Solving joint chance constrained problems using regularization and Benders' decomposition

This paper is dedicated to Marida Bertocchi

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

[1]Southern University of Science and Technology, 1088 Xueyuan Ave, Shenzhen, Guangdong, China, 518055
[2]ÚTIA, Czech Academy of Sciences, Pod Vodárenskou věží 4, 18208, Prague, Czech Republic
[3]Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75, Prague, Czech Republic
[4]Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany

July 16, 2018

Abstract

We consider stochastic programs with joint chance constraints with discrete random distribution. We reformulate the problem by adding auxiliary variables. Since the resulting problem has a non-regular feasible set, we regularize it by increasing the feasible set. We solve the regularized problem by iteratively solving a master problem while adding Benders' cuts in a slave problem. Since the number of variables of the slave problem equals to the number of scenarios, we express its solution in a closed form. We show convergence properties of the solutions. On a gas network design problem, 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 the continuous distribution.

Keywords: Stochastic programming; Chance constrained programming; Optimality conditions; Regularization; Benders decomposition; Gas networks.
AMS classification: 90C15, 90C26, 49M05.

1 Introduction

In many real-world applications, the data is 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 [56]. 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 [5], 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

*adam@utia.cas.cz

example, an ambulance has to reach its patient within a certain time limit, no matter where the patient is, this approach is usually too restrictive.

On the other hand, for the stochastic optimization [7] the distribution of the random parameters is assumed to be known. 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 $\varepsilon \in (0,1)$ it is sufficient that the constraints are satisfied with probability $1-\varepsilon$. 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 [28, 34]. However, these estimates can be too conservative, cf. [23]. Recently, [4] 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. In [11], the authors locally approximated the feasible set and applied a trust region method to solve the chance constrained problem. The local approximation allowed to consider only a few active scenarios at each iteration.

For linear constraints and finite discrete distribution, strong results and algorithms based on cutting planes for mixed-integer reformulations are available, cf. [6, 33, 35]. Recently, [30, 60] derived new strong valid inequalities based on an intersection of multiple mixing sets for the chance constrained problems with a random right-hand side.

When the random parts of constraints are separated from the decision variables, we obtain the case with a 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 [38], which generalize the notation of quantiles to the multivariate case, see also [15, 31, 32] for recent results. By adopting a dual point of view, [51] developed a solution framework based on a recent generation of bundle methods.

Nonlinear programming algorithms were suggested for chance constrained problems by [40] and further developed by [13, 14]. Recently, [17] proposed a smooth approximation approach employing an inner and an outer analytic approximation of chance constraints leading to two classes of nonlinear programming problems. In [58] the authors 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 approximating the indicator function by a more tractable function. Approximation based on conditional value at risk has been deeply investigated by [42, 43, 49]. Similar ideas were used by [21] who employed the so-called integrated chance constraints. Bernstein approximation has been introduced by [37] for constraints affine in random coefficients and further developed by [2]. Recently, algorithmic approaches based on representation using a difference of convex (DC) functions appeared in the literature, see [47, 49, 57]. A second-order cone programming reformulation was obtained by [10] for problems with linear constraints under normally distributed random coefficient and under independence and copula dependence of the rows. For these linear-Gaussian problems, [24] 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 [39, 48], it was investigated and recently verified for general problems with high probability levels by [26, 50, 55]. A special attention has been paid to the stability of the optimal values and solution with respect to the changes of the probability distribution [8, 25]. Various approximations and worst-case bounds for distributionally robust chance constraints were derived by [9, 61].

Our new algorithmic approach relies on using the Benders' decomposition. It has already been considered in conjunction with chance constraints by [33, 52, 59]. The readers can be also referred to the survey

[41] for further insights.

In this paper we extend and combine our earlier work [1] and [20]. In [1] 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. Since standard constraint qualifications are not satisfied for the relaxed problem, we regularized it by increasing the feasible set and showed the convergence properties of solutions of the regularized problem. Finally, we applied this approach to a portfolio optimization problem.

In [20] we built on the results of M. Bertocchi [3, 36] on gas networks. 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 [19] via a simple projection method.

Our contribution in this paper is the following:

- We extend the results of [1] to problems with joint chance constraints by approximating the chance constrained problem by a regularized problem.
- Even when the data is convex, the feasible set of the chance constrained problem is only a union of convex sets and thus nonconvex. However, we show that under convex data stationary points are local minima. In Theorem 5 we propose a general result which also implies that strong stationary points for MPECs or MPCCs with convex data are local minima.
- Since the number of variables in the regularized problem equals to the sum of the number of original variables and the number of scenarios, it is usually intractable. To solve this issue, we propose a Benders' decomposition algorithm for the regularized problem. The algorithm consists of two parts: The outer part solves the master problem which depends only on the original variables. The inner part solves a slave problem which adds feasibility cuts to the master problem. Since the number of variables in the slave problem depends on the number of scenarios, we derive a closed-form solution for it.
- We propose a heuristic method for obtaining a fast high-quality solution. This is based on artificially requiring that the chance constraint is satisfied with a higher probability than the prescribed level.
- Since our method requires a discrete distribution, we empirically examine its convergence by increasing the number of scenarios. We compare it with the method from [20] which works directly with a continuous distribution. On a numerical example, we show that our method may perform better.
- The codes are available online.[1]

The paper is organized as follows: In Section 2 we derive theoretical results extending our work from [1] and we briefly summarize the results of [20]. Since the proof techniques are rather similar to those of [1], we moved the proofs to the Appendix. In Section 3 we propose the numerical method of alternatively solving the master problem and adding feasibility Benders' cuts via the slave problem. We show how to get an explicit solution for the slave problem and that the number of possible cuts is finite. In Section 4 we demonstrate that we obtain a high-quality solution on a problem with 10000 scenarios within a few minutes.

[1] `http://staff.utia.cas.cz/adam/research.html`

2 Methodology and Algorithms

The joint chance constrained problem may be formulated as follows:

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

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, a finite number of scenarios is sampled and the chance constraint is replaced by its discrete approximation. In the second one, we keep the continuous distribution. The scenario approach has the following advantages:

- Lower requirements on the random distribution. For example, it suffices to know only the empirical distribution whereas for the continuous approach, a special distributional assumption such as normality is often needed.
- Lower requirements on the data (e.g., in the application in Section 4 it allows to work with cycles).

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

2.2 Discrete (scenario) approach

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

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

where χ stands for the characteristic function which equals to 1 if $\max_k g_k(x,\xi_i)\le 0$ 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

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

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 the nonlinear programming

problem

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

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:

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

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

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. The reason for this is that in both cases a small perturbation in x will not change whether the inequality is satisfied or violated. For the scenarios where the random constraint is satisfied as the equality, we select an arbitrary scenario subset such that the prescribed probability level $1-\varepsilon$ is achieved.

Even though these problems are not equivalent, see [1, Example 2.1], there are close similarities between them, as stated in the next result.

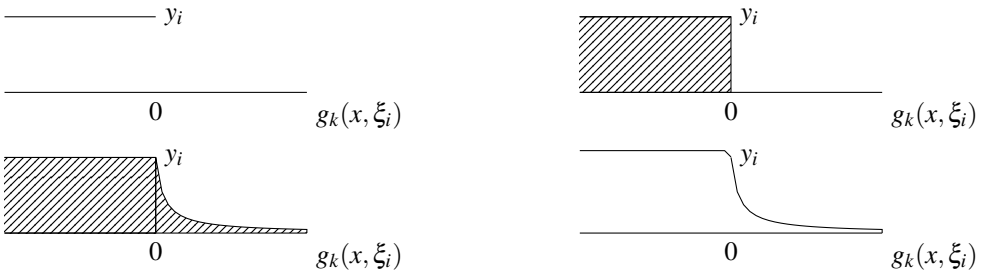

Figure 1: Feasible sets of problems (3) (top left), (4) (top right) and (8) (bottom left) and regularization ϕ_t from (10) (bottom right).

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

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

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

Assumption 1. 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 $I \in \mathscr{I}(\bar{x})$:

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

Finally, we derive the stationarity conditions for both problems.

Theorem 1. *Let $\bar{x}$ be a local minimum of problem* (2) *and let Assumption 1 be satisfied at $\bar{x}$. Then for every $I \in \mathscr{I}(\bar{x})$ there exist multipliers $\lambda_{ik},\ i \in I_0(\bar{x}),\ k \in K_0^i(\bar{x})$ and $\mu_j,\ j \in J_0(\bar{x})$ 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 1 be satisfied at $\bar{x}$, where for its second part we check system* (5) *only for $I = I_{0+}(\bar{x},\bar{y})$ and not for all $I \subset \mathscr{I}(\bar{x})$. Then there exist multipliers $\lambda_{ik},\ i \in I_0(\bar{x}),\ k \in K_0^i(\bar{x})$ and $\mu_j,\ j \in J_0(\bar{x})$ 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 1. *Consider a feasible point $\bar{x}$ of problem* (2) *and let Assumption 1 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, under convex data, stationary points are also local minima even though the problem is nonconvex. This follows directly from the general result in Theorem 5 at page 16 in the Appendix.

Theorem 2. *Let $f(\cdot)$, $g_k(\cdot,\xi_i)$ and $h_j(\cdot)$ be convex function for all i, j and k. 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 [1, 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 [46]. 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

$$\begin{array}{l} \underset{x,y}{\text{minimize}}\ f(x) \\ \text{subject to } p^\top y \geq 1 - \varepsilon, \\ \qquad 0 \leq y_i \leq 1,\ i = 1,\ldots,S, \\ \qquad y_i \leq \phi_t(g_k(x,\xi_i)),\ k = 1,\ldots,K,\ i = 1,\ldots,S, \\ \qquad h_j(x) \leq 0,\ j = 1,\ldots,J. \end{array} \tag{8}$$

Here, $\phi_t : \mathbb{R} \to \mathbb{R}$ is chosen in such a way that driving t to infinity ensures that y_i approaches zero for the violated constraint $g_k(x, \xi_i) > 0$. Namely, ϕ_t is a family of continuously differentiable decreasing functions which depend on the regularization parameter $t > 0$ and which satisfy the following properties:

$$\phi_t(0) = 1, \tag{9a}$$

$$\phi_t(z) > 0 \qquad \text{for } z \in \mathbb{R}, \tag{9b}$$

$$\phi_t(z^t) \to 0 \qquad \text{whenever } z^t \overset{t\to\infty}{\to} \bar{z} > 0 \text{ for some } \bar{z}, \tag{9c}$$

$$\frac{\phi_t'(z^t)}{\phi_t'(\tilde{z}^t)} \to 0 \qquad \text{whenever } \phi_t(z^t) \searrow 0 \text{ and } \phi_t(\tilde{z}^t) \to \bar{z} > 0 \text{ for some } \bar{z}. \tag{9d}$$

As an example of such regularizing function, we may consider

$$\phi_t(z) = \begin{cases} e^{-tz} & \text{if } z \geq 0, \quad (10a) \\ 1 - \dfrac{c}{(t+1)^2} \operatorname{arctg}\left(\dfrac{t(t+1)^2}{c} z\right) & \text{if } z < 0, \quad (10b) \end{cases}$$

where $c > 0$ is an arbitrary parameter. Note that form of ϕ_t on the negative numbers (10b) is not important for (8). Indeed, consider any scenario i and any k with $g_k(x, \xi_i) < 0$. Since $\phi_t(0) = 1$ and ϕ_t is decreasing by assumptions, this means that $\phi_t(g_k(x, \xi_i)) > 1$. But since the upper bound on y_i reads $y_i \leq \min\{1, \phi_t(g_k(x, \xi_i))\} = 1$, the exact form (10b) is not important. However, it 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 1 as certain boundedness of multipliers is needed.

Theorem 3. *Consider $(\bar{x}^t, \bar{y}^t)$ to be stationary points of problem* (8)*. Assume that the second part of Assumption 1 is satisfied at $\bar{x}$ and that $(\bar{x}^t, \bar{y}^t) \to (\bar{x}, \bar{y})$ for some $(\bar{x}, \bar{y})$ as $t \to \infty$. Then $(\bar{x}, \bar{y})$ is a stationary point of problem* (4)*. Moreover, if $f(\cdot)$, $g_k(\cdot, \xi_i)$ and $h_j(\cdot)$ are convex function for all i, j and k, it is even a local minimum of problem* (4)*.*

We summarize the previous results in Algorithm 2.1. Note that due to Theorem 3 the resulting point $(\bar{x}, \bar{y})$ is only a stationary point of the relaxed problem (4) and due to Corollary 1, $\bar{x}$ does not have to be a stationary point of the original problem (2). However, as the numerical experience in [1] suggests, it often happens that $\mathcal{I}(\bar{x})$ has only a single element and then $\bar{x}$ becomes the stationary point of the original problem (2) as well. Our conjecture is that this happens due to the regularizing properties of (8).

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** $\bar{x} = x^l$

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 sampling. It is based on the so-called *spheric-radial decomposition* of Gaussian random vectors [18] which has been successfully applied to chance constrained optimization problems [12, 19, 45].

Theorem 4. *Let ξ be a 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*

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

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

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

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 [53, 54], 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. In the numerical section, 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 intractable when the number of scenarios is large. In this section, we propose a method to eliminate y based on Benders' decomposition and derive a basic convergence analysis.

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, f is the same objective as in (8) and $v_b(x) \geq 0$ 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 $v_B(x) \leq 0$, which cuts away $\hat{x}$ from the feasible region of (11) but does not cut away any feasible point of (8), increase B and iterate. Thus the approximation stays an outer one. The feasibility Benders' cut is based on the feasible set of (8). For fixed $\hat{x}$ we define the linear slave problem

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

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

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

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 under the light shed by Lemma 1 below we construct the feasibility Benders' cut

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

add it to the master problem (11) and continue in the same way.

Lemma 1. *Assume that problem* (12) *does not have a feasible solution. Then cut* (14) *is valid, thus* $v_B(\hat{x}) > 0$ *and* $v_B(x) \le 0$ *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). Now first observe that $(\hat{u},\hat{v},\hat{w})$ is feasible for (13) and so $v_B(x) \le 0$, which concludes the proof. □

Even though (13) is a linear problem, it still depends on the number of scenarios S. We will 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 $\sum_{k=1}^{K} u_{ik} + vp_i + w_i \le 0$ 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

$$\begin{aligned} \underset{u_{ik},w_i}{\text{maximize}} \ & \sum_{k=1}^{K}\phi_t(g_k(\hat{x},\xi_i))u_{ik} + w_i \\ \text{subject to} \ & \sum_{k=1}^{K} u_{ik} + w_i \le -p_i, \\ & u_{ik} \le 0,\ \ w_i \le 0. \end{aligned}$$

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

$$\begin{aligned} \max_k g_k(\hat{x},\xi_i) > 0 &\implies \phi_t(g_{\hat{k}_i}(\hat{x},\xi_i)) < 1 \implies \hat{u}_{ik} = 0 \text{ for } k \ne \hat{k}_i,\ \hat{u}_{i\hat{k}_i} = -p_i,\ \hat{w}_i = 0, \\ \max_k g_k(\hat{x},\xi_i) \le 0 &\implies \phi_t(g_k(\hat{x},\xi_i)) \ge 1 \implies \hat{u}_{ik} = 0 \text{ for } k \ne \hat{k}_i,\ \hat{u}_{i\hat{k}_i} = 0,\ \hat{w}_i = -p_i, \end{aligned} \tag{15}$$

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

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

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

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

If (16) does not hold, $\hat{x}$ is optimal for (8). Note that the cut (17) is in an explicit form and depends only on $\hat{x}$.

3.2 Algorithm summary and convergence analysis

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 larger 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 $0 < t^1 < \cdots < t^L$

$B \leftarrow 1$
for $l = 1, \ldots, L$ **do**
  $t \leftarrow t^l$
  for $B = 1, 2, \ldots$ **do**
    find $\hat{x}^{lB}$ solving (11)
    if (16) is violated **or** termination criterion **then**
      break for
    else
      add cut v_B via (17)
    end if
  end for
  if $\hat{x}^{lB}$ is feasible for (2) **or** termination criterion is satisfied **then**
    break for
  end if
end for
return $\bar{x} = \hat{x}^{lB}$

Provided that L is finite, the number of loops in Algorithm 3.1 is finite, which follows from two facts. First, since cuts (17) depend only on the index sets $I(x)$, the number of possible cuts is finite. Second, the same cut cannot be generated twice. To see this, assume that first some x_1 generates a cut $v_1(x) \leq 0$ and later some x_2 generates the cut $v_2(x) \leq 0$ with $v_1 = v_2$. Due to Lemma 1 we have $v_2(x_2) > 0$. However, since x_2 is feasible for the regularized problem, we also have $v_1(x_2) \leq 0$. But this is a contradiction with $v_1 = v_2$. Thus, we have the finite convergence. Moreover, Lemma 5 in the Appendix states that $(\hat{u}, \hat{v}, \hat{w})$ derived in (15) is an extremal direction of (13).

Finally, from the discussion in Appendix C, 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 $I(\hat{x}) = \{1, \ldots, S\}$ 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 [47]. This cut is also close to the true feasible set which equals to

$$\sum_{i=1}^{S} p_i \max\{\phi_t(g_1(x, \xi_i)), 1\} \geq 1 - \varepsilon. \tag{18}$$

Thus, the feasibility cuts (17) seems to provide a good approximation of the feasible set of the regularized problem (8).

4 Numerical experiments

In this section, we show a good performance of our method on a gas network problem [20].

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 tree structured networks without cycles, it was shown in [20] that a random demand ξ can be satisfied if and only if

$$\begin{aligned} (p_0^{min})^2 &\le (p_k^{max})^2 + h_k(\xi),\ k = 1,\dots,|\mathscr{V}|, \\ (p_0^{max})^2 &\ge (p_k^{min})^2 + h_k(\xi),\ k = 1,\dots,|\mathscr{V}|, \\ (p_k^{max})^2 + h_k(\xi) &\ge (p_l^{min})^2 + h_l(\xi),\ k,l = 1,\dots,|\mathscr{V}|, \end{aligned} \tag{19}$$

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

$$\begin{aligned} &\underset{p^{max}}{\text{minimize}}\ c^\top p^{max} \\ &\text{subject to } \mathbb{P}(\text{system (19) is fulfilled}) \ge 1-\varepsilon. \end{aligned} \tag{20}$$

The minimal capacities p^{min} are usually considered to be fixed. Then we can reduce the number of inequalities in (19) to $(p_k^{max})^2 \ge \nu_k$, $k = 0,1,\dots,|\mathscr{V}|$, where we make use of $h_0(\xi) \equiv 0$ and where

$$\begin{aligned} \nu_0 &:= \max_{k=0,1,\dots,|\mathscr{V}|} \left\{ (p_k^{min})^2 + h_k(\xi) \right\}, \\ \nu_k &:= \nu_0 - h_k(\xi). \end{aligned} \tag{21}$$

Then problem (20) reduces to

$$\begin{aligned} &\underset{p^{max}}{\text{minimize}}\ c^\top p^{max} \\ &\text{subject to } \mathbb{P}\left((p_0^{max})^2 \ge \nu_0, \dots, (p_n^{max})^2 \ge \nu_n \right) \ge 1-\varepsilon. \end{aligned} \tag{22}$$

Note that the resulting regularized problem (8) has $n+1+S$ variables, $1+(n+1)S$ nonlinear constraints and additional $2S$ bound constraints.

4.2 Parameter description

In this section, we describe the choice of model parameters and their updates. According to [29, Chapter 13], random gas loads can often be described by combinations of Gaussian-like multivariate distributions (Gaussian, truncated Gaussian, log-normal). For simplicity, we shall assume here a multivariate Gaussian distribution as in Theorem 4 with mean μ and covariance matrix Σ. For both methods described in Section 2, we employed Quasi-Monte Carlo (QMC) sampling on the basis of Sobol sequences as a special case of low-discrepancy sequences that are included in the category of (t,m,d)-nets and (t,d) sequences [16].

For the application of our discrete (scenario) method, a QMC sample of 10000 scenarios was created according to the given distribution $\mathcal{N}(\mu,\Sigma)$ and processed as described below. For the continuous approach via spheric-radial decomposition presented in Section 2.3, a QMC sample of 100000 scenarios was created according to a standard Gaussian distribution (zero mean and identity covariance matrix). Normalizing each scenario to unit length then provides a sample of the uniform distribution on the sphere as required in the simultaneous update of values and gradients of the probability function (see the text below Theorem 4). The superiority of the QMC sampling (in particular when combined with the variance-reducing spheric-radial decomposition) over the crude Monte-Carlo of the given Gaussian distribution is highlighted in [20, p.454].

The probability level is chosen as 0.85, i.e., $\varepsilon = 0.15$. For the discrete approach, we used the 10000 QMC scenarios of the given Gaussian distribution as mentioned above. Then we applied Algorithm 3.1 to solve (20) with a fixed number of scenarios

$$S \in \{100, 300, 500, 1000, 3000, 5000, 10000\}$$

as many times as possible before exhausting the total sample. Hence, we solved the problem altogether 100 times for $S = 100$, 33 times for $S = 300$ and so on. For the presentation of results, we then averaged the results. For indicating a fairly precise value of the probability for the final solution (not for the discrete method itself), we employed the continuous method via spheric-radial decomposition (with 100000 QMC samples).

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 $t^{l+1} = 2t^l$. At every iteration we computed some (x^t, y^t). For all not-the-last t, we increased t whenever constraint $p^\top y \geq 1 - \varepsilon - 0.0005$ was satisfied or whenever we added 100 Benders' cuts. For the last $t = t^{20}$ we stop the algorithm whenever $p^\top y^t \geq 1 - \varepsilon$.

Moreover, at every iteration of Algorithm 3.1 we performed the heuristic cut reduction technique described in Appendix C. Since in (22) only the right-hand side is stochastic, it should be possible to use the scenario reduction technique called bundle preprocessing proposed by [32]. However, since our main goal is to test the performance of the above-described algorithms, we do not implement this interesting method. For algorithm details we refer to our codes available online.[1]

4.3 Numerical results

In this section, we solve problem (20) 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, small and medium, as depicted in Figure 2.

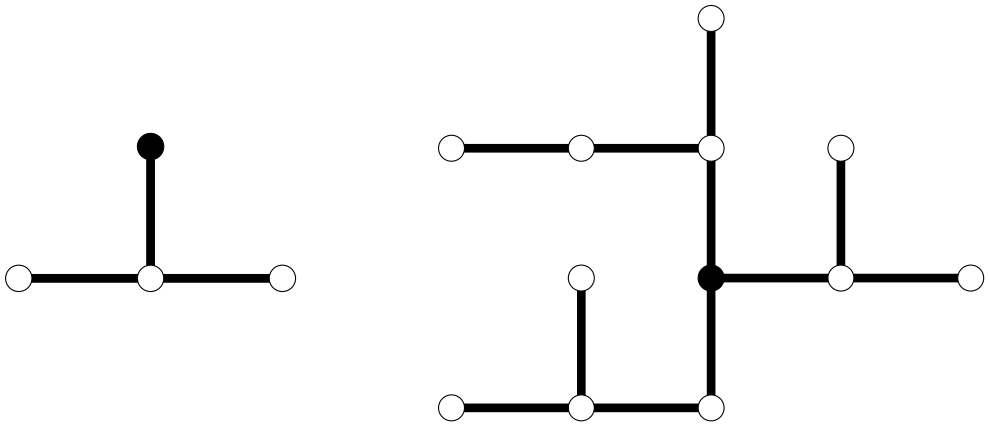

Figure 2: Network topology of small (left) and medium (right) size network examples. The entry and exit points of the considered gas transportation networks are displayed in black (entry) and white (exit), respectively.

Table 1 displays the results for both networks depicted in Figure 2. Every column corresponds to a given number of scenarios S. The rows depict the mean and standard deviation (SD) of the obtained optimal value and probability, the number of solves of the master problem (11) and the computational time. We present the reached true probability (computed for 100000 samples as described above) and the reached probability inside the sample (computed from S samples on which the optimization was run). This corresponds to the standard machine learning technique to dividing the data into training and testing sets [22].

		Number of scenarios S						
Network	Criterion	100	300	500	1000	3000	5000	10000
Small	**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 (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
	Time [s]	0.9	1.5	2.0	5.2	48.7	164.0	2868.2
Medium	**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 (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
	Time [s]	3.7	12.2	47.3	233.8	≈16h	≈1d	≈4d

Table 1: Results of Algorithm 3.1 for the small $n = K = 4$ and medium $n = K = 12$ networks depicted in Figure 2. The large computational times were fixed by a heuristic method presented in Table 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, the objective gets worse and the standard deviation decreases. This makes sense as the approximation of the true distribution gets better.
- The number of solves of (11) increases. We guess that cuts (14) provide a worse approximation of the feasible set for large S and thus the efficiency decreases.
- The computational time increases significantly. This is connected with the number of solves of (11) increases as well. Note that in every iteration problem (11) has the same number of variables but the number of constraints increases.

Since the computational time is infeasible for a large number of scenarios, for $S \geq 1000$ we employed a heuristic method to reduce it. Instead of $\varepsilon = 0.15$ all computation was performed with $\varepsilon = 0.14$ for the small network and with $\varepsilon = 0.13$ for the medium network. The only exception was the stopping criterion which was based on the original level $\varepsilon = 0.15$.

The results of this heuristic procedure are depicted in Table 2. For one out of the ten cases for $S = 1000$, the optimization failed. First note that it happened that the probability level in the sample is above 85% and that the true probability is higher than the probability in the sample. This did not happen for Algorithm 3.1 in Table 1. The reason is that while the designs presented in Table 1 are often at a local minimum, the results from Table 2 are only close to it. However, the new heuristic method is significantly faster and even for the medium network with $S = 10000$ scenarios, the required time was only around 14 minutes. This is a significant improvement from four days in Table 1.

Since all results have a different probability level, they are difficult to compare in terms of the objective value. In Figure 3 we depict the objective and the achieved probability level. The probability was computed on a QMC sample with 100000 samples. The diamond and the gray filled circles show the performance

Network	Criterion	Number of scenarios S			
		1000	3000	5000	10000
Small	**Mean objective**	740.1	739.7	741.2	739.4
	Mean probability (true) [%]	84.67	84.94	85.28	85.31
	Mean probability (sample) [%]	85.27	85.19	85.45	85.36
	SD objective	2.7	1.2	2.5	-
	SD probability (true) [%]	0.33	0.34	0.11	-
	Number of solves of (11)	91.2	94.0	98.0	94.0
	Time [s]	9.1	13.1	45.1	44.6
Medium	**Mean objective**	3148.1	3154.9	3166.2	3158.3
	Mean probability (true) [%]	84.27	84.74	85.07	85.30
	Mean probability (sample) [%]	85.23	85.24	85.27	85.56
	SD objective	16.4	3.4	1.0	-
	SD probability (true) [%]	0.61	0.42	0.00	-
	Number of solves of (11)	183.7	184.7	195.0	189.0
	Time [s]	135.9	277.7	482.3	820.7

Table 2: Results for the heuristic method based on Algorithm 3.1 for the small $n = K = 4$ and medium $n = K = 12$ networks depicted in Figure 2.

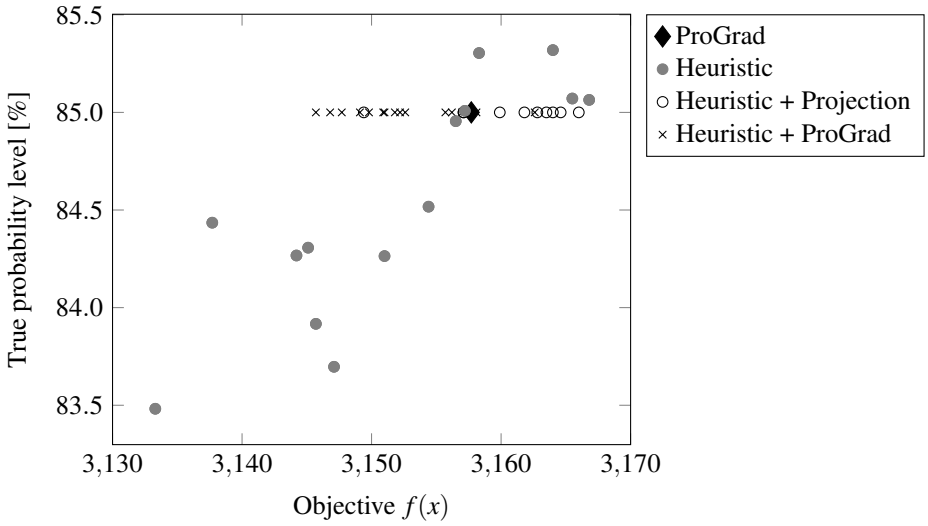

Figure 3: Selected results for the medium $n = K = 12$ network for $S \in \{1000, 3000, 5000, 10000\}$. The probabilities are computed outside of the sample on which the optimization was performed. The ProGrad method is applied to the problem with continuous Gaussian distribution. The heuristic method is a slight modification of Algorithm 3.1 and works with a finite number of samples of the same distribution. The heuristic method was also combined with a projection onto the feasible set under the Gaussian distribution and as a starting point for ProGrad.

of the ProGrad method and the heuristic method, respectively. The empty circles show the result of the heuristic method after being the projection onto the boundary of the feasible set of (1) governed by the continuous normal distribution. This projection was performed using the gradient information that is available by the spheric-radial approach, see Section 2.3. Finally, the crosses show the results when the final point of the heuristic method was used as the starting point for the ProGrad method.

For one sample the heuristic method obtained lower objective value while achieving higher probability than the ProGrad method. For the remaining samples, it either obtained a lower objective value or a

	Method				
Network	CCP (projection)	Heur (projection)	ProGrad	CCP + ProGrad	Heur + ProGrad
Small	738.56	738.51	738.40	738.40	738.40
Medium	3146.52	3149.35	3157.68	3145.47	3145.75

Table 3: Comparison of the methods described in Table 3. The CCP method refers to Algorithm 3.1.

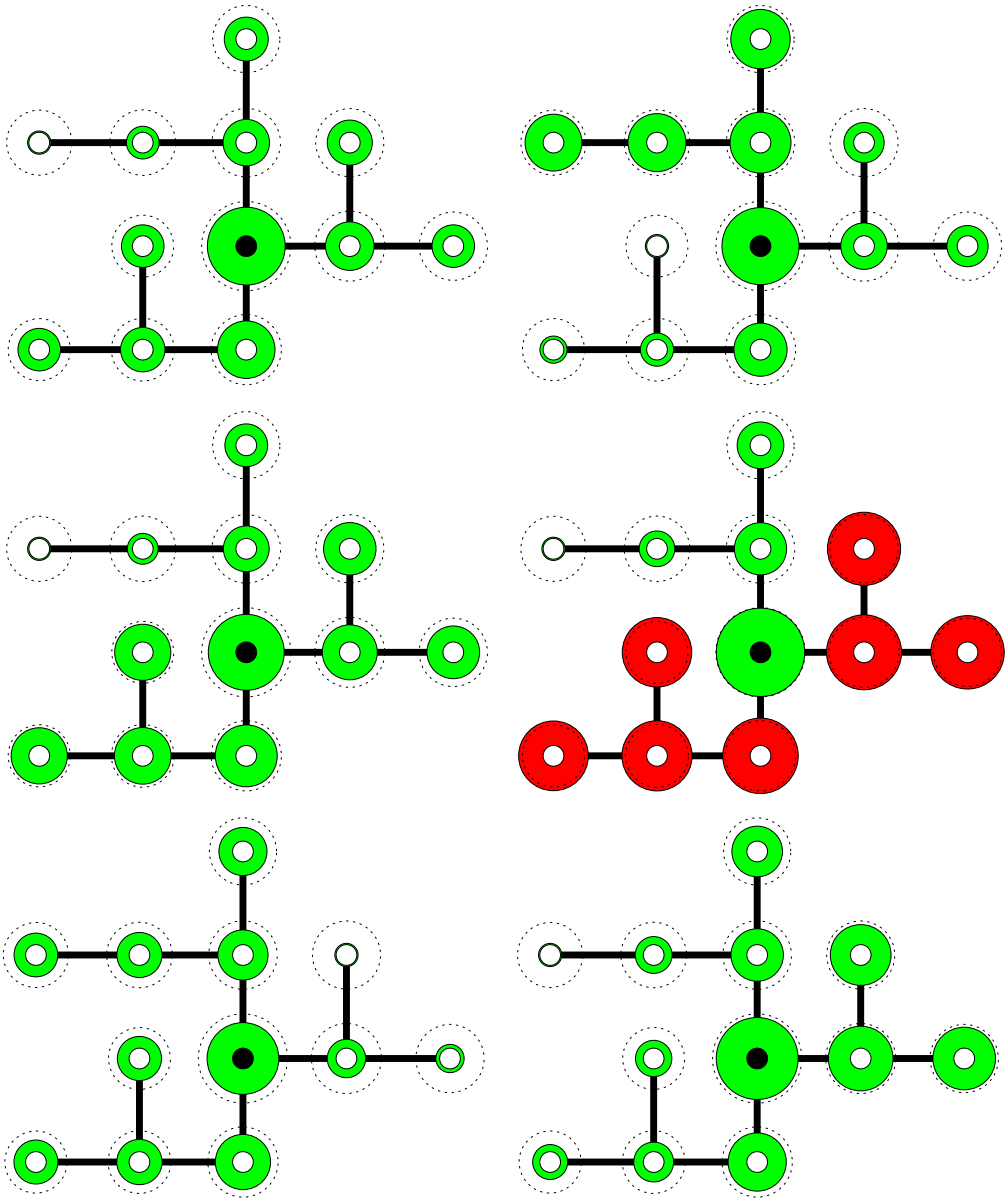

Figure 4: Six simulations of exit pressure realizations compared to the upper pressure bounds (dashed lines) provided by the numerical solution for the medium network. Feasible and infeasible pressure realizations at each node are shown in green and red, respectively.

higher probability level. The performance was further significantly improved when the point obtained by the heuristic method was used as a starting point for the ProGrad method. In Table 3 we show a similar

comparison for both networks. For all methods, we show the best obtained objective value. Note that this comparison is fair in the sense that all solutions have its probability level precisely 85%. Since the problem is highly nonconvex, we believe that the superior performance of our method is caused by the slow update in t which is able to avoid local minima, see [1, Example 3.1]. We would like to stress again that the running time for the heuristic method was in the order of minutes.

In Figure 4 we perform an 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 4, 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.

Conclusions

We have proposed a method for solving optimization problems with joint chance constraints where the random vector follows a discrete distribution. The proposed method is based on introducing binary variables, relaxing them into continuous variables, regularizing the resulted optimization problem and driving the regularization parameter to infinity. To handle a large number of scenarios, we transformed the regularized problem into a two-stage problem where the master problem is in the decision variables and the slave problem adds cuts to the master problem. Since the slave problem depends on the number of scenarios, we derived a closed-form solution for it. We proposed a heuristic method to significantly decrease the computational time. We compared our methods with the ProGrad method and showed their good performance on two instances of a gas network design problem.

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 5. *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*

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in X. \end{aligned} \tag{23}$$

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 a finite number of (possibly overlapping) convex sets, then $\bar{x}$ is a local minimum of problem (23)*.*

Proof. From the theorem statement, there are convex sets X_i, $i = 1, \dots, I$ 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 (23), we have

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

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 (23). □

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 [1]. □

Theorem 1. Denote the feasible set of problem (2) by Z and consider a point $\bar{x} \in Z$. Note that Z can be written as

$$Z = \bigcup_{I \subset \{1,\dots,S\}, \sum_{i \in I} p_i \geq 1-\varepsilon} \left\{ x : \begin{array}{c} g_k(x,\xi_i) \leq 0,\ i \in I,\ k = 1,\dots,K \\ h_j(x) \leq 0,\ j \in J \end{array} \right\}.$$

Then Z coincides locally around $\bar{x}$ with

$$\bigcup_{I \in \mathscr{I}(\bar{x})} Z_I := \bigcup_{I \in \mathscr{I}(\bar{x})} \left\{ x : \begin{array}{c} g_k(x,\xi_i) \leq 0,\ i \in I_0(\bar{x}) \cap I,\ k \in K_0^i(\bar{x}) \\ h_j(x) \leq 0,\ j \in J_0 \end{array} \right\}, \tag{24}$$

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 [44, 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 realize that Z_I can be due to (24) written as $F(x) \leq 0$ for some function F combining g_k and h_j and to use chain rule [44, Theorem 6.14].

The proof of the second part goes in a similar way. Due to [44, Theorem 6.12], the necessary optimality conditions for problem (4) read

$$0 \in \begin{pmatrix} \nabla f(\bar{x}) \\ 0 \end{pmatrix} + \hat{N}_Z(\bar{x},\bar{y}) \tag{25}$$

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

$$\begin{aligned} Z_I^x &:= \left\{ x : \begin{array}{c} g_k(x,\xi_i) \leq 0,\ i \in I \cup I_{0+}(\bar{x},\bar{y}),\ k \in K_0^i(\bar{x}) \\ h_j(x) \leq 0,\ j \in J_0 \end{array} \right\} \\ Z_I^y &:= \left\{ \begin{array}{l} p^\top y \geq 1-\varepsilon \\ y :\ y_i \in [0,1],\ i \in I \cup I_{0+}(\bar{x},\bar{y}) \cup \{i : \max_k\ g_k(\bar{x},\xi_i) < 0\} \\ \quad y_i = 0,\ i \in (I_{00}(\bar{x},\bar{y}) \setminus I) \cup \{i : \max_k\ g_k(\bar{x},\xi_i) > 0\} \end{array} \right\}. \end{aligned}$$

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 (25) is equivalent to

$$0 \in \nabla f(\bar{x}) + \bigcap_{I \subset I_{00}(\bar{x},\bar{y})} \hat{N}_{Z_I^x}(\bar{x}). \tag{26}$$

To finish the proof, it suffices to realize that the intersection in (26) is attained for $I = \emptyset$ and to use either [27, Proposition 3.4] (if the first part of Assumption 1 holds true) or [44, Theorem 6.14] (if the second part of Assumption 1 holds true). □

The proof of Theorem 3 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 first-order optimality conditions

$$0 = \nabla f(\bar{x}^t) - \sum_{i=1}^{S} \sum_{k=1}^{K} \gamma_{ik}^t \phi_t'(g_k(\bar{x}^t, \xi_i)) \nabla_x g_k(\bar{x}^t, \xi_i), \tag{27a}$$

$$0 = -\alpha^t p_i + \beta_i^t + \sum_{k=1}^{K} \gamma_{ik}^t, \; i = 1, \ldots, S \tag{27b}$$

and the complementarity conditions

$$\alpha^t (1 - \varepsilon - p^\top \bar{y}^t) = 0, \tag{28a}$$

$$\beta_i^t \begin{cases} \geq 0 & \text{if } \bar{y}_i^t = 1, \\ = 0 & \text{if } 0 < \bar{y}_i^t < 1, \\ \leq 0 & \text{if } \bar{y}_i^t = 0, \end{cases} \tag{28b}$$

$$\gamma_{ik}^t (\bar{y}_i^t - \phi_t(g_k(\bar{x}^t, \xi_i))) = 0 \tag{28c}$$

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

Lemma 2. *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;*

Proof. If $g_k(\bar{x}^t, \xi_i) < 0$, then from the definition of ϕ_t we have $\phi_t(g_k(\bar{x}^t, \xi_i)) > 1$. Since $\bar{y}_i^t \in [0, 1]$, this directly implies $\gamma_{ik}^t = 0$ due to (28c). For the second part, consider the case of $\alpha^t = 0$. Then due to (27b) we have $\beta_i^t + \sum_{k=1}^{K} \gamma_{ik}^t = 0$ for all i. Assume that $\beta_i^t < 0$. Then there exists some k such that $\gamma_{ik}^t > 0$. From (28b) we further get $\bar{y}_i^t = 0$, and thus $\bar{y}_i^t - \phi_t(g_k(\bar{x}^t, \xi_i)) < 0$. But this is a contradiction with (28c) and thus we have $\beta_i^t \geq 0$. But this together with $\gamma_{ik}^t \geq 0$ and $\beta_i + \gamma_{ik}^t = 0$ implies the second statement. □

In the following text, by $\bar{y}_i^t \searrow 0$ we understand that the sequence $\bar{y}_i^t$ is positive and converges monotonically to 0.

Lemma 3. *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*

$$\frac{-\gamma_{ik}^t \phi_t'(g_k(\bar{x}^t, \xi_i))}{-\sum_{\tilde{k}=1}^{K} \gamma_{j\tilde{k}}^t \phi_t'(g_{\tilde{k}}(\bar{x}^t, \xi_j))} \to 0. \tag{29}$$

Proof. Since $\bar{y}_i^t \searrow 0$ and since $p^\top \bar{y} = 1 - \varepsilon$, 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 by (28b). If $\gamma_{ik}^t = 0$, then the proof is finished. In the opposite case of $\gamma_{ik}^t > 0$, we realize that $\alpha^t > 0$ due to Lemma 2. Since $\gamma_{ik}^t \geq 0$ and

$\phi_t'(g_k(\bar{x}^t,\xi_i))<0$, we deduce

$$\begin{aligned}0\le\frac{-\gamma_{ik}^t\phi_t'(g_k(\bar{x}^t,\xi_i))}{-\sum_{\tilde{k}=1}^K\gamma_{j\tilde{k}}^t\phi_t'(g_{\tilde{k}}(\bar{x}^t,\xi_j))}&\le\frac{-\left(\sum_{\tilde{k}=1}^K\gamma_{i\tilde{k}}^t\right)\phi_t'(g_k(\bar{x}^t,\xi_i))}{-\sum_{\tilde{k}=1}^K\gamma_{j\tilde{k}}^t\phi_t'(g_{\tilde{k}}(\bar{x}^t,\xi_j))}\\&\le\frac{-\sum_{\tilde{k}=1}^K\gamma_{i\tilde{k}}^t}{-\sum_{\tilde{k}=1}^K\gamma_{j\tilde{k}}^t}\frac{\phi_t'(g_k(\bar{x}^t,\xi_i))}{\phi_t'(g_{\hat{k}^t}(\bar{x}^t,\xi_j))}\\&=\frac{-\alpha^t p_i}{-\alpha^t p_j}\frac{\phi_t'(g_k(\bar{x}^t,\xi_i))}{\phi_t'(g_{\hat{k}^t}(\bar{x}^t,\xi_j))}\to 0\end{aligned}\tag{30}$$

where $\hat{k}^t:=\arg\min_{\tilde{k}}-\phi_t'(g_{\tilde{k}}(\bar{x}^t,\xi_j)$, the last equality follows from (27b) as well as $\beta_i^t=\beta_j^t=0$ and the convergence follows from assumption (9d), for which we realize that $\phi_t(g_{\hat{k}^t}(\bar{x}^t,\xi_j))\ge\bar{y}_j^t$, 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 3. We will show first that $(\bar{x},\bar{y})$ is a feasible point of (4). Due to continuity, it is sufficient to show that $g_k(\bar{x},\xi_i)\bar{y}_i\le 0$. Since this relation is obvious whenever $g_k(\bar{x},\xi_i)\le 0$ for all k, we consider scenario i with $g_k(\bar{x},\xi_i)>0$ for some k. But then $g_k(\bar{x}^t,\xi_i)>0$ for sufficiently large t and thus $0\le\bar{y}_i^t\le\phi_t(g_k(\bar{x}^t,\xi_i))$. But since $g_k(\bar{x}^t,\xi_i)\to g_k(\bar{x},\xi_i)>0$, assumption (9c) implies that $\bar{y}_i=0$, and thus $(\bar{x},\bar{y})$ is a feasible point of problem (4).

Define now

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

where the nonnegativity follows from the property that ϕ_t is decreasing. If $g_k(\bar{x},\xi_i)<0$, then for all sufficiently large t we have $g_k(\bar{x}^t,\xi_i)<0$ and due to Lemma 2 we deduce $\gamma_{ik}^t=0$ and subsequently $\lambda_{ik}^t=0$. Then for sufficiently large t, optimality condition (27a) reads

$$\begin{aligned}0&=\nabla f(\bar{x}^t)+\sum_{i=1}^S\sum_{k=1}^K\lambda_{ik}^t\nabla_x g_k(\bar{x}^t,\xi_i)\\&=\nabla f(\bar{x}^t)+\sum_{\{(i,k):g_k(\bar{x},\xi_i)=0,\,\bar{y}_i>0\}}\lambda_{ik}^t\nabla_x g_k(\bar{x}^t,\xi_i)+\sum_{\{(i,k):g_k(\bar{x},\xi_i)\ge 0,\,\bar{y}_i=0\}}\lambda_{ik}^t\nabla_x g_k(\bar{x}^t,\xi_i).\end{aligned}\tag{31}$$

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

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

$$\lambda_{max}^t:=\max_{i=1,\dots,S}\sum_{k=1}^K\lambda_{ik}^t\to\infty.\tag{32}$$

Then dividing equation (31) by λ_{max}^t yields

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

When taking the limit with respect to $t\to\infty$, the first term vanishes. Consider now the third term. If $p^\top\bar{y}^t>1-\varepsilon$, from Lemma 2 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 $\phi_t(g_k(\bar{x}^t,\xi_i))>\bar{y}_i^t$, then from (28c) we have $\lambda_{ik}^t=0$. In the opposite case, we may use Lemma 3 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

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

Since $\frac{\lambda_{ik}^t}{\lambda_{max}^t}\in[0,1]$ and the numerators sum to λ_{max}^t due to (32) and the discussion above, at least one of these fractions converges to a positive number. However, the existence of such positive limit contradicts Assumption 1 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 (31), 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 2 and 3. Thus $(\bar{x},\bar{y})$ is indeed a stationary point of problem (4). □

C Cut reduction

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

Lemma 4. *Consider two points $\hat{x}^1,\hat{x}^2$ such that for all i we have*

$$\max_k g_k(\hat{x}^2,\xi_i) > 0 \implies \max_k g_k(\hat{x}^1,\xi_i) > 0,\ \arg\max_k g_k(\hat{x}^1,\xi_i) = \arg\max_k g_k(\hat{x}^2,\xi_i). \tag{34}$$

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

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

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

Then for any x we have

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

Proof. Due to (34) 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 (34) 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 the following heuristic 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 (34) is satisfied. Alternatively, we remove all previous cuts, for which (34) is violated only for a small number of scenarios.

Finally, the next lemma states that the added cuts in (17) 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 5. *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

$$(\hat{u},\hat{v},\hat{w}) = \kappa(u^1,v^1,w^1) + (1-\kappa)(u^2,v^2,w^2). \tag{35}$$

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 the first equation in (15). Define

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

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 (35) 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 (15). □

References

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

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

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

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

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

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

[7] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2011.

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

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

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

[11] F. Curtis, A. Wächter, and V. Zavala. A sequential algorithm for solving nonlinear optimization problems with chance constraints. *SIAM Journal on Optimization*, 28(1):930–958, 2018.

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

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

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

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

[16] J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, New York, NY, USA, 2010.

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

[18] A. Genz and F. Bretz. *Computation of multivariate normal and t-probabilities*, volume 195 of *Lecture Notes in Statistics*. Springer, Heidelberg, 2009.

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

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

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

[22] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

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

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

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

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

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

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

[29] T. Koch, B. Hiller, M. E. Pfetsch, and L. Schewe, editors. *Evaluating gas network capacities*, volume 21 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015.

[30] S. Küçükyavuz. On mixing sets arising in chance-constrained programming. *Mathematical Programming*, 132(1):31–56, Apr 2012.

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

[32] M. Lejeune and N. Noyan. Mathematical programming approaches for generating p-efficient points. *European Journal of Operational Research*, 207(2):590 – 600, 2010.

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

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

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

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

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

[38] A. Prékopa. 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.

[39] A. Prékopa. *Stochastic programming*. Kluwer, Dordrecht and Académiai Kiadó, Budapest, 1995.

[40] A. Prékopa. Probabilistic programming. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267–351. Elsevier, 2003.

[41] R. Rahmaniani, T. G. Crainic, M. Gendreau, and W. Rei. The benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801 – 817, 2017.

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

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

[44] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, 1998.

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

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

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

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

[49] H. Sun, H. Xu, and Y. Wang. 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.

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

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

[52] W. van Ackooij, A. Frangioni, and W. de Oliveira. Inexact stabilized benders' decomposition approaches with application to chance-constrained problems with finite support. *Computational Optimization and Applications*, 65(3):637–669, Dec 2016.

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

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

[55] W. van Ackooij and J. Malick. Eventual convexity of probability constraints with elliptical distributions. *Mathematical Programming*, Jan 2018.

[56] M. M. Wiecek and G. M. Dranichak. *Robust Multiobjective Optimization for Decision Making Under Uncertainty and Conflict*, chapter 4, pages 84–114. 2016.

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

[58] W. Xie and S. Ahmed. On quantile cuts and their closure for chance constrained optimization problems. *Mathematical Programming*, Sep 2017.

[59] Y. Zhang, J. Wang, B. Zeng, and Z. Hu. Chance-constrained two-stage unit commitment under uncertain load and wind power output using bilinear benders decomposition. *IEEE Transactions on Power Systems*, 32(5):3637–3647, Sept 2017.

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

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