Annual Reviews in Control 47 (2019) 341–363

Contents lists available at ScienceDirect

Annual Reviews in Control

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

Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization

Xinbo Geng*, Le Xie

Texas A&M University, College Station, TX, USA

A R T I C L E I N F O

Article history:
Received 1 February 2019
Revised 29 April 2019
Accepted 7 May 2019
Available online 24 May 2019

Keywords:
Data-driven
Power system
Chance constraint
Probabilistic constraint
Stochastic programming
Robust optimization
Chance-constrained optimization

A B S T R A C T

Uncertainties from deepening penetration of renewable energy resources have posed critical challenges to the secure and reliable operations of future electric grids. Among various approaches for decision making in uncertain environments, this paper focuses on chance-constrained optimization, which provides explicit probabilistic guarantees on the feasibility of optimal solutions. Although quite a few methods have been proposed to solve chance-constrained optimization problems, there is a lack of comprehensive review and comparative analysis of the proposed methods. We first review three categories of existing methods to chance-constrained optimization: (1) scenario approach; (2) sample average approximation; and (3) robust optimization based methods. Data-driven methods, which are not constrained by any particular distributions of the underlying uncertainties, are of particular interest. Key results of the analytical reformulation approach for specific distributions are briefly discussed. We then provide a comprehensive review on the applications of chance-constrained optimization in power systems. Finally, this paper provides a critical comparison of existing methods based on numerical simulations, which are conducted on standard power system test cases.

Contents

* Corresponding author.
E-mail address: xbgeng@tamu.edu (X. Geng).

https://doi.org/10.1016/j.arcontrol.2019.05.005

1. Introduction

Real-time decision making in the presence of uncertainties is a classical problem that arises in many contexts. In the context of electric energy systems, a pivotal challenge is how to operate a power grid with an increasing amount of supply and demand uncertainties. The unique characteristics of such operational problem include (1) the underlying distribution of uncertainties is largely unknown (e.g. the forecast error of demand response); (2) decisions have to be made in a timely manner (e.g. a dispatch order needs to be given by 5 minutes prior to the real-time); and (3) there is a strong desire to know the risk that the system is exposed to after a decision is made (e.g. the risk of violating transmission constraints after the real-time market clears). In response to these challenges, a class of optimization problems named “chance-constrained optimization” has received increasing attention in both operations research and practical engineering communities.

The objective of this article is to provide a comprehensive and up-to-date review of mathematical formulations, computational algorithms, and engineering implications of chance-constrained optimization in the context of electric power systems. In particular, this paper focuses on the data-driven approaches to solving chance-constrained optimization without knowing the underlying distribution of uncertainties. This paper also briefly mentions some critical results of an alternative approach, which derives equivalent forms of chance-constrained optimization problems for specific distributions. A more general class of problems, i.e. distributionally robust optimization or ambiguous chance constraint, is beyond the scope of this paper.

1.1. An overview of chance-constrained optimization

Chance-constrained optimization (CCO) is an important tool for decision making in uncertain environments. Since its birth in 1950s, CCO has found many successful applications in various fields, e.g. economics (Yaari, 1965), control theory (Calafiore, Campi et al., 2006), chemical process (Henrion et al., 2001; Sahinidis, 2004), water management (Dupačová, Gaivoronski, Kos, & Szantai, 1991) and recently in machine learning (Ben-Tal, Bhadra, Bhattacharyya, & Saketha Nath, 2011; Ben-Tal, El Ghaoui, & Nemirovski, 2009; Caramanis, Mannor, & Xu, 2012; Gabrel, Murat, & Thiele, 2014; Sra, Nowozin, & Wright, 2012; Xu, Caramanis, & Mannor, 2009). Chance-constrained optimization plays a particularly important role in the context of electric power systems (Ozturk, Mazumdar, & Norman, 2004; Wang, Guan, & Wang, 2012), applications of CCO can be found in various time-scales of power system operations and at different levels of the system.

The first chance-constrained program was formulated in Charnes, Cooper, and Symonds (1958), then was extensively studied in the following 50 years, e.g. Charnes and Cooper (1959), Charnes and Cooper (1963), Kataoka (1963), Pintér (1989), Sen (1992), Prekopa, Vizvari, and Badics (1998), Ruszczynski and Shapiro (2003), Ben-Tal et al. (2009) and Prékopa (1995). Previously, most methods to solve CCO problems deal with specific families of distributions, such as log-concave distributions (Miller & Wagner, 1965; Prékopa, 1995). Many novel methods appeared in the past ten years, e.g. scenario approach (Calafiore et al., 2006), sample average approximation (Luedtke & Ahmed, 2008; Ruszczyński, 2002) and convex approximation (Nemirovski & Shapiro, 2006). Most of them are generic methods that are not limited to specific distribution families and require very limited knowledge about the uncertainties. In spite of many successful applications of these methods in various fields, there is a lack of comprehensive review and a critical comparison.

1.2. Contributions of this paper

The main contributions of this paper are threefold:

1. We provide a detailed tutorial on the existing algorithms to solve chance-constrained programs and a survey of major theoretical results. To the best of our knowledge, there is no such review available in the literature;
2. We provide a comprehensive review on the applications of chance-constrained optimization in power systems, with focus on various interpretations of chance constraints in the context of power engineering.
3. We implement most of the reviewed methods and develop an open-source Matlab toolbox (ConvertChanceConstraint), which is available on Github.[1] We also provide a critical comparison of existing methods based numerical simulations on IEEE standard test systems.

1.3. Organization of this paper

The remainder of this paper is organized as follows. Section 2 introduces chance-constrained optimization. Section 3 summarizes the fundamental properties of chance-constrained optimization problems. An overview of how to solve chance-constrained optimization problems is described in Section 4, which outlines Sections 5–7. Three major approaches to solving chance-constrained optimization (scenario approach, sample average approximation and robust optimization based methods) are presented in Sections 5–7, respectively. Section 8 provides a comprehensive review on applications of CCO in power systems. The structure and usage of the Toolbox *ConvertChanceConstraint* is in Section 9. Section 9 also conducts numerical simulations and compares existing approaches to solving CCO problems. Concluding remarks are in Section 10.

1.4. Notations

The notations in this paper are standard. All vectors and matrices are in the real field $\mathbf{R}$. Sets are in calligraphy fonts, e.g. $\mathcal{S}$. The upper and lower bounds of a variable x are denoted by $\bar{x}$ and $\underline{x}$. The estimation of a random variable ϵ is $\hat{\epsilon}$. We use $\mathbf{1}_n$ to denote an all-one vector in $\mathbf{R}^n$, the subscript n is sometimes omitted for simplicity. The absolute value of vector x is $|x|$, and the cardinality of a set $\mathcal{S}$ is $|\mathcal{S}|$. Function $[a]_+$ returns the positive part of variable a. The indicator function $\mathbb{1}_{x>0}$ is one if $x>0$. The floor function $\lfloor a \rfloor$ returns the largest integer less than or equal to the real number a. The ceiling function $\lceil a \rceil$ returns the smallest integer greater than or equal to a. $\mathbb{E}[\xi]$ is the expectation of a random vector ξ, $\mathbb{V}(x)$ denotes the violation probability of a candidate solution x, and $\mathbb{P}_\xi(\cdot)$ is the probability taken with respect to ξ. The transpose of a vector a is $a^\intercal$. Infimum, supremum and essential supremum are denoted by inf, sup and ess sup. The element-wise multiplication of the same-size vectors a and b is denoted by $a \circ b$.

[1] github.com/xb00dx/ConvertChanceConstraint-ccc.

2. Chance-constrained optimization

2.1. Introduction

We study the following chance-constrained optimization problem throughout this paper:

$$\text{(CCO)}: \ \min_x \ c^\intercal x \tag{1a}$$

$$\text{s.t.} \ \mathbb{P}_\xi\Big(f(x,\xi) \le 0 \Big) \ge 1-\epsilon \tag{1b}$$

$$x \in \mathcal{X} \tag{1c}$$

where $x \in \mathbf{R}^n$ is the decision variable and random vector $\xi \in \mathbf{R}^d$ is the source of uncertainties. Without loss of generality,[2] we assume the objective function is linear in x and does not depend on ξ. Constraint (1b) is the *chance constraint* (or *probabilistic constraint*), it requires the inner constraint $f(x, \xi) \le 0$ to be satisfied with high probability $1-\epsilon$. The inner constraint $f(x, \xi)$: $\mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}^m$ consists of m individual constraints, i.e. $f(x,\xi) = \big(f_1(x,\xi), f_2(x,\xi), \cdots, f_m(x,\xi)\big)$. Set $\mathcal{X}$ represent the deterministic constraints. Parameter ϵ is called the *violation probability* of (CCO). Notice that $f(x, \xi)$ is random due to the randomness of ξ, the probability $\mathbb{P}$ is taken with respect to ξ. Sometimes the probability is denoted by $\mathbb{P}_\xi$ to avoid confusion.

It is worth mentioning that CCO is closely related with the theory of risk management. For example, an individual chance constraint $\mathbb{P}(f_i(x,\xi) \le 0) \ge 1-\epsilon_i$ can be equivalently interpreted as a constraint on the value at risk $\text{VaR}(f_i(x,\xi); 1-\epsilon_i) \le 0$. This connection can be directly seen from the definition.

Definition 1 (Value at Risk)**.** Value at risk (VaR) of random variable ζ at level $1-\epsilon$ is defined as

$$\text{VaR}(\zeta; 1-\epsilon) := \inf\Big\{\gamma : \mathbb{P}(\zeta \le \gamma) \ge 1-\epsilon\Big\} \tag{2}$$

More details about this can be found in Section 7.3.1, (Chen, Sim, Sun, & Teo, 2010; Rockafellar & Uryasev, 2000) and references therein.

CCO is closely related with two other major tools for decision making with uncertainties: stochastic programming and robust optimization. The idea of sample average approximation, which originated from stochastic programming, can be applied on chance-constrained programs (Section 6). Section 7 demonstrates the connection between robust optimization and CCO.

2.2. Joint and individual chance constraints

Constraint (1b) is called a *joint chance constraint* because of its multiple inner constraints (Miller & Wagner, 1965), i.e.

$$\mathbb{P}\Big(f_1(x,\xi) \le 0, f_2(x,\xi) \le 0, \cdots, f_m(x,\xi) \le 0 \Big) \ge 1-\epsilon \tag{3}$$

Alternatively, each one of the following m constraints is called an *individual chance constraint*:

$$\mathbb{P}\Big(f_i(x,\xi) \le 0 \Big) \le 1-\epsilon_i, \ i = 1,2,\cdots,m \tag{4}$$

Joint chance constraints typically have more modeling power since an individual chance constraint is a special case ($m=1$) of a joint

[2] Using the epigraph formulation as mentioned in Campi, Garatti, and Prandini (2009) and Boyd and Vandenberghe (2004).

chance constraint. But individual chance constraints are relatively easier to deal with (see Sections 3.2 and 7.3). There are several ways to convert individual and joint chance constraints between each other.

First, a joint chance constraint can be written as a set of individual chance constraints using Bonferroni inequality or Boole's inequality. Notice (3) can be represented as

$$\mathbb{P}_\xi\left(\cup_{i=1}^m \left\{f_i(x,\xi)\geq 0\right\}\right) \leq \epsilon. \tag{5}$$

Since $\mathbb{P}_\xi(\cup_{i=1}^m\{f_i(x,\xi)\geq 0\}) \leq \sum_{i=1}^m \mathbb{P}_\xi(\{f_i(x,\xi)\geq 0\})$, if $\sum_{i=1}^m \epsilon_i \leq \epsilon$, then any feasible solution to (4) is also feasible to (3). In other words, (4) is a *safe approximation* (see Definition 11) to (3) when $\sum_{i=1}^m \epsilon_i \leq \epsilon$. With appropriate $\{\epsilon_i\}_{i=1}^m$, (4) could be a good approximation of (3). However, it is usually difficult to find such $\{\epsilon_i\}_{i=1}^m$. Some other issues of this approach are discussed in Section 7.4.1.

Alternatively, a joint chance constraint (3) is equivalent to the following individual chance constraint:

$$\mathbb{P}_\xi\left(\overline{f}(x,\xi)\leq 0\right) \geq 1-\epsilon \tag{6}$$

where $\overline{f}(x,\xi): \mathbf{R}^n\times\mathbf{R}^d\to\mathbf{R}$ is the pointwise maximum of functions $\{f_i(x,\xi)\}_{i=1}^m$ over x and ξ, i.e.

$$\overline{f}(x,\xi) := \max\left\{f_1(x,\xi), f_2(x,\xi), \cdots, f_m(x,\xi)\right\}. \tag{7}$$

It is worth noting that converting $\{f_i(x,\xi)\}_{i=1}^m$ to $\overline{f}(x,\xi)$ could lose nice structures of the original constraint $f(x,\xi)\leq 0$ and cause more difficulties.

In this paper, we focus on the chance-constrained optimization problems with a *joint* chance constraint.

2.3. Critical definitions and assumptions

Theoretical results in the following sections are based on the critical definitions and assumptions below.

Definition 2 (Violation Probability)**.** Let $x^\diamond$ denote a candidate solution to (CCO), its violation probability is defined as

$$\mathbb{V}(x^\diamond) := \mathbb{P}_\xi\left(f(x^\diamond,\xi)\geq 0\right) \tag{8}$$

Definition 3. $x^\diamond$ is a *feasible* solution to (CCO) if $x^\diamond\in\mathcal{X}$ and $\mathbb{V}(x^\diamond)\leq\epsilon$. Let $\mathcal{F}_\epsilon$ denote the set of feasible solutions to the chance constraint (1b),

$$\mathcal{F}_\epsilon := \{x\in\mathbf{R}^n : \mathbb{V}(x)\leq\epsilon\} = \{x\in\mathbf{R}^n : \mathbb{P}_\xi\left(f(x,\xi)\leq 0\right)\geq 1-\epsilon\},$$

then $x^\diamond$ is *feasible* to (CCO) if $x^\diamond\in\mathcal{X}\cap\mathcal{F}_\epsilon$.

Although (CCO) seeks optimal solutions under uncertainties, it is a *deterministic* optimization problem. To better see this, (CCO) can be equivalently written as $\min_{x\in\mathcal{X}}\ c^\intercal x$, s.t. $\mathbb{V}(x)\leq\epsilon$ or $\min_{x\in\mathcal{X}\cap\mathcal{F}_\epsilon} c^\intercal x$.

Definition 4. Let $o^\star$ denote the optimal objective value of (CCO). For simplicity, we define $o^\star=+\infty$ when (CCO) is infeasible and $o^\star=-\infty$ when (CCO) is unbounded. Let $x^\star$ denote the optimal solution to (CCO) if exists, and $o^\star=c^\intercal x^\star$.

Definition 5. We say a candidate solution $x^\diamond$ is *conservative* if $\mathbb{V}(x^\diamond)\ll\epsilon$ or $c^\intercal x^\diamond\gg o^\star$.

Most existing theoretical results on (CCO) are built upon the following two assumptions.

Assumption 1. Let Ξ denote the support of the random variable ξ, the distribution $\xi\sim\Xi$ exists and is fixed.

Assumption 1 only assumes the existence of an underlying distribution, but we do not necessarily need to know it to solve (CCO). Removing Assumption 1 leads to a more general class of problem named *distributionally robust optimization* or *ambiguous chance constraints*. Section 3.4 discusses cases with Assumption 1 removed.

Assumption 2. (1) Function $f(x,\xi)$ is convex in x for every instance of ξ, and (2) the deterministic constraints define a convex set $\mathcal{X}$.

The convexity assumption above makes it possible to develop theories on (CCO). However, the feasible region $\mathcal{F}_\epsilon$ of (CCO) is often non-convex even under Assumption 2. More details are presented in Sections 3.1 and 3.2.

3. Fundamental properties

3.1. Hardness

Although CCO is an important and useful tool for decision making under uncertainties, it is very difficult to solve in general. Major difficulties come from two aspects:

- **(D1)** It is difficult to check the feasibility of a candidate solution $x^\diamond$. Namely, it is intractable to evaluate the probability $\mathbb{P}_\xi\left(f(x^\diamond,\xi)\leq 0\right)$ with high accuracy. More specifically, calculating the probability involves multivariate integration, which is NP-Hard (Khachiyan, 1989). The only general method might be Monte-Carlo simulation, but it can be computationally intractable due to the curse of dimensionality.
- **(D2)** It is difficult to find the optimal solution $x^\star$ and $o^\star$ to (CCO). Even with the convexity assumption (Assumption 2), the feasible region $\mathcal{F}_\epsilon$ of (CCO) is often non-convex except a few special cases. For example, Section 3.3 shows the feasible region of (CCO) with separable chance constraints is a union of cones, which is non-convex in general. Although researchers have proved various sufficient conditions on the convexity of (CCO), it remains challenging to solve (CCO) because of difficulty (D1). Most of times, however, we are agnostic about the properties of the feasible region $\mathcal{F}_\epsilon$.

Despite that fact that Assumptions 1 and 2 largely simplify the problem and make theoretical analysis on (CCO) possible, (D1) and (D2) still exist and pose great challenges to solve (CCO).

Theorem 1 (Luedtke, Ahmed, & Nemhauser, 2010; Qiu, Ahmed, Dey, & Wolsey, 2014)**.** *(CCO) is strongly NP-Hard.*

Theorem 2 (Ahmed, 2018)**.** *Unless* P = NP, *it is impossible to obtain a polynomial time algorithm for (CCO) with a constant approximation ratio.*

Theorem 1 formalizes the hardness results of solving (CCO), Theorem 2 further demonstrates that it is also difficult to obtain approximate solutions to (CCO): any polynomial algorithm is not able to find a solution x^* (with $o^*=c^\intercal x^*$) such that $|o^*/o^\star|$ is bounded by a constant C from above. In other words, any polynomial-time algorithm could be arbitrarily worse.

3.2. Special cases

Although (CCO) is NP-Hard to solve in general, there are several special cases in which solving (CCO) is relatively easy. The most well-known special case is (9), which was first proved in (Kataoka, 1963).

$$\min_{x\in\mathcal{X}}\ c^\intercal x \tag{9a}$$

$$\text{s.t. } \mathbb{P}(a^\intercal x + b^\intercal\xi + \xi^\intercal Dx \leq e)\geq 1-\epsilon \tag{9b}$$

Parameters $a\in\mathbf{R}^n$, $b\in\mathbf{R}^d$, $D\in\mathbf{R}^{d\times n}$ and $e\in\mathbf{R}$ are fixed coefficients. $\xi\sim\mathcal{N}(\mu,\Sigma)$ is a multivariate Gaussian random vector with mean μ and covariance Σ. Notice that (9b) is an *individual* chance constraint with multivariate Gaussian coefficients. Let $\Phi(\cdot)^{-1}$ denote the inverse cumulative distribution function (CDF) function of a standard normal distribution. It is easy to show that if $\epsilon \le 1/2$, (9) is equivalent to (10), which is a second order cone program (SOCP) and can be solved efficiently.

$$\min_{x \in \mathcal{X}} c^\intercal x \tag{10a}$$

$$\text{s.t. } e - b^\intercal \mu - (a + D^\intercal \mu)^\intercal x \ge \Phi^{-1}(1-\epsilon)\sqrt{(b+Dx)^\intercal \Sigma (b+Dx)} \tag{10b}$$

(10) also shows the possibility of deriving equivalent reformulations of (CCO), many analytical methods to solve chance-constrained optimization are built on this observation.

The case of log-concave distribution (Prékopa, 1971, 1995; Prékopa, Yoda, & Subasi, 2011) is another famous special case where chance constraint is convex. There are many other sufficient conditions on the convexity of chance constraints, e.g. Lagoa (1999), Calafiore and El Ghaoui (2006), Henrion and Strugarek (2008), Henrion and Strugarek (2011) and Van Ackooij (2015).

3.3. Feasible region

A chance-constrained program with only right hand side uncertainties (11) is considered in this section. With this example, we provide deeper understandings on the non-convexity of (CCO).

$$\min_{x \in \mathcal{X}} c^\intercal x \tag{11a}$$

$$\text{s.t. } \mathbb{P}(f(x) \le \zeta) \ge p \tag{11b}$$

In (11b), the inner function $f(x)$: $\mathbf{R}^n \to \mathbf{R}^m$ is deterministic. The only uncertainty is the right-hand side value, represented by a random vector $\zeta \in \mathbf{R}^m$. Chance constraints like (11b) are also named *separable* chance constraints (or probabilistic constraints) since the deterministic and random parts are separated. We replace $1-\epsilon$ with p in (11b) to follow the convention in the existing literature.

Definition 6 (p-efficient points (Shapiro, Dentcheva, & Ruszczyński, 2009)). Let $p \in (0, 1)$, a point $v \in \mathbf{R}^m$ is called a *p-efficient point* of the probability function $\mathbb{P}_\zeta(\zeta \le z)$, if $\mathbb{P}_\zeta(\zeta \le v) \ge p$ and there is no $z \le v$, and $z \ne v$ such that $\mathbb{P}_\zeta(\zeta \le z) \ge p$.

Theorem 3 (Prékopa, 1995; Shapiro et al., 2009). *Let $\mathcal{E}$ be the index set of p-efficient points v^i, $i \in \mathcal{E}$. Let $\mathcal{F}_p := \{x \in \mathbf{R}^n : \mathbb{P}_\zeta(f(x) \le \zeta) \ge p\}$ denote the feasible region of (11b), then it holds that*

$$\mathcal{F}_p = \cup_{i \in \mathcal{E}} K_i \tag{12}$$

where each cone K_i is defined as $K_i := v^i + \mathbf{R}^m_+$, $i \in \mathcal{E}$.

Theorem 3 shows the geometric properties of (CCO). The finite union of convex sets need not to be convex, therefore the feasible region of (CCO) is generally non-convex.

Remark 1. Many methods to solve (CCO) (e.g. Beraldi & Ruszczyński, 2002; Kress, Penn, & Polukarov, 2007; Prekopa et al., 1998) start with a partial or complete enumeration of p-efficient points. However, the number of p-efficient points could be astronomic or even infinite. See Shapiro et al. (2009) and Prékopa (1995) and references therein for the finiteness results of p-efficient points and complete theories and algorithms on p-efficient points.

3.4. Ambiguous chance constraints

Ambiguous chance constraint is a generalization of chance constraints,

$$\mathbb{P}_{\xi \sim P}\Big(f(x,\xi) \le 0\Big) \ge 1-\epsilon, \ \forall P \in \mathcal{P}. \tag{13}$$

It requires the inner chance constraint $f(x, \xi) \le 0$ holds with probability $1-\epsilon$ for any distribution P belonging to a set of pre-defined distributions $\mathcal{P}$.

Ambiguous chance constraints are particularly useful in the cases where only partial knowledge on the distribution P is available, e.g. we know only that P belongs a given family of $\mathcal{P}$. However, it is generally more difficult to solve ambiguous chance constraints, and the theoretical results rely on different assumptions of uncertainties. This paper only reviews solutions to CCO, studies on ambiguous chance constraints are beyond the scope of this paper.

4. An overview of solutions to CCO

This paper concentrates on solutions to (CCO) with the following properties: (i) dealing with both difficulties (D1) and (D2) mentioned in Section 3.1; (ii) utilizing information from data (only) without making suspicious assumptions on the distribution of uncertainties; and (iii) possessing rigorous guarantees on the feasibility and optimality of the returned solutions. Sections 4.1–4.3 explain these three properties in detail. Section 4.4 provides an overview of methods with the properties above.

4.1. Classification of solutions

Existing methods on (CCO) can be roughly classified into four categories (Ahmed & Shapiro, 2008):

- **(C1)** When both difficulties (D1) and (D2) in Section 3.1 are absent, (CCO) is convex and the probability $\mathbb{P}(f(x,\xi) \le 0)$ is easy to calculate. The only known case in this category is the individual chance constraint (9) with Gaussian distributions, which might be the only special case of (CCO) that can be easily solved;
- **(C2)** When (D1) is absent but (D2) is present, it is relatively easy to calculate $\mathbb{P}(f(x,\xi) \le 0)$ (e.g. finite distributions with not too many realizations). As shown in Theorem 3, the feasible region of (CCO) could be non-convex and solutions typically rely on integer programming and global optimization (Ahmed & Shapiro, 2008);
- **(C3)** When (D1) is present but (D2) is absent, (CCO) is proved to be convex but remains difficult to solve because of the difficulty (D1) in calculating probabilities. This case often requires approximating the probability via simulations or specific assumptions. All examples mentioned in Section 3.2 except (9) belong to this category.
- **(C4)** When both difficulties (D1) and (D2) are present, it is almost impossible to find the optimal solution $x^\star$ and $o^\star$. All existing methods attempt to obtain approximate solutions or suboptimal solutions and construct upper and lower bounds on the true objective value $o^\star$ of (CCO).

Methods associated with (C1)-(C3) are briefly mentioned in Section 3, the remaining part of this paper presents more general and powerful methods in category (C4).

4.2. Prior knowledge

In order to solve (CCO), a reasonable amount of prior knowledge on the underlying distribution $\xi \sim \Xi$ is necessary. Fig. 1 illustrates three categories of prior knowledge:

- **(K1)** We know the exact distribution $\xi \sim \Xi$ thus have *complete knowledge* on the underlying distribution;
- **(K2)** We know partially on the distribution (e.g. multivariate Gaussian distribution with bounded mean and variance) and thus have *partial knowledge*;

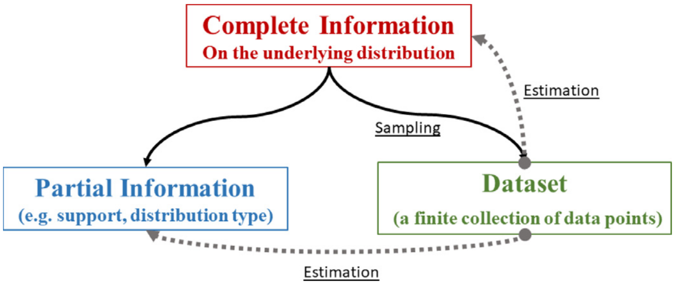


Fig. 1. Different knowledge levels to solve (CCO).

(K3) We have a finite dataset $\{\xi^i\}_{i=1}^N$, this is another case of *partial knowledge*.

It can be seen that prior information in (K2) is a strict subset of (K1), also by sampling we can construct a dataset in (K3) from the exact distribution in (K1). It seems (K1) is the best starting point to solve (CCO). However, *probability distributions are not known in practice, they are just models of reality and exist only in our imagination.* What exists in reality is *data*. Therefore (K3) might be the most practical case and becomes the focus of this paper. Almost all the data-driven methods to solve (CCO) are based on the following assumption.

Assumption 3. The samples (scenarios) ξ^i $(i = 1, 2, \cdots, N)$ in the *dataset* $\{\xi^i\}_{i=1}^N$ are independent and identically distributed (i.i.d.).

4.3. Theoretical guarantees

This paper concentrates on the theoretical aspects of the reviewed methods. In particular, we pay special attention to *feasibility guarantees* and *optimality guarantees*.

Given a candidate solution $x^\diamond$ to (CCO), the first and possibly most important thing is to check its *feasibility*, i.e. if $\mathbb{V}(x^\diamond) \le \epsilon$. Although (D1) demonstrates the difficulty in calculating $\mathbb{V}(x^\diamond)$ with high accuracy, there are various feasibility guarantees that either estimate $\mathbb{V}(x^\diamond)$ or provide upper bound on $\mathbb{V}(x^\diamond)$. The feasibility results can be classified into two categories: a-priori and a-posteriori guarantees. The *a-priori* ones typically provide prior conditions on (CCO) and the dataset $\{\xi^i\}_{i=1}^N$, the feasibility of the corresponding solution $x^\diamond$ is guaranteed *before* obtaining $x^\diamond$. Examples of this type include Corollary 1, Theorems 6,13 and 11. As the name suggests, the *a-posteriori* guarantees make effects *after* obtaining $x^\diamond$. The a-posteriori guarantees are constructed based on the observations of the structural features associated with $x^\diamond$. Examples include Theorem 7 and Proposition 1.

Given a candidate solution $x^\diamond$ and the associated objective value $o^\diamond = c^\intercal x^\diamond$, another important question to be answered is about the *optimality* gap $|o^\diamond - o^\star|$. Although finding $o^\star$ is often an impossible mission because of difficulty (D2), bounding from below on $o^\star$ is relatively easier. Sections 5.5 and 6.4 dedicate to algorithms of constructing lower bounds $\underline{o} \le o^\star$.

4.4. A schematic overview

A schematic overview of solutions to (CCO) and their relationships are presented in Fig. 2. Akin methods are plotted in similar colors, and links among two circles indicate the connection of the two methods. The tree-like structure of Fig. 2 illustrates the hierarchical relationship of the reviewed methods. Key references of each method are also provided. The root node of Fig. 2 is the "ambiguous chance constraint" or distributionally robust optimization (DRO), which is the parent node of "chance-constrained optimization". This indicates that DRO contains CCO as a special case. Similarly, for example, node "scenario approach" has three child nodes "prior", "posterior" and "sampling and discarding", this indicates the scenario approach has three major variations.

As shown in Fig. 2, CCO is a special case of ambiguous chance constraints where the set of distributions $\mathcal{P}$ is a singleton (Section 3.4). Therefore methods to solve ambiguous chance constraints can be applied on chance constraints as well. The methods and algorithms to solve CCO are the main focus of this paper, we will briefly mention the connection if some methods are related with ambiguous chance constraints.

Fig. 2 also outlines the first half of this paper, which dedicates to a review and tutorial on chance-constrained optimization. We summarize key results on the basic properties (Section 3), three main approaches to solving chance-constrained optimization problems, scenario approach (Section 5), sample average approximation (Section 6) and robust optimization (RO) based methods (Section 7).

5. Scenario approach

5.1. Introduction to the scenario approach

Scenario approach utilizes a dataset with N scenarios $\{\xi^i\}_{i=1}^N$ to approximate the chance-constrained program (1) and obtains the following *scenario problem* $(\mathrm{SP})_N$:

$$(\mathrm{SP})_N: \min_{x\in\mathcal{X}} c^\intercal x \tag{14a}$$

$$\text{s.t. } f(x,\xi^1) \le 0, \cdots, f(x,\xi^N) \le 0 \tag{14b}$$

SP_N seeks the optimal solution x_N^* which is feasible for all N scenarios. The scenario approach is a very simple yet powerful method. The most attractive feature of the scenario approach is its generality. It requires nothing except the convexity of constraints $f(x, \xi)$ and $\mathcal{X}$. It is purely data-driven and makes no assumption on the underlying distribution.

Remark 2. SP_N is a random program. Both its optimal objective value o_N^* and optimal solution x_N^* depend on the random samples $\{\xi^i\}_{i=1}^N$, therefore they are random variables. In consequence, $\mathbb{V}(x_N^*)$ is also a random variable. Let $\mathcal{N} := \{1, 2, \cdots, N\}$ denote the index set of scenarios. The optimal objective value of SP_N is denoted by $o^*(\mathcal{N})$ to emphasize its dependence on the random samples.

Theoretical results of the scenario approach are built upon the following assumption in addition to Assumptions 1–3.

Assumption 4 (Feasibility and Uniqueness (Campi & Garatti, 2008))**.** Every scenario problem $(\mathrm{SP})_N$ is feasible, and its feasibility region has a non-empty interior. Moreover, the optimal solution x_N^* of $(\mathrm{SP})_N$ exists and is unique.

If there exist multiple optimal solutions, the tie-break rules in Calafiore and Campi (2005) can be applied to obtain a unique solution.

Remark 3. (Sample Complexity N). We first provide some intuition on the scenario approach. When solving $(\mathrm{SP})_N$ with a very large number of scenarios, the solution x_N^* will be robust to almost every realization of ξ, thus the violation probability goes to zero. Although x_N^* is a feasible solution to (CCO) as $N \to +\infty$, it is overly conservative because $\mathbb{V}(x^*) \approx 0 \ll \epsilon$. On the other hand, using too few scenarios for SP_N might result in infeasible solutions x_N^* to (CCO). Notice that N is the only tuning parameter in the scenario approach, the most important question in the scenario approach theory is: *what is the right sample complexity N?* Namely, what is the smallest N such that $\mathbb{V}(x_N^*) \le \epsilon$ (with high probability)? Rigorous answers to the sample complexity question are built upon the structural properties of the scenario problem SP_N.

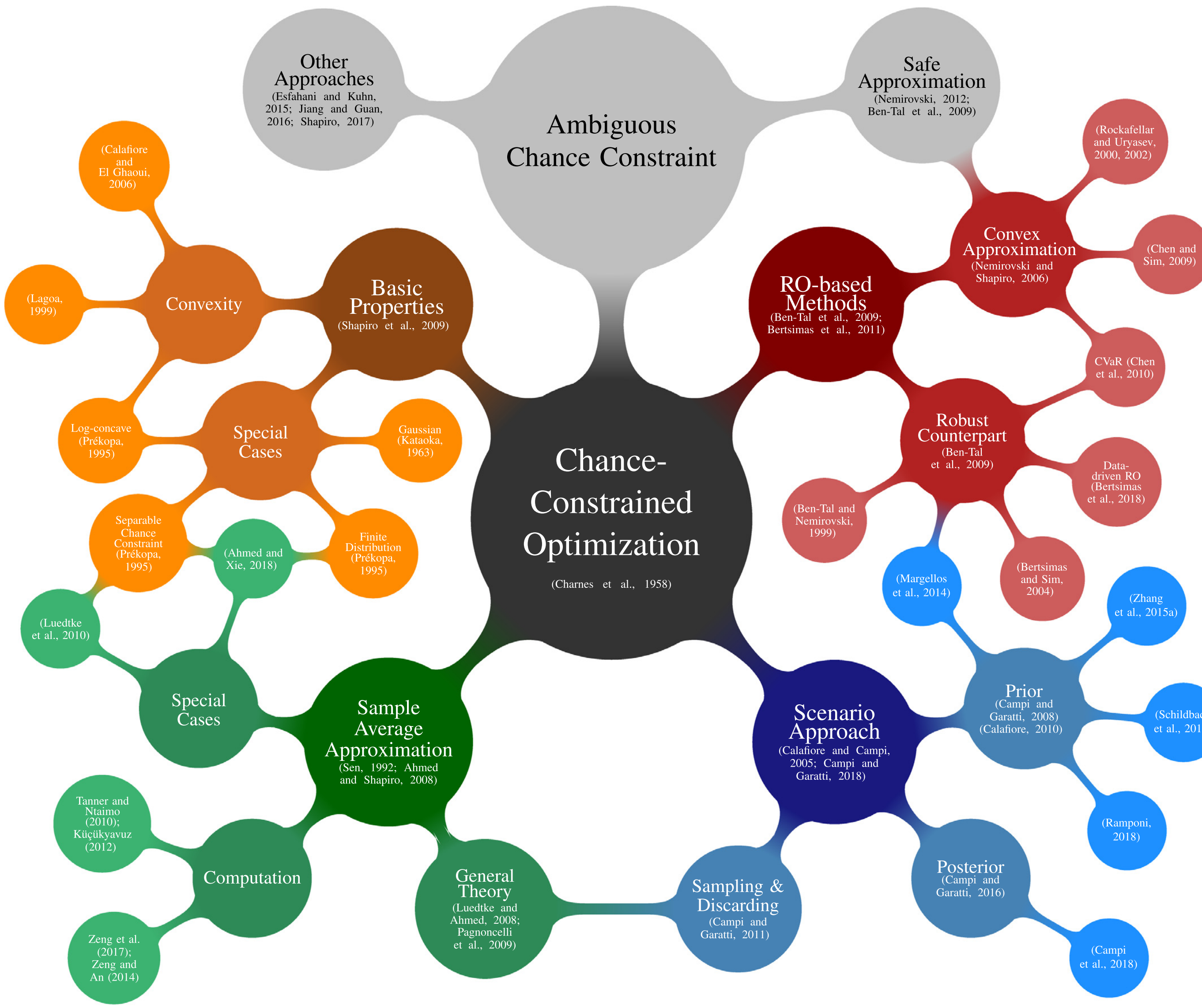


Fig. 2. A schematic overview of existing methods and algorithms to solve chance-constrained optimization problems.

5.2. Structural properties of the scenario problem

Among N scenarios in the dataset $\{\xi^i\}_{i=1}^N$, there are some important scenarios having direct impacts on the optimal solution x_N^*.

Definition 7 (Support Scenario (Calafiore & Campi, 2005)). Scenario ξ^i is a *support scenario* for $(\text{SP})_N$ if its removal changes the solution of $(\text{SP})_N$. The set of *support scenarios* of (SP_N) is denoted by $\mathcal{S}$.

Theorem 4 (Calafiore & Campi, 2005; Calafiore, 2010). *Under Assumption 2, the number of support scenarios in* SP_N *is at most* n, *i.e.* $|\mathcal{S}| \le n$.

Theorem 4 is built upon Helly's theorem and Radon's theorem (Rockafellar, 2015) in convex analysis. For non-convex problems, the number of support scenarios could be greater than the number of decision variables n. An example for non-convex problems is provided in Campi, Garatti, and Ramponi (2018).

Definition 8 (Fully-supported Problem (Campi & Garatti, 2008)). A scenario problem SP_N with $N \ge n$ is *fully-supported* if the number of support scenarios is exactly n. Scenario problems with $|\mathcal{S}| < n$ are referred as *non-fully-supported* problems.

Definition 9 (Non-degenerate Problem (Calafiore, 2010; Campi & Garatti, 2008)). Problem SP_N is said to be *non-degenerate*, if $o^*(\mathcal{N}) = o^*(\mathcal{S})$. In other words, SP_N is *non-degenerate* if the solution of $(\text{SP})_N$ with all scenarios in place coincides with the solution to the program with only the support scenarios are kept.

5.3. A-priori feasibility guarantees

Obtaining a-priori feasibility guarantees on the solution x_N^* to SP_N typically involves the following three steps:

1. Exploring the problem structure of SP_N and obtain an upper bound $\bar{h}$ on the number of support scenarios;
2. Choosing a good sample complexity $N(\epsilon, \beta, \bar{h})$ using Corollary 1, Theorem 6 or Remark 4;
3. Solving the scenario problem SP_N and obtain x_N^* and o_N^*.

Theorem 5 (Campi & Garatti, 2008)**.** *Under Assumptions 1–3, for a non-degenerate problem* SP_N*, it holds that*

$$\mathbb{P}^N\left(\mathbb{V}(x_N^*) > \epsilon\right) \le \sum_{i=1}^{n-1}\binom{N}{i}\epsilon^i(1-\epsilon)^{N-i}. \tag{15}$$

The probability $\mathbb{P}^N$ *is taken with respect to N random samples* $\{\xi^i\}_{i=1}^N$*, and the inequality is tight for fully-supported problems.*

As mentioned in Remark 2, $\mathbb{V}(x_N^*)$ is a random variable, its randomness comes from drawing scenarios $\{\xi^i\}_{i=1}^N$. For fully-supported problems, Theorem 5 shows the *exact* probability distribution of the violation probability $\mathbb{V}(x_N^*)$, i.e.

$$\mathbb{P}^N\left(\mathbb{V}(x_N^*) > \epsilon\right) = \sum_{i=1}^{n-1}\binom{N}{i}\epsilon^i(1-\epsilon)^{N-i}, \tag{16}$$

the tail of a binomial distribution. We could use Theorem 5 to answer the sample complexity question in Remark 3.

Corollary 1 (Campi & Garatti, 2008)**.** *Given a violation probability* $\epsilon \in (0, 1)$ *and a confidence parameter* $\beta \in (0, 1)$*, if we choose the number of scenarios N (the smallest such N is denoted by* N_{2008}*) such that*

$$\sum_{i=0}^{n-1}\binom{N}{i}\epsilon^i(1-\epsilon)^{N-i} \le \beta \tag{17}$$

Let x_N^* *denote the optimal solution to* SP_N*, it holds that*

$$\mathbb{P}^N\left(\mathbb{V}(x_N^*) \le \epsilon\right) \ge 1-\beta \tag{18}$$

In other words, the optimal solution x_N^* *is a feasible solution to (CCO) with probability at least* $1-\beta$*.*

Remark 2 states that the scenario approach is a randomized algorithm. Thus it is possible that the scenarios $\{\xi^i\}_{i=1}^N$ are drawn from a "bad" set and lead to infeasible solutions x_N^*, i.e. $\mathbb{V}(x_N^*) > \epsilon$. The *confidence parameter* β denotes the risk of failure associated to the randomized solution algorithm (Calafiore et al., 2006), and it bounds the probability that x_N^* is infeasible.

For fully-supported problems, N_{2008} is the tightest upper bound on sample complexity, which cannot be improved. For non-fully supported problems, it turns out N_{2008} can be further tightened. An improved sample complexity bound is provided in Theorem 6 based on the definition of Helly's dimension.

Definition 10 (Helly's Dimension (Calafiore, 2010))**.** *Helly's dimension* of SP_N is the smallest integer h such that

$$\operatorname{ess\,sup}_{\xi\in\Xi^N}|\mathcal{S}(\xi)| \le h$$

holds for any finite $N \ge 1$. The essential supremum is denoted by ess sup. We emphasize the dependence of support scenarios $\mathcal{S}$ on ξ by $\mathcal{S}(\xi)$.

Theorem 6 (Calafiore, 2010)**.** *Let h denote the Helly's dimension for* SP_N*, under Assumptions 1–3, for a non-degenerate problem* SP_N*, it holds that*

$$\mathbb{P}^N\left(\mathbb{V}(x_N^*) > \epsilon\right) \le \sum_{i=0}^{h-1}\binom{N}{i}\epsilon^i(1-\epsilon)^{N-i} \tag{19}$$

Equivalently, for a fixed confidence parameter $\beta \in (0, 1)$*, if the sample complexity N satisfies*

$$\sum_{i=0}^{h-1}\binom{N}{i}\epsilon^i(1-\epsilon)^{N-i} \le \beta \tag{20}$$

then the following probabilistic guarantee holds

$$\mathbb{P}^N\left(\mathbb{V}(x_N^*) > \epsilon\right) \le \beta \tag{21}$$

The only difference between Theorems 6 and 5 (and Corollary 1) is replacing n with Helly's dimension h in (19) and (20). Unfortunately, Helly's dimension is often difficult to calculate, while finding upper bounds $\bar{h}$ on Helly's dimension is usually a much easier task. Similarly we can replace h by $\bar{h}$ in (19) and (20), the same theoretical guarantees still hold because of the monotonicity of (19) and (20) in N and h. The support-rank defined in Schildbach, Fagiano, and Morari (2013) is an upper bound on Helly's dimension, some other upper bounds can be obtained by exploiting the structural properties of the problem, e.g. Zhang, Grammatico, Schildbach, Goulart, and Lygeros (2015).

Remark 4 (Sample Complexity Revisited)**.** A binary search type algorithm could be used to find N_{2008}. And a looser but handy upper bound is provided in (Campi et al., 2009):

$$N_{2009} := \frac{2}{\epsilon}\left(\ln(\frac{1}{\beta}) + n\right) \tag{22}$$

Notice n in (22) can be replaced by h or $\bar{h}$.

5.4. A-posteriori feasibility guarantees

When the desired violation probability ϵ is very small, the sample complexity of the a-priori guarantees grows with $1/\epsilon$ (Remark 4) and could be prohibitive. In other words, the a-priori approach is only suitable for the case where a *sufficient* amount of scenarios is always available. In many real-world applications (e.g. medical experiments, tests conducted by NASA), however, the amount of data is quite limited, and it could take months or cost a fortune to obtain a data point (experiment). Because of the limitation on the data availability, one of the most fundamental problem in data-driven decision making (e.g. system identification, quantitative finance) is to come up with good decisions or estimates with a moderate or even small amount of data. To overcome this, the scenario approach is extended towards a-posteriori feasibility guarantees.

Similar with the *a-priori* guarantees, obtaining a-posteriori guarantees typically requires taking the following three steps:

1. given dataset $\{\xi^i\}_{i=1}^N$, solve the corresponding scenario problem SP_N and obtain x_N^*;
2. find support scenarios in $\{\xi^i\}_{i=1}^N$, whose number is denoted as s_N^*;
3. calculate the posterior violation probability $\epsilon(\beta, s_N^*, N)$ using Theorem 7.

If the resulting violation probability $\epsilon(\beta, s_N^*, N)$ is greater than the acceptable level ϵ, we could repeat this process with more scenarios until reaching $\epsilon(\beta, s_N^*, N) \le \epsilon$. If the number of available scenarios is limited, then it might be impossible to obtain a solution x_N^* such that $\mathbb{V}(x_N^*) \le \epsilon$.

Theorem 7 (Wait-and-Judge (Campi & Garatti, 2016))**.** *Given* $\beta \in (0, 1)$*, for any* $k = 0, 1, \cdots, n$*, the polynomial equation in variable t*

$$\frac{\beta}{N+1}\sum_{i=k}^{N}\binom{i}{k}t^{i-k} - \binom{N}{k}t^{N-k} = 0 \tag{23}$$

has exactly one solution $\epsilon(k)$ *in the interval (0,1). Under Assumptions 1–3, for a non-degenerate problem, it holds that*

$$\mathbb{P}^N(\mathbb{V}(x_N^*) \ge \epsilon(s_N^*)) \le \beta \tag{24}$$

Theorem 7 is particularly useful in the following cases: (i) the problem is not fully-support thus difficult to calculate *a-priori* bounds on number of support scenarios; or (ii) only a moderate or small amount of data points is available, it is difficult to meet the sample complexity from the a-priori guarantees.

Given a candidate solution $x^\diamond$, the most straightforward method is to approximate $\mathbb{V}(x^\diamond)$ by the empirical estimation $\hat{\epsilon}$ through Monte-Carlo simulation with $\hat{N}$ samples, i.e.

$$\hat{\epsilon} = \frac{1}{\hat{N}} \sum_{i=1}^{\hat{N}} \mathbb{1}_{f(x^\diamond,\xi^i)>0} = \frac{\hat{V}}{\hat{N}} \tag{25}$$

where $\hat{V} := \sum_{i=1}^{\hat{N}} \mathbb{1}_{f(x^\diamond,\xi^i)>0}$ is the total number of scenarios in which x_N^* is infeasible. Although (25) only involves $f(x^\diamond, \xi^i) > 0$ which is easy to calculate, it might require an astronomical number $\hat{N}$ to have accurate estimation $\hat{\epsilon}$ because of (D1). Nemirovski and Shapiro (2006) shows a method to bound $\mathbb{V}(x^\diamond)$ from above using a dataset of a moderate size $\hat{N}$.

Proposition 1 (Nemirovski & Shapiro, 2006)**.** *Given a candidate solution $x^\diamond$ and $\hat{N}$ samples, let $\hat{V} := \sum_{i=1}^{\hat{N}} \mathbb{1}_{f(x^\diamond,\xi^i)>0}$ and $1-\rho$ be the confidence parameter.*

$$\overline{\epsilon} := \max_{\gamma\in[0,1]} \{\gamma : \sum_{i=0}^{\hat{V}} \binom{\hat{N}}{i} \gamma^i (1-\gamma)^{\hat{N}-i} \geq \rho\} \tag{26}$$

After finding an upper bound $\overline{\epsilon}$, so that if $\overline{\epsilon} \leq \epsilon$, we may be sure that $\mathbb{P}(\mathbb{V}(x^\diamond) \leq \epsilon) \geq 1-\rho$.

Remark 5. Proposition 1 is closely related with the scenario approach but with one fundamental difference. Theorem 7 holds only for solution from scenario approach, while Proposition 1 can evaluate solutions from other methods.

5.5. Optimality guarantees of scenario approach

Scenario approach together with order statistics can be used to construct lower bounds $\underline{o}$ on $o^\star$ of (CCO).

Proposition 2 (Nemirovski & Shapiro, 2006)**.** *Let $\{\xi^{i,j}\}_{i=1}^N$ $(j = 1, 2, \cdots, K)$ be K independent datasets of size N. For the jth dataset, we solve the associated scenario problem SP_N and calculate the optimal value o_j^* $(j = 1, 2, \cdots, K)$. Without loss of generality, we assume that $o_1^* \leq o_2^* \leq \cdots \leq o_K^*$.*

Given $\delta \in (0, 1)$, let us choose positive integers L, N, K in such a way that

$$\sum_{i=0}^{L-1} \binom{K}{i} (1-\epsilon)^{Ni} [1-(1-\epsilon)^N]^{K-i} \leq \delta \tag{27}$$

then with probability of at least $1-\delta$, the random quantity o_L^* gives a lower bound for the true optimal value $x^\star$.

Pagnoncelli, Ahmed, and Shapiro (2009) shows that appropriate N should be the order of $O(1/\epsilon)$ as $[1-(1-\epsilon)^N]^K \approx (1 - \exp(-\epsilon N))^K$. Typically we choose proper values for N and K first, then find out the largest positive integer L that (27) holds true.

Proposition 2 turns out to be a general framework to construct lower bounds on (CCO). Pagnoncelli et al. (2009) extends the framework towards generating bounds using sample average approximation, which is introduced in Section 6.4.

6. Sample average approximation

6.1. Introduction to sample average approximation

The idea of using sample average approximation to handle chance constraints first appeared in Sen (1992) and was subsequently improved with rigorous theoretical results in Luedtke and Ahmed (2008).

Let $\overline{f}(x,\xi) := \max\left\{f_1(x,\xi), \cdots, f_m(x,\xi)\right\}$, then (CCO) is equivalent to $\min_{x\in\mathcal{X}} c^\intercal x$, s.t. $\mathbb{P}(\overline{f}(x,\xi) \leq 0) \geq 1-\epsilon$. Sample Average Approximation (SAA) approximates the true distribution of the random variable $\overline{f}(x,\xi)$ using the empirical distribution from N samples $\{\xi^i\}_{i=1}^N$, i.e. $\mathbb{P}(\overline{f}(x,\xi) \leq 0)$ is approximated by $\frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\overline{f}(x,\xi^i)\leq 0}$.

$$\text{(SAA)}: \min_{x\in\mathcal{X}} c^\intercal x \tag{28a}$$

$$\text{s.t. } \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\overline{f}(x,\xi^i)>0} \leq \varepsilon \tag{28b}$$

(SAA) is also a chance constrained optimization problem, but with two major differences from (CCO): (i) (SAA) is based on the empirical (discrete) distribution from the true distribution of ξ as in (CCO); (ii) (SAA) has the violation probability ε instead of ϵ in (CCO).

There are two critical questions to be addressed about (SAA). What is the connection of solutions of (SSA) with that of (CCO)? How to solve (SAA)? We first answer the second question in Section 6.2, then present the theoretical results of connecting (SAA) with (CCO).

6.2. Solving sample average approximation

(SAA) can be reformulated as a mixed integer program (MIP) by introducing variables $z \in \{0, 1\}^N$ (Luedtke & Ahmed, 2008; Ruszczyński, 2002). Binary variable z_i is an indicator if $\overline{f}(x,\xi) \leq 0$ is being violated in sample i, i.e.

$$z_i = \mathbb{1}_{\overline{f}(x,\xi^i)>0} \tag{29}$$

(29) can be equivalently written as $\overline{f}(x,\xi^i) \leq Mz_i$ with a sufficiently large coefficient $M > 0$. Since $\overline{f}(x,\xi^i)$ is the maximum over m functions $\{f_j(x,\xi^i)\}_{j=1}^m$, $\overline{f}(x,\xi^i) \leq Mz_i$ implies $f_j(x,\xi^i) \leq Mz_i$, $j = 1, 2, \cdots, m$. Then (SAA) is equivalent to (30), in which $\mathbf{1}_m$ is an all one vector with size m.

$$\min_{x,z} c^\intercal x \tag{30a}$$

$$\text{s.t. } f(x,\xi^1) - Mz_1\mathbf{1}_m \leq 0 \tag{30b}$$

$$\vdots$$

$$f(x,\xi^N) - Mz_N\mathbf{1}_m \leq 0 \tag{30c}$$

$$\frac{1}{N}\sum_{i=1}^N z_i \leq \varepsilon \tag{30d}$$

$$x \in \mathcal{X}, z_i \in \{0,1\}, i = 1, 2, \cdots, N \tag{30e}$$

(30) is equivalent to (SAA) for general function $f(x, \xi)$, but formulations with big-M are typically weak formulations. Introducing big coefficients M might cause numerical issues as well. Stronger formulations of (SAA) are possible by exploiting the structural features of $f(x, \xi)$. A good example is the chance-constrained linear program with separable probabilistic constraints: $\min_{x\in\mathcal{X}} c^\intercal x$ s.t.$\mathbb{P}(Tx \geq \xi) \geq 1-\epsilon$, with a constant matrix $T \in \mathbf{R}^{d\times n}$. By introducing auxiliary variables v, an equivalent but stronger formulation without big M is (31) (Luedtke et al., 2010).

$$\min_{x\in\mathcal{X}} c^\intercal x \tag{31a}$$

$$\text{s.t. } Tx = v \tag{31b}$$

$$v + \xi_i z_i \geq \xi_i, i = 1, 2, \cdots, N \tag{31c}$$

$$\frac{1}{N}\sum_{i=1}^{N} z_i \le \varepsilon \tag{31d}$$

$$z_i \in \{0,1\}, i = 1, 2, \cdots, N \tag{31e}$$

Various strong formulations for (SAA) can be found in Luedtke et al. (2010) and references therein. (30) and (31) are mixed integer programs, some well-known techniques from integer programming theory can speed up the process of solving (SAA), e.g. adding cuts (Küükyavuz, 2012; Luedtke et al., 2010; Tanner & Ntaimo, 2010) and decompositions (Zeng & An, 2014; Zeng, An, & Kuznia, 2017).

6.3. Feasibility guarantees of SAA

Various feasibility guarantees of (SAA) are proved in Luedtke and Ahmed (2008) and Pagnoncelli et al. (2009), e.g. the asymptotic behavior of (SAA) and when $f(x, \xi)$ is Lipschitz continuous. In this section, we only present the Lipschitz case, which could be used for simulations in Section 9.

Assumption 5. There exists $L > 0$ such that

$$|\overline{f}(x,\xi) - \overline{f}(x',\xi)| \le L\|x - x'\|_\infty,\ \forall x, x' \in \mathcal{X} \text{ and } \forall \xi \in \Xi. \tag{32}$$

Theorem 8 (Luedtke & Ahmed, 2008)**.** *Suppose $\mathcal{X}$ is bounded with diameter D and $\overline{f}(x,\xi)$ is L-Lipschitz for any $\xi \in \Xi$ (Assumption 5). Let $\varepsilon \in [0, \epsilon), \theta \in (0, \epsilon - \varepsilon)$ and $\gamma > 0$. Then*

$$\mathbb{P}(\mathcal{F}_{\varepsilon,\gamma}^N \subseteq \mathcal{F}_\epsilon) \ge 1 - \left\lceil \frac{1}{\theta} \right\rceil \left\lceil \frac{2LD}{\gamma} \right\rceil^n \exp(-2N(\epsilon - \varepsilon - \theta)^2) \tag{33}$$

where the feasible region of (SAA) is defined as

$$\mathcal{F}_{\varepsilon,\gamma}^N := \{x \in \mathcal{X} : \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\overline{f}(x,\xi)+\gamma \le 0} \ge 1 - \varepsilon\}. \tag{34}$$

For fixed ϵ and ε, if we choose $\theta = (\epsilon - \varepsilon)/2$ and a small number $\gamma > 0$, then Theorem 8 suggests that using

$$N \ge \frac{2}{(\epsilon - \varepsilon)^2}\Big[\ln(\frac{1}{\beta}) + n\ln(\left\lceil \frac{2LD}{\gamma} \right\rceil) + \ln(\left\lceil \frac{2}{\epsilon - \varepsilon} \right\rceil)\Big] \tag{35}$$

number of samples, solutions of (SAA) is feasible to (CCO) with high probability $1 - \beta$, i.e. $\mathbb{P}(\mathcal{F}_{\varepsilon,\gamma}^N \subseteq \mathcal{F}_\epsilon) \ge 1 - \beta$.

The results in Theorem 8 look quite similar to those of the scenario approach (e.g. Remark 4). Indeed, (SAA) with $\varepsilon = 0$ is the same as the scenario problem SP_N. However, one major difference of Theorem 8 from the scenario approach theory is that: Theorem 8 holds for the feasible region of (SAA), i.e. $\mathcal{F}_{\varepsilon,\gamma}^N \subseteq \mathcal{F}_\epsilon$ with high probability. While the theory of the scenario approach only proves the property of the optimal solution x_N^*, i.e. x_N^* is feasible with high probability. Other feasible solutions to SP_N do not necessarily process the properties guaranteed by the scenario approach (e.g. Theorem 5).

Although Theorem 8 provides explicit sample complexity bounds for (SAA) to obtain feasible solution, it requires some efforts to be applied, e.g. tuning parameters (ε, θ) and calculation of L and D. Campi and Garatti (2011) provides a similar but more straightforward theoretical result.

Theorem 9 (Sampling & Discarding (Campi & Garatti, 2011))**.** *If we draw N samples and discard any k of them, then use the scenario approach with the remaining $N - k$ samples. If N and k satisfy*

$$\binom{k+n-1}{k} \cdot \sum_{i=0}^{k+n-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} \le \beta \tag{36}$$

then $\mathbb{P}^N\Big(\mathbb{P}_\xi(f(x_{N,k}^*, \xi) \le 0) \ge 1 - \epsilon\Big) \ge 1 - \beta.$

Given parameters N, ϵ and β, we find the largest k that (36) holds, then the solution to (SAA) with $\varepsilon = k/N$ is feasible to (CCO) with probability at least $1 - \beta$.

6.4. Optimality guarantees of sample average approximation

It is intuitive that if $\varepsilon > \epsilon$, then the objective values of (SAA) yield lower bounds to (CCO). Theorem 10 formalizes this intuition.

Theorem 10 (Luedtke & Ahmed, 2008)**.** *Let $\varepsilon > \epsilon$ and assume that (CCO) has an optimal solution. Then*

$$\mathbb{P}\Big(\hat{o}_\varepsilon^N \le o_\epsilon^\star\Big) \ge 1 - \exp(-2N(\varepsilon - \epsilon)^2). \tag{37}$$

Theorem 10 directly suggests a method to construct lower bounds on (CCO).

Proposition 3. *If we choose $\varepsilon > \epsilon$ and $N \ge \frac{1}{2(\varepsilon-\epsilon)^2}\log(\frac{1}{\delta})$, let $o_\varepsilon^{\text{SAA}}$ denote the objective value of (SAA), then o_ε is a lower bound with probability at least $1 - \delta$, i.e. $\mathbb{P}(o_{N,\varepsilon}^* \le o_\epsilon^\star) \ge 1 - \delta$.*

There is an alternative method using SAA to generate lower bounds of (CCO). Luedtke and Ahmed (2008) extends the framework in Proposition 2 towards SAA.

Proposition 4 (Luedtke & Ahmed, 2008)**.** *Take K sets of N independent samples $\{\xi^{i,j}\}_{i=1}^N$, $(j = 1, 2, \cdots, K)$. For the jth dataset $\{\xi^{i,j}\}_{i=1}^N$, we solve the associated (SAA) problem and calculate the associated objective value $o_{N,\varepsilon,j}^*$ (for simplicity o_j^* and $j = 1, 2, \cdots, K$). Without loss of generality, we assume that $o_1^* \le o_2^* \le \cdots \le o_K^*$.*

Given $\delta \in (0, 1)$, $\varepsilon \in [0, 1)$, let us choose positive integers N, L, K ($L \le K$) such that

$$\sum_{i=0}^{L-1} \binom{K}{i} \big[b(\varepsilon, \epsilon, N)\big]^i \big[1 - b(\varepsilon, \epsilon, N)\big]^{K-i} \ge \delta \tag{38}$$

where $b(\varepsilon, \epsilon, N) := \sum_{i=0}^{\lfloor \varepsilon N \rfloor} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$.

Then o_L^* serves as a lower bound to (CCO) with probability at least $1 - \delta$.

7. Robust optimization related methods

7.1. Introduction to robust optimization

The last category of solutions to (CCO) is closely related with robust optimization (RO), its typical form is shown in (39).

$$\text{(RC):}\ \min_{x\in\mathcal{X}}\ c^\intercal x \tag{39a}$$

$$\text{s.t. } f(x,\xi) \le 0,\ \forall \xi \in \mathcal{U}_\epsilon \tag{39b}$$

(39) finds the optimal solution which is feasible to all realizations of uncertainties in a predefined uncertainty set $\mathcal{U}_\epsilon$. (39) is called the Robust Counterpart (RC) of the original problem (CCO). By constructing an uncertainty set $\mathcal{U}_\epsilon$ with proper shape and size, solutions to (RC) could be suboptimal or approximate solutions to (CCO).

Designing uncertainty sets lies at the heart of robust optimization. A good uncertainty set should meet the following two standards:

- **(S1)** The resulting (RC) problem is computationally tractable.
- **(S2)** The optimal solution to (RC) is not too conservative or overly optimistic.

Unfortunately, (RC) of general robust convex problems (under Assumption 2) is not always computationally tractable. For example, (RC) of a second order cone program (SOCP) with polyhedral uncertainty set is NP-Hard (Ben-Tal & Nemirovski, 1998; Ben-Tal, Nemirovski, & Roos, 2002; Bertsimas, Brown, & Caramanis, 2011). Fortunately, robust linear programs are well-studied, and (RC) of linear programs is tractable for common choices of uncertainty sets. Most tractability results of robust linear optimization are summarized in Bertsimas et al. (2011). For tractable formulations of general convex RO problems, various solutions can be found in Bertsimas and Sim (2006) and Ben-Tal et al. (2009).

For simplicity, we present solutions to the following chance-constrained linear program (CCLP).[3]

$$\min_{x\in\mathcal{X}} c^\intercal x \tag{40a}$$

$$\text{s.t. } \mathbb{P}_\xi\left(x_0^i + \xi^\intercal x^i \le 0,\ i = 1, 2, \cdots, m\right) \ge 1 - \epsilon \tag{40b}$$

and its robust counterpart

$$\min_{x\in\mathcal{X}} c^\intercal x \tag{41a}$$

$$\text{s.t. } x_0^i + \xi^\intercal x^i \le 0,\ \forall \xi \in \mathcal{U}_\epsilon,\ i = 1, 2, \cdots, m \tag{41b}$$

In (40) and (41), decision variables are $\{x_0^i, x^i\}_{i=1}^m$, where $x_0^i \in \mathbf{R}$ and $x^i \in \mathbf{R}^n$. Uncertainties are represented by $\xi \in \mathbf{R}^d$[4] With a little abuse of notation, we use $x = [x_0^1, x^1, \cdots, x_0^m, x^m]^\intercal$ to represent all the decision variables.

Standard (S2) is directly connected with chance constraints, we show the connection between RO and CCO in Sections 7.2–7.4.

7.2. Safe approximation

Almost every RO-related solution to (CCO) is based on the idea of safe approximation.

Definition 11 (Safe Approximation)**.** Let $x \in \mathcal{F}$ and $x \in \underline{\mathcal{F}}$ denote two sets of constraints. We say $\underline{\mathcal{F}}$ is a safe approximation (or inner approximation) of $\mathcal{F}$ if $\underline{\mathcal{F}} \subseteq \mathcal{F}$.

An optimization problem (SA) is called a *safe approximation* of (CCO) if $\underline{\mathcal{F}} \subseteq \mathcal{F}_\epsilon$, where $\mathcal{F}_\epsilon$ represents the feasible region of (CCO) as in Definition 3.

$$\text{(SA):}\ \min_{x\in\mathcal{X}} c^\intercal x \tag{42a}$$

$$\text{s.t. } x \in \underline{\mathcal{F}} \tag{42b}$$

$\underline{\mathcal{F}} \subseteq \mathcal{F}_\epsilon$ indicates that every solution to (SA) is *feasible* to (CCO). Therefore every optimal solution to (SA) is *suboptimal* to (CCO) and serves as an upper bound on (CCO).

There are two major approaches to constructing safe approximations of the chance constraint $\mathbb{P}_\xi\big(f(x,\xi) \le 0\big) \ge 1 - \epsilon$: (i) constructing a function $\pi(x) \ge \mathbb{P}_\xi\big(f(x,\xi) > 0\big)$, then $\pi(x) \le \epsilon$ is a safe approximation of the chance constraint; (ii) constructing a proper uncertainty set $\mathcal{U}_\epsilon$ such that $\mathcal{F}_\epsilon \supseteq \mathcal{F}_{\mathcal{U}_\epsilon} := \{x \in \mathbf{R}^n : f(x,\xi) \le 0,\ \forall \xi \in \mathcal{U}_\epsilon\}$. Although these two approaches look quite different, Section 7.3.2 shows that they are closely related with each other.

We first review how to apply these two approaches to obtain safe approximation of individual chance constraints in Section 7.2. Safe approximations of joint chance constraints (Section 7.4) are built upon the results of individual chance constraints.

7.3. Safe approximation of individual chance constraints

RO has been quite successful in constructing safe approximations of individual chance constraints. A general form of individual chance-constrained programs is (43).

$$\min_{x\in\mathcal{X}} c^\intercal x \tag{43a}$$

$$\text{s.t. } \mathbb{P}_\xi\left(f(x,\xi) \le 0\right) \ge 1 - \epsilon \tag{43b}$$

In the individual chance constraint (43b), the inner function $f(x, \xi)$: $\mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}^1$ is a *scalar-valued* function. In Section 7.3, all $f(x, \xi)$ are scalar-valued functions if not specified.

Section 7.2 outlines two different but related approaches to constructing safe approximations. The first approach is presented in Sections 7.3.1–7.3.2. The second approach is summarized in Section 7.3.3.

7.3.1. Convex approximation

Convex approximation is a general framework to build safe approximations of individual chance constraints. The idea of convex approximation first appeared in Pintér (1989), then was completed in Nemirovski and Shapiro (2006). The convex approximation framework is based on the concept of generating function.

Definition 12 (Generating Function)**.** A function ϕ: $\mathbf{R} \to \mathbf{R}$ is called a (one-dimensional) *generating function* if it is nonnegative valued, nondecreasing, convex and satisfying the following property:

$$\phi(z) > \phi(0) = 1, \forall z > 0. \tag{44}$$

The idea of convex approximation starts from the following lemma.

Lemma 1. *For a positive constant $t \in \mathbf{R}_+$ and a random variable $z \in \mathbf{R}$, it holds that*

$$\mathbb{E}[\phi(t^{-1}z)] \ge \mathbb{E}[\mathbf{1}_{t^{-1}z \ge 0}] = \mathbb{P}_z(t^{-1}z \ge 0) = \mathbb{P}(z \ge 0) \tag{45}$$

Replace z with $f(x, \xi)$, then $\mathbb{E}[\phi(t^{-1}f(x,\xi))] \ge \mathbb{P}_\xi\Big(f(x,\xi) > 0\Big) = \mathbb{P}_\xi\Big(t^{-1}f(x,\xi) > 0\Big)$. In other words, $\mathbb{E}[\phi(t^{-1}f(x,\xi))] \le \epsilon$ is a *safe approximation* to $\mathbb{P}_\xi\Big(f(x,\xi) \le 0\Big) \ge 1 - \epsilon$.

Theorem 11 (Convex Approximation (Nemirovski & Shapiro, 2006))**.** *Let $\phi(\cdot)$ be a generating function, then (CA) is a safe approximation to (CCO).*

$$\text{(CA):}\ \min_{x\in\mathcal{X}} c^\intercal x \tag{46a}$$

$$\text{s.t. } \inf_{t>0}\Big[t\mathbb{E}_\xi[\phi(\frac{f(x,\xi)}{t})] - t\epsilon\Big] \le 0 \tag{46b}$$

Under Assumption 2, (CA) is convex in x.

Remark 6. We can get rid of the strict inequality $t > 0$ by approximating it using $t \ge \delta$, where δ is very small positive number (e.g. $\delta = 10^{-4}$). Furthermore, we can show that (CA) is equivalent to (47), which is convex in (x, t).

$$\min_{x\in\mathcal{X}, t\ge\delta} c^\intercal x \tag{47a}$$

$$\text{s.t. } t\mathbb{E}_\xi[\phi(\frac{f(x,\xi)}{t})] - t\epsilon \le 0 \tag{47b}$$

Choosing a good generating functions plays a crucial role in the convex approximation framework. Choices of generating functions include: Markov bound $\phi(z) = [1+z]_+$, Chernoff bound $\phi(z) = \exp(z)$, Chebyshev bound $\phi(z) = [z+1]_+^2$ and Traditional Chebyshev bound $\phi(z) = (z+1)^2$. The *least* conservative generating function is the Markov bound $\phi(z) = [1+z]_+$ (Föllmer & Schied, 2011; Nemirovski & Shapiro, 2006).

Definition 13 (Conditional Value at Risk)**.** Conditional value at risk (CVaR) of a random variable z at level $1-\epsilon$ is defined as

$$\text{CVaR}(z; 1-\epsilon) := \inf_\gamma(\gamma + \frac{1}{\epsilon}\mathbb{E}\big[[z-\gamma]_+\big]) \tag{48}$$

[3] A (seemingly) more general form of the linear chance constraint is $\mathbb{P}\Big(A(\xi)x \le b(\xi)\Big) \ge 1 - \epsilon$, where $A(\xi)$ and $b(\xi)$ denote affine functions of ξ. This could be equivalently represented in the form of (40b) by enforcing additional affine constraints (Chen et al., 2010).

[4] Notice $d = n$ in (40) and (41).

Proposition 5 (Chen et al., 2010; Nemirovski & Shapiro, 2006). *(CA) with Markov bound $\phi(z) = [z+1]_+$ is equivalent to (49).*

$$\min_{x \in \mathcal{X}} c^\intercal x \tag{49a}$$

$$\text{s.t. } \mathrm{CVaR}\big(f(x,\xi); 1-\epsilon\big) \le 0 \tag{49b}$$

Section 2 shows an individual chance constraint $\mathbb{P}\big(f(x,\xi) \le 0\big) \ge 1-\epsilon$ is equivalent to $\mathrm{VaR}(f(x,\xi); 1-\epsilon) \le 0$. It is well-known that $\mathrm{CVaR}(z; 1-\epsilon) \ge \mathrm{VaR}(z; 1-\epsilon)$. Therefore, $\mathrm{CVaR}(f(x,\xi); 1-\epsilon) \le 0$ implies $\mathrm{VaR}(f(x,\xi); 1-\epsilon) \le 0$. In other words, $\mathrm{CVaR}(f(x,\xi); 1-\epsilon) \le 0$ is a safe approximation to both $\mathrm{VaR}(f(x,\xi); 1-\epsilon) \le 0$ and the chance constraint (43b).

Remark 7 (Sample Approximation of CVaR). Rockafellar and Uryasev (2000) utilizes a dataset $\{\xi^i\}_{i=1}^N$ to estimate CVaR.

$$\min_{x \in X, t} c^\intercal x \tag{50a}$$

$$\text{s.t. } \frac{1}{N}\sum_{i=1}^{N} [f(x,\xi^i) + t]_+ \le t\epsilon \tag{50b}$$

By introducing N auxiliary variables, Rockafellar and Uryasev (2000) shows that (50) can be reformulated as a convex problem that is easy to solve. Detailed reformulation can be found in Rockafellar and Uryasev (2000) and the full-length version of this paper (Geng & Xie, 2019a). With a sufficient number of data points (N is large enough), (50) is a safe approximation to (CCO). However, it remains unknown about the exact requirement on the number of samples needed. The sample approximation of CVaR may not necessarily yield a safe approximation (Chen et al., 2010).

The generating function based framework in Nemirovski and Shapiro (2006) was further improved and completed in Ben-Tal et al. (2009) and Nemirovski (2012). But the methods proposed there are mainly analytical and aim at solving distributionally robust problems, which is beyond the scope of this paper. More details can be found in Fig. 2 and references therein.

7.3.2. CVaR-based convex approximation of individual chance constraints

As pointed out in Nemirovski and Shapiro (2006), calculating CVaR is computationally intractable. In order to obtain tractable forms of the CVaR-based convex approximation, one approach is the sample approximation in Remark 7. An alternative approach is to bound the CVaR function from above, e.g. finding a function $\pi(x) \ge \mathrm{CVaR}(f(x,\xi); 1-\epsilon)$, then $\pi(x) \le 0$ is a safe approximation to both $\mathrm{CVaR}(f(x,\xi); 1-\epsilon) \le 0$ and the original chance constraint (43). In the latter approach, the uncertainties $\xi \sim \Xi$ are partially characterized using directional deviations.

Definition 14 (Directional Deviations (Chen, Sim, & Sun, 2007)). Given a random variable $\xi \in \mathbf{R}$ with zero mean, the forward deviation is defined as

$$\delta_+(\xi) := \sup_{\theta > 0} \left\{ \sqrt{\frac{2\ln(\mathbb{E}[\exp(\theta\xi)])}{\theta^2}} \right\} \tag{51}$$

and the backward deviation is defined as

$$\delta_-(\xi) := \sup_{\theta > 0} \left\{ \sqrt{\frac{2\ln(\mathbb{E}[\exp(-\theta\xi)])}{\theta^2}} \right\}. \tag{52}$$

Assumption 6 (Chen & Sim, 2009). Let $\mathcal{W}$ denote the smallest closed convex set containing the support Ξ of ξ. We assume that the support set is a second-order conic representable set (e.g. polyhedral and ellipsoidal sets).

Assumption 7 (Chen & Sim, 2009). Assume the uncertainties $\{\xi_i\}_{i=1}^d$ are zero mean random variables, with a positive definite covariance matrix Σ. We define the following index set:

$$\mathcal{J}_+ := \{i : \delta_+(\xi_i) < \infty\},\ \mathcal{I}_+ := \{i : \delta_+(\xi_i) = \infty\}, \tag{53}$$

$$\mathcal{J}_- := \{i : \delta_-(\xi_i) < \infty\},\ \mathcal{I}_- := \{i : \delta_-(\xi_i) = \infty\}. \tag{54}$$

For notation simplicity, we define two matrices diagonal P and Q as:

$$P := \mathrm{diag}(\delta_+(\xi_1), \cdots, \delta_+(\xi_d)),\ Q := \mathrm{diag}(\delta_-(\xi_1), \cdots, \delta_-(\xi_d)).$$

Major results developed in Chen et al. (2007) and Chen and Sim (2009) are for the individual *linear* chance constraint (55) with decision variables $x_0 \in \mathbf{R}$, $x \in \mathbf{R}^n$:

$$\mathbb{P}_\xi\Big(x_0 + \xi^\intercal x \le 0\Big) \ge 1-\epsilon \tag{55}$$

Its convex approximation using CVaR (or Markov bound) is

$$t + \frac{1}{\epsilon}\mathbb{E}\Big[[x_0 + \xi^\intercal x - t]_+\Big] \le 0 \tag{56}$$

If we are able to find a function $\pi(x_0, x)$ as an upper bound on $\mathbb{E}\big[[x_0 + \xi^\intercal x]_+\big]$, then

$$t + \frac{1}{\epsilon}\pi(x_0 - t, x) \le 0 \tag{57}$$

is a safe approximation to (56).

Theorem 12. *(Chen & Sim, 2009) Suppose that the primitive uncertainty ξ satisfies Assumptions 6 and 7. The following functions $\pi^i(x_0, x), i = 1, \cdots, 5$ are upper bounds of $\mathbb{E}_\xi\big[[x_0 + \xi^\intercal x]_+\big]$:*

$$\pi^1(x_0, x) := \Big[x_0 + \max_{\xi \in \mathcal{W}} \xi^\intercal x\Big]_+ \tag{58}$$

$$\pi^2(x_0, x) := x_0 + \Big[-x_0 + \max_{\xi \in \mathcal{W}} (-\xi)^\intercal x\Big]_+ \tag{59}$$

$$\pi^3(x_0, x) := \frac{1}{2}\left(x_0 + \sqrt{x_0^2 + x^\intercal \Sigma x}\right) \tag{60}$$

$$\pi^4(x_0, x) := \inf_{\mu > 0}\left\{\frac{\mu}{\epsilon}\exp\left(\frac{x_0}{\mu} + \frac{u^\intercal u}{2\mu^2}\right)\right\}. \tag{61}$$

where $u_j = \max\{x_j\delta_+(\xi_j), -x_j\delta_-(\xi_j)\}, j = 1, \cdots, n$. This bound is finite if and only if $x_j \le 0,\ \forall j \in \mathcal{I}_+$ and $x_j \ge 0,\ \forall j \in \mathcal{I}_-$.

$$\pi^5(x_0, x) := x_0 + \inf_{\mu > 0}\left\{\frac{\mu}{\epsilon}\exp\left(-\frac{x_0}{\mu} + \frac{v^\intercal v}{2\mu^2}\right)\right\}. \tag{62}$$

where $v_j = \max\{-x_j\delta_+(\xi_j), x_j\delta_-(\xi_j)\}, j = 1, \cdots, n$. This bound is finite if and only if $x_j \ge 0,\ \forall j \in \mathcal{I}_+$ and $x_j \le 0,\ \forall j \in \mathcal{I}_-$.

Remark 8. The epigraphs of $\pi^i(x_0, x),\ i = 1, \cdots, 5$ can be represented as second-order cones. Explicit representations depend on the form of $\mathcal{W}$. More details about the representation of (57) with different choices of $\pi^i(x_0, y)$ can be found in Chen and Sim (2009) and Geng and Xie (2019a).

7.3.3. Constructing uncertainty sets

We consider the individual linear chance constraint (55) as in Section 7.3.2. The robust counterpart of (55) is

$$x_0 + \xi^\intercal x \le 0,\ \forall \xi \in \mathcal{U}_\epsilon \tag{63}$$

Assumption 8. $\{\xi_i\}_{i=1}^d$ are independent of each other with zero mean and take values on $[-1, 1]^d$, i.e. $\mathbb{E}[\xi_i] = 0$ and $\xi_i \in [-1, 1]$ for $i = 1, 2, \cdots, d$.

Clearly, under Assumption 8, a natural choice of uncertainty set is the box $\mathcal{U}^{\text{box}} := \{\xi \in \mathbf{R}^d : -\mathbf{1} \le \xi \le \mathbf{1}\}$. Then $\mathcal{F}_{\mathcal{U}}^{\text{box}} := \{x \in \mathbf{R}^n :$

$f(x, \xi) \le 0,\ \forall \xi \in \mathcal{U}^{\text{box}}\}$ is a safe approximation to $\mathcal{F}_\epsilon$, i.e. $\mathcal{F}_{\mathcal{U}}^{\text{box}} \subseteq \mathcal{F}_\epsilon$. However, using $\mathcal{U}^{\text{box}}$ leads to $\mathbb{P}(f(x, \xi) \ge 0) = 0 \ll \epsilon$, which causes conservativeness or even infeasibility in many cases. The following choices of uncertainty sets are less conservative.

Theorem 13 (Ben-Tal et al., 2009; Ben-Tal & Nemirovski, 1999; Bertsimas & Sim, 2004)**.** *(63) is a safe approximation to (55) if $\mathcal{U}_\epsilon$ is one of the following:*

$$\mathcal{U}_\epsilon^{\text{ball}} := \left\{\xi \in \mathbf{R}^d : \|\xi\|_2 \le \sqrt{2\ln(1/\epsilon)}\right\} \tag{64a}$$

$$\mathcal{U}_\epsilon^{\text{ball-box}} := \left\{\xi \in \mathbf{R}^d : \|\xi\|_\infty \le 1, \|\xi\|_2 \le \sqrt{2\ln(1/\epsilon)}\right\} \tag{64b}$$

$$\mathcal{U}_\epsilon^{\text{budget}} := \left\{\xi \in \mathbf{R}^d : \|\xi\|_1 \le \sqrt{2d\ln(1/\epsilon)}\right\} \tag{64c}$$

And the resulting robust counterparts (RC)s are second-order cone representable (see Chapter 2 of Ben-Tal et al. (2009) and the full-length version of this paper (Geng & Xie, 2019a)).

It turns out that constructing uncertainty set $\mathcal{U}_\epsilon$ is closely related with the convex approximation framework in Sections 7.3.1–7.3.2.

Theorem 14 (Chen et al., 2010)**.** *Suppose that $\pi(x_0, x)$ is a convex, closed and positively homogeneous, and is an upper bound to $\mathbb{E}_\xi\big[[x_0 + \xi^\intercal x]_+\big]$ with $\pi(x_0, 0) = x_0^+$. Then under Assumptions 6 and 7 and given $\epsilon \in (0, 1)$, it holds that for all (x_0, x) such that $\pi(x_0, x) < \infty$, we have*

$$\inf_t \left(t + \frac{1}{\epsilon}\pi(x_0 - t, x)\right) = x_0 + \max_{z \in \mathcal{U}_\epsilon} x^\intercal z \tag{65}$$

for some convex uncertainty set $\mathcal{U}_\epsilon$.

Given an upper bound $\pi(x_0, x)$ on $\mathbb{E}\big[[x_0 + \xi^\intercal x]_+\big]$ with required properties, the safe approximation (57) can be represented in the form of $x_0 + \max_{\xi \in \mathcal{U}_\epsilon} \xi^\intercal x$ for some $\mathcal{U}_\epsilon$. Theorem 14 only proves the existence of a corresponding uncertainty set $\mathcal{U}_\epsilon$. For the $\pi^i(x_0, x)$ functions given in Theorem 12, their corresponding uncertainty sets can be explicitly calculated.

Proposition 6 (Chen et al., 2010)**.** *For the functions $\pi^i(x_0, x), i = 1, 2, \cdots, 5$ in Theorem 12, their corresponding uncertainty sets are $\mathcal{U}_\epsilon^1 \sim \mathcal{U}_\epsilon^5$ below.*

$$\mathcal{U}_\epsilon^1 := \mathcal{W}, \tag{66}$$

$$\mathcal{U}_\epsilon^2 := \left\{\xi \in \mathbf{R}^d : \xi = (1 - \frac{1}{\epsilon})\zeta,\ \text{for some } \zeta \in \mathcal{W}\right\}, \tag{67}$$

$$\mathcal{U}_\epsilon^3 := \left\{\xi \in \mathbf{R}^d : \|\Sigma^{-\frac{1}{2}}\xi\|_2 \le \sqrt{\frac{1-\epsilon}{\epsilon}}\right\} \tag{68}$$

$$\mathcal{U}_\epsilon^4 := \Big\{\xi \in \mathbf{R}^d : \exists s, t \in \mathbf{R}^d, \xi = s - t,$$
$$\|P^{-1}s + Q^{-1}t\|_2 \le \sqrt{-2\ln(\epsilon)}\Big\}, \tag{69}$$

$$\mathcal{U}_\epsilon^5 := \Big\{\xi \in \mathbf{R}^d : \exists s, t \in \mathbf{R}^d, \xi = s - t,$$
$$\|P^{-1}s + Q^{-1}t\|_2 \le \frac{1-\epsilon}{\epsilon}\sqrt{2\ln(\frac{1}{1-\epsilon})}\Big\}. \tag{70}$$

where matrices Σ, P and Q are defined in Assumptions 6 and 7.

Theorem 14 and Proposition 6 demonstrate that the two seemingly different approaches to constructing safe approximations in Section 7.2 are equivalent in many circumstances.

7.4. Safe approximation of joint chance constraints

Although RO has been successful in approximating individual chance constraints, it is rather unsatisfactory in approximating joint chance constraints (Chen et al., 2010). We restate the joint chance constraint (1b) below

$$\mathbb{P}_\xi\Big(f(x, \xi) \le 0\Big) \ge 1 - \epsilon. \tag{71}$$

Most RO-based approaches convert a joint chance constraint to several individual chance constraints, then apply the techniques in Section 7.3 on each individual chance constraint. Results along this line are summarized in Section 7.4.1. Very few approaches directly deal with joint chance constraints, these approaches are mentioned in Section 7.4.2.

7.4.1. Conversion between joint chance constraints and individual chance constraints

Section 2.2 presents two common approaches to converting a joint chance constraint to individual chance constraints.

First, according to the Bonferroni inequality, if $\sum_{i=1}^m \epsilon_i \le \epsilon$, then the set of m individual chance constraints

$$\mathbb{P}\Big(f_i(x, \xi) \le 0\Big) \le 1 - \epsilon_i,\ \ i = 1, \cdots, m \tag{72}$$

is a safe approximation to the joint chance constraint $\mathbb{P}(f(x, \xi) \le 0) \le 1 - \epsilon$. The main issue of this approach is the choice of $\{\epsilon_i\}_{i=1}^m$. The problem becomes intractable if taking $\{\epsilon_i\}_{i=1}^m$ as decision variables (Chen et al., 2010; Nemirovski & Shapiro, 2006). It remains unclear about how to find the optimal choices of $\{\epsilon_i\}_{i=1}^m$.[5] Obviously, this approach could be quite conservative in the following two cases: (i) the individual constraints $f_i(x, \xi),\ i = 1, 2, \cdots, m$ are correlated; and (ii) the choices of $\{\epsilon_i\}_{i=1}^m$ are suboptimal. Chen et al. (2010) provides some deep observations on the limitation of this approach: the Bonferroni's inequality could still lead to conservativeness even when (i) the individual chance constraints (72) are independent; and (ii) the optimal choices of $\{\epsilon_i\}_{i=1}^m$ are found. In other words, (72) is only a safe approximation at best, it may not be equivalent to (1b) even with optimal $\{\epsilon_i\}_{i=1}^m$.

The second approach is to define the pointwise maximum of functions $\{f_i(x, \xi)\}_{i=1}^m$ over x and ξ, i.e.

$$\overline{f}(x, \xi) := \max\Big\{f_1(x, \xi), \cdots, f_m(x, \xi)\Big\}.$$

then the joint chance constraint $\mathbb{P}(f(x, \xi) \le 0) \ge 1 - \epsilon$ is equivalent to the individual chance constraint $\mathbb{P}_\xi\left(\overline{f}(x, \xi) \le 0\right) \ge 1 - \epsilon$. The advantage of this approach is that it does not require parameter tuning or induce additional conservativeness. In some cases, e.g. the scenario approximation of CVaR in Remark 7, this could lead to formulations that are easy to solve (Geng & Xie, 2019a). However, in most cases, the structure of $\overline{f}(x, \xi)$ is too complicated to apply the techniques in Section 7.3.

7.4.2. Other approaches

There might be only three RO-related approaches that directly deal with joint chance constraints. The first approach is robust conic optimization (see Chapter 5–11 of Ben-Tal et al. (2009)). The inner constraint $f(x, \xi) \le 0$ is written as a conic inequality, then tractable safe approximations of the robust conic inequality are derived and solved. This approach can model a majority of optimization problems under uncertainties. However, the main limitation is that the resulting robust counterparts are not tractable in many circumstances.

[5] Most people simply choose $\epsilon_i = \epsilon/m$ (Chen et al., 2007; Nemirovski & Shapiro, 2006), which could be quite conservative if m is a large number.

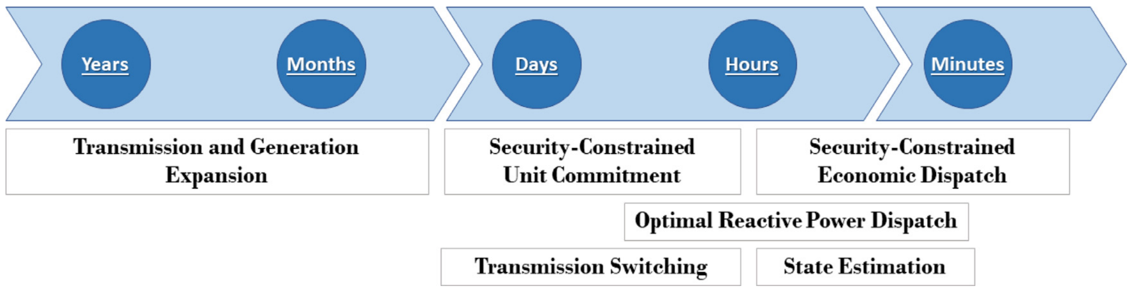


Fig. 3. Representative feed-forward decisions made in power system planning and operation.

The second approach (Chen et al., 2010) generalizes the CVaR-based convex approximation in Theorem 12 and Proposition 6. It proposes a safe approximation to the joint chance constraint (1b), and the safe approximation is second-order cone representable. The performance of this approach depends on the choice of a few tuning parameters. Although it is difficult to find the optimal setting, Chen et al. (2010) designed an algorithm that is guaranteed to improve the choice of parameters. Chen et al. (2010) also shows that it is possible to combine all the $\pi^i(x_0, x)$ functions in Theorem 12 together to reduce conservativeness.

The third approach directly dealing with joint chance constraints is the *data-driven robust optimization* proposed in Bertsimas, Gupta, and Kallus (2018). It shows that by running different hypothesis tests on datasets, it is possible to construct different uncertainty sets that lead to safe approximations of the joint chance constraint (1b) with high probability. It is worth noting that the theoretical results in Bertsimas et al. (2018) holds for non-convex functions $f(x, \xi)$, albeit the resulting (RC) is very likely to be computationally intractable.

8. Applications in power systems

A pivotal task in modern power system operation is to maintain the real-time balance of supply and demand while ensuring the system is low-cost and reliable. This pivotal task, however, faces critical challenges in the presence of rapid growth of renewable energy resources. Chance-constrained optimization, which explicitly models the risk that the system is exposed to, is a suitable conceptual framework to ensure the security and reliability of a power system under uncertainties.

There is a large body of literature adopting CCO for power system applications. Fig. 3 presents some existing applications of CCO in power systems. In the following sections, we introduce three important applications of CCO in power systems: security-constrained economic dispatch (Section 8.1), security-constrained unit commitment (Section 8.2) and generation and transmission expansion (Section 8.3).

Fig. 3 also presents a feed-forward decision making framework for power system operations. The feed-forward framework partitions the overall decision making process into several time segments. The longer-term decisions (e.g. generation expansion) are fed into shorter-term decision making processes (e.g. unit commitment). The shorter-term decisions (e.g. generation commitment from SCUC) have direct impacts on real-time operations (e.g. dispatch results in SCED). As time draws closer to the actual physical operation, information gets much sharper and the prediction about future could be significantly improved (Xie et al., 2011).

8.1. Security-constrained economic dispatch

8.1.1. Deterministic SCED

Security-constrained Economic Dispatch (SCED) lies at the center of modern electricity markets and short-term power system operations. It determines the most cost-efficient output levels of generators while keeping the real-time balance between supply and demand. Different variations of the SCED problem are all based on the direct current optimal power flow (DCOPF) problem. We present a typical form of DCOPF with wind generation.

$$\text{(det-DCOPF):} \quad \min_{g} \; c(g) \tag{73a}$$

$$\text{s.t } \mathbf{1}^\intercal g = \mathbf{1}^\intercal d - \mathbf{1}^\intercal \hat{w} \tag{73b}$$

$$f = H_g g + H_w \hat{w} - H_d d \tag{73c}$$

$$\underline{f} \le f \le \overline{f} \tag{73d}$$

$$\underline{g} \le g \le \overline{g} \tag{73e}$$

The decision variables are generation output levels $g \in \mathbf{R}^{n_g}$. The objective of (det-DCOPF) is to minimize total generation cost $c(g)$, while ensuring total generation equates total *net* demand[6] (73b). Constraints include transmission line flow limits (73c)–(73d) and generation capacity limits (73e). Transmission line flows $f \in \mathbf{R}^{n_l}$ are calculated using (73c), in which H is the power transfer distribution factor (PTDF) matrix, and $H_g \in \mathbf{R}^{n_l \times n_g}$ ($H_d \in \mathbf{R}^{n_l \times n_d}, H_w \in \mathbf{R}^{n_l \times n_w}$) denotes the submatrix formed by the columns of H corresponding to generators (loads, wind farms). (73) utilizes the expected wind generation or wind forecast $\hat{w}$, we refer to (73) as *deterministic DCOPF* (det-DCOPF) since no uncertainties are being considered.

8.1.2. Chance-constrained SCED

Many researchers advance (det-DCOPF) towards a chance-constrained formulation with wind uncertainties. A representative formulation is (74), which appears in a majority of the existing literatures, e.g. (Bienstock, Chertkov, & Harnett, 2014; Vrakopoulou, Margellos, Lygeros, & Andersson, 2013a).

(cc-DCOPF):

$$\min_{g,\eta} \; c(g) \tag{74a}$$

$$\text{s.t } \mathbf{1}^\intercal g = \mathbf{1}^\intercal d - \mathbf{1}^\intercal \hat{w} \tag{74b}$$

$$f(\hat{w}, \tilde{w}) = H_g(g - \mathbf{1}^\intercal \tilde{w}\eta) - H_d d + H_w(\hat{w} + \tilde{w}) \tag{74c}$$

$$\mathbb{P}_{\tilde{w}}\Big(\underline{f} \le f(\hat{w}, \tilde{w}) \le \overline{f} \text{ and } \underline{g} \le g - \mathbf{1}^\intercal \tilde{w}\eta \le \overline{g}\Big) \ge 1 - \epsilon \tag{74d}$$

$$\mathbf{1}^\intercal \eta = 1 \tag{74e}$$

[6] Wind generation is treated as negative loads.

$$\underline{g} \le g \le \overline{g} \tag{74f}$$

$$-\mathbf{1} \le \eta \le \mathbf{1} \tag{74g}$$

Unlike (det-DCOPF) using wind forecast $\hat{w}$, chance-constrained DCOPF (cc-DCOPF) explicitly models wind generation as a random vector $w \in \mathbf{R}^{n_w}$. The wind generation $w = \hat{w} + \tilde{w}$ is decomposed into two components: the *deterministic* wind forecast value $\hat{w} \in \mathbf{R}^{n_w}$ and the *uncertain* forecast error $\tilde{w} \in \mathbf{R}^{n_w}$. To guarantee the real-time balance of supply and demand, (cc-DCOPF) introduces an affine control policy $\eta \in [-1, 1]^{n_g}$ to proportionally allocate total wind fluctuations $\mathbf{1}^\intercal \tilde{w}$ to each generator. It is easy to verify that constraints (74b) and (74e) imply the supply-demand balance in the presence of wind uncertainties, i.e.

$$\mathbf{1}^\intercal (g - \mathbf{1}^\intercal \tilde{w}\eta) = \mathbf{1}^\intercal d - \mathbf{1}^\intercal (\hat{w} + \tilde{w}), \tag{75}$$

The affine policy vector $\eta \in \mathbf{R}^{n_g}$ is sometimes referred as participation factor or distribution vector (Vrakopoulou et al., 2013a). The (joint) chance constraint (74d) constrains the transmission flow and generation within their capacities with high probability $1 - \epsilon$ in the presence of wind uncertainties.

For simplicity, we only account for the major source of uncertainties (i.e. wind) in the real-time. Many references provides more complicated formulation of (cc-DCOPF), e.g. considering joint uncertainties from load and wind (Doostizadeh, Aminifar, Ghasemi, & Lesani, 2016; Mühlpfordt, Faulwasser, Roald, & Hagenmeyer, 2017), and contingencies of potential generator or transmission line outages (Roald, Misra, Chertkov, & Andersson, 2015).

There exist a few different but similar formulations of (cc-DCOPF). In general, policies of any form could help balance supply with demand under uncertainties. The affine policy in (cc-DCOPF) is the simplest choice and lead to optimization problems that are easy to solve. There are other papers applying different forms of policies, e.g. Jabr (2013) introduces a matrix form of the affine policy $\Upsilon \in \mathbf{R}^{n_g \times n_w}$, which specifies the corrective control of each generator on each wind farm. (cc-DCOPF) is a single snapshot dispatch problem, it is straightforward to extend it to a multi-period or look-ahead dispatch problem (Modarresi et al., 2018; Vrakopoulou et al., 2013a). Many papers evaluate the impacts of new elements in modern power systems, such as demand response (Ming, Xie, Campi, Garatti, & Kumar, 2017; Zhang, Shen, & Mathieu, 2017), ambient temperatures and meteorological quantities (Bucher, Vrakopoulou, & Andersson, 2013), and frequency control (Li & Mathieu, 2015; Zhang, Shen et al., 2017).

Although DC power flow equations have been widely accepted in modern power system operations and planning, it is only a linear approximation of the alternating current (AC) version, which is a more accurate model of the underlying physical laws. Many efforts have been made to solve the chance-constrained AC optimal power flow (cc-ACOPF) problem, e.g. Vrakopoulou, Katsampani, Margellos, Lygeros, and Andersson (2013), Roald and Andersson (2017), Venzke, Halilbasic, Markovic, Hug, and Chatzivasileiadis (2017) and Anese, Baker, and Summers (2017). Major difficulties to solve cc-ACOPF come from the non-convexity of AC power flow equations. It remains as an open question that how to ensure the feasibility of the non-convex AC power flow equations under uncertainties.

8.1.3. Solving cc-DCOPF

Table 1 summarizes various methods to solve (cc-DCOPF). The most popular one consists of two steps: (i) decomposing the joint chance constraint (74d) into individual ones $\mathbb{P}_\xi(f_i(x, \xi) \le 0) \ge 1 - \epsilon_i, i = 1, 2, \cdots, m$; (ii) deriving the deterministic equivalent form of each individual chance constraint by making the Gaussian assumption. More technical details of this method are in Section 3.2. This method is taken by many researchers for its simplicity and computationally tractable reformulation. Although the Gaussian assumption enjoys the law of large numbers, it is often an approximation or even doubtful assumption. For example, Hodge and Milligan (2011) shows that the wind forecast error is better represented by Cauchy distributions instead of Gaussian ones. The first step of this method is to decompose a joint chance constraint $\mathbb{P}_\xi(f(x, \xi) \le 0) \ge 1 - \epsilon$ into individual ones. As discussed in Sections 2.2 and 7.4.1, this step often introduces conservativeness because of the limitation of the Bonferroni inequality. The level of conservativeness could be significant when the number of constraints m is large, which is typically the case in power systems.

The scenario approach is another commonly-accepted method. It provides rigorous guarantees on the quality of the solution and does not assume the distribution is Gaussian or any particular type. Most papers adopting the scenario approach apply the *a-priori* guarantees (e.g. Theorem 5 and 6) on (cc-DCOPF) and verify the a-posteriori feasibility of solutions through Monte-Carlo simulations (25). One common observation is that the solution x_N^* is often quite conservative, i.e. $\mathbb{V}(x_N^*) \ll \epsilon$. One major source of conservativeness is the loose sample complexity bounds N.[7] Since (cc-DCOPF) is convex, Theorem 4 states that the number of decision variables n is an upper bound of the number of support scenarios $|\mathcal{S}|$ or Helly's dimension h. This upper bound, as pointed out in Modarresi et al. (2018), is indeed very loose. (Modarresi et al., 2018) reported only ~ 5 support scenarios for a chance-constrained look-ahead SCED problem with thousands of decision variables. By exploiting the structural features of (cc-DCOPF), the sample complexity bound N can be significantly improved. Unfortunately, only Modarresi et al. (2018) and Ming et al. (2017) followed this path to reduce conservativeness.

There are also many papers utilizing the robust optimization related methods to solve (cc-DCOPF). Jiang, Wang, and Guan (2012) constructs uncertainty sets with the help of probabilistic guarantees in Bertsimas and Sim (2004). References Summers, Warrington, Morari, and Lygeros (2014, 2015) incorporate the convex approximation framework and compare different choices of generating functions $\phi(z)$ on (cc-DCOPF). Although there are no explicit forms of chance constraints in Zhang and Giannakis (2013), the CVaR-oriented approach therein can be interpreted as solving cc-DCOPF using convex approximation with the choice of Markov bound.

Most papers in Table 1 aim at finding suboptimal solutions to (cc-DCOPF). However, it is somewhat surprising to note that none of them estimates how suboptimal the solution is via approaches like Proposition 2 or 4. Almost all the papers evaluate the a-posteriori feasibility by Monte-Carlo simulations with a huge sample size. Methods like Proposition 1 would be more attractive when data is limited, which is closer to the reality.

8.2. Security-constrained unit commitment

8.2.1. Deterministic SCUC

Security-Constrained Unit Commitment (SCUC) is one of the most important procedures in power system day-ahead or intra-day operations.

(det-SCUC):

$$\min_{z,u,v,g,s} \sum_{t=1}^{n_t} c_n^\intercal z^t + c_u^\intercal u^t + c_v^\intercal v^t + c_g^\intercal g^{t,0} + c_s^\intercal s^t \tag{76a}$$

$$\text{s.t. } \mathbf{1}^\intercal g^{t,k} \ge \mathbf{1}^\intercal \hat{d}^t - \mathbf{1}^\intercal \hat{w}^t \tag{76b}$$

[7] Many papers still utilize the first sample complexity bound proved in Calafiore and Campi (2005), which was significantly tightened in Campi and Garatti (2008) and following works (Calafiore, 2010).

Table 1
Power system applications of chance-constrained optimization.

	Methods	Expansion	SCUC	SCED	Other Applications
Deterministic Equivalent	Gaussian	(López et al., 2007; Manickavasagam et al., 2015; Mazadi et al., 2009; Sanghvi et al., 1982)	(Ding, Lee, Jianxue, & Liu, 2010; Pozo & Contreras, 2013; Wu et al., 2014)	(Bent, Bienstock, & Chertkov, 2013; Bienstock, Chertkov, & Harnett, 2013; 2014; Doostizadeh et al., 2016; Jabr, 2013; Li & Mathieu, 2015; Li, Vrakopoulou, & Mathieu, 2019; Lubin, Dvorkin, & Backhaus, 2016; Roald, Misra et al., 2015; Roald, Misra, Krause, & Andersson, 2017; Roald, Misra, Morrison, & Andersson, 2017; Roald, Oldewurtel, Krause, & Andersson, 2013; Vrakopoulou, Li, & Mathieu, 2019; Wang et al., 2017; Zhang, Shen, & Mathieu, 2015)	(Franco, Rider, & Romero, 2016; López, Pozo, Contreras, & Mantovani, 2015)
Scenario Approach	a-priori	-	(Geng et al., 2019; Margellos et al., 2013)	(Bucher et al., 2013; Geng & Xie, 2019b; Ming et al., 2017; Modarresi et al., 2018; Roald, Vrakopoulou, Oldewurtel, & Andersson, 2014; Roald, Vrakopoulou, Oldewurtel, & Andersson, 2015; Vrakopoulou et al., 2013a; 2013b; Zhang, Shen et al., 2015)	(Yang & Nehorai, 2014)
	a-posteriori	-	(Geng et al., 2019; Hreinsson, Vrakopoulou, & Andersson, 2015; Margellos et al., 2013)	(Geng & Xie, 2019b; Modarresi et al., 2018)	-
Sample Average Approximation	-	(Zhang, Wang, Li, & Cao, 2017)	(Bagheri et al., 2017; Tan & Shaaban, 2016; Wang et al., 2012; Wang et al., 2013; Zhang, Wang, Zeng et al., 2017; Zhao et al., 2014)	(Geng & Xie, 2019b)	-
RO-based Approach	RLO	-	Jiang et al. (2012)	(Geng & Xie, 2019b)	-
	Convex Approximation	-	-	(Geng & Xie, 2019b; Summers et al., 2014; 2015; Zhang & Giannakis, 2013)	-
Others	-	(Qiu et al., 2016; Yang & Wen, 2005)	(Martínez & Anderson, 2015; Wu, Zeng, Zhang, & Zhou, 2016)	(Bienstock et al., 2014; Doostizadeh et al., 2016; Ke, Chung, & Sun, 2016; Mühlpfordt et al., 2017; Vrakopoulou et al., 2013a; Wang et al., 2017)	-

$$\underline{f} \le H_g^{t,k} g^{t,k} - H_d^{t,k} \hat{d}^t + H_w^{t,k} \hat{w}^t \le \overline{f} \tag{76c}$$

$$\underline{r} \le g^{t,k} - g^{t-1,k} \le \overline{r} \tag{76d}$$

$$a^k \circ (g^{t,0} - s^t) \le g^{t,k} \le a^k \circ (g^{t,0} + s^t) \tag{76e}$$

$$k \in [0, n_k], t \in [1, n_t]$$

$$\underline{g} \circ z^t \le g^{t,0} \le \overline{g} \circ z^t \tag{76f}$$

$$\underline{s} \circ z^t \le s^t \le \overline{s} \circ z^t \tag{76g}$$

$$\underline{g} \circ z^t \le g^{t,0} - s^t \le g^{t,0} + s^t \le \overline{g} \circ z^t \tag{76h}$$

$$z^{t-1} - z^t + u^t \ge 0 \tag{76i}$$

$$z^t - z^{t-1} + v^t \ge 0 \tag{76j}$$

$$t \in [1, n_t]$$

$$z_i^t - z_i^{t-1} \le z_i^\iota, \ \iota \in [t+1, \min\{t + \underline{u}_i - 1, n_t\}] \tag{76k}$$

$$z_i^{t-1} - z_i^t \le 1 - z_i^\iota, \ \iota \in [t+1, \min\{t + \underline{v}_i - 1, n_t\}] \tag{76l}$$

$$i \in [1, n_g], \ t \in [2, n_t]$$

Deterministic SCUC (det-SCUC) seeks the optimal commitment and generation schedule of n_g generators for the upcoming n_t snapshot while ensuring system security in n_k contingencies. Decision variables include commitment and startup/shutdown decisions (z^t, u^t, v^t), as well as generation and reserve schedules ($g^{t,k}$, s^t). The objective of (76) is to minimize total operation costs, which include no-load costs $c_n^\intercal z^t$, startup costs $c_u^\intercal u^t$, shutdown costs $c_v^\intercal v^t$, generation costs $c_g^\intercal g^{t,0}$ and reserve costs $c_s^\intercal s^t$. Constraint (76b) assures there is enough supply to meet *net* demand. Constraints (76c), (76d) and (76g) are about transmission capacity, generation ramping capability and reserve limit in contingency scenario k at time t. In contingency scenarios, the adjusted output $g_i^{t,k}$ of generator i is bounded by its reserve s_i^t. Vector $a^k \in \{0, 1\}^{n_g}$ represents the availability of generators in contingency k. When $a_i^k = 0$, generator i is not available in contingency k, thus has zero generation output. Generation and reserve capacity constraints are in (76f) and (76g). Constraints (76f)–(76h) also ensure the consistency of generation with commitment decisions. (76i)–(76j) are the logistic constraints about commitment status, startup and shutdown decisions. Minimum on/off time constraints for all generators are presented in (76k)–(76l).

8.2.2. Chance-constrained SCUC

Many researchers proposed various advanced formulations of SCUC to deal with uncertainties, e.g. using robust optimization (Bertsimas, Litvinov, Sun, Zhao, & Zheng, 2013) and stochastic programming (Takriti, Birge, & Long, 1996). A good overview of SCUC formulations with uncertainties is in Zheng, Wang, and Liu (2015). In this paper, we formulate the chance-constrained SCUC problem. Unlike the case of SCED, there is no unified formulation of chance-constrained SCUC. We present one simplified formulation in (77). Alternative formulations of chance-constrained SCUC can be found in (Jiang et al., 2012; Wu, Shahidehpour, & Li, 2007; Zheng et al., 2015).

(cc-SCUC):

$$\min_{z,u,v,g,s} \sum_{t=1}^{n_t} c_n^\intercal z^t + c_u^\intercal u^t + c_v^\intercal v^t + c_g^\intercal g^{t,0} + c_s^\intercal s^t \tag{77a}$$

$$\begin{aligned} \text{s.t. } & (76b), (76c), (76d), (76e),\ k \in [0, n_k], t \in [1, n_t] \\ & (76f), (76g), (76h)), (76i), (76j),\ t \in [1, n_t] \\ & (76k), (76l),\ i \in [1, n_g],\ t \in [2, n_t] \end{aligned}$$

$$\mathbb{P}\Big(\mathbf{1}^\intercal g^{t,k} \geq \mathbf{1}^\intercal (\hat{d}^t + \tilde{d}^t) - \mathbf{1}^\intercal (\hat{w}^t + \tilde{w}^t), \tag{77b}$$

$$\underline{f} \leq H_g^{t,k} g^{t,k} - H_d^{t,k} (\hat{d}^t + \tilde{d}^t) + H_w^{t,k} (\hat{w}^t + \tilde{w}^t) \leq \overline{f}, \tag{77c}$$

$$k \in [0, n_k], t \in [1, n_t] \Big) \geq 1 - \epsilon \tag{77d}$$

The formulation of (cc-SCUC) is almost identical to (det-SCUC) except the chance constraint (77b)–(77d). In (cc-SCUC), wind generation $w \in \mathbf{R}^{n_w}$ is modeled as a random vector consisting of a deterministic predicted component $\hat{w} \in \mathbf{R}^{n_w}$ and a stochastic error component $\tilde{w} \in \mathbf{R}^{n_w}$. The chance constraint (77b)–(77d) ensures enough supply to meet demand and line flows within limits under uncertainties with probability at least $1 - \epsilon$ for any contingency scenario k at any time t.

The *joint* chance constraint (77b)–(77d) is sometimes written as two (joint) chance constraints:

$$\mathbb{P}\Big(\mathbf{1}^\intercal g^{t,k} \geq \mathbf{1}^\intercal (\hat{d}^t + \tilde{d}^t) - \mathbf{1}^\intercal (\hat{w}^t + \tilde{w}^t),\ k \in [0, n_k],\ t \in [1, n_t] \Big) \geq 1 - \epsilon^{\text{LOLP}} \tag{78a}$$

$$\mathbb{P}\Big(\underline{f} \leq H_g^{t,k} g^{t,k} - H_d^{t,k} (\hat{d}^t + \tilde{d}^t) + H_w^{t,k} (\hat{w}^t + \tilde{w}^t) \leq \overline{f},\ k \in [0, n_k],\ t \in [1, n_t] \Big) \geq 1 - \epsilon^{\text{TLOP}} \tag{78b}$$

An important metric to evaluate power system reliability is through the *loss of load probability* (LOLP), which is defined as the probability that the total demand is not met by the total generation (Allan & others, 2013; Qiu et al., 2016). It can be seen that (78a) is essentially ensuring the value of LOLP will not exceed a desired level ϵ^{LOLP}. Similarly, we could define the concept *transmission line overload probability* (TLOP) (Wu, Shahidehpour, Li, & Tian, 2014). Then (78b) is the same as TLOP $\leq \epsilon^{\text{TLOP}}$.

Some papers (e.g. (Wu et al., 2014)) further break down the joint chance constraint (78a)–(78b) into individual chance constraints (79a)–(79b), which can be interpreted as constraints on LOLP or TLOP for each time period t.

$$\mathbb{P}\Big(\mathbf{1}^\intercal g^{t,k} \geq \mathbf{1}^\intercal (\hat{d}^t + \tilde{d}^t) - \mathbf{1}^\intercal (\hat{w}^t + \tilde{w}^t) \Big) \geq 1 - \epsilon_{t,k}^{\text{LOLP}},\quad k \in [0, n_k],\ t \in [1, n_t]. \tag{79a}$$

$$\mathbb{P}\Big(\underline{f} \leq H_g^{t,k} g^{t,k} - H_d^{t,k} (\hat{d}^t + \tilde{d}^t) + H_w^{t,k} (\hat{w}^t + \tilde{w}^t) \leq \overline{f} \Big) \geq 1 - \epsilon_{t,k}^{\text{TLOP}},\quad k \in [0, n_k],\ t \in [1, n_t]. \tag{79b}$$

Another interesting application of chance constraints in cc-SCUC guarantees the utilization ratio of wind generation greater than a desired threshold with high probability $1 - \epsilon$ (Wang et al., 2012; Wang, Wang, & Guan, 2013; Zhao, Wang, Wang, & Guan, 2014). Different variations of the chance constraint on wind utilization ratios can be found in Wang et al. (2012).

8.2.3. Solving chance-constrained SCUC

As mentioned in Section 8.2.2, there is no uniform formulation of chance-constrained SCUC. Many references in Table 1 concentrate on exploring alternative formulations of cc-SCUC. Therefore theoretical guarantees on the solution quality is not a major concern.

Among all the reviewed methods, sample average approximation is commonly used when solving chance-constrained SCUC (Bagheri, Zhao, & Guo, 2017; Tan & Shaaban, 2016; Wang et al., 2012; Wang et al., 2013; Zhang, Wang, Zeng, & Hu, 2017; Zhao et al., 2014). Section 6 shows that SAA reformulates (CCO) to a mixed integer program, which is difficult to solve in general. Many references apply various techniques from integer programming to speed up the computation, e.g. Zhao et al. (2014) and Jiang, Guan, and Watson (2016).

Section 5.2 shows that there is no upper bound on the number of support scenarios for non-convex problems in general. Thus, a majority of results of the scenario approach cannot be directly applied on cc-SCUC. Reference Margellos et al. (2013) might be the first attempt to solve cc-SCUC with the scenario approach. Recently, Campi et al. (2018) extends the a-posteriori guarantees of the scenario approach towards non-convex problems. Geng, Modarresi, and Xie (2019) adopts the approach in Campi et al. (2018) and shows the possibility to apply the theoretical results of the scenario approach on (cc-SCUC). It is worth mentioning that some theoretical results in robust optimization still apply in spite of the non-convexity of SCUC from integer variables (z^t, u^t, v^t), e.g. Bertsimas et al. (2018). This could be an interesting direction to explore.

8.3. Generation and transmission expansion

Generation and transmission expansion (the expansion problem in short) is a critical component in *long-term* power system planning exercises. The expansion problem answers the following critical questions: (i) when to invest on new elements such as transmission lines and generators in the system; (ii) what types of new elements are necessary; and (iii) how much capacity is needed and where the best locations would be for those new elements. A typical objective of the expansion problem is to minimize (i) total cost of investment in new generators and transmission line; (ii) environmental impacts; and (iii) cost of generation. Constraints of the expansion problem often include total or individual costs within budget, capacity constraint, reliability requirement, supply-demand balance, power flow equations, and operation requirements such as generation or transmission limits.

The expansion problem typically needs to deal with uncertainties from demand, generation and transmission outages, and renewables. Chance constraints often appear as requirements on reliability metrics such as LOLP (78a) and TLOP (78b).

Among all the papers incorporating chance constraints in the expansion problem, a majority of them assume the underlying distribution is Gaussian and derive the second order cone equivalent form as in Section 3.2, e.g. Sanghvi, Shavel, and Spann (1982), López, Ponnambalam, and Quintana (2007), Mazadi, Rosehart, Malik, and Aguado (2009) and Manickavasagam, Anjos, and Rosehart (2015). A few papers design its own simulation-based iterative algorithms because of complicated problem formulations, e.g. Yang and Wen (2005) and Qiu et al. (2016). Although Monte-Carlo simulation is typically performed to evaluate the actual feasibility, there is no rigorous guarantees on these results.

Similar to the chance constrained DCOPF problem, deriving deterministic equivalent forms is the most popular choice. Considering the expansion problem is usually ultra-large-scale and involves lots of integer variables, the simplicity of deterministic equivalent form becomes particularly attractive. Additional pros and cons of this approach are analyzed in Section 8.1.3.

Similar to chance-constrained SCUC, the expansion problem includes many integer variables and is non-convex in nature. As discussed in Section 8.2.3, the scenario approach and sample average approximation can still be applied on the expansion problem. Because of the size of the expansion problem, the required sample complexity could be astronomic, which lead to major computational issues. Although the scenario approach and sample average approximation could provide better theoretical guarantees, it is essential to overcome the major obstacles in computation to apply some better methods on the expansion problem.

9. Numerical simulations

9.1. ConvertChanceConstraint (CCC): a Matlab toolbox

Most existing optimization solvers cannot directly solve (CCO). All reviewed methods in Sections 5–7 translate (CCO) to forms that can be recognized and solved by optimization solvers, e.g. SAA converts (CCO) to a mixed integer program (MIP), which can be solved by Gurobi. When solving a chance-constrained program, a typical approach is to write the converted formulation (e.g. the MIP of SAA) in the compact format that a solver recognizes then rely on the solver to get optimal solutions. This approach is unnecessarily repetitive as it needs to be repeated by different researchers on different problems. In addition, different solvers often take various input formats, thus this typical approach is limited to one specific solver. To overcome these issues, an interface or toolbox that automatically converts (CCO) to suitable forms for a variety of solvers is needed.

The remaining part of this subsection introduces the open-source Matlab toolbox *ConvertChanceConstraint* (CCC), which is developed to automate the process of converting chance constraints. CCC is written in Matlab, one of the most popular tools in engineering and many other fields. In consideration of flexibility in modeling and compatibility with existing solvers, CCC is built on YALMIP (Löfberg, 2004), a modeling language for optimization in Matlab. CCC is open-source on Github,[8] other researchers and engineers could freely use, modify and improve it.

Fig. 4 illustrates the logic flow when using CCC to solve and analyze a chance-constrained program. The problem is first formulated in the language of Matlab and YALMIP, then the chance constraint is modeled using the *prob()* function defined in CCC. After receiving the problem formulation and specified method to use (e.g. scenario approach), CCC translates the chance constraint to the formulation that YALMIP could understand. Then YALMIP interfaces with various solvers and further translates the problem for a specific solver. After optimization solver returns the optimal solution, CCC provides a few functions for result analysis, e.g. checking out-of-sample violation probability, calculating the posterior guarantees of the scenario approach.

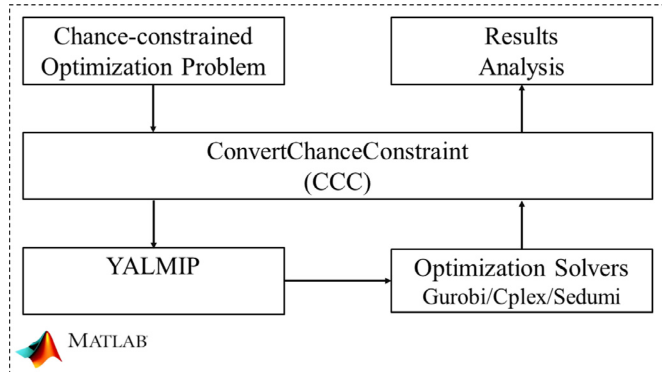


Fig. 4. Solving and analyzing a chance-constrained program via CCC.

Fig. 5 presents the structure and main functions of CCC. Three major methods to solve (CCO) are implemented: scenario approach, sample average approximation and robust optimization related methods. The implementation of RO-related methods is based on the robust optimization module (Löfberg, 2012) of YALMIP. As illustrated in Fig. 4 and 5, CCC is interfaced via YALMIP with most existing optimization solvers, e.g. Cplex (CPLEX, 2009), Gurobi (Gurobi Optimization, 2016), Mosek (Mosek, 2015) and Sedumi (Sturm, 1999).

9.2. Simulation settings

Chance-constrained DCOPF (74) serves as a benchmark problem for a critical comparison of solutions to (CCO). We provide numerical solutions of cc-DCOPF on two test systems: a 3-bus system and the IEEE 24-bus RTS test system.

[8] https://github.com/xb00dx/ConvertChanceConstraint-ccc.

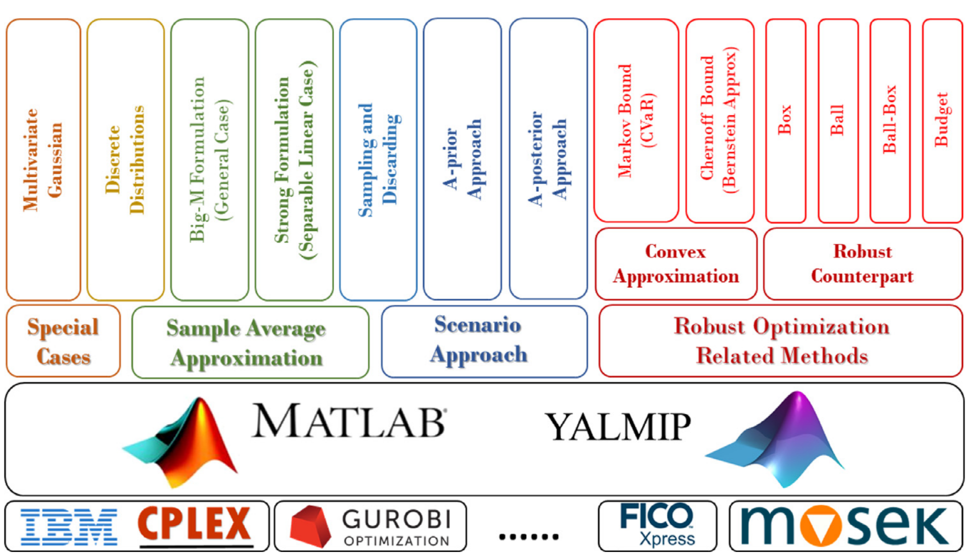


Fig. 5. Structure and main functions of ConvertChanceConstraint.

The 3-bus system is a modified version of the 3-bus system in Lesieutre, Molzahn, Borden, and DeMarco (2011). The major difference is the removal of the load at bus 2 and the synchronous condensor at bus 3 in order to visualize the feasible region and the space of uncertainties. The original 3-bus system "*case3sc.m*" is available in the Matpower toolbox (Zimmerman, Murillo-Snchez, & Thomas, 2011). The modified system in this paper can be found in the examples of CCC.[9] For simplicity, we only consider uncertainties of loads, which is modeled as Gaussian variables with 5% standard variation.

The 24-bus system in this paper is a modified version of the IEEE 24-bus RTS benchmark system (Grigg, Wong, Billinton, & others, 1999). The transmission line capacities are set to be 60% of the original capacities. We conduct two sets of simulations on the 24-bus system with different distributions of uncertainties. The first one is similar with the 3-bus case, nodal loads are modeled as independent Gaussian variables with 5% standard deviation. The second one models the errors of nodal load forecasts as independent beta-distributed random variables, with parameters $\alpha = 25.2414$ and $\beta = 25.2692$.[10]

Ten Monte-Carlo simulations are conducted on every method to examine the randomness of solutions. For the 3-bus case, each Monte-Carlo simulation uses 100 i.i.d samples to solve cc-DCOPF. 2048 points are used in each run to solve (cc-DCOPF) of the 24-bus system. The returned solutions are evaluated on an independent set of 10^4 points (Fig. 6).

We use Gurobi 7.10 (Gurobi Optimization, 2016) to get results of scenario approach and sample average approximation. Cplex 12.8 is used to solve (CCO) with robust counterpart and convex approximation.

9.3. Simulation results

We solve cc-DCOPF on the 3-bus system with eight different methods: (1) scenario approach with prior guarantees, (SA:prior, Corollary 1); (2) scenario approach with posterior guarantees (SA:posterior, Theorem 7); (3) sample average approximation,

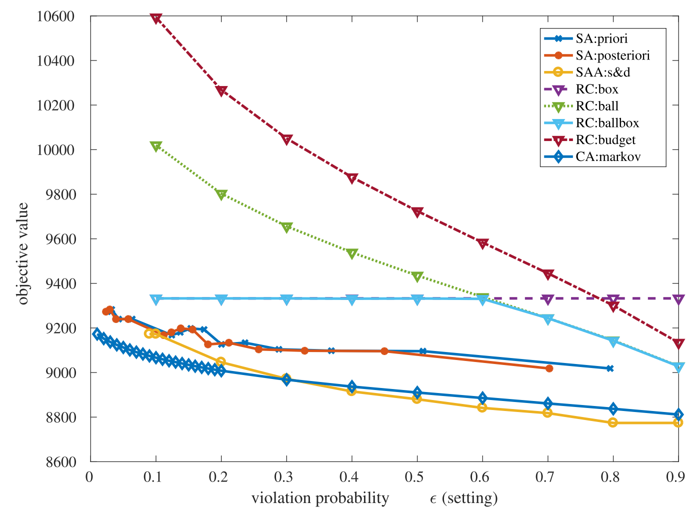


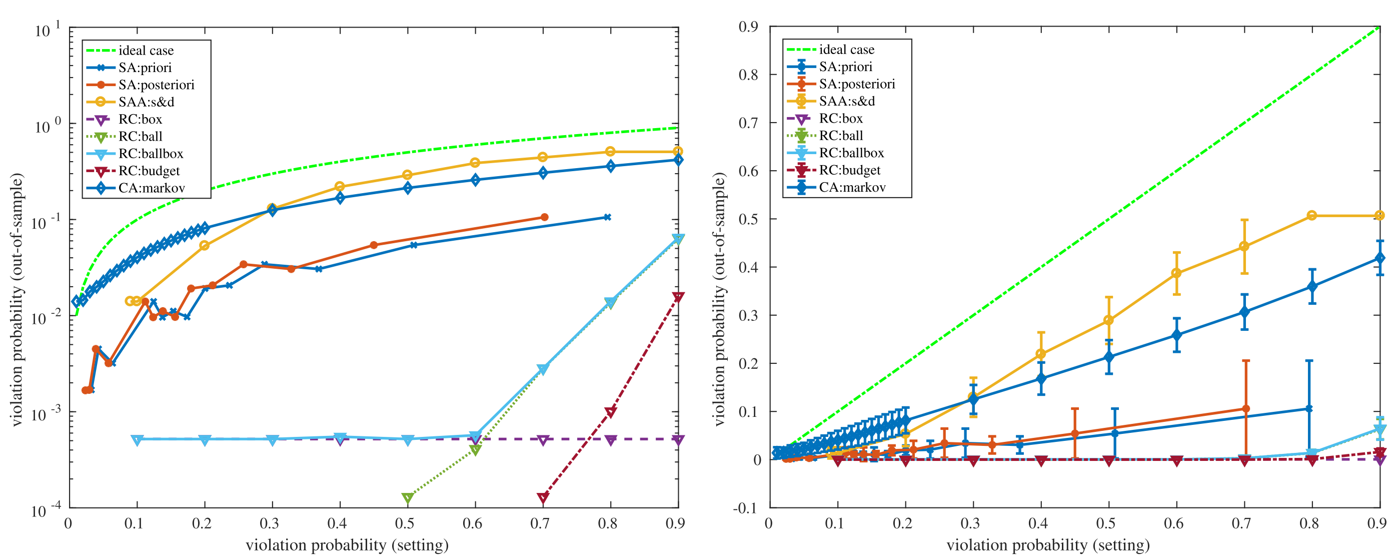
Fig. 6. Objective values (cc-DCOPF of the 3-bus System).

where N and ε are chosen based on the sampling and discarding Theorem (SAA:s&d, Theorem 9); (4–7) Robust counterpart with different uncertainty sets specified in Theorem 13: box (RC:box), ball (RC:ball), ball-box (RC:ball-box) and budget (RC:budget) uncertainty sets; (8) convex approximation with Markov bound (CA:markov, Theorem 11 and Proposition 5).

We first examine the feasibility of the returned solutions from eight algorithms. Figs. 7 and 8 show the out-of-sample violation probabilities $\hat{\epsilon}$ versus desired ϵ in the setting. The green dashed lines in Figs. 7 and 8 denote the ideal case where $\hat{\epsilon} = \epsilon$. Any points above the green dashed line indicate infeasible solutions that $\mathbb{V}(x) > \epsilon$. Clearly all methods return feasible solutions (with high probability) to (CCO). From Fig. 7, sample average approximation and convex approximation are less conservative than other methods. However, it is worth noting that when ϵ is small (e.g. 10^{-2}), the data-driven approximation of CVaR (Proposition 5) does not necessarily give a safe approximation to (CCO) (Chen et al., 2010). The robust counterpart methods are typically 10 ∼ 100

[9] github.com/xb00dx/ConvertChanceConstraint-ccc/tree/master/examples.

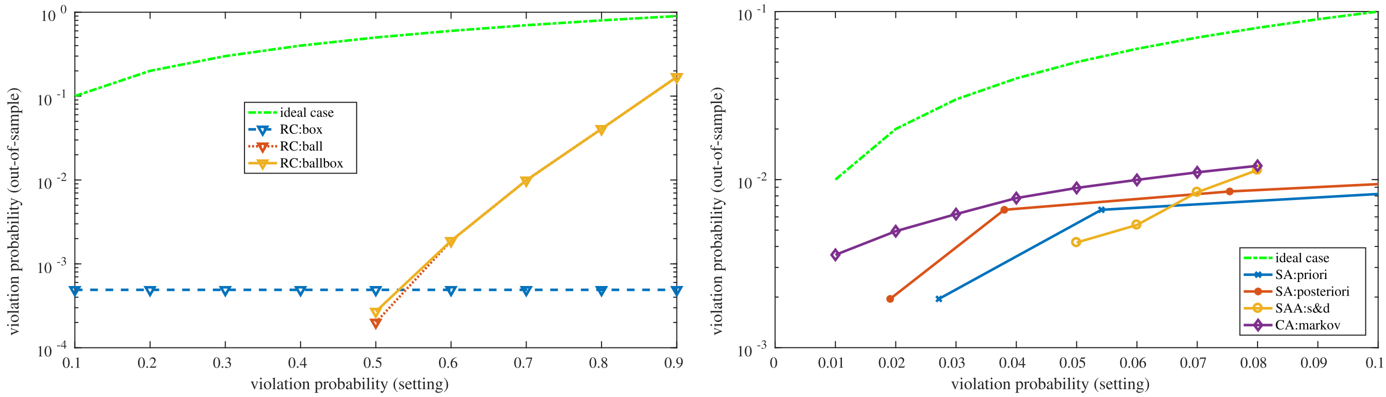
[10] This setting of beta distribution is from Hodge and Milligan (2011), and scaled from [0,1] to [−18%, 18%].



(a) with logarithmic y-axis

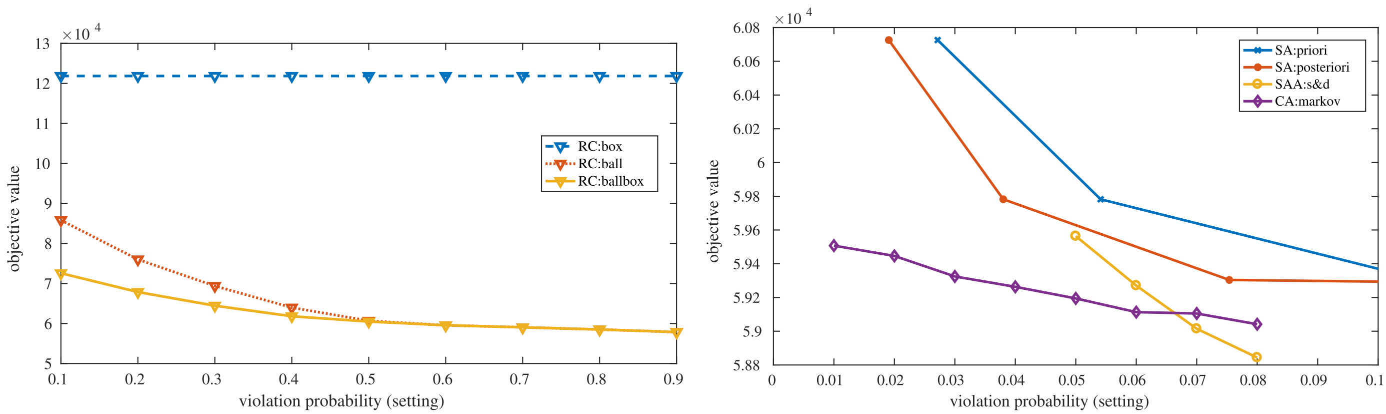
(b) with error bars showing standard deviations

Fig. 7. Violation probabilities (cc-DCOPF of the 3-bus System).



(a) robust counterpart methods (b) other methods

Fig. 8. Violation probabilities (cc-DCOPF of the 24-bus System, Gaussian Distributions).



(a) robust counterpart methods (b) other methods

Fig. 9. Objective values (cc-DCOPF of the 24-bus System, Gaussian Distributions).

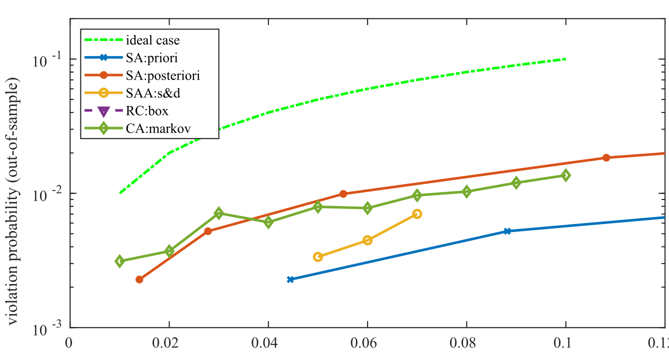


Fig. 10. Violation probabilities in logarithmic scale (cc-DCOPF of the 24-bus System, Beta Distributions).

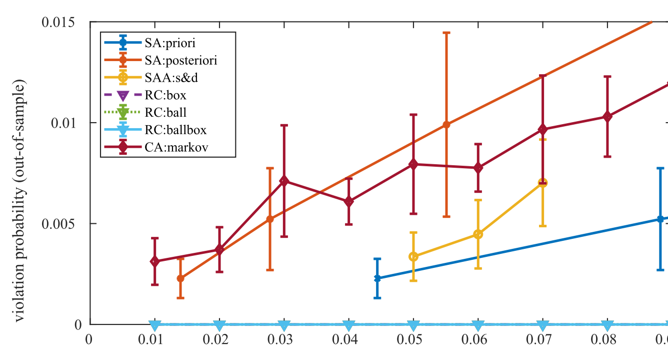


Fig. 11. Violation probabilities with error bars showing standard deviations (cc-DCOPF of the 24-bus System, Beta Distributions).

times more conservative than other methods, as illustrated in the comparison of Fig. 8a with Fig. 8b. The conservativeness could be significantly reduced by better construction of uncertainty sets, e.g. Chen et al. (2010) and Bertsimas et al. (2018). Among four different choices of uncertainty sets, the ball-box set is the least conservative one, which combines the advantages of the ball and box uncertainty sets.

Figs. 8 and 9 present the results of the 24-bus system with Gaussian distributions. Simulation results of the beta distribution are in Figs. 10–12. Observations from Figs. 10–12 are similar with the case of Gaussian distributions. Every method behaves more conservative in the case of beta distributions than the case of Gaussian distributions. It is worth noting that the RO-based methods (RC:box, RC:ball, RC:ball-box in Fig. 11) are so conservative that lead to zero empirical violation probability $\hat{\epsilon}$.

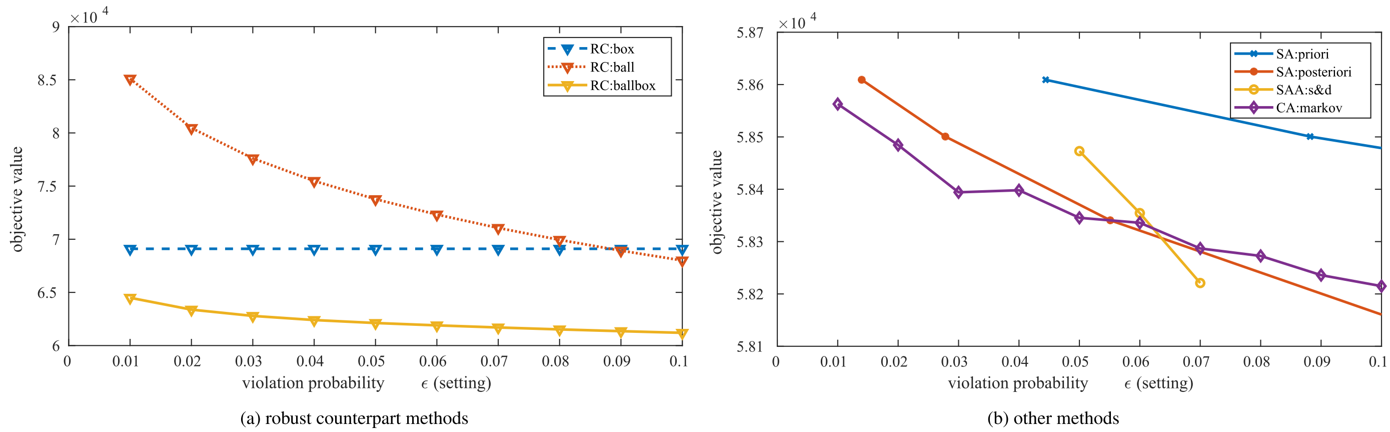


Fig. 12. Objective values (cc-DCOPF of the 24-bus System, Beta Distributions).

10. Concluding remarks

This paper consists of two parts. The first part presents a comprehensive review on the fundamental properties, key theoretical results, and three classes of algorithms for chance-constrained optimization. An open-source MATLAB toolbox ConvertChanceConstraint is developed to automate the process of translating chance constraints to compatible forms for mainstream optimization solvers. The second part of this paper presents three major applications of chance-constrained optimization in power systems. We also present a critical comparison of existing algorithms to solve chance-constrained programs on IEEE benchmark systems.

Many interesting directions are open for future research. More thorough and detailed comparisons of solutions to (CCO) on various problems with realistic datasets is needed. In terms of theoretical investigation, an analytical comparison of existing solutions to chance-constrained optimization is necessary to substantiate the fundamental insights obtained from numerical simulations. In terms of applications, many existing results can be improved by exploiting the structural properties of the problem to be solved. The application of chance-constrained optimization in electric energy systems could go beyond operational planning practices. For example, it would be worth investigating into the economic interpretation of market power issues through the lens of chance-constrained optimization.

Declaration of competing interest

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from xbgeng@tamu.edu.

Signed by all authors as follows:

Xinbo Geng, Le Xie, January 31, 2019.

References

Van Ackooij, W. (2015). Eventual convexity of chance constrained feasible sets. *Optimization, 64*(5), 1263–1284.

Ahmed, S. (2018). *Relaxations and approximations of chance constraints*. https://ismp2018.sciencesconf.org/.

Ahmed, S., & Shapiro, A. (2008). Solving chance-constrained stochastic programs via sampling and integer programming. *Tutorials in Operations Research, 10*, 261–269.

Allan, R. N., & others (2013). *Reliability evaluation of power systems*. Springer Science & Business Media.

Anese, E. D., Baker, K., & Summers, T. (2017). Chance-constrained AC optimal power flow for distribution systems with renewables. *IEEE Transactions on Power Systems*.

Bagheri, A., Zhao, C., & Guo, Y. (2017). Data-driven chance-constrained stochastic unit commitment under wind power uncertainty. In *Power & energy society general meeting, 2017 IEEE* (pp. 1–5). IEEE.

Ben-Tal, A., Bhadra, S., Bhattacharyya, C., & Saketha Nath, J. (2011). Chance constrained uncertain classification via robust optimization. *Mathematical Programming, 127*(1), 145–173. doi:10.1007/s10107-010-0415-1.

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

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

Ben-Tal, A., & Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters, 25*(1), 1–13.

Ben-Tal, A., Nemirovski, A., & Roos, C. (2002). Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization, 13*(2), 535–560.

Bent, R., Bienstock, D., & Chertkov, M. (2013). Synchronization-aware and algorithm–efficient chance constrained optimal power flow. In *Bulk power system dynamics and control-IX optimization, security and control of the emerging power grid (IREP), 2013 IREP symposium* (pp. 1–11). IEEE.

Beraldi, P., & Ruszczyński, A. (2002). The probabilistic set-covering problem. *Operations Research, 50*(6), 956–967.

Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review, 53*(3), 464–501.

Bertsimas, D., Gupta, V., & Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*.

Bertsimas, D., Litvinov, E., Sun, X. A., Zhao, J., & Zheng, T. (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems, 28*(1), 52–63.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research, 52*(1), 35–53.

Bertsimas, D., & Sim, M. (2006). Tractable approximations to robust conic optimization problems. *Mathematical Programming, 107*(1–2), 5–36.

Bienstock, D., Chertkov, M., & Harnett, S. (2013). Robust modeling of probabilistic uncertainty in smart grids: Data ambiguous chance constrained optimum power flow. In *Decision and control (CDC), 2013 IEEE 52nd annual conference on* (pp. 4335–4340). IEEE.

Bienstock, D., Chertkov, M., & Harnett, S. (2014). Chance-constrained optimal power flow: Risk-aware network control under uncertainty. *SIAM Review, 56*(3), 461–495.

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

Bucher, M. A., Vrakopoulou, M., & Andersson, G. (2013). Probabilistic N- 1 security assessment incorporating dynamic line ratings. In *Power and energy society general meeting (PES), 2013 IEEE* (pp. 1–5). IEEE.

Calafiore, G., & Campi, M. C. (2005). Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming, 102*(1), 25–46.

Calafiore, G. C. (2010). Random convex programs. *SIAM Journal on Optimization, 20*(6), 3427–3464.

Calafiore, G. C., Campi, M. C., et al. (2006). The scenario approach to robust control design. *IEEE Transactions on Automatic Control*.

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

Campi, M., & Garatti, S. (2016). Wait-and-judge scenario optimization. *Mathematical Programming*, 1–35.

Campi, M. C., & Garatti, S. (2008). The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization, 19*(3), 1211–1230.

Campi, M. C., & Garatti, S. (2011). A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications, 148*(2), 257–280.

Campi, M. C., Garatti, S., & Prandini, M. (2009). The scenario approach for systems and control design. *Annual Reviews in Control, 33*(2), 149–157.

Campi, M. C., Garatti, S., & Ramponi, F. A. (2018). A general scenario theory for non-convex optimization and decision making. *IEEE Transactions on Automatic Control*.

Caramanis, C., Mannor, S., & Xu, H. (2012). 14 Robust optimization in machine learning. *Optimization for Machine Learning*, 369.

Charnes, A., & Cooper, W. W. (1959). Chance-constrained programming. *Management Science, 6*(1), 73–79.

Charnes, A., & Cooper, W. W. (1963). Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research*.

Charnes, A., Cooper, W. W., & Symonds, G. H. (1958). Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science, 4*(3), 235–263.

Chen, W., & Sim, M. (2009). Goal-driven optimization. *Operations Research, 57*(2), 342–357.

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

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

CPLEXI. I. (2009). V12. 1: User's manual for CPLEX. *International Business Machines Corporation, 46*(53), 157.

Ding, X., Lee, W.-J., Jianxue, W., & Liu, L. (2010). Studies on stochastic unit commitment formulation with flexible generating units. *Electric Power Systems Research, 80*(1), 130–141.

Doostizadeh, M., Aminifar, F., Ghasemi, H., & Lesani, H. (2016). Energy and reserve scheduling under wind power uncertainty: An adjustable interval approach. *IEEE Transactions on Smart Grid, 7*(6), 2943–2952.

Dupačová, J., Gaivoronski, A., Kos, Z., & Szantai, T. (1991). Stochastic programming in water management: A case study and a comparison of solution techniques. *European Journal of Operational Research, 52*(1), 28–44.

Franco, J. F., Rider, M. J., & Romero, R. (2016). Robust multi-stage substation expansion planning considering stochastic demand. *IEEE Transactions on Power Systems, 31*(3), 2125–2134.

Föllmer, H., & Schied, A. (2011). *Stochastic finance: An introduction in discrete time*. Walter de Gruyter.

Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European Journal of Operational Research, 235*(3), 471–483.

Geng, X., Modarresi, S., & Xie, L. (2019). Security-constrained unit commitment with probabilistic guarantees. *IEEE Transactions on Power Systems (submitted)*.

Geng, X., & Xie, L. (2019a). *Data-driven decision making with probabilistic guarantees (Part 1): A schematic overview of chance-constrained optimization* arXiv:1903.10621.

Geng, X., & Xie, L. (2019b). *Data-driven decision making with probabilistic guarantees (Part 2): Applications of chance-constrained optimization in power systems*.

Grigg, C., Wong, P., Billinton, R., & others (1999). The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee. *IEEE Transactions on Power Systems, 14*(3), 1010–1020.

Gurobi Optimization I. (2016). *Gurobi optimizer reference manual*.

Henrion, R., Li, P., Möller, A., Steinbach, M. C., Wendt, M., & Wozny, G. (2001). Stochastic optimization for operating chemical processes under uncertainty. In *Online optimization of large scale systems* (pp. 457–478). Springer.

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

Henrion, R., & Strugarek, C. (2011). Convexity of chance constraints with dependent random variables: the use of copulae. In *Stochastic optimization methods in finance and energy* (pp. 427–439). Springer.

Hodge, B.-M., & Milligan, M. (2011). Wind power forecasting error distributions over multiple timescales. In *Power and energy society general meeting, 2011 IEEE* (pp. 1–8). IEEE.

Hreinsson, K., Vrakopoulou, M., & Andersson, G. (2015). Stochastic security constrained unit commitment and non-spinning reserve allocation with performance guarantees. *International Journal of Electrical Power & Energy Systems, 72*, 109–115.

Jabr, R. A. (2013). Adjustable robust OPF with renewable energy sources. *IEEE Transactions on Power Systems, 28*(4), 4742–4751.

Jiang, R., Guan, Y., & Watson, J.-P. (2016). Cutting planes for the multistage stochastic unit commitment problem. *Mathematical Programming, 157*(1), 121–151.

Jiang, R., Wang, J., & Guan, Y. (2012). Robust unit commitment with wind power and pumped storage hydro. *IEEE Transactions on Power Systems, 27*(2), 800.

Kataoka, S. (1963). A stochastic programming model. *Econometrica: Journal of the Econometric Society*, 181–196.

Ke, D., Chung, C., & Sun, Y. (2016). A novel probabilistic optimal power flow model with uncertain wind power generation described by customized Gaussian mixture model. *IEEE Transactions on Sustainable Energy, 7*(1), 200–212.

Khachiyan, L. (1989). The problem of calculating the volume of a polyhedron is enumerably hard. *Russian Mathematical Surveys*.

Kress, M., Penn, M., & Polukarov, M. (2007). The minmax multidimensional knapsack problem with application to a chance-constrained problem. *Naval Research Logistics (NRL), 54*(6), 656–666.

Küükyavuz, S. (2012). On mixing sets arising in chance-constrained programming. *Mathematical Programming, 132*(1–2), 31–56.

Lagoa, C. (1999). On the convexity of probabilistically constrained linear programs. In *Decision and control, 1999. Proceedings of the 38th IEEE conference on: 1* (pp. 516–521). IEEE.

Lesieutre, B. C., Molzahn, D. K., Borden, A. R., & DeMarco, C. L. (2011). Examining the limits of the application of semidefinite programming to power flow problems. In *Communication, control, and computing (Allerton), 2011 49th annual allerton conference on* (pp. 1492–1499). IEEE.

Li, B., & Mathieu, J. L. (2015). Analytical reformulation of chance-constrained optimal power flow with uncertain load control. In *PowerTech, 2015 IEEE Eindhoven* (pp. 1–6). IEEE.

Li, B., Vrakopoulou, M., & Mathieu, J. L. (2019). Chance constrained reserve scheduling using uncertain controllable loads Part II: Analytical reformulation. *IEEE Transactions on Smart Grid*.

López, J., Pozo, D., Contreras, J., & Mantovani, J. (2015). A convex chance-constrained model for reactive power planning. *International Journal of Electrical Power & Energy Systems, 71*, 403–411.

López, J. Ã., Ponnambalam, K., & Quintana, V. H. (2007). Generation and transmission expansion under risk using stochastic programming. *IEEE Transactions on Power Systems, 22*(3), 1369–1378.

Lubin, M., Dvorkin, Y., & Backhaus, S. (2016). A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Transactions on Power Systems, 31*(5), 3840–3849.

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

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

Löfberg, J. (2004). YALMIP: A Toolbox for Modeling and Optimization in MATLAB. In *In Proceedings of the CACSD conference. Taipei, Taiwan*.

Löfberg, J. (2012). Automatic robust convex programming. *Optimization Methods and Software, 27*(1), 115–129.

Manickavasagam, M., Anjos, M. F., & Rosehart, W. D. (2015). Sensitivity-based chance-constrained generation expansion planning. *Electric Power Systems Research, 127*, 32–40.

Margellos, K., Rostampour, V., Vrakopoulou, M., Prandini, M., Andersson, G., & Lygeros, J. (2013). Stochastic unit commitment and reserve scheduling: A tractable formulation with probabilistic certificates. In *Control conference (ECC), 2013 European* (pp. 2513–2518). IEEE.

Martínez, G., & Anderson, L. (2015). A risk-averse optimization model for unit commitment problems. In *System sciences (HICSS), 2015 48th Hawaii international conference on* (pp. 2577–2585). IEEE.

Mazadi, M., Rosehart, W., Malik, O., & Aguado, J. (2009). Modified chance-constrained optimization applied to the generation expansion problem. *IEEE Transactions on Power Systems, 24*(3), 1635–1636.

Miller, B. L., & Wagner, H. M. (1965). Chance constrained programming with joint constraints. *Operations Research, 13*(6), 930–945.

Ming, H., Xie, L., Campi, M., Garatti, S., & Kumar, P. (2017). Scenario-based economic dispatch with uncertain demand response. *IEEE Transactions on Smart Grid*.

Modarresi, M. S., Xie, L., Campi, M., Garatti, S., Caré, A., Thatte, A., & Kumar, P. (2018). Scenario-based economic dispatch with tunable risk levels in high-renewable power systems. *IEEE Transactions on Power Systems*.

Mosek, A. (2015). *The MOSEK optimization toolbox for MATLAB manual*. Version.

Mühlpfordt, T., Faulwasser, T., Roald, L., & Hagenmeyer, V. (2017). Solving optimal power flow with non-Gaussian uncertainties via polynomial chaos expansion. In *Decision and control (CDC), 2017 IEEE 56th annual conference on* (pp. 4490–4496). IEEE.

Nemirovski, A. (2012). On safe tractable approximations of chance constraints. *European Journal of Operational Research, 219*(3), 707–718.

Nemirovski, A., & Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*.

Ozturk, U. A., Mazumdar, M., & Norman, B. A. (2004). A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Transactions on Power Systems, 19*(3), 1589–1598.

Pagnoncelli, B., Ahmed, S., & Shapiro, A. (2009). Sample average approximation method for chance constrained programming: theory and applications. *Journal of Optimization Theory and Applications*.

Pintér, J. (1989). Deterministic approximations of probability inequalities. *Zeitschrift für Operations-Research, 33*(4), 219–239.

Pozo, D., & Contreras, J. (2013). A chance-constrained unit commitment with an $n-k$ security criterion and significant wind generation. *IEEE Transactions on Power systems, 28*(3), 2842–2851.

Prékopa, A. (1971). Logarithmic concave measures with application to stochastic programming. *Acta Scientiarum Mathematicarum, 32*, 301–316.

Prékopa, A. (1995). *Stochastic programming*.

Prekopa, A., Vizvari, B., & Badics, T. (1998). Programming under probabilistic constraint with discrete random variable. In *New trends in mathematical programming* (pp. 235–255). Springer.

Prékopa, A., Yoda, K., & Subasi, M. M. (2011). Uniform quasi-concavity in probabilistic constrained stochastic programming. *Operations Research Letters, 39*(3), 188–192.

Qiu, F., Ahmed, S., Dey, S. S., & Wolsey, L. A. (2014). Covering linear programming with violations. *INFORMS Journal on Computing, 26*(3), 531–546.

Qiu, J., Dong, Z. Y., Zhao, J., Xu, Y., Luo, F., & Yang, J. (2016). A risk-based approach to multi-stage probabilistic transmission network planning. *IEEE Transactions on Power Systems, 31*(6), 4867–4876.

Roald, L., & Andersson, G. (2017). *Chance-constrained AC optimal power flow: Reformulations and efficient algorithms* arXiv:1706.03241v1.

Roald, L., Misra, S., Chertkov, M., & Andersson, G. (2015). Optimal power flow with weighted chance constraints and general policies for generation control. In *Decision and control (CDC), 2015 IEEE 54th annual conference on* (pp. 6927–6933). IEEE.

Roald, L., Misra, S., Krause, T., & Andersson, G. (2017). Corrective control to handle forecast uncertainty: A chance constrained optimal power flow. *IEEE Transactions on Power Systems, 32*(2), 1626–1637.

Roald, L., Misra, S., Morrison, A., & Andersson, G. (2017). Optimized risk limits for stochastic optimal power flow. In *Decision and control (CDC), 2017 IEEE 56th annual conference on* (pp. 4476–4483). IEEE.

Roald, L., Oldewurtel, F., Krause, T., & Andersson, G. (2013). Analytical reformulation of security constrained optimal power flow with probabilistic constraints. In *PowerTech (POWERTECH), 2013 IEEE Grenoble* (pp. 1–6). IEEE.

Roald, L., Vrakopoulou, M., Oldewurtel, F., & Andersson, G. (2014). Risk-constrained optimal power flow with probabilistic guarantees. In *Power systems computation conference (PSCC), 2014* (pp. 1–7). IEEE.

Roald, L., Vrakopoulou, M., Oldewurtel, F., & Andersson, G. (2015). Risk-based optimal power flow with probabilistic guarantees. *International Journal of Electrical Power & Energy Systems, 72*, 66–74.

Rockafellar, R. T. (2015). *Convex analysis*. Princeton University Press.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk, 2*, 21–42.

Ruszczyński, A. (2002). Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming, 93*(2), 195–215.

Ruszczynski, A. P., & Shapiro, A. (2003). *Stochastic programming*: 10. Elsevier Amsterdam.

Sahinidis, N. V. (2004). Optimization under uncertainty: State-of-the-art and opportunities. *Computers & Chemical Engineering, 28*(6–7), 971–983.

Sanghvi, A. P., Shavel, I. H., & Spann, R. M. (1982). Strategic planning for power system reliability and vulnerability: An optimization model for resource planning under uncertainty. *IEEE Transactions on Power Apparatus and Systems*, (6), 1420–1429.

Schildbach, G., Fagiano, L., & Morari, M. (2013). Randomized solutions to convex programs with multiple chance constraints. *SIAM Journal on Optimization, 23*(4), 2479–2501.

Sen, S. (1992). Relaxations for probabilistically constrained programs with discrete random variables. In *System modelling and optimization* (pp. 598–607). Springer.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on stochastic programming: Modeling and theory*. SIAM.

Sra, S., Nowozin, S., & Wright, S. J. (2012). *Optimization for machine learning*. Mit Press.

Sturm, J. F. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software, 11*(1–4), 625–653.

Summers, T., Warrington, J., Morari, M., & Lygeros, J. (2014). Stochastic optimal power flow based on convex approximations of chance constraints. In *Power Systems Computation Conference (PSCC)*. IEEE.

Summers, T., Warrington, J., Morari, M., & Lygeros, J. (2015). Stochastic optimal power flow based on conditional value at risk and distributional robustness. *International Journal of Electrical Power & Energy Systems, 72*, 116–125.

Takriti, S., Birge, J. R., & Long, E. (1996). A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems, 11*(3), 1497–1508.

Tan, W.-S., & Shaaban, M. (2016). A hybrid stochastic/deterministic unit commitment based on projected disjunctive milp reformulation. *IEEE Transactions on Power Systems, 31*(6), 5200–5201.

Tanner, M. W., & Ntaimo, L. (2010). IIS branch-and-cut for joint chance-constrained stochastic programs and application to optimal vaccine allocation. *European Journal of Operational Research*.

Venzke, A., Halilbasic, L., Markovic, U., Hug, G., & Chatzivasileiadis, S. (2017). *Convex relaxations of chance constrained AC optimal power flow* arXiv:1702.08372v1.

Vrakopoulou, M., Katsampani, M., Margellos, K., Lygeros, J., & Andersson, G. (2013). Probabilistic security-constrained ac optimal power flow. *PowerTech (POWERTECH)*. IEEE.

Vrakopoulou, M., Li, B., & Mathieu, J. L. (2019). Chance constrained reserve scheduling using uncertain controllable loads Part I: Formulation and scenario-based analysis. *IEEE Transactions on Smart Grid*.

Vrakopoulou, M., Margellos, K., Lygeros, J., & Andersson, G. (2013a). A probabilistic framework for reserve scheduling and security assessment of systems with high wind power penetration. *IEEE Transactions on Power Systems, 28*(4), 3885–3896.

Vrakopoulou, M., Margellos, K., Lygeros, J., & Andersson, G. (2013b). Probabilistic guarantees for the N-1 security of systems with wind power generation. In *Reliability and risk evaluation of wind integrated power systems* (pp. 59–73). Springer.

Wang, Q., Guan, Y., & Wang, J. (2012). A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output. *IEEE Transactions on Power Systems, 27*(1), 206–215.

Wang, Q., Wang, J., & Guan, Y. (2013). Price-based unit commitment with wind power utilization constraints. *IEEE Transactions on Power Systems, 28*(3), 2718–2726.

Wang, Z., Shen, C., Liu, F., Wu, X., Liu, C.-C., & Gao, F. (2017). Chance-constrained economic dispatch with non-Gaussian correlated wind power uncertainty. *IEEE Transactions on Power Systems, 32*(6), 4880–4893.

Wu, H., Shahidehpour, M., Li, Z., & Tian, W. (2014). Chance-constrained day-ahead scheduling in stochastic power system operation. *IEEE Transactions on Power Systems, 29*(4), 1583–1591.

Wu, L., Shahidehpour, M., & Li, T. (2007). Stochastic security-constrained unit commitment. *IEEE Transactions on Power Systems, 22*(2), 800–811. doi:10.1109/TPWRS.2007.894843.

Wu, Z., Zeng, P., Zhang, X.-P., & Zhou, Q. (2016). A solution to the chance-constrained two-stage stochastic program for unit commitment with wind energy integration. *IEEE Transactions on Power Systems, 31*(6), 4185–4196.

Xie, L., Carvalho, P. M., Ferreira, L. A., Liu, J., Krogh, B. H., Popli, N., & Ilic, M. D. (2011). Wind integration in power systems: Operational challenges and possible solutions. *Proceedings of the IEEE, 99*(1), 214–232.

Xu, H., Caramanis, C., & Mannor, S. (2009). Robust regression and lasso. In *Advances in neural information processing systems* (pp. 1801–1808).

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies, 32*(2), 137–150.

Yang, N., & Wen, F. (2005). A chance constrained programming approach to transmission system expansion planning. *Electric Power Systems Research, 75*(2–3), 171–177.

Yang, P., & Nehorai, A. (2014). Joint optimization of hybrid energy storage and generation capacity with renewable energy. *IEEE Transactions on Smart Grid, 5*(4), 1566–1574.

Zeng, B., & An, Y. (2014). Solving bilevel mixed integer program by reformulations and decomposition. *Optimization Online*, 1–34.

Zeng, B., An, Y., & Kuznia, L. (2017). Chance constrained mixed integer program: Bilinear and linear formulations, and Benders decomposition. *Mathematical Programming*.

Zhang, X., Grammatico, S., Schildbach, G., Goulart, P., & Lygeros, J. (2015). On the sample size of random convex programs with structured dependence on the uncertainty. *Automatica, 60*, 182–188.

Zhang, Y., & Giannakis, G. B. (2013). Robust optimal power flow with wind integration using conditional value-at-risk. In *Smart grid communications (SmartGridComm), 2013 IEEE international conference on* (pp. 654–659). IEEE.

Zhang, Y., Shen, S., & Mathieu, J. (2017). Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads. *IEEE Transactions on Power Systems*.

Zhang, Y., Shen, S., & Mathieu, J. L. (2015). Data-driven optimization approaches for optimal power flow with uncertain reserves from load control. In *American control conference (ACC), 2015* (pp. 3013–3018). IEEE.

Zhang, Y., Wang, J., Li, Y., & Cao, X. (2017). *Chance-constrained transmission expansion planning with guaranteed wind power utilization. Chicago, IL.*

Zhang, Y., Wang, J., Zeng, B., & Hu, Z. (2017). Chance-constrained two-stage unit commitment under uncertain load and wind power output using bilinear benders decomposition. *IEEE Transactions on Power Systems*.

Zhao, C., Wang, Q., Wang, J., & Guan, Y. (2014). Expected value and chance constrained stochastic unit commitment ensuring wind power utilization. *IEEE Transactions on Power Systems, 29*(6), 2696–2705.

Zheng, Q. P., Wang, J., & Liu, A. L. (2015). Stochastic optimization for unit commitment— review. *IEEE Transactions on Power Systems, 30*(4), 1913–1924.

Zimmerman, R. D., Murillo-Snchez, C. E., & Thomas, R. J. (2011). MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems, 26*(1), 12–19.