslant 1/N$,

$$\begin{aligned}\mathrm{CVaR}_\alpha(X) &= \sup_{\{\mathbf{q}\in\Delta^N:\, q_i \leqslant 1/(N\alpha)\}} \mathbb{E}_{\mathbf{q}}[-X]\\ &= -\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha\rfloor} x_{(i)} - \left(\frac{N\alpha - \lfloor N\alpha\rfloor}{\lfloor N\alpha\rfloor}\right) x_{(\lceil N\alpha\rceil)}\\ &\triangleq -\sum_{i=1}^{N} q_i^\alpha x_{(i)}\end{aligned}$$

(and for $\alpha < 1/N$, we have $\mathbf{q}^\alpha = \mathbf{q}^{1/N}$). Via Lemma 4.1, we can write μ in the form (4) with $q_i = \int q_i^\alpha \nu(d\alpha)$; because $\mathbf{q}^\alpha \in \hat{\Delta}^N$, we have $\mathbf{q} \in \hat{\Delta}^N$ as well, which completes the first part of the proof.

To show finite generation of $\mathbf{q}$, consider the matrix $\mathbf{Q}_N$ with columns $\hat{\mathbf{q}}^j$, i.e.,

$$\mathbf{Q}_N = \begin{bmatrix} 1 & \cdots & \dfrac{1}{N-2} & \dfrac{1}{N-1} & \dfrac{1}{N}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \dfrac{1}{N-2} & \dfrac{1}{N-1} & \dfrac{1}{N}\\ 0 & \cdots & 0 & \dfrac{1}{N-1} & \dfrac{1}{N}\\ 0 & \cdots & 0 & 0 & \dfrac{1}{N}\end{bmatrix},$$

and define the vector $\boldsymbol{\lambda} \in \mathbb{R}^N$ as $\lambda_N = Nq_N$, $\lambda_{N-j} = (N-j)\cdot(q_{N-j} - q_{N-j+1})$ for all $j \in \{1, \ldots, N-1\}$. We have $\mathbf{q} \in \hat{\Delta}^N$, so $q_j \leqslant q_{j+1}$, and thus $\boldsymbol{\lambda} \geqslant \mathbf{0}$. In addition,

$$\sum_{j=1}^{N} \lambda_j = \sum_{j=1}^{N} q_j = 1.$$

Finally, we compute the vector $\mathbf{Q}_N\boldsymbol{\lambda}$ and see that

$$\begin{aligned}[\mathbf{Q}_N\boldsymbol{\lambda}]_i &= \sum_{j=1}^{i} \frac{1}{N-j+1}\lambda_{N-j+1}\\ &= q_N + (q_{N-1} - q_N) + \cdots + (q_{i+1} - q_{i+2})\\ &\quad + (q_i - q_{i+1})\\ &= q_i,\end{aligned}$$

which completes the proof. □

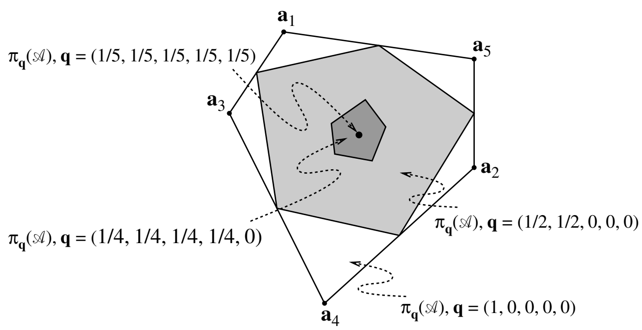
4.3. Connecting to Polyhedral Uncertainty

With the representation for distortion measures in hand, we now connect to their induced uncertainty sets. Theorem 4.2 indicates that any distortion risk measure is generated by the family

$$\mathcal{Q} = \{\mathbf{q}' \in \Delta^N\colon q'_i = q_{\sigma(i)}, \sigma \in S(N)\}$$

for some $\mathbf{q} \in \hat{\Delta}^N$. This motivates the following definition.

Figure 1. $\Pi_{\mathbf{q}}(\mathcal{A})$ for various $\mathbf{q}$ for an example with $N = 5$.



Definition 4.7. For some measure $\mathbf{q} \in \Delta^N$ and discrete set $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$ with $\mathbf{a}_i \in \mathbb{R}^n$ for all $i \in \{1, \ldots, N\}$, we define the $\mathbf{q}$-*permutohull* of $\mathcal{A}$ by

$$\Pi_{\mathbf{q}}(\mathcal{A}) = \operatorname{conv}\left(\left\{\sum_{i=1}^{N} q_{\sigma(i)}\mathbf{a}_i \colon \sigma \in S_N\right\}\right).$$

Figure 1 illustrates a family of such polyhedra. With most of the work done, we are now able to link these sets to distortion measures.

Theorem 4.3. *If a risk measure μ is a distortion risk measure, the following hold*:

$$\begin{aligned}&\{\mathbf{x} \in \mathbb{R}^n \colon \mu(\tilde{\mathbf{a}}'\mathbf{x} - b) \geqslant 0\}\\ &\quad = \{\mathbf{x} \in \mathbb{R}^n \colon \mathbf{a}'\mathbf{x} \geqslant b \ \forall\, \mathbf{a} \in \Pi_{\mathbf{q}}(\mathcal{A})\}\\ &\quad = \{\mathbf{x} \in \mathbb{R}^n \colon \exists (\mathbf{y}_1, \mathbf{y}_2) \in \mathbb{R}^N \times \mathbb{R}^N, \text{ s.t. } \mathbf{e}'\mathbf{y}_1 + \mathbf{e}'\mathbf{y}_2 \geqslant b,\\ &\qquad\qquad y_{1,i} + y_{2,j} \leqslant q_i \cdot (\mathbf{a}_j'\mathbf{x}) \ \forall\, (i,j) \in \{1, \ldots, N\}^2\}.\end{aligned}$$

Proof. The first equality follows by applying Theorems 3.1 and 4.2. To show the second one, we consider the problem for fixed $\mathbf{x}$, $\min_{\mathbf{a} \in \Pi_{\mathbf{q}}(\mathcal{A})} \mathbf{x}'\mathbf{a}$. This is equivalent to the problem

$$\begin{aligned}\text{minimize}\quad & \sum_{i,j} q_i \cdot (\mathbf{a}_j'\mathbf{x}) \cdot w_{ij}\\ \text{subject to}\quad & \sum_{j=1}^{N} w_{ij} = 1 \quad \forall\, i \in \{1, \ldots, n\},\\ & \sum_{i=1}^{N} w_{ij} = 1 \quad \forall\, j \in \{1, \ldots, n\},\\ & w_{ij} \geqslant 0 \quad \forall\, (i,j) \in \{1, \ldots, N\}^2,\end{aligned} \tag{6}$$

in variables $w_{ij} \in \mathbb{R}^{N^2}$. The dual problem is

$$\begin{aligned}\text{maximize}\quad & \mathbf{e}'\mathbf{y}_1 + \mathbf{e}'\mathbf{y}_2\\ \text{subject to}\quad & \mathbf{e}_i'\mathbf{y}_1 + \mathbf{e}_j'\mathbf{y}_2 \leqslant q_i \cdot \mathbf{a}_j'\mathbf{x},\end{aligned} \tag{7}$$

in variables $\mathbf{y}_1 \in \mathbb{R}^N$, $\mathbf{y}_2 \in \mathbb{R}^N$. Because strong duality holds between (6) and (7) (because (6) has a nonempty, bounded feasibleset), we may replace the left-hand side of the robust constraint with the objective from (7) and add in the dual constraints as well, leaving us with the desired result. □

We remark that although $\Pi_{\mathbf{q}}(\mathcal{A})$ has as many as $N!$ extreme points, the complexity of using it as an uncertainty set is polynomial in N. Specifically, the equivalent problem formulation has only $2N$ extra variables and N^2 extra constraints. In the case of CVaR, this reduces to $\mathcal{O}(N)$ additional variables and constraints.

Corollary 4.3. *The class of polytopes*

$$\mathcal{U}_\lambda(\mathcal{A}) = \operatorname{conv}\left(\left\{\sum_{j=1}^{N} \lambda_j \left[\frac{1}{j}\sum_{i=1}^{j} \mathbf{a}_{\sigma_j(i)}\right] \colon \sigma_j \in S(N),\ j \in \{1, \ldots, N\}\right\}\right), \quad \boldsymbol{\lambda} \geqslant \mathbf{0},\ \sum_{j=1}^{N} \lambda_j = 1,$$

is the class of all uncertainty sets induced by all distortion risk measures.

Proof. From Theorems 4.2 and 4.3, the possible robust constraints for distortion risk measures are of the form

$$\sum_{j=1}^{N} \lambda_j \min_{\mathbf{a} \in \Pi_{\hat{\mathbf{q}}^j}(\mathcal{A})} (\mathbf{a}'\mathbf{x}) \geqslant b, \tag{8}$$

where $\boldsymbol{\lambda} \geqslant \mathbf{0}$ and $\sum_{j=1}^{N} \lambda_j = 1$. From here, (8) holds if and only if $\sum_{j=1}^{N} \lambda_j \mathbf{v}_j'\mathbf{x} \geqslant b$ for all choices of extreme points $\mathbf{v}_j$ of $\Pi_{\hat{\mathbf{q}}^j}(\mathcal{A})$. The result then follows by recognizing that these extreme points are contained in the set of all j-point averages of all permutations of the elements of $\mathcal{A}$. □

Figure 2 highlights the result of Theorem 4.3. The $\text{CVaR}_{j/N}$ measures effectively generate the space of all distortion risk measures and, in this sense, are fundamental.

4.4. Distortion Risk Measures and Centrally Symmetric Polytopes

In this section, we study a more restricted class of generating measures $\mathbf{q}$; specifically, we study those that lead to polyhedral uncertainty sets obeying a specific symmetry property. These structures are useful because they naturally induce norm spaces that we will use in the next section to approximate arbitrary polyhedral uncertainty sets by those corresponding to distortion risk measures.

Definition 4.8. A set P is *centrally symmetric through* $\mathbf{x}_0 \in P$ if $\mathbf{x}_0 + \mathbf{x} \in P$ implies $\mathbf{x}_0 - \mathbf{x} \in P$.

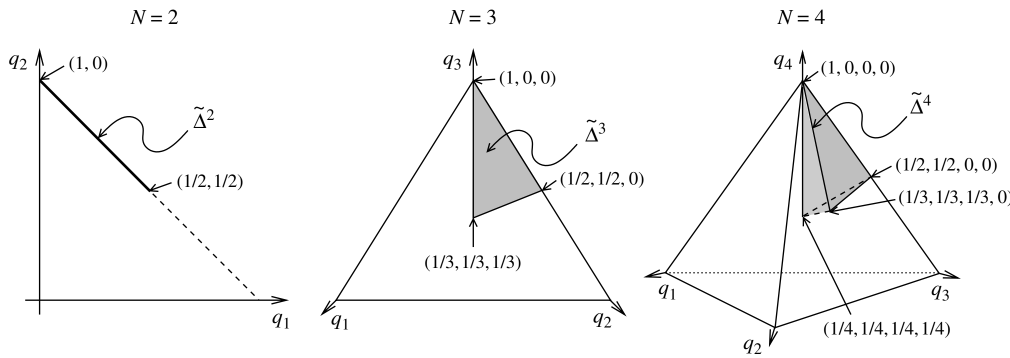
Here we will be interested in uncertainty sets that are symmetric through the sample mean of the data. It turns out that this will be true for any data if and only if $\mathbf{q}$ satisfies a certain property.

Proposition 4.1. *For $\mathbf{q} \in \hat{\Delta}^N$, $\Pi_{\mathbf{q}}(\mathcal{A})$ is centrally symmetric through $\hat{\mathbf{a}} = \mathbf{A}\mathbf{e}_N$ for any $\mathcal{A}$ if and only if*

$$\mathbf{q} = 2\mathbf{e}_N - \mathbf{q}_\sigma, \tag{9}$$

where σ is a permutation and $\mathbf{e}_N$ is the vector with $1/N$ at each entry.

Figure 2. Generation of $\hat{\Delta}^N$ by the CVaR measures for $N = 2, 3, 4$.



Proof. $\Pi_{\mathbf{q}}(\mathscr{A})$ is centrally symmetric if and only if every extreme point is centrally symmetric. An extreme point of $\Pi_{\mathbf{q}}(\mathscr{A})$ is of the form $\mathbf{A}\mathbf{q}_{\sigma'}$ for some permutation σ'. Letting $\boldsymbol{\delta} = \mathbf{A}\mathbf{q}_{\sigma'} - \mathbf{A}\mathbf{e}_N$, $\Pi_{\mathbf{q}}(\mathscr{A})$ is symmetric if and only if $\mathbf{A}\mathbf{e}_N - \boldsymbol{\delta} = 2\mathbf{A}\mathbf{e}_N - \mathbf{A}\mathbf{q}_{\sigma'} \in \Pi_{\mathbf{q}}(\mathscr{A})$ for every permutation σ'. Because this must hold for any $\mathbf{A}$, this is true if and only if there exists a permutation σ'' such that $2\mathbf{e}_N - \mathbf{q}_{\sigma'} = \mathbf{q}_{\sigma''}$ for every permutation σ'. Clearly, (9) implies this to be true. Conversely, choose σ' as the identity, which means $\mathbf{q} = 2\mathbf{e}_N - \mathbf{q}_{\sigma''}$ for some other permutation σ'', which is the same as (9). □

This motivates the following definition.

Definition 4.9. The *symmetric restricted simplex in N-dimensions* is denoted by $\hat{\Delta}^N_{\text{sym}}$ and defined by

$$\hat{\Delta}^N_{\text{sym}} = \{\mathbf{q} \in \hat{\Delta}^N \colon \mathbf{q} \text{ satisfies (9)}\}.$$

We now prove that, like the restricted simplex, $\hat{\Delta}^N_{\text{sym}}$ is generated by a finite family of distortion risk measures.

Theorem 4.4. *The class of distortion risk measures that generates centrally symmetric sets $\Pi_{\mathbf{q}}(\mathscr{A})$ for any $\mathscr{A}$ is equivalent to the class of risk measures*

$$\mu_{\lambda}(X) = \sum_{j=1}^{\widehat{N}} \lambda_j \hat{\mu}_j(X), \quad \boldsymbol{\lambda} \geqslant \mathbf{0}, \quad \sum_{j=1}^{\widehat{N}} \lambda_j = 1,$$

where $\widehat{N} = \lfloor N/2 \rfloor + 1$ and $\mu_j(X)$ is the distortion risk measure $\mu_j(X) = -\sum_{i=1}^N q_i^j x_{(i)}$ with

$$q_i^j = \begin{cases} 2/N & \text{if } i < j, \\ 1/N & \text{if } j \leqslant i \leqslant N - j + 1, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Proof. We assume that N is odd; the proof for N even is nearly identical, with some small changes on summation limits. We define the matrix $\mathbf{Q}_N \in \mathbb{R}^{N \times \widehat{N}}$ by

$$\mathbf{Q}_N = \begin{bmatrix} \hat{\mathbf{q}}^1 & \cdots & \hat{\mathbf{q}}^{\widehat{N}} \end{bmatrix}.$$

Analogous to Theorem 4.2, if we can show that the convex hull of the columns of $\mathbf{Q}_N$ is equivalent to $\hat{\Delta}^N_{\text{sym}}$, we will prove the claim.

For one direction, consider an arbitrary convex combination $\boldsymbol{\lambda} \in \mathbb{R}^{\widehat{N}}$ of the columns of $\mathbf{Q}_N$, i.e., a vector $\mathbf{q} \in \mathbb{R}^N$ such that $\mathbf{q} = \mathbf{Q}_N \boldsymbol{\lambda}$. Clearly, $\mathbf{q} \geqslant \mathbf{0}$ and $\mathbf{q}$ sums to one. We find that

$$q_i = \begin{cases} \dfrac{1}{N} + \dfrac{1}{N} \displaystyle\sum_{j=i+1}^{\widehat{N}} \lambda_j, & i < \widehat{N}, \\ \dfrac{1}{N} \displaystyle\sum_{j=1}^{2\widehat{N}-i} \lambda_j, & i \geqslant \widehat{N}. \end{cases}$$

We see that we can rearrange this to find that

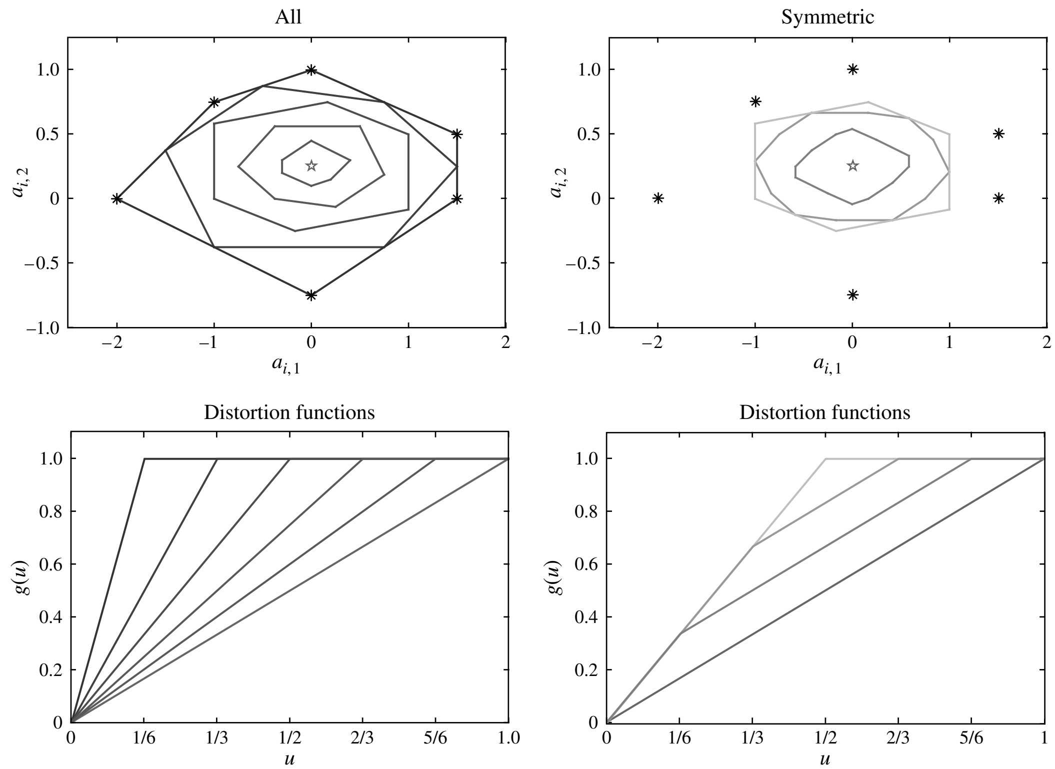
$$\begin{aligned} q_N &= \lambda_1 / N, \\ q_{N-i} &= q_{N-i+1} + \lambda_{i+1}/N, \quad 1 \leqslant i \leqslant \widehat{N} - 1, \\ q_{\widehat{N}-i} &= q_{\widehat{N}-i+1} + \lambda_{\widehat{N}-i+1}/N, \quad 1 \leqslant i \leqslant \widehat{N} - 1, \end{aligned} \tag{11}$$

from which it follows that $q_i \geqslant q_{i+1}$ as well, so $\mathbf{q} \in \hat{\Delta}^N$. We now check the symmetry condition from Proposition 4.1. We have

$$\begin{aligned} \tilde{\mathbf{q}} &= \frac{2}{N}\mathbf{e} - \mathbf{q} \\ &= \begin{cases} \dfrac{1}{N} - \dfrac{1}{N} \displaystyle\sum_{j=i+1}^{\widehat{N}} \lambda_j, & i < \widehat{N}, \\ \dfrac{2}{N} - \dfrac{1}{N} \displaystyle\sum_{j=1}^{2\widehat{N}-i} \lambda_j, & i \geqslant \widehat{N}, \end{cases} \\ &= \begin{cases} \dfrac{1}{N} \displaystyle\sum_{j=1}^{i} \lambda_j, & i < \widehat{N}, \\ \dfrac{1}{N} + \dfrac{1}{N} \displaystyle\sum_{j=2\widehat{N}-i+1}^{\widehat{N}} \lambda_j, & i \geqslant \widehat{N}, \end{cases} \end{aligned}$$

and we see that $\tilde{q}_i = q_{\sigma(i)}$ under the permutation $\sigma(i) = N - i + 1$ for all $i \in \{1, \ldots, N\}$. This shows that $\text{conv}(\mathbf{Q}_N) \subseteq \hat{\Delta}^N_{\text{sym}}$.

Figure 3. For a case with $N = 6$ data points (denoted by $*$): the six generators for $\Pi_{\mathbf{q}}(\mathcal{A})$ (upper left) and the corresponding distortion ($\text{CVaR}_{j/N}$) functions (lower left); the four generators for the centrally symmetric subclass of $\Pi_{\mathbf{q}}(\mathcal{A})$ (upper right) and the corresponding distortion functions (lower right).



For the reverse inclusion, consider any $\mathbf{q} \in \hat{\Delta}^N_{\text{sym}}$. Now construct a $\boldsymbol{\lambda} \in \mathbb{R}^{\widehat{N}}$ by reversing the construction above in (11). This leads us to

$$\lambda_1 = Nq_N,$$
$$\lambda_i = N(q_{N-i+1} - q_{N-i+2}) \quad \forall\, i \in \{2, \ldots, \widehat{N}\}.$$

From the fact that the q_i are nonincreasing, we see that $\boldsymbol{\lambda} \geqslant \mathbf{0}$. In addition, we find that

$$\sum_{i=1}^{\widehat{N}} \lambda_i = N(q_{\widehat{N}-1} + q_{\widehat{N}}) - 1 = N(2/N) - 1 = 1,$$

so $\mathbf{q} \in \text{conv}(\mathbf{Q}_N)$, and we are done. □

Figure 3 depicts a simple two-dimensional case with $N = 6$ data points. Shown are the polytopes corresponding to $\text{CVaR}_{j/N}$ that generate all $\Pi_{\mathbf{q}}(\mathcal{A})$, as well as those corresponding to the distortion risk measures that generate the space of all centrally symmetric polytopes.

4.5. Tight Distortion Risk Measure Approximations to Arbitrary Polytopes

Given an arbitrary polytope contained within the support as the uncertainty set, we can construct a corresponding coherent risk measure, but not necessarily a distortion risk measure. In this section, we consider the problem of finding the largest uncertainty set contained in a polytope that also corresponds to a distortion risk measure. This would be useful, for example, in determining how conservative a particular coherent risk measure (or, uncertainty set in the robust setting) is compared to a distortion risk measure.

In general, it is not clear how to do this algorithmically, but among the special class of distortion risk measures inducing centrally symmetric $\Pi_{\mathbf{q}}(\mathcal{A})$, we are able to do so in a tractable way. This is because these risk measures naturally induce a norm, and thus provide a convenient way of approximating more general measure structures, as we now illustrate.

We first describe the norm induced by the sets $\Pi_{\mathbf{q}}(\mathcal{A})$, where $\mathbf{q} \in \hat{\Delta}^N_{\text{sym}}$.

Proposition 4.2. *For a comonotone generator* $\mathbf{q} \in \hat{\Delta}^N_{\text{sym}}$ *and any* $\mathcal{A}$, *with* $\hat{\mathbf{a}} = \mathbf{A}\mathbf{e}_N$, *the function*

$$\|\mathbf{a} - \hat{\mathbf{a}}\|_{\mathbf{q},\mathcal{A}} = \inf\left\{\alpha > 0 \;\middle|\; \frac{\mathbf{a} - \hat{\mathbf{a}}}{\alpha} \in \tilde{\pi}_{\mathbf{q}}(\mathcal{A})\right\}, \tag{12}$$

where $\tilde{\pi}_{\mathbf{q}}(\mathcal{A})$ *is* $\Pi_{\mathbf{q}}(\mathcal{A})$ *shifted by* $-\hat{\mathbf{a}}$, *is a norm.*

Proof. $\|\cdot\|_{\mathbf{q},\mathcal{A}}$ is a form of a Minkowski function, and it is well known that this function is convex whenever the underlying set in question (here $\tilde{\pi}_{\mathbf{q}}(\mathcal{A})$) is closed and convex, which is the case in this construction. Without loss of generality, we assume that $\hat{\mathbf{a}} = \mathbf{0}$ in the remainder of this proof.

$\|\mathbf{a}\|_{\mathbf{q},\mathcal{A}} = 0$ implies that $\mathbf{a}/\epsilon \in \Pi_{\mathbf{q}}(\mathcal{A})$ for all $\epsilon > 0$. Because $\Pi_{\mathbf{q}}(\mathcal{A})$ is a bounded set, this can only be the case if $\mathbf{a} = \mathbf{0}$.

If $\beta > 0$, it is easy to see that $\|\beta\mathbf{a}\|_{\mathbf{q},\mathcal{A}} = \beta\|\mathbf{a}\|_{\mathbf{q},\mathcal{A}}$ by a simple scaling argument. If $\beta < 0$, we have $\beta\mathbf{a} \in \Pi_{\mathbf{q}}(\mathcal{A})$ if and only if $\Pi_{\mathbf{q}}(\mathcal{A})$ is centrally symmetric through zero; this is the case, however, because $\mathbf{q} \in \hat{\Delta}^N_{\text{sym}}$. Combining all this, we see that $\|\beta\mathbf{a}\|_{\mathbf{q},\mathcal{A}} = |\beta|\|\mathbf{a}\|_{\mathbf{q},\mathcal{A}}$.

Finally, noting that this function is convex, we have, for all $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^n$,

$$\begin{aligned}\|\mathbf{a}_1 + \mathbf{a}_2\|_{\mathbf{q},\mathcal{A}} &= \|2(0.5\mathbf{a}_1 + 0.5\mathbf{a}_2)\|_{\mathbf{q},\mathcal{A}}\\ &= 2\|0.5\mathbf{a}_1 + 0.5\mathbf{a}_2\|_{\mathbf{q},\mathcal{A}}\\ &\leqslant \|\mathbf{a}_1\|_{\mathbf{q},\mathcal{A}} + \|\mathbf{a}_2\|_{\mathbf{q},\mathcal{A}},\end{aligned}$$

which completes the proof that it is a norm. □

The norm $\|\cdot\|_{\mathbf{q},\mathcal{A}}$ has one particular property that is of interest.

Lemma 4.2. *Let $\mathbf{q} \in \hat{\Delta}^N_{\text{sym}}$ be any centrally symmetric generator and $\lambda \in \mathbb{R}$. Then, the vector $\tilde{\mathbf{q}} = \lambda\mathbf{q} + (1-\lambda)\mathbf{e}_N$ satisfies*

$$\|\mathbf{a} - \hat{\mathbf{a}}\|_{\tilde{\mathbf{q}},\mathcal{A}} = \frac{1}{|\lambda|}\|\mathbf{a} - \hat{\mathbf{a}}\|_{\tilde{\mathbf{q}},\mathcal{A}} \tag{13}$$

for all $\mathbf{a} \in \mathbb{R}^n$.

Proof. The proof follows by noting that $\Pi_{\tilde{\mathbf{q}}}(\mathcal{A})$ is a scaled version of $\Pi_{\tilde{\mathbf{q}}}(\mathcal{A})$ around $\hat{\mathbf{a}}$ by a factor of λ. Indeed, it is easy to see that the extreme points of $\Pi_{\tilde{\mathbf{q}}}(\mathcal{A})$ are affine combinations of the extreme points of $\Pi_{\tilde{\mathbf{q}}}(\mathcal{A})$ and $\hat{\mathbf{a}}$. Note that because of this we may as well assume that $\lambda \geqslant 0$ because $\lambda < 0$ reflects the set through $\hat{\mathbf{a}}$, and it is centrally symmetric through this point by construction. The result then follows from the definition of $\|\cdot\|_{\mathbf{q},\mathcal{A}}$. □

Now consider a robust optimization constraint with an arbitrary polytope $\mathcal{U}$ as the uncertainty set. We assume that the sample mean $\hat{\mathbf{a}} \in \mathcal{U}$. In particular, assume that $\mathcal{U}$ has the form

$$\mathcal{U} = \{\mathbf{a} \in \mathbb{R}^n\colon\ \mathbf{u}_i'\mathbf{a} \geqslant v_i,\ i \in \{1, \ldots, m\}\}. \tag{14}$$

In such cases, we can always find a distortion risk measure that leads to an inner approximation (a less conservative uncertainty set) on the robust problem with $\mathcal{U}$. We will find an inner approximation that is centrally symmetric through $\hat{\mathbf{a}}$ and use the norm $\|\cdot\|_{\mathbf{q},\mathcal{A}}$ derived in Proposition 4.2 to measure the quality of the approximation.

Lemma 4.2 immediately suggests a method for finding an inner approximation to an arbitrary uncertainty polytope $\mathcal{U}$: Begin with a centrally symmetric generator $\mathbf{q} \in \hat{\Delta}^N_{\text{sym}}$ and "mix" it with as little of the generator $\mathbf{e}_N$ such that the result is contained in $\mathcal{U}$. From Lemma 4.2, the resulting $\Pi_{\mathbf{q}^*}(\mathcal{A})$ will be the largest in the $\|\cdot\|_{\mathbf{q},\mathcal{A}}$ sense among all such mixtures contained in $\mathcal{U}$. We now show how to compute this algorithmically via linear optimization.

Theorem 4.5. *Given a centrally symmetric generator $\hat{\mathbf{q}} \in \hat{\Delta}^N_{\text{sym}}$ and an arbitrary polytope $\mathcal{U} \subseteq \mathbb{R}^n$ described by* (14) *such that $\hat{\mathbf{a}} \in \mathcal{U}$, the centrally symmetric $\Pi_{\mathbf{q}}(\mathcal{A})$, which is largest in the $\|\cdot\|_{\hat{\mathbf{q}},\mathcal{A}}$ sense among all such $\Pi_{\mathbf{q}}(\mathcal{A}) \subseteq \mathcal{U}$, is given by the solution to the linear optimization problem*

$$\begin{aligned}&\text{maximize} && \lambda\\ &\text{subject to} && \mathbf{q} = \lambda\hat{\mathbf{q}} + (1-\lambda)\mathbf{e}/N,\\ &&& \mathbf{e}'(\mathbf{s}_k + \mathbf{t}_k) \geqslant v_k \quad \forall\, k \in \{1, \ldots, m\},\\ &&& s_{k,i} + t_{k,j} \leqslant (\mathbf{u}_k'\mathbf{a}_j)q_i\\ &&& \forall\,(i,j) \in \{1, \ldots, N\}^2,\ \forall\, k \in \{1, \ldots, m\},\end{aligned} \tag{15}$$

in variables $\mathbf{s}_k \in \mathbb{R}^N$, $k \in \{1, \ldots, m\}$, $\mathbf{t}_k \in \mathbb{R}^N$, $k \in \{1, \ldots, m\}$, $\mathbf{q} \in \mathbb{R}^N$, and $\lambda \in \mathbb{R}$. Moreover, the resulting approximating uncertainty set corresponds to a distortion risk measure if and only if the optimal value λ^* of (15) *satisfies*

$$\lambda^* \leqslant \frac{1}{1 - Nq_{\min}}. \tag{16}$$

Proof. Consider a single inequality constraint $\mathbf{u}'\mathbf{a} \geqslant v$. We have $\mathbf{u}'\mathbf{a} \geqslant v$ for all $\mathbf{a} \in \Pi_{\mathbf{q}}(\mathcal{A})$ if and only if the optimal value of the problem

$$\begin{aligned}&\text{minimize} && \sum_{i=1}^N q_i \sum_{j=1}^N (\mathbf{u}'\mathbf{a}_j) y_{ij}\\ &\text{subject to} && \sum_{i=1}^N y_{ij} = 1 \quad \forall\, j \in \{1, \ldots, N\},\\ &&& \sum_{j=1}^N y_{ij} = 1 \quad \forall\, i \in \{1, \ldots, N\},\\ &&& y_{ij} \geqslant 0, \quad \forall\,(i,j) \in \{1, \ldots, N\}^2,\end{aligned}$$

is no smaller than v. We note that this is optimization over a bounded, nonempty polyhedron, and thus by strong duality the optimal value of this problem equals the optimal value of its dual:

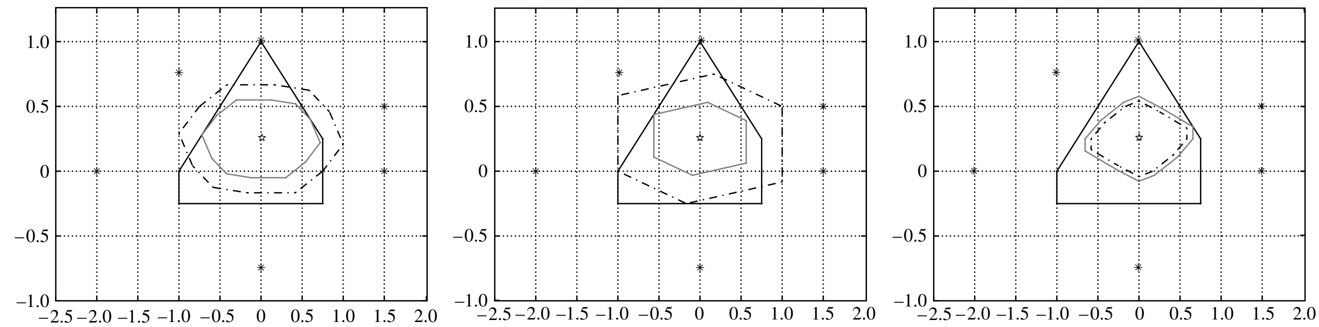
$$\begin{aligned}&\text{maximize} && \mathbf{e}'(\mathbf{s} + \mathbf{t})\\ &\text{subject to} && s_i + t_j \leqslant (\mathbf{u}'\mathbf{a}_j)q_i \quad \forall\,(i,j) \in \{1, \ldots, N\}^2.\end{aligned}$$

It is then easy to see that $\Pi_{\mathbf{q}}(\mathcal{A}) \subseteq \mathcal{U}$ if and only if there exist $\mathbf{s}_1, \ldots, \mathbf{s}_m, \mathbf{t}_1, \ldots, \mathbf{t}_m$ such that

$$\begin{aligned}&\mathbf{e}'(\mathbf{s}_k + \mathbf{t}_k) \geqslant v_k \quad \forall\, k \in \{1, \ldots, m\},\\ &s_{k,i} + t_{k,j} \leqslant (\mathbf{u}_k'\mathbf{a}_j)q_i \quad \forall\,(i,j) \in \{1, \ldots, N\}^2,\ \forall\, k \in \{1, \ldots, m\}.\end{aligned}$$

Now, to find the largest such $\Pi_{\mathbf{q}}(\mathcal{A}) \subseteq \mathcal{U}$ in the $\|\cdot\|_{\hat{\mathbf{q}},\mathcal{A}}$ sense, we can, by Lemma 4.2, set $\mathbf{q} = \lambda\hat{\mathbf{q}} + (1-\lambda)\mathbf{e}_N$ and maximize λ, which leads us to the desired linear program. The bound (16) follows by noting that in order for $\mathbf{q}$ to

Figure 4. Optimal inner approximation for a class of centrally symmetric generators for the example from Figure 3 and an arbitrary polyhedral uncertainty set.



Notes. The dashed line indicates the nonscaled version of $\Pi_{\mathbf{q}}(\mathcal{A})$ in each case, and in solid gray is the tightest inner approximation. In the first two cases, the approximations are "shrunken" and thus correspond to distortion risk measures. In the last case, the optimal approximation is actually larger than $\Pi_{\mathbf{q}}(\mathcal{A})$.

correspond to a distortion risk measure, it must have nonnegative components. □

Figure 4 shows an example of an inner approximation to an arbitrary polyhedral uncertainty set. Note that the resulting optimal measure $\mathbf{q}^*$ from this optimization problem is $\mathbf{q}^* = \lambda^*\hat{\mathbf{q}} + (1-\lambda^*)\mathbf{e}_N$; we can therefore interpret the corresponding distortion risk measure μ in the following way:

$$\begin{aligned}\mu(X) &= \lambda^*\mu_{\hat{\mathbf{q}}}(X) + (1-\lambda^*)\mathbb{E}[-X]\\ &= -\mathbb{E}[X] + \lambda^*(\mu_{\hat{\mathbf{q}}}(X) + \mathbb{E}[X])\\ &= -\mathbb{E}[X] + \lambda^*\sigma_{\hat{\mathbf{q}}}(X),\end{aligned}$$

where $\sigma_{\hat{\mathbf{q}}}(X) \triangleq \sup_{\mathbf{q}\in\Pi_{\hat{\mathbf{q}}}(\{e_1,\ldots,e_N\})} \mathbb{E}_{\mathbf{q}}[-X + \mathbb{E}[X]]$. Problem (15), then, is equivalent to the problem

$$\sup\{\lambda \in \mathbb{R}_+ \colon -\mathbb{E}[X] + \lambda\sigma_{\hat{\mathbf{q}}}(X) \geqslant \mu_{\mathcal{U}}(X)\},$$

where $\mu_{\mathcal{U}}$ is the coherent risk measure corresponding to $\mathcal{U}$ (via Theorem 3.1). This has the flavor of a Sharpe ratio problem with the usual standard deviation measure replaced by the deviation measure defined above.

5. Conclusions

In this paper, we have proposed the framework of coherent risk measures as a starting point for uncertainty set construction for robust linear optimization problems. Our focus has been on the case when the underlying probability space is discrete, which is motivated by sampling considerations. We drew a connection between distortion risk measures and polytopes of a particular structure and explored the geometry and underlying generation of these classes of risk measures and corresponding uncertainty sets.

Some possible future directions are the following:

1. *Multiple constraints and risk over vector spaces*. Obviously, we could apply our framework in a constraint-wise fashion, but there are undoubtedly more sophisticated ways to balance risk among multiple constraints. It would be interesting to explore the resulting implications for robustness under appropriately defined risk measures over more general vector spaces.

2. *More general conic problems*. We have restricted ourselves to linear optimization here. It is not immediately clear how to extend this to more general robust optimization problems.

3. *Integrating information beyond samples*. How can we introduce additional knowledge of the distribution (e.g., prior distributions, moments, or other, "physical" constraints such as no-arbitrage conditions), and what are the implications from the robust perspective?

Endnotes

1. For some more-recent work using samples in the context of chance-constrained optimization, see, e.g., Calafiore and Campi (2005) and Nemirovski and Shapiro (2005).
2. Obviously, when we impose that $|\Omega|$ must be finite and supported by finite elements, this is automatically satisfied.
3. Strictly speaking, Theorem 2.1 only holds if μ satisfies a technical condition known as the *Fatou property*, which means $\mu(X) \geqslant \limsup \mu(X_n)$ for any sequence X_n of bounded random variables converging to X in probability. If $|\Omega|$ is finite, as is our focus throughout the paper, the Fatou property is automatically satisfied. We refer the interested reader to Delbaen (2000) and Föllmer and Schied (2004) for the technical details on dealing with general probability spaces.
4. Again, this interpretation is not exact if the distribution is not atomless, but it is approximately true. We are just trying to provide intuition here.
5. Acerbi (2002) studies such risk measures and calls them "spectral" risk measures.

Acknowledgments

The authors thank two anonymous referees and an anonymous associate editor for a number of insights and observations that substantially improved this manuscript, in exposition as well as in technical rigor.

References

Acerbi, C. 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Banking Finance* **26** 1505–1581.

Acerbi, C., D. Tasche. 2002. On the coherence of expected shortfall. *J. Banking Finance* **26**(7) 1487–1503.

Artzner, P., F. Delbaen, J. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–228.

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

Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming Ser. A* **88** 411–424.

Bertsimas, D., M. Sim. 2006. Tractable approximations to robust conic optimization problems. *Math. Programming* **107**(1–2) 5–36.

Bertsimas, D., G. J. Lauprete, A. Samarov. 2004a. Shortfall as a risk measure: Properties, optimization and applications. *J. Econom. Dynam. Control* **28**(7) 1353–1381.

Bertsimas, D., D. Pachamanova, M. Sim. 2004b. Robust linear optimization under general norms. *Oper. Res. Lett.* **32** 510–516.

Calafiore, G., M. C. Campi. 2005. Uncertain convex programs: Randomized solutions and confidence levels. *Math. Programming Ser. A* **102**(1) 25–46.

Chen, W., M. Sim, J. Sun, C.-P. Teo. 2009. From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Oper. Res.* Forthcoming.

Choquet, G. 1954. Theory of capacities. *Ann. Inst. Fourier* **5** 131–295.

Delbaen, F. 2000. Coherent risk measures on general probability spaces. *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, 1–37.

Dellacherie, C. 1970. Quelque commentaires sur les prolongements de capacités. *Seminaire Probabilities V, Strasbourg, Lecture Notes in Mathematics*, Vol. 191. Springer-Verlag, Berlin.

Denneberg, D. 1990. Distorted probabilities and insurance premium. *Methods Oper. Res.* **52** 21–42.

Dhaene, J., M. Denuit, M. J. Goovaerts, R. Kaas, D. Vyncke. 2002. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Math. Econom.* **31**(2) 133–161.

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

El Ghaoui, L., F. Oustry, H. Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* **9**(1) 33–52.

Fischer, T. 2001. Examples of coherent risk measures depending on one-sided moments. Working paper, http://www.ma.hw.ac.uk/~fischer/papers/.

Föllmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6**(4) 429–447.

Föllmer, H., A. Schied. 2004. *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin.

Huber, P. 1981. *Robust Statistics*. Wiley, New York.

Jouini, E., M. Medded, N. Touzi. 2004. Vector-valued coherent risk measures. *Finance and Stochastics* **8** 531–552.

Kusuoka, S. 2001. On law invariant coherent risk measures. *Adv. Math. Econom.* **3** 83–95.

Leitner, J. 2005. A short note on second order stochastic monotonicity preserving coherent risk measures. *Math. Finance* **15**(4) 649–651.

Natarajan, K., D. Pachamanova, M. Sim. 2009. Constructing risk measures from uncertainty sets. *Oper. Res.* **57**(5) 1129–1141.

Nemirovski, A., A. Shapiro. 2005. Scenario approximations of chance constraints. G. Calafiore, F. Dabbene, eds. *Probabilistic and Randomized Methods for Design Under Uncertainty*. Springer, London, 3–48.

Nemirovski, A., A. Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM J. Optim.* **17**(4) 969–996.

Queyranne, M. 2002. An introduction to submodular functions and optimization. http://www.ima.umn.edu/optimization/seminar/queyranne.pdf.

Rockafellar, R. T., S. P. Uryasev. 2000. Optimization of conditional value-at-risk. *J. Risk* **2** 21–41.

Schmeidler, D. 1986. Integral representation without additivity. *Proc. Amer. Math. Soc.* **97**(2) 255–261.

Soyster, A. L. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper. Res.* **21**(5) 1154–1157.

Wang, S. 2000. A class of distortion operators for pricing financial and insurance risks. *J. Risk and Insurance* **67**(1) 15–36.