

Stochastic linear programming with a distortion risk constraint

Karl Mosler · Pavel Bazovkin

Published online: 24 July 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Coherent distortion risk measures are applied to capture the possible violation of a restriction in linear optimization problems whose parameters are uncertain. Each risk constraint induces an uncertainty set of coefficients, which is proved to be a weighted-mean trimmed region. Thus, given a sample of the coefficients, an uncertainty set is a convex polytope that can be exactly calculated. We construct an efficient geometrical algorithm to solve stochastic linear programs that have a single distortion risk constraint. The algorithm is available as an R-package. The algorithm's asymptotic behavior is also investigated, when the sample is i.i.d. from a general probability distribution. Finally, we present some computational experience.

Keywords Robust optimization · Weighted-mean trimmed regions · Central regions · Coherent risk measure · Spectral risk measure · Mean-risk portfolio

1 Introduction

Uncertainty in the coefficients of a linear program is often handled by probability constraints or, more generally, bounds on a risk measure. The random restrictions are then captured by imposing risk constraints on their violation. Consider the linear program

K. Mosler · P. Bazovkin (✉)
Department of Economic and Social Statistics, Universität zu Köln,
50923 Cologne, Germany
e-mail: bazovkin@statistik.uni-koeln.de

K. Mosler
e-mail: mosler@statistik.uni-koeln.de

$$\mathbf{c}'\mathbf{x} \longrightarrow \min \quad \text{s.t. } \tilde{\mathbf{A}}\mathbf{x} \geq \mathbf{b}, \tag{1}$$

and assume that $\tilde{\mathbf{A}}$ is a stochastic $m \times d$ matrix and $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^d$, $\mathbf{x} \in \mathbb{R}^d$. This is a stochastic linear optimization problem. To handle the stochastic restrictions a *joint risk constraint*,

$$\rho^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \leq 0, \tag{2}$$

may be introduced, where ρ^m is an m -variate risk measure. For instance, with $\rho^m(Y) = \text{Prob}[Y < 0] - \alpha$ the restriction (2) becomes

$$\text{Prob}[\tilde{\mathbf{A}}\mathbf{x} \geq \mathbf{b}] \geq 1 - \alpha, \tag{3}$$

and a usual *chance-constrained linear program* is obtained. Alternatively, the restrictions may be subjected to *separate risk constraints*,

$$\rho^1(\tilde{\mathbf{A}}_j\mathbf{x} - b_j) \leq 0, \quad j = 1 \dots m, \tag{4}$$

with $\tilde{\mathbf{A}}_j$ denoting the j -th row of $\tilde{\mathbf{A}}$. In (4) each restriction is subject to the same bound that limits the risk of violating the condition. A linear program that minimizes $\mathbf{c}'\mathbf{x}$ subject to one of the restrictions, (2) or (4), is called a *risk-constrained stochastic linear program*.

For stochastic linear programs (SLPs) in general and risk-constrained SLPs in particular, the reader is referred to, e.g., [Kall and Mayer \(2010\)](#). What we call a risk measure here is mentioned in that book as a *quality measure*, and useful representations of the corresponding constraints are given. As most of the literature, [Kall and Mayer \(2010\)](#) focus on classes of SLPs with chance constraints that lead to convex programming problems, since these have obvious computational advantages; see also [Prékopa \(1995\)](#). Our choice of the quality measure, besides its generality, enjoys a meaningful interpretation and, as it will be seen, enables the use of convex structures in the problem.

In the case of a single constraint ($m = 1$) notate

$$\rho(\tilde{\mathbf{a}}'\mathbf{x} - b) \leq 0. \tag{5}$$

A practically important example of an SLP with a single risk constraint (5) is the *portfolio selection problem*. Let $\tilde{r}_1, \dots, \tilde{r}_d$ be the return rates on d assets and notate $\tilde{\mathbf{r}} = (\tilde{r}_1, \dots, \tilde{r}_d)'$. A convex combination of the assets' returns is sought, $\tilde{\mathbf{r}}'\mathbf{x} = \sum_{j=1}^d \tilde{r}_j x_j$ that has maximum expectation under a risk constraint and an additional deterministic constraint,

$$\max_{\mathbf{x} \in \mathcal{C}} E[\tilde{\mathbf{r}}'\mathbf{x}], \quad \text{s.t. } \rho(\tilde{\mathbf{r}}'\mathbf{x}) \leq \rho_0, \quad \mathbf{x} \in \mathcal{C}, \tag{6}$$

where ρ is a risk measure, $\rho_0 \in \mathbb{R}$ is a given upper bound of risk (a nonnegative monetary value), and $\mathcal{C} \in \mathbb{R}^d$ is a deterministic set that restricts the vector \mathbf{x} of portfolio weights in some way. For example, if short sales are excluded, \mathcal{C} is the positive orthant in \mathbb{R}^d . The solution \mathbf{x}^* is the optimal investment under the given model. We will see that a solution, if it exists, is, as a rule, finite and unique. In our geometric approach

such a solution corresponds to the intersection of some line and some convex body that both contain the point $E[\tilde{\mathbf{r}}]$.

Regarding the choice of ρ , two special cases are well known. First, let $\rho(\tilde{\mathbf{r}}'\mathbf{x}) = \text{Prob}[\tilde{\mathbf{r}}'\mathbf{x} \leq -v_0]$ and $\rho_0 = \alpha$. Then the optimization problem (6) says: maximize the mean return $E[\tilde{\mathbf{r}}'\mathbf{x}]$ under the restrictions $\mathbf{x} \in \mathcal{C}$ and

$$V@R_\alpha(\tilde{\mathbf{r}}'\mathbf{x}) \leq v_0.$$

That is, the *value at risk* $V@R_\alpha$ of the portfolio return must not exceed the bound v_0 . Second, let

$$\rho(\tilde{\mathbf{r}}'\mathbf{x}) = -\frac{1}{\alpha} \int_0^\alpha Q_{\tilde{\mathbf{r}}'\mathbf{x}}(t) dt, \quad (7)$$

where Q_Z signifies the quantile function of a random variable Z . This means that the measured risk is the *expected shortfall* of portfolio return and shall stay below ρ_0 .

In practice, $\tilde{\mathbf{a}}$ has to be obtained from data. If the solution of the SLP is based on n observed coefficient vectors $\mathbf{a}^1, \dots, \mathbf{a}^n \in \mathbb{R}^d$, the SLP is mentioned as an *empirical risk-constrained SLP*. In other words, we assume that $\tilde{\mathbf{a}}$ follows an *empirical distribution* that gives equal mass $\frac{1}{n}$ to some observed points $\mathbf{a}^1, \dots, \mathbf{a}^n \in \mathbb{R}^d$. Rockafellar and Uryasev (2000) investigate an empirical stochastic program that arises in portfolio choice when the expected shortfall of a portfolio is minimized. They convert the objective into a function that is convex in the decision vector \mathbf{x} and optimize it by standard methods. This approach is commonly used in more recent works of these and other authors on portfolio optimization.

A more complex situation is investigated by Bertsimas and Brown (2009), who discuss the risk-constrained SLP with arbitrary coherent distortion risk measures, which also include expected shortfall. These allow for a sound interpretation in terms of expected utility with distorted probabilities. For the linear restriction an, as it is called, *uncertainty set* is constructed, consisting of all coefficients satisfying the risk constraint. Bertsimas and Brown (2009) discuss the uncertainty set that turns the SLP into a minimax problem, called *robust linear program*; however, they provide no optimal solution of this program there. The uncertainty set is a convex body and, as will be made precise below in this paper, comes out to equal a so-called weighted-mean trimmed region. Natarajan et al. (2009), on the reverse, construct similar risk measures from given polyhedral and conic uncertainty sets. As an extension, Ben-Tal et al. (2010) propose the so-called “soft robustness” model, which, as they show, can be regarded as an LP with the feasible set defined by some *convex* risk measure. Such approaches are also applicable to approximately solving a multi-stage robust convex optimization problem, where the information about the realization of uncertain parameters is adjusted on each stage; see Bertsimas and Goyal (2013).

Pflug (2006) has proposed an iterative algorithm to optimize a portfolio by using distortion functionals; in each step he adds a constraint to the problem and solves it by the simplex method. Meanwhile, many other authors have recently contributed to the development of robust linear programs related to risk-constrained optimization problems: see, e.g., Nemirovski and Shapiro (2006), Ben-Tal et al. (2009) and Chen et al. (2010). For a review of robust linear programs in portfolio optimization the reader

is referred to [Fabozzi et al. \(2010\)](#). There are also attempts to solve this problem by means of robust nonlinear models (see, for instance, [Kawas and Thiele \(2011\)](#)), which, however, are substantially less investigated in the literature than the linear ones. Other applications are surveyed in detail in [Gabrel et al. \(2014\)](#).

In this paper we contribute to this discussion in three respects:

1. The uncertainty set of an SLP under a general coherent distortion risk constraint is shown to be a *weighted-mean trimmed region*, which provides a useful visual and computable characterization of the set.
2. An *algorithm* is constructed that solves the minimax problem over the uncertainty set, hence the SLP.
3. If the data are i.i.d. from a general probability distribution, the uncertainty set and the solution of the SLP are shown to be *consistent estimators* of the uncertainty set and the SLP solution.

The paper is organized as follows. In Sect. 2 constraints on distortion risk measures are discussed. They are characterized by uncertainty sets in parameter, which, in turn, are shown to be weighted-mean trimmed regions (Theorem 1). Based on Theorem 1, which is a core result, we formulate a robust linear program, which is investigated in Sect. 3 and by which the SLP with a distortion risk constraint is solved. Section 4 introduces an algorithm for this program and discusses sensitivity issues of its solution. In Sect. 5 we address the SLP and its solution for generally distributed coefficients and investigate the limit behavior of our algorithm if based on an independent sample of coefficients. Section 6 contains first computational results and concludes the article. The Appendix gathers properties of distortion risk measures, a proof of Theorem 1 and a demonstration (Proposition 5) that the weighted-mean trimmed regions have the *coherency* property.

2 Distortion risk constraints and weighted-mean regions

2.1 Distortion risk measures

A large and versatile subclass of risk measures is the class of distortion risk measures, which have appeared first from ideas in insurance research ([Wang et al. 1997](#)). Again, let Q_Y denote the quantile function of a random variable Y .

Definition 1 (*Distortion risk measure*) Let r be an increasing function $[0, 1] \rightarrow [0, 1]$. The function ρ given by

$$\rho(Y) = - \int_0^1 Q_Y(t) dr(t) \tag{8}$$

is a *distortion risk measure* with *weight generating function* r .

Distortion risk measures are essentially the same as *spectral risk measures* ([Acerbi 2002](#)). Their properties are considered in detail in the Appendix. Here, we concentrate on their *coherency*, as this property is crucial in assessing diversified risks.

A distortion risk measure is coherent if and only if r is concave. For example, with $r(t) = 0$ if $t < \alpha$ and $r(t) = 1$ if $t \geq \alpha$, the *value at risk* $V@R_\alpha(Y) = -Q_Y(\alpha)$

is obtained, which is a noncoherent distortion risk measure. A prominent example of a coherent distortion risk measure is the *expected shortfall*, which is yielded by $r(t) = t/\alpha$ if $t < \alpha$ and $r(t) = 1$ otherwise. This measure is defined as (the negative of) the conditional expectation of Y under the condition that Y does not exceed its α -quantile, $Q_Y(\alpha)$. Roughly speaking, it is the mean of the $\alpha \cdot 100\%$ largest possible losses. Clearly, for a given α , this measure is more conservative than the value at risk, because its value cannot be smaller than $V@R_\alpha(Y)$. Also observe that, for $r(t) = t$, the risk measure (1) becomes the expectation of $-Y$.

A general distortion risk measure $\rho(Y)$ can thus be interpreted as the expectation of $-Y$ with respect to a probability distribution that has been distorted by the function r . In particular, a concave function r distorts the probabilities of lower outcomes of Y in the positive direction (the lower the more) and conversely for higher outcomes (the higher the less). In recent empirical applications, coherent distortion risk measures other than expected shortfall have been used by several authors; see, e.g., Adam et al. (2008) for a comparison of various such measures in portfolio choice.

An equivalent characterization of a distortion risk measure is that it is a law-invariant and comonotonic risk measure; see Kusuoka (2011). ρ is *comonotonic* if

$$\rho(Y + Z) = \rho(Y) + \rho(Z) \text{ for all } Y \text{ and } Z \text{ that are comonotonic,}$$

i.e., that satisfy $(Y(\omega) - Y(\omega'))(Z(\omega) - Z(\omega')) \geq 0$ for every $\omega, \omega' \in \Omega$. If Y has an empirical distribution on $y_1, \dots, y_n \in \mathbb{R}$, the definition (8) of a distortion risk measure specializes to

$$\rho(Y) = - \sum_{i=1}^n q_i y_{[i]}, \tag{9}$$

where $y_{[i]}$ are the values ordered from above and q_i are nonnegative weights adding up to 1. (Observe that $q_i = r(y_{\lfloor \frac{n+1-i}{n} \rfloor}) - r(y_{\lfloor \frac{n-i}{n} \rfloor})$.) Then, the distortion risk measure (9) is coherent if and only if the weights are ordered, i.e., $\mathbf{q} \in \Delta^n_{\leq} := \{\mathbf{q} \in \Delta^n : 0 \leq q_1 \leq \dots \leq q_n\}$.

2.2 Weighted-mean regions as uncertainty sets

If ρ is a coherent distortion risk measure, the uncertainty set \mathcal{U}_ρ has a special geometric structure, which will be explored now to visualize the optimization problem and to provide the basis for an algorithm. We will demonstrate that \mathcal{U}_ρ equals a so-called *weighted-mean trimmed region* (or, equivalently, *WM region*) of the distribution of $\tilde{\mathbf{a}}$.

Given the probability distribution F_Y of a random vector Y in \mathbb{R}^d , WM regions form a nested family of convex compact sets, $\{D_\alpha(F_Y)\}_{\alpha \in [0,1]}$, which are affine equivariant (that is $D_\alpha(F_{AY+b}) = A D_\alpha(F_Y) + b$ for any regular matrix A and $b \in \mathbb{R}^d$). By this, the regions describe the distribution with respect to its location, dispersion and shape. Weighted-mean trimmed regions have been introduced in Dyckerhoff and Mosler (2011) for empirical distributions, and in Dyckerhoff and Mosler (2012) for general ones.

For an empirical distribution on $\mathbf{a}^1, \dots, \mathbf{a}^n \in \mathbb{R}^d$, a weighted-mean trimmed region is a polytope in \mathbb{R}^d and defined as

$$D_{\mathbf{w}_\alpha}(\mathbf{a}^1, \dots, \mathbf{a}^n) = \text{conv} \left\{ \sum_{j=1}^n w_{\alpha,j} \mathbf{a}^{\pi(j)} : \pi \text{ permutation of } \{1, \dots, n\} \right\}. \tag{10}$$

Here, $\mathbf{w}_\alpha = [w_{\alpha,1}, \dots, w_{\alpha,n}]'$ is a vector of ordered weights, i.e., $\mathbf{w}_\alpha \in \Delta_{\leq}^n$, indexed by $0 \leq \alpha \leq 1$ that satisfies the implication: if $\alpha < \beta$, then

$$\sum_{j=1}^k w_{\alpha,j} \leq \sum_{j=1}^k w_{\beta,j}, \quad \forall k = 1, \dots, n. \tag{11}$$

Any such family of *weight vectors* $\{\mathbf{w}_\alpha\}_{0 \leq \alpha \leq 1}$ specifies a particular notion of weighted-mean trimmed regions. There are many types of weighted-mean trimmed regions. They contain well-known trimmed regions like the zonoid regions, the expected convex hull regions and several others. For example,

$$w_{\alpha,j} = \begin{cases} \frac{1}{n\alpha} & \text{if } j > n - \lfloor n\alpha \rfloor, \\ \frac{n\alpha - \lfloor n\alpha \rfloor}{n\alpha} & \text{if } j = n - \lfloor n\alpha \rfloor, \\ 0 & \text{if } j < n - \lfloor n\alpha \rfloor, \end{cases}$$

$0 \leq \alpha \leq 1$ defines the *zonoid regions*. However, some popular types of trimmed regions, such as Mahalanobis or halfspace regions, are no weighted-mean trimmed regions.

Now, we are ready to formulate the key theoretical result of this section, which formalizes the relation between coherent distortion risk measures and their uncertainty sets on one side, and weighted-mean trimmed regions on the other side. This is stated in the following Theorem 1, which is proved in the Appendix. Also in the Appendix, the geometrical properties of WM regions leading to such a relation are considered.

Theorem 1 *If $\tilde{\mathbf{a}}$ has an empirical distribution on $\mathbf{a}^1, \dots, \mathbf{a}^n$ and ρ is a coherent distortion risk measure, then it holds that:*

$$\{\mathbf{x} \in \mathbb{R}^d : \rho(\tilde{\mathbf{a}}' \mathbf{x} - b) \leq 0\} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}' \mathbf{x} \geq b \forall \mathbf{a} \in D_{\mathbf{w}_\alpha}(\mathbf{a}^1, \dots, \mathbf{a}^n)\}. \tag{12}$$

The reader can see that, loosely speaking, Theorem 1 provides a transition from a well-interpreted but hardly manageable risk constraint to an equivalent well-manageable constraint regarding a trimmed region that can be determined by a geometrical algorithm. In fact, note that $D_{\mathbf{w}_\alpha}(\mathbf{a}^1, \dots, \mathbf{a}^n)$ is a d -dimensional *convex polytope*, and thus the convex hull of a finite number of points (its vertices) or, equivalently, a bounded nonempty intersection of a finite number of closed halfspaces (that contain its facets). By this the calculation and representation of such a polytope can be done in two ways: either by its vertices or by its facets. A polytope’s boundary includes a finite number of *faces* having different affine dimensions. Recall that a nonempty intersection of

the boundary with a hyperplane is a *facet* if it has an affine dimension $d - 1$, and a *ridge* if it has an affine dimension $d - 2$. It is called an *edge* if it is a line segment, and a *vertex* if it is a single point. In general, each facet of a polytope in \mathbb{R}^d is itself a polytope of dimension $d - 1$ and has at least d vertices. With WM regions the number of a facet's vertices can vary considerably; it ranges between d and $d!$ (Bazovkin and Mosler 2012a). That is why in calculating WM regions a representation by facets is preferable. In the following Sect. 3 we consider the topic in detail.

3 Solving the SLP with distortion risk constraint

3.1 Calculating the uncertainty set

In the previous section we have shown that the uncertainty set \mathcal{U}_ρ equals the weighted-mean trimmed region $D_{\mathbf{w}_\alpha}$ for a properly chosen weight vector \mathbf{w}_α . Bazovkin and Mosler (2012a) provide an algorithm by which this WM region can be exactly calculated in any dimension d . The results can be visualized in dimensions two and three; for examples, see Fig. 1.

It has been already mentioned that the number of vertices of a facet can be as much as $d!$. Therefore, the representation of a WM region by its vertices appears to be less efficient than that by its facets. In the sequel, we will use the facet representation for solving the SLP.

3.2 The robust linear program

Using the result of Theorem 1, we can write the robust linear program (1) with distortion risk constraint (5) as

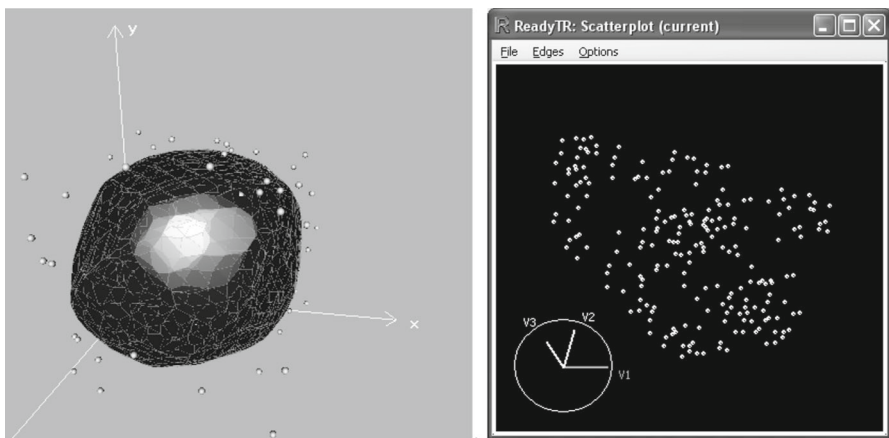


Fig. 1 Visualization of WM regions by the R package *WMRegions*. *Left panel* facets of a three-dimensional region in \mathbb{R}^3 . *Right panel* vertices of a four-dimensional region projected on a subspace of \mathbb{R}^3

$$c'x \longrightarrow \min \quad \text{s.t. } a'x \geq b \text{ for all } a \in \mathcal{U}, \tag{13}$$

where the subscript ρ has been dropped for convenience. In simple words, we get a deterministic linear program with a set of constraints whose coefficients are contained in a set \mathcal{U} that is, according to Theorem 1, a weighted-mean trimmed region. The restriction in (13) is then rewritten as

$$x \in \mathcal{X} = \bigcap_{a \in \mathcal{U}} X_a, \quad X_a = \{x : a'x \geq b\}. \tag{14}$$

Note that \mathcal{X} , as an intersection of a finite number of halfspaces, is a convex polyhedron. Therefore, a linear goal function is to be minimized on a convex polyhedron. Obviously, any optimal solution will lie on the boundary of \mathcal{X} .

3.3 Finding the optimum on the uncertainty set

In constructing an algorithm for the robust linear program, we explore the set \mathcal{X} of feasible solutions and relate it to the uncertainty set \mathcal{U} in the parameter space. It is shown that the space of solutions x and the space of coefficients a are, in some sense, *dual* to each other. The following two lemmas provide the connection between \mathcal{X} and \mathcal{U} . First, we demonstrate that \mathcal{X} is the intersection of those halfspaces whose normals are extreme points of \mathcal{U} .

Lemma 1 *It holds that*

$$\mathcal{X} = \bigcap_{a \in \mathcal{U}} \{x : a'x \geq b\} = \bigcap_{a \in \text{ext } \mathcal{U}} \{x : a'x \geq b\}.$$

Proof We show that $\bigcap_{a \in \text{ext } \mathcal{U}} X_a \subset X_{\mathbf{u}}$ for all $\mathbf{u} \in \mathcal{U}$; then $\bigcap_{a \in \text{ext } \mathcal{U}} X_a \subset \bigcap_{a \in \mathcal{U}} X_a$. The opposite inclusion is obvious. Assume $\mathbf{u} \in \mathcal{U}$. Then, as \mathcal{U} is convex and compact, \mathbf{u} is a convex combination of some points $\mathbf{a}^1, \dots, \mathbf{a}^\ell \in \text{ext } \mathcal{U}$, i.e., $\mathbf{u} = \sum_{i=1}^\ell \lambda_i \mathbf{a}^i$ with $\lambda_i \geq 0$ and $\sum_{i=1}^\ell \lambda_i = 1$, and for any $x \in \bigcap_{a \in \text{ext } \mathcal{U}} X_a$ holds as $x \in X_{\mathbf{a}^j}$ and $\mathbf{a}^j'x \geq b$ for all j ; hence, $\mathbf{u}'x = \sum_{i=1}^\ell \lambda_i \mathbf{a}^i'x \geq b$, that is, $x \in X_{\mathbf{u}}$. \square

Lemma 1 says that each facet of the set \mathcal{X} of feasible solutions corresponds to a vertex of the uncertainty set \mathcal{U} . Hence, it is sufficient to consider the extreme points of the uncertainty set.

As a generalization of Lemma 1, we may prove by recursion on k : each k -dimensional face of the feasible set corresponds to a $(d - k)$ -dimensional face of the uncertainty set in the solution space. This resembles the dual correspondence between convex sets and their *polars* (cf. Rockafellar 1997). However, in contrast to polars, this correspondence of facets is not reflexive.

From Lemma 1, it is immediately seen how the robust optimization problem contrasts with a deterministic problem, where the empirical distribution of $\tilde{\mathbf{a}}$ concentrates at some $\mathbf{a}^0 \in \mathcal{U}$. Observe that the deterministic feasible set \mathcal{X} is just a halfspace, $\mathcal{X}_{\mathbf{a}^0} = \{x : x'\mathbf{a}^0 \geq b\}$. In the general robust case, a halfspace is obtained for each

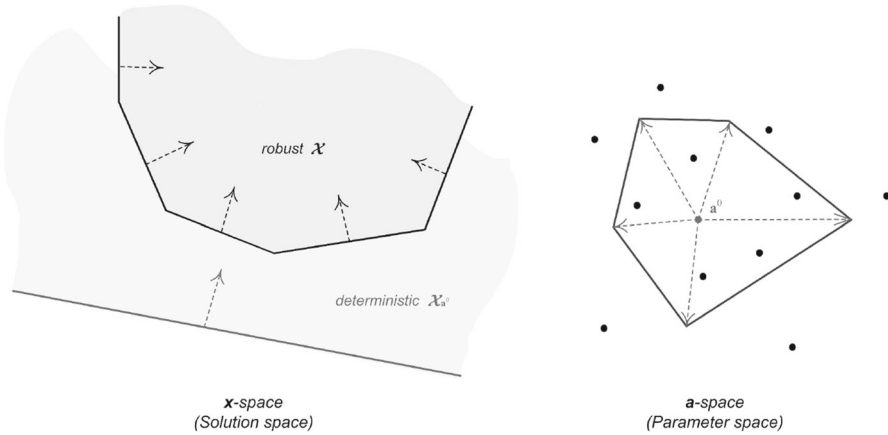


Fig. 2 Deterministic and robust cases: feasible set (*left panel*), uncertainty set (*right panel*)

$\mathbf{a} \in \text{ext}\mathcal{U}$, and the robust feasible set \mathcal{X} is their intersection. The halfspaces are bounded by hyperplanes with normals equal to $\mathbf{a} \in \text{ext}\mathcal{U}$, and their intercepts are all the same and equal to b . Consequently, the robust feasible set \mathcal{X} is always included in the deterministic feasible set $\mathcal{X}_{\mathbf{a}^0}$,

$$\mathcal{X} \subseteq \mathcal{X}_{\mathbf{a}^0} \quad \text{for any } \mathbf{a}^0 \in \mathcal{U}.$$

Moreover, the two feasible sets cannot be equal unless each element of \mathcal{U} is a scalar multiple of \mathbf{a}^0 with a factor greater than one, $\mathcal{U} \subseteq \{\mathbf{a} : \mathbf{a} = \lambda\mathbf{a}^0, \lambda > 1\}$. Consequently, the minimum value of the robust stochastic LP cannot be smaller than the value of an LP with any deterministic parameter \mathbf{a}^0 chosen from the uncertainty set. Fig. 2 (left panel) illustrates how a deterministic feasible set in dimension two compares to a general robust one: the line that bounds the halfspace $\mathcal{X}_{\mathbf{a}^0}$ ‘folds’ into a piecewise linear curve delimiting \mathcal{X} . In turn, Fig. 2 (right panel) demonstrates the same relation between the uncertainty sets in the parameter space: the deterministic uncertainty set, which is a singleton \mathbf{a}^0 , ‘enlarges’ into a nondegenerate uncertainty set containing \mathbf{a}^0 .

Let

$$U_{\mathbf{x}} = \{\mathbf{a} \in \mathbb{R}^d : \mathbf{a}'\mathbf{x} \geq b\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Lemma 2 *It holds that*

$$\mathcal{U} \subset \bigcap_{\mathbf{x} \in \mathcal{X}} U_{\mathbf{x}} \subset \bigcap_{\mathbf{x} \in \text{ext } \mathcal{X}} U_{\mathbf{x}}.$$

Moreover, each vertex $\mathbf{x} \in \text{ext } \mathcal{X}$ corresponds to a facet of \mathcal{U} .

Proof By Lemma 1, we have $\mathbf{x} \in \mathcal{X} \Leftrightarrow \mathbf{a}'\mathbf{x} \geq b$ for all $\mathbf{a} \in \mathcal{U}$. Now, let $\mathbf{a} \in \mathcal{U}$; then for any $\mathbf{x} \in \mathcal{X}$ it holds that $\mathbf{a}'\mathbf{x} \geq b$; hence, $\mathbf{a} \in U_{\mathbf{x}}$. Conclude $\mathcal{U} \subset \bigcap_{\mathbf{x} \in \mathcal{X}} U_{\mathbf{x}}$. Further, it is clear that an extreme point $\mathbf{x} \in \text{ext } \mathcal{X}$ yields a facet of \mathcal{U} . \square

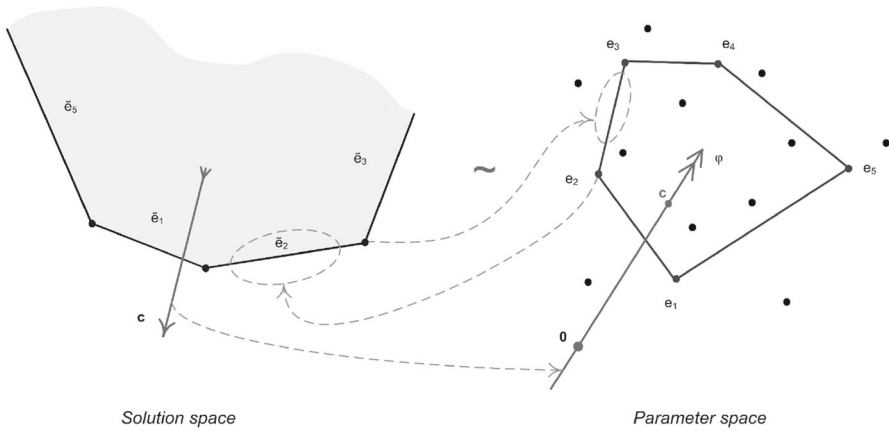


Fig. 3 Duality between spaces

Remark. While \mathcal{U} is always compact, \mathcal{X} is in general not. Therefore, neither inclusion holds with equality.

According to Lemma 1, we could now reformulate the robust LP (13), basing it on the constraints generated by the vertices of \mathcal{U} :

$$c'x \longrightarrow \min \quad \text{s.t. } a'x \geq b \text{ for all } a \in \text{ext}\mathcal{U},$$

and then apply the ordinary simplex method to it. However, usually WM regions have a very large number of vertices, because even a single facet can have up to $d!$ ones. Bazovkin and Mosler (2012a) have shown this number to lie between $\mathcal{O}(n^d)$ and $\mathcal{O}(\frac{n^{2d}}{2^d})$ depending on the type of the WM region. This obviously makes the basic straightforward approach almost inapplicable here. On the other side, WM regions can be efficiently determined by their facets. In our algorithm, we pursue still another way to find the optimal solution, namely searching it even without explicit construction of \mathcal{X} and taking advantage of the facet representation of \mathcal{U} .

To manage this task let us consider the goal function $c'x$. In the solution space the optimization vector c defines a direction, which can be also determined by a set of hyperplanes orthogonal to this direction. Clearly, all these hyperplanes are parallel and their normals are some multiples of c . For example, in dimension two for $c = (2.1, 1.4)'$ and $b = 5$, the hyperplanes $\{x : (2.1, 1.4)' \cdot x = 5\}$ and $\{x : (4.2, 2.8)' \cdot x = 5\}$ belong to the set. Recall that we have fixed the intercept of all hyperplanes in the solution space at some b , and thus can distinguish them by their normals only. In the parameter space, each of these hyperplane corresponds to the point c multiplied with a proper scaling factor. Hence, the image of all the hyperplanes in the parameter space is obtained by moving a point along a straight ray φ that starts at the origin and contains c , as shown on Fig. 3.

One of the hyperplanes touches \mathcal{X} at the optimum. All others either intersect the interior of \mathcal{X} or do not intersect it at all. This means that the touching hyperplane corresponds to a point lying on the surface of \mathcal{U} . Therefore, we have to search the

intersection of \mathcal{U} with the ray φ . Then, because of Lemma 2, the intersected facet of \mathcal{U} corresponds to the optimal vertex of \mathcal{X} .

Note that finding the intersection of a line and a polyhedron in \mathbb{R}^3 is an important problem in computer graphics (cf. Kay and Kajiya 1986). The same principle is employed for a general dimension d . The uncertainty set \mathcal{U} is the finite intersection of halfspaces \mathcal{H}_j , $j = 1 \dots J$, each being defined by a hyperplane H_j with normal \mathbf{n}_j pointing into \mathcal{H}_j and intercept d_j .

Consider some point \mathbf{u} on the ray φ that is not in \mathcal{U} . Compute $\frac{d_j}{\mathbf{u}'\mathbf{n}_j}$ for all halfspaces \mathcal{H}_j that do not include \mathbf{u} , i.e., where $(\mathbf{u}'\mathbf{n}_j - d_j) < 0$ holds. (In other words, H_j is visible from \mathbf{u} .) Find j_* at which this value is the largest. Recall that moving a point \mathbf{u} along φ is equivalent to multiplying \mathbf{u} by some constant. The farthest move is given by the largest constant. The optimal solution \mathbf{x}^* of the robust SLP has to satisfy $\mathbf{a}'\mathbf{x}^* \geq b$, which is equivalent to

$$\mathbf{a}' \left(\frac{d_{j_*}}{b} \mathbf{x}^* \right) \geq d_{j_*}.$$

Hence, to obtain \mathbf{x}^* , the normal \mathbf{n}_{j_*} has to be scaled with the constant $\frac{b}{d_{j_*}}$,

$$\mathbf{x}^* = \frac{b}{d_{j_*}} \mathbf{n}_{j_*}. \tag{15}$$

Besides the regular situation described above, two special cases can arise:

1. There is no facet visible from the origin. This means that no solution is obtained.
2. φ does not intersect \mathcal{U} . Then the whole procedure is repeated with the opposite ray $-\varphi$. If this still gives no intersection, an infinite solution exists.

Finally, we should point out that not the whole polytope \mathcal{U} needs to be calculated, but only that part of it that intersects the ray φ . In searching for the optimum, not all F facets need to be checked, but only a subset of the surface where the intersection will happen. Such a filtration makes the procedure more efficient. Next, we show how to select this subset.

Let \mathbf{x}^* be an optimal solution of the robust SLP. A subset \mathcal{U}_{eff} of \mathcal{U} will be mentioned as an efficient parameter set if

- \mathbf{x}^* remains the solution for $\bigcap_{\mathbf{a} \in \mathcal{U}_{\text{eff}}} \{\mathbf{x} : \mathbf{a}'\mathbf{x} \geq b\} \supset \mathcal{X}$ and
- $\mathbf{a}, \mathbf{d} \in \mathcal{U}_{\text{eff}}, \mathbf{a}'\mathbf{x} \geq b \Rightarrow \mathbf{d}'\mathbf{x} \geq b, \forall \mathbf{x}$ implies $\mathbf{a} = \mathbf{d}$.

That is to say, \mathcal{U}_{eff} is the minimal subset of \mathcal{U} containing all facets that can be optimal for some \mathbf{c} .

Proposition 1 \mathcal{U}_{eff} is the union of all facets of \mathcal{U} for which $d_j \geq 0$ holds.

In other words, an efficient parameter set \mathcal{U}_{eff} consists of that part of the surface of \mathcal{U} that is visible from the origin $\mathbf{0}$. The proof is obvious.

To visualize the efficient parameter set we use the augmented uncertainty set, which is defined as

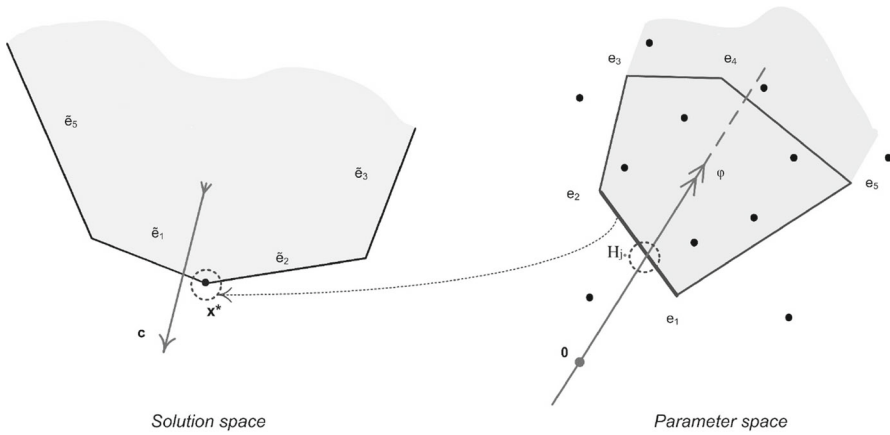


Fig. 4 Finding the optimal solution on the uncertainty set

$$\{\mathbf{a} : \mathbf{a} = \lambda \mathbf{a}^*, \lambda > 1, \mathbf{a}^* \in \mathcal{U}_{\text{eff}}\}.$$

It includes all parameters that are dominated by \mathcal{U}_{eff} ; see the shaded area in the right panel of Fig. 4.

So far, we have assumed that $b > 0$. It is easy to show that with $b < 0$ we have to construct the intersection of φ with the part of the surface of \mathcal{U} that is *invisible* from the origin $\mathbf{0}$, which is $\tilde{\mathcal{U}}_{\text{eff}}$ in this case. In the sense of Proposition 1, $\tilde{\mathcal{U}}_{\text{eff}}$ contains all facets of \mathcal{U} with $d_j \leq 0$. Obviously, $\tilde{\mathcal{U}}_{\text{eff}}$ is always nonempty in this case, which, in turn, means that the existence of a solution is guaranteed. However, the solution can be infinite if φ does not intersect $\tilde{\mathcal{U}}_{\text{eff}}$.

By the same approach, a robust linear program of maximum type can be solved. Note that

$$\mathbf{c}'\mathbf{x} \longrightarrow \max \quad \text{s.t.} \mathbf{a}'\mathbf{x} \leq b \quad \text{for all } \mathbf{a} \in \mathcal{U}, \tag{16}$$

can be rewritten as

$$(-\mathbf{c})'\mathbf{x} \longrightarrow \min \quad \text{s.t. } (-\mathbf{a})'\mathbf{x} \geq -b \quad \text{for all } \mathbf{a} \in \mathcal{U}. \tag{17}$$

Clearly, (17) is equivalent to (13).

4 The algorithm

In this section, the algorithmic procedure for obtaining an optimal solution to (13) is given in detail.

Input:

- a vector $\mathbf{c} \in \mathbb{R}^d$ of coefficients of the goal function,
- n observations $\{\mathbf{a}^1, \dots, \mathbf{a}^n\} \subset \mathbb{R}^d$ of coefficient vectors of the restriction,
- a right-hand side $b \in \mathbb{R}$ of the restriction,
- a distortion risk measure ρ (defined either by name or by a weight vector).

Output:

- the uncertainty set \mathcal{U} of parameters given by
 - facets (i.e., normals and intercepts),
 - vertices,
- the optimal solution \mathbf{x}^* of the robust LP and its value $\mathbf{c}'\mathbf{x}^*$.

Steps:

- A. Calculate the subset $\mathcal{U}_{\text{eff}} \subset \mathcal{U}$ consisting of facets $\{(\mathbf{n}_j, d_j)\}_{j \in J}$.
- B. Create a line φ passing through the origin $\mathbf{0}$ and \mathbf{c} .
- C. Search for a facet H_{j^*} of \mathcal{U}_{eff} that is intersected by φ :
 - a. Select a subset $\mathcal{U}_{\text{sel}} \subseteq \mathcal{U}_{\text{eff}}$ of facets. This may be either \mathcal{U}_{eff} itself or its part where the intersection is expected; $\mathcal{U}_{\text{sel}} = \{(\mathbf{n}_j, d_j) : j \in J_{\text{sel}}\}$. For example, we can search for the best solution on a pre-given subset of parameters. Another possible filtration consists in iterative transitions to facets having better criterion values.
 - b. Take a point $\mathbf{u} = \lambda \mathbf{c}, \lambda \geq 0$, outside the augmented uncertainty set. Find the $j^* = \arg \max_j \{\lambda_j = \frac{d_j}{\mathbf{u}'\mathbf{n}_j} : \lambda_j > 0\}_{j \in J_{\text{sel}} \subseteq J}$. For the case $b < 0$, just replace $\arg \max$ with $\arg \min$.
 - I. If φ does not intersect \mathcal{U}_{eff} , then the solution is *infinite*. If $b > 0$, then repeat C.b. with the opposite ray $-\varphi$; else stop.
 - II. If in the case $b > 0$ the inner part of \mathcal{U} contains the origin, then *no solution* exists; stop.
 - c. $\mathbf{x}^* = \frac{b}{d_{j^*}} \mathbf{n}_{j^*}$ is the optimal solution of the robust LP.

In fact, the line φ consists of points that correspond to hyperplanes whose normal is the vector \mathbf{c} in the dual space. One part of φ is dominated by points from \mathcal{U}_{eff} , while the other is not (which results from Proposition 1). The crossing point \mathbf{a}^* defines the hyperplane that touches the feasible set at the optimum as its dual.

Moreover, a typical nonnegativity side constraint $\mathbf{x} \geq \mathbf{0}$ can easily be accounted for in the algorithm. In considering this, the search for facets has just to be restricted to those having nonnegative normals.

To solve the portfolio selection problem (6) with the algorithm, we treat the realizations of the vector $-\bar{\mathbf{r}}$ of losses rates as $\{\mathbf{a}^1, \dots, \mathbf{a}^n\}$, and minimize $\mathbf{c}'\mathbf{x}$ with $\mathbf{c} = \frac{1}{n} \sum_{i=1}^n \mathbf{a}^i$. This corresponds to transforming the maximizing SLP by (17) and running the procedure outlined above. Note that both φ and \mathcal{U} contain the point $\frac{1}{n} \sum_{i=1}^n \mathbf{a}^i$, that is, they always intersect, which, in turn, guarantees the existence of a finite solution. To meet a unit budget constraint, the solution \mathbf{x}^* is finally scaled down to $\sum_{j=1}^d x_j^* = 1$. Recall that the risk measure is, by definition, scale equivariant.

4.1 Sensitivity and complexity issues

Next, we discuss how the robust SLP and its optimal solution behave when the data $\{\mathbf{a}^1, \dots, \mathbf{a}^n\}$ on the coefficients are slightly changed. From (18) it is immediately seen that the support function $h_{\mathcal{U}}$ of the uncertainty set is continuous in the data \mathbf{a}^j as well as in the weight vector \mathbf{w}_α . (Note that the support function $h_{\mathcal{U}}$ is even *uniformly*

continuous in $\mathbf{a}^1, \dots, \mathbf{a}^n$ and \mathbf{w}_α , which is tantamount to saying that the uncertainty set \mathcal{U} is Hausdorff continuous in the data and the risk weights.) Consequently, a slight perturbation of the data will only slightly change the value of the support function of \mathcal{U} , which is a practically useful result regarding the sensitivity of the uncertainty set with respect to the data. The same is true for a small change in the weights of the risk measure.

We conclude that the point \mathbf{a}^{j^*} where the line through the origin and \mathbf{c} cuts \mathcal{U} depends continuously on the data and the weights. However this is not true for the optimal solution \mathbf{x}^* , which may ‘jump’ when the cutting point moves from one facet of \mathcal{U} to a neighboring one.

The theoretical complexity in time of finding the solution is compounded from the complexity of one transition to the next facet and by the whole number of such transitions until the sought-for facet is achieved. Bazovkin and Mosler (2012a) have shown that the transition has a complexity of $\mathcal{O}(d^2n)$. In turn, in the same paper the number of facets $N(n, d)$ of a WM region is shown to lie between $\mathcal{O}(n^d)$ and $\mathcal{O}(n^{2d})$ depending on the type of the WM region. Thus, it is easily seen that an average number of facets in a facets chain of a fixed length is defined by the density of facets on the region’s surface, $\sqrt[d]{N(n, d)}$, and is estimated by a function between $\mathcal{O}(n)$ and $\mathcal{O}(n^2)$. The overall complexity is then $\mathcal{O}(d^2n^2)$ up to $\mathcal{O}(d^2n^3)$. Notice that the lower complexity is achieved for zonoid regions, namely when the expected shortfall is used for the risk measure.

4.2 Ordered sensitivity analysis

Alternative uncertainty sets that are ordered by inclusion can be also compared. From Lemma 1 it is clear that the respective sets of feasible solutions are then ordered in the reverse direction; see, e.g., Fig. 5. In particular, we can consider the robust LP for two alternative distortion risk measures based on weight vectors \mathbf{w}_α and \mathbf{w}_β , respectively, that satisfy the monotonicity restriction (11). Then the resulting uncertainty sets are nested, $\mathcal{U}_\beta \subset \mathcal{U}_\alpha$, and so are, conversely, the feasible sets, $\mathcal{X}_\beta \supset \mathcal{X}_\alpha$. This is a useful approach for visualizing the sensitivity of the robust LP against changes in risk evaluation.

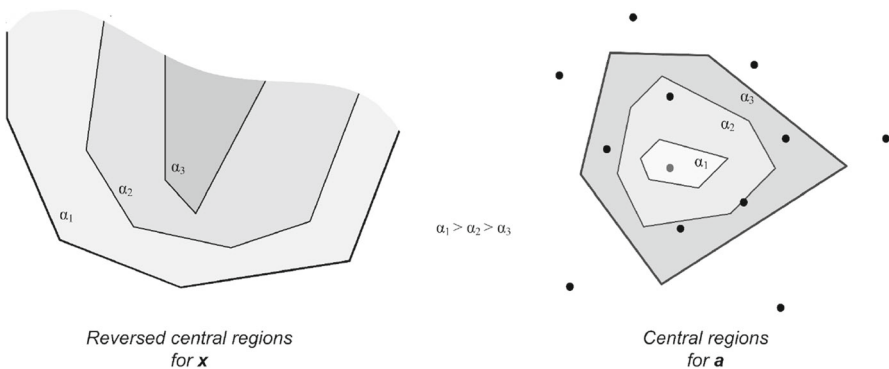


Fig. 5 Example of the ‘reversed’ central regions in the dimension 2

5 Robust SLP for generally distributed coefficients

So far, an SLP (1) has been considered where the coefficient vector $\tilde{\mathbf{a}}$ follows an empirical distribution. It has been solved on the basis of n observations $\{\mathbf{a}^1, \dots, \mathbf{a}^n\}$. In this section the SLP is addressed with a general probability distribution P of $\tilde{\mathbf{a}}$. We formulate the robust SLP in the general case and demonstrate that the solution of this SLP can be consistently estimated by random sampling from P .

Consider a distortion risk measure ρ (8) that measures the risk of a general random variable Y and has weight generating function r , $\rho(Y) = -\int_0^1 Q_Y(t)dr(t)$. As in Sect. 2.2, a convex compact \mathcal{U} in \mathbb{R}^d is constructed through its support function $h_{\mathcal{U}}$,

$$h_{\mathcal{U}}(\mathbf{p}) = \int_0^1 Q_{\mathbf{p}\tilde{\mathbf{a}}}(t) dr(t).$$

Now, let a sequence $(\tilde{\mathbf{a}}^n)_{n \in \mathbb{N}}$ of independent random vectors be given that are identically distributed with P , and consider the sequence of random uncertainty sets \mathcal{U}_n based on $\tilde{\mathbf{a}}^1, \dots, \tilde{\mathbf{a}}^n$. Dyckerhoff and Mosler (2011) have shown:

Proposition 2 (Dyckerhoff and Mosler 2011) \mathcal{U}_n converges to \mathcal{U} almost surely in the Hausdorff sense.

The proposition implies that by drawing an independent sample of $\tilde{\mathbf{a}}$ and solving the robust LP based on the observed empirical distribution, a consistent estimate of the uncertainty set \mathcal{U} is obtained. Moreover, the cutting point \mathbf{a}^{j^*} , where the line through the origin and \mathbf{c} hits the uncertainty set, is consistently estimated by our algorithm. If an ambiguous solution is possible, in particular for a discretely distributed $\tilde{\mathbf{a}}$, the algorithm calculates one of the available solutions consistently. In fact, the optimal solution \mathbf{x}^* may perform a jump when \mathbf{a}^{j^*} moves from one facet of \mathcal{U} to a neighboring one; however the algorithm for determining \mathbf{x}^* selects always a unique facet containing \mathbf{a}^{j^*} .

6 Concluding remarks

A stochastic linear program (SLP) has been investigated, where the coefficients of the linear restrictions are random. Distortion risk constraints are imposed on the random restrictions and an equivalent robust SLP is modeled, whose worst-case solution is searched over an uncertainty set of coefficients. If the risk is measured by a general coherent distortion risk measure, the uncertainty set of a restriction has been shown to be a *weighted-mean trimmed region*. This provides a comprehensive visual and computable characterization of the uncertainty set. An *algorithm* has been developed that solves the robust SLP under a single stochastic constraint, given a set of observations. It is available as an R-package *StochaTR* (Bazovkin and Mosler 2012b). Moreover, when the data are generated by an infinite i.i.d. sample, the limit behavior of the solution was investigated. The algorithm allows the introduction of additional deterministic constraints, in particular, those regarding nonnegativity.

Table 1 Running times of *StochaTR* for different n and d (in seconds)

$d \backslash n$	1,000	2,000	3,000	4,000	5,000	10,000	15,000	20,000	25,000
3	0.3	1.14	1.76	2.92	3.41	6.18	12.61	15.06	47.54
4	0.66	2.21	3.47	4.48	4.27	7.68	16.97	20.04	
5	1.85	3.09	5.68	9.28	11.03	13.52	27.34	54.86	
6	2.08	4.41	5.62	14.99	18.73	25.07	46.88		
7	2.16	6.22	13.3	25.44	28.56	52.33			
8	4.18	9.78	20.18	31.82	34.23				
9	5.18	14.75	24.11	35.94	61.14				
10	6.17	16.97	33.82	42.11	67.06				

Table 2 Running times of *StochaTR* for different n and α (in s); in parentheses running times of [Rockafellar and Uryasev \(2000\)](#)

$\alpha \backslash n$	1,000	5,000	10,000	15,000	20,000	25,000
0.10	1.1 (<5)	7.2 (6)	23.7 (20)	46	56.3 (45)	74.4
0.05	0.5 (<5)	4.7 (6)	14.0 (12)	20.0	39.8 (40)	53.2
0.01	0.3 (<5)	2.3 (6)	3.8 (6)	7.9	22.1 (50)	38.5

Table 1 reports simulated running times (in seconds) of the R-package for the 5 %-level expected shortfall and different d and n . The data are simulated by mixing the uniform distribution on a d -dimensional parallelogram with a multivariate Gaussian distribution. In light of the table, the complexity seems to grow with d and n slower than $\mathcal{O}(d^2n^2)$.

Besides this, we contrast our new procedure with the seminal approach of [Rockafellar and Uryasev \(2000\)](#), who solve the portfolio problem by optimizing the expected shortfall with a simplex-based method. In illustrating their method, they simulate three-dimensional normal returns having specified expectations and covariance matrices. We have applied our package to likewise simulated data on a 1.73GHz single-core CPU with at most 1.5 gigabytes of memory available. The computational times are exhibited in Table 2. For a comparison, some cells also contain a second value that corresponds to the [Rockafellar and Uryasev \(2000\)](#) procedure and is taken from Table 5 there. Our algorithm usually needs some dozens of iterations only, which are substantially fewer than those of the algorithm of [Rockafellar and Uryasev \(2000\)](#). Also, in contrast to the latter, where the resulting portfolio can vary between (0.42, 0.13, 0.45) for $n = 1,000$ and (0.64, 0.04, 0.32) for $n = 5,000$, we get a *stable* optimal portfolio. Our solution averages at (0.36, 0.15, 0.49), which has approximately the same V@R and expected shortfall as that in the compared study but yields a *better* value of the *expected return*. Note that the computational times reported in [Rockafellar and Uryasev \(2000\)](#) do not differ much from ours.

Finally, our approach turns out to be very flexible. In particular, nonsample information can be introduced into the procedure in an interactive way by explicitly changing and modifying the uncertainty set. More research is needed in extending the algorithm

to solve SLPs with multiple constraints (2). Also, procedures that allow for a stochastic right-hand side in the constraints and random coefficients in the goal function are yet to be explored.

Acknowledgments Pavel Bazovkin was partly supported by a grant from the German Research Foundation (DFG).

Appendix

In this Appendix we first discuss the principal properties of distortion risk measures and the characterization of a risk bound by an uncertainty set. Then, we describe WM regions fully by their projections on lines. Based on these notions, next, Theorem 1 is proved. Finally, the coherency property of WM regions is demonstrated.

Properties of distortion risk measures

Let us consider a probability space $\langle \Omega, \mathcal{F}, P \rangle$ and a set \mathcal{R} of random variables (e.g., returns of portfolios). A function $\rho : \mathcal{R} \rightarrow \mathbb{R}$ is a monetary *risk measure* if for $Y, Z \in \mathcal{R}$ it holds:

1. *Monotonicity*: If Y is pointwise larger than Z , $Y \geq Z$, then it has less risk, $\rho(Y) \leq \rho(Z)$.
2. *Translation invariance*: $\rho(Y + \gamma) = \rho(Y) - \gamma$ for all $\gamma \in \mathbb{R}$.

A risk measure is law-invariant if it holds additionally:

3. *Law-invariance*: If Y and Z have the same distribution, $P_Y = P_Z$, then $\rho(Y) = \rho(Z)$.

A law-invariant risk measure ρ is *coherent* if it is, in addition, positive homogeneous and subadditive,

4. *Positive homogeneity*: $\rho(\lambda Y) = \lambda \rho(Y)$ for all $\lambda \geq 0$,
5. *Subadditivity*: $\rho(Y + Z) \leq \rho(Y) + \rho(Z)$ for all $Y, Z \in \mathcal{R}$.

The last two restrictions imply that *diversification* is encouraged by the risk measure—a crucial property in risk management. For the theory of risk measures, see, e.g., Föllmer and Schied (2004). Loosely speaking, diversification is a natural mechanism of reducing risk by ‘not putting all the eggs into one basket’.

Note that a distortion risk measure (8) satisfies the above properties 1 to 3, and hence is a law-invariant risk measure.

A function $\rho : \mathcal{R} \rightarrow \mathbb{R}$ is said to satisfy the *Fatou property* if for any bounded sequence converging pointwise to Y , $\liminf_{n \rightarrow \infty} \rho(Y_n) \geq \rho(Y)$ holds. With the notion of coherent risk measures, we reformulate a fundamental representation result of Huber (1981):

Proposition 3 ρ is a coherent risk measure satisfying the Fatou property if and only if there exists a family \mathbb{Q} of probability measures that are dominated by P (i.e., $P(S) = 0 \Rightarrow Q(S) = 0$ for any $S \in \mathcal{F}$ and $Q \in \mathbb{Q}$) such that for all $Y \in \mathcal{R}$

$$\rho(Y) = \sup_{Q \in \mathbb{Q}} E_Q(-Y).$$

We say that the family \mathbb{Q} generates ρ . In particular, let $(\Omega, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}^d)$ and P be the probability distribution of a random vector $\tilde{\mathbf{a}}$. Huber’s Theorem implies that for any coherent risk measure ρ there exists a family \mathbb{G} of P -dominated probabilities on \mathcal{B}^d , so that

$$\begin{aligned} \rho(\tilde{\mathbf{a}}' \mathbf{x} - b) \leq 0 &\Leftrightarrow \rho(\tilde{\mathbf{a}}' \mathbf{x}) \leq -b \\ &\Leftrightarrow \inf_{G \in \mathbb{G}} E_G(\tilde{\mathbf{a}}' \mathbf{x}) \geq b \\ &\Leftrightarrow E_G(\tilde{\mathbf{a}}' \mathbf{x}) \geq b \text{ for all } G \in \mathbb{G}. \end{aligned}$$

Let us denote the unit simplex in \mathbb{R}^n by Δ^n ,

$$\Delta^n = \circ \left\{ \mathbf{q} \in \mathbb{R}^n : \sum_{k=1}^n q_k = 1, q_k \geq 0 \ \forall k \right\}.$$

Then, if $\tilde{\mathbf{a}}$ has an empirical distribution on n given points in \mathbb{R}^d , any subset \mathcal{Q} of Δ^n corresponds to a family of P -dominated probabilities and thus defines a coherent risk measure ρ . As an immediate consequence of Huber’s theorem, an equivalent characterization of the risk constraint is obtained (see also [Bertsimas and Brown 2009](#)):

Proposition 4 *Let $\rho : \mathcal{R} \rightarrow \mathbb{R}$ be a coherent risk measure and let $\tilde{\mathbf{a}}$ have an empirical distribution on $\mathbf{a}^1, \dots, \mathbf{a}^n \in \mathbb{R}^d$. Then there exists some $\mathcal{Q}_\rho \subset \Delta^n$ such that*

$$\begin{aligned} \rho(\tilde{\mathbf{a}}' \mathbf{x} - b) \leq 0 &\Leftrightarrow \mathbf{a}' \mathbf{x} \geq b \text{ for all} \\ \mathbf{a} \in \mathcal{U}_\rho &:= \text{conv} \left\{ \mathbf{a} \in \mathbb{R}^d : \mathbf{a} = \sum_{i=1}^n q_i \mathbf{a}^i, (q_1, \dots, q_n) \in \mathcal{Q}_\rho \right\}. \end{aligned}$$

Here, $\text{conv}(W)$ denotes the convex closure of a set W . Proposition 4 says that a deterministic restriction $\mathbf{a}' \mathbf{x} \geq b$ holding uniformly for all \mathbf{a} in the uncertainty set \mathcal{U}_ρ is equivalent to the risk constraint (5) on the stochastic restriction.

Characterization of WM regions

A WM region (10) is characterized by its projections on lines. Note that each $\mathbf{p} \in S^{d-1}$, where S^{d-1} is the $(d - 1)$ -variate unit sphere, yields a projection of the data $\mathbf{a}^1, \dots, \mathbf{a}^n$ on the line generated by \mathbf{p} and thus induces a permutation $\pi_{\mathbf{p}}$ of the data,

$$\mathbf{p}' \mathbf{a}^{\pi_{\mathbf{p}}(1)} \leq \mathbf{p}' \mathbf{a}^{\pi_{\mathbf{p}}(2)} \leq \dots \leq \mathbf{p}' \mathbf{a}^{\pi_{\mathbf{p}}(n)}.$$

The permutation is not necessarily unique and let $H(\mathbf{a}^1, \dots, \mathbf{a}^n)$ denote the set of all directions $\mathbf{p} \in S^{d-1}$ that induce a nonunique permutation $\pi_{\mathbf{p}}$. Recall that the support function of a closed convex set K is defined as $h(\mathbf{p}) = \sup\{\mathbf{p}' \mathbf{x} : \mathbf{x} \in K\}$,

$\mathbf{p} \in S^{d-1}$. Given a convex polytope K , an *extreme point* of K is the unique solution of $\max\{\mathbf{p}'\mathbf{x} : \mathbf{x} \in K\}$ for some $\mathbf{p} \in S^{d-1}$. Dyckerhoff and Mosler (2011) have shown that the support function h_α of $D_{\mathbf{w}_\alpha} = D_{\mathbf{w}_\alpha}(\mathbf{a}^1, \dots, \mathbf{a}^n)$ amounts to

$$h_\alpha(\mathbf{p}) = \sum_{j=1}^n w_{\alpha,j} \mathbf{p}' \mathbf{a}^{\pi_{\mathbf{p}}(j)}, \quad \mathbf{p} \in S^{d-1}. \tag{18}$$

It follows that, whenever $\pi_{\mathbf{p}}$ is unique, the polytope $D_{\mathbf{w}_\alpha}$ has an extreme point in direction \mathbf{p} , which is given by

$$\sum_{j=1}^n w_{\alpha,j} \mathbf{a}^{\pi_{\mathbf{p}}(j)} = \sum_{i=1}^n w_{\alpha,\pi_{\mathbf{p}}^{-1}(i)} \mathbf{a}^i, \quad \mathbf{p} \in S^{d-1} \setminus H(\mathbf{a}^1, \dots, \mathbf{a}^n). \tag{19}$$

Proof of Theorem 1

Now, we are moving to the proof of Theorem 1. From (9) and (19) we can see that, with $q_i = w_{\alpha,\pi_{\mathbf{p}}^{-1}(i)}$ and $y_{[i]} = -\mathbf{p}'\mathbf{a}^i$, the extreme point of the projection of $D_{\mathbf{w}_\alpha}$ on the \mathbf{p} -line is obtained by applying a \mathbf{q} -distortion risk measure to the projected data points. Now, setting

$$\mathcal{Q}_\rho = \left\{ \mathbf{q} \in \Delta^n : \mathbf{q} = (w_{\alpha,\pi_{\mathbf{p}}^{-1}(1)}, \dots, w_{\alpha,\pi_{\mathbf{p}}^{-1}(n)}), \mathbf{p} \in S^{d-1} \setminus H(\mathbf{a}^1, \dots, \mathbf{a}^n) \right\}$$

and having $\mathcal{U}_\rho = \text{conv}\{\mathbf{a} : \mathbf{a} = \sum_{i=1}^n q_i \mathbf{a}^i, \mathbf{q} \in \mathcal{Q}_\rho\}$ according to Proposition 4, we get that all extreme points of $D_{\mathbf{w}_\alpha}$ are in \mathcal{U}_ρ ; hence $D_{\mathbf{w}_\alpha} \subseteq \mathcal{U}_\rho$. On the other hand, for every $\mathbf{q} \in \mathcal{Q}_\rho$ it holds that $\sum_{i=1}^n q_i \mathbf{a}^i \in D_{\mathbf{w}_\alpha}$, which implies $\mathcal{U}_\rho \subseteq D_{\mathbf{w}_\alpha}$. We conclude that $\mathcal{U}_\rho = D_{\mathbf{w}_\alpha}$.

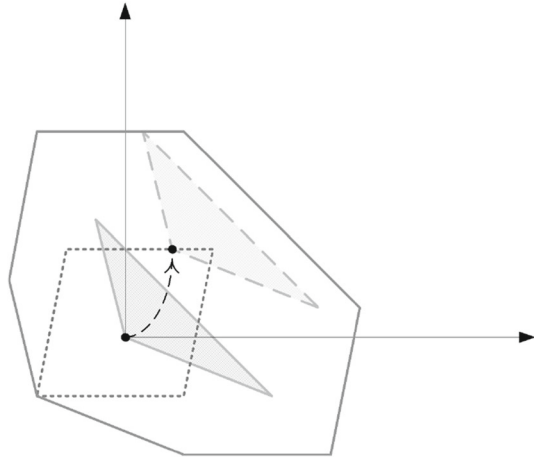
Thus, using the result from Proposition 4, we have proven the equality between the distortion risk constraint feasible set and a properly chosen WM region, which parallels Theorem 4.3 in Bertsimas and Brown (2009) and proves Theorem 1.

Finally, we emphasize a fact that is used for constructing trimmed regions as convex polytopes: the vertices of a polytope are its extreme points. From the above we know that the directions $\mathbf{p} \in S^{d-1} \setminus H(\mathbf{a}^1, \dots, \mathbf{a}^n)$ belong to vertices, while the directions $\mathbf{p} \in H(\mathbf{a}^1, \dots, \mathbf{a}^n)$ belong to parts of the boundary that have affine dimension ≥ 1 .

Risk-relevant properties of WM regions

In the context of risk measurement, it is crucial that the WM regions possess two properties that enable them to generate coherent risk measures: *monotonicity* and *subadditivity*.

Fig. 6 An illustration of the subadditivity property



Proposition 5 (Coherency properties of WM regions)

1. *Monotonicity:* If $\mathbf{z}_k \leq \mathbf{y}_k$ holds for all k (in the componentwise ordering of \mathbb{R}^d), then

$$D_{\mathbf{w}_\alpha}(\mathbf{y}_1, \dots, \mathbf{y}_n) \subset D_{\mathbf{w}_\alpha}(\mathbf{z}_1, \dots, \mathbf{z}_n) \oplus \mathbb{R}_+^d, \quad \text{and}$$

$$D_{\mathbf{w}_\alpha}(\mathbf{z}_1, \dots, \mathbf{z}_n) \subset D_{\mathbf{w}_\alpha}(\mathbf{y}_1, \dots, \mathbf{y}_n) \oplus \mathbb{R}_-^d.$$

2. *Subadditivity:*

$$D_{\mathbf{w}_\alpha}(\mathbf{y}_1 + \mathbf{z}_1, \dots, \mathbf{y}_n + \mathbf{z}_n) \subset D_{\mathbf{w}_\alpha}(\mathbf{y}_1, \dots, \mathbf{y}_n) \oplus D_{\mathbf{w}_\alpha}(\mathbf{z}_1, \dots, \mathbf{z}_n).$$

In this Proposition, the symbol \oplus is the *Minkowski addition*, $A \oplus B = \{a + b : a \in A, b \in B\}$ for A and $B \subset \mathbb{R}^d$. For a proof, see [Dyckerhoff and Mosler \(2011\)](#).

The subadditivity property of WM regions is an immediate extension of the subadditivity restriction usually imposed on univariate risk measures. In dimensions two and more, it has an interpretation as a dilation of one trimmed region by the other. To understand this better, let us consider the simple example of the Minkowski addition given in Fig. 6. The figure exhibits a solid triangle with one vertex at the origin and a dotted-border quadrangle. Now, move the triangle in such a way that its lower left corner passes all points of the quadrangle. At each point of the quadrangle, we get a copy of the initial triangle (having dashed border) shifted by the coordinate of the point. The union of all these triangles gives us the Minkowski sum of the initial two sets, which is the big heptagon in Fig. 6. Observe that, if the rectangle is moved around the triangle, the same sum is obtained. The subadditivity states that if, e.g., these two figures are WM regions $D_{\mathbf{w}_\alpha}(\mathbf{y}_1, \dots, \mathbf{y}_n)$ and $D_{\mathbf{w}_\alpha}(\mathbf{z}_1, \dots, \mathbf{z}_n)$, then $D_{\mathbf{w}_\alpha}(\mathbf{y}_1 + \mathbf{z}_1, \dots, \mathbf{y}_n + \mathbf{z}_n)$ is contained in the heptagon.

References

- Acerbi C (2002) Spectral measures of risk: a coherent representation of subjective risk aversion. *J Bank Finance* 26:1505–1581
- Adam A, Houkari M, Laurent JP (2008) Spectral risk measures and portfolio selection. *J Bank Finance* 32:1870–1882
- Bazovkin P, Mosler K (2012a) An exact algorithm for weighted-mean trimmed regions in any dimension. *J Stat Softw* 47(13)
- Bazovkin P, Mosler K (2012b) StochaTR: solving stochastic linear programs with a single risk constraint. R package version 1.0.4. <http://CRAN.R-project.org/package=StochaTR>
- Ben-Tal A, Ghaoui LE, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton
- Ben-Tal A, Bertsimas D, Brown DB (2010) A soft robust model for optimization under ambiguity. *Oper Res* 58(4):1220–1234
- Bertsimas D, Brown DB (2009) Constructing uncertainty sets for robust linear optimization. *Oper Res* 57(6):1483–1495
- Bertsimas D, Goyal V (2013) On the approximability of adjustable robust convex optimization under uncertainty. *Math Methods Oper Res* 77(3):323–343
- Chen W, Sim M, Sun J, Teo CP (2010) From cvar to uncertainty set: implications in joint chance-constrained optimization. *Oper Res* 58:470–485
- Dyckerhoff R, Mosler K (2011) Weighted-mean trimming of multivariate data. *J Multivariate Anal* 102:405–421
- Dyckerhoff R, Mosler K (2012) Weighted-mean trimming of a probability distribution. *Stat Probab Lett* 82:318–325
- Fabozzi FJ, Huang D, Zhou G (2010) Robust portfolios: contributions from operations research and finance. *Ann Oper Res* 176:191–220
- Föllmer H, Schied A (2004) Stochastic finance: an introduction in discrete time. Walter de Gruyter, Berlin
- Gabrel V, Murat C, Thiele A (2014) Recent advances in robust optimization: an overview. *Eur J Oper Res* 235(3):471–483
- Huber PJ (1981) Robust statistics. Wiley, New York
- Kall P, Mayer J (2010) Stochastic linear programming. Models, theory, and computation, 2nd edn. Springer, New York
- Kawas B, Thiele A (2011) A log-robust optimization approach to portfolio management. *OR Spectrum* 33(1):207–233
- Kay TL, Kajiya JT (1986) Ray tracing complex scenes. *SIGGRAPH Comput Graph* 20:269–278
- Kusuoka S (2001) On law invariant coherent risk measures. *Adv Math Econ* 3:83–95
- Natarajan K, Pachamanova D, Sim M (2009) Constructing risk measures from uncertainty sets. *Oper Res* 57(5):1129–1141
- Nemirovski A, Shapiro A (2006) Convex approximations of chance constrained programs. *SIAM J Optim* 17:969–996
- Pflug GC (2006) On distortion functionals. *Stat Decis* 24(1):45–60
- Prékopa A (1995) Stochastic programming. Kluwer Academic Publ., Dordrecht
- Rockafellar R, Uryasev S (2000) Optimization of conditional value-at-risk. *J Risk* 2:21–41
- Rockafellar RT (1997) Convex analysis. Princeton Univ. Press, Princeton
- Wang SS, Young VR, Panjer HH (1997) Axiomatic characterization of insurance prices. *Insur Math Econ* 21(2):173–183