Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming

Chao Ning, Fengqi You*

Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA

ABSTRACT

This paper reviews recent advances in the field of optimization under uncertainty via a modern data lens, highlights key research challenges and promise of data-driven optimization that organically integrates machine learning and mathematical programming for decision-making under uncertainty, and identifies potential research opportunities. A brief review of classical mathematical programming techniques for hedging against uncertainty is first presented, along with their wide spectrum of applications in Process Systems Engineering. A comprehensive review and classification of the relevant publications on data-driven distributionally robust optimization, data-driven chance constrained program, data-driven robust optimization, and data-driven scenario-based optimization is then presented. This paper also identifies fertile avenues for future research that focuses on a closed-loop data-driven optimization framework, which allows the feedback from mathematical programming to machine learning, as well as scenario-based optimization leveraging the power of deep learning techniques. Perspectives on online learning-based data-driven multistage optimization with a learning-while-optimizing scheme are presented.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Optimization applications abound in many areas of science and engineering (Biegler and Grossmann, 2004; Grossmann and Biegler, 2004; Sakizlis et al., 2004). In real practice, some parameters involved in optimization problems are subject to uncertainty due to a variety of reasons, including estimation errors and unexpected disturbance (Sahinidis, 2004). Such uncertain parameters can be product demands in process planning (Liu and Sahinidis, 1996), kinetic constants in reaction-separation-recycling system design (Acevedo and Pistikopoulos, 1998), and task durations in batch process scheduling (Li and Ierapetritou, 2008), among others. The issue of uncertainty could unfortunately render the solution of a deterministic optimization problem (i.e. the one disregarding uncertainty) suboptimal or even infeasible (Ben-Tal and Nemirovski, 2002). The infeasibility, i.e. the violation of constraints in optimization problems, has a disastrous consequence on the solution quality. Motivated by the practical concern, optimization under uncertainty has attracted tremendous attention from both academia and industry (Sahinidis, 2004; Pistikopoulos, 1995).

In the era of big data and deep learning, intelligent use of data has a great potential to benefit many areas. Although there is no rigorous definition of big data (John Walker, 2014), people typically characterize big data with five Vs, namely, volume, velocity, variety, veracity and value (Yin and Kaynak, 2015). Torrents of data are routinely collected and archived in process industries, and these data are becoming an increasingly important asset in process control, operations and design (Qin, 2014; Yin et al., 2015; Venkatasubramanian, 2019). Nowadays, a wide array of emerging machine learning tools can be leveraged to analyze data and extract accurate, relevant, and useful information to facilitate knowledge discovery and decision-making. Deep learning, one of the most rapidly growing machine learning subfields, demonstrates remarkable power in deciphering multiple layers of representations from raw data without any domain expertise in designing feature extractors (Goodfellow et al., 2016). More recently, dramatic progress of mathematical programming (Grossmann, 2012), coupled with recent advances in machine learning (Jordan and Mitchell, 2015), especially in deep learning over the past decade (LeCun et al., 2015), sparks a flurry of interest in data-driven optimization (Bertsimas et al., 2018; Bertsimas and Thiele, 2006; Calfa et al., 2015; Calfa et al., 2015; Campbell and How, 2015; Jiang and Guan, 2015; Ning and You, 2017a; Levi et al., 2015). In the data-driven optimization paradigm, uncertainty model is formulated based on data, thus allowing uncertainty data "speak" for themselves in the optimization algorithm. In this way, rich

* Corresponding author.
  E-mail address: fengqi.you@cornell.edu (F. You).

information underlying uncertainty data can be harnessed in an automatic manner for smart and data-driven decision making.

In this review paper, we summarize and classify the existing contributions of data-driven optimization under uncertainty, highlight the current research trends, point out the research challenges, and introduce promising methodologies that can be used to tackle these challenges. We briefly review conventional mathematical programming techniques for hedging against uncertainty, alongside their wide spectrum of applications in Process Systems Engineering (PSE). We then summarize the existing research papers on data-driven optimization under uncertainty and classify them into four categories according to their unique approach for uncertainty modeling and distinct optimization structures. Based on the literature survey, we identify three promising future research directions on optimization under uncertainty in the era of big data and deep learning and highlight respective research challenges and potential methodologies.

The rest of this paper is organized as follows. In Section 2, background on mathematical programming techniques for decision making under uncertainty is given. Section 3 presents a comprehensive literature review, where relevant research papers are summarized and classified into four categories. Section 4 discusses promising future research directions to further advance the area of data-driven optimization. Conclusions are provided in Section 5.

2. Background on optimization under uncertainty

In recent years, mathematical programming techniques for decision making under uncertainty have gained tremendous popularity among the PSE community, as witnessed by various successful applications in process synthesis and design (Pistikopoulos, 1995; Rooney and Biegler, 2003), production scheduling and planning (Li and Ierapetritou, 2008; Verderame et al., 2010), and process control (Mesbah, 2016; Krieger and Pistikopoulos, 2014; Chiu and Christofides, 2000). In this section, we present some background knowledge of methodologies for optimization under uncertainty, along with computational algorithms and applications in PSE. Specifically, we briefly review three leading modeling paradigms for optimization under uncertainty, namely stochastic programming, chance-constrained programming, and robust optimization.

2.1. Stochastic programming

Stochastic programming is a powerful modeling paradigm for decision making under uncertainty that aims to optimize the expected objective value across all the uncertainty realizations (Birge and Louveaux, 2011). The key idea of the stochastic programming approach is to model the randomness in uncertain parameters with probability distributions (Birge, 1997; Kall and Wallace, 1994). In general, the stochastic programming approach can effectively accommodate decision making processes with various time stages. In single-stage stochastic programs, there are no recourse variables and all the decisions must be made before knowing uncertainty realizations. By contrast, stochastic programming with recourse can take corrective actions after uncertainty is revealed. Among the stochastic programming approach with recourse, the most widely used one is the two-stage stochastic program, in which decisions are partitioned into "here-and-now" decisions and "wait-and-see" decisions.

The general mathematical formulation of a two-stage stochastic programming problem is given as follows (Birge and Louveaux, 2011).

$$\min_{\mathbf{x} \in X} \mathbf{c}^{\mathsf{T}}\mathbf{x} + \mathbb{E}_{\omega}[Q(\mathbf{x}, \omega)]$$
$$\text{s.t. } \mathbf{A}\mathbf{x} \geq \mathbf{d} \tag{1}$$

The recourse function $Q(\mathbf{x}, \omega)$ is defined by,

$$Q(\mathbf{x}, \omega) = \min_{\mathbf{y}(\omega) \in Y} \mathbf{b}(\omega)^{\mathsf{T}}\mathbf{y}(\omega)$$
$$\text{s.t. } \mathbf{W}(\omega)\mathbf{y}(\omega) \geq \mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x} \tag{2}$$

where $\mathbf{x}$ represents first-stage decisions made "here-and-now" before the uncertainty ω is realized, while the second-stage decisions $\mathbf{y}$ are postponed in a "wait-and-see" manner after observing the uncertainty realization. The objective of the two-stage stochastic programming model includes two parts: the first-stage objective $\mathbf{c}^{\mathsf{T}}\mathbf{x}$ and the expectation of the second-stage objective $\mathbf{b}(\omega)^{\mathsf{T}}\mathbf{y}(\omega)$. The constraints associated with the first-stage decisions are $\mathbf{A}\mathbf{x} \geq \mathbf{d}$, $\mathbf{x} \in X$, and the constraints of the second-stage decisions are $\mathbf{W}(\omega)\mathbf{y}(\omega) \geq \mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x}$ and $\mathbf{y}(\omega) \in Y$. Sets X and Y can include nonnegativity, continuity or integrality restrictions.

The resulting two-stage stochastic programming problem is computationally expensive to solve because of the growth of computational time with the number of scenarios. To this end, decomposition based algorithms have been developed in the existing literature, including Benders decomposition or the L-shaped method (Vanslyke and Wets, 1969; Laporte and Louveaux, 1993), and Lagrangean decomposition (Oliveira et al., 2013). The location of binary decision variables is critical for the design of computational algorithms. For stochastic programs with integer recourse, the expected recourse function is no longer convex, and even discontinuous, thus hindering the employment of conventional L-shaped method. As a result, research efforts have made on computational algorithms for efficient solution of two-stage stochastic mixed-integer programs (Küçükyavuz and Sen, 2017), such as Lagrangian relaxation (Caroe and Schultz, 1999), branch-and-bound scheme (Ahmed et al., 2004), and an improved L-shaped method (Li and Grossmann, 2018; You and Grossmann, 2013).

Stochastic programming has demonstrated various applications in PSE, such as design and operations of batch processes (Ierapetritou and Pistikopoulos, 1995; Bonfill et al., 2004; Bonfill et al., 2005; Chu and You, 2013), process flowsheet optimization (Steimel and Engell, 2015), energy systems (Liu et al., 2010; Peng et al., 2018; Yue and You, 2016; Gao and You, 2017), and supply chain management (Zeballos et al., 2016; Gao and You, 2015; You et al., 2009; Gebreslassie et al., 2012; Tong et al., 2014; You et al., 2011; Gupta and Maranas, 2003). Due to its wide applicability, immense research efforts have been made on the variants of stochastic programming approach. For instance, the two-stage formulation in (1) can be readily extended to a multi-stage stochastic programming setup by utilizing scenario trees. Other extensions include stochastic nonlinear programming (Li et al., 2011), and stochastic programs with endogenous uncertainties (Gupta and Grossmann, 2014; Goel and Grossmann, 2007).

2.2. Chance constrained optimization

As another powerful paradigm for optimization under uncertainty, chance constrained programming aims to optimize an objective while ensuring constraints to be satisfied with a specified probability in uncertain environment (Prékopa, 1995) {Uryasev, 2000}. As in the stochastic programming approach, probability distribution is the key uncertainty model to capture the randomness of uncertain parameters in chance constrained optimization. The chance constrained program was first introduced in the seminal work of (Charnes and Cooper, 1959), and attracted considerable attention ever since. Such chance constraints or probabilistic constraints are flexible enough to quantify the trade-off between objective performance and system reliability (Li et al., 2008).

The generic formulation of a chance constrained optimization problem is presented as follows,

$$\min_{\mathbf{x} \in X} f(\mathbf{x})$$

$$\text{s.t.} \quad \mathbb{P}\{\xi \in \Xi | G(\mathbf{x}, \xi) \leq \mathbf{0}\} \geq 1 - \varepsilon \tag{3}$$

where $\mathbf{x}$ represents the vector of decision variables, X denotes the deterministic feasible region, f is the objective function to be minimized, $\boldsymbol{\xi}$ is a random vector following a known probability distribution $\mathbb{P}$ with the support set Ξ, $G = (g_1, \ldots, g_m)$ represents a constraint mapping, $\mathbf{0}$ is a vector of all zeros, and parameter ε is a pre-specified risk level.

The chance constraint $\mathbb{P}\{\xi \in \Xi | G(\mathbf{x}, \xi) \leq \mathbf{0}\} \geq 1 - \varepsilon$ guarantees that decision $\mathbf{x}$ satisfies constraints with a probability of at least $1 - \varepsilon$. Note that when the number of constraints $m = 1$, the above optimization model is an individual chance constrained program; for $m > 1$, it is called joint chance constrained program (Miller and Wagner, 1965). A salient merit of chance constrained programs is that it allows decision makers choose their own risk levels for the improvement in objectives. To model sequential decision-making processes, two-stage chance constrained optimization with recourse was recently studied and had various applications (Liu et al., 2016; Quddus et al., 2018).

Despite of its promising modeling power, the resulting chance constrained program is generally computationally intractable for the following two main reasons. First, calculating the probability of constraint satisfaction for a given x involves a multivariate integral, which is believed to be computationally prohibitive. Second, the feasible region is not convex even if set X is convex and $G(\mathbf{x}, \boldsymbol{\xi})$ is convex in $\mathbf{x}$ for any realizations of uncertain vector $\boldsymbol{\xi}$ (Prékopa, 1995). In light of these computational challenges, a large body of related literature is devoted into the development of solution algorithms for chance constrained optimization problems, such as sample average approximation (Luedtke and Ahmed, 2008), sequential approximation (Hong et al., 2011), and convex conservative approximation schemes (Nemirovski and Shapiro, 2006). Note that chance constrained programs admit convex reformulation for some very special cases. For example, individual chance constrained programs are endowed with tractable convex reformulations for normal distributions (Birge and Louveaux, 2011). Chance constraints with right-hand-side uncertainty are convex if uncertain parameters are independent and follow log-concave distributions (Prékopa, 1995).

In the PSE community, chance constraints are usually employed for customer demand satisfaction, product quality specification, and service level of process systems {Maranas, 1997; Yue and You, 2013; Gupta et al., 2000; You and Grossmann, 2011; Chu et al., 2015; Zipkin, 2000}. Due to its practical relevance, chance constrained optimization has been applied in numerous applications, including model predictive control (Shen et al., 2018; Cannon et al., 2009), process design and operations (Li et al., 2008; Chu et al., 2015; Gong et al., 2017), refinery blend planning (Yang et al., 2017), biopharmaceutical manufacturing (Liu et al., 2016), and supply chain optimization (Mitra et al., 2008; You and Grossmann, 2011; You and Grossmann, 2008; Ye and You, 2016).

2.3. Robust optimization

As a promising alternative paradigm, robust optimization does not require accurate knowledge on probability distributions of uncertain parameters (Ben-Tal and Nemirovski, 2000; Bertsimas and Sim, 2004; Ben-Tal et al., 2009). Instead, it models uncertain parameters using an uncertainty set, which includes possible uncertainty realizations. It is worth noting that uncertainty set is a paramount ingredient in robust optimization framework (Ben-Tal et al., 2009). Given a specific uncertainty set, the idea of robust optimization is to hedge against the worst case within the uncertainty set. The worst-case uncertainty realization is defined based on different contexts: it could be the realization giving rise to the largest constraint violation, the realization leading to the lowest asset return (Gregory et al., 2011) or the one resulting in the highest regret (Assavapokee et al., 2008).

The conventional box uncertainty set is not a good choice since it includes the unlikely-to-happen scenario where uncertain parameters simultaneously increase to their highest values. The conventional box uncertainty set is defined as follows (Soyster, 1973).

$$U_{\text{box}} = \left\{ \mathbf{u} | \mathbf{u}_i^L \leq \mathbf{u}_i \leq \mathbf{u}_i^U, \forall i \right\} \tag{4}$$

where U_{box} is a box uncertainty set, $\mathbf{u}$ is a vector of uncertain parameters, $\mathbf{u}_i$ is the i-th component of uncertainty vector $\mathbf{u}$. $\mathbf{u}_i^L$ and $\mathbf{u}_i^U$ represent the lower bound and the upper bound of uncertain parameter $\mathbf{u}_i$, respectively. Box uncertainty set simply defines the range of each uncertain parameter in vector $\mathbf{u}$. One cannot easily control the size of this uncertainty set to meet his or her risk-averse attitude. To this end, researchers propose the following budgeted uncertainty set (Bertsimas and Sim, 2004).

$$U_{\text{budget}} = \left\{ \mathbf{u} | \mathbf{u}_i = \bar{\mathbf{u}}_i + \Delta \mathbf{u}_i \cdot z_i, \ -1 \leq z_i \leq 1, \ \sum_i |z_i| \leq \Gamma, \forall i \right\} \tag{5}$$

where U_{budget} denotes a budgeted uncertainty set, $\mathbf{u}$ and $\mathbf{u}_i$ have the same definitions as in (4), $\bar{\mathbf{u}}_i$ is the nominal value of $\mathbf{u}_i$, $\Delta \mathbf{u}_i$ is the largest possible deviation of uncertain parameter $\mathbf{u}_i$, z_i denotes the extent and direction of parameter deviation, and Γ is an uncertainty budget.

Traditional robust optimization approaches, also known as static robust optimization (Bertsimas et al., 2011), make all the decisions at once. This modeling framework cannot well represent sequential decision-making problems (Lorca et al., 2016; Atamtürk and Zhang, 2007; Bertsimas et al., 2013). Adaptive robust optimization (ARO) was proposed to offer a new paradigm for optimization under uncertainty by incorporating recourse decisions (Ben-Tal et al., 2004). Due to the flexibility of adjusting recourse decisions after observing uncertainty realizations, ARO typically generates less conservative solutions than static robust optimization (Ben-Tal et al., 2009). The general form of a two-stage adaptive robust mixed-integer programming model is given as follows:

$$\min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + \max_{\mathbf{u} \in U} \min_{\mathbf{y} \in \Omega(\mathbf{x}, \mathbf{u})} \mathbf{b}^T \mathbf{y}$$

$$\text{s.t.} \quad \mathbf{Ax} \geq \mathbf{d}, \mathbf{x} \in R_+^{n_1} \times Z^{n_2}$$

$$\Omega(\mathbf{x}, \mathbf{u}) = \left\{ \mathbf{y} \in R_+^{n_3} : \mathbf{Wy} \geq \mathbf{h} - \mathbf{Tx} - \mathbf{Mu} \right\} \tag{6}$$

where $\mathbf{x}$ is the first-stage decision made before uncertainty $\mathbf{u}$ is realized, while the second-stage decision $\mathbf{y}$ is postponed in a "wait-and-see" manner. $\mathbf{x}$ includes both continuous and integer variables, while $\mathbf{y}$ only includes continuous variables. $\mathbf{c}$ and $\mathbf{b}$ are the vectors of the cost coefficients. U is an uncertainty set that characterizes the region of uncertainty realizations. ARO approaches could be applied to address uncertainty in a variety of applications, including process design (Gong and You, 2018; Gong and You, 2017; Gong et al., 2016), process scheduling (Shi and You, 2016), supply chain optimization (Gong et al., 2016; Tong et al., 2014), among others. Besides the two-stage ARO framework, the multistage ARO method has attracted immense attention due to its unique feature in reflecting sequential realizations of uncertainties over time (Delage and Iancu, 2015). In multistage ARO, decisions are made sequentially, and uncertainties are revealed gradually over stages. Note that the additional value delivered by ARO over static robust optimization is its adjustability of recourse decisions based on uncertainty realizations (Ben-Tal et al., 2004). Accordingly, the multistage ARO method has demonstrated applications in process

scheduling and planning (Lappas and Gounaris, 2016; Ning and You, 2017b).

Despite popularity of the above three leading paradigms for optimization under uncertainty, these approaches have their own limitations and specific application scopes. To this end, research efforts have been made on "hybrid" methods that leverage the synergy of different optimization approaches to inherit their corresponding strengths and complement respective weaknesses (McLean and Li, 2013; Liu et al., 2016; Baringo and Baringo, 2018; Zhao and Guan, 2013; Keyvanshokooh et al., 2016; Parpas et al., 2009). For instance, stochastic programming was integrated with robust optimization for supply chain design and operation under multi-scale uncertainties (Yue and You, 2016). Robust chance constrained optimization along with global solution algorithms were developed and applied to process design under price and demand uncertainties (Parpas et al., 2009).

3. Existing methods for data-driven optimization under uncertainty

In this section, we review the recent advances in optimization under uncertainty in the era of big data and deep learning. Recent years have witnessed a rapidly growing number of publications on data-driven optimization under uncertainty, an active area integrating machine learning and mathematical programming. These publications cover various topics and can be roughly classified into four categories, namely data-driven stochastic program, data-driven chance constrained program, data-driven robust optimization, and data-driven scenario-based optimization. Unlike the conventional mathematical programming techniques, these data-driven approaches do not presume the uncertainty model is perfectly given *a priori*, rather they all focus on the practical setting where only uncertainty data are available.

3.1. Data-driven stochastic program and distributionally robust optimization

The literature review of data-driven stochastic program, also known as distributionally robust optimization (DRO), is presented in detail in this subsection. The motivation of this emerging paradigm on data-driven optimization under uncertainty is first presented, followed by its model formulation. In this modeling paradigm, the uncertainty is modeled via a family of probability distributions that well capture uncertainty data on hand. This set of probability distributions is referred to as ambiguity set. We then present and analyze various types of ambiguity sets alongside their corresponding strengths and weaknesses. Finally, the extension of DRO to the multistage decision-making setting is also discussed, as well as their recent applications in PSE.

In the stochastic programming approach, it is assumed that the probability distribution of uncertain parameters is perfectly known. However, such precise information of the uncertainty distribution is rarely available in practice. Instead, what the decision maker has is a set of historical and/or real-time uncertainty data and possibly some prior structure knowledge of the probability. Moreover, the assumed probability in conventional stochastic programming might deviate from the true distribution. Therefore, relying on a single probability distribution could result in sub-optimal solutions, or even lead to the deterioration in out-of-sample performance (Smith and Winkler, 2006). Motivated by these weaknesses of stochastic programming, DRO emerges as a new data-driven optimization paradigm which hedges against the worst-case distribution in an ambiguity set. Rather than assuming a single uncertainty distribution, the DRO approach constructs an uncertainty set of probability distributions from uncertainty data through statistical inference and big data analytics. In this way, DRO is capable of

hedging against the distribution errors, and accounts for the input of uncertainty data.

The general model formulation of data-driven stochastic programming is presented as follows (Delage and Ye, 2010).

$$\min_{x \in X} \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} \{ l(x, \xi) \} \tag{7}$$

where **x** is the vector of decision variables, X is the feasible set, l is the objective function, and ξ represents a random vector whose probability distribution $\mathbb{P}$ is only known to reside in an ambiguity set $\mathcal{D}$. The DRO approach aims for optimal decisions under the worst-case distribution, and as a result offers performance guarantee over the family of distributions.

The DRO or data-driven stochastic optimization framework enjoys two salient merits compared with the conventional stochastic programming approach. First, it allows the decision maker to incorporate partial distribution information learned from uncertainty data into the optimization. As a result, the data-driven stochastic programming approach greatly mitigates the issue of optimizer's curse and improves the out-of-sample performance. Second, data-driven stochastic programming inherits the computational tractability from robust optimization and some resulting problems can be solved exactly in polynomial time without resorting to the approximation scheme via sampling or discretization. For example, optimization problem (7) for a convex program with continuous variables and a moment-based ambiguity set is proved to be solvable in polynomial time (Delage and Ye, 2010).

The choice of ambiguity sets plays a critical role in the performance of DRO. When choosing ambiguity set, the decision maker needs to consider the following three factors, namely tractability, statistical meaning, and performance (Hanasusanto et al., 2015). First, the data-driven stochastic programming problem with the ambiguity set should be computationally tractable, meaning the resulting optimization could be formulated as linear, conic quadratic or semidefinite programs. Second, the derived ambiguity set should have clear statistical meaning. Therefore, various ways of constructing ambiguity sets based on uncertainty data were extensively studied (Delage and Ye, 2010; Esfahani and Kuhn, 2018; Shang and You, 2018). Third, the devised ambiguity set should be tight to increase the performance of resulting decisions.

One commonly used approach to constructing ambiguity set is moment-based approaches, in which first and second order information is extracted from uncertainty data using statistical inference (Calafiore and El Ghaoui, 2006). The ambiguity set that specifies the support, first and second moment information is shown as follows,

$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{M}_+ \left| \begin{array}{l} \mathbb{P}[\xi \in \Xi] = 1 \\ \mathbb{E}_{\mathbb{P}}[\xi] = \mu \\ \mathbb{E}_{\mathbb{P}}[(\xi - \mu)(\xi - \mu)^T] = \sum \end{array} \right. \right\} \tag{8}$$

where ξ represents the uncertainty vector, Ξ is the support, $\mathbb{P}$ represents the probability distribution of ξ, $\mathcal{M}_+$ denotes the set of all probability measures, $\mathbb{E}_{\mathbb{P}}$ denotes the expectation with respect to distribution $\mathbb{P}$. Parameters μ and Σ represent the mean vector and covariance matrix estimated from uncertainty data, respectively.

The ambiguity set in (8) fails to account for the fact that the mean and covariance matrix are also subject to uncertainty. To this end, an ambiguity set was proposed based on the distribution's support information as well as the confidence regions for the mean and second-moment matrix in the work of (Delage and Ye, 2010). The resulting DRO problem could be solved efficiently in polynomial time.

$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{M}_+ \left| \begin{array}{l} \mathbb{P}[\xi \in \Xi] = 1 \\ (\mathbb{E}_{\mathbb{P}}[\xi] - \mu)^T \sum^{-1} (\mathbb{E}_{\mathbb{P}}[\xi] - \mu) \le \psi_1 \\ \mathbb{E}_{\mathbb{P}}[(\xi - \mu)(\xi - \mu)^T] \le \psi_2 \sum \end{array} \right. \right\} \tag{9}$$

where ξ represents the uncertainty vector, Ξ is the support, $\mathbb{P}$ represents the probability distribution of ξ. The equality constraint $\mathbb{P}[\xi \in \Xi] = 1$ enforces that all uncertainty realizations reside in the support set Ξ. Parameters ψ_1 and ψ_2 are used to define the sizes of confidence regions for the first and second moment information, respectively.

The moment-based ambiguity sets typically enjoy the advantage of computational tractability. For example, DRO with the ambiguity set based on principal component analysis and first-order deviation functions was developed (Shang and You, 2018). Additionally, the computational effectiveness of this data-driven DRO method was demonstrated via process network planning and batch production scheduling. Recently, a data-driven DRO model was developed for the optimal design and operations of shale gas supply chains to hedge against uncertainties associated with shale well estimated ultimate recovery and product demand (Gao et al., 2019). However, the moment-based ambiguity set is not guaranteed to converge to the true probability distribution as the number of uncertainty data goes to infinity. Consequently, this type of ambiguity set suffers from the conservatism with moderate uncertainty data. To address the above issue with moment-based methods, ambiguity sets based on statistical distance between probability distributions were developed, as shown below,

$$\mathcal{D} = \{\mathbb{P} \in \mathcal{M}_+ | d(\mathbb{P}, \mathbb{P}_0) \leq \theta \} \qquad (10)$$

where $\mathbb{P}$ is the probability distribution of uncertain parameters, $\mathbb{P}_0$ represents the reference distribution such as the empirical distribution, d denotes some statistical distance between two distributions, and θ stands for the confidence level.

Ambiguity set in (10) can be further classified based on the adopted distance metric, such as Kullback–Leibler divergence (Hu and Hong, 2013) and Wasserstein distance (Esfahani and Kuhn, 2018). For example, a DRO model was proposed for lot-sizing problem, in which the chi-square goodness-of-fit test and robust optimization were combined. The ambiguity set of demand was constructed from uncertainty data by using a hypothesis test in statistics, called the chi-square goodness-of-fit test (Klabjan et al., 2013). This set is well defined by linear constraints and second order cone constraints. It is worth noting that the input of their model is histograms, which make it possible to use a finite dimensional probability vector to characterize the distribution. The adopted statistic belonged to the phi-divergences, which motivated researchers to construct distribution uncertainty set by using the phi-divergences (Bayraksan and Love, 2015).

To account for the sequential decision-making process, researchers recently developed the adaptive DRO method by incorporating recourse decision variables (Hanasusanto and Kuhn, 2018; Bertsimas et al., 2019). A general two-stage data-driven stochastic programming model is presented in the following form:

$$\min_{x \in X} c^{\mathrm{T}}x + \max_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} \{ Q(x, \xi) \}$$
$$\text{s.t. } Ax \leq d$$
$$Q(x, \xi) = \begin{cases} \min_{y \in Y} \mathbf{b}(\xi)^{\mathrm{T}} y \\ \text{s.t. } T(\xi)x + W(\xi)y \leq h(\xi) \end{cases} \qquad (11)$$

where $\mathbf{x}$ presents the vector of first-stage decision variables that need to be determined before observing uncertainty realizations, $\mathbf{y}$ denotes the vector of second-stage decision variables that can be adjustable based on the realized uncertain parameters ξ, sets X and Y can include nonnegativity, continuity or integrality restrictions, and Q represents the recourse function. The objective of the above data-driven stochastic program is to minimize the worst-case expected cost with respect to all possible uncertainty distributions $\mathbb{P}$ within the ambiguity set $\mathcal{D}$. Based on the literature, multistage data-driven DRO is becoming a rapidly evolving research direction.

Data-driven stochastic programming has several salient merits over the conventional stochastic programming approach. However, there are few papers on its PSE applications in the existing literature (Shang and You, 2018; Gao et al., 2019). As the trend of big data has fueled the increasing popularity of data-driven stochastic programming in many areas, DRO emerges as a new data-driven optimization paradigm which hedges against the worst-case distribution in an ambiguity set, and has various applications in power systems, such as unit commitment problems (Xiong et al., 2017; Chen et al., 2018; Duan et al., 2018; Zhao and Guan, 2016), and optimal power flow (Wang et al., 2018; Guo et al., 2018).

3.2. Data-driven chance constrained program

In contrast to the data-driven stochastic programming approach reviewed in Section 3.1., data-driven chance constrained programming is another paradigm focusing on chance constraint satisfaction under the worst-case probability instead of optimizing the worst-case expected objective. Although both data-driven chance constrained program and DRO adopt ambiguity sets in the uncertainty models, they have distinct model structures. Specifically, data-driven chance constrained program features constraints subject to uncertainty in probability distributions, while DRO typically only involves the worst-case expectation of an objective function with respect to a family of probability distributions. As introduced in Section 2, the chance constrained programming approach assumes the complete distribution information is perfectly known. However, the decision maker only has access to a finite number of uncertainty realizations or uncertainty data. On the one hand, such complete knowledge of distribution is usually estimated from limited number of uncertainty data or obtained from expert knowledge. On the other hand, even if the probability distribution is available, the chance constrained program is computationally cumbersome. In practice, one can only have partial information on the probability distribution of uncertainty. Therefore, data-driven chance constrained optimization emerges as another paradigm for hedging against uncertainty in the era of big data.

The general form of data-driven chance constrained program is given by,

$$\min_{\mathbf{x} \in X} f(\mathbf{x})$$
$$\text{s.t. } \min_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in \Xi | G(\mathbf{x}, \xi) \leq \mathbf{0}\} \geq 1 - \varepsilon \qquad (12)$$

where $\mathbf{x}$ represents the vector of decision variables, X denotes the deterministic feasible region, f is the objective function, ξ is a random vector following a probability distribution $\mathbb{P}$ that belongs to an ambiguity set $\mathcal{D}$. $G = (g_1, \ldots, g_m)$ represents a constraint mapping, $\mathbf{0}$ is a vector of all zeros, and parameter ε is a pre-specified risk level. The data-driven chance constraints enforce classical chance constraints to be satisfied for every probability distribution within the ambiguity set.

The computational tractability of the resulting data-driven chance constrained program can vary depending on both the ambiguity sets and the structure of the optimization problem. In the following, we summarize the relevant papers according to the adopted uncertainty set of distributions and optimization structures.

Distributionally robust individual linear chance constraints under the ambiguity set comprised of all distributions sharing the same known mean and covariance were reformulated as convex second-order cone constraints (Calafiore and El Ghaoui, 2006). The deterministic convex conditions to enforce distributionally robust chance constraints were provided under distribution families of (a) independent random variables with box-type support and (b) radially symmetric non-increasing distributions over the orthotope support. The worst-case conditional value-at-risk (CVaR)

approximation for distributionally robust joint chance constraints was studied assuming first and second moment (Zymler et al., 2013), and the resulting conservative approximation can be cast as semidefinite program. In addition to moment information, a specific structural information of distributions called unimodality was incorporated into the ambiguity set, and the corresponding ambiguous risk constraints were reformulated as a set of second second-order cone constraints (Li et al., 2017). Instead of assuming unimodality of distributions, data-driven robust individual chance constrained programs along with convex approximations were recently developed using a mixture distribution-based ambiguity set with fixed component distribution and uncertain mixture weights (Chen et al., 2018).

In real world applications, exact moment information can be challenging to obtain, and can only be estimated through confidence intervals from uncertainty realizations (Delage and Ye, 2010). To accommodate this moment uncertainty, attempts were made in the context of distributionally robust chance constraints, including constructing convex moment ambiguity set (El Ghaoui et al., 2003), employing Chebyshev ambiguity set with bounds on second-order moment (Cheng et al., 2014), characterizing a family of distributions with upper bounds on both mean and covariance (Zhang et al., 2018). Ambiguous joint chance constraints were studied where the ambiguity set was characterized by the mean, convex support, and an upper bound on the dispersion (Hanasusanto et al., 2017), and the resulting constraints were conic representable for right-hand-side uncertainty. In addition to generalized moment bounds (Wiesemann et al., 2014), structural properties of distributions, such as symmetry, unimodality, multimodality and independence, were further integrated into distributionally robust chance constrained programs leveraging a Choquet representation (Hanasusanto et al., 2015). Nonlinear extensions of distributionally robust chance constraints were made under the ambiguity sets defined by mean and variance (Yang and Xu, 2016), convex moment constraints (Xie and Ahmed, 2018), mean absolute deviation (Postek et al., 2018), and a mixture of distributions (Lasserre and Weisser, 2018).

Although moment-based ambiguity sets achieve certain success, they do not converge to the true probability distribution as the number of available uncertainty data increases. Consequently, the resulting data-driven chance-constrained programs tend to generate conservative solutions. To this end, data-driven chance-constrained programs with distance-based ambiguity set were proposed to alleviate the undesirable consequence of moment-based data-driven chance-constrained programs. The ambiguity set defined by the Prohorov metric was introduced into the distributionally robust chance constraints, and the resulting optimization problem was approximated by using robust sampled problem (Erdogan and Iyengar, 2006). Distributionally robust chance constraints with the ambiguity set containing all distributions close to a reference distribution in terms of Kullback–Leibler divergence were cast as classical chance constraints with an adjusted risk level (Hu and Hong, 2013). Data-driven chance constrained programs with ϕ-divergence based ambiguity set were proposed (Jiang and Guan, 2016), and further extensions were made using the kernel smoothing method (Calfa et al., 2015). Recently, data-driven chance constraints over Wasserstein balls were exactly reformulated as mixed-integer conic constraints (Chen et al., 2018; Ji and Lejeune, 2018). Leveraging the strong duality result (Gao and Kleywegt, 2016), distributionally robust chance constrained programs with Wasserstein ambiguity set were studied for linear constraints with both right and left hand uncertainty (Xie, 2018), as well as for general nonlinear constraints (Hota et al., 2018).

Data-driven chance constrained programs have successful applications in a number of areas, such as power system (Xie and Ahmed, 2018), stochastic control (Van Parys et al., 2016), and vehicle routing problem (Ghosal and Wiesemann, 2018).

3.3. Data-driven robust optimization

As a paramount ingredient in robust optimization, uncertainty sets endogenously determine robust optimal solutions and therefore should be devised with special care. However, uncertainty sets in the conventional robust optimization methodology are typically set *a priori* using a fixed shape and/or model without providing sufficient flexibility to capture the structure and complexity of uncertainty data. For example, the geometric shapes of uncertainty sets in (4) and (5) do not change with the intrinsic structure and complexity of uncertainty data. Furthermore, these uncertainty sets are specified by finite number of parameters, thereby having limited modeling flexibility. Motivated by this knowledge gap, data-driven robust optimization emerges as a powerful paradigm for addressing uncertainty in decision making.

A data-driven ARO framework that leverages the power of Dirichlet process mixture model was proposed (Ning and You, 2017a). The data-driven approach for defining uncertainty set was developed based on Bayesian machine learning. This machine learning model was then integrated with the ARO method through a four-level optimization framework. This developed framework effectively accounted for the correlation, asymmetry and multimode of uncertainty data, so it generated less conservative solutions. Its salient feature is that multiple basic uncertainty sets are used to provide a high-fidelity description of uncertainties. Although the data-driven ARO has a number of attractive features, it does not account for an important evaluation metric, known as regret, in decision-making (Bell, 1982). Motivated by the knowledge gap, a data-driven bi-criterion ARO framework was developed that effectively accounted for the conventional robustness as well as minimax regret (Ning and You, 2018a).

In some applications, uncertainty data in large datasets are usually collected under multiple conditions. A data-driven stochastic robust optimization framework was proposed for optimization under uncertainty leveraging labeled multi-class uncertainty data (Ning and You, 2018b). Machine learning methods including Dirichlet process mixture model and maximum likelihood estimation
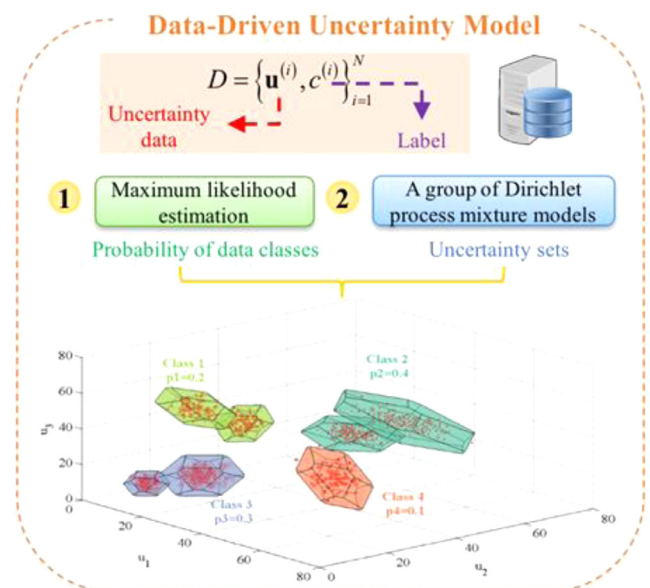

Fig. 1. The data-driven uncertainty model based on the Dirichlet process mixture model (Ning and You, 2018b).

were employed for uncertainty modeling, which is illustrated in Fig. 1. This framework was further proposed based on the data-driven uncertainty model through a bi-level optimization structure. The outer optimization problem followed the two-stage stochastic programming approach, while ARO was nested as the inner problem for maintaining computational tractability.

To mitigate computational burden, research effort has been made on convex polyhedral data-driven uncertainty set based on machine learning techniques, such as principal component analysis and support vector clustering. A data-driven robust optimization framework that leveraged the power of principal component analysis and kernel smoothing for decision-making under uncertainty was studied (Ning and You, 2018c). In this approach, correlations between uncertain parameters were effectively captured, and latent uncertainty sources were identified by principal component analysis. To account for asymmetric distributions, forward and backward deviation vectors were utilized in the uncertainty set, which was further integrated with robust optimization models. A data-driven static robust optimization framework based on support vector clustering that aims to find the hypersphere with minimal volume to enclose uncertainty data was proposed (Shang et al., 2017). The adopted piecewise linear kernel incorporates the covariance information, thus effectively capturing the correlation among uncertainties. These two data-driven robust optimization approaches utilized polyhedral uncertainty learned from data, and thus enjoying computational efficiency. Various types of data-driven uncertainty sets were developed for static robust optimization based on statistical hypothesis tests (Bertsimas et al., 2018), copula (Zhang et al., 2018), and probability density contours (Zhang et al., 2018).

To address multistage decision making under uncertainty, a data-driven approach for optimization under uncertainty based on multistage ARO and nonparametric kernel density M-estimation was developed (Ning and You, 2017b). The salient feature of the framework was its incorporation of distributional information to address the issue of over-conservatism. Robust kernel density estimation was employed to extract probability distributions from data. This data-driven multistage ARO framework exploited robust statistics to be immunized to data outliers. An exact robust counterpart was developed for solving the resulting data-driven ARO problem.

In recent years, data-driven robust optimization has been applied to a variety of areas, such as power systems (Ning and You, 2019), energy systems (Zhao et al., 2019), planning and scheduling (Ning and You, 2017b; Zhang et al., 2018), process control (Ning and You, 2018c; Shang and You, 2019), and transportation systems (Miao et al., 2019; Zhao and You, 2019).

3.4. Scenario optimization approach for chance constrained programs

A salient feature of scenario-based optimization is that it does not require the explicit knowledge of probability distribution as in the stochastic programming approach. Additionally, scenario-based optimization uses uncertainty scenarios to seek an optimal solution having a high probabilistic guarantee of constraint satisfaction instead of utilizing scenarios or samples to approximate the expectation term as in stochastic programming. Although the scenario-based optimization can be regarded as a special type of robust optimization that has a discrete uncertainty set consisting of uncertainty data, it can provide probabilistic guarantee for those unobserved uncertainty data in the testing data set. Note that the scenario-based optimization approach provides a viable and data-driven route to achieving approximate solutions of chance-constrained programs. The scenario-based optimization approach is a general data-driven optimization under uncertainty framework in which uncertainty data or random samples are utilized in a

more direct manner compared with other data-driven optimization methods. This data-driven optimization framework was first introduced in (Calafiore and Campi, 2005), and has gained great popularity within the systems and control community (Campi et al., 2009). As in data-driven chance constrained programs, the knowledge of true underlying uncertainty distribution is not required in scenario optimization but a finite number of uncertainty realizations. Specifically, the scenario approach enforces the constraint satisfaction with N independent identically distributed uncertainty data $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$. The resulting scenario optimization problem is given by,

$$\min_{\mathbf{x} \in X} \mathbf{c}^T \mathbf{x}$$
$$\text{s.t. } f\left(\mathbf{x}, \mathbf{u}^{(i)}\right) \leq 0, i = 1, \ldots, N \tag{13}$$

where $\mathbf{x}$ is the vector of decision variables, X represents a deterministic convex and closed set unaffected by uncertainty, $\mathbf{c}$ is the vector of cost coefficients, and f denotes the constraint function affected by uncertainty $\mathbf{u}$. Note that function f is typically assumed to be convex in $\mathbf{x}$, and can have arbitrarily nonlinear dependence on $\mathbf{u}$, as opposed to data-driven nonlinear chance constrained program assuming the constraint function must be quasi-convex in $\mathbf{u}$ (Yang and Xu, 2016). Additionally, scenario-based optimization can be considered as a special case of data-driven robust optimization when the uncertainty set is constructed as a union of $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$.

In the scenario optimization literature, $\omega \doteq \{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}\}$ is referred to as the multi-sample or scenario that is drawn from the product probability space. Due to the random nature of the multi-sample, the optimal solution of the scenario optimization problem (13), denoted as $\mathbf{x}^*(\omega)$, is also random. One key merit of the scenario approach is that the scenario optimization problem admits the same problem type as its deterministic counterpart, so that it can be solved efficiently by convex optimization algorithms when $f(\mathbf{x}, \mathbf{u})$ is convex in $\mathbf{x}$ (Boyd and Vandenberghe, 2004). Moreover, the optimal solution $\mathbf{x}^*(\omega)$ is guaranteed to satisfy the constraints with other unseen uncertainty realizations with a high probability (Campi and Garatti, 2008).

For the sake of clarity, we revisit the following definition and theorem (Campi and Garatti, 2008).

Definition (Violation probability) The violation probability of a given decision $\mathbf{x}$ is defined as follows:

$$V(\mathbf{x}) \doteq \mathbb{P}\{\mathbf{u} \in \Xi | f(\mathbf{x}, \mathbf{u}) > 0\} \tag{14}$$

where $V(\mathbf{x})$ denotes the probability of violation for a given $\mathbf{x}$, and Ξ represents the support of uncertainty $\mathbf{u}$. We say a decision $\mathbf{x}$ is ε-feasible if $V(\mathbf{x}) \leq \varepsilon$.

Theorem. Assuming $\mathbf{x}^*(\omega)$ is the unique optimal solution of the scenario optimization problem. It holds that

$$\mathbb{P}^N\{\omega | V(\mathbf{x}^*(\omega)) \leq \varepsilon\} \geq 1 - \sum_{i=0}^{n-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \tag{15}$$

where n is the number of decision variables, N denotes the number of uncertainty data, and $\mathbb{P}^N$ is a product probability governing the sample generation.

The above theorem implies that the optimal solution $\mathbf{x}^*(\omega)$ satisfies the corresponding chance constraint with a certain confidence level. The proof of this theorem depends on the fundamental fact that the number of support constraints, the removal of which changes the optimal solution, is upper bounded by the number of decision variables (Calafiore and Campi, 2005). Note that (15) holds with equality for the fully-supported convex optimization problem (Campi and Garatti, 2008), meaning that the probability bound is tight. Additionally, the result holds true irrespective of probability distribution information or even its support set.

By exploiting the structured dependence on uncertainty, the sample size required by the scenario optimization problem

was reduced through a tighter bound on Helly's dimension (Zhang et al., 2015). Rather than focusing on the constraint violation probability, considerable research efforts have been made on the degree of violation (Kanamori and Takeda, 2012), expected probability of constraint violation (Calafiore, 2009), and the performance bounds for objective values (Esfahani et al., 2015). To make a trade-off between feasibility and performance, the case was studied where some of the sampled constraints were allowed to be violated for improving the performance of the objective (Calafiore, 2010). Subsequent work along this direction includes a sampling-and-discarding method (Campi and Garatti, 2011). A wait-and-judge scenario optimization framework was proposed in which the level of robustness was assessed *a posteriori* after the optimal solution was obtained (Campi and Garatti, 2018). Recently, the extension of scenario-based optimization to the multistage decision making setting was made (Kariotoglou et al., 2016; Vayanos et al., 2012).

While the scenario optimization problems with continuous decision variables are extensively studied (Campi et al., 2009), the mixed-integer scenario optimization was less developed. An attempt to extend the scenario theory to random convex programs with mixed-integer decision variables was made (Calafiore et al., 2012), and the Helly dimension in the mixed-integer scenario program was proved to depend geometrically on the number of integer variables. This result suggests that the required sample size can be prohibitively large for scenario programs with many discrete variables. Along this research direction, two sampling algorithms within the framework of *S*-optimization were recently developed for solving mixed-integer convex scenario programs (De Loera et al., 2018).

In some real-world applications, the required sample size can be very large, resulting in great computational burden for scenario optimization problems with huge number of sampling constraints. One way to circumvent this difficulty is to devise sequential solution algorithms. Along this direction, sequential randomized algorithms were developed for convex scenario optimization problems (Chamanbaz et al., 2016), and fell into the framework of Sequential Probabilistic Validation (SPV) (Alamo et al., 2015). The motivation behind these sequential algorithms is that validating a given solution with a large number of samples is less computational expensive than solving the corresponding scenario optimization problem. Recently, a repetitive scenario design approach was proposed by iterating between reduced-size scenario optimization problems and the probabilistic feasibility check (Calafiore, 2017). The trade-off between the sample size and the expected number of repetitions was also revealed in the repetitive scenario design (Calafiore, 2017). Note that the classical scenario-based approach is an extreme situation in the trade-off curve, where one seeks to find the solution at one step. Another effective way to reduce the computation cost of large-scale scenario optimization is to employ distributed algorithms (Margellos et al., 2018; Carlone et al., 2014). Particularly, the sampled constraints were distributed among multiple processors of a network, and the large-scale scenario optimization problems can be efficiently solved via constraint consensus schemes (Carlone et al., 2014). Along this direction, a distributed computing framework was developed for the scenario convex program with multiple processors connected by a graph (You et al., 2018). The major advantage of this approach is that the computational cost for each processor becomes lower and the original scenario optimization problem can be solved collaboratively. Other contribution to reduce computational cost is made based on a non-iterative two-step procedure, i.e. the optimization step and detuning step (Care et al., 2014). As a consequence, the total sample complexity was greatly decreased.

Traditionally, the field of scenario optimization has focused on convex optimization problems, in which the number of support constraints is upper bounded by the number of decision variables. However, such upper bounds are no longer available in nonconvex scenario optimization problems, giving rise to research challenges of extending the scenario theory to the nonconvex setting. To date, few works have considered nonconvex uncertain program using the scenario approach. One contribution is that of (Campi et al., 2018), in which assessing the generalization of the optimal solution in a wait-and-judge manner through the concept of support subsample was proposed. The proposed approach can be employed to general nonconvex setups, including mixed-integer scenario optimization problems. Another attempt to address nonconvex scenario optimization made use of the statistical learning theory for bounding the violation probability, and devised a randomized solution algorithm (Alamo et al., 2009). The statistical learning theory-based method provided the probabilistic guarantee for all feasible solutions, as opposed to the convex scenario approach where such guarantee is valid only for the optimal solution. This unique feature regarding probabilistic guarantees for all feasible solutions granted by the statistical learning based method is of practical relevance (Calafiore et al., 2011), since it is computationally challenging to solve nonconvex optimization problems to global optimality. A class of non-convex scenario optimization problem, which has non-convex objective functions and convex constraints, was recently studied (Grammatico et al., 2016). Since the Helly's dimension for the optimal solution of such non-convex scenario program can be unbounded, the direct application of scenario approaches based on Helly's theorem is impossible. To overcome the research challenge, the feasible region was restricted to the convex hull of few optimizers, thus enabling the application of sample complexity results (Campi and Garatti, 2008).

4. Future research directions and opportunities

Several promising research directions in data-driven optimization under uncertainty are highlighted in this section. We specifically focus on some ideas on closed-loop data-driven optimization, integration of deep learning and scenario-based optimization, and learning-while-optimizing frameworks.

4.1. A "closed-loop" data-driven optimization framework with feedback from mathematical programming to machine learning

The framework of data-driven optimization under uncertainty could be considered as a "hybrid" system that integrates the data-driven system based on machine learning to extract useful and relevant information from data, and the model-based system based on the mathematical programming to derive the optimal decisions from the information. Existing data-driven optimization approaches adopt a sequential and open-loop scheme, which could be further improved by introducing feedback steps from the model-based system to data-driven system. A "closed-loop" data-driven optimization paradigm that explores the information feedback to fully couple upper-stream machine learning and downstream mathematical programming could be a more effective and rigorous approach.

4.1.1. The issues of conventional "open-loop" data-driven optimization methods

It is widely recognized that data-driven optimization is a promising way to hedging against uncertainty in the era of big data and deep learning. Such promise hinges heavily on the organic integration and effective interaction between machine learning and mathematical programming. In existing data-driven optimization frameworks, the tasks performed by the data-driven

system and the model-based system are treated separately in a sequential fashion. More specifically, data serve as input to a data-driven system. After that, useful, accurate and relevant uncertainty information is extracted through the data-driven system and further passed along to the model-based system based on mathematical programming for rigorous and systematic optimization under uncertainty, using paradigms such as robust optimization and stochastic programming. However, due to the sequential connection between these two systems, the machine learning model is trained without interacting with the "downstream" mathematical programming. Accordingly, from a control theoretical perspective, such "hybrid" systems in the existing data-driven optimization literature are essentially open loop. In contrast to open-loop systems, closed-loop systems using the feedback control strategy deliver amazingly superior system performance (e.g. stability, robustness to disturbances, and safety) in virtually every area of science and engineering, such as biological systems, social networks, and mechanical systems (Aström and Kumar, 2014). Therefore, there should be a "feedback" channel for information flow returning from the model-based system to the data-driven system, in addition to the information flow that is fed into the mathematical programming problem from the machine learning results. The design of such feedback loops from mathematical programming to machine learning deserves further attention in future research. Although there are closed-loop machine learning methods in the case of reinforcement learning (Shin and Lee, 2019), to the best of our knowledge, there are few works on developing a closed-loop strategy for data-driven mathematical programming under uncertainty. Different from mathematical programming, reinforcement learning is a kind of machine learning that aims to find an action policy to increase an agent's performance in terms of reward by interacting with an environment.

4.1.2. A "closed-loop" data-driven optimization framework

Due to its critical role in the training of machine learning models, loss functions could provide a foundation for considering feedback steps in future research efforts. Instead of using a mathematical-programming-agnostic loss function (e.g. logistic or squared-error loss), a loss function that incorporates the objective function of mathematical programming could be used to train the machine learning model. Specifically, a weighted sum of the conventional loss function and the objective function in the mathematical programming problem should be useful in handling issues experienced with current "open-loop" data-driven frameworks. An iterative scheme between machine learning and mathematical programming offers an alternative promising path to close the loop of the data-driven system and the model-based system. Fig. 2 presents the potential schematic of the closed-loop data-driven mathematical programming system. From the figure, we can see that the feedback from the model-based system serves as input to

the data-driven system. In this way, the "hybrid" system becomes a closed-loop one in which information can be transmitted in both directions. Such feedback strategy should be beneficial to the "hybrid" system and could provide an effective way to organically integrate machine learning and mathematical programming.

Research challenges emerge from the feedback step in the "hybrid" system. In typical PSE applications, the problem size of mathematical programs tends to be large. Such large-scale mathematical programming problems in conjunction with big data could pose a computational challenge for the training of machine learning. Additionally, how to design an effective feedback strategy to "close the loop" poses another key challenge to be addressed.

4.1.3. Incorporating "prior" knowledge in the data-driven optimization framework

In addition to uncertainty data, some available domain-specific knowledge or "prior" knowledge could serve as another informative input to the data-driven system. Relying solely on the data to develop the uncertainty model could unfavorably influence the downstream mathematical programming. The prior knowledge depicts what the decision maker knows about the uncertainty, and it can come in different forms. For example, the prior knowledge could be the structural information of probability distributions, upper and lower bounds of uncertain parameters or certain correlation relationship among uncertainties. Incorporating such "prior" knowledge in the data-driven optimization framework could be substantially useful and provides more reliable results in the face of messy data.

4.2. Leveraging deep learning techniques for hedging against uncertainty in data-driven optimization

Recently, deep learning has shown great promise due to its amazing power in hierarchical representation of data (Goodfellow et al., 2016). The deep learning techniques are now shaping and revolutionizing many areas of science and engineering (LeCun et al., 2015). In recent years, deep learning has a wide array of applications in the PSE domain, such as process monitoring (Zhu et al., 2019; Zhang and Zhao, 2017), refinery scheduling (Gao et al., 2014), and soft sensor (Shang et al., 2014). For extensive surveys on deep learning in the PSE area, we refer the reader to the review papers on this subject (Venkatasubramanian, 2019; Lee et al., 2018). In real applications, uncertainty data exhibit very complex and highly nonlinear characteristic. Therefore, it should be promising to explore the potential opportunities of leveraging deep neural networks with various architectures to uncover useful patterns of uncertainty data for mathematical programming.

In this section, a variety of deep learning techniques are first summarized along with their unique features from a practical
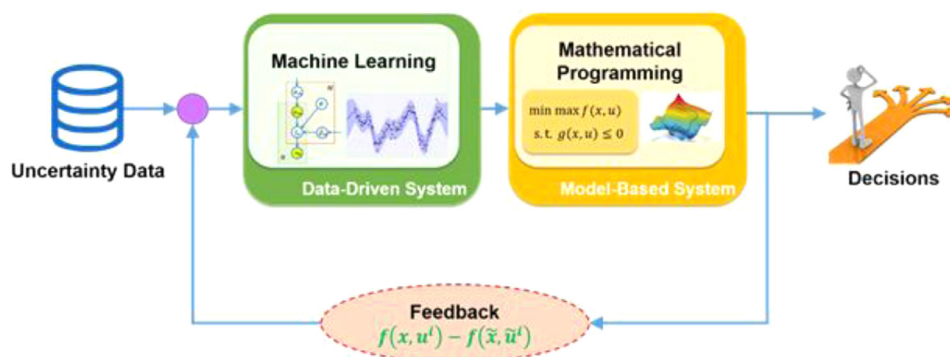

Fig. 2. The schematic of "closed-loop" data-driven mathematical programming framework.

point of view, and future research directions on how to leverage the power of deep learning in optimization under uncertainty are further suggested. Research opportunities of integrating data-driven scenario-based optimization with deep generative models are then presented.

4.2.1. Various types of deep learning techniques and their potentials

In this subsection, we present three types of deep learning techniques, including deep belief networks, convolutional neural networks, and recurrent neural networks, and explore their potential applications in data-driven optimization under uncertainty.

- Deep belief networks

Among deep learning techniques, deep belief networks (DBNs) are becoming increasingly popular primarily because its unique feature in capturing a hierarchy of latent features (Mohamed et al., 2012). DBNs essentially belong to probabilistic graphical models and are structured by stacking a series of restricted Boltzmann machines (RBMs). This specific network structure is designed based on the fact that a single RBM with only one hidden layer fall shorts of capturing the intrinsic complexities in high-dimensional data. As the building blocks for DBNs, RBMs are characterized as two layers of neurons, namely hidden layer and visible layer. Note that the hidden layer can be regarded as the abstract representation of the visible layer. There are undirected connections between these two layers, while there exist no intra-connections within each layer. The training process of DBNs typically involves the pre-training and fine-tuning procedures in a layer-wise scheme. Armed with multiple layers of hidden variables, DBNs enjoy unique power in extracting a hierarchy of latent features automatically, which is desirable in many practical applications. As a result, DBNs have been applied in a wide spectrum of areas, including fault diagnosis (Zhang and Zhao, 2017), soft sensor (Shang et al., 2014), and drug discovery (Gawehn et al., 2016). DBNs can decipher complicated nonlinear correlation among uncertain parameters. Recently, deep Gaussian process model was proposed as a special type of DBN based Gaussian process mappings. Due to its unique advantage in nonlinear regression, deep Gaussian process model should be used to characterize the relationship between uncertain parameters, such as product price and demand.

- Convolutional neural networks

Convolutional neural networks (CNNs) are one specialized version of deep neural networks (Krizhevsky et al., 2017), and they have become increasingly popular in areas such as image classification, speech recognition, and robotics. Inspired by the visual neuroscience, CNNs are designed to fully exploit the three main ideas, namely sparse connectivity, weight sharing, and equivariant representations (Goodfellow et al., 2016). This kind of neural network is suited for processing data in the form of multiple arrays, particularly two-dimensional image data. The architecture of a CNN typically consists of convolution layers, nonlinear layers, and pooling layers. In convolution layers, feature maps are extracted by performing convolutions between local patch of data and filters. The filters share the same weights when moving across the dataset, leading to reduced number of parameters in networks. The obtained results are further passed through a nonlinear activation function, such as rectified linear unit (ReLU). After that, pooling layers, such as max pooling and average pooling, are applied to aggregate semantically similar features. Such different types of layers are alternatively connected to extract hierarchical features with various abstractions. For the purpose of classification, a fully connected layer is stacked after extracting the high-level features. Although CNNs are mainly used for image classification, they have been used to learn spatial features of traffic flow data at nearby locations which exhibit strong spatial correlations (Wu et al., 2018).

Given its unique power in spatial data modeling, CNNs hold the potential to model uncertainty data with large spatial correlations, such as demand data in different adjacent market locations. In addition, the CNNs can be trained for the labeled multi-class uncertainty data to perform the task of classification. Therefore, the output of the CNN potentially acts as the probability weights used in the data-driven stochastic robust optimization framework.

- Recurrent neural networks

Besides the aforementioned models for spatial data, recurrent neural networks (RNNs) are widely recognized as the state-of-the-art deep learning technique for processing time series data, especially those from language and speech (Graves et al., 2013). RNNs can be considered as feedforward neural networks if they are unfolded in time scale. The architecture of neural networks in a RNN possesses a unique structure of directed cycles among hidden units. In addition, the inputs of the hidden unit come from both the hidden unit of previous time and the input unit at current time. Accordingly, these hidden units in the architecture of RNNs constitute the state vectors and store the historical information of past input data. With this special architecture, RNNs are well-suited for feature learning for sequential data and demonstrate successful applications in various areas, including natural speech recognition (Graves et al., 2013), and load forecasting (Vermaak and Botha, 1998). However, one drawback of RNNs is its weakness in storing long-term memory due to gradient vanishing and exploding problems. To address this issue, research efforts have been made on variants of RNNs, such as long short-term memory (LSTM) and gated recurrent unit (GRU) (Hochreiter and Schmidhuber, 1997). By explicitly incorporating input, output and forget gates, LSTM enhances the capability of memorizing the long-term dependency among sequential data. In sequential mathematical programming under uncertainty, massive time series of uncertain parameters are collected. Uncertainty data realized at different time stages often exhibit temporal dynamics. To this end, deep learning techniques, such as deep RNNs and LSTM, could be leveraged to decipher the temporal dynamics and trajectories of uncertainty over time stages. Therefore, exploring the integration between deep learning and multistage optimization under uncertainty is another promising research direction.

4.2.2. Deep generative models for scenario-based optimization

Despite the various successful applications of scenario-based optimization, this type of data-driven optimization framework has its own limitations. In general, scenario-based optimization enjoys computational efficiency by constraint sampling and provides the feasibility guarantee regardless of probability types. These advantages of scenario-based optimization rely heavily on the key assumption that sufficient amount of uncertainty data is available. However, in most practical cases, this assumption does not hold, and on the contrary the amount of uncertainty data sampled from the underlying true distribution is quite limited. Moreover, acquiring uncertainty data could be extremely expensive or time consuming in some specific cases, which greatly hinders the applicability of the scenario-based approach (Gupta and Rusmevichientong, 2017). Existing studies of the scenario-based optimization neglect the aforementioned practical situation (Campi et al., 2009; Zhang et al., 2015; Calafiore, 2017). The practical challenge of handling insufficient amount of data requires further research attention, and data-driven scenario-based optimization frameworks addressing this issue need to be developed.

This knowledge gap could be potentially filled by leveraging the power of deep generative models for the data-driven scenario-based optimization, whose schematic is shown in Fig. 3. Instead of assuming unlimited uncertainty scenarios sampled from the true distribution, deep generative models could be leveraged to learn
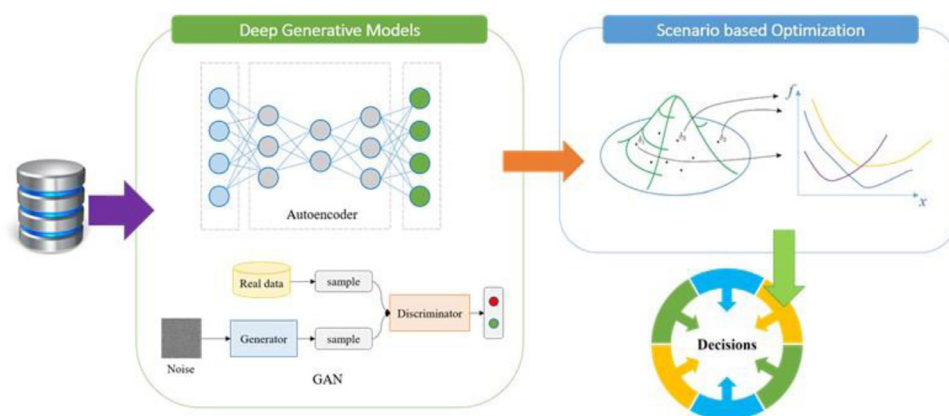
Fig. 3. The schematic of the scenario-based optimization framework based on deep learning.

the intrinsic useful patterns from the available uncertainty data and to generate synthetic data. These synthetic uncertainty data generated by the deep learning techniques mimic the real uncertainty data, and should be potentially useful in the scenario-based optimization model. Deep generative models can be utilized to generate synthetic uncertainty data with the aim for better decision with insufficient uncertainty data. To be more precise, in deep generative models, the true data distribution is learned either explicitly or implicitly, and then the learned distribution is used to generate new data points referred to as synthetic data. One of the most commonly used deep generative models is variational autoencoders (VAEs) (Goodfellow et al., 2016). VAEs generate new data samples through the architecture of synthesizing an encoder network and a decoder network in an unsupervised fashion. The function of encoders is to reduce the dimension of input data and extracts the latent features, while the decoder network aims to reconstruct data given the latent variables. In this way, the VAE model learns the complicated target distribution by maximizing the lower bound of the data log-likelihood. The advantage of this technique is that its quality is easily evaluated via log-likelihood or importance sampling. However, researchers have found out that VAEs typically tend to generate blurry images, meaning a noticeable difference between the true distribution and the learned one (Goodfellow et al., 2016). Recently, an emerging deep generative model named generative adversarial networks (GANs) was proposed and has become increasingly popular in various areas, such as image processing (Ledig et al., 2017), renewable scenario generation (Chen et al., 2018), and molecular designs (Sanchez-Lengeling and Aspuru-Guzik, 2018). Different from VAEs, GANs implicitly learn the data distribution through a zero-sum game between two competing neural networks, namely generator network and discriminator network (Goodfellow et al., 2014). Given the noise input, the generator network competes against the discriminator network by generating plausible synthetic data. On the contrary, the discriminator network attempts to distinguish the real uncertainty data from the synthetic data. These two networks compete against each other. Accordingly, the data distribution resulted from the generator network will be the true distribution once the Nash equilibrium is achieved. The required sample size for random convex programs scales linearly with the number of decision variables (Alamo et al., 2015), implying that the "small data" regime should be frequently encountered for large-scale optimization problems. Consequently, data-driven scenario-based optimization tends to suffer severely from the issue of insufficient uncertainty data. Leveraging the power of deep generative models could be a promising way to addressing this challenge. The required sample size to guarantee the constraint satisfaction could become large for optimization problems with a huge number of

decision variables and a small value of risk level (Alamo et al., 2015). Therefore, the available amount of uncertainty data might not be enough for the purpose of probabilistic guarantee. However, the number of uncertainty data can still be sufficient for training generative models.

Although deep learning could be a silver bullet in many areas, a lot of research challenges still persist in organically integrating the state-of-the-art deep learning techniques with optimization under uncertainty. The discussion in Section 4.2. is aimed to serve as a good starting point to promote the employment of deep learning in the field of data-driven optimization under uncertainty.

4.3. Online learning-based data-driven optimization: a learning-while-optimizing paradigm for addressing uncertainty

In conventional data-driven optimization frameworks, a batch of uncertainty data serves as input to the data-driven system, in which learning typically takes place only once and is termed as batch machine learning. Most, if not all, of the papers on data-driven optimization under uncertainty are restricted to such learning of data (Ning and You, 2017a; Esfahani and Kuhn, 2018; Jiang and Guan, 2016; Shang et al., 2017), so they fail to account for real-time data. For example, in data-driven robust optimization methods, uncertainty sets are learned from a batch of uncertainty data. Once these data-driven uncertainty sets are obtained, they remain fixed for the model-based system based on mathematical programming and are not updated or refined. Additionally, probability distributions of uncertainties and their support sets could be time variant and evolves gradually, rendering the data-driven system "outdated". Such obsolete data-driven system inevitably deteriorates the resulting solution quality of the mathematical programming problem. In many practical settings, uncertainty data are collected sequentially in an online fashion (Shalev-Shwartz, 2012). Although previous works have explored the online learning of uncertainty sets (Shang and You, 2019), they typically re-train the data-driven system from scratch using the existing and new addition of data, thus making these approaches suitable only for systems with slow dynamics. Therefore, few studies to date investigate the real-time data analytics for systems with fast dynamics such as those encountered in chemical processes, establishing research opportunities for the PSE community.

An online-learning-based data-driven optimization paradigm, in which learning takes place iteratively to account for real-time data, could be a promising research direction. More specifically, a learning-while-optimizing scheme could be explored by taking advantage of deep reinforcement learning. On the one hand, the uncertainty model should be time varying to accommodate real-time uncertainty data. On the other hand, decisions are made sequen-

tially under uncertainty. After decisions are made, uncertainties are realized and then collected in the database. There are research challenges associated with such online-learning-based frameworks. Updating the data-driven system in an online fashion is paramount in implementing the learning-while-optimizing scheme and poses a key research challenge. Additionally, developing efficient algorithms to solve the resulting online-learning-based mathematical programming problems creates the computational challenge. There exist some theoretical research challenges as well. One theoretical challenge is to investigate the convergence of solutions when the probability distribution shifts to a new one. Another challenge is to provide theoretical bounds for computational complexity and required memory for the online-learning-based data-driven optimization.

5. Conclusions

Although conventional stochastic programming, robust optimization, and chance constrained optimization are the most recognized modeling paradigms for hedging against uncertainty, it is foreseeable that in the near future data-driven mathematical programming frameworks would experience a rapid growth fueled by big data and deep learning. We reviewed recent progress of data-driven mathematical programming under uncertainty in terms of systematic uncertainty modeling, organic integration of machine learning and mathematical programming, and efficient computational algorithms for solving the resulting mathematical programming problems. The advantages and disadvantages of different data-driven uncertainty models were also analyzed in detail. Future research could be directed toward devising feedback steps to close the loop of the data-driven system and the model-based system, leveraging the power of deep generative models for the data-driven scenario-based optimization, and developing data-driven mathematical programming frameworks with online learning for real-time data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2019.03.034.

References

Aström, K.J., Kumar, P.R., 2014. Control: a perspective. Automatica 50, 3–43.
Acevedo, J., Pistikopoulos, E.N., 1998. Stochastic optimization based algorithms for process synthesis under uncertainty. Comput. Chem. Eng. 22, 647–671.
Ahmed, S., Tawarmalani, M., Sahinidis, N.V., 2004. A finite branch-and-bound algorithm for two-stage stochastic integer programs. Math. Program. 100, 355–377.
Alamo, T., Tempo, R., Camacho, E.F., 2009. Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems. IEEE Trans. Autom. Control 54, 2545–2559.
Alamo, T., Tempo, R., Luque, A., Ramirez, D.R., 2015. Randomized methods for design of uncertain systems: sample complexity and sequential algorithms. Automatica 52, 160–172.
Assavapokee, T., Realff, M.J., Ammons, J.C., 2008. Min-max regret robust optimization approach on interval data uncertainty. J. Optim. Theory Appl. 137, 297–316.
Atamtürk, A., Zhang, M., 2007. Two-stage robust network flow and design under demand uncertainty. Oper. Res. 55, 662–673.
Baringo, L., Baringo, A., 2018. A stochastic adaptive robust optimization approach for the generation and transmission expansion planning. IEEE Trans. Power Syst. 33, 792–802.
Bayraksan, G., Love, D.K., 2015. Data-Driven Stochastic Programming Using Phi-Divergences. The Operations Research Revolution, pp. 1–19.
Bell, D.E., 1982. Regret in decision making under uncertainty. Oper. Res. 30, 961–981.
Ben-Tal, A., Nemirovski, A., 2000. Robust solutions of linear programming problems contaminated with uncertain data. Math. Program. 88, 411.
Ben-Tal, A., Nemirovski, A., 2002. Robust optimization—methodology and applications. Math. Program. 92, 453–480.
Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A., 2004. Adjustable robust solutions of uncertain linear programs. Math. Program. 99, 351–376.
Ben-Tal, A., Ghaoui, L.E., Nemirovski, A., 2009. Robust Optimization. Princeton University Press.

Ben-Tal, A., El Ghaoui, L., Nemirovski, A., 2009. Robust Optimization. University Press, Princeton.
Bertsimas, D., Sim, M., 2004. The price of robustness. Oper. Res. 52, 35.
Bertsimas, D., Thiele, A., 2006. Robust and data-driven optimization: modern decision-making under uncertainty. INFORMS Tutor. Oper. Res. 95–122.
Bertsimas, D., Brown, D.B., Caramanis, C., 2011. Theory and applications of robust optimization. SIAM Rev. 53, 464–501.
Bertsimas, D., Litvinov, E., Sun, X.A., Zhao, J., Zheng, T., 2013. Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem. IEEE Trans. Power Syst. 28, 52–63.
Bertsimas, D., Gupta, V., Kallus, N., 2018. Data-driven robust optimization. Math. Program. 167, 235–292.
Bertsimas, D., Sim, M., Zhang, M., 2019. Adaptive distributionally robust optimization. Manag. Sci. 0 null.
Biegler, L.T., Grossmann, I.E., 2004. Retrospective on optimization. Comput. Chem. Eng. 28, 1169–1192.
Birge, J.R., Louveaux, F., 2011. Introduction to Stochastic Programming. Springer Science & Business Media.
Birge, J.R., 1997. State-of-the-Art-Survey—stochastic programming: computation and applications. INFORMS J. Comput. 9, 111–133.
Bonfill, A., Bagajewicz, M., Espuña, A., Puigjaner, L., 2004. Risk management in the scheduling of batch plants under uncertain market demand. Ind. Eng. Chem. Res. 43, 741–750.
Bonfill, A., Espuña, A., Puigjaner, L., 2005. Addressing robustness in scheduling batch processes with uncertain operation times. Ind. Eng. Chem. Res. 44, 1524–1534.
Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.
Calafiore, G., Campi, M.C., 2005. Uncertain convex programs: randomized solutions and confidence levels. Math. Program. 102, 25–46.
Calafiore, G.C., El Ghaoui, L., 2006. On distributionally robust chance-constrained linear programs. J. Optim. Theory Appl. 130, 1–22.
Calafiore, G., Dabbene, F., Tempo, R., 2011. Research on probabilistic methods for control system design. Automatica 47, 1279–1293.
Calafiore, G., Lyons, D., Fagiano, L., 2012. On mixed-integer random convex programs. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), pp. 3508–3513.
Calafiore, G., 2009. On the expected probability of constraint violation in sampled convex programs. J. Optim. Theory Appl. 143, 405–412.
Calafiore, G.C., 2010. Random convex programs. SIAM J. Optim. 20, 3427–3464.
Calafiore, G., 2017. Repetitive scenario design. IEEE Trans. Autom. Control 62, 1125–1137.
Calfa, B.A., Agarwal, A., Bury, S.J., Wassick, J.M., Grossmann, I.E., 2015. Data-driven simulation and optimization approaches to incorporate production variability in sales and operations planning. Ind. Eng. Chem. Res. 54, 7261–7272.
Calfa, B.A., Grossmann, I.E., Agarwal, A., Bury, S.J., Wassick, J.M., 2015. Data-driven individual and joint chance-constrained optimization via kernel smoothing. Comput. Chem. Eng. 78, 51–69.
Campbell, T., How, J.P., 2015. Bayesian nonparametric set construction for robust optimization. In: American Control Conference (ACC), 2015, pp. 4216–4221.
Campi, M.C., Garatti, S., 2008. The exact feasibility of randomized solutions of uncertain convex programs. SIAM J. Optim. 19, 1211–1230.
Campi, M.C., Garatti, S., 2011. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. J. Optim. Theory Appl. 148, 257–280.
Campi, M.C., Garatti, S., 2018. Wait-and-judge scenario optimization. Math. Program. 167, 155–189.
Campi, M.C., Garatti, S., Prandini, M., 2009. The scenario approach for systems and control design. Ann. Rev. Control 33, 149–157.
Campi, M.C., Garatti, S., Ramponi, F.A., 2018. A general scenario theory for non-convex optimization and decision making. IEEE Trans. Autom. Control 63, 4067–4078.
Cannon, M., Kouvaritakis, B., Wu, X.J., 2009. Probabilistic constrained MPC for multiplicative and additive stochastic uncertainty. IEEE Trans. Autom. Control 54, 1626–1632.
Care, A., Garatti, S., Campi, M.C., 2014. FAST-Fast algorithm for the scenario technique. Oper. Res. 62, 662–671.
Carlone, L., Srivastava, V., Bullo, F., Calafiore, G.C., 2014. Distributed random convex programming via constraints consensus. SIAM J. Control Optim. 52, 629–662.
Caroe, C.C., Schultz, R., 1999. Dual decomposition in stochastic integer programming. Oper. Res. Lett. 24, 37–45.
Chamanbaz, M., Dabbene, F., Tempo, R., Venkataramanan, V., Wang, Q.G., 2016. Sequential randomized algorithms for convex optimization in the presence of uncertainty. IEEE Trans. Autom. Control 61, 2565–2571.
Charnes, A., Cooper, W.W., 1959. Chance-constrained programming. Manag. Sci. 6, 73–79.
Chen, Y.W., Guo, Q.L., Sun, H.B., Li, Z.S., Wu, W.C., Li, Z.H., 2018. A Distributionally robust optimization model for unit commitment based on Kullback-Leibler divergence. IEEE Trans. Power Syst. 33, 5147–5160.
Chen, Z., Peng, S., Liu, J., 2018. Data-driven robust chance constrained problems: a mixture model approach. J. Optim. Theory Appl. 179, 1065–1085.
Chen, Y.Z., Wang, Y.S., Kirschen, D., Zhang, B.S., 2018. Model-free renewable scenario generation using generative adversarial networks. IEEE Trans. Power Syst. 33, 3265–3275.
Chen, Z., Kuhn, D., Wiesemann, W., 2018. Data-Driven Chance Constrained Programs Over Wasserstein Balls arXiv:1809.00210.

Cheng, J., Delage, E., Lisser, A., 2014. Distributionally robust stochastic knapsack problem. *SIAM J. Optim.* 24, 1485–1506.

Chiu, T.Y., Christofides, P.D., 2000. Robust control of particulate processes using uncertain population balances. *AIChE J.* 46, 266–280.

Chu, Y., You, F., 2013. Integration of scheduling and dynamic optimization of batch processes under uncertainty: two-stage stochastic programming approach and enhanced generalized benders decomposition algorithm. *Ind. Eng. Chem. Res.* 52, 16851–16869.

Chu, Y., You, F., Wassick, J.M., Agarwal, A., 2015. Simulation-based optimization framework for multi-echelon inventory systems under uncertainty. *Comput. Chem. Eng.* 73, 1–16.

Chu, Y., You, F., Wassick, J.M., Agarwal, A., 2015. Integrated planning and scheduling under production uncertainties: bi-level model formulation and hybrid solution method. *Comput. Chem. Eng.* 72, 255–272.

De Loera, J.A., La Haye, R.N., Oliveros, D., Roldan-Pensado, E., 2018. Chance-constrained convex mixed-integer optimization and beyond: two sampling algorithms within s-optimization. *J. Convex Anal.* 25, 201–218.

Delage, E., Iancu, D.A., 2015. In: Aleman, DM, Thiele, AC (Eds.), Robust Multistage Decision Making. INFORMS Tutorials in Operations Research, Catonsville, MD, pp. 20–46.

Delage, E., Ye, Y.Y., 2010. Distributionally Robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58, 595–612.

Duan, C., Jiang, L., Fang, W.L., Liu, J., 2018. Data-driven affinely adjustable distributionally robust unit commitment. *IEEE Trans. Power Syst.* 33, 1385–1398.

El Ghaoui, L., Oks, M., Oustry, F., 2003. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Oper. Res.* 51, 543–556.

Erdogan, E., Iyengar, G., 2006. Ambiguous chance constrained problems and robust optimization. *Math. Program.* 107, 37–61.

Esfahani, P.M., Kuhn, D., 2018. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.* 171, 115–166.

Esfahani, P.M., Sutter, T., Lygeros, J., 2015. Performance bounds for the scenario approach and an extension to a class of non-convex programs. *IEEE Trans. Autom. Control* 60, 46–58.

Gao, R., Kleywegt, A.J., 2016. Distributionally Robust Stochastic Optimization With Wasserstein Distance arXiv:1604.02199.

Gao, J., You, F., 2015. Deciphering and handling uncertainty in shale gas supply chain design and optimization: novel modeling framework and computationally efficient solution algorithm. *AIChE J.* 61, 3739–3755.

Gao, J., You, F., 2017. Modeling framework and computational algorithm for hedging against uncertainty in sustainable supply chain design using functional-unit-based life cycle optimization. *Comput. Chem. Eng.* 107, 221–236.

Gao, X.Y., Shang, C., Jiang, Y.H., Huang, D.X., Chen, T., 2014. Refinery scheduling with varying crude: a deep belief network classification and multimodel approach. *AIChE J.* 60, 2525–2532.

Gao, J., Ning, C., You, F., 2019. Data-driven distributionally robust optimization of shale gas supply chains under uncertainty. *AIChE J.* 65, 947–963.

Gawehn, E., Hiss, J.A., Schneider, G., 2016. Deep learning in drug discovery. *Mol. Inf.* 35, 3–14.

Gebreslassie, B.H., Yao, Y., You, F., 2012. Design under uncertainty of hydrocarbon biorefinery supply chains: multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk. *AIChE J.* 58, 2155–2179.

Ghosal, S., Wiesemann, W., 2018. The Distributionally Robust Chance Constrained Vehicle Routing Problem Available Optimization Online.

Goel, V., Grossmann, I.E., 2007. A class of stochastic programs with decision dependent uncertainty. *Math. Program.* 108, 355–394.

Gong, J., You, F., 2017. Optimal processing network design under uncertainty for producing fuels and value-added bioproducts from microalgae: two-stage adaptive robust mixed integer fractional programming model and computationally efficient solution algorithm. *AIChE J.* 63, 582–600.

Gong, J., You, F., 2018. Resilient design and operations of process systems: nonlinear adaptive robust optimization model and algorithm for resilience analysis and enhancement. *Comput. Chem. Eng.* 116, 231–252.

Gong, J., Garcia, D.J., You, F., 2016. Unraveling optimal biomass processing routes from bioconversion product and process networks under uncertainty: an adaptive robust optimization approach. *ACS Sustain. Chem. Eng.* 4, 3160–3173.

Gong, J., Yang, M., You, F., 2017. A systematic simulation-based process intensification method for shale gas processing and NGLs recovery process systems under uncertain feedstock compositions. *Comput. Chem. Eng.* 105, 259–275.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al., 2014. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.). Generative Adversarial Nets, 27. Advances in Neural Information Processing Systems 27.

Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning, 1. MIT Press Cambridge.

Grammatico, S., Zhang, X.J., Margellos, K., Goulart, P., Lygeros, J., 2016. A scenario approach for non-convex control design. *IEEE Trans. Autom. Control* 61, 334–345.

Graves, A., Mohamed, A.R., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics. Speech and Signal Processing, pp. 6645–6649.

Gregory, C., Darby-Dowman, K., Mitra, G., 2011. Robust optimization and portfolio selection: the cost of robustness. *Eur. J. Oper. Res.* 212, 417–428.

Grossmann, I.E., Biegler, L.T., 2004. Part II. Future perspective on optimization. *Comput. Chem. Eng.* 28, 1193–1218.

Grossmann, I.E., 2012. Advances in mathematical programming models for enterprise-wide optimization. *Comput. Chem. Eng.* 47, 2–18.

Guo, Y., Baker, K., Dall'Anese, E., Hu, Z., Summers, T., 2018. Stochastic optimal power flow based on data-driven distributionally robust optimization. In: 2018 Annual American Control Conference (ACC), pp. 3840–3846.

Gupta, V., Grossmann, I.E., 2014. A new decomposition algorithm for multistage stochastic programs with endogenous uncertainties. *Comput. Chem. Eng.* 62, 62–79.

Gupta, A., Maranas, C.D., 2003. Managing demand uncertainty in supply chain planning. *Comput. Chem. Eng.* 27, 1219–1227.

Gupta, V., Rusmevichientong, P., 2017. Small-Data, Large-Scale Linear Optimization.

Gupta, A., Maranas, C.D., McDonald, C.M., 2000. Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management. *Comput. Chem. Eng.* 24, 2613–2621.

Hanasusanto, G.A., Kuhn, D., 2018. Conic Programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Oper. Res.* 66, 849–869.

Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W., 2015. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Math. Program.* 151, 35–62.

Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W., 2017. Ambiguous joint chance constraints under mean and dispersion information. *Oper. Res.* 65, 751–767.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.

Hong, L.J., Yang, Y., Zhang, L.W., 2011. Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach. *Oper. Res.* 59, 617–630.

Hota, A.R., Cherukuri, A., Lygeros, J., 2018. Data-Driven Chance Constrained Optimization Under Wasserstein Ambiguity Sets arXiv:1805.06729.

Hu, Z., Hong, L.J., 2013. Kullback-Leibler Divergence Constrained Distributionally Robust Optimization Available at Optimization Online.

Ierapetritou, M.G., Pistikopoulos, E.N., 1995. Design of multiproduct batch plants with uncertain demands. *Comput. Chem. Eng.* 19, S627–S632.

Ji, R., Lejeune, M., 2018. Data-driven distributionally robust chance-constrained programming with Wasserstein metric.

Jiang, R., Guan, Y., 2015. Data-driven chance constrained stochastic program. *Math. Program.* 158, 291–327.

Jiang, R.W., Guan, Y.P., 2016. Data-driven chance constrained stochastic program. *Math. Program.* 158, 291–327.

John Walker, S., 2014. Big Data: A Revolution That will Transform how we live, work, and think. Taylor & Francis.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260.

Küçükyavuz, S., Sen, S., 2017. An Introduction to Two-Stage Stochastic Mixed-Integer Programming. Leading Developments From INFORMS Communities, pp. 1–27. INFORMS.

Kall, P., Wallace, S.W., 1994. Stochastic Programming. John Wiley and Sons Ltd.

Kanamori, T., Takeda, A., 2012. Worst-case violation of sampled convex programs for optimization with uncertainty. *J. Optim. Theory Appl.* 152, 171–197.

Kariotoglou, N., Margellos, K., Lygeros, J., 2016. On the computational complexity and generalization properties of multi-stage and stage-wise coupled scenario programs. *Syst. Control Lett.* 94, 63–69.

Keyvanshokooh, E., Ryan, S.M., Kabir, E., 2016. Hybrid robust and stochastic optimization for closed-loop supply chain network design using accelerated Benders decomposition. *Eur. J. Oper. Res.* 249, 76–92.

Klabjan, D., Simchi-Levi, D., Song, M., 2013. Robust stochastic lot-sizing by means of histograms. *Prod. Oper. Manag.* 22, 691–710.

Krieger, A., Pistikopoulos, E.N., 2014. Model predictive control of anesthesia under uncertainty. *Comput. Chem. Eng.* 71, 699–707.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet Classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90.

Laporte, G., Louveaux, F.V., 1993. The integer L-shaped method for stochastic integer programs with complete recourse. *Oper. Res. Lett.* 13, 133–142.

Lappas, N.H., Gounaris, C.E., 2016. Multi-stage adjustable robust optimization for process scheduling under uncertainty. *AIChE J.* 62, 1646–1667.

Lasserre, J., Weisser, T., 2018. Distributionally Robust Polynomial Chance-Constraints Under Mixture Ambiguity Sets.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al., 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. CVPR, p. 4.

Lee, J.H., Shin, J., Realff, M.J., 2018. Machine learning: overview of the recent progresses and implications for the process systems engineering field. *Comput. Chem. Eng.* 114, 111–121.

Levi, R., Perakis, G., Uichanco, J., 2015. The data-driven newsvendor problem: new bounds and insights. *Oper. Res.* 63, 1294–1306.

Li, C., Grossmann, I.E., 2018. An improved L-shaped method for two-stage convex 0–1 mixed integer nonlinear stochastic programs. *Comput. Chem. Eng.* 112, 165–179.

Li, Z., Ierapetritou, M.G., 2008. Process scheduling under uncertainty: review and challenges. *Comput. Chem. Eng.* 32, 715–727.

Li, P., Arellano-Garcia, H., Wozny, G., 2008. Chance constrained programming approach to process optimization under uncertainty. *Comput. Chem. Eng.* 32, 25–45.

Li, P., Arellano-Garcia, H., Wozny, G., 2008. Chance constrained programming ap-

proach to process optimization under uncertainty. *Comput. Chem. Eng.* 32, 25–45.

Li, X., Tomasgard, A., Barton, P.I., 2011. Nonconvex generalized benders decomposition for stochastic separable mixed-integer nonlinear programs. *J. Optim. Theory Appl.* 151, 425–454.

Li, B., Jiang, R., Mathieu, J.L., 2017. Ambiguous risk constraints with moment and unimodality information. *Math. Program.* 173, 151–192.

Liu, M.L., Sahinidis, N.V., 1996. Optimization in process planning under uncertainty. *Ind. Eng. Chem. Res.* 35, 4154–4165.

Liu, P., Pistikopoulos, E.N., Li, Z., 2010. Decomposition based stochastic programming approach for polygeneration energy systems design under uncertainty. *Ind. Eng. Chem. Res.* 49, 3295–3305.

Liu, X., Kucukyavuz, S., Luedtke, J., 2016. Decomposition algorithms for two-stage chance-constrained programs. *Math. Program.* 157, 219–243.

Liu, S.S., Farid, S.S., Papageorgiou, L.G., 2016. Integrated optimization of upstream and downstream processing in biopharmaceutical manufacturing under uncertainty: a chance constrained programming approach. *Ind. Eng. Chem. Res.* 55, 4599–4612.

Liu, C., Lee, C., Chen, H., Mehrotra, S., 2016. Stochastic robust mathematical programming model for power system optimization. *IEEE Trans. Power Syst.* 31, 821–822.

Lorca, Á., Sun, X.A., Litvinov, E., Zheng, T., 2016. Multistage adaptive robust optimization for the unit commitment problem. *Oper. Res.* 61, 32–51.

Luedtke, J., Ahmed, S., 2008. A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.* 19, 674–699.

Maranas, C.D., 1997. Optimization accounting for property prediction uncertainty in polymer design. *Comput. Chem. Eng.* 21, S1019–S1024.

Margellos, K., Falsone, A., Garatti, S., Prandini, M., 2018. Distributed constrained optimization and consensus in uncertain networks via proximal minimization. *IEEE Trans. Autom. Control* 63, 1372–1387.

McLean, K., Li, X., 2013. Robust scenario formulations for strategic supply chain optimization under uncertainty. *Ind. Eng. Chem. Res.* 52, 5721–5734.

Mesbah, A., 2016. Stochastic model predictive control an overview and perspectives for future research. *IEEE Control Syst. Mag.* 36, 30–44.

Miao, F., Han, S., Lin, S., Wang, Q., Stankovic, J.A., Hendawi, A., et al., 2019. Data–driven robust taxi dispatch under demand uncertainties. *IEEE Trans. Control Syst. Technol.* 27, 175–191.

Miller, B.L., Wagner, H.M., 1965. Chance constrained programming with joint constraints. *Oper. Res.* 13, 930.

Mitra, K., Gudi, R.D., Patwardhan, S.C., Sardar, G., 2008. Midterm supply chain planning under uncertainty: a multiobjective chance constrained programming framework. *Ind. Eng. Chem. Res.* 47, 5501–5511.

Mohamed, A.R., Dahl, G.E., Hinton, G., 2012. Acoustic modeling using deep belief networks. *IEEE Trans. Audio Speech Lang. Process.* 20, 14–22.

Nemirovski, A., Shapiro, A., 2006. Convex approximations of chance constrained programs. *SIAM J. Optim.* 17, 969–996.

Ning, C., You, F., 2017a. Data-driven adaptive nested robust optimization: general modeling framework and efficient computational algorithm for decision making under uncertainty. *AIChE J.* 63, 3790–3817.

Ning, C., You, F., 2017b. A data-driven multistage adaptive robust optimization framework for planning and scheduling under uncertainty. *AIChE J.* 63, 4343–4369.

Ning, C., You, F., 2018a. Adaptive robust optimization with minimax regret criterion: multiobjective optimization framework and computational algorithm for planning and scheduling under uncertainty. *Comput. Chem. Eng.* 108, 425–447.

Ning, C., You, F., 2018b. Data-driven stochastic robust optimization: general computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era. *Comput. Chem. Eng.* 111, 115–133.

Ning, C., You, F., 2018c. Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. *Comput. Chem. Eng.* 112, 190–210.

Ning, C., You, F., 2019. Data-driven adaptive robust unit commitment under wind power uncertainty: a Bayesian nonparametric approach. *IEEE Trans. Power Syst.* doi:10.1109/TPWRS.2019.2891057.

Oliveira, F., Gupta, V., Hamacher, S., Grossmann, I.E., 2013. A Lagrangean decomposition approach for oil supply chain investment planning under uncertainty with risk considerations. *Comput. Chem. Eng.* 50, 184–195.

Parpas, P., Rustem, B., Pistikopoulos, E., 2009. Global optimization of robust chance constrained problems. *J. Global Optim.* 43, 231–247.

Peng, X., Root, T.W., Maravelias, C.T., 2018. Optimization-based process synthesis under seasonal and daily variability: application to concentrating solar power. *AIChE J.* doi:10.1002/aic.16458.

Pistikopoulos, E.N., 1995. Uncertainty in process design and operations. *Comput. Chem. Eng.* 19, 553–563.

Postek, K., Ben-Tal, A., den Hertog, D., Melenberg, B., 2018. Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Oper. Res.* 66, 814–833.

Prékopa, A., 1995. Stochastic Programming, Volume 324 of Mathematics and Its Applications. Kluwer Academic Publishers Group, Dordrecht.

Qin, S.J., 2014. Process data analytics in the era of big data. *AIChE J.* 60, 3092–3100.

Quddus, M.A., Chowdhury, S., Marufuzzaman, M., Yu, F., Bian, L.K., 2018. A two-stage chance-constrained stochastic programming model for a bio-fuel supply chain network. *Int. J. Prod. Econ.* 195, 27–44.

Rooney, W.C., Biegler, L.T., 2003. Optimal process design with model parameter uncertainty and process variability. *AIChE J.* 49, 438–449.

Sahinidis, N.V., 2004. Optimization under uncertainty: state-of-the-art and opportunities. *Comput. Chem. Eng.* 28, 971–983.

Sakizlis, V., Perkins, J.D., Pistikopoulos, E.N., 2004. Recent advances in optimization-based simultaneous process and control design. *Comput. Chem. Eng.* 28, 2069–2086.

Sanchez-Lengeling, B., Aspuru-Guzik, A., 2018. Inverse molecular design using machine learning: generative models for matter engineering. *Science* 361, 360–365.

Shalev-Shwartz, S., 2012. Online learning and online convex optimization. *Found. Trends Mach. Learn.* 4, 107–194.

Shang, C., You, F., 2018. Distributionally robust optimization for planning and scheduling under uncertainty. *Comput. Chem. Eng.* 110, 53–68.

Shang, C., You, F., 2019. A data-driven robust optimization approach to stochastic model predictive control. *J. Process Control* 75, 24–39.

Shang, C., Yang, F., Huang, D.X., Lyu, W.X., 2014. Data-driven soft sensor development based on deep learning technique. *J. Process Control* 24, 223–233.

Shang, C., Huang, X., You, F., 2017. Data-driven robust optimization based on kernel learning. *Comput. Chem. Eng.* 106, 464–479.

Shen, W., Li, Z., Huang, B., Jan, N.M., 2018. Chance-constrained model predictive control for SAGD process using robust optimization approximation. *Ind. Eng. Chem. Res.* doi:10.1021/acs.iecr.8b03207.

Shi, H., You, F., 2016. A computational framework and solution algorithms for two-stage adaptive robust scheduling of batch manufacturing processes under uncertainty. *AIChE J.* 62, 687–703.

Shin, J., Lee, J.H., 2019. Multi-timescale, multi-period decision-making model development by combining reinforcement learning and mathematical programming. *Comput. Chem. Eng.* 121, 556–573.

Smith, J.E., Winkler, R.L., 2006. The optimizer's curse: skepticism and postdecision surprise in decision analysis. *Manag. Sci.* 52, 311–322.

Soyster, A.L., 1973. Technical note—convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper. Res.* 21, 1154–1157.

Steimel, J., Engell, S., 2015. Conceptual design and optimization of chemical processes under uncertainty by two-stage programming. *Comput. Chem. Eng.* 81, 200–217.

Tong, K., Gong, J., Yue, D., You, F., 2014. Stochastic programming approach to optimal design and operations of integrated hydrocarbon biofuel and petroleum supply chains. *ACS Sustain. Chem. Eng.* 2, 49–61.

Tong, K., You, F., Rong, G., 2014. Robust design and operations of hydrocarbon biofuel supply chain integrating with existing petroleum refineries considering unit cost objective. *Comput. Chem. Eng.* 68, 128–139.

Uryasev, S., 2000. Conditional value-at-risk: Optimization algorithms and applications. In: Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr)(Cat. No. 00TH8520). IEEE, pp. 49–57.

Van Parys, B.P.G., Kuhn, D., Goulart, P.J., Morari, M., 2016. Distributionally robust control of constrained stochastic systems. *IEEE Trans. Autom. Control* 61, 430–442.

Vanslyke, R.M., Wets, R., 1969. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM J. Appl. Math.* 17, 638.

Vayanos, P., Kuhn, D., Rustem, B., 2012. A constraint sampling approach for multi-stage robust optimization. *Automatica* 48, 459–471.

Venkatasubramanian, V., 2019. The promise of artificial intelligence in chemical engineering: is it here, finally? *AIChE J.* 65, 466–478.

Verderame, P.M., Elia, J.A., Li, J., Floudas, C.A., 2010. Planning and scheduling under uncertainty: a review across multiple sectors. *Ind. Eng. Chem. Res.* 49, 3993–4017.

Vermaak, J., Botha, E.C., 1998. Recurrent neural networks for short-term load forecasting. *IEEE Trans. Power Syst.* 13, 126–132.

Wang, C., Gao, R., Qiu, F., Wang, J., Xin, L., 2018. Risk-based distributionally robust optimal power flow with dynamic line rating. *IEEE Trans. Power Syst.* 33, 6074–6086.

Wiesemann, W., Kuhn, D., Sim, M., 2014. Distributionally robust convex optimization. *Oper. Res.* 62, 1358–1376.

Wu, Y., Tan, H., Qin, L., Ran, B., Jiang, Z., 2018. A hybrid deep learning based traffic flow prediction method and its understanding. *Transp. Res. Part C* 90, 166–180.

Xie, W.J., Ahmed, S., 2018. On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM J. Optim.* 28, 1151–1182.

Xie, W., Ahmed, S., 2018. Distributionally robust chance constrained optimal power flow with renewables: a conic reformulation. *IEEE Trans. Power Syst.* 33, 1860–1867.

Xie, W., 2018. On Distributionally Robust Chance Constrained Program with Wasserstein Distance arXiv:1806.07418.

Xiong, P., Jirutitijaroen, P., Singh, C., 2017. A distributionally robust optimization model for unit commitment considering uncertain wind power generation. *IEEE Trans. Power Syst.* 32, 39–49.

Yang, W.Z., Xu, H., 2016. Distributionally robust chance constraints for non-linear uncertainties. *Math. Program.* 155, 231–265.

Yang, Y., Vayanos, P., Barton, P.I., 2017. Chance-constrained optimization for refinery blend planning under uncertainty. *Ind. Eng. Chem. Res.* 56, 12139–12150.

Ye, W., You, F., 2016. A computationally efficient simulation-based optimization method with region-wise surrogate modeling for stochastic inventory management of supply chains with general network structures. *Comput. Chem. Eng.* 87, 164–179.

Yin, S., Kaynak, O., 2015. Big data for modern industry: challenges and trends. *Proc. IEEE* 103, 143–146.

Yin, S., Li, X., Gao, H., Kaynak, O., 2015. Data-based techniques focused on modern industry: an overview. *IEEE Trans. Ind. Electron.* 62, 657–667.

You, F., Grossmann, I.E., 2008. Mixed-integer nonlinear programming models and algorithms for large-scale supply chain design with stochastic inventory management. *Ind. Eng. Chem. Res.* 47, 7802–7817.

You, F., Grossmann, I.E., 2011. Stochastic inventory management for tactical process planning under uncertainties: MINLP models and Algorithms. *AIChE J.* 57, 1250–1277.

You, F., Grossmann, I.E., 2011. Balancing responsiveness and economics in process supply chain design with multi-echelon stochastic inventory. *AIChE J.* 57, 178–192.

You, F., Grossmann, I.E., 2013. Multicut Benders decomposition algorithm for process supply chain planning under uncertainty. *Ann. Oper. Res.* 210, 191–211.

You, F., Wassick, J.M., Grossmann, I.E., 2009. Risk management for a global supply chain planning under uncertainty: models and algorithms. *AIChE J.* 55, 931–946.

You, F., Pinto, J.M., Grossmann, I.E., Megan, L., 2011. Optimal distribution-inventory planning of industrial gases. II. MINLP models and algorithms for stochastic cases. *Ind. Eng. Chem. Res.* 50, 2928–2945.

You, K., Tempo, R., Xie, P., 2018. Distributed algorithms for robust convex optimization via the scenario approach. *IEEE Trans. Autom. Control* 64 1-1.

Yue, D., You, F., 2013. Planning and scheduling of flexible process networks under uncertainty with stochastic inventory: MINLP models and algorithm. *AIChE J.* 59, 1511–1532.

Yue, D., You, F., 2016. Optimal supply chain design and operations under multi-scale uncertainties: nested stochastic robust optimization modeling framework and solution algorithm. *AIChE J.* 62, 3041–3055.

Zeballos, L.J., Méndez, C.A., Barbosa-Povoa, A.P., 2016. Design and planning of closed-loop supply chains: a risk-averse multistage stochastic approach. *Ind. Eng. Chem. Res.* 55, 6236–6249.

Zhang, Z.P., Zhao, J.S., 2017. A deep belief network based fault diagnosis model for complex chemical processes. *Comput. Chem. Eng.* 107, 395–407.

Zhang, X.J., 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., Jiang, R., Shen, S., 2018. Ambiguous chance-constrained binary programs under mean-covariance information. *SIAM J. Optim.* 28, 2922–2944.

Zhang, Y., Jin, X.Z., Feng, Y.P., Rong, G., 2018. Data-driven robust optimization under correlated uncertainty: a case study of production scheduling in ethylene plant (Reprinted from computers and Chemical Engineering, vol 109, pg 48-67, 2017). *Comput. Chem. Eng.* 116, 17–36.

Zhang, Y., Feng, Y.P., Rong, G., 2018. Data-driven rolling-horizon robust optimization for petrochemical scheduling using probability density contours. *Comput. Chem. Eng.* 115, 342–360.

Zhao, C.Y., Guan, Y.P., 2013. Unified stochastic and robust unit commitment. *IEEE Trans. Power Syst.* 28, 3353–3361.

Zhao, C.Y., Guan, Y.P., 2016. Data-driven stochastic unit commitment for integrating wind generation. *IEEE Trans. Power Syst.* 31, 2587–2596.

Zhao, S., You, F., 2019. Resilient supply chain design and operations with decision-dependent uncertainty using a data-driven robust optimization approach. *AIChE J.* 65, 1006–1021.

Zhao, L, Ning, C., You, F., 2019. Operational optimization of industrial steam systems under uncertainty using data-driven adaptive robust optimization. *AIChE J.* doi:10.1002/aic.16500.

Zhu, W., Ma, Y., Benton, M.G., Romagnoli, J.A., Zhan, Y., 2019. Deep learning for pyrolysis reactor monitoring: from thermal imaging toward smart monitoring system. *AIChE J.* 65, 582–591.

Zipkin, P.H., 2000. Foundations of inventory management. McGraw-Hill.

Zymler, S., Kuhn, D., Rustem, B., 2013. Distributionally robust joint chance constraints with second-order moment information. *Math. Program.* 137, 167–198.