Robust optimization with ambiguous stochastic constraints under mean and dispersion information

Krzysztof Postek
CentER and Department of Econometrics and Operations Research, Tilburg University, k.postek@tilburguniversity.edu

Aharon Ben-Tal
Department of Industrial Engineering and Management, Technion - Israel Institute of Technology;
Shenkar College, Ramat Gan;
CentER Extramural Fellow, Tilburg University, abental@ie.technion.ac.il

Dick den Hertog, Bertrand Melenberg
CentER and Department of Econometrics and Operations Research, Tilburg University,
d.denhertog@tilburguniversity.edu,b.melenberg@tilburguniversity.edu

In this paper we consider ambiguous stochastic constraints under partial information consisting of means and dispersion measures of the underlying random parameters. Whereas the past literature used the variance as the dispersion measure, here we use the mean absolute deviation from the mean (MAD). This makes it possible to use the 1972 result of Ben-Tal and Hochman (BH) in which *tight* upper and lower bounds on the expectation of a convex function of a random variable are given. First, we use these results to treat ambiguous *expected feasibility* constraints. This approach requires, however, the independence of the random variables and, moreover, may lead to an exponential number of terms in the resulting robust counterparts. We then show how upper bounds can be constructed that alleviate the independence restriction and require only a linear number of terms, by exploiting models in which random variables are linearly aggregated. Moreover, using the BH bounds we derive new safe tractable approximations of *chance constraints*. In a numerical study, we demonstrate the efficiency of our methods in solving stochastic optimization problems under mean-MAD ambiguity.

Key words: robust optimization; ambiguity; stochastic programming; chance constraints

1. Introduction

Consider an optimization problem with a constraint

$$f(\mathbf{x}, \mathbf{z}) \leq 0,$$

where $\mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}}$ is the decision vector, $\mathbf{z} \in \mathbb{R}^{n_{\mathbf{z}}}$ is an uncertain parameter vector, and $f(\cdot, \mathbf{z})$ is assumed to be convex for all $\mathbf{z}$. There are three principal ways to address such constraints. One of them is *Robust Optimization*. In this approach, $\mathcal{U}$ is a user-provided convex compact uncertainty set and the constraint is to hold for all $\mathbf{z} \in \mathcal{U}$, i.e., $\mathbf{x}$ is robust feasible if:

$$\sup_{\mathbf{z} \in \mathcal{U}} f(\mathbf{x}, \mathbf{z}) \leq 0. \tag{1}$$

The key issue in this approach is to reformulate (1) to an equivalent, computationally tractable form (Ben-Tal and Nemirovski (1998), Ben-Tal et al. (2009, 2015)).

In the other approaches, which go under the name of *Distributionally Robust Optimization* (DRO), $\mathbf{z}$ is a *random* parameter vector whose distribution $\mathbb{P}_{\mathbf{z}}$ belongs to a set $\mathcal{P}$ (the so-called *ambiguity set*). A typical example for $\mathcal{P}$ is a set of all distributions with given values of the first two moments. In such a setting, there are two principal constraint types: the *worst-case expected feasibility constraints*:

$$\sup_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\mathbf{z}}} f(\mathbf{x}, \mathbf{z}) \leq 0, \tag{2}$$

and *chance constraints*:

$$\sup_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{P}_{\mathbf{z}} \left(f(\mathbf{x}, \mathbf{z}) > 0 \right) \leq \epsilon. \tag{3}$$

For constraint (2) the key challenge is, for a given ambiguity set $\mathcal{P}$, to obtain a tractable exact form of the worst-case expectation, or a good upper bound. Constraint (2) is also used in the construction of safe approximation of the ambiguous chance constraint (3), where by a *safe approximation* is meant a system $\mathcal{S}$ of computationally tractable constraints, such that $\mathbf{x}$ feasible for $\mathcal{S}$ is also feasible for constraint (3).

In this paper, we consider problems with ambiguity sets consisting of distributions having given mean-dispersion measures. The literature of this type of problem started with the paper by Scarf (1958). Under mean-variance information, Scarf derived the exact worst-case expectation formula for a single-variable piecewise linear objective function used in the newsvendor problem. Later, his result has been extended to more elaborate cases of inventory and newsvendor problems by, e.g., Gallego (1992), Gallego et al. (2001), and Perakis and Roels (2008). In a paper by Popescu (2007), it has been proved that for a wide class of increasing concave utility functions the problem of maximizing the worst-case expected utility under mean-variance distributional information reduces to solving a parametric quadratic optimization problem.

In a broader context, the idea of constructing an approximation of the worst-case expectation of a given function by a discrete distribution falls into the category of bounding strategies based on distributional approximation, see Edirisinghe (2011) who provide a broad overview of results obtained in this field. Rogosinsky (1958) and Karr (1983) show that the worst-case probability distributions corresponding to the moment problems are discrete, with a number of points corresponding to the number of moment conditions. Shapiro and Kleywegt (2002) develop a duality theory for stochastic programs where the saddle points are also vectors of discrete probabilities. Dupačová (1966) and Gassmann and Ziemba (1986) give convex upper bounds on the expectation of a convex function under first-moment conditions over a polyhedral support, based on the dual of the related moment problem. Birge and Wets (1987) and Edirisinghe and Ziemba (1994a) extend this approach to distributions with unbounded support. Dulá (1992) provide a bound for the expectation of a simplicial function of a random vector using first moments and the sum of all variances. His approach is extended by Kall (1991) demonstrating that the related moment problems can be solved using nonsmooth optimization problems with linear constraints. Other notable works in this field include Frauendorfer (1988), Edirisinghe and Ziemba (1992), and Edirisinghe and Ziemba (1994b). For a general discussion we refer the reader to Edirisinghe (2011) and references therein.

Despite numerous works, to the best of our knowledge, no *closed-form tight upper bounds* are known on the expectations of *general convex functions* under mean-variance information. Surprisingly, already in 1972 a result of Ben-Tal and Hochman (1972) (from now on referred to as BH) was available, providing exact upper and lower bounds on the expectation of a general convex $f(\mathbf{x}, \cdot)$ for the case where $\mathcal{P}$ consists of all distributions of componentwise independent $\mathbf{z}$ with known supports, means, but with another dispersion measure: the *mean absolute deviation from the mean (MAD)*.

The information about the supports, means, and MADs of the z_i's, can easily be obtained from past data, making the method suitable for data-driven settings. Moreover, the MAD has several desirable properties from an application's point of view, for example, its suitability to situations when the deviations of z_i are small. Some properties of the MAD and exact formulas for its value for several known distributions are given in Appendix D. Several practical advantages of using the MAD as a dispersion measure are given in the paper by El Amir (2012) and references therein.

Our contributions can be summarized as follows:

- We propose a new method of optimizing the *exact* worst-case-expected performance under mean-MAD information in problems involving constraints (2) with both convex and concave $f(\mathbf{x}, \cdot)$, or a mixture of these.
- We derive new safe tractable approximations of chance constraint (3) under mean-MAD information. These results apply to the case of independent random variables z_i.
- In problems where the random variables are linearly aggregated, i.e.

$$f(\mathbf{x}, \mathbf{a}^T\mathbf{z}) \text{ or } f(\mathbf{a}(\mathbf{x})^T\mathbf{z}),$$

  we derive upper bounds which do not require the independence of the random variables and which are computationally tractable.

- The above results are used to treat problems in which convexity in the uncertain parameter (usually an intractable case in classical RO) appears. This occurs, for example as a result of applying linear decision rules or when the uncertainty is due to implementation error.
- Moreover, in case of existence of multiple RO-optimal solutions, we show that the proposed approach can be used as a second-stage method of improving the *average* performance of the RO solutions.
- Our numerical study shows that minimization of $\sup_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\mathbf{z}}} f(\mathbf{x}, \mathbf{z})$ over $\mathbf{x}$ using mean-MAD information can also lead to a downward shift of $\inf_{\mathbb{P}_{\mathbf{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\mathbf{z}}} f(\mathbf{x}, \mathbf{z})$, compared to the solution obtained by classical RO.

We mention that there are alternative ways of specifying the set $\mathcal{P}$, for example, as sets of distributions deviating from a known distribution according to a certain distance measure (see for example Ben-Tal et al. (2013)). For a broad overview of types of ambiguity sets we refer the reader to Postek et al. (2015) and Hanasusanto et al. (2015). There are cases of such alternative settings for which exact reformulations are possible both for the expected feasibility constraints and for the chance constraints. For the first case, examples of such settings are given in Ben-Tal et al. (2013), Wiesemann et al. (2014) and Esfahani and Kuhn (2015). Settings in which exact reformulations are possible for individual chance constraints are given, for example, in Calafiore and El Ghaoui (2006) and Jiang and Guan (2015).

Approximation results or efficient solution methods when the components of the random vector $\mathbf{z}$ are not independent, are obtained for limited classes of function $f(\mathbf{x}, \mathbf{z})$ in Delage and Ye (2010), Goh and Sim (2010), and Zymler et al. (2013). Chen et al. (2007) propose to use so-called forward and backward deviations as characteristics of the moment generating functions of random variables to approximate chance constraints.

Wiesemann et al. (2014) have recently introduced a class of quite general ambiguity sets for which they derive computationally tractable counterparts of (2) for specific cases of $f(\mathbf{x}, \cdot)$. However, in their framework the components of $\mathbf{z}$ are unrestricted in their dependence, and taking their independence into account is not straightforward. In Appendix C, we illustrate the marked difference between theirs and our robust counterparts when $f(\mathbf{x}, \mathbf{z}) = \exp(\mathbf{x}^T \mathbf{z})$ where, without the assumption of independence, one has to reformulate a robust constraint that is *strictly convex* in the uncertain parameter. In Section 4.4 we provide a numerical comparison of their and our method on an example where both approaches can be applied without the independence assumption.

The remainder of the paper is structured as follows. In Section 2, we describe the mean-MAD results of BH, providing statistical background on estimation of the relevant parameters. In Section 3, we show how the mean-MAD results can be used to optimization problems involving stochastic constraints (2), including numerical examples. In Section 4 we outline the result for the case of linearly aggregated random variable. Section 5 includes new results on safe tractable approximations of chance constraints (3), illustrated also with a numerical study. Section 6 concludes the paper.

2. Bounds on the expectation of a convex function of a random variable

2.1. Introduction

In this section we introduce the results of BH on exact upper and lower bounds on the expected value of a convex function of a componentwise independent $\mathbf{z} = (z_1, \ldots, z_{n_\mathbf{z}})^T$. From now on we drop the subscript $\mathbf{z}$ from $\mathbb{P}_\mathbf{z}$ and the probability distribution applies to $\mathbf{z}$. The pieces of partial distributional information on z_i's constituting the ambiguity sets in BH are:

(i) support including intervals: $\text{supp}(z_i) \subseteq [a_i, b_i]$, where $-\infty < a_i \leq b_i < \infty, i = 1, \ldots, n_\mathbf{z}$. BH show also that their bounds hold in cases where $a_i = -\infty$ and/or $b_i = +\infty$. We illustrate this in Remark 3. In the remainder of the paper, however, we concentrate on the bounded case, with RO applications in mind,

(ii) means: $\mathbb{E}_\mathbb{P}(z_i) = \mu_i$,

(iii) mean absolute deviations from the means (MAD): $\mathbb{E}_\mathbb{P}|z_i - \mu_i| = d_i$. The MAD is known to satisfy the bound (BH, Lemma 1):

$$0 \leq d_i \leq d_{i,\max} = \frac{2(b_i - \mu_i)(\mu_i - a_i)}{(b_i - a_i)}, \quad i = 1, \ldots, n_\mathbf{z}, \tag{4}$$

(iv) probabilities of z_i's being greater than or equal to μ_i: $\mathbb{P}(z_i \geq \mu_i) = \beta_i$. For example, in the case of continuous symmetric distribution of z_i we know that $\beta_i = 0.5$. This quantity is known to satisfy the bounds:

$$\frac{d_i}{2(b_i - \mu_i)} = \underline{\beta}_i \leq \beta_i \leq \overline{\beta}_i = 1 - \frac{d_i}{2(\mu_i - a_i)}, \quad i = 1, \ldots, n_\mathbf{z}. \tag{5}$$

Using these building blocks, we define two types of ambiguity set $\mathcal{P}$:

- the (μ, d) *ambiguity set*, consisting of the distributions with known (i), (ii), and (iii) for each z_i:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P}: \quad \text{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_\mathbb{P}(z_i) = \mu, \quad \mathbb{E}_\mathbb{P}|z_i - \mu_i| = d_i, \quad \forall i, \quad z_i \perp\!\!\!\perp z_j, \quad \forall i \neq j\}, \tag{6}$$

where $z_i \perp\!\!\!\perp z_j$ denotes the stochastic independence of components z_i and z_j,

- the (μ, d, β) *ambiguity set*, consisting of the distributions with known (i), (ii), (iii), and (iv) for each z_i:

$$\mathcal{P}_{(\mu,d,\beta)} = \left\{\mathbb{P}: \quad \mathbb{P} \in \mathcal{P}_{(\mu,d)}, \quad \mathbb{P}(z_i \geq \mu_i) = \beta_i, \quad \forall i\right\}. \tag{7}$$

In the following, we present the results of BH on $\max\limits_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_\mathbb{P} f(\mathbf{z})$, $\max\limits_{\mathbb{P} \in \mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_\mathbb{P} f(\mathbf{z})$ and $\min\limits_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_\mathbb{P} f(\mathbf{z})$, $\min\limits_{\mathbb{P} \in \mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_\mathbb{P} f(\mathbf{z})$, where $f: \mathbb{R}^{n_\mathbf{z}} \to \mathbb{R}$ is convex. We note that in the case of concave $f(\cdot)$ the upper bounds become lower bounds and vice versa.

2.2. One-dimensional z

We begin with the simpler and more illustrative case of one-dimensional random variable z. For that reason, we drop the subscript i.

Upper bounds. BH shows that:

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = p_1 f(a) + p_2 f(\mu) + p_3 f(b), \tag{8}$$

where:

$$p_1 = \frac{d}{2(\mu-a)}, \quad p_2 = 1 - \frac{d}{2(\mu-a)} - \frac{d}{2(b-\mu)}, \quad p_3 = \frac{d}{2(b-\mu)}. \tag{9}$$

Hence, the worst-case distribution is a three-point distribution on $\{a,\mu,b\}$. The same bound holds for the (μ,d,β) ambiguity set.

Remark 1. A special case of (9) is the upper bound on $f(z)$ when only the interval $[a,b]$ and the mean μ are known. Such a bound is known as the Edmundson-Madansky bound (Edmundson 1956, Madansky 1959):

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu)}} f(z) = \frac{b-\mu}{b-a} f(a) + \frac{\mu-a}{b-a} f(b) \text{ where } \mathcal{P}_{(\mu)} = \{\mathbb{P}: \quad \text{supp}(z)\subseteq[a,b], \quad \mathbb{E}_{\mathbb{P}} z = \mu\}. \tag{10}$$

Indeed, inserting the biggest possible value of MAD (see (4)) equal to $d_{\max} = 2(b-\mu)(\mu-a)/(b-a)$ into (9) yields the probability of outcome μ equal to 0. ■

Lower bounds. To obtain a closed-form lower bound on $\mathbb{E}_{\mathbb{P}} f(z)$, additional information is needed in the form of the parameter β. Then, it holds that:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(z) = \beta f\left(\mu + \frac{d}{2\beta}\right) + (1-\beta) f\left(\mu - \frac{d}{2(1-\beta)}\right). \tag{11}$$

In case β is not known, BH shows that:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = \min_{\underline{\beta}\le\beta\le\overline{\beta}} \left\{ \beta f\left(\mu + \frac{d}{2\beta}\right) + (1-\beta) f\left(\mu - \frac{d}{2(1-\beta)}\right)\right\}, \tag{12}$$

where the minimization over β is a convex problem in β and for a strictly convex $f(\cdot)$ there is a unique optimal solution.

Remark 2. In case of no knowledge about d, the lower bound is obtained at $d^* = 0$, which corresponds to the well-known Jensen bound (Jensen 1906). ■

Remark 3. In case where $a = -\infty$ and/or $b = +\infty$, bounds can still be obtained under additional conditions, namely that the limits $\lim_{t\to\pm\infty} f(t)/t$ exist and are finite, with the '+' corresponding to

$b=+\infty$, and the '$-$' corresponding to $a=-\infty$. We illustrate this on the example $a\in\mathbb{R}, b=+\infty$. Assume that $\lim_{t\to+\infty} f(t)/t=\gamma$. We then have:

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = \max_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(z) = \lim_{b\to\infty}\left\{\tfrac{d}{2(\mu-a)}f(a)+\left(1-\tfrac{d}{2(\mu-a)}-\tfrac{d}{2(b-\mu)}\right)f(\mu)+\tfrac{d}{2(b-\mu)}f(b)\right\}$$
$$= \tfrac{d}{2(\mu-a)}f(a)+\left(1-\tfrac{d}{2(\mu-a)}\right)f(\mu)+\tfrac{d}{2}\gamma,$$

and for the lower bound we have:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} f(z)=\frac{d}{2}\gamma+f\left(\mu-\frac{d}{2}\right).$$

The lower bound for the (μ,d,β) ambiguity set is the same as (11). ∎

2.3. Multidimensional z

Upper bounds. For $n_{\mathbf{z}}>1$, the worst-case probability distribution under (μ,d) information is a componentwise counterpart of (9):

$$p_1^i=\frac{d_i}{2(\mu_i-a_i)},\quad p_2^i=1-\frac{d_i}{2(\mu_i-a_i)}-\frac{d_i}{2(b_i-\mu_i)},\quad p_3^i=\frac{d_i}{2(b_i-\mu_i)},\quad i=1,\dots,n_{\mathbf{z}}. \tag{13}$$

The worst-case expectation of $f(\mathbf{z})$ is obtained by applying the bound (8) for each z_i, i.e., by enumerating over all $3^{n_{\mathbf{z}}}$ permutations of outcomes a_i,μ_i,b_i of components z_i. It holds then that (BH):

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})=\sum_{\boldsymbol{\alpha}\in\{1,2,3\}^{n_{\mathbf{z}}}}\prod_{i=1}^{n_{\mathbf{z}}} p^i_{\alpha_i} f(\tau^1_{\alpha_1},\dots,\tau^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}}), \tag{14}$$

where

$$\tau_1^i=a_i,\quad \tau_2^i=\mu_i,\quad \tau_3^i=b_i\quad \text{for}\quad i=1,\dots,n_{\mathbf{z}}. \tag{15}$$

Again, the same upper bound holds for the (μ,d,β) ambiguity set.

Lower bounds. Similar to the one-dimensional case, the closed-form lower bound under (μ,d) information requires known $\boldsymbol{\beta}=(\beta_1,\dots,\beta_{n_{\mathbf{z}}})^T$:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})=\sum_{\boldsymbol{\alpha}\in\{1,2\}^{n_{\mathbf{z}}}}\prod_{i=1}^{n_{\mathbf{z}}} q^i_{\alpha_i} f(\upsilon^1_{\alpha_1},\dots,\upsilon^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}}), \tag{16}$$

where $\underline{\boldsymbol{\beta}}=(\underline{\beta}_1,\dots,\underline{\beta}_{n_{\mathbf{z}}})^T, \overline{\boldsymbol{\beta}}=(\overline{\beta}_1,\dots,\overline{\beta}_{n_{\mathbf{z}}})^T$ and

$$q_1^i=\beta_i,\quad q_2^i=1-\beta_i,\quad \upsilon_1^i=\mu_i+d_i/2\beta_i,\quad \upsilon_2^i=\mu_i-d_i/2(1-\beta_i). \tag{17}$$

If $\boldsymbol{\beta}$ is unknown, the bound is obtained by minimization:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{z})=\inf_{\underline{\boldsymbol{\beta}}\le\boldsymbol{\beta}\le\overline{\boldsymbol{\beta}}}\sum_{\boldsymbol{\alpha}\in\{1,2\}^{n_{\mathbf{z}}}}\prod_{i=1}^{n_{\mathbf{z}}} q^i_{\alpha_i} f(\upsilon^1_{\alpha_1},\dots,\upsilon^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}}). \tag{18}$$

In the multidimensional case, minimization over $\boldsymbol{\beta}$ is a nonconvex problem - it is only convex in β_i when other $\beta_j, j\neq i$ are fixed. A statistical procedure for estimating the parameters μ, d, and β is provided in Appendix A.

3. Robust counterparts of expected feasibility constraints

3.1. Reformulations

In this section we demonstrate how the results of BH can be used to solve problems

$$\text{Val} = \min_{\mathbf{x}} \max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x},\mathbf{z}), \tag{19}$$

where $f(\cdot,\mathbf{z})$ is convex. When $f(\mathbf{x},\cdot)$ is convex and $\mathcal{P} = \mathcal{P}_{(\mu,d)}$, the exact solution of the inner problem is given due to the BH upper bound:

$$g_U(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} p^i_{\alpha_i} f(\mathbf{x}, \tau^1_{\alpha_1}, \dots, \tau^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}}), \tag{20}$$

with $p^i_{\alpha_i}, \tau^i_{\alpha_i}$ defined as in (13) and (15). As we can see, $g_U(\cdot)$ in (20) inherits the convexity in $\mathbf{x}$ from $f(\cdot,\mathbf{z})$ and its functional form depends only on the form of $f(.,\mathbf{z})$.

When $f(\mathbf{x},\cdot)$ is concave and $\mathcal{P} = \mathcal{P}_{(\mu,d,\beta)}$, the exact solution of the inner problem is given due to the BH lower bound:

$$g_L(\mathbf{x}) = \sum_{\boldsymbol{\alpha}\in\{1,2\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} q^i_{\alpha_i} f(\mathbf{x}, \upsilon^1_{\alpha_1}, \dots, \upsilon^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}}), \tag{21}$$

with $q^i_{\alpha_i}, \upsilon^i_{\alpha_i}$ defined by (17).

For the case of convexity (concavity) of $f(\mathbf{x},\cdot)$, a lower bound on $\mathbb{E}_{\mathbb{P}} f(\mathbf{x},\mathbf{z})$ is given by $g_L(\mathbf{x})$ ($g_U(\mathbf{x})$ respectively). The upper and lower bound give a closed interval in which Val lies. That is, for the convex case it is guaranteed that Val lies in the interval $[g_L(\mathbf{x}), g_U(\mathbf{x})]$ and in the concave case in the interval $[g_U(\mathbf{x}), g_L(\mathbf{x})]$. The above result applies also to the case of the ambiguous constraints (2).

There are two difficulties associated with the bounds (20) and (21). One is the computational difficulty: when $n_{\mathbf{z}}$ is large, formulas (20) and (21) include an exponential number of terms. Second is the independence assumption on z_i's: when the independence hypothesis is rejected, the solutions obtained using (20) and (21) might underperform significantly. In Section 4 we discuss a wide class of functions $f(\mathbf{x},\mathbf{z})$ for which both of these difficulties are alleviated. Here, we discuss cases where these difficulties are not present or can be alleviated using existing techniques.

Dimensionality.

- In certain applications the number $n_{\mathbf{z}}$ of random variables is small (less than 10).
- An important special case is when $f(\mathbf{x},\mathbf{z})$ is a sum of functions

$$f(\mathbf{x},\mathbf{z}) = \sum_{j=1}^{n_c} f^{(j)}\left(\mathbf{x},\mathbf{z}^{(j)}\right),$$

where $f^{(j)}(\mathbf{x},\cdot)$ have small numbers n_j of uncertain variables.

- An important special case is the function $f(\mathbf{x},\mathbf{z}) = \exp(\mathbf{x}^T\mathbf{z})$. Upper bounds on moment generating functions $\mathbb{E}\exp(\mathbf{x}^T\mathbf{z})$ are a key tool in constructing safe tractable approximations of chance constraints. As we show in Section 5, the properties of the $\exp(\cdot)$ allow for a simple, closed-form formula for its worst-case expectation under (μ, d) information and for which the number of terms is linear in $n_{\mathbf{z}}$.
- If the dimensionality remains an issue, for problems with linear and piecewise linear $f(\mathbf{x},\cdot)$ one can use, for example, the Stochastic Decomposition method (Higle and Sen 1996) where scenarios (in our case support points) are added iteratively until the current model is a good enough approximation of the original model. In cutting-plane methods, the verification of the ambiguous constraint can exploit the tree-structure of the worst-case distribution support. In this tree structure, each outcome of z_1 leads to 3 (or 2 for the concave case) outcomes of z_2, each of these leads to another 3 outcomes of z_3 etc. Then, one can determine if the constraint holds already after investigating the first few layers of the tree, which may lead to a verification of much less than all $3^{n_{\mathbf{z}}}$ scenarios. Other approximate approaches are the sample average approximation (Shapiro et al. 2009) or the scenario reduction technique (Dupačová et al. 2003).

Dependence.

If the random uncertain vector $\mathbf{z}$ contains dependent components, it can be decomposed by means of factor analysis, for example, based on Principal Component Analysis (see Jolliffe (2002)), into linear combinations of a limited number of uncorrelated factors. For example, in a situation of portfolio optimization problem with 25 assets, it is natural to decompose them into 3-4 uncorrelated risk factors (see, for example Baillie et al. (2002)), whose empirical distribution provides information also about their support, means, and MADs. Even though uncorrelatedness can be much weaker than independence, such a technique is often a practical solution.

3.2. The use of the BH bounds in some general applications

In this section we present three cases where the reformulations of the worst-case expected feasibility constraints presented in Section 3.1 can be used.

Average-case enhancement of RO solutions. The first application lies in finding worst-case-optimal solutions with good average-case performance to the following RO problem:

$$\begin{aligned} \min_{\mathbf{x},t}\ & t \\ \text{s.t. } & \sup_{z\in\mathcal{Z}} f(\mathbf{x},\mathbf{z}) \le t, \\ & \sup_{z\in\mathcal{Z}} g_i(\mathbf{x},\mathbf{z}) \le 0, \quad i = 1,\ldots,m. \end{aligned} \tag{22}$$

It happens frequently that there exist multiple optimal solutions to (22), see Iancu and Trichakis (2013), De Ruiter et al. (2016). Whereas the worst-case performance of such solutions is the same, their average-case performance may differ dramatically. For that reason, once the optimal value $\bar{t}$ for (22) is known, a second optimization step may be used to select one of the optimal solutions to provide good average-case behavior. Since the results of BH provide exact bounds on the worst-case expectations, they can be used in such a step. In the following, we describe such a two-step procedure:

1. Solve problem (22) and denote its optimal value by $\bar{t}$.
2. Solve the following problem, minimizing the worst-case expectation of the objective value, with the worst-case value of $f(\mathbf{x},\mathbf{z})$ less than or equal to $\bar{t}$:

$$\begin{aligned} \min_{\mathbf{x},u} \ & u \\ \text{s.t.} \ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x},\mathbf{z}) \leq u \\ & \sup_{z\in\mathcal{Z}} f(\mathbf{x},\mathbf{z}) \leq \bar{t}, \\ & \sup_{z\in\mathcal{Z}} g_i(\mathbf{x},\mathbf{z}) \leq 0, \quad i=1,\ldots,m. \end{aligned} \tag{23}$$

In case of multiple optimal solutions to (22), the two-step procedure is expected to select the optimal solution with good average-case performance for its focus on the worst-case expectation among the best worst-case solutions. If the uncertainty is present only in the constraints involving functions $g_i(\cdot,\cdot)$, a similar two-step approach can be designed to maximize the worst-case expected slack in the worst-case constraints in (22), see Iancu and Trichakis (2013). We note that following the theory of Iancu and Trichakis (2013), there might exist multiple optimal solutions to (23) and one may need to include another 'enhancement step' to choose among them.

An alternative approach to enhancing robust solutions is to sample a number S of scenarios for $\mathbf{z}$ to find a solution that optimizes the average of the objective value over the sample.[1]. This approach, however, has as shortcoming that the outcome might depend on the choice of sample size S and the sample itself. For that reason, the DRO methods can provide a good alternative to enhancing the quality of RO solutions. In our paper, we test the application of the (μ, d) bounds to enhance average-case performance in an inventory management problem in Section 3.4.

Implementation error. The second application we consider is when the decision variables cannot be implemented with the designed value due to implementation error in the following problem:

$$\begin{aligned} \min_{\mathbf{x},t} \ & t \\ \text{s.t.} \ & f(\mathbf{x}) \leq t, \\ & g_i(\mathbf{x}) \leq 0, \quad i=1,\ldots,m. \end{aligned} \tag{24}$$

[1] As a special case, one can choose only one scenario, corresponding to the nominal values of the uncertain parameters (Iancu and Trichakis 2013)

In case of the existence of an additive implementation error $\mathbf{z}$ the implemented value is $\mathbf{x} = \overline{\mathbf{x}} + \mathbf{z}$, where $\overline{\mathbf{x}}$ is the designed value and $\mathbf{z} = (z_1, \ldots, z_{n_{\mathbf{x}}})^T$ is the error. Then, the problem becomes:

$$\begin{aligned} \min_{\overline{\mathbf{x}},t} \ & t \\ \text{s.t.} \ & \sup_{z \in \mathcal{Z}} f(\overline{\mathbf{x}} + \mathbf{z}) \leq t, \\ & \sup_{z \in \mathcal{Z}} g_i(\overline{\mathbf{x}} + \mathbf{z}) \leq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{25}$$

Since $f(\mathbf{x})$ is convex in $\mathbf{x}$, in (25) the function $f(\overline{\mathbf{x}} + \mathbf{z})$ is convex in $\mathbf{z}$. For that reason, optimization of the worst-case value of the objective function could be difficult, as typically RO techniques rely on the constraint being concave in the uncertain parameter (see Ben-Tal et al. (2009, 2015)). Therefore, optimizing the worst-case values of convex constraints under implementation error is a problem leading to computational intractability, apart from special cases such as linear constraints (see Ben-Tal et al. (2015)) or (conic) quadratic constraints with simultaneously diagonizable quadratic forms defining the constraint and the uncertainty set for the error (see Ben-Tal and den Hertog (2011)).

Because of the above, it may be an alternative to optimize the worst-case expectation of the objective function, for which our DRO method applies under the corresponding distributional assumptions on $\mathbf{z}$, i.e., that the ambiguity set for the distribution of $\mathbf{z}$ is $\mathcal{P}_{(\mu,d)}$. Then, the problem becomes:

$$\begin{aligned} \min_{\overline{\mathbf{x}},t} \ & t \\ \text{s.t.} \ & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\overline{\mathbf{x}} + \mathbf{z}) \leq t, \\ & \sup_{z \in \mathcal{Z}} g_i(\overline{\mathbf{x}} + \mathbf{z}) \leq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{26}$$

The first constraint in (26) is convex in $\mathbf{z}$ and one can apply the reformulation (20). For (26) to be tractable, the functions $g_i(\overline{\mathbf{x}} + \mathbf{z})$ need to be affine in $\mathbf{z}$ or belong to one of the special cases considered in Ben-Tal and den Hertog (2011). Similarly, one can reformulate a problem where multiplicative error occurs, i.e., where $\mathbf{x} = (\overline{x}_1 z_1, \ldots, \overline{x}_{n_{\mathbf{x}}} z_{n_{\mathbf{x}}})^T$.

Convex constraints and linear decision rules. The third application of our DRO approach comes when the constraints of a problem are convex in z as a result of applying linear decision rules. To show how such a situation occurs, we consider a two-stage RO problem:

$$\begin{aligned} \min_{\mathbf{x}_1,\mathbf{x}_2,t} \ & t \\ \text{s.t.} \ & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z}) \leq t \\ & \sup_{z \in \mathcal{Z}} g_i(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z}) \leq 0, \quad i = 1, \ldots, m, \end{aligned} \tag{27}$$

where $\mathbf{x}_1 \in \mathbb{R}^{n_{\mathbf{x}_1}}$ is implemented at before $\mathbf{z}$ is known (time 1) and $\mathbf{x}_2 \in \mathbb{R}^{n_{\mathbf{x}_2}}$ is implemented after $\mathbf{z}$ is known (time 2), i.e. $\mathbf{x}_2 = \mathbf{x}_2(z)$. In such cases, it is possible to define the time-2 decisions as a linear function $\mathbf{x}_2(z) = \mathbf{v} + \mathbf{V}\mathbf{z}$ of the uncertain parameter z (see Ben-Tal et al. (2004)), to provide adjustability of decisions at time 2.[2] The problem is then:

$$\begin{aligned} \min_{\mathbf{x}_1,\mathbf{v},\mathbf{V},t} \ & t \\ \text{s.t.} \ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}_1, \mathbf{v}+\mathbf{V}\mathbf{z}, \mathbf{z}) \le t \\ & \sup_{\mathbf{z}\in\mathcal{Z}} g_i(\mathbf{x}_1, \mathbf{v}+\mathbf{V}\mathbf{z}, \mathbf{z}) \le 0, \quad i = 1,\dots,m. \end{aligned} \tag{28}$$

Since $f(\mathbf{x}_1, \mathbf{x}_2(\mathbf{z}), \mathbf{z})$ is convex in $\mathbf{x}_2$, the first constraint in (28) may also be convex in $\mathbf{z}$. In such a case, a further reformulation of problem (27) can be conducted as in Section 3.2. We combine linear decision rules with (μ, d) information in the inventory problem of Sections 3.3 and 3.4.

3.3. Application 1: Inventory management - average case performance

Introduction. In this section we consider an application of the (μ, d) method to minimization of the average-case costs in inventory management. The main research questions are:

1. How does minimizing the worst-case expectation affect the best-case expectation under the given distributional assumptions?
2. What is the average-case performance of solutions minimizing the worst-case expectation compared to the robust solutions, minimizing the worst-case outcome of the objective values?

To answer them, we adapt the numerical example from Ben-Tal et al. (2005) with a single product and where inventory is managed periodically over T periods. At the beginning of each period t the decision maker has an inventory of size x_t and he orders a quantity q_t for unit price c_t. The customers then place their demands z_t. The retailer's status at the beginning of the planning horizon is given through the parameter x_1 (initial inventory). Apart from the ordering costs, the following costs are incurred over the planning horizon:

- holding cost $h_t \max\{0, x_{t+1}\}$, where h_t are the unit holding costs,
- shortage costs $p_t \max\{0, x_{t+1}\}$, where p_t are the unit shortage costs.

Inventory x_{T+1} left at the end of period T has a unit salvage value s. Also, one must impose $h_T - s \ge -p_T$ to maintain the problem's convexity. Practical interpretation of this constraint is that in the last period it is more profitable to satisfy the customer demand rather than to be left with excessive amount of inventory. The constraints in the model include:

- balance equations linking the inventory in each period to the inventory, order quantity, and demand in the preceding period,
- upper and lower bounds on the order quantities in each period $L_t \le q_t \le U_t$,

[2] One may also use other decision rules. However, we limit ourselves to the analysis of the linear case as the linear decision rules are very often a powerful enough tool, see Bertsimas et al. (2011). Moreover, the (non)convexity of the problem resulting from application of linear decision rules is easy to verify, see Boyd and Vandenberghe (2004).

- upper and lower bounds on cumulative order quantities in each period $\widehat{L}_t \leq \sum_{\tau=1}^{t} q_\tau \leq \widehat{U}_t$.

With ordering decisions $\mathbf{q}(\mathbf{z}) = (q_1, q_2(\mathbf{z}^1), \ldots, q_T(\mathbf{z}^{T-1}))^T$, where $\mathbf{z}^t = (z_1, \ldots, z_t)^T$, the objective function value for a given demand vector $\mathbf{z}$ is

$$f(\mathbf{q}(\mathbf{z}), \mathbf{z}) = -s \max\left\{x_{T+1,0}(\mathbf{z}^T)\right\} + \sum_{t=1}^{T} \left(c_t q_t(\mathbf{z}^{t-1}) + h_t \max\left\{x_{t+1}(\mathbf{z}^t), 0\right\} + p_t \max\left\{-x_{t+1}(\mathbf{z}^t), 0\right\}\right).$$

The optimization problem to be solved is given by the following, two-variant formulation where the minimized quantity is the worst-case value or the worst-case expectation of the objective function:

$$\begin{aligned} \min_{\mathbf{q}(\cdot),u} \ & u \\ \text{s.t. } & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \text{ or } \sup_{\mathbf{z} \in \mathcal{Z}} f(\mathbf{q}(\mathbf{z}), \mathbf{z}) \leq u \\ & x_{t+1}(\mathbf{z}^t) = x_t(\mathbf{z}^{t-1}) + q_t - z_t, \quad t = 1, \ldots, T \\ & L_t \leq q_t(\mathbf{z}^{t-1}) \leq U_t, \quad t = 1, \ldots, T \\ & \widehat{L}_t \leq \sum_{\tau=1}^{t} q_\tau(\mathbf{z}^{\tau-1}) \leq \widehat{U}_t, \quad t = 1, \ldots, T, \end{aligned} \tag{29}$$

where $\mathcal{Z}$ is the uncertainty set for $\mathbf{z}$ and $\mathcal{P}$ is the ambiguity set of probability distributions with support being a subset of $\mathcal{Z}$. The objective function in (29) has the sum-of-maxima form which typically is problematic in RO due to the difficulty of maximizing a convex function. It is of no concern as the BH results only require that the function at hand is convex in the uncertain parameter.

We assume that the uncertainty set $\mathcal{Z}$ is $\mathcal{Z} = \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_T$, where $\mathcal{Z}_t = [a_t, b_t]$, $t = 1, \ldots, T$, which corresponds to $\mathbf{z}$ being a random variable with independent components. The worst-case form of problem (29) has to be solved by enumerating all vertices of the uncertainty set $\mathcal{Z}$. For the worst-case expectation form of (29) we assume that $\mu_t = \frac{a_t + b_t}{2}$, and that $d_t = \mathbb{E}_{\mathbb{P}}|z_t - \mu_t| = \theta(b_t - a_t)$, yielding the following ambiguity set:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P}: \quad \text{supp}(\mathbb{P}) \subset [\mathbf{a}, \mathbf{b}], \quad \mathbb{E}_{\mathbb{P}}\mathbf{z} = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}|\mathbf{z} - \boldsymbol{\mu}| = \mathbf{d}, \quad z_i \perp\!\!\!\perp z_j \quad \forall i \neq j\},$$

where $\mathbf{a} = (a_1, \ldots, a_T)^T$, $\mathbf{b} = (b_1, \ldots, b_T)^T$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)^T$, $\mathbf{d} = (d_1, \ldots, d_T)^T$. The ordering decisions are assumed to be linear functions of the past demand: $q_{t+1}(\mathbf{z}^t) = q_{t+1,0} + \sum_{j=1}^{t} q_{t+1,j} z_j$ and require that $q_{t+1}(\mathbf{z}^t) \geq 0$ for all $\mathbf{z} \in \mathcal{Z}$, for $t = 2, \ldots, T+1$. We solve the following two variants of problem (29):

- RO solution - the objective function in (29) is preceded by $\sup_{\mathbf{z} \in \mathcal{Z}}$,
- (μ, d) solution - the objective function in (29) is preceded by $\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}$.

We conduct an experiment with $T = 6$ and 50 problem instances. We set $\theta = 0.25$, corresponding to the mean absolute deviation of the uniform distribution. The ranges for the uniform sampling of parameters are given in Table 1.

Upper and lower bounds for the expectation of the objective function. We consider now the first research question of this section. For each inventory problem instance and the optimal solution $\overline{\mathbf{q}}(\cdot)$, we compute the following quantities:

Table 1 Ranges for parameter sampling in the inventory experiment.

Parameter	Range	Parameter	Range
a_t	$[0,20]$	x_1	$[20,50]$
b_t	$[a_t, a_t+100]$	L_t	0
c_t, p_t	$[0,10]$	U_t	$[50,70]$
h_t	$[0,5]$	$\hat{L}_t$	0
s	0	$\hat{U}_t$	$0.8\sum_{t=1}^{T} U_t$

Table 2 Results of the inventory management - worst-case costs and ranges for the expectation of the objective over $\mathcal{P}_{(\mu,d,\beta)}$. All numbers are averages.

Objective type	β	Minimum cost	
		RO	(μ,d)
Worst-case value	-	1950	2384
Expectation range	0.25	[1255,1280]	[1004,1049]
Expectation range	0.5	[1223,1280]	[970,1049]
Expectation range	0.75	[1230,1280]	[994,1049]

- the worst-case expected cost under (μ,d) information: $\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\overline{\mathbf{q}}(\mathbf{z}),\mathbf{z})$
- the best-case expected cost $\inf_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\overline{\mathbf{q}}(\mathbf{z}),\mathbf{z})$ with three possibilities for the skewness of the demand distribution, i.e., with $\beta_t = \beta \in \{0.25, 0.5, 0.75\}$, corresponding to left-skewness, symmetry, and right-skewness of the demand distribution in all periods, respectively.

The two values provide us with information about the interval within which the expected objective function value lies under three different assumptions on the parameter β. Additionally, for each solution we compute the worst-case cost $\sup_{\mathbf{z}\in\mathcal{Z}} f(\overline{\mathbf{q}}(\mathbf{z}),\mathbf{z})$ to verify how the minimization of the worst-case expectation affects the worst-case performance of the solution.

Table 2 presents the results. As can be expected, the RO solution yields the best worst-case value of 1950 which is far better than the (μ,d) solution, whose worst-case value is 2384. Rows 2 to 4 show that the (μ,d) solution not only yields better upper bounds on the expected value of the solution, but also leads to an improvement of the best-case expectation for all β. For example, for $\beta = 0.5$ the interval for the expected cost related to the RO solution is given by $[1255, 1280]$, whereas for the (μ,d) solution it is $[970, 1049]$. That means that the worst-case expectation obtained by the (μ,d) solution is better than the worst-case expectation obtained by the RO solution.

Simulation results. We now answer the second research question by conducting a simulation study. Since the solutions are obtained with different objective functions, comparing their average-case performance in a 'fair' way is difficult. We compare their performance using two samples of demand vectors $\hat{\mathbf{z}}$:

- uniform sample - demand scenarios $\hat{\mathbf{z}}$ are sampled from a uniform distribution on $\mathcal{Z}$,

Table 3 Simulation results for the first inventory problem. Numbers in brackets denote the % change compared to the RO solution.

Objective type	Demand sample type	Cost	
		RO	(μ, d)
Objective mean	Uniform sample	**1230**	**994 (-19.6%)**
Objective standard deviation	Uniform sample	157	259 (+65%)
Objective mean	(μ, d) sample	**1246**	**1003 (-19.5%)**
Objective standard deviation	(μ, d) sample	160	265 (+65.6%)

- (μ, d) sample - demand scenarios $\hat{\mathbf{z}}$ are sampled from a distribution $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$. That is, first, a discretized distribution $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$ is sampled using the hit-and-run method. This method is implemented here as follows. For the $[0, 1]$-interval we construct a grid of 50 equidistant points. For a fixed (μ, d) the set of probability masses assigned to these points satisfying the μ and d values is a polytope. We sample 10 probability distributions uniformly from this polytope with the classical hit-and-run method (mixing algorithm) of Smith (1984), where we choose the starting point to be the analytic center of the polytope and we use only every 20th vector sampled with the mixing algorithm. Then, each component of the vector $\hat{\mathbf{z}}$ is sampled randomly from a randomly chosen distribution $\hat{\mathbb{P}}$.

For each instance, we sample 10^4 demand scenarios, with both of the sampling methods. Table 3 presents the results. The averages of the objective function values over the two sample types over all instance are put in bold. For example (row 1), the (μ, d) solutions perform better on average in the uniform sample, with values 994 and 1230, respectively. A similar observation holds for the (μ, d) sample (row 3). In Figure 1 we present a comparison of the empirical cumulative distribution functions of the total costs incurred by the RO and (μ, d) solutions. In both samples we can see that the costs of the RO solution stochastically dominate over the ones from the (μ, d) solutions. Thus, we conclude that the (μ, d) solutions are superior to the RO solution.

3.4. Application 2: Inventory management - enhancement of RO solutions

With the good average-case performance of the (μ, d) solutions in the previous experiment, we investigate now the following question: can the (μ, d) method be used to enhance the average-case performance of RO solutions? That is, is it possible, in cases where there are multiple optimal solutions to the RO problem, to find the worst-case optimal solution that has a better average cost than the initial worst-case optimal solution? To verify this, for each of the problem instances of the previous subsection we apply the two-step procedure of Section 3.2.

We consider three enhancement types, corresponding to three different objective functions:

- (μ, d) enhancement: $\min \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\mathbf{q}(\mathbf{z}), \mathbf{z})$,

Figure 1 Empirical cumulative distribution functions of the total costs of the solutions under the uniform sample (left) and (μ, d) sample (right). Aggregated from all problem instances.

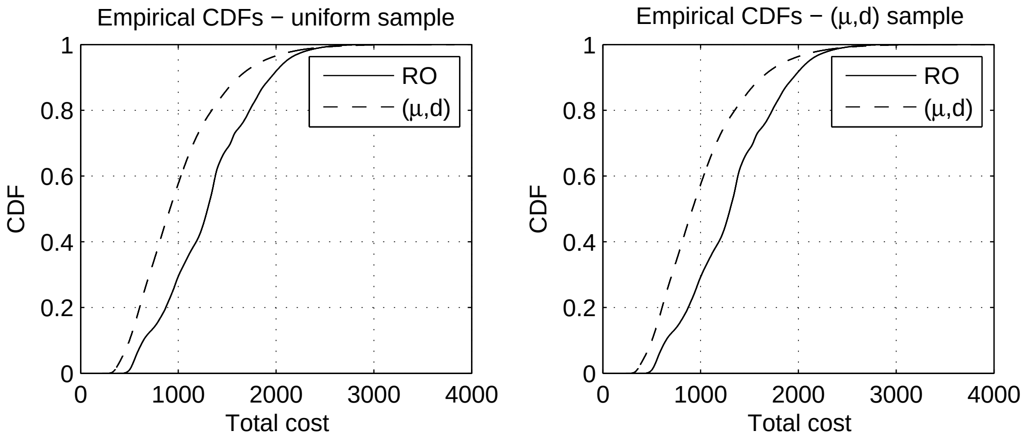


Table 4 Results of the inventory management - enhancement of RO solutions example. All numbers are averages. Numbers in brackets denote the % change compared to the initial solution with no enhancement (first column).

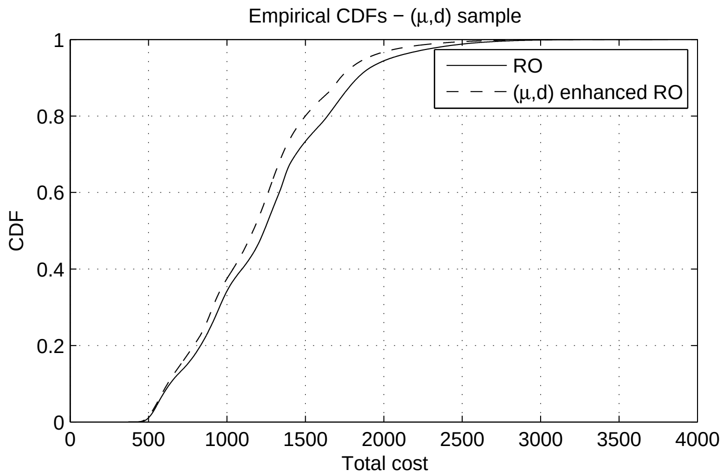
Objective type		Cost			
	Enhancement type	-	(μ, d)	Sample	Nominal
Objective mean	Uniform sample	**1230**	**1168 (-5.04%)**	1168 (-5.04%)	1180 (-4.06%)
Objective standard deviation	Uniform sample	157	158 (+0.63%)	156 (-0.63%)	161 (+2.54%)
Objective mean	(μ, d) sample	**1246**	**1172 (-5.93%)**	1172 (-5.93%)	1184 (-4.97%)
Objective standard deviation	(μ, d) sample	160	161 (+0.62%)	160 (0.00%)	164 (+2.50%)

- sample enhancement: $\min \frac{1}{S}\sum_{j=1}^{S} f(\mathbf{q}(\hat{\mathbf{z}}_j), \hat{\mathbf{z}}_j)$, where $\hat{\mathbf{z}}_j$ are $S = 200$ demand scenarios sampled uniformly from $\mathcal{Z}$,
- nominal enhancement: $\min f(\mathbf{q}(\boldsymbol{\mu}), \boldsymbol{\mu})$ considered by Iancu and Trichakis (2013).

Table 4 presents the results. In the uniform sample (row 1) the (μ, d)-enhanced solution yields an average cost of 1168, compared to 1230 for the non-enhanced solution, that is 5.04% less. For the (μ, d) sample (row 3) the corresponding number is 5.93%. The nominal enhancement turns out to be slightly worse than the (μ, d) and sample enhancements, compare e.g. the means 1180 and 1168 for the uniform sample and the higher standard deviations of the nominal enhancement.

In Figure 2 we present the empirical cumulative distribution functions of the non-enhanced and (μ, d)-enhanced solutions in a sample problem in the uniform demand sample. In this plot, it is clear that the total cost incurred by the non-enhanced solutions stochastically dominates the one from the (μ, d) enhanced solutions.

Figure 2 Empirical cumulative distribution functions of the total costs in simulation in the uniform demand sample for the non-enhanced RO solution and the (μ, d)-enhanced RO solution. Aggregated from all problem instances.



4. Extension - aggregated random vectors

4.1. Introduction

Up to now, we have been deriving exact worst-case expectations which (i) relied on the assumption of independence of the components of $\mathbf{z}$, and (ii) resulted in bounds involving $3^{n_\mathbf{z}}$ terms. In this section, we consider practically relevant cases where both of these difficulties can be alleviated.

In many cases, uncertainty appears in a linearly aggregated way such as $y = \mathbf{a}^T\mathbf{z}$ or $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^T\mathbf{z}$. Then, instead of considering the worst-case expectation of $f(\mathbf{x}, \mathbf{a}^T\mathbf{z})$ or $f(\mathbf{a}(\mathbf{x})^T\mathbf{z})$ (or, more generally, $\sum_i f(\mathbf{x}, \mathbf{a}_i^T\mathbf{z})$ or $\sum_i f(\mathbf{a}_i(\mathbf{x})^T\mathbf{z})$), it is possible to consider the worst-case expectations of $f(\mathbf{x}, y)$, where y is a single-dimensional ambiguous random variable equal either to $\mathbf{a}^T\mathbf{z}$ or to $\mathbf{a}(\mathbf{x})^T\mathbf{z}$ while the uncertainty is still specified in terms of the entire vector $\mathbf{z}$. Without loss of generality, we assume that the support of $\mathbf{z}$ is given by $\|\mathbf{z}\|_\infty \leq 1$.

4.2. Fixed vector $\mathbf{a}$

We begin our analysis with the case where the vector $\mathbf{a}$ is not dependent on the decision variables, motivated by the following example.

EXAMPLE 1. In the inventory problem of Section 3.3 the holding and backlogging costs at time $t+1$ depend on the state of inventory x_{t+1}. If the ordering decisions $q_t(\mathbf{z}^{t-1})$ are static (non-adjustable), then:

$$x_{t+1} = x_1 + \sum_{s=1}^{t} q_s - \sum_{s=1}^{t} z_s = x_1 + \sum_{s=1}^{t} q_s - \mathbf{1}^T\mathbf{z}^t,$$

and the aggregated random variable is $y = \mathbf{1}^T\mathbf{z}^t$ not depending on the decision variables q_t. ■

In order to use the results of BH to construct the worst-case expectation of $f(\mathbf{x}, y)$, we need to extract the distributional information on $y = \mathbf{a}^T\mathbf{z}$ from the information on $\mathbf{z}$. Then, we have that

$$\begin{aligned} \text{supp}(\mathbf{a}^T\mathbf{z}) = [\min_{\mathbf{z}} \mathbf{a}^T\mathbf{z}, \max_{\mathbf{z}} \mathbf{a}^T\mathbf{z}] &= [-\|\mathbf{a}\|_1, \|\mathbf{a}\|_1] \\ \mathbb{E}_{\mathbb{P}}\left(\mathbf{a}^T\mathbf{z}\right) &= \mathbf{a}^T\boldsymbol{\mu}. \end{aligned} \tag{30}$$

As for the MAD $M(\mathbf{a}^T\mathbf{z})$, only upper bounds on $M(\mathbf{a}^T\mathbf{z})$ can be obtained in terms of information on $\mathbf{z}$. Nevertheless, any upper bound on $M(\mathbf{a}^T\mathbf{z})$ will generate an upper bound on $\mathbb{E}_{\mathbb{P}} f(\mathbf{x}, y)$ due to the fact that the BH upper bound (8) is a nondecreasing function of d, as stated by the following Proposition.

Proposition 1 *The worst-case expectation (8) is a nondecreasing function of d.*

Proof. The worst-case expectation (8) is:

$$F(d) = \frac{d}{2(\mu - a)} f(a) + \left(1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}\right) f(\mu) + \frac{d}{2(b - \mu)} f(b).$$

We have

$$F'(d) = \frac{1}{2(\mu - a)} f(a) - \left(\frac{1}{2(\mu - a)} + \frac{1}{2(b - \mu)}\right) f(\mu) + \frac{1}{2(b - \mu)} f(b) \geq 0.$$

Multiplying the last inequality by $2(b - \mu)(\mu - a)/(b - a)$ and using

$$\mu = \frac{b - \mu}{b - a} a + \frac{\mu - a}{b - a} b$$

we obtain the inequality:

$$\frac{b - \mu}{b - a} f(a) + \frac{\mu - a}{b - a} f(b) \geq f\left(\frac{\mu - a}{b - a} b + \frac{b - \mu}{b - a} a\right),$$

which is valid by convexity of $f(\cdot)$. ■

In the following, we present four ways to obtain upper bounds on the MAD $M(\mathbf{a}^T\mathbf{z})$. The first three of them do not use the assumption of independence of the components of $\mathbf{z}$ and are hence computable in polynomial time. The last one, on the other hand, requires the independence of z_i's, but the computation involving $3^{n_{\mathbf{z}}}$ terms can be done in a pre-processing step, without affecting the size of the optimization problem.

No independence - simple bounds. Two upper bounds that we use here are:

$$\mathbb{E}_{\mathbb{P}}\left|y - \mathbf{a}^T\boldsymbol{\mu}\right| = \mathbb{E}_{\mathbb{P}}\left|\sum_{i=1}^{n_{\mathbf{z}}} a_i z_i - \sum_{i=1}^{n_{\mathbf{z}}} a_i \mu_i\right| \leq \sum_{i=1}^{n_{\mathbf{z}}} \mathbb{E}_{\mathbb{P}}\left|a_i z_i - a_i \mu_i\right| = \sum |a_i| d_i = |\mathbf{a}|^T\mathbf{d}, \tag{31}$$

which gives the following ambiguity set for the distribution of y:

$$\mathcal{P}_y^{\mathbf{d}} = \left\{\mathbb{P}_y : \quad \text{supp}(\mathbb{P}_y) \subseteq [-\|\mathbf{a}\|_1, \|\mathbf{a}\|_1], \quad \mathbb{E}_{\mathbb{P}_y} y = \mathbf{a}^T\boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}_y}|y - \mathbf{a}^T\boldsymbol{\mu}| \leq |\mathbf{a}|^T\mathbf{d}\right\}, \tag{32}$$

and the second bound, based on the covariance matrix $\Sigma_{\mathbf{z}}$, is

$$\mathbb{E}_{\mathbb{P}}\left|y-\mathbf{a}^T\boldsymbol{\mu}\right| \leq \sqrt{\mathbb{E}\left(y-\mathbf{a}^T\boldsymbol{\mu}\right)^2} = \sqrt{\mathbb{E}\left(\mathbf{a}^T\mathbf{z}-\mathbf{a}^T\boldsymbol{\mu}\right)^2} = \sqrt{\mathrm{Var}(\mathbf{a}^T\mathbf{z})} = \sqrt{\mathbf{a}^T\Sigma_{\mathbf{z}}\mathbf{a}}, \tag{33}$$

which gives the following ambiguity set for the distribution of y:

$$\mathcal{P}_y^{\mathrm{Cov}} = \left\{\mathbb{P}_y: \quad \mathrm{supp}(\mathbb{P}_y) \subseteq [-\|\mathbf{a}\|_1, \|\mathbf{a}\|_1], \quad \mathbb{E}_{\mathbb{P}_y} y = \mathbf{a}^T\boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}_y}|y-\mathbf{a}^T\boldsymbol{\mu}| \leq \sqrt{\mathbf{a}^T\Sigma_{\mathbf{z}}\mathbf{a}}\right\}. \tag{34}$$

BH bounds obtained using ambiguity sets (32) and (34) do not require the components of $\mathbf{z}$ to be independent and require only three terms, thus using (32) :

$$\begin{aligned}\sup_{\mathbb{P}_y\in\mathcal{P}_y^{\mathbf{d}}} \mathbb{E}_{\mathbb{P}_q} f(\mathbf{x},y) \leq &\frac{|\mathbf{a}|^T\mathbf{d}}{2(\mathbf{a}^T\boldsymbol{\mu}+\|\mathbf{a}\|_1)} f(\mathbf{x},-\|\mathbf{a}\|_1) + \left(1-\frac{|\mathbf{a}|^T\mathbf{d}}{2(\mathbf{a}^T\boldsymbol{\mu}+\|\mathbf{a}\|_1)} - \frac{|\mathbf{a}|^T\mathbf{d}}{2(\|\mathbf{a}\|_1-\mathbf{a}^T\boldsymbol{\mu})}\right) f(\mathbf{x},\mathbf{a}^T\boldsymbol{\mu}) + \\ &+\frac{|\mathbf{a}|^T\mathbf{d}}{2(\|\mathbf{a}\|_1-\mathbf{a}^T\boldsymbol{\mu})} f(\mathbf{x},\|\mathbf{a}\|_1).\end{aligned}$$

Since $\mathbf{a}$ does not depend on $\mathbf{x}$, the resulting expression is convex in $\mathbf{x}$.

No independence - exact bound using Wiesemann et al. (2014). It is also possible to obtain an exact upper bound on the MAD of y using the results of Wiesemann et al. (2014), which requires, though, solving an optimization problem. We present it on the example of (μ, d) information about $\mathbf{z}$ but it can also use the mean-covariance and some other types of information on $\mathbf{z}$. The problem to solve is:

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left|\mathbf{a}^T\mathbf{z}-\mathbf{a}^T\boldsymbol{\mu}\right|. \tag{35}$$

It can be proved (see Appendix EC.2 to this paper) using Theorem 1 of Wiesemann et al. (2014) that (35) is equivalent to:

$$\begin{array}{rl}\min\limits_{\boldsymbol{\phi}_1,\boldsymbol{\phi}_2\geq 0, w, \boldsymbol{\beta}, \kappa} & w \\ \text{s.t.} & \mathbf{b}^T\boldsymbol{\beta}+\kappa \leq w \\ & \mathbf{c}^T\boldsymbol{\phi}_1 - \mathbf{a}^T\boldsymbol{\mu} \leq \kappa \\ & \mathbf{c}^T\boldsymbol{\phi}_2 + \mathbf{a}^T\boldsymbol{\mu} \leq \kappa \\ & \mathbf{C}^T\boldsymbol{\phi}_1 + \mathbf{A}^T\boldsymbol{\beta} = \mathbf{a} \\ & \mathbf{C}^T\boldsymbol{\phi}_2 + \mathbf{A}^T\boldsymbol{\beta} = -\mathbf{a} \\ & \mathbf{D}^T\boldsymbol{\phi}_1 + \mathbf{B}^T\boldsymbol{\beta} = 0 \\ & \mathbf{D}^T\boldsymbol{\phi}_2 + \mathbf{B}^T\boldsymbol{\beta} = 0,\end{array} \tag{36}$$

where $\mathbf{A},\mathbf{B}\in\mathbb{R}^{2n_{\mathbf{z}}\times n_{\mathbf{z}}}$, $\mathbf{b}\in\mathbb{R}^{2n_{\mathbf{z}}}$, $\mathbf{C},\mathbf{D}\in\mathbb{R}^{6n_{\mathbf{z}}\times n_{\mathbf{z}}}$, and $\mathbf{c}\in\mathbb{R}^{6n_{\mathbf{z}}}$ are defined as:

$$\mathbf{A} = \begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix}\boldsymbol{\mu}\\ \mathbf{d}\end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix}\mathbf{I}\\ -\mathbf{I}\\ \mathbf{I}\\ -\mathbf{I}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ -\mathbf{I}\\ -\mathbf{I}\\ \mathbf{I}\\ -\mathbf{I}\end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix}\mathbf{1}\\ \mathbf{1}\\ \boldsymbol{\mu}\\ -\boldsymbol{\mu}\\ \mathbf{1}\\ \mathbf{0}\end{bmatrix}.$$

The optimal value to (36) is thus a tight upper bound on the MAD of $\mathbf{z}$ and at least as good as bound (31). With the optimal value to (36), one can construct an ambiguity set similar to (32) and analogously, build up the worst-case expectation corresponding to it. This MAD bound and the one of (31) are identical if, for example, $\mu_i = \mu_j$ and $d_i = d_j$ for all $i \neq j$.

In this way, the method of Wiesemann et al. (2014) can be used to enhance our method for aggregated random variables. Building up the upper bound on the MAD of $\mathbf{a}^T\mathbf{z}$ via problem (36) is preferable to bound (31) if the need to solve the optimization problem is not burdensome.

Independence - $3^{n_{\mathbf{z}}}$ terms in a pre-processing step. As a last bound, we note that the function $|\mathbf{a}^T\mathbf{z} - \mathbf{a}^T\boldsymbol{\mu}|$, whose expectation is equal to $M(\mathbf{a}^T\mathbf{z})$ is convex in $\mathbf{z}$. This means that its worst-case expectation can be computed using BH:

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}\,|y - \mathbf{a}^T\boldsymbol{\mu}| &\leq \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\,|y - \mathbf{a}^T\boldsymbol{\mu}| \\ &= \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^{n_{\mathbf{z}}}} \prod_{i=1}^{n_{\mathbf{z}}} p^i_{\alpha_i} |\mathbf{a}^T\mathbf{z}(\boldsymbol{\alpha}) - \mathbf{a}^T\boldsymbol{\mu}|, \end{aligned} \tag{37}$$

with $\mathbf{z}(\boldsymbol{\alpha}) = (\tau^1_{\alpha_1}, \ldots, \tau^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}})$ as in (15). This gives rise to another ambiguity set, which can be constructed analogously to (32). The computation in (37) can be conducted before the optimization problem is set up. Therefore, optimization problem involving constraint $f(\mathbf{x}, \mathbf{a}^T\mathbf{z})$ is easier than would be the case if formula (20) of Section 3.1 were used.

4.3. Vector $\mathbf{a}(\mathbf{x})$ depends on $\mathbf{x}$.

We assume that $\mathbf{a}(\mathbf{x})$ is a linear vector-valued function of $\mathbf{x}$ and the function whose worst-case expectation we seek is $f(\mathbf{a}(\mathbf{x})^T\mathbf{z})$, where $f(\cdot)$ is convex and $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^T\mathbf{z}$. This assumption also holds for the inventory problem of Section 3.3.

Example 2. Using linear decision rules $q_{t+1}(\mathbf{z}^t) = q_{t+1,0} + \sum_{j=1}^t q_{t+1,j} z_j$ for the ordering decisions, the state of inventory at time $t+1$ is

$$x_{t+1} = x_1 + \sum_{s=1}^t \left(q_{t,0} + \sum_{j=1}^t q_{t,j} z_j \right) - \sum_{j=1}^t z_t = x_1 + \sum_{s=1}^t q_{t,0} + \sum_{s=1}^t \left(\sum_{j=s+1}^t q_{j,s} - 1 \right) z_s.$$

Therefore, here in each time period the aggregated random variable is $\sum_{s=1}^t \left(\sum_{j=s+1}^t q_{j,s} - 1 \right) z_s$, which indeed depends on the decision variables. ■

Similarly to the previous case, one can consider the worst-case expectation of $f(y(\mathbf{x}))$ where $y(\mathbf{x}) = \mathbf{a}(\mathbf{x})^T\mathbf{z}$, which for the set (32) is:

$$\begin{aligned} \sup_{\mathbb{P}_{y(\mathbf{x})}\in\mathcal{P}_{y(\mathbf{x})}} \mathbb{E}_{\mathbb{P}_{y(\mathbf{x})}} f(y(\mathbf{x})) \leq{}& \frac{|\mathbf{a}(\mathbf{x})|^T\mathbf{d}}{2(\mathbf{a}(\mathbf{x})^T\boldsymbol{\mu} + \|\mathbf{a}(\mathbf{x})\|_1)} f(-\|\mathbf{a}(\mathbf{x})\|_1) + \frac{|\mathbf{a}(\mathbf{x})|^T\mathbf{d}}{2(\|\mathbf{a}\|_1 - \mathbf{a}(\mathbf{x})^T\boldsymbol{\mu})} f(\|\mathbf{a}(\mathbf{x})\|_1) + \\ &+ \left(1 - \frac{|\mathbf{a}(\mathbf{x})|^T\mathbf{d}}{2(\mathbf{a}(\mathbf{x})^T\boldsymbol{\mu} + \|\mathbf{a}(\mathbf{x})\|_1)} - \frac{|\mathbf{a}(\mathbf{x})|^T\mathbf{d}}{2(\|\mathbf{a}(\mathbf{x})\|_1 - \mathbf{a}(\mathbf{x})^T\boldsymbol{\mu})} \right) f(\mathbf{a}(\mathbf{x})^T\boldsymbol{\mu}). \end{aligned} \tag{38}$$

This expression is not neccessarily convex in $\mathbf{x}$. However, for the important special case where $\boldsymbol{\mu}=0$, satisfied for example, if $\mathbf{z}$ is considered to be a distortion with expected value 0 such as the white noise in signal processing, (38) can be bounded as follows:

$$\begin{aligned}\sup_{\mathbb{P}_{y(\mathbf{x})}\in\mathcal{P}_y^{\mathbf{d}}(\mathbf{x})} \mathbb{E}_{\mathbb{P}_{y(\mathbf{x})}} f(y(\mathbf{x})) &\leq \left(1-\frac{\sum_i |a_i(\mathbf{x})|d_i}{\|\mathbf{a}(\mathbf{x})\|_1}\right) f(0) + \frac{\sum_i |a_i(\mathbf{x})|d_i}{2\|a(\mathbf{x})\|_1}\left(f(-\|\mathbf{a}(\mathbf{x})\|_1)+f(\|\mathbf{a}(\mathbf{x})\|_1)\right) \\ &= \left(\frac{\sum_i |a_i(\mathbf{x})|(1-d_i)}{\|\mathbf{a}(\mathbf{x})\|_1}\right) f(0) + \frac{\sum_i |a_i(\mathbf{x})|d_i}{2\|\mathbf{a}(\mathbf{x})\|_1}\left(f(-\|\mathbf{a}(\mathbf{x})\|_1)+f(\|\mathbf{a}(\mathbf{x})\|_1)\right) \\ &\leq \max_i (1-d_i) f(0) + \max_i d_i \left(\frac{1}{2} f(-\|\mathbf{a}(\mathbf{x})\|_1) + \frac{1}{2} f(\|\mathbf{a}(\mathbf{x})\|_1)\right). \qquad (39)\end{aligned}$$

Quality of the bound (39) depends now on the dispersion of the MADs d_i - if they are equal, the second inequality is tight. Tractability of (39) depends on convexity of the sum

$$f(-\|\mathbf{a}(\mathbf{x})\|_1)+f(\|\mathbf{a}(\mathbf{x})\|_1), \qquad (40)$$

which turns out to be the case, as the following proposition shows.

Proposition 2 *For affine* $\mathbf{a}(\mathbf{x})$ *and convex* $f:\mathbb{R}\to\mathbb{R}$*, the function* $f(-\|\mathbf{a}(\mathbf{x})\|_1)+f(\|\mathbf{a}(\mathbf{x})\|_1)$ *is convex.*

Proof. Define $g(t)=f(t)+f(-t)$ for $t\in\mathbb{R}_+$ and $h(\mathbf{x})=\|\mathbf{a}(\mathbf{x})\|_1$. Then we have that:

$$f(-\|\mathbf{a}(\mathbf{x})\|_1)+f(\|\mathbf{a}(\mathbf{x})\|_1)=g(h(\mathbf{x})).$$

The function $g(h(\mathbf{x}))$ is convex if $g(t)$ is convex and nondecreasing and $h(\mathbf{x})$ is convex. Convexity of $h(\mathbf{x})$ is clear as it is a norm of an affine function of $\mathbf{x}$. Also, convexity of $g(\cdot)$ follows from convexity of $f(\cdot)$. We need to show that $g(\cdot)$ is nondecreasing, i.e., that

$$g(t+\alpha)\geq g(t), \quad \forall t\geq 0, \alpha\geq 0.$$

Consider the subgradients $v_1\in\partial f(-t)$ and $v_2\in\partial f(t)$. By properties of subgradients we have that

$$v_1 \leq \frac{f(t)-f(-t)}{2t} \leq v_2.$$

From this, it follows that for $\alpha\geq 0$:

$$\begin{aligned} g(t+\alpha) &= f(-t-\alpha)+f(t+\alpha) \\ &\geq f(-t) + \sup_{v\in\partial f(-t)} (-\alpha v) + f(t) + \sup_{v\in\partial f(t)} (\alpha v) \\ &\geq f(-t) + (-\alpha v_1) + f(t) + (\alpha v_2) \\ &= g(t) + \alpha(v_2 - v_1) \\ &\geq g(t). \end{aligned}$$

■

With respect to the efficient implementation, it can be added that due to $f(-t)+f(t)$ being nondecreasing on the nonnegative ray we have:

$$f(-\|\mathbf{a}(\mathbf{x})\|_1)+f(\|\mathbf{a}(\mathbf{x})\|_1)\leq 0 \quad \Leftrightarrow \quad \begin{cases} f(-\mathbf{1}^T\mathbf{w})+f(\mathbf{1}^T\mathbf{w})\leq 0 \\ \mathbf{w}\geq \mathbf{a}(\mathbf{x}) \\ \mathbf{w}\geq -\mathbf{a}(\mathbf{x}), \end{cases}$$

where $\mathbf{w}$ is an additional analysis variable.

4.4. Inventory experiment revisited - independent demand

In this section we re-visit our inventory experiment using the results for aggregated random vectors of the previous section. The inventory experiment can be studied in this way, since the objective function is:

$$\sum_{t=1}^{T}\left(c_t q_t(\mathbf{z}^{t-1})+h_t \max\left\{x_{t+1}(\mathbf{z}^t),0\right\}+p_t \max\left\{-x_{t+1}(\mathbf{z}^t),0\right\}\right),$$

where we dropped the first term as we assumed in the numerical experiment that the salvage value is zero, i.e., $s=0$. The objective consists of T terms

$$f_t(\mathbf{q},\mathbf{z})=c_t q_t(\mathbf{z}^{t-1})+h_t \max\left\{x_{t+1}(\mathbf{z}^t),0\right\}+p_t \max\left\{-x_{t+1}(\mathbf{z}^t),0\right\}, \tag{41}$$

that depend on the state of inventory x_{t+1} each. Therefore, in line with Examples 1 and 2, we can use our results for aggregated random variables to build worst-case expectations of $f_t(\mathbf{q},\mathbf{z})$ and to use them in the optimization problem.

Since the methods of this section are also aimed at an alleviation of the independence assumption and the piecewise linear objective function is tractable using the results of Wiesemann et al. (2014), we compare our solutions also to Wiesemann et al. (2014), whose results do not rely on the independence assumption either. We note here that to use their results to obtain an exact reformulation, we need to formulate the sum-of-maximums objective function as a maximum of linear functions which leads to 2^T terms. For the case where the independence assumption is satisfied, we refer the reader to Appendix C where the difference between their and our reformulations is illustrated.

In the following, we consider seven solutions to the same 50 instances as in Sections 3.3 and 3.4. The first four solutions consider the ordering decisions to be static:

- S-1: using BH to evaluate the true worst-case expectation of the objective with 3^T terms in the problem formulation.
- S-2: using the aggregated random variables $y=\mathbf{1}^T\mathbf{z}$ for which an upper bound on the MAD is computed under the independence assumption with (37) (T terms in the problem formulation).
- S-3: using the methodology of Wiesemann et al. (2014) (2^T terms in the problem formulation).

Table 5 Results of the inventory management with aggregated random vectors - worst-case costs and ranges for the expectation of the objective over $\mathcal{P}_{(\mu,d,\beta)}$ (with independence assumption in computing the bounds). The lower bounds of the intervals are obtained *ex-post* using BH results, after the solutions are found. All numbers are averages.

β	Cost expectation range						
	S-1	S-2	S-3	S-4	LDR-1	LDR-2	LDR-3
0.25	[1093,1175]	[1124,1205]	[1131,1190]	[1131,1190]	[1004,1049]	[1058,1093]	[1058,1092]
0.5	[1038,1175]	[1065,1205]	[1061,1190]	[1061,1190]	[970,1049]	[1029,1093]	[1028,1092]
0.75	[1079,1175]	[1117,1205]	[1106,1190]	[1106,1190]	[994,1049]	[1047,1093]	[1047,1092]

Results for LDR-1 are readily taken from Table 2.

- S-4: using ambiguity set (32) without the independence assumption on z_i's ($3T$ terms in the problem formulation).

The other three solutions consider the ordering decisions to be linear functions of past demand:

- LDR-1: using BH to evaluate the true worst-case expectation of the objective with 3^T terms.
- LDR-2: using methodology of Wiesemann et al. (2014), (2^T terms in the problem formulation).
- LDR-3: using approximation (39) to obtain upper bound on the worst-case expectation of the objective ($3T$ terms in the problem formulation).

Table 5 presents the worst-case and best-case expectations for all 7 solutions, computed under the assumption of independence of z_i (in the same way as in the main experiment). For the solutions with static decisions we can see that the new solutions S-2 and S-4 based on the aggregation technique are only slightly worse than the original solution based on BH bound with 3^T terms. For example, their worst-case expectations are 1205 and 1190, respectively, whereas solution S-1 yields 1175, which makes them only 2% worse than the exact formulation.

For the solutions with linear decision rules, we see that the expectation intervals overlap for $\beta = 0.5$ and $\beta = 0.75$. The worst-case expectation of the new LDR-3 solution is 1092, whereas for the old S-1 solution it is 1049. That is, the new solution is about 4% worse than the exact formulation LDR-1.

We observe that the intervals obtained by our aggregated solutions are very similar to the ones obtained by the methodology of Wiesemann et al. (2014), compare solutions S-3 vs S-4 and LDR-2 vs LDR-3. Solutions S-3 and S-4 are identical for all instances since our aggregation technique is exact in this case, just as the method of Wiesemann et al. (2014) - this is because in each case, the mean of the uncertain demand is in the middle of the support and the proportion of the MADs of z_t's to the support intervals' widths is the same for all t, see the description of the setting in Section 3.3. However, the similarity of the intervals for LDR-2 and LDR-3 comes as a surprise since bound (39) is just an approximation, whereas the method of Wiesemann et al. (2014) is exact and involves 2^T terms in the problem formulation (their results rely on formulating the objective as a

Table 6 Simulation results for the inventory problem with aggregation technique. Numbers in brackets denote the % change compared to the S-1/LDR-1 solution, respectively.

Objective type	Sample	Cost						
		S-1	S-2	S-3	S-4	LDR-1	LDR-2	LDR-3
Mean	Uniform	1088	1182 (+8.6%)	1109 (+1.9%)	1109 (+1.9%)	994	1051 (+5.7%)	1049 (+5.5%)
Standard deviation	Uniform	317	325 (+2.5%)	321 (+1.36%)	321 (+1.36%)	259	252 (−2.7%)	251 (−3.1%)
Mean	(μ, d)	1094	1124 (+2.7%)	1115(+1.9%)	1115(+1.9%)	1003	1056 (+ 5.5%)	1018 (+1.5%)
Standard deviation	(μ, d)	325	333 (+2.5%)	328 (+0.9%)	328 (+0.9%)	265	257 (−3.0%)	224 (−15.5%)

Results for LDR-1 are readily taken from Table 3.

maximum over a finite number of affine functions, which for our sum-of-max objective requires 2^T terms to consider all cases).

Table 6 presents the results of simulation in the same setting as in Table 3, with demands from different periods being independent. As we can see, the new S-2, S-4 and LDR-3 solutions perform slightly worse on average than the exact S-1 and LDR-1 solutions. An important observation, however, is that the solution LDR-2 based on the rather conservative MAD bound performs better than the exact S-1 solution which utilizes static decisions.

We can see that the solutions S-3 (Wiesemann et al. 2014) and S-4 (using (31)) are indeed the same since the simulation results are identical for both. Comparing the solutions LDR-2 (Wiesemann et al. 2014) and LDR-3 (using (39)) we see that despite the similarity of intervals in Table 5 the solutions do differ, since the obtained results are not the same. The large difference between the two for the (μ, d) sample (means 1056 and 1018, respectively) is a sample-specific issue and, having repeated the experiment in multiple samples, we conclude that the two solutions give very similar results on average.

To conclude on the results of this section, we can say that for problems with aggregated random variables $\mathbf{a}^T\mathbf{z}$ our approach is preferable if there is no assumption of independence of $\mathbf{z}$ and (i) the need to compute the exact worst-case MAD of $\mathbf{a}^T\mathbf{z}$ via solving problem (36) would be too burdensome, or (ii) the complexity of the function $f(\mathbf{x}, \mathbf{a}^T\mathbf{z})$ is not tractable in the sense of requirements of Wiesemann et al. (2014). In other cases, it is preferable either to estimate the exact worst-case MAD of $\mathbf{a}^T\mathbf{z}$ via (36) or to apply the results of Wiesemann et al. (2014) directly. For the case of $\mathbf{a}(\mathbf{x})^T\mathbf{z}$, our method is preferable if there is no assumption of independence of $\mathbf{z}$ and the complexity of the function $f(\mathbf{a}(\mathbf{x})^T\mathbf{z})$ is not tractable in the sense of requirements of Wiesemann et al. (2014). Otherwise, the approach of Wiesemann et al. (2014) is preferred.

4.5. Inventory experiment revisited - dependent demand

In the previous subsection the solutions were evaluated using demand samples in which the demands from subsequent periods were independent. One may ask: how do the solutions perform when the realized demand sample exhibits some dependence pattern?

To investigate this, we run an experiment where the demands $\mathbf{z}$ are sampled using copulas that couple a multivariate distribution function to its marginal distributions. Separating the dependence structure between random variables from their marginal distributions makes them a premier tool for simulating correlated random variables when particular marginal distributions are desired (Sklar 1959). In our case, we want the marginal distributions to come from our (μ, d) sample (results for the uniform marginals are very similar).

In our experiment, we use the T-dimensional Gaussian copula[3] and assume that the dependence between the z_i's follows an autocorrelative pattern where the correlations between the random variables used in the copula from periods t_1 and t_2 is equal to $\rho^{|t_1-t_2|}$, where $\rho \in \{0.1, 0.2, \ldots, 0.9\}$, that is, for the copula sampling we use the correlation matrix:

$$\begin{bmatrix} 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^1 & 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 \\ \vdots & & \ddots & & & \vdots \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

For conciseness, we focus only on the LDR-1 and LDR-3 solutions under (μ, d) sample.

Figure 3 presents the results, with the respective means and standard deviations plotted against the correlation strength ρ. The left panel shows that as the degree of correlation among the demands increases with ρ (on the horizontal axis), the mean costs obtained by solution LDR-3 approaches the one of LDR-1 and eventually becomes smaller, with the crossing point approximately around $\rho = 0.6$. In the right panel we see that the standard deviation of the costs obtained by LDR-3 is is smaller than the one of LDR-1 for all values of ρ.

These results provide a strong argument that the LDR-3 solution, though being constructed on the basis of approximation (39) and not assuming the independence of z_i's might be better than LDR-1 solution based on the full (μ, d) information in situations when the realized demand exhibits dependence among its components from different periods. More generally, this indicates that the solutions based on aggregated random vectors, while being less computationally burdensome than the 'exact solutions' can yield better performance when the true random variables deviate from the independence assumption.

5. Safe approximations of chance constraints

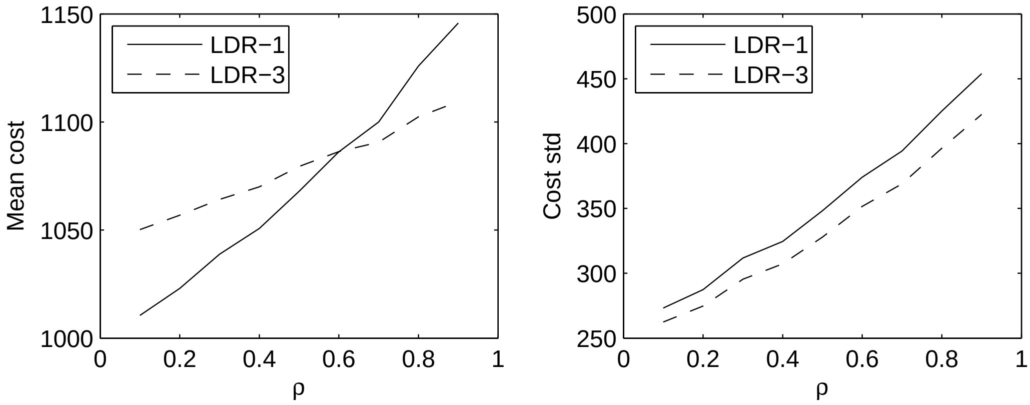
5.1. Introduction

In this section we show how the results of Ben-Tal and Hochman (1972) can be used to construct safe tractable approximations of scalar chance constraints:

$$\mathbb{P}\left(\mathbf{a}^T(\mathbf{z})\mathbf{x} > \mathbf{b}(\mathbf{z})\right) \leq \epsilon, \quad \forall \mathbb{P} \in \mathcal{P}_{(\mu,d)}, \text{ where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0]. \tag{42}$$

[3] We use the MATLAB function copularnd(). As this function simulates only the CDFs of the marginal distributions, we need to convert them into the respective uniform and (μ, d) sample by inverting their distribution functions.

Figure 3 Results of the simulation experiment for inventory solutions LDR-1 and LDR-3 with dependent (μ, d) demand sample. The left panel involves results on the means and the right panel on the standard deviations.



Without loss of generality we assume that the components z_i have a support contained in $[-1,1]$ and mean 0:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P}: \quad \text{supp}(z_i) \subseteq [-1,1], \quad \mathbb{E}_{\mathbb{P}} z_i = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d_i, \quad i=1,\ldots,n_{\mathbf{z}}, \quad z_i \perp\!\!\!\perp z_j, \quad \forall i \neq j\}.$$

To construct the safe tractable approximations, we use the mathematical framework of Ben-Tal et al. (2009). In this framework, the key step consists of bounding from above the *moment-generating function* of $z_i, i=1,\ldots,n_{\mathbf{z}}$:

$$\mathbb{E}_{\mathbb{P}} \exp(wz_i) = \int \exp(wz_i) d\mathbb{P}_i(z_i)$$

and then using the resulting bound in combination with the Markov inequality to obtain upper bounds on the probability $\mathbb{P}(\mathbf{a}^T(\mathbf{z})\mathbf{x} > \mathbf{b}(\mathbf{z}))$ - often referred to as the *Bernstein approximation.*

A strong motivation for using the ambiguity set $\mathcal{P}_{(\mu,d)}$ is due to the fact that a tight explicit bound on $\mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T\mathbf{z})$ is obtained easily in this setting by the BH results described in Section 2. Indeed, due to the independence of $z_1,\ldots,z_{n_{\mathbf{z}}}$ we have:

$$\begin{aligned}\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{z}^T\mathbf{w}) &= \prod_{i=1}^{n_{\mathbf{z}}} \sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp(z_i w_i) \\ &= \prod_{i=1}^{n_{\mathbf{z}}} \left(\frac{d_i}{2}\exp(-w_i) + \frac{d_i}{2}\exp(w_i) + (1-d_i)\exp(0) \right) \\ &= \prod_{i=1}^{n_{\mathbf{z}}} \left(d_i \cosh(w_i) + 1 - d_i\right). \end{aligned} \tag{43}$$

The worst-case expectation is evaluated separately for each component of $\mathbf{z}$, avoiding the computational burden of summation of $3^{n_{\mathbf{z}}}$ terms as in (14). In Appendix C we show that in the setting of Wiesemann et al. (2014) without independence of z_i's, obtaining the tight upper bound on $\exp(\mathbf{w}^T\mathbf{z})$ requires solving an optimization problem involving an uncertain constraint on a convex function. This requires an exponential number of constraints for an exact reformulation.

5.2. Safe approximations - results

We now show how (43) can be used to obtain safe approximations of (42). First, we present two simple safe approximations in order of increasing tightness. Later, we show that the (μ, d) information is particularly suitable for obtaining even tighter safe approximations, based on the exponential polynomials.

The first approximation requires the use of Theorem 2.4.4 of Ben-Tal et al. (2009), repeated in Appendix B.1.

Theorem 1 *If for a given vector* $\mathbf{x}$ *there exist* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n_\mathbf{z}+1}$ *such that* $(\mathbf{x}, \mathbf{u}, \mathbf{v})$ *satisfies the constraint system*

$$\begin{cases} (\mathbf{a}^i)^T\mathbf{x} - b_i = u_i + v_i, 0 \le i \le n_\mathbf{z} \\ u_0 + \sum\limits_{i=1}^{n_\mathbf{z}} |u_i| \le 0 \\ v_0 + \sqrt{2\log(1/\epsilon)}\sqrt{\sum\limits_{i=1}^{n_\mathbf{z}} \sigma_i^2 v_i^2} \le 0, \end{cases} \tag{44}$$

where

$$\sigma_i = \sup_{t\in\mathbb{R}} \sqrt{\frac{2\log\left(d_i \cosh(t) + 1 - d_i\right)}{t^2}}, \tag{45}$$

then $\mathbf{x}$ *is feasible to (42), that is, constraint system (44) is a safe approximation of (42). Moreover, (44) is the robust counterpart of the following robust constraint*

$$\mathbf{a}^T(\mathbf{z})\mathbf{x} \le \mathbf{b}(\mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{U}, \ \textit{where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_\mathbf{z}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0], \tag{46}$$

and

$$\mathcal{U} = \left\{ \mathbf{z} \in \mathbb{R}^{n_\mathbf{z}} : \quad \sqrt{\sum_{i=1}^{n_\mathbf{z}} \frac{z_i^2}{\sigma_i^2}} \le \sqrt{2\log(1/\epsilon)}, \quad -1 \le z_i \le 1, \quad i = 1, \dots, n_\mathbf{z} \right\}.$$

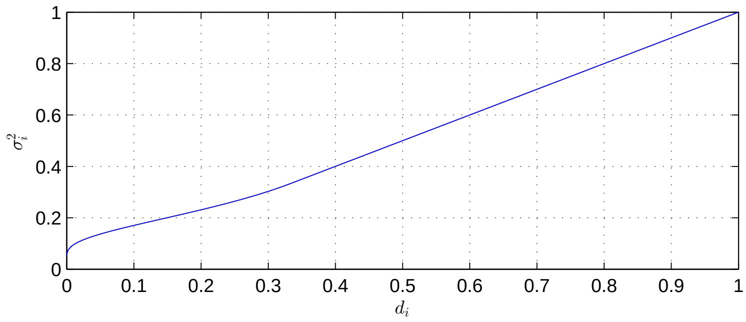
Proof. The proof follows the steps leading to Theorem 2.4.4 from Ben-Tal et al. (2009). First, we need to find scalars $\mu_i^-, \mu_i^+, \sigma_i$, where $i = 1, \dots, n_\mathbf{z}$ such that:

$$\int_{-1}^{1} \exp(tz_i) d\mathbb{P}_i(z_i) \le \exp\left(\max\{\mu_i^-, \mu_i^+\} + \frac{\sigma_i^2}{2}\right), \quad \forall t \in \mathbb{R}, \quad \forall \mathbb{P} \in \mathcal{P}_{(\mu,d)}.$$

By (43) we have $\sup\limits_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \left\{ \int_{-1}^{1} \exp(tz_i) d\mathbb{P}_i(z_i) \right\} = d_i \cosh(t) + 1 - d_i$. Thus, for each i we need to find $\mu_i^-, \mu_i^+, \sigma_i$ such that:

$$d_i \cosh(t) + 1 - d_i \le \exp\left(\max\{\mu_i^- t, \mu_i^+ t\} + \frac{\sigma_i^2 t^2}{2}\right), \quad \forall t \in \mathbb{R}.$$

Figure 4 Plot of σ_i^2 as a function of d_i.



Setting $\mu_i^- = \mu_i^+ = 0$, we then need σ_i such that

$$d_i \cosh(t) + 1 - d_i \leq \exp\left(\frac{\sigma_i^2 t^2}{2}\right), \quad \forall t \in \mathbb{R} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \sigma_i^2 \geq g_i(t) = \frac{2}{t^2} \log\left(d_i \cosh(t) + 1 - d_i\right), \quad \forall t \in \mathbb{R}.$$

Thus, we look for the maximum value of $g_i(t)$ over the real axis. From the definition of $g_i(t)$ we know that it is finite, nonnegative, and differentiable everywhere except for 0. By de l'Hôpital rule we have that $\lim_{t\to 0} g_i(t) = d_i$. It holds that $\lim_{t\to\pm\infty} g_i(t) = 0$. The value of σ_i can be obtained by means of a numerical analysis. Figure 4 presents the plot of σ_i^2 as a function of d_i.

From here, by inserting the values $\mu_i^- = \mu_i^+ = 0$, and σ_i into Theorem 2.4.4 of Ben-Tal et al. (2009) (see Appendix B.1), we obtain that robust constraint (46) with $\mathcal{U}$ defined as above, is a safe tractable approximation of chance constraint (42). By the same theorem, it holds that (44) is precisely the robust counterpart of (46). □

Constraint system (44) involves only linear and second-order conic constraints, making it highly tractable even for large-dimensional problems.

The second safe approximation is tighter and relies on the somewhat more complicated mathematical machinery of Ben-Tal et al. (2009).

Theorem 2 *If there exists $\alpha > 0$ such that $(\mathbf{x}, \alpha)$ satisfies the constraint*

$$(\mathbf{a}^0)^T\mathbf{x} - b_0 + \alpha \log\left(\sum_{i=1}^{n_\mathbf{z}} \left(d_i \cosh\left(\frac{(\mathbf{a}^i)^T\mathbf{x} - b_i}{\alpha}\right) + 1 - d_i\right)\right) + \alpha \log(1/\epsilon) \leq 0, \tag{47}$$

then $\mathbf{x}$ satisfies constraint (42). That, is (47) is a safe approximation of (42).

Proof. See Appendix B.2. □

Similar to Theorem 1, one can construct an explicit convex uncertainty set $\mathcal{U}$ for which (47) is the robust counterpart of (46) corresponding to $\mathcal{U}$. Constraint (47) is convex in $(\mathbf{x}, \alpha)$, being a sum of a linear function and $n_\mathbf{z}$ perspective functions of the convex log-sum-exp function, see Boyd and Vandenberghe (2004). For that reason, it can be handled with convex optimization algorithms such as Interior Point Methods.

5.3. Towards better safe approximations - exponential polynomials

Ben-Tal et al. (2009) discuss the fact that the bounds obtained using a single exponential function can still be improved by, instead of the moment-generation function, constructing the worst-case expectation of exponential polynomials:

$$\gamma(s) = \sum_{\nu=0}^{L} c_\nu \exp\{\omega_\nu s\}, \tag{48}$$

to bound the probability of constraint violation, where $c_\nu, \omega_\nu, \nu = 0, \ldots, L$ are complex numbers and

$$\gamma(\cdot) \text{ is convex and nondecreasing, } \gamma(s) \geq 0, \quad \gamma(0) \geq 0, \quad \gamma(s) \to 0, \quad s \to -\infty. \tag{49}$$

The worst-case expectation of the exponential polynomial $\gamma(s)$, similar to the worst-case expectation of the moment-generating function (43), can then be used to obtain better upper bounds on $\mathbb{P}\left(\mathbf{a}^T(\mathbf{z})\mathbf{x} > \mathbf{b}(\mathbf{z})\right)$. In fact, the bound found in Theorem 2 is obtained using a special case of (48), where $L = 0$, $c_0 = \omega_0 = 1$. The difficulty of using general polynomials (48) lies in the (un)availability of tight, computationally tractable upper bounding function $\Psi(\mathbf{w})$ on (48):

$$\mathbb{E}_{\mathbb{P}} \gamma \left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i \right) \leq \Psi(\mathbf{w}), \quad \forall \mathbb{P} \in \mathcal{P}.$$

In the following, we show that under (μ, d) information, the result of BH can be easily applied in this case as well. Indeed, the corresponding supremum over $\mathcal{P}_{(\mu,d)}$ is given by:

$$\begin{aligned} \Psi(\mathbf{w}) &= \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \gamma \left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} z_i w_i \right) \\ &= \sum_{\nu=0}^{L} c_\nu \exp\{\omega_\nu w_0\} \prod_{i=1}^{n_{\mathbf{z}}} \left(d_i \sinh(\omega_\nu w_i) + 1 - d_i \right). \end{aligned} \tag{50}$$

Now, we can use Proposition 4.3.1 from Ben-Tal et al. (2009) to obtain the following result.

Theorem 3 *Consider an exponential polynomial $\gamma(s)$ satisfying (49), the corresponding function $\Psi(\mathbf{w})$ and the set Γ_ϵ such that:*

$$\Gamma_\epsilon = \{\mathbf{x} : \quad \exists \alpha > 0 : \quad \Psi(\alpha \mathbf{w}) \leq \epsilon\}, \qquad w_i = (\mathbf{a}^i)^T \mathbf{x} - b_i, \quad i = 1, \ldots, n_{\mathbf{z}}. \tag{51}$$

Then, any $\mathbf{x} \in \mathrm{cl}\Gamma_\epsilon$ is also feasible for the chance constraint (42).

Proof. See Appendix B.3. □

It is also important to note that constraint (51) is convex representable in $(\mathbf{x}, \alpha)$. Theorem 3 extends the results of Ben-Tal et al. (2009), which provides a safe approximation using only known supports and means of the components z_i.

5.4. Safe tractable approximations - simple experiment

We illustrate now the differences between (i) the power of the three approximations of the previous sections, and (ii) knowing and not knowing the MAD. We consider here the following problem from Section 4.3.6.2 of Ben-Tal et al. (2009):

$$\begin{aligned} &\max_{x_0} \; x_0 \\ &\text{s.t.} \sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{P}\left(x_0 + \sum_{i=1}^{n_{\mathbf{z}}} x_i z_i > 0\right) \leq \epsilon \\ &\qquad x_i = 1, \quad i = 1,\ldots,n_{\mathbf{z}}. \end{aligned} \tag{52}$$

We solve this problem using all three safe tractable approximations of the chance constraint, for two different cases:

- no information about d - which corresponds to setting $d_i = 1, i = 1,\ldots,n_{\mathbf{z}}$ (the largest possible value for d_i, see Remark 1, page 6, about the Edmundson-Madansky bound when d is maximum possible),
- knowing that $d_i = d = 0.5, i = 1,\ldots,n_{\mathbf{z}}$.

We consider three probability levels: $\epsilon \in \{10^{-1}, 10^{-2}, 10^{-3}\}$ and $n_{\mathbf{z}} = 128$. Whereas safe approximations corresponding to Theorems 1 and 2 are well-defined by the theorems, we need to choose the exponential polynomial used in the approximation of Theorem 3. As Ben-Tal et al. (2009), we use the polynomial

$$\gamma_{d,T}(s) = \exp(s)\chi_{c^*}(s),$$

where

$$\chi_{c^*}(s) = \sum_{\nu=0}^{d} \left(c^*_\nu \exp(\imath\pi\nu s/T) + \overline{c^*_\nu}\exp(-\imath\pi\nu s/T)\right)$$

is an optimal solution of the best uniform approximation problem:

$$\mathbf{c}^* \in \arg\min\left\{\max_{-T\leq s\leq T} |\exp(s)\chi_{\mathbf{c}}(s) - \max\{1+s,0\}| : \quad 0 \leq \chi_{\mathbf{c}}(s) \leq \chi_{\mathbf{c}}(0) = 1, \quad \forall s \in \mathbb{R}\right\}$$

and $\exp(s)\chi_{\mathbf{c}}(s)$ is convex nondecreasing on $[-T,T]$, with parameter values $d = 11$ ('degree of approximation' of the function $\max\{1+s,0\}$), $T = 8$ ('window width' on which the function $\max\{1+s,0\}$ is approximated).

Table 7 presents the results. First, for all safe approximations and all security levels, one can observe a substantial value of having the information about parameters d_i. For example, for $\epsilon = 0.01$ and safe approximation according to Theorem 3, the optimal solution obtained without knowledge of d is -30.55, whereas the corresponding number for known $d = 0.5$ is -21.69. A similar pattern can be observed for other values of ϵ and other approximations.

Secondly, one can see the increasing power of the safe tractable approximations that use exactly the same information. For example, for $\epsilon = 10^{-3}$ and $d = 0.5$ the subsequent optimal values are

Table 7 Maximum values of x_0 in problem (52), depending on the safe tractable approximation used, probability bound, and the assumptions on the knowledge about d.

ϵ		Maximum x_0		
	Safe approximation	Theorem 1	Theorem 2	Theorem 3
10^{-1}	Unknown d	-24.28	-24.21	-20.43
	$d = 0.5$	-17.16	-17.14	-14.48
10^{-2}	Unknown d	-34.34	-34.13	-30.55
	$d = 0.5$	-24.27	-24.20	-21.69
10^{-3}	Unknown d	-42.05	-41.67	-38.34
	$d = 0.5$	-29.73	-29.60	-27.25

-29.73, -29.60 and -27.25. For all values of ϵ and d there is a bigger difference between the second and third tractable approximation than between the first and second.

This example illustrates the extra power due to the knowledge of d, giving a strong reason to estimate this quantity in order to obtain better chance constraint approximations. Also, the difference between the quality of safe tractable approximations of Theorems 1, 2, and 3 illustrates that the power of exponential polynomial-based approximations make them an attractive tool if the parameters a, b, μ, and d can be estimated with sufficient precision.

5.5. Antenna array - chance constraints

Here we consider an application of our safe tractable approximations to scalar chance constraints to an antenna design problem under implementation error uncertainty.

Antenna is a device for sending and receiving electromagnetic signals. The signal emitted by an antenna corresponds to a function called *diagram*. An antenna array is a system of several antennas whose diagram is the sum of the diagrams of the individual components. In designing the antenna array the engineer can amplify the power sent to each of the antennas so as to obtain an array whose diagram satisfies some desired properties. For more information we refer the reader to Section 3.3 of Ben-Tal et al. (2009).

In our example, the setting is as follows. There are $n = 40$ ring-shaped antennas belonging to the XY plane in $\mathbb{R}^3$. The radius of the k-th antenna is defined as k/n and the diagram $D(\phi)$ of the antenna array is defined as a sum of diagrams $D_k(\phi)$ of the antennas:

$$D_k(\phi) = \frac{1}{2}\int_0^{2\pi} \cos\left(\frac{2\pi k}{40}\cos(\phi)\cos(\gamma)\right) d\gamma, \quad k = 1, \ldots, 40.$$

The objective of the problem is to minimize the maximum of the diagram modulus in the sidelobe angle $0 \leq \phi \leq 70°$:

$$\max_{0\leq\phi\leq 70°} \left| \sum_{k=1}^{n} x_k D_k(\phi) \right|,$$

where x_k are the decision variables - the amplification weights, subject to the restrictions that:

- the diagram in the interval $77° \leq \phi \leq 90°$ is nearly uniform:

$$77° \leq \phi \leq 90° \quad \Rightarrow \quad 0.9 \leq \sum x_k D_k(\phi) \leq 1,$$

- the diagram in other angles is not too large:

$$\left| \sum_{k=1}^{n} x_k D_k(\phi) \right| \leq 1, \quad 70° \leq \phi \leq 77°.$$

We assume that the implementation error affects the weight of the k-th antenna in the following fashion:

$$x_k \mapsto \tilde{x}_k = (1 + z_k \rho) x_k, \quad k = 1, \ldots, n,$$

where z_k, $k = 1, \ldots, n$, are independent random variables with supports contained in $[-1, 1]$, with mean 0 and MAD equal to d:

$$\mathcal{P} = \{\mathbb{P} : \quad \text{supp}(z_i) \subset [-1, 1], \quad \mathbb{E}_{\mathbb{P}}(z_i) = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d, \quad z_i \perp\!\!\!\perp z_j, \quad \forall i \neq j\}.$$

The problem to be solved is:

$$\begin{array}{lllll} \min\limits_{\tau, \mathbf{x}} & \tau & & & \\ \text{s.t.} & \mathbb{P}\left(\sum D_k(\phi_i)\tilde{x}_k \leq \tau\right) & \geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 0 \leq \phi_i \leq 70° \\ & \mathbb{P}\left(\sum D_k(\phi_i)\tilde{x}_k \geq -\tau\right) & \geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 0 \leq \phi_i < 70° \\ & \mathbb{P}\left(\sum D_k(\phi_i)\tilde{x}_k \leq 1\right) & \geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 70° \leq \phi_i < 77° \\ & \mathbb{P}\left(\sum D_k(\phi_i)\tilde{x}_k \geq -1\right) & \geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 70° \leq \phi_i \leq 77° \\ & \mathbb{P}\left(\sum D_k(\phi_i)\tilde{x}_k \leq 1\right) & \geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 77° \leq \phi_i \leq 90° \\ & \mathbb{P}\left(\sum D_k(\phi_i)\tilde{x}_k \geq 0.9\right) & \geq 1 - \epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 77° \leq \phi_i \leq 90°, \end{array} \tag{53}$$

where $\phi_1, ..., \phi_N$ is a 'fine grid' of equidistant points in $[0°, 90°]$.

In the numerical experiment we set $N = 400, d = 0.5$. The chance constraints are reformulated using the ball-box uncertainty set of Theorem 1. We solve the problem in the following settings:

- nominal solution, with $\rho = 0$ (no implementation error),
- robust solutions where both ϵ and ρ can get values in $\{0.001, 0.01, 0.05\}$.

In total, we obtain 10 solutions. For each of them we report the optimal (worst-case) objective value. Next to that, we conduct a simulation study for each solution, where the realized error magnitude $\widehat{\rho}$ can take the values in $\{0.001, 0.01, 0.05\}$. In this study, for each solution we sample 10^4 scenarios of the implementation error $\hat{\mathbf{z}} \in [-1, -1]^n$ and we report on:

- the percentage of samples for which at least one of the constraints of the problem (53) is violated,
- the perturbed objective function value $\hat{\tau} = \max\limits_{0° \leq \phi_i \leq 70°} |\sum \hat{x}_k D_k(\phi_i)|$.

Table 8 Minimum worst-case τ^* and mean simulated values of $\widehat{\tau}$ for each of the solutions. ϵ and ρ denote the parameter values used in problem (53) to obtain a given solution, and $\widehat{\rho}$ denotes the error magnitude of the given sample of 10^4 implementation error vectors $\mathbf{z}$.

	ϵ	-	0.001			0.01			0.05		
	ρ	0	0.001	0.01	0.05	0.001	0.01	0.05	0.001	0.01	0.05
Worst-case $\tau \times 10^{-2}$	-	2.68	5.86	8.12	34.05	5.77	7.92	16.04	5.68	7.71	11.80
Average $\hat{\tau} \times 10^{-2}$	$\widehat{\rho} = 0.001$	4270	5.64	7.66	31.91	5.59	7.53	14.70	5.54	7.38	10.84
	$\widehat{\rho} = 0.01$	42706	7.03	7.78	31.91	6.96	7.65	14.70	6.91	7.51	10.85
	$\widehat{\rho} = 0.05$	213534	14.00	8.61	31.91	13.73	8.46	14.94	13.64	8.36	11.12

Table 9 Empirical probabilities of violating at least one constraint. 'Violation probability (%)' denotes the percentage of simulated implementation error vectors for which at least one of the constraints of the problem (53) has been violated.

	ϵ	-	0.001			0.01			0.05		
	ρ	0	0.001	0.01	0.05	0.001	0.01	0.05	0.001	0.01	0.05
Violation probability (%)	$\widehat{\rho} = 0.001$	100	0.00	0.00	0.00	0.02	0.00	0.00	0.18	0.00	0.00
	$\widehat{\rho} = 0.01$	100	84.39	0.00	0.00	84.87	0.00	0.00	85.67	0.20	0.00
	$\widehat{\rho} = 0.05$	100	99.57	63.89	0.00	99.66	62.97	0.03	99.61	67.35	0.43

Results are given in Tables 8 and 9. The nominal solution becomes senseless already with the implementation error $\widehat{\rho} = 0.001$. At the same time, the robust solutions yield good performance even with the largest $\widehat{\rho}$, both in terms of the $\hat{\tau}$ values and the percentage of drawings for which at least one constraint is violated.

The difference between the nominal and robust solutions can be seen in Figures 5 and 6, where the diagrams are plotted for the situations (i) with no implementation error, and (ii) with a single sample of implementation error $\widehat{\rho} = 0.001$. In both cases, the solutions yield good 'desired' diagrams in the no-error case. However, in the situation with implementation error (lower panels), the robust solution still 'fits' into the desired bounds, which is completely not the case for the nominal solution.

6. Summary

In this paper, we have considered two types of ambiguous stochastic constraints - expected feasibility constraints and chance constraints. In contrast to previous research, which employs the variance as a dispersion measure, we use the mean absolute deviation. This allows us to use the 1972 results of BH on tight upper and lower bounds on the expectation of a convex function of a random variable, and thus, to provide tractable exact robust counterparts for expected feasibility constraint and to obtain safe tractable approximations of ambiguous chance constraint. Numerical examples show the proposed methodology to perform well and, in particular, to offer

Figure 5 Nominal solution - diagram plots. Upper panel - situation without implementation error. Lower panel - implementation error, single trajectory.

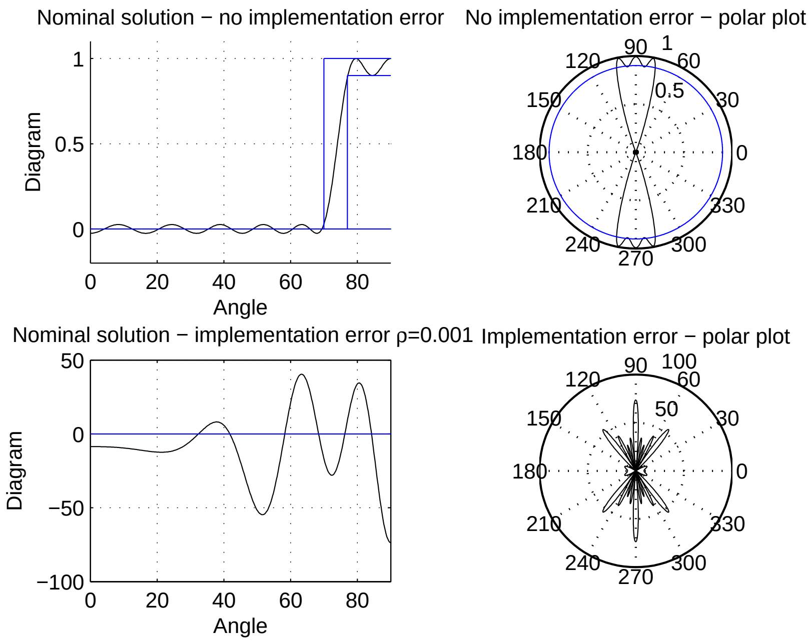
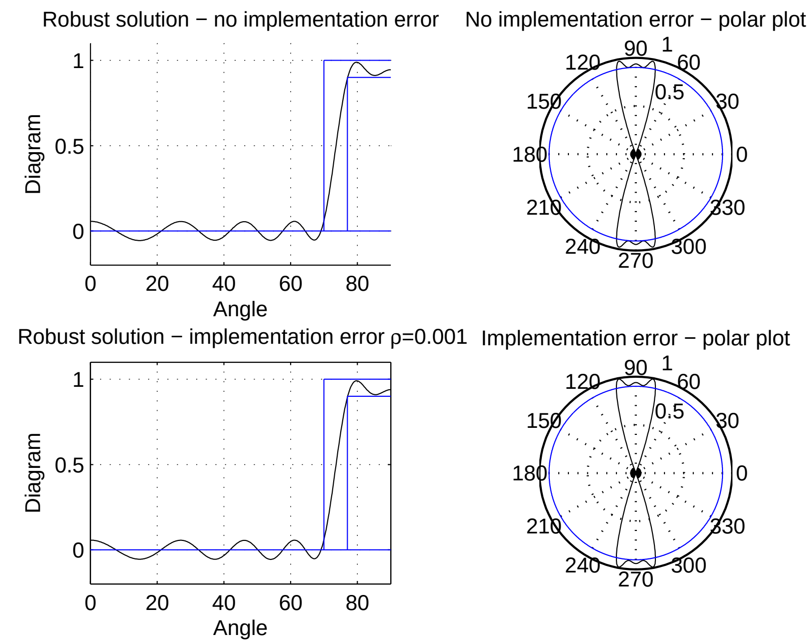


Figure 6 Robust solution - diagram plots. Upper panel - situation without implementation error. Lower panel - implementation error, single trajectory.



substantial improvements in the worst-case expected performance and probabilistic guarantees on constraints' feasibility. In particular, for the worst-case expected feasibility constraints we identify an important class of functions for which we can relax the assumption of independence of random variables needed by BH, and for which we construct highly computationally tractable approximations. Numerical experiments show that these approximations yield good practical performance and can be preferred in settings where the independence assumption on the random variables does not hold.

References

Baillie, R. T., G. G. Booth, Y. Tse, T. Zabotina. 2002. Price discovery and common factor models. *Journal of Financial Markets* **5(3)** 309–321.

Ben-Tal, A., D. den Hertog. 2011. Immunizing conic quadratic optimization problems against implementation errors. *Mathematical Programming* **149(1-2)** 265–299.

Ben-Tal, A., D. den Hertog, A. de Waegenaere, B. Melenberg, G. Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* **59(2)** 341–357.

Ben-Tal, A., D. den Hertog, J.-Ph. Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* **149(1-2)** 265–299.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust optimization*. Princeton University Press.

Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99(2)** 351–376.

Ben-Tal, A., E. Hochman. 1972. More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability* **9** 803–812.

Ben-Tal, A., E. Hochman. 1985. Approximation of expected returns and optimal decisions under uncertainty using mean and mean absolute deviation. *Zeitschrift für Operations Research* **29(7)** 285–300.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23(4)** 769–805.

Bertsimas, D., V. Goyal, X. A. Sun. 2011. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research* **36(1)** 24–54.

Birge, J. R., R. J.-B. Wets. 1987. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research* **12(1)** 149–162.

Boyd, S., L. Vandenberghe. 2004. *Convex optimization*. Cambridge University Press.

Calafiore, G. C., L. El Ghaoui. 2006. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications* **130(1)** 1–22.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55(6)** 1058–1071.

De Ruiter, F., D. den Hertog, R. Brekelmans. 2016. The impact of the existence of multiple adjustable robust solutions. *To appear in Mathematical Programming.*

Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58(3)** 595–612.

Dulá, J. H. 1992. An upper bound on the expectation of simplicial functions of multivariate random variables. *Mathematical Programming* **55(1-3)** 69–80.

Dupačová, J. 1966. On minimax solutions of stochastic linear programming problems. *Časopis Pro Pětováni Matematiky* **91** 423–430.

Dupačová, J., N. Gröwe-Kuska, W. Römisch. 2003. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming* **95** 493–511.

Edirisinghe, N. C. P. 2011. *Stochastic Programming: The State of the Art In Honor of George B. Dantzig*, chap. Stochastic Programming Approximations Using Limited Moment Information, with Application to Asset Allocation. Springer New York, 97–138.

Edirisinghe, N. C. P., W. T. Ziemba. 1992. Tight bounds for stochastic convex programs. *Operations Research* **40** 660–677.

Edirisinghe, N. C. P., W. T. Ziemba. 1994a. Bounding the expectation of a saddle function with application to stochastic programming. *Mathematics of Operations Research* **19** 314–340.

Edirisinghe, N. C. P., W. T. Ziemba. 1994b. Bounds for two-stage stochastic programs with fixed recourse. *Mathematics of Operations Research* **19** 292–313.

Edmundson, H. P. 1956. Bounds on the expectation of a convex function of a random variable. *Acta Mathematica* **30** 175–193.

El Amir, E. A. H. 2012. On uses of mean absolute deviation: decomposition, skewness and correlation coefficients. *Metron* **70(2-3)** 145–164.

Esfahani, P. M., D. Kuhn. 2015. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Available at: `http://www.optimization-online.org/DB_FILE/2015/05/4899.pdf`.

Frauendorfer, K. 1988. Slp recourse problems with arbitrary multivariate distributionsthe dependent case. *Mathematics of Operations Research* **13** 377–394.

Gallego, G. 1992. A minmax distribution free procedure for the (q, r) inventory model. *Operations Research Letters* **11(1)** 55–60.

Gallego, G., J. K. Ryan, D. Simchi-Levi. 2001. Minimax analysis for finite-horizon inventory models. *IIE Transactions* **33(10)** 861–874.

Gassmann, H. I., W. T. Ziemba. 1986. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Mathematical Programming Study* **27** 39–53.

Gastwirth, J. L. 1974. Large sample theory of some measures of income inequality. *Econometrica: Journal of the Econometric Society* **42(1)** 191–196.

Ghoudi, K., R. J. Kulperger, B. Rémillard. 2001. A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis* **79**(2) 191–218.

Goh, J., M. Sim. 2010. Distributionally robust optimization and its tractable approximations. *Operations Research* **58(4)** 902–917.

Hanasusanto, G. A., V. Roitch, D. Kuhn, W. Wiesemann. 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming* **151(1)** 35–62.

Higle, J.L., S. Sen. 1996. *Stochastic decomposition: A statistical method for large scale stochastic linear programming*. Kluwer, Dordrecht.

Iancu, D. A., N. Trichakis. 2013. Pareto efficiency in robust optimization. *Management Science* **60(1)** 130–147.

Jensen, J. L. W. V. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* **30(1)** 175–193.

Jiang, R., Y. Guan. 2015. Data-driven chance constrained stochastic program. `http://link.springer.com/article/10.1007%2Fs10107-015-0929-7`. To appear in *Mathematical Programming*.

Jolliffe, I. 2002. *Principal component analysis*. John Wiley & Sons, Ltd.

Kall, P. 1991. An upper bound for slp using first and total second moments. *Annals of Operations Research* **30** 267–276.

Karr, A. F. 1983. Extreme points of certain sets of probability measures, with applications. *Mathematics of Operations Research* **8** 74–85.

Madansky, A. 1959. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics* **30** 743–746.

Perakis, G., G. Roels. 2008. Regret in the newsvendor model with partial information. *Operations Research* **56(1)** 188–203.

Pinkse, J. 1998. A consistent nonparametric test for serial independence. *Journal of Econometrics* **84**(2) 205–231.

Popescu, I. 2007. Robust mean-covariance solutions for stochastic optimization. *Operations Research* **55(1)** 98–112.

Postek, K., D. den Hertog, B. Melenberg. 2015. Computationally tractable counterparts of distributionally robust constraints on risk measures. *To appear in SIAM Review.*

Rogosinsky, W. W. 1958. Moments of non-negative mass. *Proceedings of the Royal Society of London, Series A* **245** 1–27.

Scarf, H. E. 1958. A min-max solution of an inventory problem. K. J. Arrow, S. Karlin, H. E. Scarf, eds., *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia.

Shapiro, A., A. J. Kleywegt. 2002. Minimax analysis of stochastic problems. *Optimization Methods and Software* **17** 523–542.

Sklar, M. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8** 229–231.

Smith, R. L. 1984. Efficient monte-carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research* **32** 1296–1308.

Wiesemann, W., D. Kuhn, M. Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62(6)** 1358–1376.

Zymler, S., D. Kuhn, B. Rustem. 2013. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming* **137(1-2)** 167–198.

Appendix A: Estimating μ, d, and β

As the bounds on the expectation of a random variable depend on the parameters a, b, μ, d, and β, it is necessary to know or estimate these parameters, and decide 'how much information is actually available'. Here, we provide the reader with a simple procedure to achieve this.

First of all, it is necessary to verify the independence of the components of $\mathbf{z}$. This can be achieved using the nonparametric tests of Pinkse (1998) and Ghoudi et al. (2001). If the independence hypothesis is rejected, factorization techniques mentioned in Section 3.1 can be used to decompose the random variable into a combination of factors.

Assuming that the independence holds or is achieved by factorization, we operate here with the one-dimensional case for z, and the multi-dimensional case follows straightforwardly due to the independence of the components of $\mathbf{z}$. Appendix D describes the properties of the MAD in relation to the variance and formulas for the MAD of several important classes of probability distribution.

First, we introduce estimators of μ, d, and β and discuss their asymptotic properties. Based on these results, we provide a procedure that can be used to assess whether the amount of information available is sufficient to use the results for the (μ, d) ambiguity set or the (μ, d, β) ambiguity set.

Let $z^{(1)}, \ldots, z^{(n)}$ be a random sample of the values of z. We assume the interval $[a, b]$ to be fixed by the user. As estimators for μ, d, and β we consider

- $\widehat{\mu} = \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$, the sample mean;
- $\widehat{d} = \frac{1}{n}\sum_{i=1}^{n} |z_i - \bar{z}|$, the sample MAD;
- $\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} 1_{\bar{z},\infty}(z_i)$, the sample analogue of β.

Let $\widehat{\theta} = \left(\widehat{\mu}, \widehat{d}, \widehat{\beta}\right)^\top$ and $\theta = (\mu, d, \beta)^T$. Then we have

$$\sqrt{n}(\widehat{\theta} - \theta) = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}(z_i) + o_p(1),$$

with $\widetilde{\psi}(z) = (\widetilde{\psi}_\mu(z), \widetilde{\psi}_d(z), \widetilde{\psi}_\beta(z))^T$ defined by

$$\begin{aligned}
\widetilde{\psi}_\mu(z) &= z - \mu, \\
\widetilde{\psi}_d(z) &= 2\left((z-\mu) + ([z 1_{(\mu,\infty)}(z) - z\beta] - \frac{1}{2}d) - \mu(1_{(\mu,\infty)}(z) - \beta)\right), \\
\widetilde{\psi}_\beta(z) &= (1_{(\mu,\infty)}(z) - \beta) - (z-\mu)p(\mu),
\end{aligned}$$

where $p(\mu)$ stands for the density function of z evaluated at μ, assuming that $\mathbb{P}$ represents a continuous distribution (in which case $p(\cdot)$ is assumed to be continuous in a neighborhood of μ). The expression for $\widetilde{\psi}_\mu(z)$ is standard. The expression $\widetilde{\psi}_d(z)$ is based on Gastwirth (1974). The expression $\widetilde{\psi}_\beta(z)$ follows from arguments presented in Gastwirth (1974). As a consequence, we find for the limit distribution of $\widehat{\theta}$:

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d N(0, \text{cov}(\widetilde{\psi})). \tag{54}$$

The asymptotic covariance matrix $\text{cov}(\widetilde{\psi}) = E(\widetilde{\psi}(z)\widetilde{\psi}(z)^T)$ can be estimated consistently by

$$\widehat{\text{cov}}(\psi) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\widetilde{\psi}}(z_i)\widehat{\widetilde{\psi}}(z_i)^T,$$

with $\widehat{\widetilde{\psi}}(z_i)$ obtained from $\widetilde{\psi}(z_i)$ by replacing μ, d, and β by their estimates $\widehat{\mu}$, $\widehat{d}$, and $\widehat{\beta}$, and where $p(\cdot)$ is replaced by some (appropriately chosen) consistent estimator $\widehat{p}(\cdot)$.

We now proceed to the proper estimation of the parameters of the distribution of z. The parameters satisfy the bounds

$$a \le \mu \le b, \quad 0 \le d \le d_{\max}, \quad \underline{\beta} \le \beta \le \overline{\beta},$$

with

$$r_\beta = \overline{\beta} - \underline{\beta} = 4 - \frac{1}{2}\frac{d(b-a)}{(\mu-a)(b-\mu)}.$$

We can estimate $d_{\max}$ consistently by $\widehat{d}_{\max}$ (by estimating μ by $\widehat{\mu}$) and r_β consistently by $\widehat{r}_\beta$ (by estimating μ by $\widehat{\mu}$ and d by $\widehat{d}$). If $\widehat{d}_{\max}$ is not significantly different from 0, then there is not much empirical support for assuming that we 'know' d. Similarly, if $\widehat{r}_\beta$ is not significantly different from 0, then there is not much empirical support for assuming that we 'know' β. The (asymptotic) accuracy of $\widehat{d}_{\max}$ and $\widehat{r}_\beta$ can easily be quantified using the 'delta method', resulting in $\sqrt{n}\left(\widehat{d}_{\max} - d_{max}\right) \to_d N(0, \sigma^2_{d_{\max}})$ and $\sqrt{n}\,(\widehat{r}_\beta - r_\beta) \to_d N(0, \sigma^2_{r_\beta})$.[4] With these definitions, we present now our procedure for estimation of the information basis for the use of the bounds:

[4] The 'delta method' yields $\sigma^2_{d_{max}} = r^2 \text{var}(\widetilde{\psi}_\mu)$, with $r = \frac{\partial d_{\max}}{\partial \mu} = \frac{2(b+a-2\mu)}{b-a}$. Similarly, we have $\sigma^2_{r_\beta} = s^T \text{cov}((\widetilde{\psi}_\mu, \widetilde{\psi}_d)^T)s$, with $s = \frac{\partial r_\beta}{\partial (\mu, d)^T} = \left(-\frac{d(b-a)(b+a-2\mu)}{2((\mu-a)(b-\mu))^2}, -\frac{b-a}{2(\mu-a)(b-\mu)}\right)^T$.

1. Estimate μ by $\widehat{\mu}$, and quantify the accuracy of the latter (using the limit distribution given in (54)). Decide whether the accuracy is high enough to proceed under the assumption of a 'known' μ. If so, go to step 2.
2. Test the hypothesis $H_0 : d_{\max} = 0$ against $H_1 : d_{\max} > 0$, using as test statistic $\widehat{d}_{\max}/\sqrt{\widehat{\sigma}_{d_{\max}}/n}$. This is a one-sided test. If H_0 is rejected (H_1 accepted), go to step 3.
3. Estimate d by $\widehat{d}$, and quantify the accuracy of the latter (using the limit distribution given in (54)). Decide whether the accuracy is high enough to proceed under the assumption of a 'known' d. If so, go to step 4.
4. Test the hypothesis $H_0 : r_\beta = 0$ against $H_1 : r_\beta > 0$, using as test statistic $\widehat{r}_\beta/\sqrt{\widehat{\sigma}_{r_\beta}/n}$. This is a one-sided test. If H_0 is rejected (H_1 accepted), go to step 5.
5. Estimate β by $\widehat{\beta}$, and quantify the accuracy of the latter (using the limit distribution given in (54)). Decide whether the accuracy is high enough to proceed under the assumption of a 'known' β.

It may turn out that credible information is available only about the support, or support and the mean of z. In the first case, when only the support-including interval $[a, b]$ is known, a larger sample is needed to estimate other parameters. In case the support $[a, b]$ and μ are known, one may use the results of Edmundson-Madansky for the upper bound (see Remark 1) and Jensen for the lower bounds (see Remark 2).

Appendix B: Safe approximations of chance constraints

In this Appendix we list the relevant results from Ben-Tal et al. (2009) used to prove Theorems 1 and 2, and 3 adopted to the notation of this paper.

B.1. Safe approximation in Theorem 1

In the proof of Theorem 1 the following result is used.

Theorem 4 (Ben-Tal et al. (2009), Theorem 2.4.4) *Assume that:*

P.1. z_i, $i, \ldots, n_{\mathbf{z}}$ *are independent random variables such that* $supp(z_i) \subseteq [a_i^-, a_i^+]$, $i = 1, \ldots, n_{\mathbf{z}}$,

P.2. the distributions $\mathbb{P}_i$ *of the components* z_i *are such that*

$$\int \exp(ts) d\mathbb{P}_i(s) \leq \exp\left(\max\{\mu_i^+ t, \mu_i^- t\} + \frac{1}{2}\sigma_i^2 t\right), \quad \forall t \in \mathbb{R}, \tag{55}$$

with known constants $\mu_i^- \leq \mu_i^+$.

Then, the robust constraint

$$\mathbf{a}^T(\mathbf{z})\mathbf{x} \leq \mathbf{b}(\mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{U}, \ \textit{where } [\mathbf{a}(\mathbf{z}); \mathbf{b}(\mathbf{z})] = [\mathbf{a}^0; \mathbf{b}^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}_i^0; \mathbf{b}_i^0], \tag{56}$$

and

$$\mathcal{U} = \left\{ \mathbf{z} \in \mathbb{R}^{n_\mathbf{z}} : \exists \mathbf{u} \in \mathbb{R}^{n_\mathbf{z}} : \begin{array}{c} \mu_i^- \leq z_i - u_i \leq \mu_i^+, \quad i = 1, \ldots, n_\mathbf{z} \\ \sqrt{\sum\limits_{i=1}^{n_\mathbf{z}} \frac{u_i^2}{\sigma_i^2}} \leq \sqrt{2\log(1/\epsilon)} \\ a_i^- \leq z_i \leq a_i^+, \quad i = 1, \ldots, n_\mathbf{z} \end{array} \right\},$$

is a safe approximation of (42). Moreover, $\mathbf{x}$ *satisfies (56) if and only if there exist* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n_\mathbf{z}+1}$ *such that* $\mathbf{x}, \mathbf{u}, \mathbf{v}$ *satisfy the following set of constraints::*

$$\begin{cases} (\mathbf{a}^i)^T \mathbf{x} - b^i = u_i + v_i, \quad i = 0, \ldots, n_\mathbf{z} \\ u_0 + \sum\limits_{i=1}^{n_\mathbf{z}} \max\left\{a_i^+ u_i, a_i^- u_i\right\} \leq 0 \\ v_0 + \sum\limits_{i=1}^{n_\mathbf{z}} \max\left\{\mu_i^+ v_i, \mu_i^- v_i\right\} + \sqrt{2\log(1/\epsilon)} \sqrt{\sum\limits_{i=1}^{n_\mathbf{z}} \sigma_i^2 v_i^2} \leq 0. \end{cases}$$

B.2. Safe approximation in Theorem 2

The proof of Theorem 2 relies on the following result from Ben-Tal et al. (2009).

Theorem 5 (Ben-Tal et al. (2009), Proposition 4.2.2) *Assume that the distribution* $\mathbb{P}$ *of the random perturbation* $\mathbf{z}$ *is such that*

$$\log\left(\mathbb{E}\exp\left(\mathbf{w}^T \mathbf{z}\right)\right) \leq \Phi(\mathbf{w}),$$

where $\mathbf{w} = (w_1, \ldots, w_{n_\mathbf{z}})$ *for some known convex function* $\Phi(\cdot)$ *that is finite everywhere and satisfies* $\Phi(0) = 0$*. Then, any* $(w_0, \mathbf{w})$ *feasible for*

$$\inf_{\beta > 0} \left\{ w_0 + \beta \Phi\left(\beta^- 1 \mathbf{w}\right) + \beta \log(1/\epsilon) \right\} \leq 0$$

is feasible for the chance constraint

$$\mathbb{P}\left(w_0 + \sum_{i=1}^{n_\mathbf{z}} w_i z_i > 0 \right) \leq \epsilon.$$

Proof of Theorem 2. We show that the function $\Phi(\mathbf{w})$:

$$\Phi(\mathbf{w}) = \log(\Psi(\mathbf{w})), \quad \Psi(\mathbf{w}) = \sup_{\mathbb{P} \in \mathcal{P}_{(\mu, \mathbf{d})}} \mathbb{E}_\mathbb{P} \exp(\mathbf{w}^T \mathbf{z}) = \prod_{l=1}^{n_\mathbf{z}} \left(d_i \cosh(w_i) + 1 - d_i\right).$$

satisfies the conditions of Theorem 5. Indeed, from BH we know that $\Psi(\mathbf{w})$ gives a tight upper bound on $\mathbb{E}_\mathbb{P} \exp(\mathbf{w}^T \mathbf{z})$. Also, the function $\Phi(\mathbf{w})$ is convex as it is the log-sum-exp function, see Boyd and Vandenberghe (2004), and it holds that $\Phi(0) = 0$. Thus, it is sufficient to substitute

$$w_i := (\mathbf{a}^i)^T \mathbf{x} - b_i, \quad i = 0, \ldots, n_\mathbf{z},$$

to arrive at constraint (47) from Theorem 2. □

B.3. Safe approximation in Theorem 3

Theorem 3 follows from the following result from Ben-Tal et al. (2009):

Theorem 6 (Ben-Tal et al. (2009), Proposition 4.3.1) *Consider a generating function $\gamma(s)$ satisfying (49). Let $\Psi(\mathbf{w})$ be a finite convex function satisfying*

$$\Psi(\mathbf{w}) \geq \mathbb{E}_{\mathbb{P}}\left(\gamma\left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i\right)\right), \quad \Psi(\mathbf{w} + t[-1, 0, \ldots, 0]) \to 0, \ \textit{when } t \to \infty.$$

Then, the inequality

$$\inf_{\beta > 0} \left(\beta \Psi(\beta^{-1}\mathbf{w}) - \beta\epsilon\right) \leq 0$$

is a safe approximation of the chance constraint

$$\mathbb{P}\left(w_0 + \sum_{i=1}^{n_{\mathbf{z}}} w_i z_i > 0\right) \leq \epsilon.$$

Proof of Theorem 3. The result follows from using $\Psi(\mathbf{w})$ defined as in (50). This function clearly satisfies the conditions of Theorem 6. Then, the only remaining part is substituting the relevant terms for $w_i, i = 0, \ldots, n_{\mathbf{z}}$. □

Appendix C: Worst-case expectation of $\exp(\mathbf{w}^T\mathbf{z})$ without independent components

We now consider obtaining an upper bound on $\exp(\mathbf{w}^T\mathbf{z})$ using the results of Wiesemann et al. (2014), where the components of the random variable $\mathbf{z}$ are not assumed to be independent. For that reason, the distributional uncertainty set is given by:

$$\mathcal{P}' = \{\mathbb{P}: \quad \text{supp}(z_i) \subseteq [-1, 1], \quad \mathbb{E}_{\mathbb{P}} z_i = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d_i, \quad i = 1, \ldots, n_{\mathbf{z}}\}.$$

To obtain the worst-case expectation, one needs to solve the following problem:

$$\begin{aligned} \min_t \ & t \\ \text{s.t. } & \mathbb{E}_{\mathbb{P}} \exp(\mathbf{z}^T\mathbf{w}) \leq t, \quad \forall \mathbb{P} \in \mathcal{P}' \end{aligned} \tag{57}$$

The uncertainty set for the distributions $\mathbb{P}$ in their framework is:

$$\mathcal{P}' = \left\{ \mathbb{P}: \begin{array}{c} \mathbb{E}_{\mathbb{P}}\left(\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{z} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \mathbf{u}\right) = \begin{bmatrix} \mathbf{0} \\ \mathbf{d} \end{bmatrix} \\ \mathbb{P}((\mathbf{z}, \mathbf{u}) \in \mathcal{C}) = 1 \end{array} \right\}, \tag{58}$$

where $\mathcal{C} = \{(\mathbf{z}, \mathbf{u}): \quad -\mathbf{1} \leq \mathbf{z} \leq \mathbf{1}, \quad \mathbf{u} \geq \mathbf{z}, \quad \mathbf{u} \geq -\mathbf{z}, \quad \mathbf{u} \leq \mathbf{1}, \quad \mathbf{u} \geq \mathbf{0}\}$. Then, the problem to solve is equivalent to:

$$\begin{aligned} \min_{\boldsymbol{\kappa}, \boldsymbol{\lambda} \geq 0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, t} \ & t \\ \text{s.t. } & \boldsymbol{\beta}_2^T \mathbf{d} + \mathbf{1}^T(\boldsymbol{\kappa} - \boldsymbol{\lambda}) \leq t \\ & \mathbf{z}^T \boldsymbol{\beta}_1 + \mathbf{u}^T \boldsymbol{\beta}_2 + \mathbf{1}^T(\boldsymbol{\kappa} - \boldsymbol{\lambda}) \geq \exp(\mathbf{z}^T\mathbf{w}), \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C} \end{aligned} \tag{59}$$

The last line of (59) involves a constraint on the function $\exp(\mathbf{z}^T\mathbf{x})$ over $\mathcal{C}$. Since $\exp(\mathbf{z}^T\mathbf{x})$ is strictly convex in $\mathbf{z}$, an equivalent reformulation of such a constraint would have to take into account all $3^{n_{\mathbf{z}}}$ vertices of $\mathcal{C}$. The number $3^{n_{\mathbf{z}}}$ comes from the fact that per component, the uncertainty set is a triangle $\mathcal{C}_i = \{(z_i, u_i): \quad -1 \leq z_i \leq 1, \quad u_i \geq z_i, \quad u_i \geq -z_i, \quad u_i \leq 1\}$.

Appendix D: Properties of the MAD

In this Appendix we provide some properties of the MAD and the formulas for several well-known probability distributions, based on Ben-Tal and Hochman (1985).

If we denote by σ^2 the variance of the random variable z, whose distribution is known to belong to the set $\mathcal{P}_{(\mu,d)}$ (see 6, page 5), then it holds that:

$$\frac{d^2}{4\beta(1-\beta)} \leq \sigma^2 \leq \frac{d(b-a)}{2}.$$

In particular, since

$$d^2 \leq 4\beta(1-\beta)\sigma^2 \leq \sigma^2,$$

it holds that $d \leq \sigma$. For a proof, we refer the reader to Ben-Tal and Hochman (1985). For several specific distibutions, an explicit formula for d is available:

- Uniform distribution on $[a,b]$:
$$d = \frac{1}{4}(b-a)$$
- Normal distribution $N(\mu,\sigma^2)$:
$$d = \sqrt{\frac{2}{\pi}}\sigma$$
- Gamma distribution with parameters λ and k (for which $\mu = k/\lambda$):
$$d = \frac{2k^k}{\Gamma(k)\exp(k)}\frac{1}{\lambda}.$$

Ben-Tal and Hochman (1985) provide an explicit formula for the MAD for general *stable distributions*. A stable distribution is defined via its *location parameter* κ, *scale parameter* $D > 0$, *measure of skewness* $-1 \leq \lambda \leq 1$, and *characteristic exponent* $0 < \alpha \leq 2$. Important distributions belonging to this class are, for example, the normal and Cauchy distributions. The characteristic function of a stable distribution is given by

$$\log \Psi_z(t) = \log \mathbb{E}\exp(\imath tz) = \imath\kappa t - D|t|^\alpha \left(1 + \imath\lambda \text{sign}(t)\tan\left(\frac{1}{2}\pi\alpha\right)\right).$$

Stable distributions are are the only possible limiting laws for sums of independent identically distributed random variables. For properties of the stable distributions we refer the reader to Ben-Tal and Hochman (1985), who prove that for $1 < \alpha \leq 2$ the MAD of a stable random variable is given by:

$$d = D^{1/\alpha}H(\lambda,\alpha),$$

where

$$H(\lambda,\alpha) = \frac{2}{\pi}\frac{\Gamma(1-1/\alpha)}{(1+A^2)^{(\alpha-1)/2a}}\left(\cos\left((1-1/\alpha)\arctan A\right) + A\sin\left((1-1/\alpha)\arctan A\right)\right),$$

and $A = \lambda\tan\left(\frac{1}{2}\alpha\pi\right)$. In case of $\alpha \leq 1$ the mean of the random variable z does not exist.

Electronic Companion

Appendix EC.1: Worst-case expectations: synthesis of antenna array

In this companion we illustrate the use of the (μ, d) results in the context of incorporation of the implementation error in problems with nonlinear constraints. We consider the antenna design problem from Section 7.1.2 of Ben-Tal et al. (2009). We first introduce some necessary properties of antenna design.

The directional distribution (radiation pattern) of energy sent by a single antenna can be described in terms of an *antenna diagram* which is a complex-valued function. Its interpretation (in polar coordinates) is that the modus of the diagram stands for the amplitude of the radiation intensity at a given (fixed) distance whereas the angle of the complex number stands for the wave length (frequency). The modulus of the diagram can be changed by the amount of power allocated to the given antenna. If a device consists of more than one antenna, its diagram is a sum of the diagrams of the particular antennas. Therefore, it is possible to manipulate the power allocated to multiple antennas so that the diagram of an entire device is as close to (some) desired function as possible.

In this problem, n harmonic oscillators are placed at the points $k\mathbf{i}$, $k = 1, \dots, n$, with $\mathbf{i}$ being the unit vector in the direction of the x-axis in $\mathbb{R}^3$. The objective is to concentrate the energy sent by the antennas within a certain region of the 3-D space, defined using the angle that 3-D directions make with the x axis. The diagram of the k-th antenna sent in direction e is given by:

$$D_k(\phi) = \exp\left(2\pi\imath\cos(\phi)k/\lambda\right),$$

where ϕ is the angle between direction e and the direction $\mathbf{i}$ of the X-axis, λ is the wavelength, and $\imath$ is the imaginary unit. With complex weights vector $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{C}^n$, the diagram of the antenna array is the sum of diagrams of the antennas:

$$D(\phi) = \sum_{k=1}^{n} x_k D_k(\phi).$$

Energy sent by an antenna in the direction given by an angle ϕ from the x-axis is proportional to the L_2 norm of the diagram. The objective is to send as much energy as possible into the region $\phi \in [0, \Delta]$ by minimizing the weighted L_2 norm of the diagram $D(\cdot)$ in the sidelobe angle $(\mathcal{SA})$ $\Delta \le \phi \le \pi$:

$$\|D(\cdot)\|_{\mathcal{SA}} = \left(\frac{1}{1+\cos(\Delta)} \int_{\Delta}^{\pi} |D(\phi)| \sin(\phi) d\phi \right)^{1/2}.$$

The quantity $\|D(\cdot)\|_{\mathcal{SA}}$ can also be formulated as $\|\mathbf{A}\mathbf{x}\|$ where $\mathbf{A} \in \mathbb{C}^{n\times n}$ is such that

$$\mathbf{A} = \mathbf{H}^{1/2}, \quad \mathbf{H} \in \mathbb{C}^{n\times n} : H_{pq} = \frac{1}{1+\cos\Delta} \int_{\Delta}^{\pi} D_p(\phi)\overline{D_q(\phi)} \sin(\phi) d\phi.$$

For the problem to be bounded, a normalization restriction is added: $\Re\left(D(0)\right) \geq 1$, where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of a complex number. Weights x_k represent the electric power sent to each of the antennas and as such, are subject to implementation error. We assume that the weights x_k are distorted by the relative implementation error $\eta_k \in \mathbb{C}$ in the following fashion:

$$x_k \mapsto (1+\eta_k)x_k.$$

We assume that the real and imaginary parts of the implementation error are independent random variables with supports included in the interval $[-\rho,\rho]$, with mean 0 and MAD equal to $\theta\rho$:

$$\begin{aligned} \mathcal{P} = \{\mathbb{P} : \operatorname{supp}(\Im(\eta_k)), \operatorname{supp}(\Re(\eta_k)) \subset [-\rho,\rho], \quad & \mathbb{E}_{\mathbb{P}}\Re(\eta_k) = \mathbb{E}_{\mathbb{P}}\Re(\eta_k) = 0, \\ & \mathbb{E}_{\mathbb{P}}|\Im(\eta_k)| = \mathbb{E}_{\mathbb{P}}|\Re(\eta_k)| = \theta\rho, \quad \Im(\eta_k) \perp \Re(\eta_k), \quad k = 1,\dots,n\}. \end{aligned}$$

The optimization problem is:

$$\begin{aligned} \min_{\tau,\mathbf{x}} \ & \tau \\ \text{s.t.} \ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\|\mathbf{A}\mathbf{x}(\boldsymbol{\eta})\| \leq \tau \\ & \Re\left(\sum_{k=1}^{n} x_k(1+\eta_k)D_k(0)\right) \geq 1, \quad \forall \boldsymbol{\eta} \in \operatorname{supp}(\boldsymbol{\eta}), \end{aligned} \tag{EC.1}$$

where $\mathbf{x}(\boldsymbol{\eta}) = [x_1(1+\eta_1),\dots,x_n(1+\eta_n)]^T$, $\operatorname{supp}(\boldsymbol{\eta}) = \operatorname{supp}(\eta_1) \times \ldots \times \operatorname{supp}(\eta_n)$. The second constraint in the problem can be reformulated as a deterministic constraint:

$$\Re\left(\sum_{k=1}^{n} x_k D_k(0)\right) \geq 1 + \rho\sum_{k=1}^{n} |\Re\left(x_k D_k(0)\right)| + \rho\sum_{k=1}^{n} |\Im\left(x_k D_k(0)\right)|.$$

We solve problem (EC.1) with $n = 5$ antennas, wavelength $\lambda = 8$ and $\Delta = \pi/6$ in two ways:

- **nominal:** in this case we assume $\rho = 0$ (no implementation error)
- **robust:** we assume $\rho = 0.01$ (that is, implementation error of 1%) and $\theta = 0.5$.

To compare the nominal and robust solutions, we sample uniformly 10^4 random perturbations $\hat{\boldsymbol{\eta}}$ from the set $\mathcal{E}(\widehat{\rho}) = \{\boldsymbol{\eta} : -\widehat{\rho} \leq \Re(\eta_k), \Im(\eta_k) \leq \widehat{\rho}, \quad k = 1,\dots,n\}$, with $\widehat{\rho} \in \{0.01, 0.03, 0.05, 0.1\}$ and compute the value $\|D(\cdot)\|_{\mathcal{SA}}$ for $\mathbf{x}(\hat{\boldsymbol{\eta}})$. Since the normalization condition may not hold with perturbation, we normalize the diagrams $D(\cdot)$ in such a way that $|D(0)| = 1$. Table EC.1 presents the results.

The nominal solution performs well only in case of no implementation error, yielding an average value of 0.204, compared to 0.260 for the robust solution. However, already with the relative

Table EC.1 Results of the antenna design experiment. The numbers in the columns are the mean values of simulated $\|D(\cdot)\|_{\mathcal{SA}}$ (to be minimized in the optimization problem). The numbers in brackets are standard deviations.

	Simulated $\|D(\cdot)\|_{\mathcal{SA}}$	
Solution	Nominal	Robust
$\widehat{\rho}=0$	**0.204 (0.00)**	**0.260 (0.00)**
$\widehat{\rho}=0.01$	**0.424 (0.19)**	**0.262 (0.00)**
$\widehat{\rho}=0.03$	1.107 (1.41)	0.278 (0.01)
$\widehat{\rho}=0.05$	1.223 (1.32)	0.308 (0.03)
$\widehat{\rho}=0.1$	1.277 (1.78)	0.424 (0.13)

implementation error equal to 1%, the robust solution performs significantly better, yielding an average value 0.262 (st. dev. 0.0016), compared to 0.424 (0.19) for the nominal solution. This relationship grows even bigger for larger error values, compare 1.277 (1.78) to 0.424 (0.13) in case of 10% relative implementation error. This illustrates that the (μ, d) results provide a good way of tackling the implementation error in nonlinear constraints in a distributionally robust way.

Appendix EC.2: Evaluating the MAD of $\mathbf{a}^T\mathbf{z}$ using the results of Wiesemann et al. (2014)

Problem (35) is equivalent, in line with the methodology of Wiesemann et al. (2014), to:

$$\sup_{\mathbb{P}_{(\mathbf{z},\mathbf{u})}\in\mathcal{P}'} \mathbb{E}_{\mathbb{P}_{(\mathbf{z},\mathbf{u})}} \max\{\mathbf{a}^T\mathbf{z}-\mathbf{a}^T\boldsymbol{\mu}, -\mathbf{a}^T\mathbf{z}+\mathbf{a}^T\boldsymbol{\mu}\} \tag{EC.2}$$

where

$$\mathcal{P}' = \left\{\mathbb{P} : \begin{array}{c}\mathbb{E}_{\mathbb{P}}\left(\mathbf{A}\mathbf{z}+\mathbf{B}\mathbf{u}\right)=\mathbf{b}\\ \mathbb{P}((\mathbf{z},\mathbf{u})\in\mathcal{C})=1\end{array}\right\}, \quad \mathcal{C}=\{(\mathbf{z},\mathbf{u}): \quad \mathbf{C}\mathbf{z}+\mathbf{D}\mathbf{u}\leq\mathbf{c}\},$$

where the vector $\mathbf{u}\in\mathbf{R}^{n_\mathbf{z}}$ consists of components u_i, each of which is an auxiliary analysis variable corresponding to the MAD of z_i. The first (moment condition) in the definition of $\mathcal{P}'$ should ensure that the first moment of $\mathbf{z}$ is equal to $\boldsymbol{\mu}$ and the first moment of $\mathbf{u}$ is equal to $\mathbf{d}$. We define thus:

$$\mathbf{A}=\begin{bmatrix}\mathbf{I}\\ \mathbf{0}\end{bmatrix}, \quad \mathbf{B}=\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}, \quad \mathbf{b}=\begin{bmatrix}\boldsymbol{\mu}\\ \mathbf{d}\end{bmatrix}.$$

The second (support) condition in the definition of $\mathcal{P}'$ should ensure that the support of $\mathbf{z}$ is the unit box and that $\mathbf{u}$ indeed corresponds to the deviation of $\mathbf{u}$. We need to ensure thus that:

$$\|\mathbf{z}\|_\infty\leq 1, \quad \mathbf{u}\geq\mathbf{z}-\boldsymbol{\mu}, \quad \mathbf{u}\geq\boldsymbol{\mu}-\mathbf{z}, \quad \mathbf{u}\geq 0, \quad \mathbf{u}\leq 1,$$

where the last condition ensures boundedness of $\mathcal{C}$, required by Wiesemann et al. (2014). We ensure these conditions by setting:

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \\ \mathbf{I} \\ -\mathbf{I} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{I} \\ -\mathbf{I} \\ \mathbf{I} \\ -\mathbf{I} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \\ \boldsymbol{\mu} \\ -\boldsymbol{\mu} \\ \mathbf{1} \\ \mathbf{0} \end{bmatrix}.$$

Wiesemann et al. (2014) prove (Theorem 1 in their paper) that under mild conditions satisfied in our case, (EC.2) is equivalent to the following LP:

$$\begin{aligned} \min_{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \geq 0, w, \boldsymbol{\beta}, \kappa} \quad & w \\ \text{s.t.} \quad & \mathbf{b}^T \boldsymbol{\beta} + \kappa \leq w \\ & \mathbf{c}^T \boldsymbol{\phi}_1 - \mathbf{a}^T \boldsymbol{\mu} \leq \kappa \\ & \mathbf{c}^T \boldsymbol{\phi}_2 + \mathbf{a}^T \boldsymbol{\mu} \leq \kappa \\ & \mathbf{C}^T \boldsymbol{\phi}_1 + \mathbf{A}^T \boldsymbol{\beta} = \mathbf{a} \\ & \mathbf{C}^T \boldsymbol{\phi}_2 + \mathbf{A}^T \boldsymbol{\beta} = -\mathbf{a} \\ & \mathbf{D}^T \boldsymbol{\phi}_1 + \mathbf{B}^T \boldsymbol{\beta} = 0 \\ & \mathbf{D}^T \boldsymbol{\phi}_2 + \mathbf{B}^T \boldsymbol{\beta} = 0. \end{aligned} \tag{EC.3}$$