Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Solving joint chance constrained problems using regularization and Benders' decomposition

Lukáš Adam[1,2], Martin Branda[2,3], Holger Heitsch[4], René Henrion[4]

submitted: January 19, 2018

[1] Southern University of Science
and Technology
1088 Xueyuan Ave
Shenzhen, Guangdong, 518055
China

[2] ÚTIA, Czech Academy of Sciences
Pod Vodárenskou věží 4
18208 Prague
Czech Republic
E-Mail: adam@utia.cas.cz

[3] Faculty of Mathematics and Physics
Charles University
Sokolovská 83
186 75 Prague
Czech Republic
E-Mail: branda@karlin.mff.cuni.cz

[4] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: holger.heitsch@wias-berlin.de
rene.henrion@wias-berlin.de

Solving joint chance constrained problems using regularization and Benders' decomposition

Lukáš Adam, Martin Branda, Holger Heitsch, René Henrion

Abstract

In this paper we investigate stochastic programms with joint chance constraints. We consider discrete scenario set and reformulate the problem by adding auxiliary variables. Since the resulting problem has a difficult feasible set, we regularize it. To decrease the dependence on the scenario number, we propose a numerical method by iteratively solving a master problem while adding Benders cuts. We find the solution of the slave problem (generating the Benders cuts) in a closed form and propose a heuristic method to decrease the number of cuts. We perform a numerical study by increasing the number of scenarios and compare our solution with a solution obtained by solving the same problem with continuous distribution.

1 Introduction

In many real world applications the data are inherently random. This randomness many stem from many different sources: Let us mention imprecise parameter measurements, the necessity to consider future weather conditions or random people behavior, see the discussion in [48]. In such cases, deterministic problem formulation may perform subpar and it may be advantageous to consider stochastic (random) formulation where some parameters are not considered fixed but random. There are two basic approaches.

In the robust optimization [9], only all possible realizations of the random parameters are given and a solution performing best for the worst-case scenario is sought for. Even though this may be the correct concept, for example ambulance has to reach its patient within certain time limit, no matter where the patient is and what the traffic conditions are, this approach is usually too restrictive.

On the other hand, for the stochastic optimization [11] the distribution of the random parameters is also assumed to be known. Then there are many approaches how to handle this problem: For example to optimize the objective for a nominal scenario or in expectation if the uncertainties are in the objective. In this paper, we will consider *Chance-constrained problems (CCP)*. For this class of problems, the randomness appears only in the constraints and instead of requiring the constraints to be satisfied for all scenarios, it allows small violation of the constraints. Namely, we require that for some small $\varepsilon > 0$ it is sufficient that the constraints are satisfied with probability $1 - \varepsilon$, see (1). This provides a compromise between good system performance and satisfying the random constraints.

First, we give an overview of the main results concerning CCP. A general approach called sample (empirical) approximation is based on substituting the underlying continuous distribution by a finite sample and on reformulation as a (large) mixed-integer programming problem. The crucial question is the choice of the sample size, which is usually based on the exponential rates of convergence derived, e.g., by [30, 33]. However, these estimates can be too conservative, cf. [25]. Recently, [8] employed the importance sample technique to solve a chance constrained telecommunications problem

with Bernoulli input distributions. Exploiting its structure, they derived conditions to ensure a uniform variance reduction.

For linear constraints and finite discrete distribution, strong results and algorithms based on cutting planes for mixed-integer reformulations are available, cf. [10, 32, 34]. Recently, [51] derived new strong valid inequalities based on intersection of multiple mixing sets for the chance constrained problems with random right-hand side.

When the random parts of constraints are separated from the decision variables, we obtain the case with random right-hand side. In this case, the basic approach to individual chance constraints is to use quantiles and to reformulate the chance constraints in a deterministic way. This approach can be extended to joint chance constraints under discrete distribution using p-level efficient points (pLEPs) introduced by [37], which generalize the notation of quantiles to the multivariate case, see also [18, 31] for recent results. By adopting a dual point of view, [2] developed a solution framework based on a recent generation of bundle methods.

Nonlinear programming algorithms were suggested for chance constrained problems by [39] and further developed by [16, 17]. Recently, [19] proposed a smooth approximation approach employing an inner and an outer analytic approximation of chance constraints leading to two classes of nonlinear programming problems. [50] introduced quantile cuts which can be obtained as a projection of the mixing inequalities valid for the MINLP reformulation onto the original problem space. The paper further shows that a recursive application of quantile closure operations recovers the convex hull of the nonconvex chance constrained set in the limit.

A wide class of approaches is based on approximation of the indicator function by a more tractable function. Approximation based on conditional value at risk has been deeply investigated by [40, 41, 47]. Similar idea was used by [23] who employed the so-called integrated chance constraints. Bernstein approximation has been introduced by [36] for constraints affine in random coefficients and further developed by [6]. Recently, algorithmic approaches based on representation using difference of convex (DC) functions appeared in the literature, see [45, 47, 49]. A second-order cone programming reformulation was obtained by [14] for problems with linear constraints under normally distributed random coefficient and under independence and copula dependence of the rows. For these linear-Gaussian problems, [26] provided an explicit gradient formula and derived an efficient solution procedure.

Convexity is a desirable property which is often violated for CCP. Apart from well known cases based on log-concavity and its generalizations, cf. [38, 46], it was investigated and recently verified for general problems with high probability levels by [28, 1]. A special attention has been paid to the stability of the optimal values and solution with respect to the changes of the probability distribution, see, e.g., [12, 27]. Various approximations and worst-case bounds for distributionally robust chance constraints were derived by [13, 52].

In this paper we extend and combine our earlier work [5] and [22]. In [5] we considered discrete distribution and proposed a method for solving CCP with individual chance constraints. We introduced binary variables for each scenario, relaxed these variables and derived necessary optimality conditions and equivalence between the original and the relaxed problems. Finally, for numerical solution we proposed a regularization technique and successfully applied it to a portfolio optimization problem.

For application of stochastic optimization to gas networks under uncertainty we refer to, e.g., M. Bertocchi [7, 35]. In this paper we built on the results of [22]. We considered a gas network with random demands at all nodes. We showed an equivalent condition for satisfying all demands. For normally distributed scenarios we proposed a method for computing values and gradients of chance constraints based on the spheric-radial decomposition of normal random vectors. Finally, we optimized

the network design via a simple projection method.

Our contribution in this paper is the following:

- We extend the results of [5] to problems with joint chance constraints.

- Even though the problem is non-convex, we show that under convex data stationary points are local minima. In Theorem 11 we propose a general results for the same implication which can be used for hierarchical problems such as MPECs or MPCCs.

- We propose a Benders decomposition algorithm for the regularized problem. The algorithm is based on two cycles: The outer iterations make the regularization more tighter by increasing the regularization parameter whereas the inner cycle solves the regularized problem for a given value of the parameter. We employ feasibility cuts based on an explicit solution of a linear slave problem.

- We present a heuristic algorithm for a cut reduction. This is based on proving that certain cuts are naturally "included" in others.

- Since our method requires a discrete distribution, we empirically examine its convergence by increasing scenario number. Finally, we compare it with the method from [22] which works directly on a continuous distribution.

- We make the codes available online at `staff.utia.cas.cz/adam/research.html`

The paper is organized as follows: In Section 2 we derive theoretical results extending our work from [5] and we briefly summarize the results of [22]. Since the proof techniques are rather similar to those of [5], we moved the proofs to the Appendix. As the regularized problem includes an auxiliary variable with cardinality equalling to the number of scenarios, in Section 3 we propose a two-stage method to remove this auxiliary variable. This is based on iteratively solving a master problem while adding feasibility Benders cuts. For the slave problem we find an explicit solution, thus obtaining a theoretical independence on the number of scenarios. Unfortunately, this independence is not present in Section 4 where we apply both methods to design a gas network with random demands. However, we still show a very good performance of our method even for a large number of scenarios.

2   Methodology and Algorithms

The joint chance constrained problem may be formulated as follows:

$$
\begin{aligned}
&\underset{x}{\text{minimize}} \; f(x) \\
&\text{subject to } \mathbb{P}(g_1(x,\xi) \le 0, \dots, g_K(x,\xi) \le 0) \ge 1 - \varepsilon, \\
&\qquad\qquad h_j(x) \le 0, \; j = 1, \dots, J.
\end{aligned}
\tag{1}
$$

Here $x \in \mathbb{R}^n$ is the decision variable, $0 < \varepsilon < 1$ is a prescribed probabilistic level, $f : \mathbb{R}^n \to \mathbb{R}$, $g_k : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ and $h_j : \mathbb{R}^n \to \mathbb{R}$ are functions which are continuously differentiable in variable x and finally $\xi \in \mathbb{R}^d$ is a random vector with known probability distribution $\mathbb{P}$.

2.1   Comparison of discrete and continuous approaches

As we have mentioned in the introduction, concerning the probabilistic distribution there are two basic approaches to solve (1). In the first one, finite number of scenarios is sampled and the chance con-

straint is replaced by its discrete approximation. In the second one, we keep the original problem and work with this (usually) infinite dimensional formulation. The scenario approach is more general in the following way:

- Possible model simplification (in the application in Section 4 it allows to remove some constraints which became redundant after discretization).
- Lower requirements on the data (in the application in Section 4 it allows to work with cycles).
- Lower requirements on the random distribution. For the continuous approach a special assumption such as normality is often needed.

However, its biggest disadvantage is that it can handle only restricted number of scenarios, thus precision of the solution will be limited compared with the approach based on underlying continuous distribution.

2.2 Discrete (scenario) approach

In this part, we discretize the continuous distribution into possible realizations $\xi_1, \ldots, \xi_S$. Assuming that these realization may have different probabilities, we denote these probabilities by $p_1, \ldots, p_S$. We may then reformulate problem (1) into

$$\begin{aligned}
&\underset{x}{\text{minimize }} f(x) \\
&\text{subject to } \sum_{i=1}^{S} p_i \chi\left(\max_k g_k(x, \xi_i) \leq 0\right) \geq 1 - \varepsilon, \\
&\qquad\qquad h_j(x) \leq 0, \ j = 1, \ldots, J,
\end{aligned} \tag{2}$$

where χ stands for the characteristic function which equals to 1 if $\max_k g_k(x, \xi_i) \leq 0$ and to 0 otherwise. Introducing artificial binary variable $y \in \{0, 1\}^S$ to deal with χ, we obtain the following mixed–integer nonlinear problem

$$\begin{aligned}
&\underset{x,y}{\text{minimize }} f(x) \\
&\text{subject to } p^\top y \geq 1 - \varepsilon, \\
&\qquad\qquad y_i \in \{0, 1\}, \ i = 1, \ldots, S, \\
&\qquad\qquad g_k(x, \xi_i) y_i \leq 0, \ k = 1, \ldots, K, i = 1, \ldots, S, \\
&\qquad\qquad h_j(x) \leq 0, \ j = 1, \ldots, J.
\end{aligned} \tag{3}$$

Since this problem is difficult to tackle by mixed-integer (nonlinear) programming techniques in any of the previous forms, we relax binary variable $y_i \in \{0, 1\}$ into $y_i \in [0, 1]$ to obtain nonlinear programming problem

$$\begin{aligned}
&\underset{x,y}{\text{minimize }} f(x) \\
&\text{subject to } p^\top y \geq 1 - \varepsilon, \\
&\qquad\qquad 0 \leq y_i \leq 1, \ i = 1, \ldots, S, \\
&\qquad\qquad g_k(x, \xi_i) y_i \leq 0, \ k = 1, \ldots, K, i = 1, \ldots, S, \\
&\qquad\qquad h_j(x) \leq 0, \ j = 1, \ldots, J.
\end{aligned} \tag{4}$$

In the subsequent text, we denote (4) as the relaxed problem and (2) as the original problem.

The feasible relations for y_i and $g_k(x, \xi_i)$ are depicted in the top row of Figure 1. The feasible set of problem (2) can be written as a union of "nice" sets. The following index powersets will locally describe this decomposition:

$$\tilde{\mathscr{I}}(x) := \left\{ \tilde{I} \subset \{1, \ldots, S\} : \begin{array}{l} \max_k g_k(x, \xi_i) < 0 \implies i \in \tilde{I}, \\ \max_k g_k(x, \xi_i) > 0 \implies i \notin \tilde{I}, \end{array} \sum_{i \in \tilde{I}} p_i \geq 1 - \varepsilon \right\},$$

$$\mathscr{I}(x) := \text{minimal elements of } \tilde{\mathscr{I}}(x) \text{ with respect to set inclusion,}$$

Note that any $I \in \mathscr{I}(x)$ always contains all scenario indices where the random constraint is strictly satisfied and never contains scenario indices where the random constraint is not satisfied. For the scenarios where the random constraint is satisfied exactly, we select an arbitrary scenario subset such that the prescribed probability level $1 - \varepsilon$ is achieved.

Even though these problems are not equivalent, see [5, Example 2.1], there are close similarities between them, as stated in the next result.
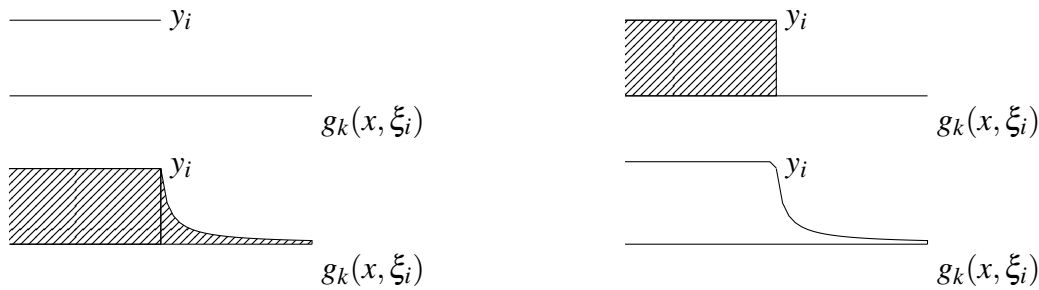

Figure 1: Feasible sets of problems (3) (top left), (4) (top right) and (8) (bottom left) and regularization ϕ_t from (10) (bottom right).

Proposition 1 *A point $\bar{x}$ is a global minimum of problem* (2) *if and only if there exists $\bar{y}$ such that $(\bar{x}, \bar{y})$ is a global minimum of problem* (4). *A point $\bar{x}$ is a local minimum of problem* (2) *if and only if for all $I \in \mathscr{I}(\bar{x})$ the point $(\bar{x}, \bar{y})$ is a local minimum of problem* (4), *where $\bar{y}_i = \chi(i \in I)$.*

When actually solving an optimization problem, we usually search for stationary points instead of local minima. To derive stationary conditions for problems (2) and (4), we define the following sets of active indices

$$\begin{array}{ll} I_0(x) := \{i : \max_k g_k(x, \xi_i) = 0\}, & I_{00}(x, y) := \{i : i \in I_0(x), y_i = 0\}, \\ K_0^i(x) := \{k : g_k(x, \xi_i) = 0\}, & I_{0+}(x, y) := \{i : i \in I_0(x), 0 < y_i \leq 1\}, \\ J_0(x) := \{j : h_j(x) = 0\} \end{array}$$

and the following constraint qualification. Note that its second part is close to the standard Mangasarian-Fromovitz constraint qualification in its dual form.

Assumption 2 *Let $\bar{x}$ be a feasible point of problem* (2). *Assume that at least one of the following two conditions is satisfied:*

■ *function $g_k(\cdot, \xi_i)$ and h_j are affine linear.*

■ *the following implication is satisfied for all $I \in \mathscr{I}(\bar{x})$:*

$$
\left.
\begin{aligned}
&\sum_{i \in I_0(\bar{x})} \sum_{k \in K_0^i(\bar{x})} \lambda_{ik} \nabla_x g_k(\bar{x}, \xi_i) + \sum_{j \in J_0(\bar{x})} \mu_j \nabla h_j(\bar{x}) = 0 \\
&\lambda_{ik} \geq 0, \ i \in I_0(\bar{x}) \cap I, \ k \in K_0^i(\bar{x}) \\
&\lambda_{ik} = 0, \ i \in I_0(\bar{x}) \setminus I, \ k \in K_0^i(\bar{x}) \\
&\mu_j \geq 0, \ j \in J_0(\bar{x})
\end{aligned}
\right\}
\implies
\begin{aligned}
&\lambda_{ik} = 0, \ i \in I_0(\bar{x}), \ k \in K_0^i(\bar{x}), \\
&\mu_j = 0, \ j \in J_0(\bar{x}).
\end{aligned}
\tag{5}
$$

Finally, we derive the stationarity conditions for both problems.

Theorem 3 *Let $\bar{x}$ be a local minimum of problem* (2) *and let Assumption 2 be satisfied at $\bar{x}$. Then for every $I \in \mathscr{I}(\bar{x})$ there exist multipliers λ_{ik}, $i \in I_0(\bar{x})$, $k \in K_0^i(\bar{x})$ and μ_j, $j \in J_0(\bar{x})$ such that*

$$
\nabla f(\bar{x}) + \sum_{i \in I_0(\bar{x})} \sum_{k \in K_0^i(\bar{x})} \lambda_{ik} \nabla_x g_k(\bar{x}, \xi_i) + \sum_{j \in J_0(\bar{x})} \mu_j \nabla h_j(\bar{x}) = 0,
\tag{6a}
$$

$$
\lambda_{ik} \geq 0, \ i \in I_0(\bar{x}) \cap I, \ k \in K_0^i(\bar{x}),
\tag{6b}
$$

$$
\lambda_{ik} = 0, \ i \in I_0(\bar{x}) \setminus I, \ k \in K_0^i(\bar{x}),
\tag{6c}
$$

$$
\mu_j \geq 0, \ j \in J_0(\bar{x}).
\tag{6d}
$$

Let $(\bar{x}, \bar{y})$ be a local minimum of problem (4) *and let Assumption 2 be satisfied at $\bar{x}$, where for its second part we check system* (5) *only for $I = I_{0+}(\bar{x}, \bar{y})$. Then there exist multipliers λ_{ik}, $i \in I_0(\bar{x})$, $k \in K_0^i(\bar{x})$ and μ_j, $j \in J_0(\bar{x})$ such that*

$$
\nabla f(\bar{x}) + \sum_{i \in I_0(\bar{x})} \sum_{k \in K_0^i(\bar{x})} \lambda_{ik} \nabla_x g(\bar{x}, \xi_i) + \sum_{j \in J_0(\bar{x})} \mu_j \nabla h_j(\bar{x}) = 0,
\tag{7a}
$$

$$
\lambda_{ik} \geq 0, \ i \in I_{0+}(\bar{x}, \bar{y}), \ k \in K_0^i(\bar{x}),
\tag{7b}
$$

$$
\lambda_{ik} = 0, \ i \in I_{00}(\bar{x}, \bar{y}), \ k \in K_0^i(\bar{x}),
\tag{7c}
$$

$$
\mu_j \geq 0, \ j \in J_0(\bar{x}).
\tag{7d}
$$

We briefly comment on these conditions. First, $\bar{y}$ enters system (7) only through index sets $I_{00}(\bar{x}, \bar{y})$ and $I_{0+}(\bar{x}, \bar{y})$. Second, the difference between (6) and (7) is only in the b) and c) part, where the signs are prescribed for different indices. This leads to the following result.

Corollary 4 *Consider a feasible point $\bar{x}$ of problem* (2) *and let Assumption 2 be satisfied at it. Then $\bar{x}$ is a stationary point of problem* (2) *if and only if for all $I \in \mathscr{I}(\bar{x})$ the point $(\bar{x}, \bar{y})$ with $\bar{y}_i = \chi(i \in I)$ is a stationary point of problem* (4).

Interestingly, when the data are convex, stationary points are also local minima even though the problem is nonconvex. This follows directly from a general result in Theorem 11 in the Appendix.

Theorem 5 *Let all data be convex in x. If a feasible point $\bar{x}$ of* (4) *satisfies the stationary conditions* (7)*, then it is a local minimum of problem* (4).

Unfortunately, as argued in [5, Remark 2.1], the Mangasarian-Fromovitz constraint qualification is often not satisfied for problem (4). For these reasons, we propose an additional technique which is based on a well–known solution approach to mathematical problems with complementarity constraints, see [44].

This technique enlarges the feasible set and solves the resulting regularized problem while driving the regularization parameter t to infinity. Thus, we consider regularized problem

$$
\begin{aligned}
\operatorname*{minimize}_{x,y} \quad & f(x) \\
\text{subject to} \quad & p^\top y \geq 1 - \varepsilon, \\
& 0 \leq y_i \leq 1, \ i = 1, \ldots, S, \\
& y_i \leq \phi_t(g_k(x, \xi_i)), \ k = 1, \ldots, K, \ i = 1, \ldots, S, \\
& h_j(x) \leq 0, \ j = 1, \ldots, J,
\end{aligned}
\tag{8}
$$

where $\phi_t : \mathbb{R} \to \mathbb{R}$ are continuously differentiable decreasing functions which depend on the regularization parameter $t > 0$ and which satisfy the following properties:

$$
\begin{aligned}
& \phi_t(0) = 1, && \tag{9a} \\
& \phi_t(z) > 0 && \text{for } z \in \mathbb{R}, \tag{9b} \\
& \phi_t(z^t) \to 0 && \text{whenever } z^t \stackrel{t \to \infty}{\to} \bar{z} > 0, \tag{9c} \\
& \frac{\phi_t'(z^t)}{\phi_t'(\tilde{z}^t)} \to 0 && \text{whenever } \phi_t(z^t) \searrow 0 \text{ and } \phi_t(\tilde{z}^t) \to \bar{z} > 0. \tag{9d}
\end{aligned}
$$

As an example of such regularizing function, we may consider

$$
\phi_t(z) = \begin{cases} e^{-tz} & \text{if } z \geq 0, \\ 1 - \frac{c}{(t+1)^2} \operatorname{arctg} \frac{t(t+1)^2}{c} z & \text{if } z < 0, \end{cases}
\tag{10}
$$

where $c > 0$ is an arbitrary parameter. Note that form of ϕ_t on the negative line is not important for (8) as both $y_i \leq 1$ and $y_i \leq \phi_t(g_k(x, \xi_i))$ are imposed. However, this form will play a crucial role later when the feasible region is approximated via Benders cuts; we comment more on this in the next section. The feasible relation for y_i and $g_k(x, \xi_i)$ for (8) and the regularizing function ϕ_t from (10) are depicted in the bottom row of Figure 1.

Now we justify the use of the regularized problem and show that it is a good approximation of the original problem. Note that we have to impose the second part of Assumption 2 as certain boundedness of multipliers is needed.

Theorem 6 *Consider $(\bar{x}^t, \bar{y}^t)$ to be stationary points of problem* (8). *Assume that the second part of Assumption 2 is satisfied at $\bar{x}$ and that $(\bar{x}^t, \bar{y}^t) \to (\bar{x}, \bar{y})$ as $t \to \infty$. Then $(\bar{x}, \bar{y})$ is a stationary point of problem* (4). *Moreover, if all data are convex in x, it is even a local minimum of problem* (4).

We summarize the previous results in Algorithm 2.1. Note that due to Theorem 6 the resulting point is only a stationary point of the relaxed problem (4) and due to Corollary 4 it does not have to be a stationary point of the original problem (2). However, as the numerical experience in [5] suggests, this is almost always the case due to the regularizing properties of (8).

2.3 Continuous approach

For the sake of comparison we present an alternative numerical solution approach addressing Gaussian and Gaussian-like distributions of the random vector without discretization. It is based on the so-called *spheric-radial decomposition* of Gaussian random vectors (see, e.g., [20]) which has been successfully applied to chance constrained optimization problems, e.g., [15, 21, 43].

Algorithm 2.1 for solving problem (1)

Input: starting point (x^0, y^0), regularization parameters $0 < t^1 < \cdots < t^L$

1: **for** $l = 1, \ldots, L$ **do**
2:      find (x^l, y^l) solving (8) with t^l and starting point (x^{l-1}, y^{l-1})
3:      **if** (x^l, y^l) is feasible for (4) **or** termination criterion is satisfied **then**
4:          **break**
5:      **end if**
6: **end for**
7: **return** $\bar{x} = x^l$

Theorem 7 *Let ξ be an d-dimensional Gaussian random vector distributed according to $\xi \sim \mathcal{N}(\mu, \Sigma)$. Then for any Borel measurable subset $A \subset \mathbb{R}^d$ it holds that*

$$\mathbb{P}(\xi \in A) = \int_{v \in \mathbb{S}^{d-1}} \mu_\chi \{r \geq 0 \,|\, (rLv + \mu) \cap A \neq \emptyset\} d\mu_\eta$$

where L is such that $\Sigma = LL^T$ (e.g., Cholesky decomposition), μ_χ is the Chi-distribution with d degrees of freedom and μ_η is the uniform distribution over the Euclidean unit sphere $\mathbb{S}^{d-1}$.

Accordingly, the x-dependent probability in (1) can be represented as

$$\mathbb{P}(g_1(x, \xi) \leq 0, \ldots, g_K(x, \xi) \leq 0) = \int_{v \in \mathbb{S}^{d-1}} \mu_\chi \{r \geq 0 \,|\, \max_{k=1,\ldots,K} g_k(x, rLv + \mu) \leq 0\} d\mu_\eta$$

Numerically, the chance constraint in (1) is then approximated as a finite sum

$$\sum_{i=1}^N \mu_\chi \{r \geq 0 \,|\, \max_{k=1,\ldots,K} g_k(x, rLv^i + \mu) \leq 0\} \geq 1 - \varepsilon,$$

where $\{v^1, \ldots, v^N\}$ is a sample (e.g. extracted from a Quasi Monte-Carlo sequence) approximating the uniform distribution on the unit sphere. In order to set up a nonlinear optimization algorithm solving problem (1) subject to a Gaussian random vector ξ, one has not only to compute (approximate) the probabilities above but also their gradients with respect to the decision variable x. As shown in [3, 4], the gradients can be represented as spheric integrals too (just with different integrands), so that one and the same sample v^i can be employed in order to update values and gradients of the probabilities above. We embedded this strategy into a simple projected gradient method.

3   Numerical method

The biggest disadvantage of solving (8) is that variables x and y are treated in an equal manner. Since y corresponds to the scenarios, this nonconvex problem becomes numerically untractable when the number of scenarios is large. In this section we first propose a method to eliminate y based on Benders decomposition, then derive some properties for these feasibility cuts and finally propose a method for their reduction.

3.1 Cut generation for fixed t

In this part we consider a fixed t and derive an outer approximation of the feasible set of (8). To this aim, consider the master problem

$$
\begin{aligned}
&\underset{x}{\text{minimize}} \ f(x) \\
&\text{subject to } v_b(x) \leq 0, \ b = 1, \ldots, B-1, \\
&\qquad\qquad h_j(x) \leq 0, \ j = 1, \ldots, J.
\end{aligned}
\tag{11}
$$

Here, $v_b(x) \geq 0$ are cuts which provide an outer approximation of the feasible region of (8) in the x dimension. If an optimal solution $\hat{x}$ of (11) is a feasible point for (8), then $(\hat{x}, y)$ is also an optimal solution of (8) for some y. In the opposite case, we generate a Benders cut v_B, which cuts away $\hat{x}$ from the feasible region of (11) but does not cut away any feasible point of (8). Thus the approximation is still the outer one. This cut is based on the linear slave problem for fixed $\hat{x}$:

$$
\begin{aligned}
&\underset{y}{\text{minimize}} \ 0 \\
&\text{subject to } p^{\top} y \geq 1 - \varepsilon, \\
&\qquad\qquad 0 \leq y_i \leq 1, \ i = 1, \ldots, S, \\
&\qquad\qquad y_i \leq \phi_t(g_k(\hat{x}, \xi_i)), \ k = 1, \ldots, K, i = 1, \ldots, S,
\end{aligned}
\tag{12}
$$

Its optimal value is 0 if $\hat{x}$ is feasible for (8) and $+\infty$ otherwise. The dual problem for (12) reads

$$
\begin{aligned}
&\underset{u,v,w}{\text{maximize}} \ \sum_{k=1}^{K} \sum_{i=1}^{S} \phi_t(g_k(\hat{x}, \xi_i)) u_{ik} + v(1 - \varepsilon) + \sum_{i=1}^{S} w_i \\
&\text{subject to } \sum_{k=1}^{K} u_{ik} + v p_i + w_i \leq 0, \ i = 1, \ldots, S, \\
&\qquad\qquad u_{ik} \leq 0, \ v \geq 0, \ w_i \leq 0.
\end{aligned}
\tag{13}
$$

Since the feasible region of (13) is a cone, its optimal value is zero or it is unbounded. In the former case, the optimal value of (12) is also zero, which means that $\hat{x}$ is optimal for (8). In the latter case, the dual slave problem (13) is unbounded in some direction $(\hat{u}, \hat{v}, \hat{w})$ and we construct the feasibility Benders cut

$$
v_B(x) := \sum_{k=1}^{K} \sum_{i=1}^{S} \phi_t(g_k(x, \xi_i)) \hat{u}_{ik} + \hat{v}(1 - \varepsilon) + \sum_{i=1}^{S} \hat{w}_i \leq 0,
\tag{14}
$$

add it to the master problem (11) and continue in the same way. We justify these cuts in the next statement.

Lemma 8 *Assume that problem* (12) *does not have a feasible solution. Then cut* (14) *is viable, thus* $v_B(\hat{x}) > 0$ *and* $v_B(x) \leq 0$ *if x is feasible for* (8).

Proof. Since (13) is unbounded in direction $(\hat{u}, \hat{v}, \hat{w})$, we immediately get $v_B(\hat{x}) > 0$. Consider now any feasible point x of (8). Then the optimal value of (12) with $\hat{x}$ replaced by x equals to zero and the same holds for (13). But this implies $v_B(x) \leq 0$, which concludes the proof. ∎

Even though (13) is a linear problem, it still depends on the number of scenarios S. Here, we derive its explicit solution, which means that we are able to generate the cuts quickly even for large S. Since

the feasible set of (13) is a cone and the objective is linear, we may consider only $v \in \{0,1\}$. If $v = 0$, then constraint $\sum_{k=1}^K u_{ik} + v p_i + w_i \leq 0$ is redundant and we may decompose (13) into S problems, from which we obtain $w_i = 0$. Since ϕ_t is positive, we also obtain $u_{ik} = 0$. Consider thus $v = 1$. Then we may again decompose (13) into S problems

$$\operatorname*{maximize}_{u_{ik}, w_i} \sum_{k=1}^K \phi_t(g_k(\hat{x}, \xi_i)) u_{ik} + w_i$$

$$\text{subject to } \sum_{k=1}^K u_{ik} + w_i \leq -p_i,$$

$$u_{ik} \leq 0, \ \ w_i \leq 0.$$

This problem has, together with $\hat{v} = 1$, the explicit solution

$$\max_k g_k(\hat{x}, \xi_i) > 0 \implies \phi_t(g_{\hat{k}_i}(\hat{x}, \xi_i)) < 1 \implies \hat{u}_{ik} = 0 \text{ for } k \neq \hat{k}_i, \ \hat{u}_{i\hat{k}_i} = -p_i, \ \hat{w}_i = 0, \quad (15a)$$

$$\max_k g_k(\hat{x}, \xi_i) \leq 0 \implies \phi_t(g_k(\hat{x}, \xi_i)) \geq 1 \implies \hat{u}_{ik} = 0 \text{ for } k \neq \hat{k}_i, \ \hat{u}_{i\hat{k}_i} = 0, \ \hat{w}_i = -p_i, \quad (15b)$$

where $\hat{k}_i := \arg\max_k g_k(\hat{x}, \xi_i)$. Then if the objective of (13) is positive, thus if

$$\sum_{i \in I(\hat{x})} p_i \phi_t(g_{\hat{k}_i}(\hat{x}, \xi_i)) < 1 - \varepsilon - \sum_{i \notin I(\hat{x})} p_i, \quad (16)$$

where $I(\hat{x}) := \{i | \ \max_k g_k(\hat{x}, \xi_i) > 0\}$, then problem (13) is unbounded and cut (14) amounts to

$$\sum_{i \in I(\hat{x})} p_i \phi_t(g_{\hat{k}_i}(x, \xi_i)) \geq 1 - \varepsilon - \sum_{i \notin I(\hat{x})} p_i, \quad (17)$$

In the opposite case $\hat{x}$ is optimal for (8).

3.2 Cut reduction

To propose a method for cut reduction, we start with the following technical lemma.

Lemma 9 *Consider two points $\hat{x}^1, \hat{x}^2$ such that for all i we have*

$$\max_k g_k(\hat{x}^2, \xi_i) > 0 \implies \max_k g_k(\hat{x}^1, \xi_i) > 0, \ \arg\max_k g_k(\hat{x}^1, \xi_i) = \arg\max_k g_k(\hat{x}^2, \xi_i). \quad (18)$$

Define now for $j = 1, 2$ mappings $\mathcal{K}^j : \{1, \ldots, S\} \to \{0, 1, \ldots, K\}$ and $v^j : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathcal{K}^j(i) := \begin{cases} 0 & \text{if } \max_k g_k(\hat{x}^j, \xi_i) \leq 0, \\ \arg\max_k g_k(\hat{x}^j, \xi_i) & \text{otherwise}, \end{cases}$$

$$v^j(x) := \sum_{\{i | \ \mathcal{K}^j(i) > 0\}} p_i \phi_t(g_{\mathcal{K}^j(i)}(x, \xi_i)) - 1 + \varepsilon + \sum_{\{i | \ \mathcal{K}^j(i) = 0\}} p_i.$$

Then for any x we have

$$v^1(x) \geq 0 \implies v^2(x) \geq -(\sup \phi_t(\cdot) - 1).$$

Proof. Due to (18) we have $\{i|\ \mathscr{K}^2(i)>0\}\subset\{i|\ \mathscr{K}^1(i)>0\}$ and subsequently

$$
\begin{aligned}
v^2(x) = v^1(x) - \sum_{\{i|\ \mathscr{K}^1(i)>0,\ \mathscr{K}^2(i)=0\}} p_i\phi_t(g_{\mathscr{K}^1(i)}(x,\xi_i)) + \sum_{\{i|\ \mathscr{K}^1(i)>0,\ \mathscr{K}^2(i)=0\}} p_i \\
\geq v^1(x) - \sum_{\{i|\ \mathscr{K}^1(i)>0,\ \mathscr{K}^2(i)=0\}} p_i(\sup\phi_t(\cdot)-1) \geq v^1(x) - (\sup\phi_t(\cdot)-1),
\end{aligned}
$$

which finishes the proof.   ∎

Note that the cuts generated by (17) equal to $\{x|\ v^j(x)\geq 0\}$. This lemma states that if (18) holds true, then the cut generated by $\hat{x}^1$ is tighter than the one generated by $\hat{x}^2$ up to a margin $\sup\phi_t(\cdot)-1$. Since $\phi_t(0)=1$, this margin may be made arbitrarily small by a proper choice of ϕ_t, the best cut in (17) is generated whenever $\max_k g_k(\hat{x},\xi_i)>0$ for all i.

This also gives rise to a cut reduction technique. If we add a new cut generated by some $\hat{x}^1$, we remove all the previously included cuts, generated by $\hat{x}^2$, for which (18) is satisfied. Alternatively, we remove all previous cuts, for which (18) is violated only for a small number of scenarios.

3.3   Algorithm summary

Denote the projections of the feasible sets of (8) and (4) into the x dimension by Z^t and Z^∞, respectively. For any $t_1 < t_2$ we have $Z^{t_1} \supset Z^{t_2} \supset Z^\infty$. This means that cuts generated for (8) for $t = t_1$ are also valid for problem (8) with $t = t_2$. In other words, when we pass to a greater t, it is not necessary to delete cuts from (11). We summarize the whole procedure in Algorithm 3.1.

Algorithm 3.1 for solving problem (1)

Input: starting point (x^0, y^0), regularization parameters $0 < t^1 < \cdots < t^L$

1: $B \leftarrow 1$
2: **for** $l = 1, \ldots, L$ **do**
3:      $t \leftarrow t^l$
4:      **while true do**
5:          find $\hat{x}^{lB}$ solving (11)
6:          set $\hat{v} = 1$ and $(\hat{u}, \hat{w})$ according to (15)
7:          **if** (16) is violated **or** termination criterion **then**
8:              **break while**
9:          **else**
10:             add cut v_B via (17)
11:             remove all cuts generated by $\hat{x}^2$ for which (18) is satisfied with $\hat{x}^1 = \hat{x}^{lB}$
12:             update B to equal to the current cut number
13:         **end if**
14:     **end while**
15:     **if** $\hat{x}^{lB}$ is feasible for (2) **or** termination criterion is satisfied **then**
16:         **break for**
17:     **end if**
18: **end for**
19: **return** $\bar{x} = \hat{x}^{lB}$

3.4 Convergence analysis

First note that the number of possible cuts in (17) is finite and thus Algorithm 3.1 stops after finite number of iterations. Moreover, from the discussion above, there are K^S best cuts. This has a direct consequence for individual chance constraints with $K = 1$, where there is only one best cut which in (17) corresponds to $I(\hat{x}) = \{1, \dots, S\}$ and takes the form

$$\sum_{i=1}^{S} p_i \phi_t(g_1(x, \xi_i)) \geq 1 - \varepsilon.$$

which is nothing else than direct smoothing of chance constraints, see [45]. This cut is also close to the true feasible set which equals to

$$\sum_{i=1}^{S} p_i \max\{\phi_t(g_1(x, \xi_i)), 1\} \geq 1 - \varepsilon. \tag{19}$$

Finally, from the next lemma it follows that the added cuts are optimal as they cannot be linear combination of each other. Recall that a direction d is an extremal direction of a cone C if there do not exist directions $d_1, d_2 \in C$ different from d and a scalar $\kappa \in (0,1)$ such that $d = \kappa d_1 + (1 - \kappa)d_2$.

Lemma 10 *The direction defined in* (15) *is an extremal direction of the feasible set of* (13).

Proof. Denote the feasible set of problem (13) by Z. For contradiction assume that $(\hat{u}, \hat{v}, \hat{w})$ is not an extremal direction of Z. Then there are some $(u^1, v^1, w^1) \in Z$ and $(u^2, v^2, w^2) \in Z$ different from $(\hat{u}, \hat{v}, \hat{w})$ and some $\kappa \in (0,1)$ such that

$$(\hat{u}, \hat{v}, \hat{w}) = \kappa(u^1, v^1, w^1) + (1 - \kappa)(u^2, v^2, w^2). \tag{20}$$

Note now that the role of $\hat{u}_{i\hat{k}_i}$ and $\hat{w}_i$ is symmetric in (15) and thus, it suffices to consider only (15b). Define

$$Z_1 := \{(v, w_1) \mid vp_1 + w_1 \leq 0, \ v \geq 0, \ w_1 \leq 0\}$$

and observe that $(\hat{v}, \hat{w}_1), (v^1, w_1^1), (v^2, w_1^2) \in Z_1$. But this due to (20) means that $(\hat{v}, \hat{w}_1)$ is not an extremal direction of Z_1. But since the extremal directions of Z_1 amount to $(0, -p_1)$ and $(1, -p_1)$, this is a contradiction with (15b). ∎

4 Numerical experiments

In this section we show a very good performance of the proposed discrete method on a gas network problem from [22].

4.1 Application to gas network design problem

We consider a gas network described as follows:

- ◼ withdrawal points (exit nodes) $\mathscr{V} = \{1, \dots, n\}$ with random exit loads $\xi = (\xi^1, \dots, \xi^n)$,
- ◼ one injection point 0 corresponding to the root,

- directed edges (pipes) $\mathscr{E}$ with $e = (i, j) \subset \mathscr{V} \times \mathscr{V}$ and coefficient of the pressure drop in Φ_e
- lower and upper pressure bounds p_k^{min}, p_k^{max}.

For trees-connected networks without cycles, it was shown in [22] that a random demand ξ can be satisfied if and only if

$$
\begin{aligned}
(p_0^{min})^2 &\leq (p_k^{max})^2 + h_k(\xi), \ k = 1, \ldots, |\mathscr{V}|, \\
(p_0^{max})^2 &\geq (p_k^{min})^2 + h_k(\xi), \ k = 1, \ldots, |\mathscr{V}|, \\
(p_k^{max})^2 + h_k(\xi) &\geq (p_l^{min})^2 + h_l(\xi), \ k, l = 1, \ldots, |\mathscr{V}|,
\end{aligned}
\tag{21}
$$

Here, functions $h_k(\xi)$ can be computed by

$$
h_k(\xi) = \sum_{e \in \Pi(k)} \Phi_e \left(\sum_{l \in \mathscr{V}, \ l \succeq \pi(e)} \xi^l \right)^2 ,
$$

where $\Pi(k)$ denotes the unique directed path (edges) from the root to node k, $k \succeq l$ means that the unique path from root to k passes through l and $\pi(e)$ is the end node of edge e.

There are many ways of defining the objective. The simplest way is to minimize the upper pressure bounds, which with a cost vector c, results in

$$
\begin{aligned}
&\underset{p^{max}}{\text{minimize}} \ c^\top p^{max} \\
&\text{subject to } \mathbb{P}(\text{system (21) is fulfilled}) \geq 1 - \varepsilon.
\end{aligned}
\tag{22}
$$

The minimal capacities p^{min} are usually considered to be fixed. Then we can reduce the number of inequalities in (21) to $(p_k^{max})^2 \geq v_k, k = 0, 1, \ldots, |\mathscr{V}|$, where we make use of $h_0(\xi) \equiv 0$ and where

$$
\begin{aligned}
v_0 &:= \max_{k = 0, 1, \ldots, |\mathscr{V}|} \left\{ (p_k^{min})^2 + h_k(\xi) \right\}, \\
v_k &:= v_0 - h_k(\xi).
\end{aligned}
\tag{23}
$$

Then problem (22) reduces to

$$
\begin{aligned}
&\underset{p^{max}}{\text{minimize}} \ c^\top p^{max} \\
&\text{subject to } \mathbb{P}\left((p_0^{max})^2 \geq v_0, \ldots, (p_n^{max})^2 \geq v_n \right) \geq 1 - \varepsilon.
\end{aligned}
\tag{24}
$$

4.2 Parameter description

In this short section we describe parameter choice and their update. For the probability level we chose $\varepsilon = 0.15$. Whenever we needed to compute the true probability, we employed the technique from Section 2.3 with 100000 samples.

For the discrete (scenario) approach, we randomly generated 10000 scenarios of ξ with quasi-Monte Carlo sampling. Then we used Algorithm 3.1 to solve (22) with fixed scenario number $S \in \{100, 300, 500, 1000, 3000, 5000, 10000\}$. We made use of all generated scenarios, thus we solved it altogether 100 times for $S = 100$, 33 times for $S = 300$ and so on. For the result presentation we then averaged the results.

We solved the master problem (11) by the SQP method implemented in MATLAB's `fmincon`. We set the initial $t^1 = 10^{-5}$ and employed the simplest update rule $t^{l+1} = 2t^l$. For all not-the-last t, we increased it whenever constraint

$$p^\top y \geq 1 - \varepsilon - 0.0005$$

was satisfied or whenever we added 100 Benders cuts. For the last t^{20} we used the usual stopping criterion $p^\top y \geq 1 - \varepsilon$. We removed cuts based on the strategy presented in Section 3.2 whenever at most 2 indices differed for $S \leq 1000$ or whenever at most 20 indices differed for $S > 1000$. Even though this is a highly heuristic strategy, it performed very well. For algorithm details we refer to our codes available online.

4.3 Numerical results

In this section we solve problem (22) by both the discrete and continuous approach, thus by the procedures described in Algorithm 3.1 and Section 2.3, respectively. We considered two networks depicted in Figure 2.

Figure 2: Network topology of small (left) and medium (right) size network examples. The entry and exit points of the considered gas transportation networks are displayed in black (entry) and white (exit), respectively.

Table 1 displays the results for the small network whereas Table 2 corresponds to the larger network. Every column corresponds to a given scenario count S. In the first part of both tables we have the mean and standard deviation of the obtained optimal value and probability. For the probability we present the reached true probability and the reached probability inside the sample. This corresponds to the standard machine learning technique to dividing the data into training and testing samples, see [24]. In the second part of the table, we depict the number of solves of (11), number of cuts B and the number of eliminated cuts via the procedure described in Section 3.2.

There are several things worth mentioning. Namely, with increasing S:

■ The probability in sample stays close to the desired 85%. This means that Algorithm 3.1 always either managed to find a solution or was close to it.

■ The true probability increases. This makes sense as the approximation of the true distribution gets better.

Scenario number S	100	300	500	1000	3000	5000	10000
Mean objective	728.1	734.1	735.6	737.2	737.5	737.7	738.0
Mean probability (true) [%]	79.94	82.97	83.60	84.14	84.50	84.69	84.84
Mean probability (in sample) $[\%]$	84.97	84.96	84.99	85.00	84.99	84.96	84.98
SD objective	5.4	3.0	1.4	1.6	0.5	0.5	-
SD probability (true) [%]	1.72	0.87	0.45	0.29	0.13	0.21	-
Number of solves of (11)	137.8	177.8	211.3	253.1	341.7	385.0	647.0
Number of cuts B	75.2	126.8	162.5	209.3	216.0	248.0	537.0
Cut reduction $[\%]$	39.01	22.31	17.41	12.21	33.47	32.05	14.63

Table 1: Network I: Results for small network with $n = K = 4$.

Scenario number S	100	300	500	1000	3000	5000	10000
Mean objective	3022.2	3079.6	3097.0	3112.2	3127.5	3131.4	3132.3
Mean probability (true) [%]	77.32	81.08	82.22	83.20	84.09	84.37	84.42
Mean probability (in sample) $[\%]$	84.98	85.00	85.00	84.99	84.96	84.84	84.80
SD objective	45.5	24.9	18.0	11.8	5.4	2.7	-
SD probability (true) [%]	2.51	1.32	0.77	0.52	0.23	0.09	-
Number of solves of (11)	572.8	716.8	789.1	876.6	2011.3	2613.0	2976.0
Number of cuts B	513.8	678.3	756.6	839.8	1961.7	2603.5	2970.0
Cut reduction $[\%]$	8.38	4.00	3.03	3.35	2.01	0.02	0.00

Table 2: Network II: Results for medium size network with $n = K = 12$.

■ The objective gets worse and the standard deviation decreases. This also make sense for the same reason as in the previous bullet.

■ The number of solves of (11) increases. Even though it should be constant theoretically, our guess is that cuts (14) provide worse approximation of the feasible set for large S and thus the efficiency decreases.

■ Cut reduction decreases. Empirically we found out that most of the cut reduction happens at the end of the algorithm (and seems to be necessary for its convergence). For large S the algorithm either needed large number of cuts to get close to the solution or was only able to get close to it but not actually find it.

Not surprisingly the larger network needed more iterations.

Method	CCP (projection)	ProGrad	CCP + ProGrad
Network I	738.56	738.40	738.40
Network II	3146.52	3157.68	3145.47

Table 3: Comparison of the best CCP solution (projected onto the feasibility set) with the solution of the continuous approach based on spheric-radial decomposition (ProGrad). In addition, a combination of both methods has been performed, where the CCP solution serves as the starting point for the ProGrad method (CCP + ProGrad).

In Table 3 we compare the discrete method with the continuous one for both Network I and Network II. We have chosen the best discrete approximation which is obtained for 10000 samples. Since its obtained probability level is less then the required 85%, we have projected it onto the feasible set using the gradient information that is available by the spheric-radial approach, see Section 2.3. This

optimal value is displayed in the column "CCP (projection)". The next column shows the results for the continuous method when applying a simple projected gradient method (ProGrad). Finally, a combination of both methods is performed, where the computed CCP solution serves as a starting point for the ProGrad computation. In both examples, the best result is obtained by the combination of CCP and ProGrad. For the larger network, CCP performed better than ProGrad. Since the problem is highly nonconvex, we believe that this superiority is caused by the slow update in t which is able to evade local minima, see [5, Example 3.1].
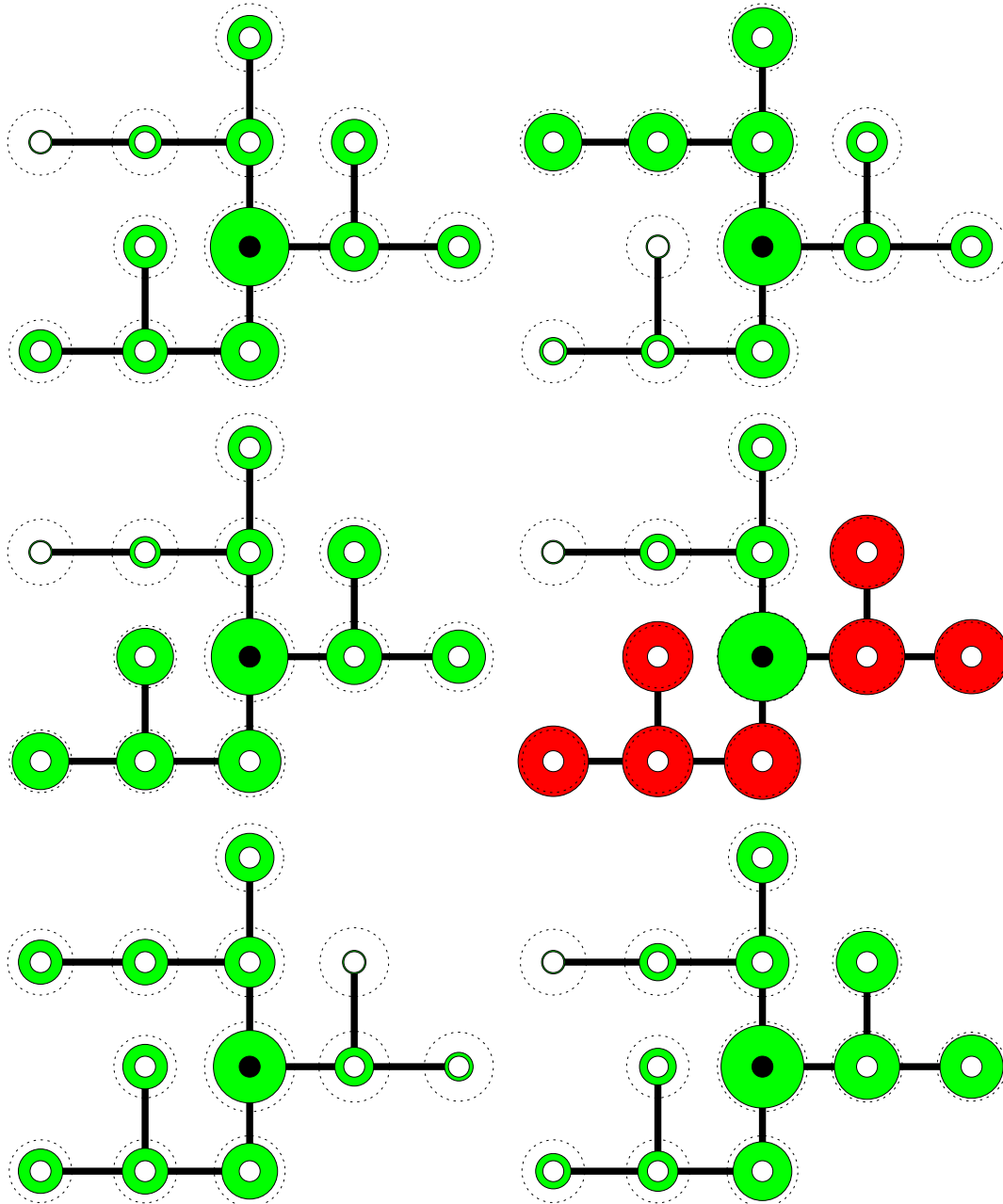
Figure 3: Six simulations of exit pressure realizations compared to the upper pressure bounds (dashed lines) provided by the numerical solution for Network II. Feasible and infeasible pressure realizations at each node are shown in green and red, respectively.

In Figure 3 we perform a posterior check of the computed solution. We simulate six sets of exit loads according to the given Gaussian distribution and check whether the corresponding minimal pressure

is feasible with respect to the computed upper pressure limits. Feasible pressures are displayed in green whereas violated pressures in red. If all nodes are shown in green, this load scenario is feasible. According to Figure 3, five out of six gas demand scenarios are feasible which agrees with the prescribed probability level 85%. Moreover, the method seems to be rather stable as for the infeasible scenario, the minimal needed pressures did not exceed the upper pressure bounds by a large margin.

A New result for sufficient optimality conditions for hierarchical problems

In this short section we present a new result which may play a crucial role in deriving sufficient optimality conditions for hierarchical problems such as bilevel problems, mathematical problems with equilibrium/complementarity/vanishing constraints and so on. In these problems, the feasible set is usually rather nasty but may be written as a finite union of nice sets. We show that if these nice sets are convex, then strong stationary points are immediately local minima.

Theorem 11 *Consider a convex differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, a set $X \subset \mathbb{R}^n$, a point $\bar{x} \in X$ and an optimization problem*

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in X. \end{aligned} \tag{25}$$

Assume that $\bar{x}$ is its S-stationary point, thus a point with

$$0 \in \nabla f(\bar{x}) + \hat{N}_X(\bar{x}),$$

where $\hat{N}_X(\bar{x})$ stands for the Fréchet normal cone of X at $\bar{x}$. If X can be locally around $\bar{x}$ written as a union of finite number of (possibly overlapping) convex sets, then $\bar{x}$ is a local minimum of problem (25).

Proof. From the theorem statement, there are convex sets X_i, $i = 1, \ldots, I$ such that locally around $\bar{x}$ we have that X coincides with $\cup_{i=1}^I X_i$. First, we realize that

$$\hat{N}_X(\bar{x}) = (T_X(\bar{x}))^* = (\cup_{i=1}^I T_{X_i}(\bar{x}))^* = \cap_{i=1}^I (T_{X_i}(\bar{x}))^* = \cap_{i=1}^I N_{X_i}(\bar{x}).$$

Since $\bar{x}$ is a S-stationary point of (25), we have

$$0 \in \nabla f(\bar{x}) + \hat{N}_X(\bar{x}) = \nabla f(\bar{x}) + \bigcap_{i=1}^I N_{X_i}(\bar{x}) = \bigcap_{i=1}^I \left(\nabla f(\bar{x}) + N_{X_i}(\bar{x}) \right).$$

Fix now any i. From the equation above we obtain that $\bar{x}$ is a stationary point of

$$\begin{aligned} \text{minimize } & f(x) \\ \text{subject to } & x \in X_i. \end{aligned}$$

Due to the data convexity, it is a local minimum of the above problem and thus for all $x \in X_i$ sufficiently close to $\bar{x}$ we have $f(x) \geq f(\bar{x})$. But since i was chosen arbitrarily, we obtain that $\bar{x}$ is a local minimum of problem (25). ∎

B   Proofs

In this section we collect the skipped proofs from Section 2.

Proposition 1. It is sufficient to follow the proof of Lemma 3.1 in [5].   ∎

Theorem 3. Denote the feasible set of problem (2) by Z and consider a point $\bar{x} \in Z$. Then Z coincides locally around $\bar{x}$ with

$$\bigcup_{I \in \mathscr{I}(\bar{x})} Z_I := \bigcup_{I \in \mathscr{I}(\bar{x})} \left\{ x : \begin{array}{l} g_k(x, \xi_i) \leq 0, \ i \in I_0(\bar{x}) \cap I, \ k \in K_0^i(\bar{x}) \\ h_j(x) \leq 0, \ j \in J_0 \end{array} \right\},$$

which means that

$$\hat{N}_Z(\bar{x}) = (T_Z(\bar{x}))^* = \left(\bigcup_{I \in \mathscr{I}(\bar{x})} T_{Z_I}(\bar{x}) \right)^* = \bigcap_{I \in \mathscr{I}(\bar{x})} \hat{N}_{Z_I}(\bar{x}).$$

By [42, Theorem 6.12] we obtain that $0 \in \nabla f(\bar{x}) + \hat{N}_Z(\bar{x})$ is a necessary optimality condition for chance constrained problem (2). To obtain the first statement, it suffices to use chain rule [42, Theorem 6.14].

The proof of the second part goes in a similar way. Due to [42, Theorem 6.12], the necessary optimality conditions for problem (4) read

$$0 \in \begin{pmatrix} \nabla f(\bar{x}) \\ 0 \end{pmatrix} + \hat{N}_Z(\bar{x}, \bar{y}) \tag{26}$$

where Z is the feasible set of problem (4). For the computation of the normal cone, realize first that Z locally around $(\bar{x}, \bar{y})$ coincides with the union of $Z_I := Z_I^x \times Z_I^y$ with respect to all $I \subset I_{00}(\bar{x}, \bar{y})$, where

$$Z_I^x := \left\{ x : \begin{array}{l} g_k(x, \xi_i) \leq 0, \ i \in I \cup I_{0+}(\bar{x}, \bar{y}), \ k \in K_0^i(\bar{x}) \\ h_j(x) \leq 0, \ j \in J_0 \end{array} \right\}$$

$$Z_I^y := \left\{ y : \begin{array}{l} p^\top y \geq 1 - \varepsilon \\ y_i \in [0,1], \ i \in I \cup I_{0+}(\bar{x}, \bar{y}) \cup \{i : \max_k g_k(\bar{x}, \xi_i) < 0\} \\ y_i = 0, \ i \in (I_{00}(\bar{x}, \bar{y}) \setminus I) \cup \{i : \max_k g_k(\bar{x}, \xi_i) > 0\} \end{array} \right\}.$$

As before, we have

$$\hat{N}_Z(\bar{x}, \bar{y}) = \bigcap_{I \subset I_{00}(\bar{x}, \bar{y})} \hat{N}_{Z_I}(\bar{x}, \bar{y}) = \bigcap_{I \subset I_{00}(\bar{x}, \bar{y})} \hat{N}_{Z_I^x}(\bar{x}) \times \bigcap_{I \subset I_{00}(\bar{x}, \bar{y})} \hat{N}_{Z_I^y}(\bar{y}).$$

Since zero always belongs to a normal cone, the optimality condition (26) is equivalent to

$$0 \in \nabla f(\bar{x}) + \bigcap_{I \subset I_{00}(\bar{x}, \bar{y})} \hat{N}_{Z_I^x}(\bar{x}). \tag{27}$$

To finish the proof, it suffices to realize that the intersection in (27) is manifested for $I = \emptyset$ and to use either [29, Proposition 3.4] (if the first part of Assumption 2 holds true) or [42, Theorem 6.14] (if the second part of Assumption 2 holds true).   ∎

The proof of Theorem 6 is more complicated. For notational simplicity, we consider only the case of $J_0(\bar{x}) = \emptyset$. First, we write down the stationarity conditions of (8), then show two preliminary lemmas and only then proof the theorem itself.

The necessary optimality conditions for problem (8) at a point $(\bar{x}^t, \bar{y}^t)$ read as follows: there exist multipliers $\alpha^t \in \mathbb{R}$, $\beta^t \in \mathbb{R}^S$ and $\gamma^t \in \mathbb{R}^{SK}$ such that the optimality conditions

$$0 = \nabla f(\bar{x}^t) - \sum_{i=1}^{S} \sum_{k=1}^{K} \gamma_{ik}^t \phi_t'(g_k(\bar{x}^t, \xi_i)) \nabla_x g_k(\bar{x}^t, \xi_i), \tag{28a}$$

$$0 = -\alpha^t p_i + \beta_i^t + \sum_{k=1}^{K} \gamma_{ik}^t, \quad i = 1, \dots, S \tag{28b}$$

and the complementarity conditions

$$\alpha^t (1 - \varepsilon - p^\top \bar{y}^t) = 0, \tag{29a}$$

$$\beta_i^t \begin{cases} \geq 0 & \text{if } \bar{y}_i^t = 1, \\ = 0 & \text{if } 0 < \bar{y}_i^t < 1, \\ \leq 0 & \text{if } \bar{y}_i^t = 0, \end{cases} \tag{29b}$$

$$\gamma_{ik}^t (\bar{y}_i^t - \phi_t(g_k(\bar{x}^t, \xi_i))) = 0 \tag{29c}$$

are satisfied. Moreover, the sign restrictions $\alpha^t \geq 0$ and $\gamma_{ik}^t \geq 0$ hold true.

Lemma 12 *Assume that $(\bar{x}^t, \bar{y}^t)$ is a stationary point of problem* (8). *Then the following assertions hold true:*

1 $g_k(\bar{x}^t, \xi_i) < 0 \implies \gamma_{ik}^t = 0$;

2 $\alpha^t = 0 \implies \beta_i^t = \gamma_{ik}^t = 0$ for all i and k;

Lemma 13 *If for all t we have $p^\top \bar{y}^t = 1 - \varepsilon$ and for some i and k we have $\phi_t(g_k(\bar{x}^t, \xi_i)) = \bar{y}_i^t \searrow 0$, then there exists a subsequence in t such that $\gamma_{ik}^t = 0$ for all t or such that there exists index j such that*

$$\frac{-\gamma_{ik}^t \phi_t'(g_k(\bar{x}^t, \xi_i))}{-\sum_{\tilde{k}=1}^{K} \gamma_{j\tilde{k}}^t \phi_t'(g_{\tilde{k}}(\bar{x}^t, \xi_j))} \to 0, \tag{30}$$

Proof. Due to the assumptions there exists index j, and possibly a subsequence in t, such that $\bar{y}_j^t$ is strictly increasing. This implies that $0 < \bar{y}_j^t < 1$ and $\beta_i^t = \beta_j^t = 0$ for all t. If $\gamma_{ik}^t = 0$, then the proof is finished. In the opposite case, we realize that $\alpha^t > 0$, $\gamma_{ik}^t \geq 0$ and $\phi_t'(g_k(\bar{x}^t, \xi_i)) < 0$ to deduce

$$
\begin{aligned}
\frac{-\gamma_{ik}^t \phi_t'(g_k(\bar{x}^t, \xi_i))}{-\sum_{\tilde{k}=1}^{K} \gamma_{j\tilde{k}}^t \phi_t'(g_{\tilde{k}}(\bar{x}^t, \xi_j))} &\leq \frac{-\left(\sum_{\tilde{k}=1}^{K} \gamma_{i\tilde{k}}^t\right) \phi_t'(g_k(\bar{x}^t, \xi_i))}{-\sum_{\tilde{k}=1}^{K} \gamma_{j\tilde{k}}^t \phi_t'(g_{\tilde{k}}(\bar{x}^t, \xi_j))} \\
&\leq \frac{-\sum_{\tilde{k}=1}^{K} \gamma_{i\tilde{k}}^t}{-\sum_{\tilde{k}=1}^{K} \gamma_{j\tilde{k}}^t} \frac{\phi_t'(g_k(\bar{x}^t, \xi_i))}{\phi_t'(g_{\hat{k}^t}(\bar{x}^t, \xi_j))} \\
&= \frac{-\alpha^t p_i}{-\alpha^t p_j} \frac{\phi_t'(g_k(\bar{x}^t, \xi_i))}{\phi_t'(g_{\hat{k}^t}(\bar{x}^t, \xi_j))} \to 0
\end{aligned}
\tag{31}
$$

where $\hat{k}^t := \arg\max_{\tilde{k}} \phi_t'(g_{\tilde{k}}(\bar{x}^t, \xi_j))$, the last equality follows from (28b) and the convergence follows from assumption (9d), for which we realize that $\phi_t(g_{\tilde{k}^t}(\bar{x}^t, \xi_j)) \geq \bar{y}_j^t$, the fact that $\bar{y}_j^t$ is a strictly increasing sequence and the assumed convergence $\phi_t(g_k(\bar{x}^t, \xi_i)) = \bar{y}_i^t \searrow 0$. ∎

Theorem 6. We will show first that $(\bar{x}, \bar{y})$ is a feasible point of (4). Due to continuity, it is sufficient to show that $g_k(\bar{x}, \xi_i)\bar{y}_i \leq 0$. Since this relation is obvious whenever $g_k(\bar{x}, \xi_i) \leq 0$ for all k, we consider scenario i with $g_k(\bar{x}, \xi_i) > 0$ for some k. But then $g_k(\vec{x}^t, \xi_i) > 0$ for sufficiently large t and thus $0 \leq \bar{y}_i^t \leq \phi_t(g_k(\vec{x}^t, \xi_i))$. But since $g_k(\vec{x}^t, \xi_i) \to g_k(\bar{x}, \xi_i) > 0$, assumption (9c) implies that $\bar{y}_i = 0$, and thus $(\bar{x}, \bar{y})$ is a feasible point of problem (4).

Define now

$$\lambda_{ik}^t := -\gamma_{ik}^t \phi_t'(g_k(\vec{x}^t, \xi_i)) \geq 0,$$

where the nonnegativity follows from the property that ϕ_t is decreasing. Then for a subsequence in t, optimality condition (28a) reads

$$
\begin{aligned}
0 &= \nabla f(\vec{x}^t) + \sum_{i=1}^{S}\sum_{k=1}^{K} \lambda_{ik}^t \nabla_x g_k(\vec{x}^t, \xi_i) \\
&= \nabla f(\vec{x}^t) + \sum_{\{(i,k):g_k(\bar{x},\xi_i)=0,\bar{y}_i>0\}} \lambda_{ik}^t \nabla_x g_k(\vec{x}^t, \xi_i) + \sum_{\{(i,k):g_k(\bar{x},\xi_i)\geq0,\ \bar{y}_i=0\}} \lambda_{ik}^t \nabla_x g_k(\vec{x}^t, \xi_i).
\end{aligned}
\tag{32}
$$

Here we can omit pairs of indices (i,k) with $g_k(\bar{x}, \xi_i) < 0$ due to Lemma 12.

We claim now that $\sum_{k=1}^{K} \lambda_{ik}^t$ is uniformly bounded in i and t. If this is not the case, then we have

$$\lambda_{max}^t := \max_{i=1,\dots,S} \sum_{k=1}^{K} \lambda_{ik}^t \to \infty.$$

Then dividing equation (32) by λ_{max}^t yields

$$0 = \frac{1}{\lambda_{max}^t}\nabla f(\vec{x}^t) + \sum_{\{(i,k):g_k(\bar{x},\xi_i)=0,\bar{y}_i>0\}} \frac{\lambda_{ik}^t}{\lambda_{max}^t}\nabla_x g_k(\vec{x}^t, \xi_i) + \sum_{\{(i,k):g_k(\bar{x},\xi_i)\geq0,\ \bar{y}_i=0\}} \frac{\lambda_{ik}^t}{\lambda_{max}^t}\nabla_x g_k(\vec{x}^t, \xi_i). \tag{33}$$

When taking limit $t \to \infty$, the first term vanishes. Consider now the third term. If $p^\top \bar{y}^t > 1 - \varepsilon$, from Lemma 12 we have $\lambda_{ik}^t = 0$ for all i and k. Assume thus that $p^\top \bar{y}^t = 1 - \varepsilon$ for all t. If $\phi_t(g_k(\vec{x}^t, \xi_i)) > \bar{y}_i^t$, then from Lemma 12 we have $\lambda_{ik}^t = 0$. In the opposite case, we may use Lemma 13 to obtain again that $\lambda_{ik}^t = 0$ or there exists j such that $\frac{\lambda_{ik}^t}{\sum_{\tilde{k}=1}^K \lambda_{j\tilde{k}}^t} \to 0$. But this implies that the last term in (33) vanishes as well. This means that

$$0 = \sum_{\{(i,k):g_k(\bar{x},\xi_i)=0,\bar{y}_i>0\}} \lim_{t\to\infty} \frac{\lambda_{ik}^t}{\lambda_{max}^t}\nabla_x g_k(\vec{x}^t, \xi_i).$$

Since $\frac{\lambda_{ik}^t}{\lambda_{max}^t} \in [0,1]$ and the numerators sum to one, at least one of these fractions converges to a positive number. However, the existence of such positive limit contradicts Assumption 2 and thus $\sum_{k=1}^{K}\lambda_{ik}^t$ is indeed bounded. Since it is a sum of nonnegative elements, these elements λ_{ik}^t are uniformly bounded in i, k and t.

This means that we may pass to a converging subsequence, say $\lambda_{ik}^t \to \lambda_{ik}$. Since $\lambda_{ik}^t \geq 0$ for all t, the same property holds for λ_{ik}. In the light of (32), to finish the proof it suffices to show that $\lambda_{ik} = 0$ for all pairs (i,k) such that $g_k(\bar{x}, \xi_i) \geq 0$ and $\bar{y}_i = 0$. But this may be shown as in the previous paragraph via applying Lemmas 12 and 13. Thus $(\bar{x}, \bar{y})$ is indeed a stationary point of problem (4). ∎

References

[1] van Ackooij, W.: Eventual convexity of chance constrained feasible sets. Optimization **64**(5), 1263–1284 (2015)

[2] van Ackooij, W., Berge, V., de Oliveira, W., Sagastizábal, C.: Probabilistic optimization via approximate p-efficient points and bundle methods. Computers & Operations Research **77**, 177–193 (2017)

[3] van Ackooij, W., Henrion, R.: Gradient formulae for nonlinear probabilistic constraints with gaussian and gaussian-like distributions. SIAM Journal on Optimization **24**, 1864–1889 (2014)

[4] van Ackooij, W., Henrion, R.: (Sub-) gradient formulae for probability functions of random inequality systems under gaussian distribution. SIAM/ASA J. Uncertainty Quantification **5**, 63–87 (2017)

[5] Adam, L., Branda, M.: Nonlinear chance constrained problems: Optimality conditions, regularization and solvers. J. Optim. Theory Appl. **170**(2), 419–436 (2016)

[6] Ahmed, S.: Convex relaxations of chance constrained optimization problems. Optimization Letters **8**(1), 1–12 (2014)

[7] Allevi, E., Bertocchi, M.I., Vespucci, M.T., Innorta, M.: A mixed integer nonlinear optimization model for gas sale company. Optimization Letters **1**(1), 61–69 (2007)

[8] Barrera, J., Homem-de Mello, T., Moreno, E., Pagnoncelli, B.K., Canessa, G.: Chance-constrained problems and rare events: an importance sampling approach. Mathematical Programming **157**(1), 153–189 (2016)

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

[10] Beraldi, P., Bruni, M.: An exact approach for solving integer problems under probabilistic constraints with random technology matrix. Annals of Operations Research **177**(1), 127–137 (2010)

[11] Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer (2011)

[12] Branda, M., Dupačová, J.: Approximation and contamination bounds for probabilistic programs. Annals of Operations Research **193**(1), 3–19 (2012)

[13] Chen, W., Sim, M., Sun, J., Teo, C.P.: From cvar to uncertainty set: Implications in joint chance-constrained optimization. Operations Research **58**(2), 470–485 (2010)

[14] Cheng, J., Lisser, A.: A second-order cone programming approach for linear programs with joint probabilistic constraints. Operations Research Letters **40**(5), 325–328 (2012)

[15] Déak, I.: Subroutines for computing normal probabilities of sets-computer experiences. Annals of Operations Research **100**, 103–122 (2000)

[16] Dentcheva, D., Martinez, G.: Augmented Lagrangian method for probabilistic optimization. Annals of Operations Research **200**(1), 109–130 (2012)

[17] Dentcheva, D., Martinez, G.: Regularization methods for optimization problems with probabilistic constraints. Mathematical Programming **138**(1-2), 223–251 (2013)

[18] Dentcheva, D., Prékopa, A., Ruszczyński, A.: Concavity and efficient points of discrete distributions in probabilistic programming. Mathematical Programming **89**(1), 55–77 (2000)

[19] Geletu, A., Hoffmann, A., Kläppel, M., Li, P.: An inner-outer approximation approach to chance constrained optimization. SIAM Journal on Optimization **27**(3), 1834–1857 (2017)

[20] Genz, A., Bretz, F.: Computation of multivariate normal and t-probabilities, *Lecture Notes in Statistics*, vol. 195. Springer, Heidelberg (2009)

[21] Gonzalez Grandon, T., Heitsch, H., Henrion, R.: A joint model of probabilistic /robust constraints for gas transport management in stationary networks. Computational Management Science **14**, 443–460 (2017)

[22] Gotzes, C., Heitsch, H., Henrion, R., Schultz, R.: On the quantification of nomination feasibility in stationary gas networks with random load. Mathematical Methods of Operations Research **84**(2), 427–457 (2016)

[23] Haneveld, W., van der Vlerk, M.: Integrated chance constraints: Reduced forms and an algorithm. Computational Management Science **3**(4), 245–269 (2006)

[24] Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Springer (2001)

[25] Henrion, R.: A critical note on empirical (Sample Average, Monte Carlo) approximation of solutions to chance constrained programs. In: D. Hömberg, F. Tröltzsch (eds.) System Modeling and Optimization, *IFIP Advances in Information and Communication Technology*, vol. 391, pp. 25–37. Springer Berlin Heidelberg (2013)

[26] Henrion, R., Möller, A.: A gradient formula for linear chance constraints under Gaussian distribution. Mathematics of Operations Research **37**(3), 475–488 (2012)

[27] Henrion, R., Römisch, W.: Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints. Mathematical Programming **100**(3), 589–611 (2004)

[28] Henrion, R., Strugarek, C.: Convexity of chance constraints with independent random variables. Computational Optimization and Applications **41**(2), 263–276 (2008)

[29] Ioffe, A.D., Outrata, J.: On metric and calmness qualification conditions in subdifferential calculus. Set-Valued Anal. **16**(2-3), 199–227 (2008)

[30] Kaňková, V.: On the convergence rate of empirical estimates in chance constrained stochastic programming. Kybernetika **26**(6), 449–461 (1990)

[31] Lejeune, M.: Pattern definition of the p-efficiency concept. Annals of Operations Research **200**(1), 23–36 (2012)

[32] Luedtke, J.: A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. Mathematical Programming **146**(1-2), 219–244 (2014)

[33] Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM Journal on Optimization **19**(2), 674–699 (2008)

[34] Luedtke, J., Ahmed, S., Nemhauser, G.: An integer programming approach for linear programs with probabilistic constraints. Mathematical Programming **122**(2), 247–272 (2010)

[35] Maggioni, F., Bertocchi, M., Giacometti, R., Vespucci, M.T., Innorta, M., Allevi, E.: A stochastic optimization model for gas retail with temperature scenarios and oil price parameters. IMA Journal of Management Mathematics **21**(2), 149–163 (2010)

[36] Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM Journal on Optimization **17**(4), 969–996 (2007)

[37] Prékopa, A.: Dual method for the solution of a one-stage stochastic programming problem with random RHS obeying a discrete probability distribution. Zeitschrift für Operations Research **34**(6), 441–461 (1990)

[38] Prékopa, A.: Stochastic programming. Kluwer, Dordrecht and Académiai Kiadó, Budapest (1995)

[39] Prékopa, A.: Probabilistic programming. In: A. Ruszczyński, A. Shapiro (eds.) Stochastic Programming, *Handbooks in Operations Research and Management Science*, vol. 10, pp. 267–351. Elsevier (2003)

[40] Rockafellar, R., Uryasev, S.: Optimization of conditional value-at-risk. Journal of Risk **2**, 21–41 (2000)

[41] Rockafellar, R., Uryasev, S.: Conditional value-at-risk for general loss distributions. Journal of Banking & Finance **26**(7), 1443–1471 (2002)

[42] Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer (1998)

[43] Royset, J., Polak, E.: Extensions of stochastic optimization results to problems with system failure probability functions. Journal of Optimization Theory and Applications **133**, 1–18 (2007)

[44] Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM J. Optim. **11**(4), 918–936 (2001)

[45] Shan, F., Zhang, L., Xiao, X.: A smoothing function approach to joint chance-constrained programs. Journal of Optimization Theory and Applications **163**(1), 181–199 (2014)

[46] Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming. MOS-SIAM Series on Optimization. SIAM–Society for Industrial and Applied Mathematics, Philadelphia (2009)

[47] Sun, H., Xu, H., Wang, Y.: Asymptotic analysis of sample average approximation for stochastic optimization problems with joint chance constraints via conditional value at risk and difference of convex functions. Journal of Optimization Theory and Applications **161**(1), 257–284 (2014)

[48] Wiecek, M.M., Dranichak, G.M.: Robust Multiobjective Optimization for Decision Making Under Uncertainty and Conflict, chap. 4, pp. 84–114 (2016)

[49] Wozabal, D., Hochreiter, R., Pflug, G.C.: A difference of convex formulation of value-at-risk constrained optimization. Optimization **59**(3), 377–400 (2010)

[50] Xie, W., Ahmed, S.: On quantile cuts and their closure for chance constrained optimization problems. Mathematical Programming (2017)

[51] Zhao, M., Huang, K., Zeng, B.: A polyhedral study on chance constrained program with random right-hand side. Mathematical Programming **166**(1), 19–64 (2017)

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