European Journal of Operational Research 219 (2012) 707–718

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

On safe tractable approximations of chance constraints

Arkadi Nemirovski

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Dr., NW, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Available online 12 November 2011

Keywords:
Uncertainty modeling
Convex programming
Optimization under uncertainty
Chance constraints
Robust Optimization

ABSTRACT

A natural way to handle optimization problem with data affected by stochastic uncertainty is to pass to a chance constrained version of the problem, where candidate solutions should satisfy the randomly perturbed constraints with probability at least $1-\epsilon$. While being attractive from modeling viewpoint, chance constrained problems "as they are" are, in general, computationally intractable. In this survey paper, we overview several simulation-based and simulation-free computationally tractable approximations of chance constrained convex programs, primarily, those of chance constrained linear, conic quadratic and semidefinite programming.

1. Introduction

A typical optimization problem can be represented as

$$\min_x \{f(x) : F(x,\zeta) \leqslant 0\}, \tag{1}$$

where $x \in \mathbf{R}^n$ is the decision vector, $\zeta \in \mathbf{R}^d$ is problem's *data*, and $F(x,\zeta)$ is a vector-valued mapping taking values in certain $\mathbf{R}^m$; note that we do not restrict generality by assuming the objective to be "standard" – not affected by the data.

From now on, we restrict ourselves with *convex* problems (1), meaning that $f(x)$, same as all components of $F(x,\zeta)$ for every fixed ζ, are efficiently computable convex and continuous functions of x.

In applications more often than not the data is *uncertain* – not known exactly when the problem is being solved. If there are reasons to model the uncertain data as random vector with a known distribution P, the standard way to treat the uncertainty is to pass to the *chance constrained* version of the problem

$$\min_x \{f(x) : \mathrm{Prob}_{\zeta\sim P}\{F(x,\zeta) \leqslant 0\} \geqslant 1-\epsilon\}, \tag{2}$$

where $\epsilon \ll 1$ is a given tolerance. In many situations ζ does have stochastic nature, but the probability distribution of ζ is not known exactly; all we know is that this distribution belongs to a given family $\mathcal{P}$ of probability distributions on $\mathbf{R}^d$. In thus case, the natural way to treat the data uncertainty is to pass to the *ambiguously chance constrained* version

$$\min_x \{f(x) : \mathrm{Prob}_{\zeta\sim P}\{F(x,\zeta) \leqslant 0\} \geqslant 1-\epsilon \ \forall P \in \mathcal{P}\} \tag{3}$$

of (1).

The chance constraint based approach to modeling data uncertainty in Optimization was proposed as early as in 1958 by Charnes et al. [11] and was extended further by Miller and Wagner [20] and Prékopa [24]. There is a huge literature devoted to this topic (see [15,23,25–27,13,28,12,17,19,29] and references therein). While being quite natural, this approach, unfortunately, has a somehow restricted field of applications, since a chance constraint, even as simple looking as

$$q_P(x) := \mathrm{Prob}_{\zeta\sim P}\left\{w_0(x) + \sum_{i=1}^d \zeta_i w_i(x) \leqslant 0\right\} \geqslant 1-\epsilon, \tag{4}$$

where the functions $w_0(x),\ldots,w_d(x)$ are affine in x, is in general computationally intractable. The reason is twofold:

- the feasible set of (4) can be nonconvex, which makes problematic subsequent optimization over this set;
- even when convex, the feasible set of (4) can be intractable, since the left hand side of the constraint can be difficult to compute.

  This, e.g., is the case when ζ is uniformly distributed in the unit box. Here the feasible set of (4) is convex whenever $\epsilon \leqslant 1/2$[1]; however, it was shown by Khachiyan [18] that unless P = NP, one cannot compute the volume of the set $\left\{\zeta : w_0 + \sum_{i=1}^d \zeta_i w_i \leqslant 0\right\}$ within accuracy $\delta > 0$ in time polynomial in $\ln(1/\delta)$ and the total bit size of the coefficients $w_0,\ldots,w_d$ (assumed to be integer). Thus, even with $\zeta \sim \mathrm{Uniform}([0,1]^d)$, the constraint (4) is difficult to handle when ϵ is small.

Essentially, the only generic case when the outlined difficulties do not arise is the case of chance constraint (4) with Gaussian distribution P. Whenever chance constraint

$$\mathrm{Prob}_{\zeta\sim P}\{\zeta : F(x,\zeta) \leqslant 0\} \geqslant 1-\epsilon \quad \forall P \in \mathcal{P} \tag{5}$$

E-mail address: nemirovs@isye.gatech.edu

[1] This is so whenever P possesses a logarithmically concave and symmetric w.r.t. some point density [19].

0377-2217/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.11.006

"as it is" is computationally intractable, a natural course of actions is to replace this intractable constraint with its *safe tractable approximation* defined as a system $\mathcal{S}$ of efficiently computable convex constraints in variables x and perhaps additional slack variables u such that the projection of the feasible set of the system on the space of x-variables is contained in the feasible set of (5) ("safety"). Given such an approximation, we can replace the intractable problem of interest (3) with the its computationally tractable approximation

$$\min_{x,u}\{f(x) : (x,u) \text{ satisfies } \mathcal{S}\}; \tag{6}$$

due to the origin of the latter problem, the x-component of a feasible solution to the approximation is feasible for the problem of interest.

In this paper we survey some recent results on tractable approximations of chance constrained convex problems, with emphasis on "well structured" chance constraints, primarily, on chance constrained versions of linear, conic quadratic and semidefinite inequalities with the "natural data" affinely depending on ζ. The main body of the paper is organized as follows. We start with the most general simulation-based approach ("scenario approximation," Section 2) which allows to handle *arbitrary* chance constrained convex problems, provided that ϵ is not too small and $\mathcal{P}$ is a singleton. We then switch to chance constrained versions of "well structured" convex constraints – linear (Section 3) and conic quadratic and semidefinite (Section 4) and to simulation-free "analytical" safe tractable approximations of these constraints capable to handle arbitrarily small values of ϵ and/or only partial knowledge of the probability distribution of uncertain data. Finally, in Section 5 we revisit simulation-based approximation schemes, now – in the case of well-structured constraints, and demonstrate that here one can get rid, to some extent, of the limitations ("not too small ϵ, no ambiguity") of the plain scenario-based approach.

2. Scenario approximations

Consider chance constrained problem (2) (or, which is the same, problem (3) with a singleton $\mathcal{P} = \{P\}$) and assume that we can sample from the distribution P. In this case, one can associate with (2) its *scenario approximation*

$$\min_{x\in\mathbf{R}^n}\{f(x) : F(x,\zeta^t) \leqslant 0,\ 1 \leqslant t \leqslant N\}, \tag{7}$$

where the *scenarios* ζ^t, $1 \leqslant t \leqslant N$, are, independently of each other, sampled from the distribution P. Under our standing assumption of convexity, all problems of the form (7) are convex with efficiently computable objective and constraints, and thus are computationally tractable. Nevertheless, the (random!) approximation (7) not necessarily is safe, and one can easily present examples where the feasible set of (7), independently of the value of N, with probability 1 contains points which are infeasible for the chance constrained problem of interest (2). Note, however, that our typical goal is to use the optimal solution to (7) as a suboptimal solution to the problem of interest. Whenever this is the case, we are interested in a property which is weaker that safety: all we need is the feasibility, in terms of the "true" chance constrained problem, of the optimal solution to (7). The issue of when the latter property takes place was resolved by de Farias and Van Roy [12] and by Calafiore and Campi [7–9]. Here is the relevant result from [9] (see also [10]):

Theorem 2.1. *Assume that for every scenario sample* $\overrightarrow{\zeta_N} = (\zeta^1, \ldots, \zeta^N)$, *the scenario approximation* (7) *is either solvable with a unique optimal solution* $x_*(\overrightarrow{\zeta_N})$, *or infeasible. Given* $\beta \in (0,1)$, let the sample size N satisfy the relation

$$N \geqslant N^* := \mathrm{Ceil}(2n\epsilon^{-1}\log(12/\epsilon) + 2\epsilon^{-1}\log(2/\beta) + 2n), \tag{8}$$

Then

$$\mathrm{Prob}_{\overrightarrow{\zeta_N}\sim P\times\ldots\times P}\{x_*(\overrightarrow{\zeta_N}) \text{ is either undefined, or is feasible for (2)}\} \geqslant 1-\beta.$$

The outlined result is really striking: all what matters is the convexity of f and F in x; there are no restrictions on how the data is distributed and on how it enters the constraints; the sample size is nearly independent of the "unreliability" β, while the dependence on ϵ, $N = O(\epsilon^{-1})$, is the best possible under circumstances. At the same time, the scenario approach as presented so far has two intrinsic limitations:

- first, the (in general, unimprovable) dependence $N = O(\epsilon^{-1})$ of the "reliable" sample size on ϵ makes the approach prohibitively time consuming when ϵ is really small, like 10^{-5} or less. While "really small" values of ϵ are of no interest in decision making applications (e.g., in Finance), they become a must in engineering;
- second, it is unclear how to apply the scenario approach to ambiguously chance constrained problems; given that in applications, especially of the decision making origin, a precise, or nearly so, knowledge of P is an exception rather than a rule, this indeed is a serious drawback. While there are ways to adjust somehow the scenario approach to the ambiguously chance constrained case, the known so far results in this direction [17], with all due respect to their technical aspects, are by far less general than Theorem 2.1.

The main focus of this paper is on simulation-free "analytical" approximations of chance constraints. These approximation impose strong structural restrictions on the chance constraint in question: the latter should be of the form

$$\mathrm{Prob}_{\zeta\sim P}\{A(x,\zeta) \in \mathbf{K}\} \geqslant 1-\epsilon \quad \forall P \in \mathcal{P}, \tag{9}$$

where $A(x,\zeta)$ is affine in x, ζ being fixed, and is affine in ζ, x being fixed, and $\mathbf{K}$ is either a non-negative orthant, or the direct product of finitely many Lorentz cones, or a semidefinite cone. In other words, we intend to focus on chance constrained linear, conic quadratic and semidefinite problems with the "natural data" affinely parameterized by random vector ζ with partially known distribution. Given that linear, conic quadratic and semidefinite programming cover nearly all applications of convex optimization, this "restricted scope" still is wide enough. At the same time, it turns out that the *conic chance constraints* – those of the form (9) – under favorable circumstances admit simulation-free safe tractable approximations which allow to handle, at the same computational effort, arbitrarily small values of ϵ, and are well suited for handling the case of partially known P.

3. Approximating scalar chance constraints

In the simplest case of $\mathbf{K} = -\mathbf{R}^m_+$ the chance constraint (9) reads

$$\mathrm{Prob}_{\zeta\sim P}\left\{w_0^\ell(x) + \sum\nolimits_{i=1}^d \zeta_i w_i^\ell(x) \leqslant 0, 1 \leqslant \ell \leqslant m\right\} \geqslant 1-\epsilon \quad \forall P \in \mathcal{P} \tag{!}$$

with affine in x scalar functions $w_i^\ell(x)$. This constraint can be safely approximated by the system of scalar chance constraints

$$\mathrm{Prob}_{\zeta\sim P}\left\{w_0^\ell(x) + \sum\nolimits_{i=1}^d \zeta_i w_i^\ell(x) \leqslant 0\right\} \geqslant 1-\epsilon_\ell \quad \forall P \in \mathcal{P}, \quad 1 \leqslant \ell \leqslant m \tag{$*$}$$

with $\epsilon_\ell \geqslant 0$ satisfying $\sum_\ell \epsilon_\ell \leqslant \epsilon$ (e.g., $\epsilon_\ell = \epsilon/m$ for all ℓ). Thus, with some conservatism added, safe tractable approximation of a chance constrained *system* of scalar linear inequalities reduces to a much simpler task of safe tractable approximation of a *single* chance constrained scalar linear inequality. In this section we focus on this latter problem, the scalar chance constrained inequality in question being written down in the equivalent form

$$p_P(w) := \mathrm{Prob}_{\zeta\sim P}\left\{\zeta : \zeta^w := w_0 + \sum\nolimits_{i=1}^d \zeta_i w_i > 0\right\} \leqslant \epsilon \quad \forall P \in \mathcal{P}. \quad (10)$$

Note that we lose nothing when assuming that $w_0, w_1, \ldots, w_d$ are decision variables rather than affine functions of the decision variables x.

"**Solvable case**." Essentially, the only generic situation when (10) is tractable is one where $\mathcal{P}$ is comprised of Gaussian distributions $\mathcal{N}(\mu, \Sigma)$ on $\mathbf{R}^d$ with the "parameter" $(\mu, \Sigma = [\sigma_{ij}])$ running through a computationally tractable closed convex set $\mathcal{Z} \subset \mathbf{R}^d \times \mathbf{S}^n_+$, where $\mathbf{S}^n_+$ is the cone of positive semidefinite $n \times n$ matrices. Assuming $\epsilon \leqslant 1/2$, (10) is equivalent to the convex constraint

$$F(w) := w_0 + \max_{(\mu,\Sigma=[\sigma_{ij}])\in\mathcal{Z}} \left\{\sum\nolimits_i \mu_i w_i + \mathrm{ErfInv}(\epsilon)\sqrt{\sum\nolimits_{i,j=1}^d \sigma_{ij} w_i w_j}\right\} \leqslant 0, \quad (11)$$

where ErfInv is the inverse error function:

$$\frac{1}{\sqrt{2\pi}}\int_{\mathrm{ErfInv}(s)}^{\infty} \exp\{-r^2/2\}dr = s, \quad 0 < s < 1. \quad (12)$$

The function F clearly is convex and efficiently computable (the latter – since the value of F at a point is the maximum of an efficiently computable concave function over a computationally tractable convex set), so that (11) is a computationally tractable convex constraint.

In the sequel, we focus on the non-Gaussian case where (10) indeed needs approximation.

3.1. Generator-based approximation scheme

Following [22,2], we present here a conceptually simple safe convex approximation scheme for (10).

Let $\gamma(s)$ be a *generator* – a convex function on the axis such that $\gamma(-\infty) := \lim_{s\to-\infty}\gamma(s) = 0$ and $\gamma(0) \geqslant 1$, so that $\gamma(\cdot)$ is an upper bound on the step function "0 on $\{s \leqslant 0\}$, 1 on $\{s > 0\}$". It follows that if P is a probability distribution on the axis, then, setting $\Gamma_P(w) = \mathbf{E}_{\zeta\sim P}\{\gamma(\zeta^w)\}$, we get a convex function of w such that $\Gamma_P(w/\alpha) \geqslant p_P(w)$ for all w and all $\alpha > 0$. Therefore if $\Gamma(w)$ is a convex upper bound of the (convex, by its origin), function $\Gamma^{\mathcal{P}}(w) := \sup_{P\in\mathcal{P}}\Gamma_P(w)$, then the implication

$$\alpha > 0, G(w, \alpha) := \alpha[\Gamma(w/\alpha) - \epsilon] \leqslant 0 \Rightarrow p_P(w) \leqslant \epsilon \quad (!)$$

takes place. The function $G(w,\alpha)$ is convex in w, α in the domain $\alpha > 0$, so that the convex constraints $G(w,\alpha) \leqslant 0, \alpha > 0$ form a safe convex approximation of (10). With minimal effort, this observation can be converted to the following

Proposition 1 ([22,2]). *For every convex upper bound $\Gamma(\cdot)$ on $\Gamma^{\mathcal{P}}(\cdot)$, the function $H(w) = \inf_{\alpha>0}\alpha(\Gamma(\alpha^{-1}w) - \epsilon)$ is convex, and the convex constraint*

$$H(w) \leqslant 0 \quad (13)$$

is a safe convex approximation of the ambiguous chance constraint (10). This approximation is tractable, provided that Γ is efficiently computable.

3.1.1. The least conservative approximation: CVaR

We have defined a family, parameterized by a generator $\gamma(\cdot)$, of safe convex approximations of (10). It is easy to see [22,2] that the least conservative approximation in this family is yielded by the generator $\gamma(s) = \max[1 + s, 0]$.[2] When $\mathcal{P} = \{P\}$ is a singleton and $\Gamma \equiv \Gamma_P$, the associated constraint (13) can be easily rewritten as

$$\mathrm{CVaR}_\epsilon(\zeta^w) := \min_\beta \left[\beta + \frac{1}{\epsilon}\mathbf{E}_{\zeta\sim P}\{\max[\zeta^w - \beta, 0]\}\right] \leqslant 0. \quad (14)$$

The left hand side here is the famous *conditional value* of ζ^w *at risk* ϵ, see [16,28]. Unfortunately, the CVaR-approximation not always is computationally tractable. Indeed, seemingly the only case where $\mathrm{CVaR}_\epsilon(\zeta^w)$ admits a closed form analytical expression is the one where P is a known distribution supported on a finite set of moderate cardinality; the "ambiguous case" extension of this observation is that when all distributions from $\mathcal{P}$ are supported on a common set of moderate cardinality and the convex hull of $\mathcal{P}$ (which now is a finite-dimensional convex set) is computationally tractable, the generator $\max[1 + s, 0]$ leads to a computationally tractable safe convex approximation of (10). Beyond this favorable case, the CVaR approximation, even for a singleton $\mathcal{P}$, can be intractable, since with non-discrete multivariate P, a closed form analytic expression for CVaR typically is out of question, while computing CVaR via Monte Carlo simulation becomes completely impractical when ϵ is "really small."

3.1.2. Tractable case: Bernstein approximation

The most general safe *tractable* approximation of (10) is the *Bernstein approximation* originating from [23] and further improved in [22,2]. This scheme is applicable when the following two assumptions hold:

B.1. The entries ζ_i, $1 \leqslant i \leqslant d$, of ζ are independent with distributions P_i known to belong to given families $\mathcal{P}_i$ of probability distributions on the axis. Thus, $\mathcal{P} = \{P_1 \times \cdots \times P_d : P_i \in \mathcal{P}_i, 1 \leqslant i \leqslant d\}$;

B.2. We have at our disposal efficiently computable convex upper bounds $\Phi_i(t)$ on the (convex by their origin) worst-case, over $P_i \in \mathcal{P}_i$, logarithmic moment-generating functions $\Phi^{\mathcal{P}_i}(t) = \sup_{P_i\in\mathcal{P}_i}\ln(\mathbf{E}_{s\sim P_i}\{\exp\{ts\}\})$.

Remark 3.1. Note that if ζ is given by "factor model" $\zeta = A\eta + a$ with deterministic A, a and random η with independent entries η_j ("the factors"), one loses nothing by treating as the uncertain data η rather than ζ, thus ensuring **B.1**.

Choosing $\exp\{s\}$ as the generator $\gamma(s)$ and applying the above construction "in the logarithmic scale", one arrives at the following

Proposition 2 ([22,2]). *Let* **B.1–2** *take place. Then the function*

$$\mathcal{H}(w) = \inf_{\alpha>0}\left[w_0 + \alpha\sum\nolimits_{i=1}^d \Phi_i(w_i/\alpha) + \alpha\ln(1/\epsilon)\right]$$

is convex and efficiently computable, and the constraint

$$\mathcal{H}(w) \leqslant 0 \quad (15)$$

is a safe tractable approximation of the ambiguous chance constraint (10).

The functions $\Phi^{\mathcal{P}_i}(\cdot)$ are readily available for a wide variety of families $\mathcal{P}_i$ of "light tail" univariate probability distributions, e.g., those given by bounds on support, the first and the second moments, requirements of unimodality and symmetry, etc., see examples in [22,2]. Whenever it is the case, the Bernstein approximation in its least conservative form (i.e., with $\Phi_i = \Phi^{\mathcal{P}_i}$) is easy to use. The actual bottleneck in applying this scheme is the independence assumption **B.1**.

[2] Same as by $\gamma(s) = \max[1 + \beta s, 0]$, $\beta > 0$, since generators $\gamma(s)$ and $\gamma(\beta s)$, $\beta > 0$, yield the same constraint (13).

Example 1. Let ζ_i be independent, supported on given (finite) segments and with expectations belonging to given segments. Applying scaling, the situation reduces to the case when ζ_i are supported on $[-1,1]$ with expectations from the segments $[\mu_i^-,\mu_i^+]\subset[-1,1]$. Here

$$\Phi_i(t) := \Phi^{\mathcal{P}_i}(t) = \ln(\cosh(t) + \max\left[\mu_i^-\sinh(t), \mu_i^+\sinh(t)\right]) \leqslant \max\left[\mu_i^+ t, \mu_i^- t\right] + \frac{1}{2}t^2. \quad (16)$$

Consequently, every one of the convex constraints

$$\begin{aligned}(a)\quad & w_0 + \inf_{\alpha>0}\alpha\left[\sum_{i=1}^d \ln\left(\cosh\left(\tfrac{w_i}{\alpha}\right) + \max\left[\mu_i^-\sinh\left(\tfrac{w_i}{\alpha}\right), \mu_i^+\sinh\left(\tfrac{w_i}{\alpha}\right)\right]\right)\right.\\ & \left. + \ln(1/\epsilon)\right] \leqslant 0,\\ (b)\quad & w_0 + \sum_{i=1}^d \max\left[\mu_i^- w_i, \mu_i^+ w_i\right] + \sqrt{2\ln(1/\epsilon)}\sqrt{\sum_{i=1}^d w_i^2} \leqslant 0\end{aligned} \quad (17)$$

is a safe tractable approximation of (10).

3.1.3. Bridging Bernstein and CVaR approximations

Under favorable circumstances, one can "bridge" the Bernstein and the CVaR approximations of (10), keeping the approximation tractable and reducing its conservatism as compared to the one of the "plain" Bernstein approximation. Specifically, assume that ζ_i are independent with *known* distributions P_i such that the functions $G_i(t) = \int \exp\{ts\}dP_i(s)$, regarded as functions of complex variable t, are efficiently computable. In this case one can compute efficiently the quantity $G(w) = \mathbf{E}_{\zeta\sim P_1\times\cdots\times P_d}\{\gamma(w_0 + \sum_i w_i\zeta_i)\}$ for every *exponential polynomial* $\gamma(s) = \sum_{\ell=1}^L c_\ell \exp\{\omega_\ell s\}$ with complex coefficients c_ℓ, ω_ℓ. We now can select as $\gamma(\cdot)$ an exponential polynomial which is a generator (i.e., is convex on the real axis and satisfies $\gamma(-\infty) = 0$, $\gamma(0) \geqslant 1$) and, in addition, is close to the "best possible" generator $\max[1+s,0]$ on a large enough segment, and use this generator in the approximation scheme from the beginning of Section 3.1. Since the associated function $\Gamma_P(w) := \mathbf{E}_{\zeta\sim P_1\times\ldots\times P_d}\{\gamma(w_0 + \sum_i w_i\zeta_i)\}$ is efficiently computable:

$$\Gamma_P(w) = \sum_\ell c_\ell \exp\{\omega_\ell w_0\}\prod_{i=1}^d G_i(\omega_\ell w_i),$$

the resulting safe convex approximation of (10) is tractable. For a natural implementation of this scheme, see [2, Section 4.3.6]. Note that the outlined approach is not well suited for the ambiguous case. Indeed, assuming that ζ_i are independent with distributions P_i varying in given families $\mathcal{P}_i$, implementation of our scheme would require the ability to compute efficiently the function

$$\Gamma^{\mathcal{P}}(w) = \sup_{\substack{P_i\in\mathcal{P}_i,\\ 1\leqslant i\leqslant d}} \sum_\ell c_\ell \exp\{\omega_\ell w_0\}\prod_{i=1}^d \int \exp\{\omega_\ell w_i s\}dP_i(s),$$

which seems to be impossible in general. However, for the most interesting families $\mathcal{P}_i$ – those given by the requirement that P_i are supported in given segments with expectations satisfying known upper and lower bounds (cf. Example 1), $\Gamma^{\mathcal{P}}$ can be computed efficiently, see [2, Section 4.3.6].

3.2. Relation to Robust Optimization

Given scalar linear inequality in variables w:

$$w_0 + \sum_{i=1}^d \zeta_i w_i \leqslant 0 \quad (18)$$

with uncertain data ζ, Stochastic Programming treats the uncertainty ζ as stochastic with probability distribution P known to belong to a given family $\mathcal{P}$, and associates with (18) the chance constraint (10). In contrast to this, Robust Optimization assumes the uncertain data ζ to run through a given convex compact uncertainty set $\mathcal{Z}$ and associates with the uncertain constraint (18) its Robust Counterpart (RC)

$$w_0 + \sum_{i=1}^d \zeta_i w_i \leqslant 0 \quad \forall \zeta\in\mathcal{Z}; \quad (19)$$

feasible solutions of the RC are called *robust feasible* solutions to (18).

As it was already explained, chance constraint (10) typically is computationally intractable (although in good situations admits safe tractable approximations), while the RC (19) of (18) is computationally tractable, provided $\mathcal{Z}$ is so. A natural question is: are there any links between the chance constrained version (10) of the uncertain linear constraint (18) and the RC of this uncertain constraint? It turns out that such a link exists, and this is exactly what safe tractable approximations of (10) are about. Specifically, observe that the feasible set W_* of (10) (a) is conic (that is, $tw\in W_*$ whenever $t\geqslant 0$ and $w\in W_*$), (b) is closed, and (c) under extremely mild regularity assumptions contains the point $e = [-1;0;\ldots;0]$ in its interior.[3] Let us call a safe convex approximation of (10) *normal*, if its feasible set W in the w-space inherits the properties (a)–(c) of the "true" feasible set W_*. Since our approximation is convex, W is convex, which combines with (a)–(c) to imply that W is a proper (i.e., $W\neq\mathbf{R}^{d+1}$) closed convex cone containing the vector $e = [-1;0;\ldots;0]$ in its interior. From these properties it easily follows that setting $\mathcal{Z} = \{z\in\mathbf{R}^d : [1;z]^T w\leqslant 0\ \forall w\in W\}$, one gets a nonempty convex compact set such that $W = \{[w_0;w_1;\ldots;w_d] : w_0 + \sum_{i=1}^d w_i z_i \leqslant 0\ \forall z\in\mathcal{Z}\}$. We arrive at the following

Proposition 3 [2, Chapter 4]. *Every normal approximation of chance constraint (10) is of the Robust Counterpart form: its feasible set W is nothing but the set of all robust feasible solutions to (18), the uncertainty set being the just defined convex compact set $\mathcal{Z}$. The latter set is computationally tractable, provided that the normal approximation in question is tractable.*

Note that under mild regularity assumptions, the generator-based safe convex approximation of (10) from the beginning of Section 3.1 is normal. Invoking [2, Lemma B.1.1, Theorem B.1.2], this is the case for approximation (13), provided that the underlying convex function $\Gamma(\cdot)$ is lower semicontinuous, finite in a neighborhood of the origin, and $\Gamma(te)\to 0$ as $t\to\infty$. Assuming, in addition, that Γ is finite everywhere and is given by a Fenchel-type representation:

$$\Gamma(w) = \max_u\{w^T[Bu+b] - g_*(u)\},$$

where $g_*(\cdot)$ is a lower semicontinuous convex function with bounded level sets, the set $\mathcal{Z} := \{z = Bu + b : g_*(u) \leqslant -\epsilon\}\subset\mathbf{R}^{d+1}$ is a convex compact set, and the approximation (13) is equivalent to the Robust Counterpart

$$\sum_{i=0}^d z_i w_i \leqslant 0 \quad \forall z\in\mathcal{Z} \quad (20)$$

of the uncertainty-affected inequality (18), the uncertainty set being $\mathcal{Z}$.[4]

Example 2 (*Robust Counterpart representation of CVaR approximation*). *Let $\mathcal{P}$ in (10) be a singleton {P}, let the distribution P be supported on a finite set $\{\zeta^1,\ldots,\zeta^N\}$, and let $P\{\zeta^j\} = \pi_j$. In this case, the Robust Counterpart form of the CVaR safe approximation of (10) is (see [2, Proposition 4.3.3])*

[3] To ensure (c), it suffices to assume that all distributions from $\mathcal{P}$ admit for every $\delta > 0$ common bounded $(1-\delta)$-support, or, equivalently, that $\sup_{P\in\mathcal{P}} P\{\zeta : \|\zeta\| > R\}\to 0$ as $R\to\infty$.

[4] Note that in (18), the only certain coefficient in (10) – the coefficient 1 at w_0 – is treated as uncertain; this aesthetic drawback can be easily cured, see [2, Section 4.3.2].

$$w_0 + \sum_{i=1}^d \zeta_i w_i \leqslant 0 \quad \forall \zeta \in \mathcal{Z} = \left\{\sum_{j=1}^N u_j \zeta^j : 0 \leqslant u_j \leqslant \pi_j/\epsilon \;\forall j, \sum_j u_j = 1\right\}.$$

The RC form of Bernstein approximation. Similarly to the general generator-based case, under assumptions **B.1–2**, the Bernstein approximation of (10) is normal, provided that the functions Φ_i are convex, lower semicontinuous and finite in a neighborhood of the origin [2, Lemma B.1.1]. Assuming, in addition, that $\Phi_i(\cdot)$ are finite everywhere and the function $\Phi(w_1,\ldots,w_d) = \sum_{i=1}^d \Phi_i(w_i)$ admits a Fenchel-type representation $\Phi(w_1,\ldots,w_d) = \sup_u\{[w_1;\ldots;w_d]^T[Bu+b] - g_*(u)\}$ with lower semicontinuous convex function $g_*(\cdot)$ possessing bounded level sets, the Bernstein approximation of (10) can be represented in the Robust Counterpart form

$$w_0 + \sum_{i=1}^d z_i w_i \leqslant 0 \quad \forall z \in \mathcal{Z} = \{Bu + b : g_*(u) \leqslant \ln(1/\epsilon)\},$$

see [2, Theorem 4.2.5].

Illustration: Example 1 revisited. Consider the ambiguous chance constraint (10), and let $\mathcal{P}$ be as in Example 1. By this example, the best under circumstances Bernstein approximation of (10) is given by the efficiently computable convex inequality (17.*a*). The Robust Counterpart form of this approximation is (see [2, Example 4.2.8])

$$w_0 + \sum_{i=1}^d z_i w_i \leqslant 0 \quad \forall z \in \mathcal{Z}^{\rm Brn} = \left\{z : \sum_{i=1}^d \phi_i(z_i) \leqslant 2\ln(1/\epsilon)\right\},$$
$$\phi_i(z_i) = \begin{cases} (1+z_i)\ln\left(\frac{1+z_i}{1+\mu_i^-}\right) + (1-z_i)\ln\left(\frac{1-z_i}{1-\mu_i^-}\right), & -1 \leqslant z_i \leqslant \mu_i^-, \\ 0, & \mu_i^- \leqslant z_i \leqslant \mu_i^+, \\ \frac{1}{2}(1+z_i)\ln\left(\frac{1+z_i}{1+\mu_i^+}\right) + (1-z_i)\ln\left(\frac{1-z_i}{1-\mu_i^+}\right), & \mu_i^+ \leqslant z_i \leqslant 1, \\ +\infty, & |z_i| > 1. \end{cases} \tag{21}$$

It is easily seen that $\phi_i(z_i) \geqslant \min_{s\in[\mu_i^-,\mu_i^+]}|z_i - s|^2$, meaning that the above *Bernstein uncertainty set* $\mathcal{Z}^{\rm Brn}$ is contained in the *Ball-Box uncertainty set*

$$\mathcal{Z}^{\rm BallBox} = \underbrace{[\{z : \mu^- \leqslant z \leqslant \mu^+\} + \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}\}]}_{\mathcal{Z}^{\rm Ball}} \bigcap \{z : \|z\|_\infty \leqslant 1\} \tag{22}$$

(cf. "Ball-Box uncertainty" from [1]); the explicit form of the corresponding safe tractable approximation of (10) is given by the convex inequality

$$w_0 + \sum_{i=1}^d \max\left[\mu_i^- u, \mu_i^+ u\right] + \sqrt{2\ln(1/\epsilon)}\|u\|_2 + \|u - [w_1;\ldots;w_d]\|_1 \leqslant 0$$

in original decision variables $w_0,\ldots,w_d$ and slack variables $u_1,\ldots,u_d$. This approximation of (10), while being slightly more conservative than (21), can be straightforwardly converted to a system of conic quadratic inequalities and thus is computationally simpler than the Bernstein approximation (17.*a*). We can further extend the ball-box uncertainty set to gain in computational simplicity of the approximation at the price of certain growth in its conservatism. One way to do so is to extend the ball-box uncertainty set to the *ball uncertainty* $\mathcal{Z}^{\rm Ball}$, see (22); the explicit form of the associated approximation is nothing but (17.*b*). Another way to simplify the ball-box approximation is to extend in (22) the Euclidean ball $\{z \in \mathbf{R}^d : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}\}$ to the larger and simpler set, the ℓ_1-ball $\{z \in \mathbf{R}^d : \|z\|_1 \leqslant \sqrt{2d\ln(1/\epsilon)}\}$, thus arriving at the *budgeted uncertainty set*

$$\mathcal{Z}^{\rm Budget} = [\{z : \mu^- \leqslant z \leqslant \mu^+\} + \{z : \|z\|_1 \leqslant \sqrt{2d\ln(1/\epsilon)}\}] \bigcap \{z : \|z\|_\infty \leqslant 1\} \tag{23}$$

(cf. [3]). The associated safe tractable approximation of (10) is given by the polyhedral constraint

$$w_0 + \sum_{i=1}^d \max\left[\mu_i^- u, \mu_i^+ u\right] + \sqrt{2d\ln(1/\epsilon)}\|u\|_\infty + \|u - [w_1;\ldots;w_d]\|_1 \leqslant 0$$

in variables w and slack variables $u \in \mathbf{R}^d$; this constraint can be straightforwardly converted to a system of linear inequalities in w and additional variables. Finally, we can extend the Bernstein uncertainty set to the embedding box $\mathcal{Z}^{\rm Box} = \{z : \|z\|_\infty \leqslant 1\}$, thus ending up with the trivial, especially simple, safe tractable approximation

$$w_0 + \|[w_1;\ldots:w_d]\|_1 \leqslant 0$$

of (10); this approximation ignores all a priori information on the distribution of the uncertain data, except for the domain information $\|\zeta\|_\infty \leqslant 1$.

Discussion. The outlined results on the Robust Counterpart representation of safe normal approximations of the ambiguous chance constraint (10) can be summarized as follows: *in order to immunize a candidate solution* $w = [w_0;\ldots;w_d]$ *to the uncertainty-affected scalar linear inequality* (18) *against "nearly all"* (*namely,* up to probability ϵ) *realizations of random data* $\zeta \sim P \in \mathcal{P}$*, it suffices to immunize* w *against all realizations of* ζ *from a properly chosen convex compact uncertainty set* $\mathcal{Z}$. By itself, this statement is nearly trivial. Indeed, assuming (and this is a really weak assumption) that the distributions from $\mathcal{P}$ for every $\epsilon > 0$ admit a common bounded $(1-\epsilon)$-support Z_ϵ, the above claim would clearly take place when choosing as $\mathcal{Z}$ the closed convex hull of Z_ϵ. A non-evident, although simple, fact is that *when applicable, the Bernstein approximation typically results in uncertainty sets* $\mathcal{Z}$ *which are much smaller than common* $(1-\epsilon)$*-supports of all distributions from* $\mathcal{P}$*, at least when* d *is large.* To illustrate this point, consider the situation of Example 1 when the expectations of ζ_i are known exactly, say, $\mu_i^\pm = 0$ for all i. Here $\mathcal{P}$ contains all product-type distributions with zero means supported on the unit box $B = [-1,1]^d$; when $\epsilon < 1/2$, every common $(1-\epsilon)$-support $\mathcal{Z}$ of the distributions from B is, essentially, "as large as the unit box," e.g., the Euclidean diameter $\mathcal{Z}$ is at least $O(\sqrt{d})$, and the volume – at least $O(2^d)$. In contrast to this, the Ball-Box uncertainty set (22) is the intersection of B and the centered at the origin Euclidean ball of the radius $r(\epsilon) = \sqrt{2\ln(1/\epsilon)}$ which, for all meaningful ϵ, is just a moderate constant (say, $r(10^{-8}) \approx 6.07$); when d is large, this intersection, both in terms of its diameter and in terms of its volume, is "incomparably less" than any common $(1-\epsilon)$-support of distributions from $\mathcal{P}$. For example, when $d > r^2(\epsilon)$ and $P \in \mathcal{P}$ is the uniform distribution on the vertices of the unit box, $\mathcal{Z}^{\rm BallBox}$ does not contain a *single* realization of $\zeta \sim P$. Similar conclusions hold true for all other uncertainty sets we have presented in connection with Example 1 (except, of course, the set $\mathcal{Z}^{\rm Box}$). The bottom line is that interplay of linearity of (18) with independence of entries in ζ allows in the large scale case to build a Robust Counterpart type safe tractable approximation of (10) by using a small, as compared to the support of ζ, uncertainty set $\mathcal{Z}$. The geometry of this set depends on $\mathcal{P}$ and can be rather nontrivial, and approximating chance constraints of the form (10) is the major source of "unexpected" uncertainty sets in Robust Optimization.

Illustration. To give an impression of what can be achieved with the outlined techniques, consider the chance constrained "optimization problem"

$$\tau_*(\epsilon) = \min_\tau \left\{\tau : {\rm Prob}_{\zeta\sim P}\left\{-\tau + \sqrt{\frac{3}{256^3}}\sum_{i=1}^{256} i\zeta_i \leqslant 0\right\} \geqslant 1-\epsilon \;\forall P \in \mathcal{P}\right\}, \tag{24}$$

where $\mathcal{P}$ is the family of all product-type distributions with zero mean marginals supported on $[-1,1]$ (cf. Example 1). In other words, given independent zero mean random variables $\zeta_1,\ldots,\zeta_{256}$ taking values in $[-1,1]$, we want to bound from above the maximal, over the distributions of ζ compatible with just stated a priori information, upper ϵ-quantile $\tau_*(\epsilon)$ of the random variable $\widehat{\zeta} = \sum_i w_i\zeta_i$ with $w_i = i\sqrt{3/d^3}$, $1 \leqslant i \leqslant d = 256$ (so that $\|[w_1;\ldots;w_d]\|_2 \approx 1$). Table 1 contains the optimal values of optimization problems which we get when replacing the chance constraint in (24) by its safe tractable approximations listed above. Note that the situation is so simple that the least conservative generator-based approximation – the CVaR one – becomes tractable. Indeed, it is easily seen that whatever be a generator $\gamma(\cdot)$, the maximum over $P \in \mathcal{P}$ of the expectation $\mathbf{E}_{\zeta\sim P}\{\gamma(-\tau + \sum_i w_i\zeta_i)\}$ is achieved when $P = P_*$ is the uniform distribution on the vertices of the unit box (i.e., when ζ_i take, independently of each other, values ±1 with probabilities 1/2). In other words, as far as our generator-based approximation scheme is concerned, we lose nothing when passing from $\mathcal{P}$ to the singleton P_*, thus getting rid of the ambiguity of the chance constraint in question. Now, with $\zeta \sim P_*$, the distribution of the random variable $\widehat{\zeta}^k = \sum_{i=1}^k i\zeta_i$ is easy to compute numerically; indeed, this variable takes only integral values form the segment $[-k(k+1)/2, k(k+1)/2]$, and the corresponding probabilities can be computed by a simple recurrence in k. As a result, we can compute both the upper ϵ-quantiles $\underline{\tau}_*(\epsilon)$ of $\widehat{\zeta}^d$ (they are lower bounds on $\tau_*(\epsilon)$) and the CVaR-upper bounds on these quantiles; these quantities are displayed in the table.

The data in the table show that all safe convex approximations, except, perhaps, for the Budgeted one, possess moderate conservatism which decreases as $1-\epsilon$ approaches 1. Pay attention to the closeness between the least conservative, within our approach, CVaR approximation and its "tractable counterpart" – the bridged Bernstein-CVaR approximation.

3.3. Beyond the scope of affinely perturbed chance constraints with independent perturbations

The safe *tractable* approximations of scalar linear chance constraints presented so far are restricted to the case when the random data ζ enters the body of the constraint affinely, and the entries ζ_i in ζ are independent. Now consider a more general case where the independence assumption is removed and, in addition, the uncertain data enters the body of the (still scalar and linear in the decision variables) chance constraint quadratically. A generic form of such a chance constraint is

$$\mathrm{Prob}_{\zeta\sim P}\{\mathrm{Tr}(WZ[\zeta]) \leqslant 0\} \geqslant 1-\epsilon \quad \forall P \in \mathcal{P}, \quad Z[\zeta] = \begin{bmatrix} 1 & \zeta^T \\ \zeta & \zeta\zeta^T \end{bmatrix}, \tag{25}$$

where $\zeta \in \mathbf{R}^d$, and the symmetric matrix W is affine in the decision variables; we lose nothing by assuming that W itself is the decision variable.

The approach we are about to present originates from [5,4], while the presentation itself is a slightly refined version of [2, Section 4.5]. We start with the description of $\mathcal{P}$. Specifically, we assume that our a priori information on the distribution P of the uncertain data can be summarized as follows:

- $\mathcal{P}$.1 We know that the marginals P_i of P belong to given families $\mathcal{P}_i$ of probability distributions on the axis;
- $\mathcal{P}$.2 The matrix $V_P = \mathbf{E}_{\zeta\sim P}\{Z[\zeta]\}$ of the first and the second moments of P is known to belong to a given convex closed subset $\mathcal{V}$ of the positive semidefinite cone;
- $\mathcal{P}$.3 P is supported on a set S given by a finite system of quadratic constraints:

$$S = \{\zeta : \mathrm{Tr}(A_\ell Z[\zeta]) \leqslant 0,\ 1 \leqslant \ell \leqslant L\}.$$

We believe that the above assumptions model well enough a priori information on uncertain data in typical applications of decision-making origin.

The approach, which we in the sequel refer to as *Lagrangian approximation*, is of the same spirit as the one developed in Section 3.1. Assume that given W, we have a systematic way to generate pairs $(\gamma(\cdot): \mathbf{R}^d \to \mathbf{R}, \theta > 0)$ such that (I) $\gamma(\cdot) \geqslant 0$ in S, (II) $\gamma(\cdot) \geqslant \theta$ in the part of S where $\mathrm{Tr}(WZ[\zeta]) \geqslant 0$, and (c) we have at our disposal a functional $\Gamma[\gamma]$ such that $\Gamma[\gamma] \geqslant \Gamma_*[\gamma] := \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\{\gamma(\zeta)\}$. Then the condition

$$\Gamma[\gamma] \leqslant \theta\epsilon$$

clearly is sufficient for the validity of (25), and we can further optimize this sufficient condition over the pairs γ, θ produced by our hypothetic mechanism. The question is, how to build the required mechanism, and here is an answer. Let us start with building $\Gamma[\cdot]$. Under our assumptions on $\mathcal{P}$, the most natural family of functions $\gamma(\cdot)$ for which one can bound from above $\Gamma_*[\gamma]$, is comprised of functions of the form

$$\gamma(\zeta) = \mathrm{Tr}(QZ[\zeta]) + \sum\nolimits_{i=1}^d \gamma_i(\zeta_i) \tag{26}$$

with $Q \in \mathbf{S}^{d+1}$. For such a γ, one can set

$$\Gamma[\gamma] = \sup_{V\in\mathcal{V}} \mathrm{Tr}(QV) + \sum\nolimits_{i=1}^d \sup_{P_i\in\mathcal{P}_i} \int \gamma_i(\zeta_i)dP_i(\zeta_i).$$

Further, the simplest way to ensure (I) is to use Lagrangian relaxation, specifically, to require from the function $\gamma(\cdot)$ given by (26) to be such that with properly chosen $\mu_\ell \geqslant 0$ one has

$$\mathrm{Tr}(QZ[\zeta]) + \sum\nolimits_{i=1}^d \gamma_i(\zeta_i) + \sum\nolimits_{\ell=1}^L \mu_\ell \mathrm{Tr}(A_\ell Z[\zeta]) \geqslant 0 \quad \forall \zeta \in \mathbf{R}^d.$$

In turn, the simplest way to ensure the latter relation is to impose on $\gamma_i(\zeta_i)$ the restrictions

$$\begin{aligned} &(a) \quad \gamma_i(\zeta_i) \geqslant p_i\zeta_i^2 + 2q_i\zeta_i + r_i \quad \forall \zeta_i \in \mathbf{R},\\ &(b) \quad \mathrm{Tr}(QZ[\zeta]) + \sum\nolimits_{i=1}^d [p_i\zeta_i^2 + 2q_i\zeta_i + r_i] + \sum\nolimits_{\ell=1}^L \mu_\ell \mathrm{Tr}(A_\ell Z[\zeta]) \geqslant 0\\ &\qquad \forall \zeta \in \mathbf{R}^d; \end{aligned} \tag{27}$$

note that (b) reduces to Linear Matrix Inequality (LMI)

$$Q + \begin{bmatrix} \sum_i r_i & q^T \\ q & \mathrm{Diag}\{p\} \end{bmatrix} + \sum\nolimits_{\ell=1}^L \mu_\ell A_\ell \succeq 0 \tag{28}$$

in variables $p = [p_1;\ldots;p_d]$, $q = [q_1;\ldots;q_d]$, $r = [r_1;\ldots;r_d]$, Q, $\{\mu_\ell \geqslant 0\}_{\ell=1}^L$. Similarly, a sufficient condition for (II) is the existence of $p', q', r' \in \mathbf{R}^d$ and nonnegative ν_ℓ such that

$$\begin{aligned} &(a) \quad \gamma_i(\zeta_i) \geqslant p'_i\zeta_i^2 + 2q'_i\zeta_i + r'_i \quad \forall \zeta_i \in \mathbf{R},\\ &(b) \quad \mathrm{Tr}(QZ[\zeta]) + \sum_{i=1}^d [p'_i\zeta_i^2 + 2q'_i\zeta_i + r'_i] + \sum_{\ell=1}^L \nu_\ell \mathrm{Tr}(A_\ell Z[\zeta]) - \mathrm{Tr}(WZ[\zeta])\\ &\qquad \geqslant \theta \quad \forall \zeta \in \mathbf{R}^d \end{aligned} \tag{29}$$

with (b) reducing to the LMI

$$Q + \begin{bmatrix} \sum_i r'_i - \theta & [q']^T \\ q' & \mathrm{Diag}\{p'\} \end{bmatrix} + \sum_{\ell=1}^L \nu_\ell A_\ell - W \succeq 0 \tag{30}$$

in variables W, p', q', $r' \in \mathbf{R}^d$, Q, $\{\nu_i \geqslant 0\}_{i=1}^L$ and $\theta > 0$. Finally, observe that under the restrictions (27.*a*), (29.*a*), the best – resulting in the smallest possible $\Gamma[\gamma]$ – choice of $\gamma_i(\cdot)$ is

$$\gamma_i(\zeta_i) = \max\left[p_i\zeta_i^2 + 2q_i\zeta_i + r_i,\ p'_i\zeta_i^2 + 2q'_i\zeta_i + r'_i\right].$$

Table 1
Lower bound $\underline{\tau}_*(\epsilon)$ and upper bounds on $\tau_*(\epsilon)$ given by various approximation schemes. Percents: excess of the bound over $\underline{\tau}_*(\epsilon)$.

Bounding scheme	ϵ								
	1.0e−1	5.0e−2	1.0e−2	5.0e−3	1.0e−3	1.0e−4	1.0e−5	1.0e−6	1.0e−7
$\underline{\tau}_*(\epsilon)$	1.286	1.650	2.330	2.578	3.088	3.706	4.239	4.712	5.140
CVaR	1.759 37%	2.067 25%	2.666 14%	2.891 12%	3.360 9%	3.940 6%	4.446 5%	4.899 4%	5.312 3%
Bridged	1.771 38%	2.078 26%	2.676 15%	2.901 13%	3.370 9%	3.950 7%	4.456 5%	4.911 4%	5.325 4%
Bernstein	2.146 67%	2.446 48%	3.027 30%	3.244 26%	3.698 20%	4.258 15%	4.747 12%	5.186 10%	5.586 9%
Ball-Box	2.152 67%	2.455 49%	3.044 31%	3.265 27%	3.728 21%	4.305 16%	4.813 14%	5.272 12%	5.694 11%
Ball	2.152 67%	2.455 49%	3.044 31%	3.265 27%	3.728 21%	4.305 16%	4.813 14%	5.272 12%	5.694 11%
Budgeted	3.475 170%	3.924 138%	4.768 105%	5.076 97%	5.703 85%	6.451 74%	7.081 67%	7.627 62%	8.108 58%

We have arrived at the following result:

Proposition 4. *Let* $(\mathcal{S})$ *be the system of constraints in variables* $W,\ p,\ q,\ r,\ p',\ q',\ r' \in \mathbf{R}^d,\ Q \in \mathbf{S}^{d+1},\ \{\nu_\ell \in \mathbf{R}\}_{\ell=1}^L,\ \theta \in \mathbf{R}$, $\{\mu_\ell \in \mathbf{R}\}_{\ell=1}^L$ *comprised of the LMIs (28), (30) augmented by the constraints*

$$\mu_\ell \geqslant 0\ \forall \ell, \quad \nu_\ell \geqslant 0\ \forall \ell, \quad \theta > 0 \tag{31}$$

and

$$\sup_{V\in\mathcal{V}} \mathrm{Tr}(QV) + \sum_{i=1}^d \sup_{P_i\in\mathcal{P}_i} \int \max\left[p_i\zeta_i^2 + 2q_i\zeta_i + r_i, p_i'\zeta_i^2 + 2q_i'\zeta_i + r_i'\right] dP_i \leqslant \epsilon\theta. \tag{32}$$

$(\mathcal{S})$ is a safe convex approximation of the chance constraint *(25)*, $\mathcal{P}$ *being given by* $\mathcal{P}.1$–$\mathcal{P}.3$. *This approximation is tractable, provided that the suprema in (32) are efficiently computable.*

Note that the strict inequality $\theta > 0$ in (31) can be expressed by the nonstrict LMI $\begin{bmatrix}\theta & 1\\ 1 & \lambda\end{bmatrix} \succeq 0$, where λ is an additional variable.

Illustration. Consider the situation as follows: there are $d = 15$ assets with yearly returns $r_i = 1 + \mu_i + \sigma_i\zeta_i$, where μ_i is the expected profit of ith return, σ_i is the return's variability, and ζ_i is random factor with zero mean supported on $[-1,1]$. The quantities μ_i, σ_i used in our illustration are shown on the left plot on Fig. 1. The goal is to distribute \$1 between the assets in order to maximize the value-at-1% risk (the lower 1%-quantile) of the yearly profit. This is the ambiguously chance constrained problem

$$\mathrm{Opt} = \max_{t,x}\left\{t : \mathrm{Prob}_{\zeta\sim P}\left\{\sum_{i=1}^{15}\mu_i x_i + \sum_{i=1}^{15}\zeta_i\sigma_i x_i \geqslant t\right\} \geqslant 0.99 \quad \forall P\in\mathcal{P},\ x \geqslant 0,\ \sum_{i=1}^{15} x_i = 1\right\}. \tag{33}$$

Consider three hypotheses A, B, C about $\mathcal{P}$. In all of them, ζ_i are zero mean and supported on $[-1,1]$, so that the domain information is given by the quadratic inequalities $\zeta_i^2 \leqslant 1,\ 1 \leqslant i \leqslant 15$; this is exactly what is stated by C. In addition, A says that ζ_i are independent, and B says that the covariance matrix of ζ is proportional to the unit matrix. Thus, the sets $\mathcal{V}$ associated with the hypotheses are, respectively, $\left\{V \in \mathbf{S}_+^{d+1} : V_{ii} \leqslant V_{00} = 1, V_{ij} = 0, i \neq j\right\}$, $\left\{V \in \mathbf{S}_+^{d+1} : 1 = V_{00} \geqslant V_{11} = V_{22} = \cdots = V_{dd}, V_{ij} = 0, i \neq j\right\}$, and $\left\{V \in \mathbf{S}_+^{d+1} : V_{ii} \leqslant V_{00} = 1, V_{0j} = 0, 1 \leqslant j \leqslant d\right\}$, where $\mathbf{S}_+^k$ is the cone of positive semidefinite symmetric $k \times k$ matrices. Solving the associated safe tractable approximations of the problem, specifically, the Bernstein approximation in the case of A, and the Lagrangian approximations in the cases of B, C, we arrive at the results displayed in Table 2 and on Fig. 1.

Note that in our illustration, the (identical to each other) single-asset portfolios yielded by the Lagrangian approximation under hypotheses B, C are *exactly optimal* under circumstances. Indeed, on a close inspection, there exists a distribution P_* compatible with hypothesis B (and therefore – with C as well) such that the probability of "crisis," where all ζ_i simultaneously are equal to -1, is $\geqslant 0.01$. It follows that under hypotheses B, C, the worst-case, over $P \in \mathcal{P}$, profit at 1% risk of *any* portfolio cannot be better than the profit of this portfolio in the case of crisis, and the latter quantity is maximized by the single-asset portfolio depicted on Fig. 1. Note that the Lagrangian approximation turns out to be "intelligent enough" to discover this phenomenon and to infer its consequences. A couple of other instructive observations is as follows:

- the diversified portfolio yielded by the Bernstein approximation in the case of crisis exhibits *negative* profit, meaning that under hypotheses B, C its worst-case profit at 1% risk is negative;
- assume that the yearly returns are observed on a year-by-year basis, and the year-by-year realizations of ζ are independent and identically distributed. It turns out that it takes over 60 years to distinguish, with reliability 0.99, between hypothesis A and the "bad" distribution P_* via the historical data.

To put these observations into proper perspective, note that it is extremely time-consuming to identify, to reasonable accuracy and with reasonable reliability, a multi-dimensional distribution directly from historical data, so that in applications one usually postulates certain parametric form of the distribution with a relatively small number of parameters to be estimated from the historical data. When $\dim\zeta$ is large, the requirement on the distribution to admit a low-dimensional parameterization usually results in postulating some kind of independence. While in some applications (e.g., in telecommunications) this independence in many cases can be justified via the "physics" of the uncertain data, in Finance and other decision-making applications postulating independence typically is an "act of faith" which is difficult to justify experimentally, and we believe a decision-maker should be well aware of the dangers related to these "acts of faith."

3.3.1. A modification

When building safe tractable approximation of (10), one, essentially, is interested in efficient bounding from above the quantity $p(w) = \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\left\{f(w_0 + \sum_{i=1}^d \zeta_i w_i)\right\}$ for a very specific function

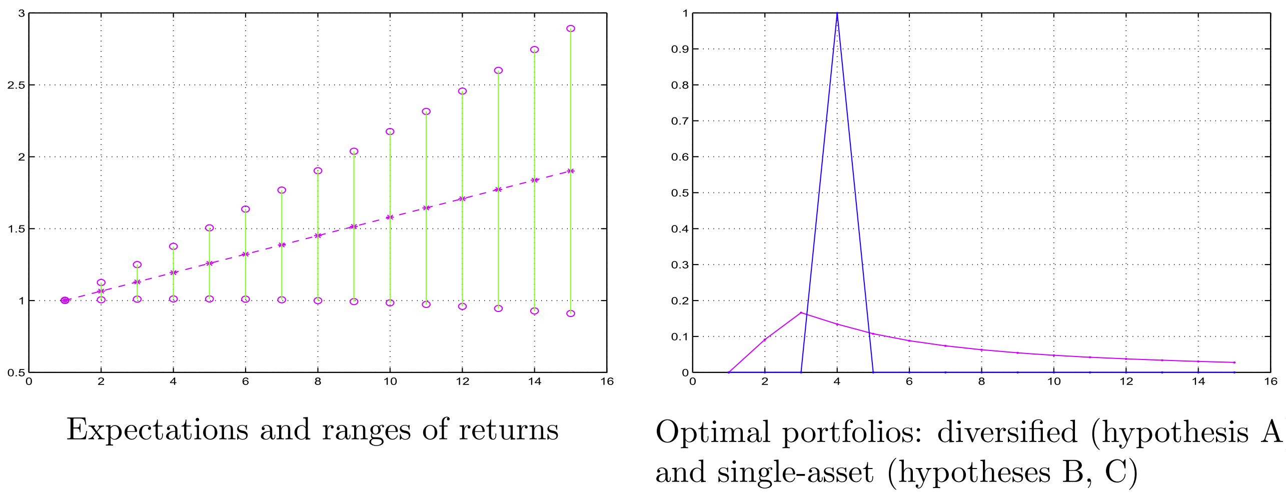

Fig. 1. Data and results for portfolio allocation.

Table 2
Optimal values in various approximations of (33).

Hypothesis	Approximation	Guaranteed profit-at-1%-risk
A	Bernstein	0.0552
B, C	Lagrangian	0.0101

$f(s)$, namely, equal to 0 when $s \leqslant 0$ and equal to 1 when $s > 0$. There are situations when we are interested in bounding similar quantity for other functions f, specifically, piecewise linear convex function $f(s) = \max_{1\leqslant j\leqslant J}[a_i + b_i s]$, see, e.g., [14]. Here again one can use Lagrange relaxation, which in fact is able to cope with a more general problem of bounding from above the quantity

$$\Psi[W] = \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\{f(W,\zeta)\}, \quad f(W,\zeta) = \max_{1\leqslant j\leqslant J} \mathrm{Tr}(W^j Z[\zeta]);$$

here the matrices $W^j \in \mathbf{S}^{d+1}$ are affine in the decision variables and $W = [W^1,\ldots,W^J]$. Specifically, with the assumptions $\mathcal{P}$.1–3 in force, observe that if, for a given W, a matrix $Q \in \mathbf{S}^{d+1}$ and vectors p^j, q^j, $r^j \in \mathbf{R}^d$, $1 \leqslant j \leqslant J$ are such that the relations

$$\mathrm{Tr}(QZ[\zeta]) + \sum_{i=1}^d \left[p_i^j\zeta_i^2 + 2q_i^j\zeta_i + r_i^j\right] \geqslant \mathrm{Tr}(W^j Z[\zeta]) \quad \forall \zeta \in S \qquad (\mathrm{I_j})$$

take place for $1 \leqslant j \leqslant J$, then the function

$$\gamma(\zeta) = \mathrm{Tr}(QZ[\zeta]) + \sum_{i=1}^d \max_{1\leqslant j\leqslant J} \left[p_i^j\zeta_i^2 + 2q_i^j\zeta_i + r_i^j\right] \qquad (34)$$

upper-bounds $f(W,\zeta)$ when $\zeta \in S$, and therefore the quantity

$$F(W,Q,p,q,r) = \sup_{V\in\mathcal{V}} \mathrm{Tr}(QV) + \sum_{i=1}^d \sup_{P_i\in\mathcal{P}_i} \times \int \max_{1\leqslant j\leqslant J} \left[p_i^j\zeta_i^2 + 2q_i^j\zeta_i + r_i^j\right] dP_i(\zeta_i) \qquad (35)$$

is an upper bound on $\Psi[W]$. Using Lagrange relaxation, a sufficient condition for the validity of (I_j), $1 \leqslant j \leqslant J$, is the existence of nonnegative $\mu_{j\ell}$ such that

$$Q + \begin{bmatrix} \sum_i r_i^j & [q^j]^T \\ q^j & \mathrm{Diag}\{p^j\} \end{bmatrix} - W^j + \sum_{\ell=1}^L \mu_{j\ell} A_\ell \succeq 0, \quad 1 \leqslant j \leqslant J. \qquad (36)$$

We have arrived at the result as follows:

Proposition 5. *Consider the system $\mathcal{S}$ of constraints in the variables $W = [W^1,\ldots,W^d]$, t and slack variables $Q \in \mathbf{S}^{d+1}$, $\{p^j, q^j, r^j \in \mathbf{R}^d : 1 \leqslant j \leqslant J\}$, $\{\mu_{ij} : 1 \leqslant i \leqslant d, 1 \leqslant j \leqslant J\}$ comprised of LMIs (36) augmented by the constraints*

$$\begin{aligned} &(a) \quad \mu_{ij} \geqslant 0, \quad \forall i,j, \\ &(b) \quad \max_{V\in\mathcal{V}} \mathrm{Tr}(QV) + \sum_{i=1}^d \sup_{P_i\in\mathcal{P}_i} \int \max_{1\leqslant j\leqslant J} \left[p_i^j\zeta_i^2 + 2q_i^j\zeta_i + r_i^j\right] dP_i(\zeta_i) \leqslant t. \end{aligned} \qquad (37)$$

The constraints in the system are convex; they are efficiently computable, provided that the suprema in (37.b) are efficiently computable, and whenever W, t can be extended to a feasible solution to $\mathcal{S}$, one has $\Psi[W] \leqslant t$. In particular, when the suprema in (37.b) are efficiently computable, the efficiently computable quantity

$$\mathrm{Opt}[W] = \min_{t,Q,p,q,r,\mu} \{t : W,t,Q,p,q,r,\mu \text{ satisfy } \mathcal{S}\} \qquad (38)$$

is a convex in W upper bound on $\Psi[W]$.

Illustration. Consider a special case of the above situation where all we know about ζ are the marginal distributions P_i of ζ_i with well defined first order moments; in this case, $\mathcal{P}_i = \{P_i\}$ are singletons, and we lose nothing when setting $\mathcal{V} = \mathbf{S}_+^{d+1}$, $S = \mathbf{R}^d$. Let a piecewise linear convex function on the axis:

$$f(s) = \max_{1\leqslant j\leqslant J} [a_j + b_j s]$$

be given, and let our goal be to bound from above the quantity

$$\Psi(w) = \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\{f(\zeta^w)\}, \quad \zeta^w = w_0 + \sum_{i=1}^d \zeta_i w_i.$$

This is a special case of the problem we have considered, corresponding to

$$W^j = \begin{bmatrix} a_j + b_j w_0 & \frac{1}{2}b_j w_1 & \ldots & \frac{1}{2}b_j w_d \\ \frac{1}{2}b_j w_1 & & & \\ \vdots & & & \\ \frac{1}{2}b_j w_d & & & \end{bmatrix}.$$

In this case, system $\mathcal{S}$ from Proposition 5, where we set $p_i^j = 0$, $Q = 0$, reads

$$\begin{aligned} &(a) \quad 2q_i^j = b_j w_i, 1 \leqslant i \leqslant d, q \leqslant j \leqslant J, \quad \sum_{i=1}^d r_i^j \geqslant a_j + b_j w_0, \\ &(b) \quad \sum_{i=1}^d \int \max_{1\leqslant j\leqslant J} \left[2q_i^j\zeta_i + r_i^j\right] dP_i(\zeta_i) \leqslant t \end{aligned}$$

so that the upper bound (38) on $\sup_{P\in\mathcal{P}}\mathbf{E}_{\zeta\sim P}\{f(\zeta^w)\}$ implies that

$$\mathrm{Opt}[w]=\min_{\{r_i^j\}}\left\{\sum_{i=1}^d\int\max_{1\leqslant j\leqslant J}\left[b_jw_i\zeta_i+r_i^j\right]dP_i(\zeta_i):\ \sum_i r_i^j=a_j+b_jw_0,1\leqslant j\leqslant J\right\} \tag{39}$$

is an upper bound on $\Psi(w)$. A surprising fact is that *in the situation in question* (i.e., when $\mathcal{P}$ is comprised of all probability distributions with given marginals $P_1,\ldots,P_d$), *the upper bound* Opt$[w]$ *on* $\Psi(w)$ *is equal to* $\Psi(w)$ [2, Poposition 4.5.4]. This result offers an alternative (and simpler) proof for the following remarkable fact established by Dhaene et al. [14]: *if f is a convex function on the axis and $\eta_1,\ldots,\eta_d$ are scalar random variables with distributions $P_1,\ldots,P_d$ possessing first order moments, then $\sup_{P\in\mathcal{P}}\mathbf{E}_{\eta\sim P}\{f(\eta_1+\ldots+\eta_d)\}$, $\mathcal{P}$ being the family of all distributions on $\mathbf{R}^d$ with marginals $P_1,\ldots,P_d$, is achieved when $\eta_1,\ldots,\eta_d$ are comonotone, that is, are deterministic monotone transformation of a single random variable uniformly distributed on* $[0,1]$.

4. Approximating chance constrained conic quadratic and linear matrix inequalities

Next we pass to approximating systems of chance constrained systems of conic quadratic and semidefinite inequalities, that is, chance constrained conic inequalities

$$\mathrm{Prob}_{\zeta\sim P}\{\zeta:A(x,\zeta)\in\mathbf{K}\}\geqslant 1-\epsilon\quad\forall P\in\mathcal{P}, \tag{40}$$

where $x\in\mathbf{R}^n$, $\zeta\in\mathbf{R}^d$, $\mathcal{P}$ is a family of probability distributions on $\mathbf{R}^d$, $A(x,\zeta)$ is vector-valued function affine in x for ζ fixed and affine in ζ for x fixed, and $\mathbf{K}$ is either a direct product of Lorenz cones

$$\mathbf{L}^{k_\ell}=\left\{y\in\mathbf{R}^{k_\ell}:y_1\geqslant\sqrt{\sum_{i=2}^{k_\ell}y_i^2}\right\}$$

over $\ell=1,\ldots,L$, or a positive semidefinite cone $\mathbf{S}^m_+$ – the set of all $m\times m$ positive semidefinite matrices "living" in the Euclidean space $\mathbf{S}^m$ of $m\times m$ symmetric matrices.

It is well known that a vector $y\in\mathbf{R}^k$ belongs to the Lorentz cone $\mathbf{L}^k$ if and only if the "arrow matrix" $\begin{bmatrix}y_1&y_2&\cdots&y_k\\ y_2&y_1&&\\ \vdots&&\ddots&\\ y_k&&&y_1\end{bmatrix}$ is positive semidefinite. It follows that the chance constrained system of conic quadratic inequalities (i.e., the case of $K=\mathbf{L}^{k_1}\times\ldots\times\mathbf{L}^{k_L}$) reduces to chance constrained LMI (i.e., to the case when $\mathbf{K}$ is a semidefinite cone). By this reason, we from now on focus solely on chance constrained LMIs. In the LMI case, the mapping $A(x,\zeta)$ can be written as $W_0+\sum_{i=1}^d\zeta_iW_i$, where $W_0,\ldots,W_d$ are matrices from certain $\mathbf{S}^m$ affinely depending on the decision variables x, and we lose nothing when assuming that these matrices are themselves our decision variables. Thus, we are interested in approximating a chance constrained LMI

$$\mathrm{Prob}_{\zeta\sim P}\left\{W_0+\sum_{i=1}^d\zeta_iW_i\succeq 0\right\}\geqslant 1-\epsilon\quad\forall P\in\mathcal{P} \tag{41}$$

in decision variables $W=[W_0,\ldots,W_d]\in[\mathbf{S}^m]^{d+1}$. In the sequel, we assume that $\mathcal{P}$ is comprised of all probability distributions on $\mathbf{R}^d$ such that

- **L.1** $\zeta_1,\ldots,\zeta_d$ are independent, and
- **L.2** either ζ_i have zero means and are supported on $[-1,1]$ ("bounded case"), or $\zeta_i\sim\mathcal{N}(0,\sigma_i^2)$ with $0\leqslant\sigma_i\leqslant 1$ ("Gaussian case").

Note that the "factor model" $\zeta=B\eta+b$ with independent zero mean and bounded almost surely entries in η clearly reduces to the bounded case, with properly scaled η in the role of ζ. Similarly, the case of $\zeta\sim\mathcal{N}(\mu,\Sigma)$ clearly reduces to the Gaussian case.

4.1. Main result

Guessing the approximation. Assume that $\epsilon<1/2$ and that the distribution P of ζ belongs to $\mathcal{P}$ and is symmetric w.r.t. the origin. In this case, feasibility of W for (41) implies that $W_0\succeq 0$. Assuming slightly more, namely, that $W_0\succ 0$, (41) is equivalent to

$$\mathrm{Prob}_{\zeta\sim P}\left\{I+\sum_{i=1}^d\zeta_iB_i\succeq 0\right\}\geqslant 1-\epsilon,\quad B_i=W_0^{-1/2}W_iW_0^{-1/2}, \tag{42}$$

whence also

$$\mathrm{Prob}_{\zeta\sim P}\left\{-I\preceq\sum_{i=1}^d\zeta_iB_i\preceq I\right\}\geqslant 1-2\epsilon\Longleftrightarrow\mathrm{Prob}_{\zeta\sim P}\left\{\left\|\sum_i\zeta_iB_i\right\|\leqslant 1\right\}\geqslant 1-2\epsilon, \tag{43}$$

where $\|\cdot\|$ is the usual matrix norm (maximal singular value). It follows that

$$\mathrm{Prob}_{\zeta\sim P}\left\{\|\left[\sum_{i=1}^d\zeta_iB_i\right]e\|_2^2\leqslant 1\right\}\geqslant 1-2\epsilon\quad\forall(e:\|e\|_2\leqslant 1).$$

It is natural to guess that when ϵ is small and ζ_i are independent zero mean "light tail" random variables, the latter relation should imply that

$$\begin{aligned}&\textstyle\sum_i\mathbf{E}_{\zeta\sim P}\{\zeta_i^2\}\|B_ie\|_2^2=\mathbf{E}_{\zeta\sim P}\{\|\sum_i\zeta_iB_ie\|_2^2\}\leqslant O(1),\quad\forall(e:\|e\|_2\leqslant 1),\\ \Longleftrightarrow\ &\textstyle\sum_i\mathbf{E}_{\zeta\sim P}\{\zeta_i^2\}B_i^2\preceq O(1)I_m\end{aligned} \tag{!}$$

(from now on, $O(1)$ stands for properly chosen positive absolute constants). On a close inspection, this guess indeed is true when $P=\mathcal{N}(0,I_d)$ or P is the uniform distribution on the vertices of the unit box, the "small enough" ϵ in both cases being $O(1)$. We conclude that *in the case of* **L.1–2**, *the condition*

$$\sum_{i=1}^dB_i^2\preceq O(1)I_m \tag{44}$$

with properly chosen O(1) is necessary for the validity of (43). Now, if B_i were deterministic reals rather than deterministic symmetric matrices, the condition like (44), but with somehow reduced right hand side, namely, the condition

$$\sum_{i=1}^dB_i^2\leqslant\Upsilon^{-2}(\epsilon)$$

with $\Upsilon(\epsilon)$ going slowly to $+\infty$ as $\epsilon\to+0$, would be *sufficient* for the validity of (42), provided **L.1–2** take place.[5] One could hope that similar fact holds true in the matrix case as well.

Main result. A simple corollary of nontrivial facts of functional analysis due to Lust-Picard, Pisier, Buhholz (see [30] and references therein) is that the hope we have expressed can indeed be justified.

Theorem 4.1 [2, Theorems 10.1.1, 10.1.2, Proposition B.5.2]. Let $0<\epsilon<1/4$, let the distribution P of ζ satisfy **L.1–2**, and let $B_1,\ldots,B_d$ be deterministic matrices from $\mathbf{S}^m$ such that

$$\begin{aligned}&\sum_{i=1}^dB_i^2\preceq\Upsilon^{-2}(m,\epsilon)I_m,\\ &\Upsilon(m,\epsilon)=4\sqrt{\ln(\max(m,3))}+\begin{cases}4\sqrt{\ln\left(\frac{4}{3\epsilon}\right)}, & \text{bounded case},\\ \mathrm{ErfInv}(\epsilon)-\mathrm{ErfInv}(1/4), & \text{Gaussian case}.\end{cases}\end{aligned} \tag{45}$$

Then

[5] By Hoeffding Inequality, in the bounded case it suffices to set $\Upsilon(\epsilon)=\sqrt{2\ln(1/\epsilon)}$, while in the Gaussian case one can take $\Upsilon(\epsilon)=\mathrm{ErfInv}(\epsilon)\leqslant\sqrt{2\ln(1/\epsilon)}$.

$$\mathrm{Prob}_{\zeta\sim P}\left\{I+\sum\nolimits_{i=1}^{d}\zeta_i B_i \succeq 0\right\} \geqslant 1-\epsilon. \quad (46)$$

In particular, the LMI

$$\begin{bmatrix} W_0 & \Upsilon(m,\epsilon)W_1 & \dots & \Upsilon(m,\epsilon)W_d \\ \Upsilon(m,\epsilon)W_1 & W_0 & & \\ \vdots & & \ddots & \\ \Upsilon(m,\epsilon)W_d & & & W_0 \end{bmatrix} \succeq 0 \quad (47)$$

in variables $W = [W_0, W_1, \ldots, W_d]$ is a safe tractable approximation of chance constrained LMI (41).

Note that the "In particular" part of Theorem stems from the fact that with $W_0 \succ 0$, and $B_i = W_0^{-1/2} W_i W_0^{-1/2}$, matrix inequality (45) is equivalent to the LMI (47). The latter relation, being sufficient for the validity of (41) in the case of $W_0 \succ 0$, by continuity argument remains so when $W_0 \succeq 0$.

5. Scenario approximation revisited: Gaussian majorization

In this section, we, following [21] and [2, Section 10.3], show that under favorable circumstances one can overcome, to some extent, the drawbacks of the scenario approximation as presented in Section 2, that is, its impracticality in the case of "really small" ϵ and difficulties with the ambiguous case.

5.1. Gaussian majorization

Preliminaries: convex dominance. Let $\mathcal{R}_m$ be the family of all probability distributions on $\mathbf{R}^m$ with zero mean. With slight abuse of terminology, we say that a random vector ξ belongs to $\mathcal{R}_m$ if the distribution of ξ is in $\mathcal{R}_m$. Given random vectors ξ, η belonging to $\mathcal{R}_m$, the respective distributions being P, Q, we say that (the distribution P of) ξ is *convexly dominated* by (the distribution Q of) η, if $\mathbf{E}\{f(\xi)\} \leqslant \mathbf{E}\{f(\eta)\}$ for every *convex* function $f : \mathbf{R}^m \to \mathbf{R}$ with linear growth: $|f(x)| \leqslant a_f + b_f\|x\|_2$ for all x. We write down this fact as $\xi \preceq_c \eta (\iff \eta \succeq_c \xi)$, or as $P \preceq_c Q (\iff Q \succeq_c P)$. Convex dominance possesses a number of useful properties, in particular, as follows (for justification, see, e.g., [21,2]):

- $\succeq_c$ is a partial order on $\mathcal{R}_m$.
- If $P_1, \ldots, P_k, Q_1, \ldots, Q_k \in \mathcal{R}_m$, and $P_i \preceq_c Q_i$ for every i, then $\sum_i \lambda_i P_i \preceq_c \sum_i \lambda_i Q_i$ for all nonnegative λ_i with unit sum.
- If $\xi \in \mathcal{R}_m$ and $t \geqslant 1$ is deterministic, then $t\xi \succeq_c \xi$.
- Let $P_1, Q_1 \in \mathcal{R}_r$, $P_2, Q_2 \in \mathcal{R}_s$ be such that $P_i \preceq_c Q_i$, $i = 1, 2$. Then $P_1 \times P_2 \preceq_c Q_1 \times Q_2$. In particular, if $2k$ random vectors ξ_i, $\eta_i \in \mathcal{R}_{m_i}$, $1 \leqslant i \leqslant k$, are independent and $\xi_i \preceq_c \eta_i$ for every i, then $[\xi_1; \ldots; \xi_k] \preceq_c [\eta_1; \ldots; \eta_k]$.
- If $\xi_1, \ldots, \xi_k, \eta_1, \ldots, \eta_k \in \mathcal{R}_m$ are independent random variables, $\xi_i \preceq_c \eta_i$ for every i, and $S_i \in \mathbf{R}^{m' \times m}$ are deterministic matrices, then $\sum_i S_i \xi_i \preceq_c \sum_i S_i \eta_i$.
- Let $\xi \in \mathcal{R}_1$ be supported on $[-1,1]$ and $\eta \sim \mathcal{N}(0, \pi/2)$. Then $\eta \succeq_c \xi$.
- If ξ has unimodal w.r.t. 0 distribution and is supported on $[-1,1]$ and $\eta \sim \mathcal{N}(0, 2/\pi)$, then $\eta \succeq_c \xi$.
- Assume that $\xi \in \mathcal{R}_m$ is supported in the unit cube $\{u : \|u\|_\infty \leqslant 1\}$ and is "absolutely symmetrically distributed," meaning that if J is a diagonal matrix with diagonal entries ±1, then $J\xi$ has the same distribution as ξ. Let also $\eta \sim \mathcal{N}(0, (\pi/2)I_m)$. Then $\xi \preceq_c \eta$.
- Let ξ, $\eta \in \mathcal{R}_m$, $\xi \sim \mathcal{N}(0, \Sigma)$, $\eta \sim \mathcal{N}(0, \Theta)$ with $\Sigma \preceq \Theta$. Then $\xi \preceq_c \eta$.

Gaussian Majorization. Our working horse in the rest of this section is the following concentration result readily given by [21,2, Theorem 10.3.3] (these results, in turn, are simple consequences of the results of Borell [6]):

Theorem 5.1. *Let $\mathbf{Q} \subset \mathbf{R}^m$ be a closed convex set, $w_0, w_1, \ldots, w_d$ be deterministic vectors from $\mathbf{R}^m$, and let ζ, $\eta \in \mathcal{R}_d$ be such that $\zeta \preceq_c \eta$ and $\eta \sim \mathcal{N}(0, \Sigma)$. When $\epsilon \leqslant 0.01$, the relation*

$$\mathrm{Prob}\left\{w_0 + \gamma(\epsilon)\sum\nolimits_{i=1}^{d}\eta_i w_i \in \mathbf{Q}\right\} \geqslant \frac{3}{4}, \quad \gamma(\epsilon) = \frac{9}{5}\mathrm{ErfInv}(\epsilon), \quad (48)$$

implies that

$$\mathrm{Prob}\left\{w_0 + \sum\nolimits_{i=1}^{d}\zeta_i w_i \in \mathbf{Q}\right\} \geqslant 1-\epsilon. \quad (49)$$

5.2. Applications

Chance constrained LMI beyond the scope of assumptions L.1–2. One application of Gaussian Majorization is as follows. Consider chance constrained LMI (41) and, instead of assumptions **L.1–2**, assume that all distributions from $\mathcal{P}$ are convexly dominated by a known Gaussian distribution. This is a weaker assumption which allows, e.g., for certain dependence between the entries in ζ. Now, with appropriate linear scaling of ζ, we can assume the dominating Gaussian distribution to be the standard one, $\mathcal{N}(0, I_d)$. By Theorem 5.1, every feasible solution W to the chance constrained LMI

$$\mathrm{Prob}_{\eta\sim\mathcal{N}(0,I_d)}\left\{W_0 + \gamma(\epsilon)\sum\nolimits_{i=1}^{d}\eta_i W_i \succeq 0\right\} \geqslant 3/4 \quad (50)$$

is a feasible solution to the chance constraint of interest (41). It follows that a safe tractable approximation of (50) given by Theorem 4.1 (this theorem is applicable, since the uncertain data in (50) obey the standard Gaussian distribution), or whatever else safe tractable approximation of (50), is a safe tractable approximation of the chance constraint of interest (41) as well.

Analysis problem. Consider the situation as follows: we are given a chance constraint

$$p_P(x) := \mathrm{Prob}\{F(x,\zeta) \in \mathbf{Q}\} \geqslant 1-\epsilon \quad \forall P \in \mathcal{P}; \quad (51)$$

where $\mathbf{Q} \subset \mathbf{R}^m$ is a closed convex set, along with a candidate solution $\bar{x}$, and want to certify that $\bar{x}$ is feasible for (51). How could we handle this task when the probabilities in question are difficult to compute? When $\mathcal{P} = \{P\}$ is a singleton, the standard way to fulfil our task is to choose a confidence level $1-\beta$ and to use Monte Carlo simulation in order to get a (depending on the sample and thus random) estimate $\widehat{p}$ of $p_P(\bar{x})$ such that "up to probability $\leqslant 1-\beta$ of bad sampling" it holds $\widehat{p} - p_P(\bar{x}) \leqslant \epsilon'$ for some chosen in advance $\epsilon' < \epsilon$, say, $\epsilon' = \epsilon/2$. Now, if $\widehat{p} \geqslant 1-(\epsilon-\epsilon')$, we, with confidence $1-\beta$, can conclude that $\bar{x}$ satisfies (51). This approach, however, requires samples of the length of order of ϵ^{-1}, and thus becomes impractical when ϵ is "really small;" besides this, the approach does not work in the ambiguous case. The question is, what to do when the outlined "crude" Monte Carlo approach is either impractical, or inapplicable. In the case of "well structured constraints," a (partial) answer is as follows. Assume that $F(x,\zeta)$ is affine in ζ:

$$F(x,\zeta) = w_0(x) + \sum\nolimits_{i=1}^{d}\zeta_i w_i(x)$$

and that all distributions from $\mathcal{P}$ have common expectation, which we w.l.o.g. assume to be 0. Further, assume that we can point out a Gaussian distribution $\mathcal{N}(0, \Sigma)$ on $\mathbf{R}^d$ which convexly dominates all distributions $P \in \mathcal{P}$.

Whatever restrictive, these assumptions indeed are satisfied in many applications. For example, assume that F is affine in ζ and all distributions from $\mathcal{P}$ are either (a) product type zero mean ones supported on the unit cube, or (b) supported on the unit cube and absolutely symmetric, or (c) zero mean Gaussian distributions with uniformly bounded covariance matrices. Under these assumptions all distributions from $\mathcal{P}$ indeed are convexly dominated by an appropriate zero mean Gaussian distribution (which can be found using relevant items from the above list).

Under the above assumptions, we can use crude Monte Carlo simulation to certify, with a desired confidence $1-\beta$, the validity of the condition

$$\mathrm{Prob}_{\eta\sim\mathcal{N}(0,\Sigma)}\left\{w_0(\bar{x})+\gamma(\epsilon)\sum_{i=1}^d\eta_i w_i(\bar{x})\right\}\geqslant\frac{3}{4},\tag{52}$$

with $\gamma(\epsilon)$ given by (48). All we need here is to sample from the distribution $\mathcal{N}(0,\Sigma)$, with the sample size *depending solely on* β (and "small" – $O(1)\ \ln(\beta^{-1})$). When crude Monte Carlo simulation says that (52) does take place, we can use Theorem 5.1 to conclude, with confidence $\geqslant 1-\beta$, that $\bar{x}$ is feasible for (51). Note that the outlined approach is nothing but a kind of importance sampling, the difference being that in the usual importance sampling, one wants to *estimate* probability of certain event by sampling from a reference distribution, while we use the same technique to *bound from above* this probability. In applications we are aiming at, the former goal, in contrast to the latter one, would be completely unrealistic. For example, the distributions of two zero mean Gaussian vectors ξ, $\eta=9\xi$ in $\mathbf{R}^{100}$ are "nearly mutually singular" (specifically, admit non-overlapping $(1-\delta)$-supports with δ as small as 1.e−43), so that it seems completely unrealistic to estimate $\mathrm{Prob}\{\xi\notin Q\}$ by sampling η and taking the empirical expectation of the corresponding importance sampling ratio. At the same time, taking into account that $\frac{9}{5}\mathrm{ErfInv}(10^{-6})<9$, Theorem 5.1 says that if $\mathrm{Prob}\{\eta\notin Q\}\leqslant 1/4$ (which we can certify reliably by a short-sample simulation) and Q is closed and convex, then $\mathrm{Prob}\{\xi\notin Q\}\leqslant 10^{-6}$, although we have no idea whether the latter probability is 10^{-6} or 10^{-16}.

The outlined approach has various applications in chance constrained optimization; for example, one can use it to certify "true feasibility" of a candidate solution to a chance constrained problem when this candidate solution is yielded by an *un*safe approximation of this problem.

"Short sample" scenario approximations. Consider affinely perturbed chance constrained problem

$$\min_{x\in X}\left\{f(x):\mathrm{Prob}_{\zeta\sim P}\left\{w_0(x)+\sum_{i=1}^d\zeta_i w_i(x)\in\mathbf{Q}\right\}\geqslant 1-\epsilon\quad\forall P\in\mathcal{P}\right\},\tag{53}$$

where $X\subset\mathbf{R}^n$ is closed and convex, $f(x):X\to\mathbf{R}$ is convex and continuous function with bounded level sets $\{x\in X:f(x)\leqslant a\}$, $w_0(x),\ldots,w_d(x)$ are affine in x vector-valued functions taking values in certain $\mathbf{R}^m$, $\mathbf{Q}$ is a closed and convex subset of $\mathbf{R}^m$, and $\mathcal{P}$, as always, is certain family of probability distributions on $\mathbf{R}^d$. Assume that

D: All distributions from $\mathcal{P}$ are convexly dominated by a given Gaussian distribution $\mathcal{N}(0,\Sigma)$.

Let us associate with (53) its scenario approximation

$$\min_{x\in X}\left\{f(x):w_0(x)+\gamma\sum_{i=1}^d\eta_i^t w_i(x)\in\mathbf{Q},1\leqslant t\leqslant N\right\},\qquad (P[\overrightarrow{\eta_N}])$$
$$\gamma=\gamma(\epsilon)=\tfrac{9}{5}\mathrm{ErfInv}(\epsilon),$$

where $\overrightarrow{\eta_N}=(\eta^1,\ldots,\eta^N)$, and $\eta^1,\ldots,\eta^N$ are, independently of each other, sampled from the distribution $\mathcal{N}(0,\Sigma)$. The difference with scenario approximations from Section 2 is that now we are sampling from the distribution $\mathcal{N}(0,\Sigma)$ rather than from the true distribution of ζ, and, in addition, "amplify" the simulated data perturbations by factor γ. Note that since f has bounded level sets, $(P[\overrightarrow{\eta_N}])$, depending on $\overrightarrow{\eta_N}$, is either infeasible, or solvable. We have the following analogy of the result of Calafiore and Campi (Theorem 2.1):

Theorem 5.2. *In addition to assumption* **D**, *assume that whenever scenario approximation* $(P[\overrightarrow{\eta_N}])$ *is feasible (and thus solvable), the optimal solution to the approximation is unique (e.g., f is strictly convex). Let a confidence level* $1-\beta$, $\beta\in(0,0.1)$, be given, and let

$$N\geqslant\mathrm{Ceil}(5n\ln(n)+4\ln(\beta^{-1}))\tag{54}$$

(note that the right hand side is independent of ϵ*). Then the probability of the event*

$$\mathcal{H}=\{\overrightarrow{\eta_N}:\text{the optimal solution to } (P[\overrightarrow{\eta_N}])\text{ is well defined and is infeasible for }(53)\}$$

does not exceed β.

Proof 1. The proof follows, with minor modifications, the original proof of Theorem 2.1 due to Calafiore and Campi. Let $\mathcal{I}$ be the family of all subsets of $\{1,\ldots,N\}$ of cardinality $n+1$. For $I=\{t_1<t_2<\ldots<t_{n+1}\}\in\mathcal{I}$, let $\overrightarrow{\eta_{N,I}}=(\eta^{t_1},\eta^{t_2},\ldots,\eta^{t_{n+1}})$, and let $f_*(\overrightarrow{\eta_{N,I}})\in\mathbf{R}\cup\{+\infty\}$ be the optimal value of the scenario problem $(P[\overrightarrow{\eta_{N,I}}])$ given by the scenarios $\eta^{t_1},\ldots,\eta^{t_{n+1}}$. When the latter problem is feasible, let $x_*(\overrightarrow{\eta_{N,I}})$ be the (unique, by assumption) optimal solution to this problem; when $P[\overrightarrow{\eta_{N,I}}]$ is infeasible, we set $x_*(\overrightarrow{\eta_{N,I}})$ to a fictitious value $*\notin\mathbf{R}^n$. Similarly, let $f_*(\overrightarrow{\eta_N})$ be the optimal value of $(P[\overrightarrow{\eta_N}])$, and $x_*(\overrightarrow{\eta_N})$ be the optimal solution to this problem, if the problem is feasible, and be $*$ otherwise.

Let us fix $\overrightarrow{\eta_N}$. We claim that $x_*:=x_*(\overrightarrow{\eta_N})$ is either $*$, or one of the points $x_*^I:=x_*(\overrightarrow{\eta_{N,I}}),\ I\in\mathcal{I}$.

Indeed, there is nothing to verify when $x_*=*$. Now assume that $x_*\neq *$. Then x_* is feasible for every problem $(P[\overrightarrow{\eta_{N,I}}]),\ I\in\mathcal{I}$, so that for every $I\in\mathcal{I}$ one has $x_*^I\neq *$ and $f(x_*^I)\leqslant f(x_*)$. Taking into account that the optimal solution to a scenario problem is unique and that x_* is feasible for every problem $(P[\overrightarrow{\eta_{N,I}}])$, all we need in order to prove that x_* is one of the points x_*^I is to verify that at least one of the inequalities $f(x_*^I)\leqslant f(x_*),\ I\in\mathcal{I}$, is an equality. Assume that this is not the case, so that there exists $\delta>0$ such that $f(x_*)>f(x_*^I)+\delta$ for all $I\in\mathcal{I}$, and let us lead this assumption to a contradiction. To this end, let

$$X^t=\{x\in X:f(x)\leqslant f(x_*)-\delta,w_0(x)+\gamma\sum_{i=1}^d\eta_i^t w_i(x)\in\mathbf{Q}\},\quad 1\leqslant t\leqslant N.$$

By construction, $X^1,\ldots,X^N$ are closed convex sets which in fact are bounded (since f has bounded level sets). In addition, every $n+1$ of the sets X^t, $t=1,\ldots,N$, have a point in common: when $1\leqslant t_1<t_2<\cdots<t_{n+1}\leqslant N$, a common point of $X^{t_1},\ldots,X^{t_{n+1}}$ is nothing but the point $x_*^{\{t_1,\ldots,t_{n+1}\}}$. By Helley Theorem, it follows that all the sets $X^1,\ldots,X^N$ have a point in common, let it be $\bar{x}$. Recalling the definition of X^t, $\bar{x}$ is a feasible solution to $(P[\overrightarrow{\eta_N}])$ and $f(\bar{x})\leqslant f(x_*)-\delta<f(x_*)$, which is impossible; we have arrived at a desired contradiction.

Now we can complete the proof. Since $x_*(\overrightarrow{\eta_N})$ is either $*$, or one of the points $x_*(\overrightarrow{\eta_{N,I}}),\ I\in\mathcal{I}$, all we need in order to prove that $\mathrm{Prob}\{\mathcal{H}\}\leqslant\beta$ is that

$$p:=\mathrm{Prob}\{\overrightarrow{\eta_N}:\exists I\in\mathcal{I}:x_*(\overrightarrow{\eta_{N,I}})\text{ is infeasible for }(53)\text{ and is feasible for }(P[\overrightarrow{\eta_N}])\}\leqslant\beta.\tag{55}$$

Given $I=\{t_1<t_2<\ldots<t_{n+1}\}\in\mathcal{I}$, let us set

$$p_I:=\mathrm{Prob}\{\mathcal{H}_I\},$$
$$\mathcal{H}_I=\{\overrightarrow{\eta_N}:x_*(\overrightarrow{\eta_{N,I}})\text{ is infeasible for }(53)\text{ and is feasible for }(P[\overrightarrow{\eta_N}])\};$$

note that $\mathcal{H}\subset\bigcup_{I\in\mathcal{I}}\mathcal{H}_I$. If $\overrightarrow{\eta_N}\in\mathcal{H}_I$, then $x_*^I:=x_*(\overrightarrow{\eta_{N,I}})\neq *$ (since otherwise x_*^I would be infeasible for $(P[\overrightarrow{\eta_N}])$) and

$$\pi:=\mathrm{Prob}_{\eta\sim\mathcal{N}(0,\Sigma)}\left\{w_0(x_*^I)+\gamma\sum_{i=1}^d\eta_i w_i(x_*^I)\in\mathbf{Q}\right\}<\frac{3}{4},\tag{56}$$

since otherwise x_*^I would be feasible for (53) by Theorem 5.1. On the other hand, when $\overrightarrow{\eta_N}\in\mathcal{H}_I$, x_*^I is feasible for $(P[\overrightarrow{\eta_N}])$, so that the event $\mathcal{E}_I$ defined as "$w_0(x_*^I)+\gamma\sum_{i=1}^d\eta_i^t w_i(x_*^I)\in\mathbf{Q}$ for all

$t \in J := \{1, \ldots, N\} \setminus I$" takes place. Now, x_*^I is uniquely defined by $\eta^{t_1}, \ldots, \eta^{t_{n+1}}$, and the conditional, given $\eta^{t_1}, \ldots, \eta^{t_{n+1}}$ such that x_*^I satisfies (56), probability of $\mathcal{E}_I$ is $<(3/4)^{N-(n+1)}$ (since $\eta^1, \ldots, \eta^N$ are independent). We conclude that $p_I \leqslant (3/4)^{N-n-1}$, whence, by the union bound, $p \leqslant \binom{N}{n+1}(3/4)^{N-n-1}$. Invoking (54), we conclude that $p \leqslant \beta$. □

Note that the "safe" sample size as given by the right hand side of (54) is independent of ϵ (and nearly linear in n and $\ln(\beta^{-1})$), which is a huge difference as compared to (8).

5.3. Scenario approximation of a chance constrained LMI

Theorems 2.1 and 5.2 provide sufficient conditions for "true feasibility" (i.e., feasibility in the chance constrained problem of interest), with probability $\geqslant 1-\beta$, of the optimal solution to the scenario approximation; these theorems, however, say nothing on whether the entire feasible set of the approximation is, with probability close to 1, feasible for the chance constrained problem of interest. It turns out that in the case of a chance constrained LMI, the latter can be ensured as well. Specifically, consider the chance constrained LMI (41) and assume that

D′: All distributions from $\mathcal{P}$ are convexly dominated by the standard Gaussian distribution $\mathcal{N}(0, I_d)$.

Note that whenever **D** holds true, appropriate linear scaling of ζ brings the situation into the scope of **D**′. The result to follow is a slightly refined version of the "Gaussian case" of [21, Theorem 11]; it can be proved in a fashion completely similar to the one of the prototype theorem, see [21]:

Theorem 5.3. *Given chance constrained LMI (41) with $\epsilon \leqslant 0.01$, $W_i \in \mathbf{S}^m$ and $\mathcal{P}$ satisfying* **D**′, consider the scenario approximation

$$W_0 + \gamma(\epsilon)\sum_{i=1}^d \eta_i^t W_i \succeq 0, 1 \leqslant t \leqslant N, \quad (C[\overrightarrow{\eta_N}])$$

of (41), where $\gamma(\epsilon)$ is given by (48), $\overrightarrow{\eta_N} = (\eta^1, \ldots, \eta^N)$, and $\eta^1, \ldots, \eta^N$ are, independently of each other, sampled from the distribution $\mathcal{N}(0, I_d)$, and let

$$\mathcal{W}_*[\overrightarrow{\eta_N}] = \{W = [W_0, \ldots, W_d] : W \text{ is feasible for } (C[\overrightarrow{\eta_N}]) \& \ W_0 \succeq 0\}.$$

Assume, further, that a confidence level $1-\beta$, $0 < \beta \leqslant 0.1$, is given, and let

$$N \geqslant C_1(m^2 d \ln(C_2 d \ln(\epsilon^{-1})) + \ln(\beta^{-1})) \tag{57}$$

with properly chosen absolute constants C_1, C_2. Then

$$\text{Prob}\{\overrightarrow{\eta_N} : \mathcal{W}_*[\overrightarrow{\eta_N}] \text{ is not contained in the feasible set of } (41)\} \leqslant \beta.$$

Acknowledgement

Research was supported by NSF Grants DMI-0619977, DMS-0914785 and ONR Grant N000140811104.

References

[1] A. Ben-Tal, A. Nemirovski, Robust solutions of uncertain linear programs, OR Letters 25 (1999) 1–13.
[2] A. Ben-Tal, L. El Ghaoui, A. Nemirovski, Robust Optimization, Princeton University Press, 2009.
[3] D. Bertsimas, M. Sim, Robust discrete optimization and network flows, Mathematical Programming Series B 98 (2003) 49–71.
[4] D. Bertsimas, I. Popescu, J. Sethuraman, Moment problems and semidefinite programming, in: H. Wolkovitz, R. Saigal (Eds.), Handbook on Semidefinite Programming, Kluwer Academic Publishers, 2000, pp. 469–509.
[5] D. Bertsimas, I. Popescu, Optimal inequalities in probability theory: A convex optimization approach, SIAM Journal on Optimization 15 (3) (2005) 780–804.
[6] C. Borell, The Brunn–Minkowski inequality in Gauss space, Inventiones Mathematicae 30 (2) (1975) 207–216.
[7] M.C. Campi, G. Calafiore, Decision making in an uncertain environment: the scenario based optimization approach, in: J. Andrysek, M. Karny, J. Kracik (Eds.), Multiple Participant Decision Making, Advanced Knowledge International, 2004.
[8] G. Calafiore, M.C. Campi, Uncertain convex programs: Randomized solutions and confidence levels, Mathematical Programming 102 (2005) 25–46.
[9] G. Calafiore, M.C. Campi, Sampled convex programs and probabilistically robust design. Scenario approximations of chance constraints, in: G. Calafiore, F. Dabbene (Eds.), Probabilistic and Randomized Methods for Design Under Uncertainty, Springer, London, 2006.
[10] M.C. Campi, S. Garatti, The exact feasibility of randomized solutions of robust convex programs, SIAM Journal on Optimization 19 (2008) 1211–1230.
[11] A. Charnes, W.W. Cooper, G.H. Symonds, Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil, Management Science 4 (1958) 235–263.
[12] D.P. de Farias, B. Van Roy, On constraint sampling in the linear programming approach to approximate dynamic programming, Mathematics of Operations Research 29 (2004) 462–478.
[13] D. Dentcheva, A. Prékopa, A. Ruszczynski, Concavity and efficient points of discrete distributions in probabilistic programming, Mathematical Programming 89 (2000) 55–77.
[14] J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, D. Vyncke, The concept of comonotonicity in actuarial science and finance: Theory, Insurance: Mathematics and Economics 31 (2002) 3–33.
[15] J. Dupačová, The minimax approach to stochastic programming and an illustrative application, Stochastics 20 (1987) 73–88.
[16] W.K. Klein Haneveld, Duality in stochastic linear and dynamic programming, Lecture Notes in Economics and Mathematical systems, vol. 274, Springer, 1986.
[17] G. Iyengar, E. Erdogan, Ambiguous chance constrained problems and robust optimization, Mathematical Programming 107 (2006) 17–31.
[18] L.G. Khachiyan, The problem of calculating the volume of a polyhedron is enumerably hard, Russian Mathematical Surveys 44 (1989) 199–200.
[19] C.M. Lagoa, X. Li, M. Sznaier, Probabilitically constrained linear programs and risk-adjusted controller design, SIAM Journal on Optimization 15 (2005) 938–951.
[20] L.B. Miller, H. Wagner, Chance-constrained programming with joint constraints, Operations Research 13 (1965) 930–945.
[21] A. Nemirovski, A. Shapiro, Scenario approximations of chance constraints, in: G. Calafiore, F. Dabbene (Eds.), Probabilistic and Randomized Methods for Design Under Uncertainty, Springer, London, 2006.
[22] A. Nemirovski, A. Shapiro, Convex approximations of chance constrained programs, SIAM Journal on Optimization 17 (2006) 969–996.
[23] J. Pinter, Deterministic approximations of probability inequalities, ZOR Methods and Models of Operations Research, Series Theory 33 (1989) 219–239.
[24] A. Prékopa, On probabilistic constrained programming, in: Proceedings of the Princeton Symposium on Mathematical Programming, Princeton University Press, Princeton, 1970, pp. 113–138.
[25] A. Prékopa, Stochastic Programming, Kluwer, Dordrecht, 1995.
[26] A. Prékopa, B. Vizvari, T. Badics, Programming under probabilistic constraint with discrete random variables, in: L. Grandinetti et al. (Eds.), New Trends in Mathematical Programming, Kluwer, Dordrecht, 1997, pp. 235–257.
[27] A. Prékopa, Probabilistic programming, in: A. Ruszczyński, A. Shapiro (Eds.), Stochastic Programming, Handbook in OR & MS, vol. 10, North-Holland, Publishing Company, Amsterdam, 2003.
[28] R.T. Rockafellar, S.P. Uryasev, Optimization of conditional value-at-risk, The Journal of Risk 2 (2000) 21–41.
[29] R.T. Rockafellar, S.P. Uryasev, M. Zabarankin, Generalized deviations in risk analysis, Finance and Stochastics 10 (2006) 51–74.
[30] J.A. Tropp, The random paving property for uniformly bounded matrices, Studia Mathematica 185 (2008) 67–82.