

## A general solution for robust linear programs with distortion risk constraints

Pavel Bazovkin · Karl Mosler

Received: date / Accepted: date

**Abstract** Linear optimization problems are investigated that have random parameters in their  $m \geq 1$  constraints. In constructing a robust solution  $\mathbf{x} \in \mathbb{R}^d$ , we control the risk arising from violations of the constraints. This risk is measured by set-valued risk measures, which extend the usual univariate coherent distortion (= spectral) risk measures to the multivariate case. To obtain a robust solution in  $d$  variables, the linear goal function is optimized under the restrictions holding uniformly for all parameters in a  $d$ -variate uncertainty set. This set is built from uncertainty sets of the single constraints, each of which is a weighted-mean trimmed region in  $\mathbb{R}^d$  and can be efficiently calculated. Furthermore, a possible substitution of violations between different constraints is investigated by means of the admissible set of the multivariate risk measure. In the case of no substitution, we give an exact geometric algorithm, which possesses a worst-case polynomial complexity. We extend the algorithm to the general substitutability case, that is, to robust polyhedral optimization. The consistency of the approach is shown for generally distributed parameters. Finally, an application of the model to supervised machine learning is discussed.

**Keywords** Robust optimization · data depth · weighted-mean trimmed regions · coherent risk measure · multivariate risk measure · robust classification · algorithm

**Mathematics Subject Classification (2000)** 90C15 · 90C25 · 52B11 · 62H30

---

Pavel Bazovkin  
Institute of Econometrics and Statistics, University of Cologne, D-50923 Cologne, Germany  
Tel.: +49-221-4706561  
Fax: +49-221-4705084  
E-mail: bazovkin@wiso.uni-koeln.de

Karl Mosler  
Institute of Econometrics and Statistics, University of Cologne, D-50923 Cologne, Germany  
Tel.: +49-221-4702982  
Fax: +49-221-4705084  
E-mail: mosler@statistik.uni-koeln.de

## 1 Introduction

### 1.1 The robust model

In the last decade, much progress has been made in the field of *robust linear optimization*, that is, in finding worst-case solutions under uncertain side conditions. A wide spectrum of models and methods has been proposed. Recent developments in theory and practice are reviewed by Gabrel et al (2014). For a systematized collection of most significant ones, the reader is referred to Bertsimas et al (2011). An important part of the literature uses *risk measures* to quantify the uncertainty of violations of side conditions; see, e.g., Mosler and Bazovkin (2014), Bertsimas and Brown (2009), and Natarajan et al (2009). Such risk measures are flexible and allow an immediate interpretation. They can be properly selected and tuned to control the relevant sources of uncertainty. Then, essentially, the goal function is optimized under the restriction that the risk of violation stays within acceptable bounds.

The present paper contributes to this strand. We generalize the approach of Mosler and Bazovkin (2014), where a single-constraint optimization was solved, to much more general restrictions. In doing this, for each linear restriction of a given linear program a so-called *uncertainty set* of parameters is constructed. It consists of all possible values of the unknown coefficients that are acceptable at a specified risk level of constraint violation. We employ *multivariate coherent distortion risk measures*, which are extensions of the usual univariate coherent distortion risk measures: The uncertainty sets of these measures are convex bodies and come out to coincide with *weighted-mean (WM) trimmed regions*. WM regions, as recently developed by Dyckerhoff and Mosler (2011), describe a multivariate distribution by regions of different *depth* (= centrality). They can be exactly calculated in any dimension; see Bazovkin and Mosler (2012).

Various other notions of uncertainty sets have been proposed in the recent literature; a review in the context of portfolio optimization is given in Fabozzi et al (2010). These approaches define uncertainty sets like confidence sets, describing the uncertainty in parameters of parametric distributional assumptions (e.g. Delage and Ye (2010)), or by special functionals (e.g.  $\varphi$ -divergences in Ben-Tal et al (2013)). In our approach we avoid such assumptions and represent the uncertainty of constraint violations in a purely nonparametric way, *viz.* by depth-based central regions. To obtain a numerical solution of a given optimization problem, we propose an algorithm that employs the well developed geometric machinery of central regions calculation.

Originally, we consider the following linear program,

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

and assume that  $\tilde{\mathbf{A}}$  is an  $m \times d$  matrix having stochastic entries and that  $\tilde{\mathbf{b}} \in \mathbb{R}^m$  may be stochastic, too. From now on, we will mark random variables with a tilde: For example,  $\tilde{\mathbf{b}}$  will denote the stochastic right-hand side vector, while  $\mathbf{b}$  is used for a deterministic value of it.

The linear program (1) is called a *stochastic linear optimization problem*, or *stochastic linear program (SLP)*. Of course, whether the stochastic side condition is satisfied or not for some  $\mathbf{x}$ , depends on the realizations of  $\tilde{\mathbf{A}}$  and  $\tilde{\mathbf{b}}$ . Thus an

optimal solution to (1) becomes a random vector itself. One possibility to cope with this randomness is to quantify the degree of violation of the side condition by some risk measure and to substitute the condition by a bound on the measure.

Risk measures are widely used in financial mathematics, where they assess the ex-ante risk of a financial position. The uncertain position is generally modeled by a random vector  $\mathbf{Y}$ . A univariate risk measure basically provides the value  $\mathbf{c}$  of a monetary deposit that, being added to the given uncertain position  $\mathbf{Y}$ , cancels its risk. The latter means that the location of the distribution of  $\mathbf{Y} + \mathbf{c}$  satisfies certain requirements that are set by, say, a regulator. Similarly, an  $m$ -variate risk measure maps an  $m$ -variate random vector to a deterministic  $m$ -vector of monetary deposits.

For example, if  $\mathbf{Y}$  is univariate, the regulator may require that some  $\alpha$ -quantile  $Q_Y(\alpha)$  of  $\mathbf{Y}$  be non-negative. If we add the constant  $-Q_Y(\alpha)$  to  $\mathbf{Y}$ , which can be interpreted as an insurance deposit to the distribution, we make the condition hold. In other words, only the worst  $\alpha \cdot 100\%$  of outcomes of the insured position are expected to be negative. This risk measure is called the *value-at-risk* at  $\alpha$  ( $V@R_\alpha$ ) and widely used in finance. Actually, many risk measures have been proposed and investigated in the literature, each controlling a particular aspect of the outcome distribution. Another popular univariate risk measure is the *expected shortfall*. Others are the *expected minimum* and the *entropic risk measure*. These measures easily transfer from financial losses to losses incurred by violating the side conditions of a general stochastic program.

In what follows, the stochastic side condition in (1) is handled by means of *risk constraints*,

$$\rho^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \leq \mathbf{0}, \quad \text{resp.} \quad (2)$$

$$\rho^1(\tilde{\mathbf{A}}_j\mathbf{x} - b_j) \leq 0, \quad j = 1 \dots m, \quad (3)$$

where  $\tilde{\mathbf{A}}_j$  denotes the  $j$ -th row of  $\tilde{\mathbf{A}}$ , and  $\rho^m$  is a risk measure taking values in  $\mathbb{R}^m$ ,  $m \geq 1$ . An SLP that minimizes  $\mathbf{c}'\mathbf{x}$  subject to the restrictions (2) or (3) is called a *risk-constrained stochastic linear program*.

Alternatively, in place of a risk constraint, the probability of satisfying all restrictions may be controlled by a *joint chance constraint*,

$$\text{Prob}[\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b} \geq \mathbf{0}] \geq 1 - \alpha. \quad (4)$$

Limiting the violation probability by some fixed  $\alpha$  the well-known *chance-constrained linear program* is obtained. Sometimes, neglecting eventual stochastic dependencies among the constraints, *separate chance constraints*,

$$\text{Prob}[\tilde{\mathbf{A}}_j\mathbf{x} - b_j \geq 0] \geq 1 - \alpha_j, \quad j = 1 \dots m, \quad (5)$$

are considered. Here, a chance constraint limits the risk of each single side condition by imposing a maximum probability  $\alpha_j$  on its violation.

Obviously, when using the constraints (2) or (3), the choice of the risk measure is a crucial point. In this paper we consider no single risk measure but a comprehensive class of such measures, the *coherent distortion risk measures*.

**Definition 1 (Coherent distortion risk measure)** Let  $r : [0, 1] \rightarrow [0, 1]$  be a given function that is increasing and concave with  $r(0) = 0$  and  $r(1) = 1$ . For any univariate random variable  $Y$  define

$$\rho(Y) = - \int_0^1 Q_Y(t) dr(t), \quad (6)$$

provided the integral exists. Here again,  $Q_Y$  denotes the quantile function of  $Y$ . Then  $\rho$  is mentioned as a *coherent distortion risk measure*, and  $r$  as its *weight generating function*.

If  $Y$  has an empirical distribution on the ordered values  $y_{(1)}, y_{(2)}, \dots, y_{(n)}$ , this definition becomes

$$\rho(Y) = - \sum_{i=1}^n w_i y_{(i)}, \quad (7)$$

with weights  $w_i = r(\frac{i}{n}) - r(\frac{i-1}{n})$ \* decreasing in  $i$ . Coherent distortion risk measures possess certain desired properties: *monotonicity*, *translation invariance*, *law invariance*, *positive homogeneity* and *subadditivity*. The last two properties imply a most important postulate of risk measurement: *coherence*, that means that the risk measure decreases with diversification. An example of a coherent distortion risk measure is the *expected shortfall*, which is defined by choosing

$$r(t) = \begin{cases} t/\alpha & \text{if } 0 \leq t < \alpha, \\ 1 & \text{if } \alpha \leq t \leq 1. \end{cases} \quad (8)$$

Note that the value-at-risk  $V@R_\alpha$  satisfies (6) with  $r(t) = 0$  for  $t < \alpha$  and  $r(t) = 1$  otherwise, which is not a concave function. The  $V@R_\alpha$  is an example of a *non-coherent* distortion risk measure.

In a more general context, a risk measure can be regarded as a *quality measure* (cf. Kall and Mayer (2010)). Our choice of the quality measure, besides its generality, possesses a clear interpretation and always generates a convex program. Later we will demonstrate that our approach in fact suites an even more general robust program that not only copes with linear stochastic restrictions, but also with those of a *robust polyhedral* type, which include robust conic restrictions as a special case.

In applications, information about the stochastic parameters has to be extracted from data. Here we assume that a sample of coefficient matrices  $\mathbf{A}^1, \dots, \mathbf{A}^n \in \mathbb{R}^{m \times d}$  has been observed and the solution of the SLP is based on this data. The data is mentioned as an *empirical distribution* giving equal mass  $\frac{1}{n}$  to  $\mathbf{A}^1, \dots, \mathbf{A}^n$ , and with this data the SLP is named an *empirical risk-constrained SLP*. Similarly, when also the right hand side is stochastic, a joint sample of  $(\tilde{\mathbf{A}}, \tilde{\mathbf{b}})$  is considered.

The next subsection outlines the single-constraint algorithm of Mosler and Bazovkin (2014) and provides preliminaries for its extension.

## 1.2 The single-constraint problem

In case of a single stochastic constraint ( $m = 1$ ) the SLP (1) specializes to

$$\mathbf{c}' \mathbf{x} \longrightarrow \min \quad \text{s.t.} \quad \tilde{\mathbf{a}} \mathbf{x} \geq b, \quad (9)$$

where  $\tilde{\mathbf{a}}$  is a random vector of parameters in  $\mathbb{R}^d$ . To solve (9), an *uncertainty set*  $\mathcal{U}$  is constructed that yields a *robust linear program*:

$$\mathbf{c}'\mathbf{x} \longrightarrow \min \quad \text{s.t. } \mathbf{a}'\mathbf{x} \geq b \text{ for all } \mathbf{a} \in \mathcal{U}. \quad (10)$$

The uncertainty set  $\mathcal{U}$  consists of all parameters that are taken into account by the decision maker to make his or her risk acceptable. It was found that the solution of (10) corresponds to a parameter  $\mathbf{a}$  that lies on the surface of  $\mathcal{U}$ .

To transform (9) into (10) we have to construct the uncertainty set  $\mathcal{U}$ , depending on the data and the risk measure. For a distortion risk measure  $\rho$  on an empirical distribution it turns out that the single restriction (3) can be equivalently formulated as the restriction in (10) with  $\mathcal{U}$  being a so-called *weighted-mean region* whose weights depend on  $\rho$ .

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

$$D_{\mathbf{w}_\alpha}(\mathbf{a}^1, \dots, \mathbf{a}^n) = \text{conv} \left\{ \sum_{j=1}^n w_{\alpha,j} \mathbf{a}^{\pi(j)} : \pi \text{ permutation of } \{1, \dots, n\} \right\}. \quad (11)$$

Here  $\mathbf{w}_\alpha = [w_{\alpha,1}, \dots, w_{\alpha,n}]'$  is a vector of ordered non-negative weights that add up to 1. Additionally, for all  $k = 1, \dots, n$ , the sum  $\sum_{j=1}^k w_{\alpha,j}$  has to increase with  $\alpha$ . Any such family of *weight vectors*  $\{\mathbf{w}_\alpha\}_{0 \leq \alpha \leq 1}$  defines a notion of WM regions. There are many notions of weighted-mean trimmed regions, among them well-known families like the zonoid regions and the expected convex hull regions. For example, the *zonoid regions* are given by

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

$0 \leq \alpha \leq 1$ . For details on WM regions and their extension to general probability measures, see Dyckerhoff and Mosler (2011) and Dyckerhoff and Mosler (2012). The connection between univariate coherent distortion risk measures and WM regions is established in the following Proposition:

**Proposition 1 (Mosler and Bazovkin (2014))** *Let  $\rho$  be a univariate coherent distortion risk measure and let the random vector  $\tilde{\mathbf{a}}$  have an empirical distribution on  $\mathbf{a}^1, \dots, \mathbf{a}^n \in \mathbb{R}^d$ . Then there exists a weight vector  $\mathbf{w}_\alpha$  such that  $D_{\mathbf{w}_\alpha}$  is a WM region and it holds*

$$\{\mathbf{x} \in \mathbb{R}^d \mid \rho(\tilde{\mathbf{a}}'\mathbf{x} - b) \leq 0\} \quad (12)$$

$$= \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{a}'\mathbf{x} \geq b \forall \mathbf{a} \in D_{\mathbf{w}_\alpha}(\mathbf{a}^1, \dots, \mathbf{a}^n)\}. \quad (13)$$

Specifically, this Proposition states that the class of univariate restrictions (3) involving coherent distortion risk measures corresponds to weighted-mean trimmed regions in  $\mathbb{R}^d$  as uncertainty sets. Calculating  $\mathcal{U}$  turns out to be computationally feasible due to the direct connection between the uncertainty set and a trimmed region, which can be efficiently determined by the algorithm of Bazovkin and Mosler (2012).

According to (7), the risk measure  $\rho$  defines the vector  $\mathbf{w}_\alpha$  uniquely\* and, by this, the trimmed region  $D_{\mathbf{w}_\alpha}$ . For instance, for the expected shortfall at level  $\alpha$  we generate the the following weights by the formula (8):

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

and see that the obtained weight vector is equivalent, up to the transposition of components, to the one defining the zonoid regions. Obviously, knowing  $n$  and the parameter  $\alpha$  we can calculate  $\mathbf{w}_\alpha$  in advance\* and use it as an input for the algorithm.

The further steps of the algorithm of Mosler and Bazovkin (2014) are essentially standing on the search for an intersection of a ray (in the direction of  $\mathbf{c}$ ) with the uncertainty set  $\mathcal{U} = D_{\mathbf{w}_\alpha}$ . The solution is characterized by the normal to the facet of  $\mathcal{U}$  that is intersected by the line in direction of the vector  $\mathbf{c}$ .

The goal of this paper is to develop a similar approach to (2) and more general SLPs. Specifically, we

1. *generalize* the analysis of the single-constraint SLP to multiple risk constraints ( $m \geq 2$ );
2. *construct* a geometric algorithm to solve the multi-constraint problem (if constraints cannot be compensated by each other, i.e. in the unsubstitutability case, the algorithm operates in the same dimension  $d$  as the single-constraint procedure does);
3. *extend* the robust multi-constraint linear optimization to *robust polyhedral optimization* (which covers the substitutability case);
4. *estimate* the uncertainty set and the robust solution consistently.

The further material is organized as follows. The construction of the solution for the multi-constraint SLP is described in Section 2. The formal algorithm is given in the subsequent Section 3. In the same Section the algorithm is extended to a program with stochastic right-hand side. The consistency of the solution for generally distributed data is proven in Subsection 3.5. Finally, Section 4 is devoted to a discussion of the algorithm and an application to supervised learning.

## 2 Multiple constraints

### 2.1 A general model

Consider the SLP (1) with  $m \geq 2$  constraints and deterministic right-hand side  $\mathbf{b}$ . We aim at generalizing Proposition 1 and eliminating uncertainty by a robust linear program as follows:

$$\mathbf{c}'\mathbf{x} \longrightarrow \min \quad s.t. \quad \mathbf{A}\mathbf{x} \geq \mathbf{b} \quad \text{for all } \mathbf{A} : \delta(\mathbf{A}) \in \mathcal{U}, \quad (14)$$

where  $\delta(\mathbf{A}) = (\mathbf{A}_1, \dots, \mathbf{A}_m)' \in \mathbb{R}^{m \cdot d}$  is the vectorized matrix  $\mathbf{A}$ .

Again, a proper uncertainty set  $\mathcal{U}$  has to be constructed. In doing so, we specify a risk measure that is no single vector but a whole set of such vectors in  $\mathbb{R}^m$ . Such

\* 2 RM and weights correspondence

a *set-valued risk measure* is interpreted as the set of deterministic vectors in  $\mathbb{R}^m$  which, when being added to the random variable  $\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}$ , cannot completely shift its  $\alpha$ -central region<sup>1</sup> out of the negative orthant  $\mathbb{R}_-^d$ . In other words, which cannot, at the given acceptance level, guarantee to avoid a strictly negative outcome of the shifted random variable.

Cascos and Molchanov (2007) have shown that certain multivariate risk measures correspond to trimmed regions of the considered distribution. In particular, the  $m$ -variate set-valued analogue  $\mu^m$  of a spectral risk measure is defined through a WM region  $D_{\mathbf{w}_\alpha}$  in the following way:

$$\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) = -\left(D_{\mathbf{w}_\alpha}(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \oplus \mathbb{R}_+^m\right) \subset \mathbb{R}^m, \quad (15)$$

where  $\oplus$  denotes the Minkowski sum.

We should mention that spectral risk measures are equivalent to coherent distortion risk measures. It is easy to see, that in dimension one  $\mu^1(\tilde{\mathbf{a}}'\mathbf{x} - b)$  is the half-line bounded above by  $\rho^1(\tilde{\mathbf{a}}'\mathbf{x} - b)$ , where  $\rho^1$  is a coherent distortion measure of a univariate risk. Thus, (15) can be regarded as a set-valued extension of a univariate distortion risk measure to multiple dimensions<sup>2</sup>. In this paper, we use  $\mu$  for denoting set-valued, and  $\rho$  for denoting vector-valued or real-valued measures.

If a linear program (1) has more than one stochastic constraints, we must consider not only that some or all of them may be violated, but also that the degree of violation of a restriction may be offset against that of another restriction. That is, the decision maker, in evaluating a possible solution, may compensate the missing strictness of one constraint by the fact that another constraint or a group of constraints is more strictly satisfied. In this, the *values* of single constraint satisfaction are regarded as *substitutable* by the decision maker, and his or her task in selecting a solution includes some kind of diversification regarding the constraints. Note that this possible value compensation between the constraints has nothing to do with a potential stochastic dependency among the parameters of different constraints.

To include the possibility of value substitution in our model, we introduce a multivariate utility (= negative loss) function  $u : \mathbb{R}^m \rightarrow \mathbb{R}$  that evaluates the violations  $v_1, \dots, v_m$  of the  $m$  constraints. Consider  $\mathcal{F} = \{\mathbf{v} : u(\mathbf{v}) \geq 0\}$  as the set of admissible violations. If  $u$  is a quasiconcave function,  $\mathcal{F}$  is convex. Later we will specialize  $\mathcal{F}$  to be a convex polyhedron, see (17). The marginals may or may not substitute each other. This fact actually affects the form of  $\mathcal{F}$ .

<sup>1</sup> An  $\alpha$ -central region is a set of possible outcome vectors that are taken into reasonable account by the decision maker (dependent on his or her risk posture  $\alpha$ ).

<sup>2</sup> Rüschemdorf (2013) proposes a different notion of a multivariate distortion risk measure, which is scalar-valued: Given a  $d$ -variate distribution having p.d.f.  $F$ , he considers the level set  $Q(t)$  of  $F$  at level  $t$  and defines some scalar measure of  $Q(t)$  as the  $t$ -quantile. Then, based on these scalar-valued quantiles, he introduces multivariate risk measures in the same way as univariate ones.

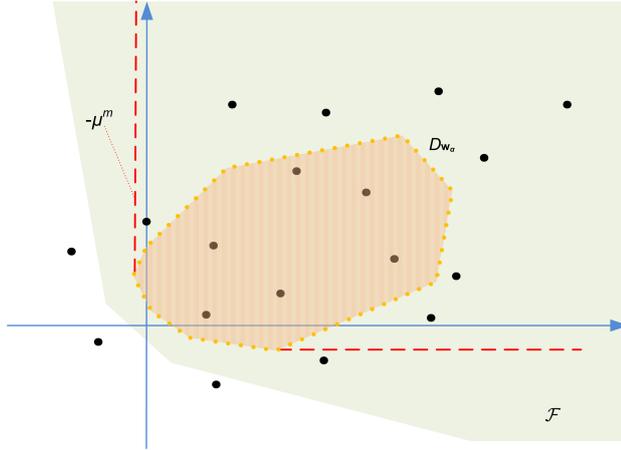


Fig. 1 Illustration of the measure  $\mu^m$  on a plane.

Using  $\mathcal{F}$ , we rewrite the joint risk constraint (2) as follows<sup>3</sup>:

$$-\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \subset \mathcal{F}. \quad (16)$$

The above is illustrated on Figure 1 in a simplified representation. There the orange dotted trimmed region  $D_{w_\alpha}(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})$  augmented by the Minkowski sum with  $\mathbb{R}_+^m$  (the red dashed lines) forms the set  $-\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})$ , which, in turn, should be contained in the green set  $\mathcal{F}$ .

Generally, the set  $\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})$  is contained in the set of all violation vectors  $\mathbf{v} \in \mathbb{R}^m$  that are admitted. If substitution between constraints is possible, the level of substitutability may vary from full substitutability to *unsubstitutability*. It is easy to show, that the first extremal case leads to an equivalent single-constraint SLP.

Full substitutability means that  $u$  is additive,  $u(\mathbf{v}) = \sum_j u_j(v_j)$ , and marginal utilities  $u_j$  are linear. In this case we obtain

$$\mathcal{F}_{\text{sub}} = \{\mathbf{v} : u(\mathbf{v}) = \sum_j u_j(v_j) = \sum_j k_j \cdot v_j = \mathbf{k}'\mathbf{v} \geq 0\}$$

with some  $\mathbf{k} \geq \mathbf{0}$ . This reduces to a problem with a single constraint  $(\mathbf{k}'\tilde{\mathbf{A}})\mathbf{x} \geq \mathbf{k}'\mathbf{b}$ . Here the admissible set  $\mathcal{F}_{\text{sub}}$  is a halfspace bordered by the hyperplane passing the origin and having normal  $\mathbf{k}$ . In the second extreme case (unsubstitutability) the admissible set is the positive orthant,  $\mathcal{F}_{\text{unsub}} = \mathbb{R}_+^m$ . Solving the SLP in

<sup>3</sup> Here,  $\mathcal{F}$  coincides with the admissible set of the vector-valued multivariate risk measure  $\nu$  described in Bazovkin (2014). This measure consists in the smallest vector which, when being added to the random returns vector  $\mathbf{Y}$ , puts  $-\mu^m(\mathbf{Y} + \nu(\mathbf{Y}))$  into the admissible set. The latter is the set of returns which appear to be acceptable to the risk taker (or the regulator). In this model, we obtain (16) with only  $\mu^m$  and  $\mathcal{F}$  being involved:

$$\nu(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \leq \mathbf{0} \iff -\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \subset \mathcal{F}.$$

this case will be the subject of Section 3. In the intermediate case of the partial substitutability  $\mathcal{F}$  is a set lying in  $\mathcal{F}_{\text{sub}}$  and containing  $\mathcal{F}_{\text{unsub}}$ .

To sum up, for obtaining a general solution of the multi-constraint SLP we have to consider different levels of substitutability among the violation of constraints. It turns out that the general substitutability case can be reduced to unsubstitutability via a special transformation of the model. In the next subsection we will define the transformation and demonstrate this fact. After that, to manage the complete task, we solve the SLP with unsubstitutability.

## 2.2 The general substitutability case

Now we consider the case that violations of constraints can be balanced against each other. Let us assume that the set  $\mathcal{F}$  of feasible violations is a convex polyhedron being characterized by  $K$  linear inequalities,

$$\mathcal{F} = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{p}'_k \mathbf{y} \geq d_k, \quad k = 1 \dots K\} \quad (17)$$

with some  $\mathbf{p}_1, \dots, \mathbf{p}_K \in \mathbb{R}_+^m$  and  $d_1, \dots, d_K \in \mathbb{R}$ , that is,  $\mathcal{F}$  is an *upper convex polytope*. Obviously, this assumption includes the extreme cases of full substitutability and unsubstitutability, and can be seen as an approximation for the intermediate cases.

**Proposition 2** Let  $\rho^1(Z)$  denote the upper border of the halfline  $\mu^1(Z) \subset \mathbb{R}$ . Then it holds:  $-\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b}) \subset \mathcal{F}$  if and only if

$$\rho^1(\mathbf{p}'_k \tilde{\mathbf{A}}\mathbf{x} - \mathbf{p}'_k \mathbf{b}) \leq -d_k, \quad \text{for all } k = 1 \dots K. \quad (18)$$

**Proof.** Let  $h_S^m$  denote the support function<sup>4</sup>. For any  $\mathbf{y} \in -\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})$  and  $k = 1 \dots K$  we obtain:

$$\begin{aligned} \mathbf{p}'_k \mathbf{y} &\geq \min_{\mathbf{z} \in -\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})} \{\mathbf{p}'_k \cdot \mathbf{z}\} \\ &= h_{-\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})}^m(-\mathbf{p}_k) \\ &= h_{\mathbf{p}'_k \cdot \mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})}^1(1) = h_{\mu^1(\mathbf{p}'_k \tilde{\mathbf{A}}\mathbf{x} - \mathbf{p}'_k \mathbf{b})}^1(1) \\ &= -\rho^1(\mathbf{p}'_k \tilde{\mathbf{A}}\mathbf{x} - \mathbf{p}'_k \mathbf{b}). \end{aligned}$$

Consequently, we have  $\mathbf{y} \in \mathcal{F}$  if  $-\rho^1(\mathbf{p}'_k \tilde{\mathbf{A}}\mathbf{x} - \mathbf{p}'_k \mathbf{b}) \geq d_k$ . Hence the “if” part of the proposition is proved. On the other hand, there exists some  $\mathbf{y} \in -\mu^m(\tilde{\mathbf{A}}\mathbf{x} - \mathbf{b})$  so that the first-line inequality is met with equality. Hence the “only if” part holds, too.  $\square$

Proposition 2 leads to the following Theorem, which for every general model provides an equivalent unsubstitutability model:

**Theorem 1** The SLP with the  $m$ -fold risk constraint (16) and  $\mathcal{F}$  defined by (17) is equivalent to an SLP with  $K$  unsubstitutable constraints

$$\mathbf{p}'_k \tilde{\mathbf{A}}\mathbf{x} \geq \mathbf{p}'_k \mathbf{b} \quad \text{for } k = 1 \dots K. \quad (19)$$

<sup>4</sup> The support function of a set  $S$  in  $\mathbb{R}^m$  is defined as  $h_S^m(\mathbf{p}) = \sup\{\mathbf{x}'\mathbf{p} : \mathbf{x} \in S\}$ ,  $\mathbf{p} \in \mathbb{R}^m$ .

**Proof.** According to Proposition 2, the SLP with joint risk constraint (16) is equivalent to the SLP with constraints (18). On the other hand, an SLP with constraints (19) produces the same risk constraints.  $\square$

Theorem 1 enables us to determine a generalized uncertainty set in the multi-constraint case. Notice \* that we can discriminate between the  $K$  constraints by choosing different distortion risk measures  $\rho^1$  for each of them. Keeping in mind the one-to-one correspondence (7) between such a measure and a weight vector  $\mathbf{w}_\alpha$ , we may use a weight vector  $\mathbf{w}_{\alpha_j}$  for the  $j$ -th constraint. For example, it can be the expected shortfall with proper  $\alpha_1, \dots, \alpha_K$  chosen for each constraint separately. However, clearly, we are not limited to using only the expected shortfall for all constraints.

**Corollary 1** *The uncertainty set  $\mathcal{U}$  of the matrix  $\tilde{\mathbf{A}}$  in an SLP with risk constraint (16) and violations admissible set  $\mathcal{F}$  (17) equals*

$$\mathcal{U} = \left\{ \mathbf{A} : \delta([\mathbf{p}'_k \mathbf{A}]_{k=1\dots K}) \in \mathbf{X}_{j=1\dots K} D_{\mathbf{w}_{\alpha_j}}(\mathbf{p}'_j \tilde{\mathbf{A}}) \right\}. \quad (20)$$

**Proof.** Here  $\mathbf{X}$  denotes the  $K$ -fold Cartesian product of WM regions. Due to Theorem 1, we can transform the general model into a model with unsubstitutability. In this case single uncertainty sets will not affect each other. In turn, each single uncertainty set is calculated as for the single-constraint SLP.  $\square$

### 2.3 The equivalent unsubstitutability case

We see that  $\mathcal{U}$  splits up into  $K$  parts, which can be calculated individually. It leads to the following key representation theorem:

**Theorem 2** *The SLP (1) with admissible set  $\mathcal{F}$  satisfying (17) and joint risk constraint (16) is equivalent to the following problem:*

$$\mathbf{c}'\mathbf{x} \longrightarrow \min \quad \text{s.t. } \mathbf{A}\mathbf{x} \geq \mathbf{b} \text{ for all } \mathbf{A} \in \mathcal{U}, \quad (21)$$

where  $\mathcal{U}$  is defined as in Corollary 1.

**Proof.** Follows from Theorem 1, Corollary 1 and Proposition 1.  $\square$

Note that if all  $d_k$  equal zero, we get the general risk-constrained *robust conic program*. Besides this, again, note that unsubstitutability does not imply stochastic independence of the constraints.

## 3 Unsubstitutable violations risks: The optimal solution

### 3.1 Generalizing the single-constraint approach

By Theorem 1 the general risk-constrained SLP is reduced to an SLP with unsubstitutable constraints. Therefore, it suffices to construct an algorithm for solving the latter SLP. Moreover, Theorem 2 reformulates the stochastic constraints in terms of an uncertainty set.

In this subsection we pursue the following idea: We want to reformulate our problem in a way that makes it similar to the single-constraint case. In doing so,

we first define a *convolution set* that will play the same role as the uncertainty set in the single-constraint SLP algorithm. Then, we show how to construct an equivalent to the optimization line. Having proved the equivalence of the elements, we are able to formulate the generalization of the algorithm to the multi-constraint case.

Let us write  $\mathcal{X}_i$  for the feasible set generated by the constraint  $i$ ,  $i = 1 \dots K$ , and  $\mathcal{X}$  for the common feasible set,

$$\mathcal{X} = \bigcap_{i=1}^K \mathcal{X}_i. \quad (22)$$

We aim at solving the SLP by a geometric procedure in the parameter space. For each constraint, the parameter space has dimension  $d$ , while the uncertainty set  $\mathcal{U}$  lives in  $\mathbb{R}^{d \cdot K}$ . We will construct a set  $\mathcal{G} \subset \mathbb{R}^d$  such that  $\mathcal{X}$  can be rewritten as follows:

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \geq \mathbf{b} \text{ whenever } \delta(\mathbf{A}) \in \mathcal{U}\} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a}'\mathbf{x} \geq 1 \forall \mathbf{a} \in \mathcal{G}\}. \quad (23)$$

$\mathcal{G}$  is obtained by convolving the general uncertainty set  $\mathcal{U}$  into  $\mathbb{R}^d$ ; we will call  $\mathcal{G}$  the parameter *convolution set*.  $\mathcal{U}$  is then decomposed into the sets  $\mathcal{U}_i, i = 1, \dots, K$ , that can be separately calculated. This dimension reducing construction is possible as the constraints are not substitutable.

The proper way here is to represent  $\mathcal{G}$  as an image of  $\mathcal{X}$  in the parameter space. According to (22), all  $\mathcal{X}_i$  are combined in one space. Combining  $\mathcal{U}_i$  in the parameter space becomes possible if any parameter  $\mathbf{a}$  contained by other uncertainty sets corresponds to the same  $X_{\mathbf{a}} = \{\mathbf{x} : \mathbf{a}'\mathbf{x} \geq b\}$  in each case. This condition holds if the right-hand sides  $b_i$  are the same for all constraints. If  $\mathbf{b} > \mathbf{0}$ , we multiply all sets  $\mathcal{U}_i$  with  $\frac{1}{b_i}$  and obtain  $b = 1$ , without changing the set of feasible solutions. If  $\mathbf{b} \not> \mathbf{0}$  we apply the transform (26), which is described in the next subsection.

$\mathcal{G}$  contains the union of  $\frac{1}{b_i}\mathcal{U}_i, i = 1 \dots K$ . Thus, according to (22) and similar to Lemma 1 in Mosler and Bazovkin (2014):

$$\mathcal{X} = \bigcap_{\mathbf{a} \in \mathcal{G}} X_{\mathbf{a}} = \bigcap_{\mathbf{a} \in \text{ext } \mathcal{G}} X_{\mathbf{a}},$$

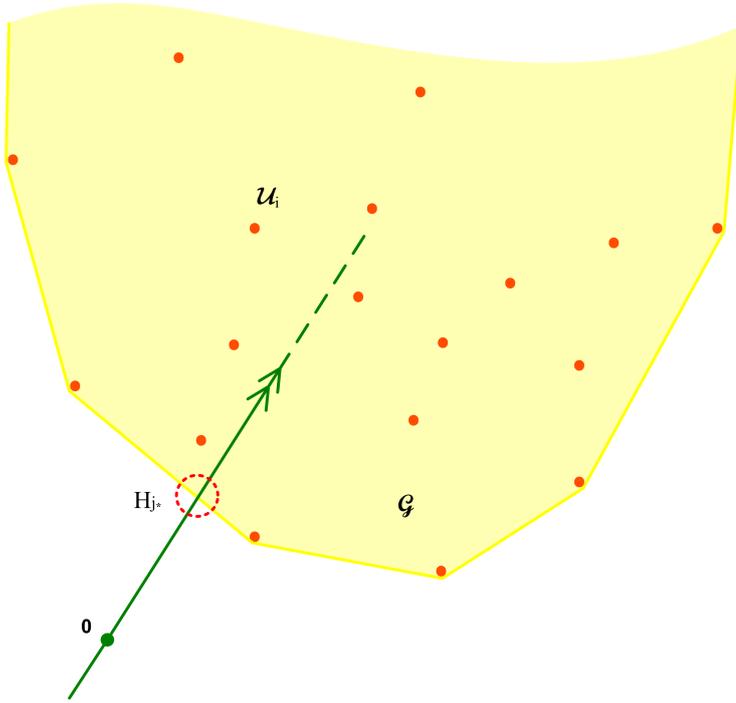
where  $\text{ext } \mathcal{G}$  denotes the set of *extreme points* of  $\mathcal{G}$ , which for a convex body corresponds to its set of vertices.

This proves that  $\mathcal{G}$  has similar properties as the uncertainty set of the single-constraint SLP. In particular, any convex combination of two points in  $\mathcal{G}$  belongs to  $\mathcal{G}$ . We obtain:

$$\mathcal{G} = \text{conv} \left\{ \bigcup_{i=1}^K \frac{1}{b_i} \cdot \mathcal{U}_i \right\}. \quad (24)$$

This transformation of  $\mathcal{X}$  is familiar to polar duality (see, e.g., Grünbaum (2003)). In our robust problem we need no explicit representation of  $\mathcal{X}$  and profit from the ready machinery of WM regions that allows the direct construction of  $\mathcal{G}$  in the parameter space.

The next step is constructing the optimization line, which is equivalent to the single-constraint case and is the line passing through the origin in direction  $\mathbf{c}$ .



**Fig. 2** Alternative to the simplex algorithm

Actually,  $\varphi$  is the locus of points that are dual to hyperplanes with the normal  $\mathbf{c}$ . However, combining the  $\mathcal{U}_i$  in one parameter space requires individual transforms of each constraint's parameter spaces, thus resulting in different  $\varphi_i$ . Observe that

- $\varphi_i \equiv s \cdot \varphi_i$  for any  $s \neq 0$ ,
- $\varphi_i \equiv \varphi$  for all  $i$ .

Hence all lines  $\varphi$  coincide. Further, they are invariant to the affine transform in (24). Therefore, the search of the optimum on  $\mathcal{G}$  equals the search on  $\mathcal{U}$  in the single-constraint SLP.

In concluding this subsection, we turn to the deterministic case, where each  $\mathcal{U}_i$  degenerates to a one-point-set  $\{\mathbf{a}_i\}$ .<sup>5</sup> The steps of our procedure stay the same but allow some simplifications. For example, calculating  $\mathcal{G}$  according to (24) reduces to the simple *quickhull* routine.<sup>6</sup>

In many practical tasks one has to consider some additional deterministic constraints, in particular, those of nonnegativity type. In such a setting, a group of trivial constraints  $x_k \geq 0, k \in \mathcal{J}$ , is mapped to a finite cloud of points in the parameter space without calculating the WM region. The latter implies that such constraints do not significantly influence the algorithm's computational complexity.

<sup>5</sup> Consequently we do not need to calculate WM regions (see Figure 2).

<sup>6</sup> Thus, the above approach can be also employed as an alternative to the regular simplex method.

### 3.2 Relaxing the right-hand side

In this subsection we show that the model (1) of the SLP also covers the case of a random coefficient vector  $\mathbf{b}$ , denoted by  $\tilde{\mathbf{b}}$ , as the restriction  $\tilde{\mathbf{A}}\mathbf{x} \geq \tilde{\mathbf{b}}$  is equivalent to

$$\begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{1} - \tilde{\mathbf{b}} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \geq \mathbf{1}, \quad (25)$$

where  $\mathbf{1}$  is a  $K$ -vector of ones. By this we obtain an SLP with deterministic right-hand side equal to  $\mathbf{1}$ . Now the solution vector has  $d + 1$  components, the last of which is fixed to 1. Geometrically the solution is searched at the intersection of the feasible set with the  $d$ -dimensional hyperplane  $\{(\mathbf{x}', x_{d+1})' \in \mathbb{R}^{d+1} : x_{d+1} = 1\}$ , that is, the task is actually solved on a convex polytope of affine dimension  $d$ .

Formally speaking, we have to solve an SLP of the form (1) with an additional equality constraint. For this we propose a similar geometrical approach. The idea is to relax the equality to a regular inequality constraint, thus obtaining the SLP in canonical form. The vector  $\mathbf{c}$  is accordingly modified.

First, we replace  $\mathbf{c}$  with the vector  $\bar{\mathbf{c}}$  that equals the normal of the facet lying on the hyperplane  $\{(\mathbf{x}', 1)' \in \mathbb{R}^{d+1}\}$ . By this, we get a non-unique solution, which also contains the sought-for optimum  $\mathbf{x}^*$ . Looking simultaneously to the parameter space, we see an intersection of  $\varphi$  with  $\mathcal{G}$  (which will be defined in the next section) at the point  $(0, \dots, 0, 1)'$ . To get rid of this non-uniqueness, we perform a small rotation of  $\bar{\mathbf{c}}$  towards  $\mathbf{x}^*$ . The latter amounts to rotating  $\bar{\mathbf{c}}$  towards  $(\mathbf{c}', 0)'$ , that is, to some new vector  $\mathbf{c}_\epsilon = (1 - \epsilon)\bar{\mathbf{c}} + \epsilon \cdot (\mathbf{c}', 0)'$  with some  $0 < \epsilon < 1$ .

**Proposition 3** *Let  $\tilde{\mathbf{b}}$  be a stochastic vector of dimension  $K$ . Then (1) is equivalent to the following model: There exists  $\bar{\epsilon} > 0$  such that*

$$\mathbf{c}'_\epsilon \begin{bmatrix} \mathbf{x} \\ x_{d+1} \end{bmatrix} \rightarrow \min \quad (26a)$$

$$s.t. \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{1} - \tilde{\mathbf{b}} \\ \mathbf{0}' & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ x_{d+1} \end{bmatrix} \geq \mathbf{1}, \quad (26b)$$

where  $\mathbf{c}_\epsilon = (1 - \epsilon)(0, \dots, 0, 1)' + \epsilon \cdot (\mathbf{c}', 0)'$ ,  $0 < \epsilon < \bar{\epsilon}$ .

**Proof.** The constraints of (1) are equivalent to (25). Obviously, if  $x_{d+1} = 1$  holds, (26b) is also equivalent to (25). To show that there exists such a small  $\epsilon > 0$  that the last inequality in (26b) necessarily turns into an equality and, as a consequence, (26b) transforms into (25), we first set  $\epsilon$  to zero. The corresponding program has a trivial non-unique solution, because the vector  $\mathbf{c}_\epsilon$  equals  $\bar{\mathbf{c}}$ , which was shown above to pick a whole facet of the feasible set for a solution.

Some small rotation of  $\mathbf{c}_\epsilon$  shifts the solution to the border of this facet, however remaining on it. This rotation is given by setting  $\epsilon > 0$ , i.e. combining the initial vector with the augmented vector  $\mathbf{c}$ , namely  $(\mathbf{c}', 0)'$ . Both facts together guarantee that  $x_{d+1}$  is fixed to 1 while optimizing (26) with  $\mathbf{c}_\epsilon$ . In this situation, (26b) reduces to (25). Moreover, the objective function turns into  $\epsilon \mathbf{c}'\mathbf{x} + 1 - \epsilon$ , which is, up to a constant, equivalent to the objective of (1). This proves the equivalence of (26) and (1).  $\square$

Let us now imagine the procedure shown above in the parameter space (see Figure 3). The additional  $\mathcal{U}_{K+1}$  is just a point  $(0, \dots, 0, 1)'$ . The virtual optimization vector  $\mathbf{c}_\epsilon$  generates a line  $\varphi_\epsilon = \epsilon \cdot \varphi + (1 - \epsilon) \cdot \{\mathbf{x} : x_j = 0, j = 1 \dots d\}$ ,

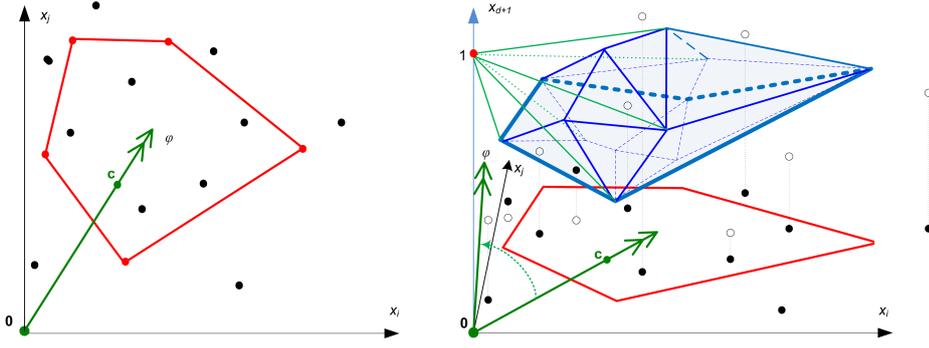


Fig. 3 Adding a dimension

that is, an equivalent affine combination of  $\varphi$  and an axis passing through the point  $(0, \dots, 0, 1)'$ .  $\varphi_\epsilon$  intersects necessarily the cone (part of the surface of  $\mathcal{G}$ ) having  $(0, \dots, 0, 1)'$  as its apex. The respective facet of  $\mathcal{G}$  determines the optimum  $\mathbf{x}^*$  similar to step (D.c.) of the algorithm in Subsection 3.3 below.

### 3.3 The algorithm

#### Prerequisites:

In solving the *general robust polyhedral optimization* problem with an arbitrary  $\mathcal{F}$ , we first modify our set of constraints and the vector  $\mathbf{b}$  according to Theorem 1, thus obtaining the  $K \times d$  matrix  $\tilde{\mathbf{A}}$  and the  $K$ -dimensional vector  $\tilde{\mathbf{b}}$ .

Without loss of generality, the algorithm is applied to a minimization problem with all constraints of the same type “ $\geq$ ”. If either  $\mathbf{b} \notin \mathbb{R}_+^K$  or  $\mathbf{b}$  is stochastic, the pretransformation of Proposition 3 should be applied first. In the sequel both modifications (including that for stochastic right-hand side  $\tilde{\mathbf{b}}$ ) are assumed to be done if necessary. Hence, we have to solve an SLP of form (26),

$$\begin{aligned} & \mathbf{c}'_\epsilon \begin{bmatrix} \mathbf{x} \\ x_{d+1} \end{bmatrix} \rightarrow \min \\ \text{s.t. } & \begin{bmatrix} \tilde{\mathbf{A}} & \mathbf{1} - \tilde{\mathbf{b}} \\ \mathbf{0}' & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ x_{d+1} \end{bmatrix} \geq \mathbf{1}, \end{aligned}$$

with  $\mathbf{c}_\epsilon = (1 - \epsilon)(0, \dots, 0, 1)' + \epsilon \cdot (\mathbf{c}', 0)'$ ,  $0 < \epsilon < \bar{\epsilon}$ , where  $\bar{\epsilon}$  is a small positive constant.

#### Input:

- a vector  $\mathbf{c} \in \mathbb{R}^d$  of coefficients of the goal function,
- a combined sample from  $\tilde{\mathbf{A}}$  and  $\tilde{\mathbf{b}} : \begin{bmatrix} \mathbf{a}^{d,i} \\ 1 - b_j^i \end{bmatrix}_{j=1 \dots K}, i = 1, \dots, n$ ,
- a distortion *risk measure*  $\mu^K$ , given by a name or an explicit weight vector.

#### Output:

- A part of the convolution set  $\mathcal{G}$  that includes the optimal solution,
- the *optimal solution*  $\mathbf{x}^*$ .

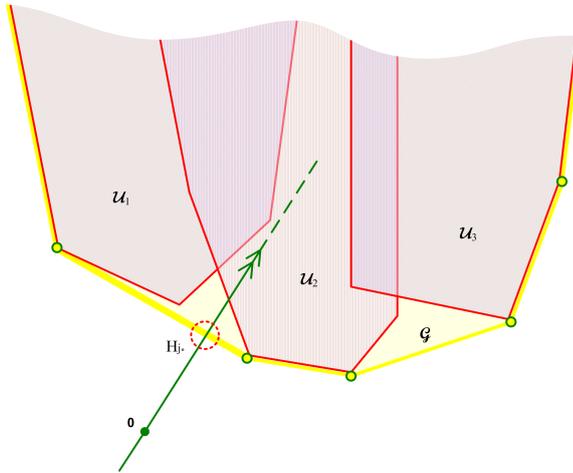
*Steps:*

- A. Construct the individual uncertainty sets:
  - a. Determine, by the algorithm of Bazovkin and Mosler (2012),  $\{\mathcal{U}_j\}_{j=1\dots K}$ , i.e. the WM regions  $\{D_{\mathbf{w}_{\alpha_j}}\}_{j=1\dots K}$  of each random vector  $\begin{bmatrix} \mathbf{a}^j \\ 1-b_j \end{bmatrix}$  having an empirical distribution on  $\begin{bmatrix} \mathbf{a}^{j1} & \dots & \mathbf{a}^{jn} \\ 1-b_j^1 & \dots & 1-b_j^n \end{bmatrix}$ ,  $j = 1 \dots K$ .
  - b. Add the uncertainty set for the  $(K+1)$ -th deterministic constraint in (26b):  $\mathcal{U}_{K+1} = \{(\mathbf{0}', 1)'\}$ .
  - c. For each nonnegativity restriction  $x_j \geq 0$ , add  $d$  point-sets  $\mathcal{N}_j = \{(\mathbf{e}'_j, 1)'\}$  to the uncertainty sets, where  $\mathbf{e}_j$  is the  $j$ -th unit vector in  $\mathbb{R}^d$ .
- B. Calculate the convolution set  $\mathcal{G}$  represented by its facets  $\{(\mathbf{n}_j; d_j)\}_{j \in G}$ :
  - a. Take the representation of  $\{\mathcal{U}_i\}_{i=1\dots K}$  by their vertices and put them into the same space after having rescaled them according to (24).
  - b. Calculate the convex hull of the set using the standard *quickhull* (Barber et al, 1996) or some *divide-and-conquer* algorithm (Grünbaum, 2003).
- C. Impose the optimization ordering on the space of parameters, creating the dual representation of the optimization vector  $\mathbf{c}$ . It is a line  $\varphi$  connecting the origin  $(\mathbf{0}', 0)'$  and the point  $(\epsilon \cdot \mathbf{c}', 1 - \epsilon)'$ , with a small  $\epsilon > 0$ .
- D. Search for a facet  $H_{j_*}$  of  $\mathcal{G}$  that is intersected by  $\varphi$  (see Figure 4). Its dual defines the sought-for optimal solution  $\mathbf{x}^*$ :
  - a. Define a set of facets  $\mathcal{G}_{sel}$  to be analysed: This may be either  $\mathcal{G}$  itself or its part where the intersection is expected;  $\mathcal{G}_{sel} = \{(\mathbf{n}_j, d_j) : j \in J_{sel}\}$ .
  - b. Take some  $\mathbf{u} = \lambda \mathbf{c}, \lambda \geq 0$ , outside the augmented  $\mathcal{G}$ . Find the  $j_* = \arg \max_j \left\{ \frac{d_j}{\mathbf{u}' \mathbf{n}_j} \right\}_{j \in J_{sel}}$ .
  - c.  $\mathbf{x}^* = \frac{1}{d_{j_*}} \cdot \mathbf{n}_{j_*}$  is the optimal *solution*.
  - d. If there is no intersection, the *solution is at infinity*.
  - e. If  $(\mathbf{0}', 1)' \in \mathcal{G}$ , there is *no solution*.

For efficient calculation of the intersection of  $\varphi$  with  $\mathcal{G}$  we can apply an approximative procedure which converges to the precise solution. The procedure finds a facet of some scaled  $\mathcal{U}_i$  that is closest from outside to the sought-for facet of  $\mathcal{G}$ . Obviously, if the facet is part of some uncertainty set, the obtained solution is optimal.

### 3.4 Complexity

The complexity of the algorithm is firstly determined by the routine for the convex hull. Before analyzing the general algorithm, we take a look at its deterministic counterpart. In solving a deterministic LP we have to calculate the convex hull of  $K+1$  points, which has a worst-case complexity of  $\mathcal{O}(K \log K + K^{\lceil \frac{d}{2} \rceil})$ ; see Chazelle (1993). The naive linear search on the set of  $n$  facets has a complexity of  $\mathcal{O}(nd)$ . Consequently, the worst-case complexity of our algorithm amounts to  $\mathcal{O}(d \cdot K^{\lceil \frac{d}{2} \rceil})$ , which is polynomial. It is well known (e.g. Borgwardt (2001)) that no variant of the simplex method exists that solves an LP with polynomial complexity. Only ellipsoid and interior-point methods (using randomizations) can achieve polynomial complexity.



**Fig. 4** Obtaining the solution using the convex hull computation routine.

For the general algorithm it is very natural to use "divide and conquer" algorithms, which construct the convex hull of the whole data out of convex hulls of subsets of data. Such procedures have best complexities in dimensions 2 and 3, namely  $\mathcal{O}(n \log n)$ . For example, the standard quickhull algorithm has complexity between  $\mathcal{O}(n \log n)$  and  $\mathcal{O}(n^2)$ , depending on the input. However, to our knowledge, such "divide and conquer" algorithms are available in the literature only for dimensions up to 5 (see, e.g., Buckley (1988)).

In fact, we need not calculate the whole convex hull, because we are interested only in the normal to the hyperplane at the intersection (even the actual facet is not interesting for us). That is why we can substantially reduce the complexity by cutting off the region of the intersection by an ellipsoid or a cylinder having axis  $\varphi$ . The worst-case complexity of the general algorithm is also polynomial, however of a high power, which can cause difficulties for large-scale problems.

### 3.5 Robust SLP for generally distributed coefficients

In the previous sections we assumed the random parameters  $\tilde{\mathbf{A}}$  and  $\tilde{\mathbf{b}}$  to follow an empirical distribution based on observations  $\{\mathbf{A}^1, \dots, \mathbf{A}^n\}$  and  $\{\mathbf{b}^1, \dots, \mathbf{b}^n\}$ . Now we want to consider an SLP (1), where  $(\tilde{\mathbf{A}}, \tilde{\mathbf{b}})$  follows a general probability distribution  $P$  and realizations are randomly sampled from this distribution.

Actually, Mosler and Bazovkin (2014) have shown that the individual uncertainty set  $\mathcal{U}_j$  is a consistent estimator of its population counterpart. Convergence is almost surely in the Hausdorff sense, which is based on the law of large numbers for weighted-mean regions (Dyckerhoff and Mosler (2012)). Here, our general algorithm constructs  $\mathcal{G}_n$  as the convex union of individual uncertainty sets, which, obviously, also converges almost surely in the Hausdorff sense to  $\mathcal{G}$ :

**Proposition 4**  $\mathcal{G}_n$  converges to  $\mathcal{G}$  almost surely in the Hausdorff sense.

Also the cutting point, where the line  $\varphi$  hits the convolution set  $\mathcal{G}$ , is consistently estimated by our algorithm. A potential complication lies in the fact that the surface of  $\mathcal{G}$  is, in general, not smooth. That is why the optimal solution  $\mathbf{x}^*$ , which, obviously, is defined by the tangent hyperplane at the cutting point can be ambiguous. However, even in such situations, the algorithm automatically selects a unique facet of  $\mathcal{G}$  determining  $\mathbf{x}^*$ .

#### 4 Conclusion and an application to supervised learning

A new geometric algorithm is proposed for robust linear optimization under distortion risk constraints. The algorithm constructs an uncertainty set in the parameter space, which measures the risk arising from non-deterministic parameters in the original linear constraints. The randomness may affect the coefficient matrix  $\mathbf{A}$  as well as the right hand side  $\mathbf{b}$ . In our setting a multivariate coherent distortion risk measure is applied to the joint distribution of the parameters. This results in uncertainty sets for each single constraint, which are so called weighted-mean trimmed regions. A multi-constraint uncertainty set then comes out as the convex hull of the union of rescaled single-constraint uncertainty sets. It is determined by calculating the relevant parts of weighted-mean trimmed regions, which is done by the algorithm of Bazovkin and Mosler (2012). Note that the uncertainty set needs not be determined from a sample; alternatively, it can be introduced explicitly by the optimizer. In this case the algorithm starts with step C.

The algorithm can be applied to multi-constraint as well as single-constraint problems. Also, as a special case, deterministic linear optimization problems are solved by the algorithm. To cope with substitution in evaluating the violation of different constraints, a variant of the model is introduced, which is mentioned as *robust polyhedral optimization*.

At this point it is reasonable to compare the present framework with that of chance-constrained problems. Such SLPs (5) with individual constraints are extremely critical to distributional assumptions: in most cases the program turns out to be non-convex (see, e.g., Kall and Mayer (2010)) and computationally intractable. Plausible results are recently obtained only for elliptically distributed random coefficients. If we have a joint chance-constraint (4) the difficulty increases. A straight-forward approach, which distributes the common violation probability equally among the individual constraints, tends to give poor results, especially if the constraints are stochastically dependent. An approximative solution based on the Bonferroni inequality has been improved by Chen and Sim (2009) and Chen et al (2010), who propose more efficient bounds for the individual probabilities of constraint violation using results from order statistics. But altogether these approximative methods still lack strong interpretation and universality. In contrast, our approach leads to a natural decomposition of (15), while remaining jointly constrained. Any stochastic dependency among parameters (and, thus, constraints) is completely feasible.

A powerful flexibilization of our model consists in varying the  $\alpha$ 's in (20), if, for instance, we are using the expected shortfall\*. This is practical, since some constraints can be more tolerant to violations while others, on the contrary, are rather strict or even exact (e.g., a non-negativity requirement).

<sup>5</sup> Flexible alphas

We conclude the investigation with an efficient application of our optimization model and algorithm to classification problems. Our procedure can be applied to supervised machine learning as a robust alternative to the support vector machine (SVM). The basic problem is: Two classes of points are given in the Euclidean  $d$ -space  $Q_1 = \{x_1, \dots, x_{n_1}\}$  and  $Q_2 = \{y_1, \dots, y_{n_2}\}$ . A rule has to be constructed by which any new point  $x$  is classified to one of those classes  $Q_1$  and  $Q_2$ . The classical SVM of Vapnik (1998) determines a hyperplane that discriminates the two classes linearly in a higher-dimensional space and serves as a separator for classifying new points. Technically, this approach results in a convex *quadratic program*. To tackle the problem in a robust way, mostly methods of replacing each point by its neighbourhood are proposed in the literature; see e.g. Ben-Tal et al (2009). In contrast, we consider no single points of the training classes as uncertain, but the whole classes, and observe the points as a sample of the random variables defining the classes. Besides having a better interpretation, we obviously can expect getting less constraints. In fact, it turns out that the robust SVM can be represented as a robust linear program with two risk constraints. To achieve these representation, we start with the following linear program:

$$C \rightarrow \max_{\omega, C}$$

$$s.t. \begin{cases} \omega' \mathbf{x}_i + b \geq C, & \mathbf{x}_i \in Q_1, \\ \omega' \mathbf{y}_j + b \leq -C, & \mathbf{y}_j \in Q_2. \end{cases} \quad (\text{SVM1})$$

Generally, the solution of (SVM1) does not coincide with the solution of the usual *quadratic* program. Like Vapnik's SVM, (SVM1) produces a central separating hyperplane that lies between the classes. However, it does not necessarily minimize the Euclidean distances between support vectors, as the Euclidean distance is not the only possible criterion here. Note that the Euclidean distance is not easily interpreted when the data have been transformed into a higher-dimensional feature space, as it is done by SVM.

To control the quality of our proposed solution, we rewrite the model (SVM1) as a stochastic linear program:

$$\mathbf{0}' \cdot \mathbf{y} - 1 \cdot z \rightarrow \min_{\mathbf{y}, z}$$

$$s.t. \begin{cases} \tilde{\mathbf{a}}_1' \cdot \mathbf{y} - 1 \cdot z \geq -1, \\ (-\tilde{\mathbf{a}}_2)' \cdot \mathbf{y} - 1 \cdot z \geq 1, \end{cases} \quad (\text{SVM2})$$

where  $\tilde{\mathbf{a}}_1 \sim Q_1, \tilde{\mathbf{a}}_2 \sim Q_2$ .

$\tilde{\mathbf{a}} \sim Q$  means that  $\tilde{\mathbf{a}}$  has an empirical distribution on the finite set  $Q$ . Following our approach, we next remove the negative values in  $\mathbf{b} = (-1 \ 1)'$ . After applying the transformation of Proposition 3, we get:

$$[\mathbf{0}' \ \epsilon \ 1 - \epsilon] (\mathbf{y}' \ z \ y_{d+1})' \rightarrow \min_{\mathbf{y}, z, y_{d+1}}$$

$$s.t. \begin{bmatrix} \tilde{\mathbf{a}}_1' & -1 & 2 \\ -\tilde{\mathbf{a}}_2' & -1 & 0 \\ \mathbf{0}' & 0 & 1 \end{bmatrix} \cdot \begin{pmatrix} \mathbf{y} \\ z \\ y_{d+1} \end{pmatrix} \geq \mathbf{1}, \quad (\text{SVM3})$$

where  $\tilde{\mathbf{a}}_1 \sim Q_1, \tilde{\mathbf{a}}_2 \sim Q_2$ .

The origin  $\mathbf{0}$  must not be situated between the classes, otherwise we may obtain an infinite solution. We extend this application as follows: To control the width of the margin we make  $\mathbf{b}$  stochastic (instead of fixing it at  $(-1 \ 1)'$ ). The more uncertain  $\mathbf{b}$ , the wider is the margin of the separating hyperplane.

A soft margin is introduced as usual; see Vapnik (1998). However, in contrast to the classical approach, the additional margin variable  $\xi$  appears to be particularly natural in our stochastic linear program:

$$\begin{aligned} & \mathbf{0}' \cdot \mathbf{y} - 1 \cdot z + M \cdot \xi \rightarrow \min_{\mathbf{y}, z, \xi} \\ s.t. & \begin{cases} \tilde{\mathbf{a}}_1' \cdot \mathbf{y} - 1 \cdot z + \xi \geq -1, \\ (-\tilde{\mathbf{a}}_2)' \cdot \mathbf{y} - 1 \cdot z + \xi \geq 1, \end{cases} & (\text{SVM-soft}) \\ & \text{where } \tilde{\mathbf{a}}_1 \sim Q_1, \tilde{\mathbf{a}}_2 \sim Q_2. \end{aligned}$$

Our soft-margin model has the advantage that, if we are unsure about the proper class labels of the training points, we can introduce a random coefficient for  $\xi$  that describes the level of certainty in labeling. It is also clear that we can use the kernel trick here, because the inner product and the induced norm are sufficient for all calculations in the algorithm.

Further, our robust optimization model and the new algorithm can be applied in many other fields of operations research. In particular, it is well suited to formalize problems in supply chain management, like the management of an inventory. This will be the topic of future work.

**Acknowledgements** The detailed suggestions of two referees have much improved the clarity and readability of the paper; they are gratefully acknowledged. Pavel Bazovkin was partly supported by a grant of the German Research Foundation (DFG).

**There are 5 comments in the text!**

## References

- Barber CB, Dobkin DP, Huhdanpaa H (1996) The quickhull algorithm for convex hulls. *ACM Trans Math Softw* 22:469–483
- Bazovkin P (2014) A geometrical framework for portfolio optimization, working paper, University of Cologne
- Bazovkin P, Mosler K (2012) An exact algorithm for weighted-mean trimmed regions in any dimension. *Journal of Statistical Software* 47(13)
- Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math Programming Ser A* 88:411–424
- Ben-Tal A, Ghaoui E, Nemirovski A (2009) *Robust Optimization*. Princeton University Press
- Ben-Tal A, den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59:341–357
- Bertsimas D, Brown DB (2009) Constructing uncertainty sets for robust linear optimization. *Operations Research* 57:1483–1495
- Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev* 53:464–501

- Borgwardt K (2001) *Optimierung, Operations Research, Spieltheorie*. Birkhäuser
- Buckley CE (1988) A divide-and-conquer algorithm for computing 4-dimensional convex hulls. In: *Proceedings International Workshop on Computational Geometry on Computational Geometry and its Applications*, Springer-Verlag, New York, pp 113–135
- Cascos I, Molchanov I (2007) Multivariate risks and depth-trimmed regions. *Finance and Stochastics* 11:373–397
- Chazelle B (1993) An optimal convex hull algorithm in any fixed dimension. *Discrete and Computational Geometry* 10:377–409
- Chen W, Sim M (2009) Goal-driven optimization. *Operations Research* 57:342–357
- Chen W, Sim M, Sun J, Teo CP (2010) From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research* 58:470–485
- Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 58:595–612
- Delbaen F (2002) Coherent risk measures on general probability spaces. In: *Advances in Finance and Stochastics*, Springer-Verlag, Berlin, pp 1–37
- Dyckerhoff R, Mosler K (2011) Weighted-mean trimming of multivariate data. *Journal of Multivariate Analysis* 102:405–421
- Dyckerhoff R, Mosler K (2012) Weighted-mean trimming of a probability distribution. *Statistics and Probability Letters* 82:318–325
- Fabozzi FJ, Huang D, Zhou G (2010) Robust portfolios: contributions from operations research and finance. *Annals of Operations Research* 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 235:471–483
- Grünbaum B (2003) *Convex Polytopes*, 2nd edn. Springer-Verlag
- Huber PJ (1981) *Robust statistics*. Wiley, New York
- Kall P, Mayer J (2010) *Stochastic Linear Programming. Models, Theory, and Computation*, 2nd edn. Springer-Verlag, New York
- Mosler K, Bazovkin P (2014) Stochastic linear programming with a distortion risk constraint. *OR Spectrum* 36(4):949–969
- Natarajan K, Pachamanova D, Sim M (2009) Constructing risk measures from uncertainty sets. *Operations Research* 57:1129–1141
- Rüschendorf L (2013) *Mathematical Risk Analysis*. Springer-Verlag
- Vapnik VN (1998) *Statistical Learning Theory*. John Wiley, New York