Computers and Chemical Engineering 110 (2018) 53–68

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Distributionally robust optimization for planning and scheduling under uncertainty

Chao Shang, Fengqi You*

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

ARTICLE INFO

Article history:
Received 26 September 2017
Revised 1 December 2017
Accepted 3 December 2017
Available online 8 December 2017

Keywords:
Distributionally robust optimization
Decision-making under uncertainty
Multi-stage decision-making
Process scheduling
Process planning
Big data

ABSTRACT

Distributionally robust optimization (DRO) is an emerging and effective method to address the inexactness of probability distributions of uncertain parameters in decision-making under uncertainty. We propose an effective DRO framework for planning and scheduling under demand uncertainties. A novel data-driven approach is proposed to construct ambiguity sets based on principal component analysis and first-order deviation functions, which help excavating accurate and useful information from uncertainty data. Moreover, it leads to mixed-integer linear reformulations of planning and scheduling problems. To account for the multi-stage sequential decision-making structure in process operations, we further develop multi-stage DRO models and adopt affine decision rules to address the computational issue. Applications in industrial-scale process network planning and batch process scheduling demonstrate that, the proposed DRO approach can effectively leverage uncertainty data information, better hedge against distributional ambiguity, and yield more profits.

1. Introduction

With increasingly fierce competitions and stringent requirements in the global market, greater challenges have been encountered in planning and scheduling, which are important decision-making problems in process industries (Chu and You, 2015; Grossmann et al., 2016; Li and Ierapetritou, 2008; Sahinidis, 2004). A crucial issue of particular interests is to devise effective robust decision-making tools in order to balance between profitability and robustness of process operations in an uncertain environment. In the past two decades, stochastic programming (SP) and robust optimization (RO) have been extensively adopted in process planning and scheduling for better immunization against uncertainties (Acevedo and Pistikopoulos, 1998; Ahmed and Sahinidis, 1998; Gong et al., 2016; Grossmann et al., 2016; Lin et al., 2004; Tong et al., 2014; Yue and You, 2016; Zhang et al., 2016). As an emerging approach to modeling and tackling parameter uncertainties in optimization problems, distributionally robust optimization (DRO) could effectively hedge against the ambiguity of probability distribution of uncertainties, and has attracted immense research attentions in operations research community (Delage and Ye, 2010; Hanasusanto et al., 2017; Liu et al., 2017; Wiesemann et al., 2014).

Intuitively, DRO can be deemed as optimizing the worst-case expected performance on a set constituted by an infinite number of distributions, typically referred to as the "ambiguity set". Analogues to RO considering the worst-case parameter realization in the uncertainty set, DRO seeks to protect against the worst-case probability distribution in the ambiguity set. Fig. 1 illustratively visualizes the differences and connections between various optimization regimes including deterministic optimization, RO, SP, and DRO. Because SP seeks to optimize the expected value, an exact distribution of uncertainties is required, which cannot be accurately estimated with empirical data. By contrast, RO asks for the support of uncertain parameters only; however, the worst-case realization that RO hedges against might be unrealistic in practice. Different from SP and RO, DRO could hedge against *the inexactness of probability distributions* in virtue of the ambiguity set. The ambiguity set is typically constructed based on partial distributional information, such as support set and moment statistics, which can be readily obtained from available empirical data (Delage and Ye, 2010). In this way, DRO could effectively leverage data information and optimize the expected value without any presumption about the probability distribution of uncertainties, which SP typically relies on. On the other hand, DRO can avoid over-conservative solutions by incorporating partial stochastic information, which is disregarded by RO. Because of such merits, a number of applications of DRO have emerged recently in various fields, such as power systems (Wei et al., 2016; Xiong et al., 2017; Zhao and Jiang, 2017) and service systems (Jiao et al., 2017; Nakao et al., 2017; Zhang et al., 2017).

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

https://doi.org/10.1016/j.compchemeng.2017.12.002

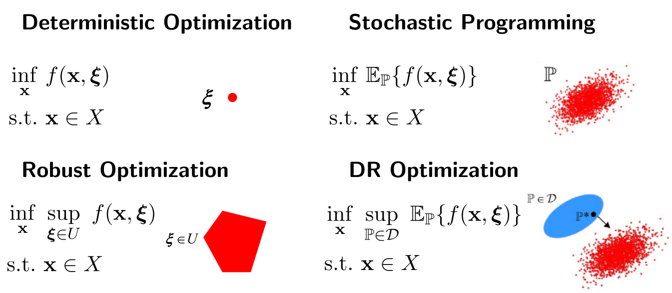


Fig. 1. Differences and connections between different optimization schemes. The uncertainty set in RO is expressed as a red polytope including different parameter configurations, while the ambiguity set in DRO is expressed as a blue ellipsoid that contains an infinite number of distributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Up to now, process planning and scheduling applications have not been benefited from recent advances in DRO. The orientation of this work is hence to fill this knowledge gap by developing a novel DRO framework for process planning and scheduling. To achieve this goal, several research challenges must be appropriately addressed. First, a DRO problem must admit a tractable and efficient reformulation that can be handled by numerical methods in practical applications. Although DRO problems based on classic ambiguity sets can be generally translated into convex optimization problems, they inevitably involve semi-definite positive (SDP) programs, which could be very challenging to solve for large-scale application problems on process planning and scheduling. The second challenge is how to determine parameters of the ambiguity set in a systematic way such that high-order statistical moment information within data can be effectively extracted. In the existing literature, such parameters are usually assumed as known *a priori*. Third, decisions in process operations are not made at the same stage, and some decisions shall be implemented in a "wait-and-see" manner after uncertainty realizations to reduce conservatism (Ben-Tal et al., 2004). In this sense, we need to integrate a complicated multi-stage and multi-level decision-making structure into the DRO framework, and devise efficient solution algorithms.

To address these research challenges, in this paper we first propose a systematic data-driven approach for ambiguity set constructions based on principal component analysis (PCA) and first-order deviation moment functions, whose merits lie in the following two aspects. First, the DRO problem can be readily translated into a classic RO problem with additional linear constraints, which turns out to be computationally beneficial in solving large-scale process operation problems. Specifically, if the deterministic problem is a mixed-integer linear program (MILP), the corresponding DRO version can also be cast as an MILP, which can be solved efficiently by branch-and-cut algorithms implemented in solvers like CPLEX and GUROBI. Second, both second-order and high-order statistical information within data could be well integrated into the optimization model. Then, we further extend single-stage DRO models by considering uncertainties revealed in a sequential manner and some adjustable "wait-and-see" decisions to be implemented after uncertainties are revealed, which is in a similar spirit to the adaptive RO (ARO) (Ben-Tal et al., 2004; Ning and You, 2018). The affine decision rule (ADR) approximation is adopted to provide a tractable, albeit conservative solution towards multi-stage DRO problems. To demonstrate the efficacy of the proposed DRO framework, two representative applications in process network planning and multi-purpose batch process scheduling are presented, with each of them posed in a two-stage setting and a multi-stage setting, respectively.

The remainder of this article is organized as follows. Section 2 revisits the background of DRO, including the concept of ambiguity sets and tractable reformulations. In Section 3, a systematic data-driven approach to parameter tuning of ambiguity sets is proposed to capture high-order moment information from data. In Section 4, two-stage and multi-stage DRO problems are developed, and the solution issue is addressed by employing ADRs. Sections 5 and 6 are devoted to two industrial-scale decision-making problems for process network design and batch process scheduling, respectively. Finally the conclusion is drawn.

2. Basics of distributionally robust optimization

In this section, we first introduce the concept of ambiguity set and revisit some formulations widely adopted in literature. After that, a tractable deterministic reformulation of the single-stage problem is derived. All materials in this section come from existing literature (Bertsimas et al., 2010; 2017b; Delage and Ye, 2010; Ghaoui et al., 2003; Shapiro, 2001; Wiesemann et al., 2014; Yue et al., 2006; Zymler et al., 2013), which will be useful for further discussions.

2.1. The concept of ambiguity set

In the context of SP, one often seeks a risk-neutral decision by minimizing the expected value of a random function $L(\mathbf{x}, \boldsymbol{\xi})$:

$$\min_{\mathbf{x}} \mathbb{E}_{\mathbb{P}}\left\{l(\mathbf{x}, \boldsymbol{\xi})\right\}. \tag{1}$$

where $\boldsymbol{\xi} = [\xi_1 \ \cdots \ \xi_M]^{\mathrm{T}}$ denotes an M-dimensional random vector representing uncertain parameters, and $\mathbf{x}$ denotes decision variables. However, in the absence of exact knowledge about the distribution $\mathbb{P}$ of uncertainties $\boldsymbol{\xi}$, solving (1) becomes intractable. Instead of considering an exact distribution $\mathbb{P}$, the objective function of DRO optimizes the worst-case expectation of $l(\mathbf{x}, \boldsymbol{\xi})$ under all possible distributions $\mathbb{P}$ in the ambiguity set $\mathcal{D}$.

$$\min_{\mathbf{x}} \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{l(\mathbf{x}, \boldsymbol{\xi})\right\}. \tag{2}$$

The ambiguity set includes a family of probability distributions, which share common statistical properties that can be conveniently estimated from historical data. A commonly used formulation of the ambiguity set $\mathcal{D}$ is to consider all distributions with the exact mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ (Ghaoui et al., 2003; Yue et al., 2006):

$$\mathcal{D} = \left\{ \mathbb{P}_{\xi} \in \mathcal{M}_+^M \left| \begin{array}{l} \mathbb{E}_{\mathbb{P}}\{\boldsymbol{\xi}\} = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\{(\boldsymbol{\xi}-\boldsymbol{\mu})(\boldsymbol{\xi}-\boldsymbol{\mu})^{\mathrm{T}}\} = \boldsymbol{\Sigma} \end{array} \right. \right\}, \tag{3}$$

where $\mathcal{M}_+^M$ stands for the set of all valid probability distributions on $\mathbb{R}^M$. In practice, the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ shall be estimated from data. To accommodate estimation errors of mean vector and covariance matrix, Delage and Ye (2010) put forward a generalized ambiguity set as follows:

$$\mathcal{D} = \left\{ \mathbb{P}_{\xi} \in \mathcal{M}_+^M \left| \begin{array}{l} \mathbb{P}\{\boldsymbol{\xi} \in \Xi\} = 1 \\ (\mathbb{E}_{\mathbb{P}}\{\boldsymbol{\xi}\} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbb{E}_{\mathbb{P}}\{\boldsymbol{\xi}\} - \boldsymbol{\mu}) \le \gamma_1 \\ \mathbb{E}_{\mathbb{P}}\{(\boldsymbol{\xi}-\boldsymbol{\mu})(\boldsymbol{\xi}-\boldsymbol{\mu})^{\mathrm{T}}\} \preceq \gamma_2 \boldsymbol{\Sigma} \end{array} \right. \right\}, \tag{4}$$

where $\mathbb{P}\{\boldsymbol{\xi} \in \Xi\} = 1$ enforces the realization of uncertainties to be constrained within a given set Ξ, termed as the "support set". Parameters $\gamma_1 \ge 0$ and $\gamma_2 \ge 1$ are responsible for controlling the "volume" of the ambiguity set. In this way, solving the problem (2) can help protecting against the inexactness of distributions, which is measured by differences in second-order moment information. The reformulation of DRO problem (2) based on the ambiguity set (3) or (4) typically involves additional linear matrix inequalities (LMIs) (Bertsimas et al., 2010; Delage and Ye, 2010; Zymler et al., 2013), which are essentially convex optimization problems. Nonetheless, the presence of LMIs prohibits applications to large-scale process operations, in which the resulting deterministic problems are cast in the form of general mixed-integer nonlinear programs (MINLP).

Wiesemann et al. (2014) put forward the following class of ambiguity sets as a general formulation:

$$\mathcal{D} = \left\{ \mathbb{P}_{\xi} \in \mathcal{M}_{+}^{M} \middle| \begin{array}{l} \mathbb{P}\{\boldsymbol{\xi} \in \Xi\} = 1 \\ \mathbb{E}_{\mathbb{P}}\{g_i(\boldsymbol{\xi})\} \le \gamma_i, \ i = 1, \cdots, I \end{array} \right\}. \tag{5}$$

The first constraint in (5) ensures that $\mathcal{D}$ only contains valid distributions supported over the support set Ξ. The second constraint in (5) characterizes moment information of uncertainties via I functions $\{g_i(\cdot)\}$, and enforces the generalized moment $\mathbb{E}_{\mathbb{P}}\{g_i(\boldsymbol{\xi})\}$ cannot exceed a given threshold γ_i. In fact, the ambiguity set $\mathcal{D}$ can be reexpressed as the projection of an extended ambiguity set $\bar{\mathcal{D}}$ by introducing an I-dimensional auxiliary random vector $\boldsymbol{\varphi}$:

$$\bar{\mathcal{D}} = \left\{ \mathbb{Q}_{\xi,\varphi} \in \mathcal{M} \middle| \begin{array}{l} \mathbb{P}\{(\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi}\} = 1 \\ \mathbb{E}_{\mathbb{Q}}\{\boldsymbol{\varphi}\} \le \boldsymbol{\gamma} \end{array} \right\}, \tag{6}$$

where the domain of uncertainties is extended to a lifted support set $\bar{\Xi}$:

$$\bar{\Xi} = \left\{ (\boldsymbol{\xi}, \boldsymbol{\varphi}) \middle| \begin{array}{l} \boldsymbol{\xi} \in \Xi \\ g_i(\boldsymbol{\xi}) \le \varphi_i, \ i = 1, \cdots, I \end{array} \right\}. \tag{7}$$

It has been proved by Bertsimas et al. (2017b) that the ambiguity set $\mathcal{D}$ is essentially tantamount to the set including all marginal distribution of $\boldsymbol{\xi}$ under $\mathbb{Q} \in \bar{\mathcal{D}}$. To make this article self-contained, a formal proof will be provided in the Appendix A. It indicates that the DRO problem (2) is essentially identical to:

$$\min_{\mathbf{x}} \sup_{\mathbb{Q} \in \bar{\mathcal{D}}} \mathbb{E}_{\mathbb{Q}}\left\{ l(\mathbf{x}, \boldsymbol{\xi}) \right\}. \tag{8}$$

Note that the expectation constraints in $\bar{\mathcal{D}}$ only contain linear functions of uncertainties ($\boldsymbol{\xi}$, $\boldsymbol{\varphi}$), which are helpful for deriving tractable reformulations. As to be clarified in the sequel, the lifting technique is particularly helpful for enhancing the expressiveness of decision rules in a multi-stage setting.

2.2. Reformulation of the worst-case expectation problem

Next we deal with the tractability issue of the DRO problem (8). The inner maximization problem $\sup_{\mathbb{Q} \in \bar{\mathcal{D}}} \mathbb{E}_{\mathbb{Q}}\left\{ l(\mathbf{x}, \boldsymbol{\xi}) \right\}$ in (8) adds significant difficulties in computations because a probability measure $\mathbb{Q}$ typically has infinite dimensions. Fortunately, we could dualize the inner problem and transform it into a minimization problem. The methodological details are described as follows. An explicit expression of the inner problem is given by:

$$\begin{aligned} \sup_{\mathbb{Q}} \quad & \int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) l(\mathbf{x}, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} \\ \text{s.t.} \quad & \int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} = 1 \\ & \int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) \boldsymbol{\varphi} \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} \le \boldsymbol{\gamma} \end{aligned} \tag{9}$$

Here the decision variable is the joint probability density function $p(\boldsymbol{\xi}, \boldsymbol{\varphi})$ or the probability measure $\mathbb{Q}$. By associating Lagrangian multipliers η and $\boldsymbol{\beta}$ with the last two constraints in (9), the Lagrangian is given by

$$\begin{aligned} & L(\mathbb{Q}, \eta, \boldsymbol{\beta}) \qquad (10) \\ & = \int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) l(\mathbf{x}, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} - \eta \left(\int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} - 1 \right) \\ & \quad - \boldsymbol{\beta}^{\mathrm{T}} \left(\int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) \boldsymbol{\varphi} \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} - \boldsymbol{\gamma} \right) \\ & = \int_{\bar{\Xi}} p(\boldsymbol{\xi}, \boldsymbol{\varphi}) \left(l(\mathbf{x}, \boldsymbol{\xi}) - \eta - \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\varphi} \right) \mathrm{d}\boldsymbol{\xi} \mathrm{d}\boldsymbol{\varphi} + \eta + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\gamma} \qquad (10) \end{aligned}$$

The Lagrange dual function is given by

$$g(\eta, \boldsymbol{\beta}) = \sup_{\mathbb{Q}} L(\mathbb{Q}, \eta, \boldsymbol{\beta}). \tag{11}$$

If there exists $(\boldsymbol{\xi}^*, \boldsymbol{\varphi}^*) \in \bar{\Xi}$ such that $l(\mathbf{x}, \boldsymbol{\xi}^*) > \eta + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\varphi}^*$, then the dual function $g(\eta, \boldsymbol{\beta})$ will be unbounded above because $p(\boldsymbol{\xi}, \boldsymbol{\varphi}) \ge 0, \ \forall (\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi}$. Thus, $g(\eta, \boldsymbol{\beta}) = +\infty$ except when $\eta + \boldsymbol{\varphi}^{\mathrm{T}} \boldsymbol{\beta} \ge l(\mathbf{x}, \boldsymbol{\xi}), \ \forall (\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi}$. Then the dual function can be analytically expressed as

$$g(\eta, \boldsymbol{\beta}) = \begin{cases} \eta + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\beta}, & \text{if } \eta + \boldsymbol{\varphi}^{\mathrm{T}} \boldsymbol{\beta} \ge l(\mathbf{x}, \boldsymbol{\xi}), \ \forall (\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi} \\ +\infty, & \text{otherwise} \end{cases} \tag{12}$$

Finally, we can form a dual problem of (9):

$$\begin{aligned} \min_{\eta, \boldsymbol{\beta}} \quad & g(\eta, \boldsymbol{\beta}) \\ \text{s.t.} \quad & \boldsymbol{\beta} \ge 0 \end{aligned} \tag{13}$$

which is equivalent to:

$$\begin{aligned} \min_{\eta, \boldsymbol{\beta}} \quad & \eta + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\beta} \\ \text{s.t.} \quad & \boldsymbol{\beta} \ge 0 \\ & \eta + \boldsymbol{\varphi}^{\mathrm{T}} \boldsymbol{\beta} \ge l(\mathbf{x}, \boldsymbol{\xi}), \ \forall (\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi} \end{aligned} \tag{14}$$

In the light of the conic duality (Shapiro, 2001), strong duality holds and hence the optimal value of the dual problem is identical to that of the primal problem (9). Therefore, by further considering $\mathbf{x}$, problem (2) can be recast as the following finite-dimensional minimization problem:

$$\begin{aligned} \min_{\mathbf{x}, \eta, \boldsymbol{\beta}} \quad & \eta + \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{\beta} \\ \text{s.t.} \quad & \boldsymbol{\beta} \ge 0 \\ & \eta + \boldsymbol{\varphi}^{\mathrm{T}} \boldsymbol{\beta} \ge l(\mathbf{x}, \boldsymbol{\xi}), \ \forall (\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi} \end{aligned} \tag{15}$$

It can be seen that the last constraint in (15) is essentially a robust constraint with uncertainty set $\bar{\Xi}$. Hence the problem above can be indeed regarded as a classic robust optimization problem.

3. A data-based ambiguity set based on first-order deviation moment functions

3.1. Ambiguity set formulation

In this article, we restrict our attentions to the general formulation (5) for ambiguity sets, which has sufficient flexibility in choosing the support Ξ and generalized moment functions $\{g_i(\cdot)\}$. More specifically, we adopt upper and lower bounds in each dimension to specify the support Ξ of uncertainties:

$$\Xi = \left\{ \boldsymbol{\xi} \middle| \xi_m^{\min} \le \xi_m \le \xi_m^{\max}, \ m = 1, \cdots, M \right\}, \tag{16}$$

which is similar to the box uncertainty set adopted in RO (Soyster, 1973). As for the moment functions $\{g_i(\cdot)\}$, we adopt the following piecewise linear formulation:

$$g_i(\boldsymbol{\xi}) = \max\left\{ \mathbf{f}_i^{\mathrm{T}} \boldsymbol{\xi} - q_i, 0 \right\}, \ i = 1, \cdots, I \tag{17}$$

which can be understood as the first-order deviation of uncertain parameters along a certain projection direction $\mathbf{f}_i$ truncated at q_i. It is a special case of convex piecewise linear moment functions suggested by Bertsimas et al. (2017b), in which a variety of functions have been recommended as generalized moment functions. We point out in this paper that the choice (17) has some exclusive advantages. It can encode first-order deviation information along different directions without assuming any prior knowledge about probability distributions; most importantly, equivalent robust counterpart problems can be induced, which are computationally tractable, thereby benefiting decision-making in complicated process operation problems. In addition, we are able to derive a

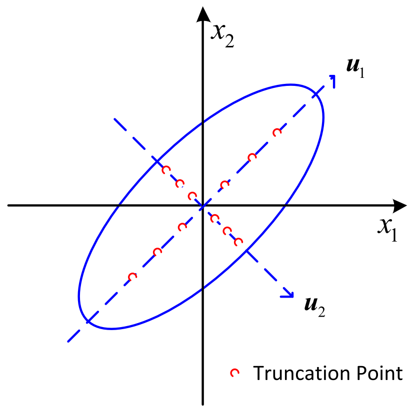


Fig. 2. A two-dimensional illustration of the proposed two-step approach to data-driven construction of ambiguity sets. In the first step, projection directions are determined by PCA, shown as blue dash lines. Then in the second step, a series of truncation points (red circles) are assigned in each projection direction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

systematic data-driven approach to determine $\{\mathbf{f}_i\}$ and $\{q_i\}$. These two aspects will be clarified in the sequel.

With the support set Ξ and functions $\{g_i(\cdot)\}$ determined by (16) and (17), we can easily rewrite the lifted support set $\bar{\Xi}$ as a set of linear inequalities:

$$\bar{\Xi} = \left\{ (\boldsymbol{\xi}, \boldsymbol{\varphi}) \left| \begin{array}{l} \boldsymbol{\xi} \le \boldsymbol{\xi}^{\max} \\ \boldsymbol{\xi}^{\min} \le \boldsymbol{\xi} \\ 0 \le \varphi_i, \ i = 1, \cdots, I \\ \mathbf{f}_i^{\mathrm{T}} \boldsymbol{\xi} - q_i \le \varphi_i, \ i = 1, \cdots, I \end{array} \right. \right\}, \tag{18}$$

which can be further concisely expressed in a matrix form:

$$\bar{\Xi} = \left\{ (\boldsymbol{\xi}, \boldsymbol{\varphi}) \middle| \mathbf{C}\boldsymbol{\xi} + \mathbf{D}\boldsymbol{\varphi} \le \mathbf{r} \right\}. \tag{19}$$

Here matrices **C, D** and the vector **r** are given by:

$$\mathbf{C} = \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \\ \mathbf{0} \\ \mathbf{F}^{\mathrm{T}} \end{bmatrix}, \ \mathbf{D} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{I} \\ -\mathbf{I} \end{bmatrix}, \ \mathbf{r} = \begin{bmatrix} \boldsymbol{\xi}^{\max} \\ -\boldsymbol{\xi}^{\min} \\ \mathbf{0} \\ \mathbf{q} \end{bmatrix}, \tag{20}$$

where **I** denotes unitary matrix with appropriate dimension, and **F** and **q** encompass parameters of piecewise linear functions $\{g_i(\cdot)\}$:

$$\mathbf{F} = [\mathbf{f}_1 \ \mathbf{f}_2 \ \cdots \mathbf{f}_I], \ \mathbf{q} = [q_1 \ q_2 \ \cdots \ q_I]^{\mathrm{T}}. \tag{21}$$

3.2. A data-based strategy for parameter determination

As a major contribution of this work, we propose in this subsection a systematic two-step procedure for determining **F** and **q** in order to capture meaningful information from available data.

A basic property of an ambiguity set $\mathcal{D}$ is that some correlation information should be incorporated. This is generally achieved by choosing quadratic functions $\{g_i(\cdot)\}$ (Bertsimas et al., 2017b; Wiesemann et al., 2014). As for the ambiguity set introduced previously, out basic idea is to first determine the directions $\{\mathbf{f}_i\}$ based on PCA such that data space becomes "decorrelated" along each direction and the information overlap between different directions is slight. After that, several truncation points $\{q_i\}$ are set along each direction $\mathbf{f}_i$. Fig. 2 depicts a two-dimensional illustration of this two-step strategy. According to the physical interpretation of PCA, the data space has the most variations along the first principle direction $\mathbf{f}_1$, and has the second most variations along the second principle direction $\mathbf{f}_2$, etc., with each direction being orthogonal. In this way, $\{\mathbf{f}_i\}$ bear clear physical meanings, and correlation information can be seamlessly encompassed in $\{\mathbf{f}_i\}$, based on which more subtle statistical information along each individual direction can be captured on its own.

Next we elaborate on this two-step procedure for ambiguity set constructions. Assuming that we are given N data samples $\{\boldsymbol{\xi}^{(n)}, \ n = 1, \cdots, N\}$, unbiased estimates of the mean vector and the covariance matrix are given by:

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{\xi}^{(n)}, \tag{22}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \left[\sum_{n=1}^{N} \boldsymbol{\xi}^{(n)} \left(\boldsymbol{\xi}^{(n)} \right)^{\mathrm{T}} - \left(\sum_{n=1}^{N} \boldsymbol{\xi}^{(n)} \right) \left(\sum_{j=1}^{N} \boldsymbol{\xi}^{(n)} \right)^{\mathrm{T}} \right]. \tag{23}$$

Then performing PCA onto $\hat{\boldsymbol{\Sigma}}$ yields (Wold et al., 1987):

$$\hat{\boldsymbol{\Sigma}} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^{\mathrm{T}}, \tag{24}$$

where **U** stands for the orthogonal transformation matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix including the variance information after transformations. Therefore, projection onto the mth principal direction can be expressed as:

$$z_m = \mathbf{u}_m^{\mathrm{T}} \boldsymbol{\xi}, \ m = 1, \cdots, M, \tag{25}$$

where $\mathbf{u}_m$ is the mth column of **U**. In this way, different projection $\{z_m\}$ carry different statistical information. Next, for each direction $\mathbf{u}_m$ we choose $2K+1$ well-distributed truncation points. To be more specific, our strategy is to set the first truncation point as the mean value $\hat{\mu}_m$, and then specify the remaining $2K$ ones around the mean μ_m symmetrically based on a fixed step-size, which can be determined according to statistical information. For example, a reasonable way is to set the step-size as the variance λ_n along the mth direction, which well integrates the scaling information along different projection directions. The variance λ_m is in fact the mth diagonal element of $\boldsymbol{\Lambda}$ in (24). Mathematically, $2K+1$ truncation points along the mth principal direction are expressed as:

$$\hat{\mu}_m, \ \hat{\mu}_m - \lambda_m, \ \hat{\mu}_m + \lambda_m, \ \cdots, \ \hat{\mu}_m - K\lambda_m, \ \hat{\mu}_m + K\lambda_m. \tag{26}$$

In this way, we will have $M(2K+1)$ piecewise functions $\{g_i(\cdot)\}$ in total in the ambiguity set (5). Accordingly, the matrix **F** and the vector **q** in (20) can be obtained by stacking different projection directions $\{\mathbf{u}_m\}$ and the associated truncation points:

$$\mathbf{F} = \mathbf{1}_{2K+1}^{\mathrm{T}} \otimes \mathbf{U} \in \mathbb{R}^{M \times M(2K+1)}, \tag{27}$$

$$\mathbf{q} = \begin{bmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\boldsymbol{\mu}} - \boldsymbol{\lambda} \\ \hat{\boldsymbol{\mu}} + \boldsymbol{\lambda} \\ \vdots \\ \hat{\boldsymbol{\mu}} - K\boldsymbol{\lambda} \\ \hat{\boldsymbol{\mu}} + K\boldsymbol{\lambda} \end{bmatrix} \in \mathbb{R}^{M(2K+1)}, \tag{28}$$

where $\otimes$ stands for the Kronecker product, $\mathbf{1}_{2K+1}$ is a $(2K+1)$ -dimensional vector with all elements being one, and $\boldsymbol{\lambda} = [\lambda_1 \ \cdots \ \lambda_M]^{\mathrm{T}}$. Notice that the transformation matrix **U** emerges $2K+1$ times in the expression of matrix **F** since for each direction multiple truncation points are adopted.

Intuitively, the parameter K can be deemed as the "size" of the ambiguity set $\mathcal{D}$, which can be manipulated to adjust the conservatism of the model. The more truncation points we have, the more statistical information will be incorporated, which leads to a smaller ambiguity set as well as a less conservative solution. Indeed, truncation points can be determined in different ways. For example, one could opt for a small step-size and more truncation points to better capture the statistical information.

After determining the value of $\{\mathbf{f}_i\}$ and $\{q_i\}$, a next step is to estimate the size parameters $\boldsymbol{\gamma}$ in (6) empirically based on N available data samples:

$$\hat{\gamma}_i = \frac{1}{N}\sum_{j=1}^{N} \max\left\{\mathbf{f}_i^{\mathrm{T}}\boldsymbol{\xi}^{(j)} - q_i, 0\right\}. \tag{29}$$

Intuitively, with the values of size parameters $\{\gamma_i\}$ increasing, the DRO model becomes more conservative. For the ambiguity set $\mathcal{D}$ in (4), Delage and Ye (2010) developed an approach for determining γ_1 and γ_2 such that $\mathcal{D}$ contains the true distribution with high probability based on some distributional assumptions. Recent works on data-driven robust optimization (Bertsimas et al., 2017a; Ning and You, 2017a; Shang et al., 2017) proposed to employ hypothesis tests to determine the size parameters so that the ambiguity set becomes statistically interpretable. Unfortunately, these strategies cannot be applied trivially to the ambiguity set used in this work. The proposed principle (29) could guarantee that $\mathcal{D}$ must include the empirical probability density $\mathbb{P}\{\boldsymbol{\xi} = \boldsymbol{\xi}^{(j)}\} = 1/N,\ j = 1, \cdots, N.$ Indeed, one could further specify $\{\gamma_i\}$ in (5) in different ways to adjust the conservatism of the ambiguity set, which is worth investigations in future.

4. Problem setup of multi-stage distributionally robust optimization

4.1. Two-stage DRO problems

Based on the single-stage DRO problem, we further consider the following two-stage DRO problem:

$$\begin{array}{ll} \min_{\mathbf{x}} & \mathbf{c}^{\mathrm{T}}\mathbf{x} + \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{l(\mathbf{x}, \boldsymbol{\xi})\right\} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \le \mathbf{b} \end{array} \tag{30}$$

where $\mathbf{x}$ denotes the first-stage decision variables, which may include both binary variables and continuous variables. We consider a linear programming formulation of the second-stage subproblem, which can be compactly expressed in the matrix form:

$$l(\mathbf{x}, \boldsymbol{\xi}) = \begin{cases} \min_{\mathbf{y}} & \mathbf{d}^{\mathrm{T}}\mathbf{y} \\ \text{s.t.} & \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{y} \le \mathbf{h}(\boldsymbol{\xi}) \end{cases} \tag{31}$$

where continuous variables $\mathbf{y}$ denotes the second-stage decisions (resource) to be determined after uncertainty $\boldsymbol{\xi}$ is uncovered. The constraint in (31) indicates that both the coefficient matrix $\mathbf{T}$ and the right-hand-side vector $\mathbf{h}$ are affinely influenced by uncertainties, expressed as follows:

$$\mathbf{T}_j(\boldsymbol{\xi}) = \mathbf{T}_j^0 + \sum_{m=1}^{M} \xi_m \cdot \mathbf{T}_{jm}^{\xi},\ \ j = 1, \cdots, J \tag{32}$$

$$\mathbf{h}(\boldsymbol{\xi}) = \mathbf{h}^0 + \sum_{m=1}^{M} \xi_m \cdot \mathbf{h}_m^{\xi} \tag{33}$$

where $\mathbf{T}_j^{\mathrm{T}}(\boldsymbol{\xi})$ denotes the jth column of the matrix $\mathbf{T}(\boldsymbol{\xi})$. $\{\mathbf{T}_j^0\}$ and $\mathbf{h}^0$ denote constant terms that are free from uncertainties, while $\{\mathbf{T}_{jm}^{\xi}\}$ and $\mathbf{h}_m^{\xi}$ denote coefficients of the affine dependence on the mth uncertain parameter ξ_m.

4.2. Reformulation based on affine decision rules

In a two-stage setting, deriving an explicit expression of optimal recourse policy and calculating the worst-case expectation are generally intractable, since they involve enumeration of all realizations of uncertainties within the lifted support set $\bar{\Xi}$ (Bertsimas et al., 2010; Goh and Sim, 2010). A pragmatic strategy to circumvent the intractability issue is to employ the ADR (Ben-Tal et al., 2004; Chen et al., 2008), which enforces the recourse decision $\mathbf{y}$ to be affinely varying with the uncertainties:

$$y_k(\boldsymbol{\xi}, \boldsymbol{\varphi}) = y_k^0 + \sum_m y_{km}^{\xi}\xi_m + \sum_i y_{ki}^{\varphi}\varphi_i. \tag{34}$$

Recently, ADR has been adopted in solving multi-stage optimization problems in the RO setting in process systems engineering (Ning and You, 2017b; Zhang et al., 2016). Here, the affine dependence is expressed for each recourse variable y_k. Rewriting ADR for all recourse variables leads to the following matrix form:

$$\mathbf{y}(\boldsymbol{\xi}, \boldsymbol{\varphi}) = \mathbf{y}^0 + \mathbf{Y}^{\xi}\boldsymbol{\xi} + \mathbf{Y}^{\varphi}\boldsymbol{\varphi}. \tag{35}$$

where $\mathbf{y}^0$ denotes the constant, $\mathbf{Y}^{\xi}$ and $\mathbf{Y}^{\varphi}$ are coefficient matrices associated with the random variables $\boldsymbol{\xi}$ and the auxiliary random variables $\boldsymbol{\varphi}$, respectively. By substituting the ADR approximation for the optimal decision policy, a conservative approximation to the two-stage DRO problem (30) can be derived:

$$\begin{array}{ll} \min\limits_{\mathbf{x}} \min\limits_{\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi})} & \mathbf{c}^{\mathrm{T}}\mathbf{x} + \sup\limits_{\mathbb{Q}\in\bar{\mathcal{D}}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi}, \boldsymbol{\varphi})\right\} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{y}(\boldsymbol{\xi}, \boldsymbol{\varphi}) \le \mathbf{h}(\boldsymbol{\xi}),\ \forall(\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi} \end{array} \tag{36}$$

Here, coefficients $\{\mathbf{y}^0, \mathbf{Y}^{\xi}, \mathbf{Y}^{\varphi}\}$ of the decision rule (35) are absorbed into first-stage variables. Based on the reformulation in Section 2.2, we could arrive at the following identical RO problem by dualizing the inner-most maximization problem:

$$\min\ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\beta} \tag{37}$$

$$\text{s.t.}\ \ \boldsymbol{\beta} \ge \mathbf{0} \tag{38}$$

$$\eta + \boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\beta} \ge \mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi}, \boldsymbol{\varphi}),\ \forall(\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi} \tag{39}$$

$$\mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{y}(\boldsymbol{\xi}, \boldsymbol{\varphi}) \le \mathbf{h}(\boldsymbol{\xi}),\ \forall(\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi} \tag{40}$$

$$\mathbf{A}\mathbf{x} \le \mathbf{b} \tag{41}$$

Note that constraints (39) and (40) can be regarded as two generic robust constraints on the polytopic uncertainty set $\bar{\Xi}$. We can then adopt existing results from RO literature to convert these two infinite-dimensional constraints into their corresponding robust counterparts, leading to the following deterministic optimization problem:

$$\min\ \mathbf{c}^{\mathrm{T}}\mathbf{x} + \eta + \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\beta} \tag{42}$$

$$\text{s.t.}\ \ \boldsymbol{\beta} \ge \mathbf{0} \tag{43}$$

$$\eta - \mathbf{d}^{\mathrm{T}}\mathbf{y}^0 - \boldsymbol{\pi}_0^{\mathrm{T}}\mathbf{r} \ge 0 \tag{44}$$

$$\boldsymbol{\pi}_0^{\mathrm{T}}\mathbf{C}_s = \sum_n d_n y_{ns}^{\xi},\ \ \forall s \in \mathcal{S} \tag{45}$$

$$\boldsymbol{\pi}_0^{\mathrm{T}}\mathbf{D}_i = \sum_n d_n y_{ni}^{\varphi} - \beta_i,\ \ \forall i \in \mathcal{I} \tag{46}$$

$$\left(\mathbf{T}_m^0\right)^{\mathrm{T}}\mathbf{x} + \mathbf{W}_m^{\mathrm{T}}\mathbf{y}^0 - h_m^0 + \mathbf{r}^{\mathrm{T}}\boldsymbol{\pi}_m \le 0, \tag{47}$$

$$\boldsymbol{\pi}_m^{\mathrm{T}}\mathbf{C}_s = \left(\mathbf{T}_{ms}^{\xi}\right)^{\mathrm{T}}\mathbf{x} + \sum_n W_{mn}y_{ns}^{\xi} - h_{ms}^{\xi} \tag{48}$$

$$\boldsymbol{\pi}_m^{\mathrm{T}}\mathbf{D}_i = \sum_n W_{mn}y_{ni}^{\varphi} \tag{49}$$

$$\mathbf{A}\mathbf{x} \le \mathbf{b} \tag{50}$$

$$\boldsymbol{\pi}_0 \ge 0,\ \ \boldsymbol{\pi}_m \ge 0 \tag{51}$$

where $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_m$ are dual variables associated with robust constraints (39) and (40). The scale of the above problem depends on the number of truncation points K, and remains unchanged when the amount of historical data change.

It is worth noting that in Bertsimas et al. (2017b) quadratic functions $g_i(\boldsymbol{\xi}) = (\mathbf{f}_i^{\mathrm{T}}\boldsymbol{\xi} - q_i)^2$ are suggested to address the cross-correlation between different parameters. However, due to the penetration of quadratic terms, it leads to a second-order cone (SOC) formulation of the lifted support $\bar{\Xi}$ as well as a mixed-integer SOC programming reformulation of the resulting robust counterpart problem, which is more computationally expensive to solve than the obtained MILP (43)–(51). In contrast, the proposed method not only incorporates cross-correlation information, but also preserves a more tractable MILP reformulation (43)–(51) of the robust counterpart problem when first-stage decisions $\mathbf{x}$ include binary variables. So it is convenient to solve using the state-of-the-art branch-and-cut methods implemented in solvers such as CPLEX. Such an advantage fundamentally owes to both ADR and the first-order deviation functions in the ambiguity set formulation.

Meanwhile, using quadratic functions only includes second-order information into the ambiguity set. In contrast, our proposed approach is able to extract high-order statistical information by using piecewise linear functions, and hence it can characterize more intricate properties of uncertainties such as non-Gaussianity.

4.3. Discussions on the suboptimality and feasibility of ADR

Bertsimas et al. (2017b) have pointed out that, the ADR in (35) enjoys improved flexibility in approximating the optimal recourse policy due to the penetration of auxiliary random vector $\boldsymbol{\varphi}$. Therefore, it can help reducing the conservatism induced by classic ADR. To see this, we first define

$$f_1 \triangleq \begin{cases} \min_{\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi})} \sup_{\mathbb{Q}\in\bar{\mathcal{D}}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi})\right\} \\ \text{s.t.} \quad \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi}) \le \mathbf{h}(\boldsymbol{\xi}), \ \ \forall(\boldsymbol{\xi},\boldsymbol{\varphi}) \in \bar{\Xi} \end{cases} \tag{52}$$

and

$$f_2 \triangleq \begin{cases} \min_{\mathbf{y}(\boldsymbol{\xi})} \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi})\right\} \\ \text{s.t.} \quad \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{y}(\boldsymbol{\xi}) \le \mathbf{h}(\boldsymbol{\xi}), \ \ \forall\boldsymbol{\xi} \in \Xi \end{cases} \tag{53}$$

as optimal values of the second-stage problem under the worst-case distribution, where $\mathbf{y}(\boldsymbol{\xi})$ in (53) stands for the classic ADR dependent only on the primitive uncertainty $\boldsymbol{\xi}$. Since any feasible solution to (53) is still feasible for (52) by setting $\mathbf{Y}^{\varphi}$ in (35) to be zero, we could easily obtain $f_1 \le f_2$. Hence, by lifting the ambiguity set and introducing auxiliary variables, the suboptimality of ADR can be reduced.

However, the implication of the dependence of ADR on auxiliary variables is still unclear. In fact, solving (52) essentially provides a *feasible* solution to the following problem based on generalized ADR:

$$f_0 \triangleq \begin{cases} \min_{\mathbf{y}(\boldsymbol{\xi},\mathbf{g}(\boldsymbol{\xi}))} \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\mathbf{g}(\boldsymbol{\xi}))\right\} \\ \text{s.t.} \quad \mathbf{T}(\boldsymbol{\xi})\mathbf{x} + \mathbf{W}\mathbf{y}(\boldsymbol{\xi},\mathbf{g}(\boldsymbol{\xi})) \le \mathbf{h}(\boldsymbol{\xi}), \ \ \forall\boldsymbol{\xi} \in \Xi \end{cases} \tag{54}$$

The corresponding generalized decision rule is given by:

$$\mathbf{y}(\boldsymbol{\xi},\mathbf{g}(\boldsymbol{\xi})) = \mathbf{y}^0 + \mathbf{Y}^{\xi}\boldsymbol{\xi} + \mathbf{Y}^{\varphi}\mathbf{g}(\boldsymbol{\xi}), \tag{55}$$

which has richer representability than the classic ADR $\mathbf{y}(\boldsymbol{\xi})$ in (53) because of the nonlinear functions $\{g_i(\cdot)\}$. However, it is challenging to attack (54) directly. Notice that

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\mathbf{g}(\boldsymbol{\xi}))\right\} = \sup_{\mathbb{Q}\in\tilde{\mathcal{D}}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi})\right\}. \tag{56}$$

where $\tilde{\mathcal{D}} = \left\{p(\boldsymbol{\xi},\boldsymbol{\varphi}) \middle| p(\boldsymbol{\xi},\boldsymbol{\varphi}) = p(\boldsymbol{\xi}) \cdot \delta_{\mathbf{g}(\boldsymbol{\xi})}(\boldsymbol{\varphi}), \ \ p(\boldsymbol{\xi}) \in \mathcal{D}\right\}$ is a new ambiguity set and hence $\tilde{\mathcal{D}} \subseteq \bar{\mathcal{D}}$ holds. This yields

$$\sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\mathbf{g}(\boldsymbol{\xi}))\right\} = \sup_{\mathbb{Q}\in\tilde{\mathcal{D}}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi})\right\} \le \sup_{\mathbb{Q}\in\bar{\mathcal{D}}} \mathbb{E}_{\mathbb{P}}\left\{\mathbf{d}^{\mathrm{T}}\mathbf{y}(\boldsymbol{\xi},\boldsymbol{\varphi})\right\}. \tag{57}$$

Therefore, $f_0 \le f_1$ holds since any feasible solution to (52) is still feasible for (54).

In summary, we have $f_0 \le f_1 \le f_2$. It implies that the formulation (52) adopted in this work essentially provides *a feasible yet conservative solution* for (54) with nonlinear functions incorporated in the decision rule, which is difficult to solve directly. Meanwhile, it is less conservative than the classic ADR that is dependent on primitive uncertainties only.

4.4. Extensions to multi-stage problems

In a multi-stage setting, uncertainties are unfolded over stages, and a series of recourse decisions are made in a sequential manner. This can be expressed mathematically using the following nested formulation:

$$\min_{\mathbf{x}} \left\{\mathbf{c}^{\mathrm{T}}\mathbf{x} + \sup_{\mathbb{P}_1\in\mathcal{D}_1} \mathbb{E}\left\{\min_{\mathbf{y}_1\in\Omega_1(\boldsymbol{\xi}_1)} \mathbf{d}_1^{\mathrm{T}}\mathbf{y}_1 + \sup_{\mathbb{P}_2\in\mathcal{D}_2} \mathbb{E}\left\{\cdots + \min_{\mathbf{y}_T\in\Omega_T(\boldsymbol{\xi}_T)} \mathbf{d}_T^{\mathrm{T}}\mathbf{y}_T\right\}\right\}\right\}. \tag{58}$$

Here $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_T$ are revealed over Stages $1, \cdots, T$, and $\mathbf{y}_1, \cdots, \mathbf{y}_T$ are recourse decisions made in Stages $1, \cdots, T$, respectively. More precisely, recourse variables $\mathbf{y}_t$ shall be made in response to uncertainties $\boldsymbol{\xi}_t$ observed at Stage t. $\Omega_t(\boldsymbol{\xi}_t)$ denotes the feasible region of recourse variables at Stage t, where we suppress the dependence of $\Omega_t(\boldsymbol{\xi}_t)$ on $\mathbf{x}, \mathbf{y}_1, \cdots, \mathbf{y}_{t-1}$ for notational convenience. The above problem setup is sufficiently general to represent process planning and scheduling problems encountered in process systems engineering, in which binary "here-and-now" decision variables are involved and continuous "wait-and-see" decisions are made after uncertainties are revealed.

We denote all recourse decisions and uncertainties as

$$\mathbf{y} = [\mathbf{y}_1^{\mathrm{T}} \ \mathbf{y}_2^{\mathrm{T}} \ \cdots \mathbf{y}_T^{\mathrm{T}}]^{\mathrm{T}}, \ \ \boldsymbol{\xi} = [\boldsymbol{\xi}_1^{\mathrm{T}} \ \boldsymbol{\xi}_2^{\mathrm{T}} \ \cdots \boldsymbol{\xi}_T^{\mathrm{T}}]^{\mathrm{T}}. \tag{59}$$

By applying the ADR with non-anticipativity constraints, we could still obtain a conservative solution to the multi-stage problem (58). The non-anticipativity constraints enforce the recourse variable $\mathbf{y}_t$ at Stage t to be dependent on the uncertainty realizations $\boldsymbol{\xi}_{1:t} = [\boldsymbol{\xi}_1^{\mathrm{T}} \ \cdots \ \boldsymbol{\xi}_t^{\mathrm{T}}]^{\mathrm{T}}$ up to Stage t only. With the auxiliary random variables $\boldsymbol{\varphi}$ such an affine correspondence can be described as follows (Ben-Tal et al., 2004):

$$\mathbf{y}_t(\boldsymbol{\xi}_{1:t}, \boldsymbol{\varphi}_{1:t}) = \mathbf{y}_t^0 + \mathbf{Y}_t^{\xi}\boldsymbol{\xi}_{1:t} + \mathbf{Y}_t^{\varphi}\boldsymbol{\varphi}_{1:t}, \ \ \forall t. \tag{60}$$

By stacking above equations from Stage 1 to T, we could determine the affine dependence relationship in (35), with some coefficients enforced to be zero, and further translate the multi-stage DRO problem into the tractable MILP (43)–(51).

4.5. An data-driven DRO framework

Up to now, we have arrived at a holistic multi-stage optimization framework based on DRO, which integrates information from historical data into final decisions. The entire flowchart is shown in Table 1.

Before closing this section, we provide a step-by-step illustration of this framework based on a simple example. Consider the following two-stage optimization problem:

$$\begin{array}{ll} \min_{\mathbf{x}} & 2x_1 + 3x_2 + \sup_{\mathbb{P}\in\mathcal{D}} \mathbb{E}_{\mathbb{P}}\left\{l(\mathbf{x},\boldsymbol{\xi})\right\} \\ \text{s.t.} & x_1 \ge 0, \ \ x_2 \in \{0,1\} \\ & x_1 + x_2 \le 3 \end{array} \tag{61}$$

Table 1
A data-driven optimization framework for multi-stage decision-making.

- *Step 1.* Collect historical data from a database.
- *Step 2.* Construct the ambiguity set $\mathcal{D}$ by applying the algorithm in Section 3.
- *Step 3.* Formulate the multi-stage DRO model based on ADR approximations and derive its tractable reformulation.
- *Step 4.* Solve the resulting optimization problem to determine the distributionally robust optimal decisions.

where the second-stage problem is given by:

$$l(\mathbf{x}, \boldsymbol{\xi}) = \begin{cases} \min_y & y \\ \text{s.t.} & 4x_1 + y \geq \xi_1 \\ & x_2 + y \geq \xi_2 \\ & y \geq 0 \end{cases} \tag{62}$$

Here there are two first-stage decision variables $\mathbf{x} \in \mathbb{R}^2$, and only one recourse decision variable $y \in \mathbb{R}$. In the second-stage problem, the r.h.s. of constraints are affected by uncertainties $\boldsymbol{\xi} = [\xi_1 \ \xi_2]^{\mathrm{T}}$.

In Step 1, we first collect a series of samples $\{\boldsymbol{\xi}^{(j)}\}_{j=1}^{N}$ of uncertain data. Then in Step 2, the data-driven procedure in Section 3 is carried out to determine the parameters of the ambiguity set $\mathcal{D}$ based on available data samples. To be more concrete, the mean vector $\hat{\boldsymbol{\mu}}$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}$ shall be first estimated from data according to (22) and (23). PCA is then performed onto the estimated covariance matrix $\hat{\boldsymbol{\Sigma}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}^{\mathrm{T}}$. Given the number of truncation points $2K+1$, the matrix $\mathbf{F}$ and the vector $\mathbf{q}$ shall be specified according to (27) and (28), which parameterize the ambiguity set. Notice that there are $2(2K+1)$ moment constraints in total, thereby resulting in a $2(2K+1)$ -dimensional auxiliary random vector $\boldsymbol{\varphi}$. Therefore, in Step 3, the ADR has the form of (35), where the coefficient $\mathbf{y}_0 \in \mathbb{R}$ is a scalar, $\mathbf{Y}^{\xi} \in \mathbb{R}^{1\times 2}$, $\mathbf{Y}^{\varphi} \in \mathbb{R}^{1\times 2(2K+1)}$. Then a tractable reformulation (43)–(51) could be easily derived by introducing extra dual decision variables $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_m$ that have appropriate dimensions. In Step 4, one only needs to implement this optimization problem with off-the-shelf solvers and obtain the final solution, which includes both here-and-now and wait-and-see decisions.

In a multi-stage setting, one could still proceed in a similar manner. The only difference lies in that additional non-anticipativity constraints shall be included in the reformulated optimization problem.

5. Application to process scheduling

In this section, the proposed DRO framework is applied to an industrial multi-purpose pharmaceutical batch production process in The Dow Chemical Company (Chu et al., 2013; Wassick et al., 2012), with its resource task network (RTN) portrayed in Fig. 3. It includes one mixer, two reactors, one finishing system and one drumming line to accomplish one preparation task, six reaction tasks, two packing tasks and two drumming tasks (Shi and You, 2016). Next, we develop a two-stage DRO-based process scheduling model to demonstrate the effectiveness of the proposed method.

5.1. Two-stage DRO-based process scheduling model

We employ the global event-based, continuous-time process scheduling model developed by Castro et al. (2004) in this application. The deterministic model formulation is given by:

$$\max \sum_s p_s \cdot S_{sN} - \sum_s q_s \cdot snp_s \cdot sint_s - \sum_i \sum_n \sum_{m\in\mathcal{N}_n^+} (fc_i \cdot W_{inm} + vc_i \cdot B_{inm}) \tag{63}$$

$$\text{s.t. } R_{rn} = R_{r(n-1)} + \sum_{i\in I_r} \left(\sum_{m\in\mathcal{N}_n^+} W_{inm} - \sum_{m\in\mathcal{N}_n^-} W_{imn} \right), \ \forall n \in \mathcal{N}, \ r \in \mathcal{R} \tag{64}$$

$$R_{rn} \leq 1, \ \forall n \in \mathcal{N}, \ r \in \mathcal{R} \tag{65}$$

$$b_i^{\min} \cdot W_{inm} \leq B_{inm} \leq b_i^{\max} \cdot W_{inm}, \ \forall n \in \mathcal{N}, \ m \in \mathcal{N}_n^+, \ i \tag{66}$$

$$S_{sn} = S_{s(n-1)} + \sum_{i\in TO_s} \rho_{is}^{O} \sum_{m\in\mathcal{N}_n^-} B_{imn} - \sum_{i\in TI_s} \rho_{is}^{I} \sum_{m\in\mathcal{N}_n^+} B_{inm}, \ \forall n \in \mathcal{N}, \ s \in \mathcal{S} \tag{67}$$

$$S_{s1} = snp_s \cdot sint_s - \sum_{i\in TI_s} \rho_{is}^{I} \sum_{m\in\mathcal{N}_n^+} B_{inm}, \ \forall s \in \mathcal{S} \tag{68}$$

$$S_{sn} \leq st_s^{\max}, \ \forall n \in \mathcal{N}, \ s \in \mathcal{S} \tag{69}$$

$$S_{sN} \geq \mathrm{dem}_s, \ \forall s \in \mathcal{S} \tag{70}$$

$$T_m - T_n \geq \sum_{i\in I_r} (\alpha_{in} \cdot W_{inm} + \beta_{in} \cdot B_{inm}), \ r \in \mathcal{R}, \ n \in \mathcal{N}, \ m \in \mathcal{N}_n^+ \tag{71}$$

$$W_{inm} \in \{0, 1\}, \ \forall i, n, m, \ W_{inm} = 0, \ \forall i, n, m \in \mathcal{N}\backslash\mathcal{N}_n^+ \tag{72}$$

$$R_{rn} \geq 0, \forall n, r, \ R_{rN} = 0, \ \forall r \in \mathcal{R} \tag{73}$$

$$B_{inm} = 0, \ \forall i, n, m \in \mathcal{N}\backslash\mathcal{N}_n^+ \tag{74}$$

$$T_1 = 0, \ T_N = H \tag{75}$$

$$B_{inm}, S_{sn}, T_n \geq 0, \ \forall i, n, m, s \tag{76}$$

The objective (63) is to maximize the revenue from selling products minus the cost of purchasing raw materials, fixed and variable running costs of equipments. Binary variables $\{W_{inm}\}$ are introduced to indicate whether task i starts at time point n and finishes before time point m. Continuous variables R_{rn} stand for whether the equipment r is operated or not at time point n, which is constrained by (65). The corresponding batch sizes are denoted by $\{B_{inm}\}$, which are constrained by (66). Eqs. (67)–(69) describe storage relationships and limits for each state s, while (70) makes the requirement for final demand satisfactions. Start/end times of tasks and periods are specified by (71) and (75). The entire deterministic problem admits an MILP formulation. Detailed parameter configurations are given in Appendix B (See Supplementary).

In this study, we consider random product demands $\boldsymbol{\xi} \triangleq \{\mathrm{dem}_s\}$, whose distribution is unknown but assumed to belong to an ambiguity set $\mathcal{D}$, while keeping other parameters as deterministic. The scheduling problem is cast in a two-stage setting. Binary variables

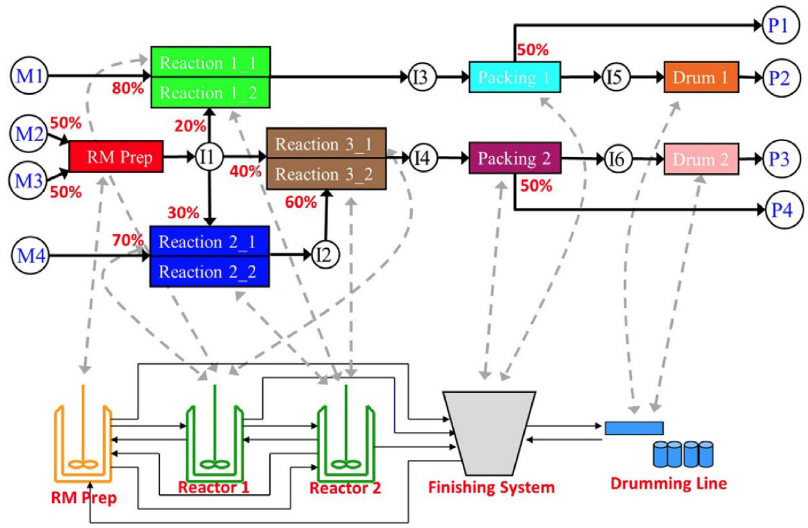


Fig. 3. Resource-task network (RTN) of the multi-purpose batch process.

Table 2
Optimization results in the batch process scheduling case.

	Deterministic Scheduling	Two-Stage DRO	Two-Stage ARO
Cont. var.	1315	87,028	20,066
Bin. var.	1100	1100	1100
Constraints	2772	55,642	30,163
CPU time (s)	2.50	1960.00	235.25
Objective ($)	2451	1513	400

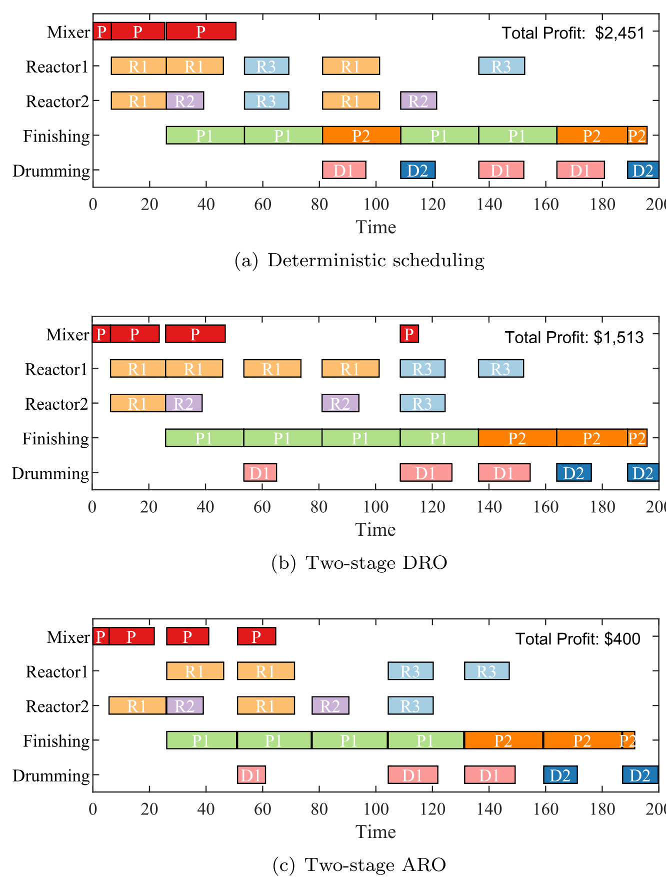


Fig. 4. Gantt charts of scheduling results based on different models under the nominal realization of demand.

$\{W_{inm}\}$, continuous variables $\{R_{rn}\}$ and initial purchase of materials $sint_s$ should be determined as "here-and-now" decisions $\mathbf{x}$ before random demands are known, while batch sizes $\{B_{inm}\}$, inventory levels $\{S_{sn}\}$, and start times of all periods $\{T_n\}$ are set as "wait-and-see" decisions $\mathbf{y}$ that are adjustable with random demands. Based on the deterministic formulation of the scheduling problem (63)–(76), we can immediately obtain its two-stage version in the form of (30). To be more specific, the second-stage problem is given by:

$$L(\mathbf{x}, \boldsymbol{\xi}) = \begin{cases} \min_{\mathbf{y}} & \sum_i \sum_n \sum_{m \in \mathcal{N}_n^+} vc_i \cdot B_{inm} - \sum_s p_s \cdot S_{sN} \\ \text{s.t.} & \text{Constraints } (66)-(71),\ (74)-(76) \end{cases} \tag{77}$$

The first-stage cost is given by:

$$\mathbf{c}^{\mathrm{T}}\mathbf{x} \triangleq \sum_s q_s \cdot snp_s \cdot sint_s + \sum_i \sum_n \sum_{m \in \mathcal{N}_n^+} fc_i \cdot W_{inm}, \tag{78}$$

and constraints for the first-stage variables correspond to (64), (65), (72) and (73).

We collect 500 historical samples of random demands, and then compute parameters of the ambiguity set $\mathcal{D}$ based on the data-driven approach proposed in Section 3. Here three truncation points ($K = 1$) are specified along each projection direction, and the step-size of truncation points is chosen as the variance λ_n along each projection direction. One could also use samples from a historical database to develop the ambiguity set, which enables an automatic schema of data-based decision making.

5.2. Results and discussions

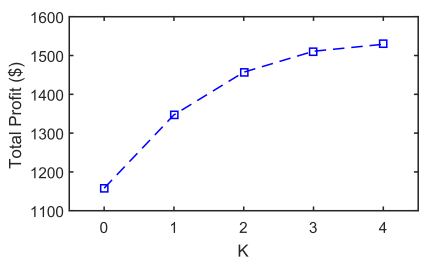
To make fair comparisons, we additionally model the scheduling problem with (i) the deterministic MILP model with nominal values and (ii) the two-stage ARO model. For the deterministic model, all parameters are set at their nominal values. The box uncertainty set is adopted in the ARO model with parameters estimated based on 500 data samples. All problems are programmed in GAMS 24.7.4, in which all induced MILPs are solved using CPLEX 12.7.0 with optimality gaps set as 0.1%. All computations are performed on a Windows 10 machine with Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz and 32GB memory.

Table 3
Expected profits ($) based on Monte Carlo simulation tests under different levels of uncertainties .

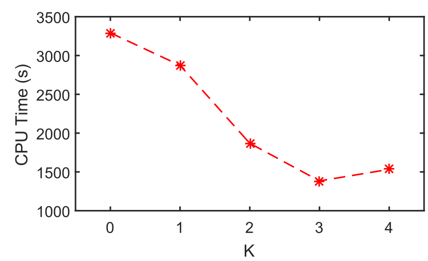
G	0.25		0.5		0.75		1	
Profit	DRO	ARO	DRO	ARO	DRO	ARO	DRO	ARO
Total	2416	2416	2222	2131	2100	1816	1959	1470
1st-Stage	− 10,241	− 10,240	− 10,434	− 10,386	− 10,534	− 10,388	− 10,633	− 10,247
2nd-Stage	12,657	12,656	12,656	− 12,517	12,634	12,204	12,592	11,717

Table 4
Purchase amounts (kg) of initial materials as first-stage decisions.

G	0.25		0.5		0.75		1	
	DRO	ARO	DRO	ARO	DRO	ARO	DRO	ARO
M1	608.00	608.00	608.00	601.60	608.00	588.64	608.00	571.18
M2	198.76	198.74	203.73	203.70	207.86	207.83	211.98	211.40
M3	198.76	198.74	203.73	203.70	207.86	207.83	211.98	211.40
M4	177.80	177.76	186.16	186.10	194.53	194.43	202.90	202.77



(a) Objective value



(b) Computational cost

Fig. 5. Optimization performances of two-stage DRO-based scheduling approach with different numbers of truncation points.

Table 5
Worst-case profits ($) based on Monte Carlo simulation tests under different levels of uncertainties.

G	0.25	0.5	0.75	1
DRO	2148	1586	951	250
ARO	2148	1601	1042	444

The computational results and problem sizes of three optimization problems are listed in Table 2. It can be seen that, the computational time of the deterministic optimization problem is much shorter than the others thanks to a much smaller problem size. The two-stage DRO problem takes about half an hour to solve to optimality, which is acceptable in practice. The relatively long solution time is due to a significant number of constraints and variables involved in the reformulated MILP problem. Meanwhile, the objective values are also computed and compared. It is shown that the deterministic model achieves the highest objective, and the total profit of ARO is much lower than that of DRO because of different criterion used in the second-stage.

We plot Gantt charts of scheduling decisions made by solving different optimization problems under nominal values of demands in Fig. 4, in which the X-axis and Y-axis denote the time horizon and equipment units, respectively. It can be observed that in comparison with the other two models, the DRO-based model schedules more tasks in the entire horizon (four times for Preparation and five times for Reaction 1). This provides sufficient flexibility in the face of random demands to adjust the recourse actions (batch sizes) for pursuing a high profit, because a schedule with more tasks can afford heavier manufacturing workload. In contrast, because of the impact of worst-case realization of demands, solving the two-stage RO gives a conservative scheduling result, in which the task of Reaction 1 is executed four times only. As with the deterministic approach, its sequencing decisions may become infeasible when randomness is imposed on demands, especially when demands higher than nominal values must be satisfied. Next we perform quantitative evaluations to shed further light on the conservatism of two robust scheduling methods. In a two-stage setting, explicit expressions of *optimal* policies for recourse variables are unavailable in both DRO and ARO. When solving two-stage DRO, recourse decisions are approximated by ADR, which is suboptimal. When solving two-stage ARO, only the recourse decisions under the worst-case realization are attained. Therefore, if we have already implemented first-stage decisions $\mathbf{x}$ (obtained by either DRO or ARO) in practice, recourse decisions $\mathbf{y}$ shall be optimally determined by solving the subproblem (77) after the uncertainties $\boldsymbol{\xi}$ are observed. It implies that we are able to quantify performances of different approaches based on an independent test dataset and evaluating the expected profits in an empirical way, giving rise to a reasonable and comprehensive comparison. Here we consider four cases with different magnitudes of uncertainties, which are obtained by multiplying the variations of random samples with a factor $G = 0.25, 0.5, 0.75, 1$, respectively. In each case, the following procedure is carried out individually.

Step 1. Collect 500 random samples, solve problems of two-stage DRO and ARO respectively, and fix first-stage decisions $\mathbf{x}$.

Step 2. Collect additional 1000 random samples of demands, which are sufficiently representative.

Step 3. For each sample, solve the corresponding deterministic second-stage problem (77) with first-stage decisions determined by DRO and ARO, derive the second-stage decisions and calculate total profits.

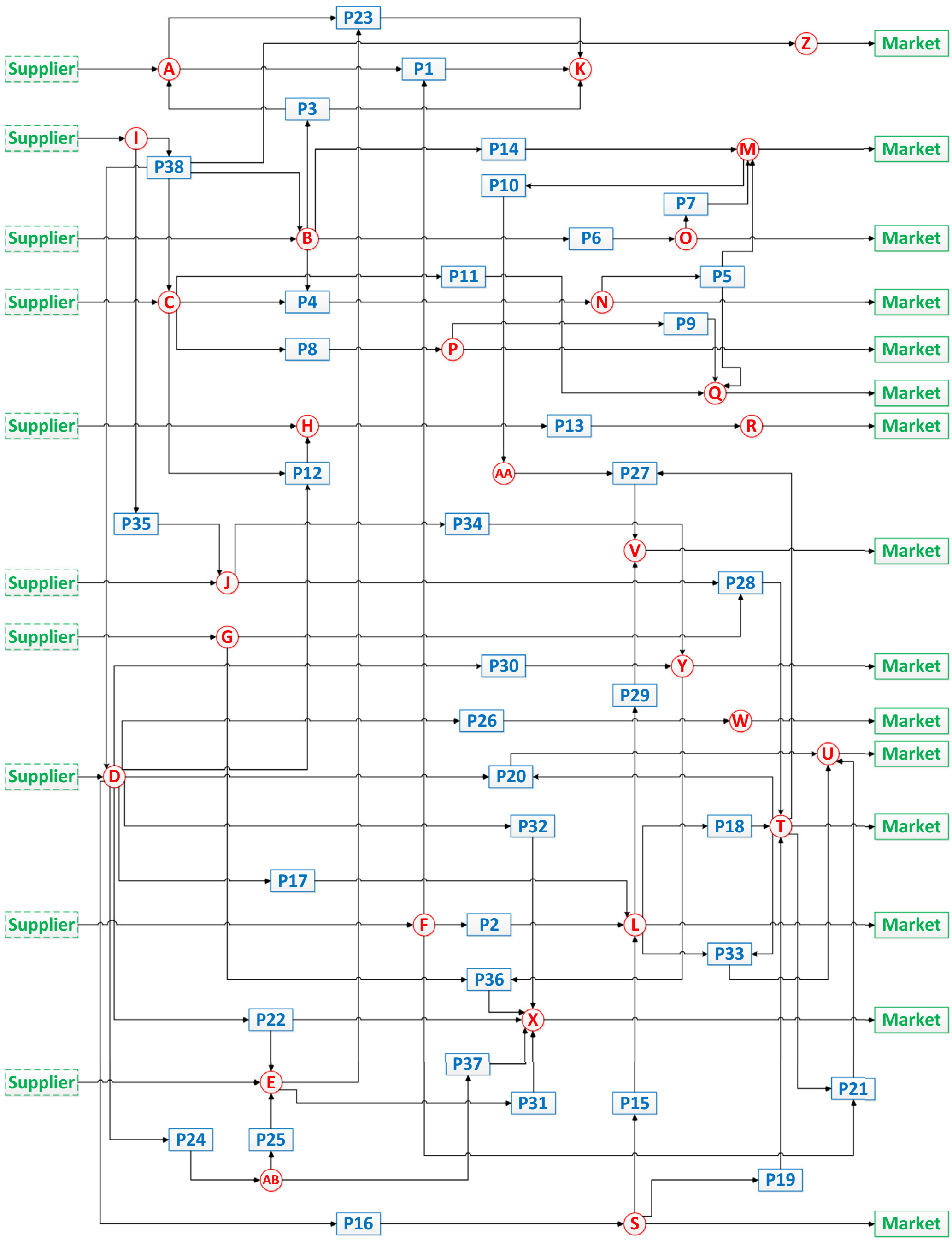


Fig. 6. The chemical process network.

Step 4. Calculate the mean value of total profits over all random samples.

In this way, the performances of DRO and ARO can be compared under the same lens, and the associated results are reported in Table 3. It can be observed that when the magnitude of uncertainties is large, DRO significantly overtakes ARO, although the ARO problems have been solved to global optimality and an approximate solution is obtained for DRO problems through ADRs. In the case of weak uncertainties ($G = 0.25$), both DRO and ARO yield nearly the same level of profits. This is reasonable because in such case, the worst-case realization of uncertainties that ARO protects against is not excessively pessimistic. By further scrutinizing the first-stage profit and the second-stage profit, we can see that optimal decisions determined by DRO tend to make more investment in the first-stage. This is also verified by Gantt charts in Fig. 4, as well as results of purchase amounts in Table 4, which show that more initial materials, notably M1, are bought by the two-stage DRO-based approach under different levels of uncertainties. Such a high investment in turn pays off, since higher second-stage profits as well as more total profits can be obtained on average under demand uncertainties. It well demonstrates that, the proposed two-stage DRO approach leads to less conservative solutions by utilizing statistical information from data and achieves better expected performance under uncertainties.

Moreover, we examine the worst total profits over all random samples obtained by two different optimization approaches. The corresponding results are reported in Table 5. It can be seen that the two-stage ARO approach achieves better worst-case perfor-

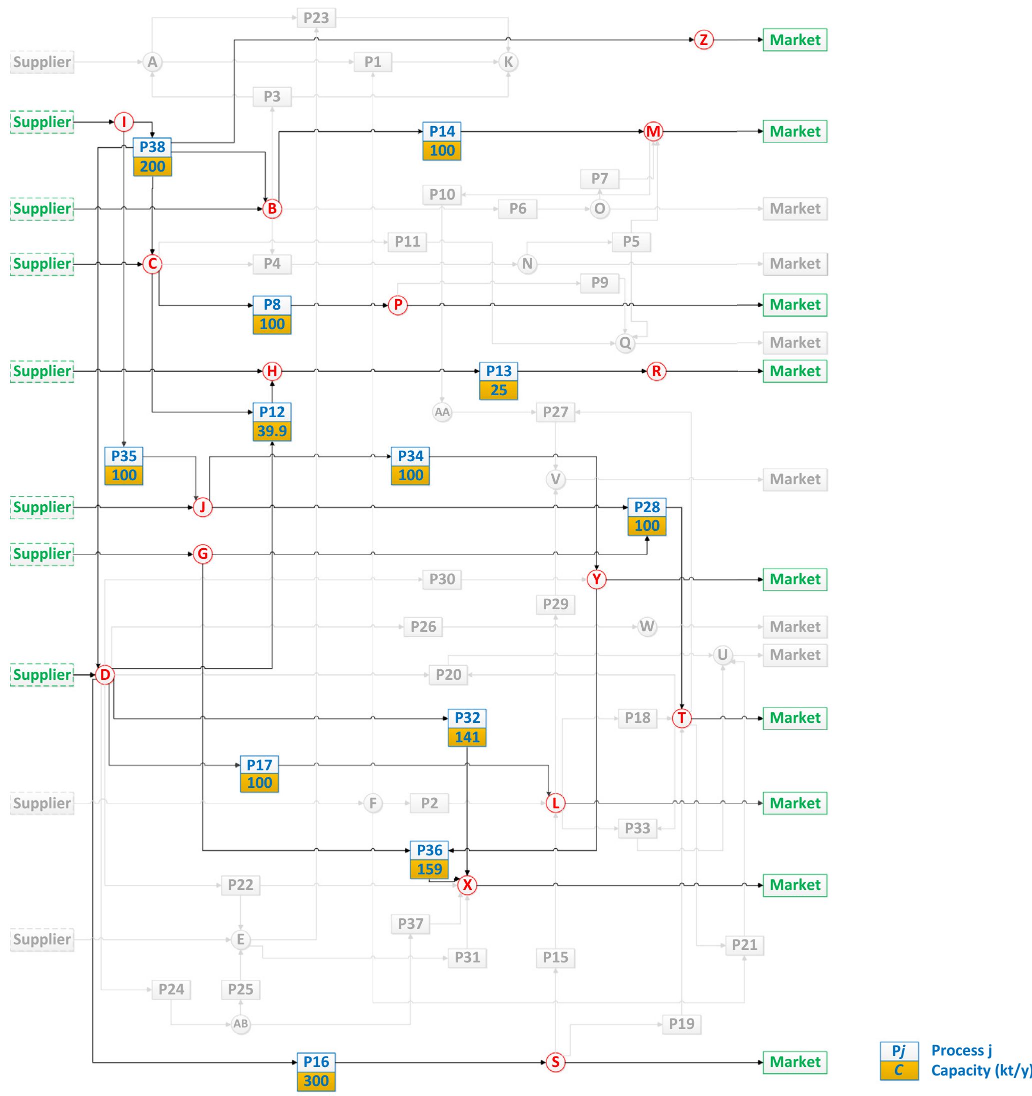


Fig. 7. Optimal design and planning decisions in time period 5 determined by the deterministic model.

mances than DRO under different level of uncertainties. This is primarily due to different optimization criteria adopted by two approaches. Specifically, when the magnitude of uncertainty is small ($G = 0.25$), two approaches tend to yield similar performances in a worst-case sense. As the magnitude of uncertainty increases, the gap between two approaches becomes significant. It implies that DRO may have some limitations in protecting against the worst-case realization of uncertainty, and is thus particularly preferred when one cares more about the expected performance rather than the worst-case performance.

Before closing this section, we further investigate the impact of number $I = 2K + 1$ of truncation points $\{\mathbf{q}_i\}$ in (5) on the optimization performance. Intuitively, when more truncation points are involved in the ambiguity set $\mathcal{D}$ (5), more statistical information can be incorporated into the optimization problem, and hence the final solution becomes less conservative. Here we adopt a smaller step-size $\lambda_n/2$ to induce a more dense distribution of truncation points, and then vary the value of K from zero to four. The optimization results in terms of computational costs as well as objective values are reported in Fig. 5. We can observe that when the number of truncation points increases, a higher profit can be obtained and conservatism of solutions is reduced expectedly, since the ambiguity set becomes more compact. Meanwhile, when K begins to increase from one, the computational cost appears to decrease, in spite of an increasing model complexity. This may be explained by the fact that a smaller "size" of the ambiguity set is obtained, which leads to a smaller search space of binary first-stage variables and renders the branch-and-cut algorithm more efficient. However, when K increases from three to four, such an phenomenon becomes insignificant, the effect of model complexity dominates and hence computational burden begins to increase. These results can provide useful guidelines for users to achieve a desirable trade-off between the model conservatism and computational costs.

6. Application to process network planning

In this section, the proposed DRO approach is applied to a multi-period process network planning problem, which is cast in a multi-stage decision-making setting. In chemical complexes, a number of interconnected processes and versatile chemicals are involved. By synthesizing all factors and modeling the entire problem mathematically, a rational decision can be made from multiple manufacturing options for producing a certain chemical (Yue and You, 2013). Moreover, capacity expansions are allowed in each period to maximize the overall profit. Here we consider a

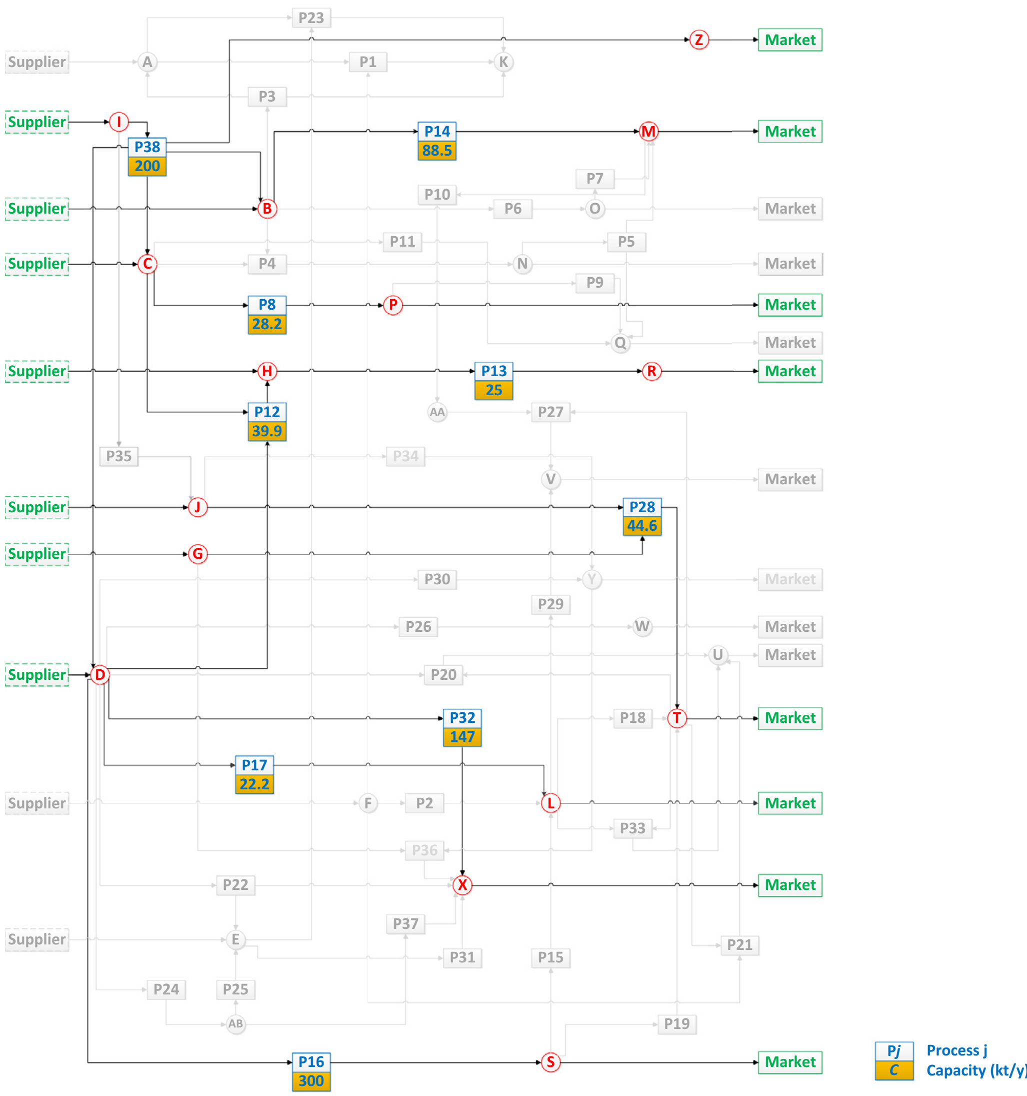


Fig. 8. Optimal design and planning decisions in time period 5 determined by the multi-stage DRO model.

large-scale process network involving 38 processes, 28 chemicals, 10 suppliers and 16 markets, whose structure is depicted in Fig. 6. The chemicals can be classified into raw materials (A-J), intermediates (AA,AB), and final products (K-Z). A ten-year planning horizon is considered here, which consists of five time periods in total, and each time period has two years. We first establish the deterministic process network planning problem, which is generally formulated as the following MILP (Liu and Sahinidis, 1996; You and Grossmann, 2011):

$$\max_{QE_{it}, Y_{it}, P_{jt}, Q_{it}, S_{jt}, W_{it}} \ -\sum_{i\in I}\sum_{t\in T}(\alpha_{it}\cdot QE_{it} + \beta_{it}\cdot Y_{it} + \delta_{it}\cdot W_{it}) + \sum_{j\in J}\sum_{t\in T}\left(\upsilon_{jt}\cdot S_{jt} - \tau_{jt}\cdot P_{jt}\right) \tag{79}$$

$$\text{s.t.} \qquad qe_{it}^{L}\cdot Y_{it} \le QE_{it} \le qe_{it}^{U}\cdot Y_{it}, \ \ \forall i\in I, t\in T \tag{80}$$

$$Q_{it} = Q_{i(t-1)} + QE_{it}, \ \ \forall i\in I, t\in T \tag{81}$$

$$\sum_{t\in T} Y_{it} \le ce_i, \ \ \forall i\in I \tag{82}$$

$$\sum_{i\in I}(\alpha_{it}\cdot QE_{it} + \beta_{it}\cdot Y_{it}) \le ci_t, \ \ \forall t\in T \tag{83}$$

$$W_{it} \le Q_{it}, \ \ \forall i\in I, t\in T \tag{84}$$

$$P_{jt} - \sum_{i}\kappa_{ij}\cdot W_{it} - S_{jt} = 0, \ \ \forall j\in J, t\in T \tag{85}$$

$$P_{jt} \le su_{jt}, \ \ \forall j\in J, t\in T \tag{86}$$

$$S_{jt} \le du_{jt}, \ \ \forall j\in J, t\in T \tag{87}$$

$$QE_{it}, Q_{it}, P_{jt}, W_{it}, S_{jt} \ge 0, \ \ \forall i\in I, j\in J, t\in T \tag{88}$$

$$Y_{it}\in\{0,1\}, \ \ \forall i\in I, t\in T \tag{89}$$

The objective (79) intends to maximize the net present value (NPV) of the process network over the entire planning horizon,

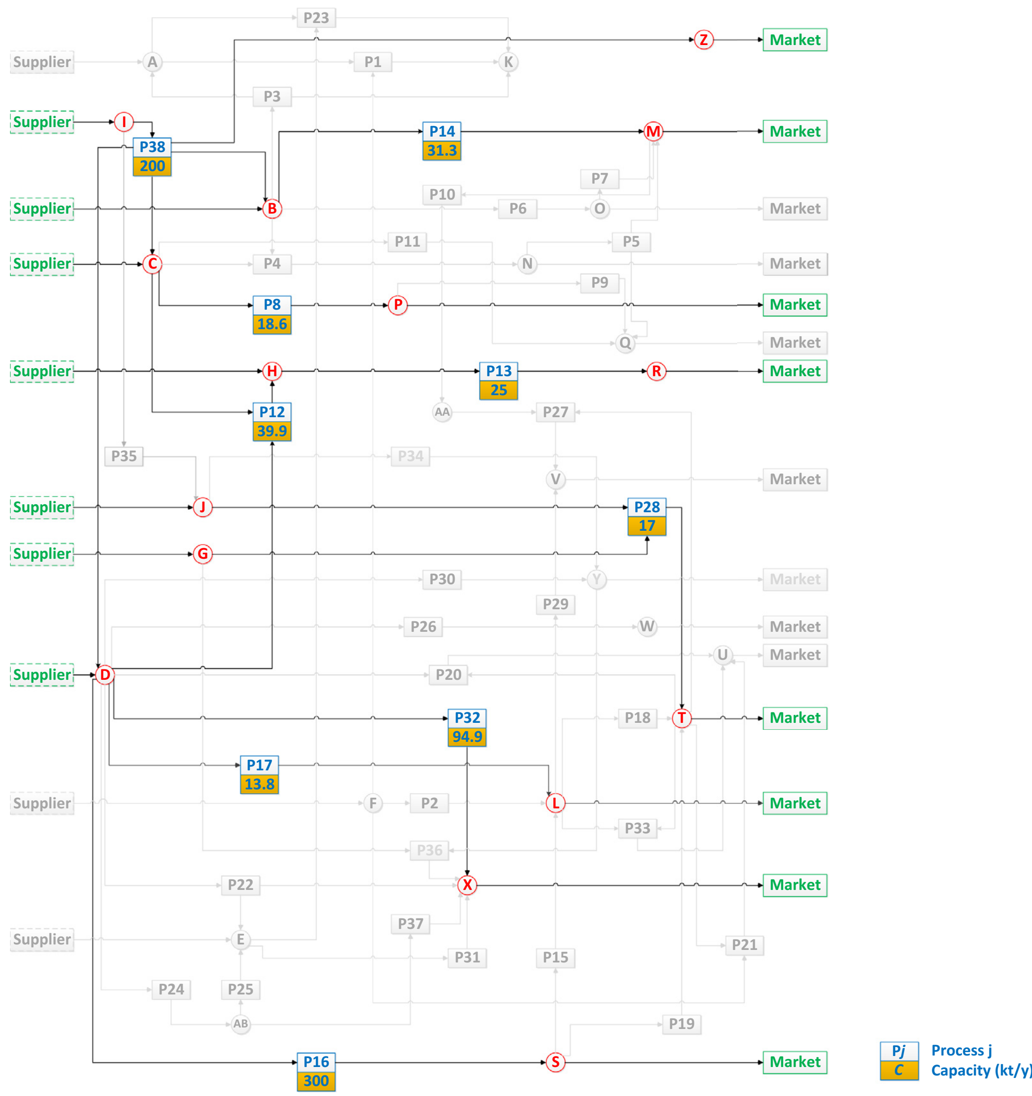


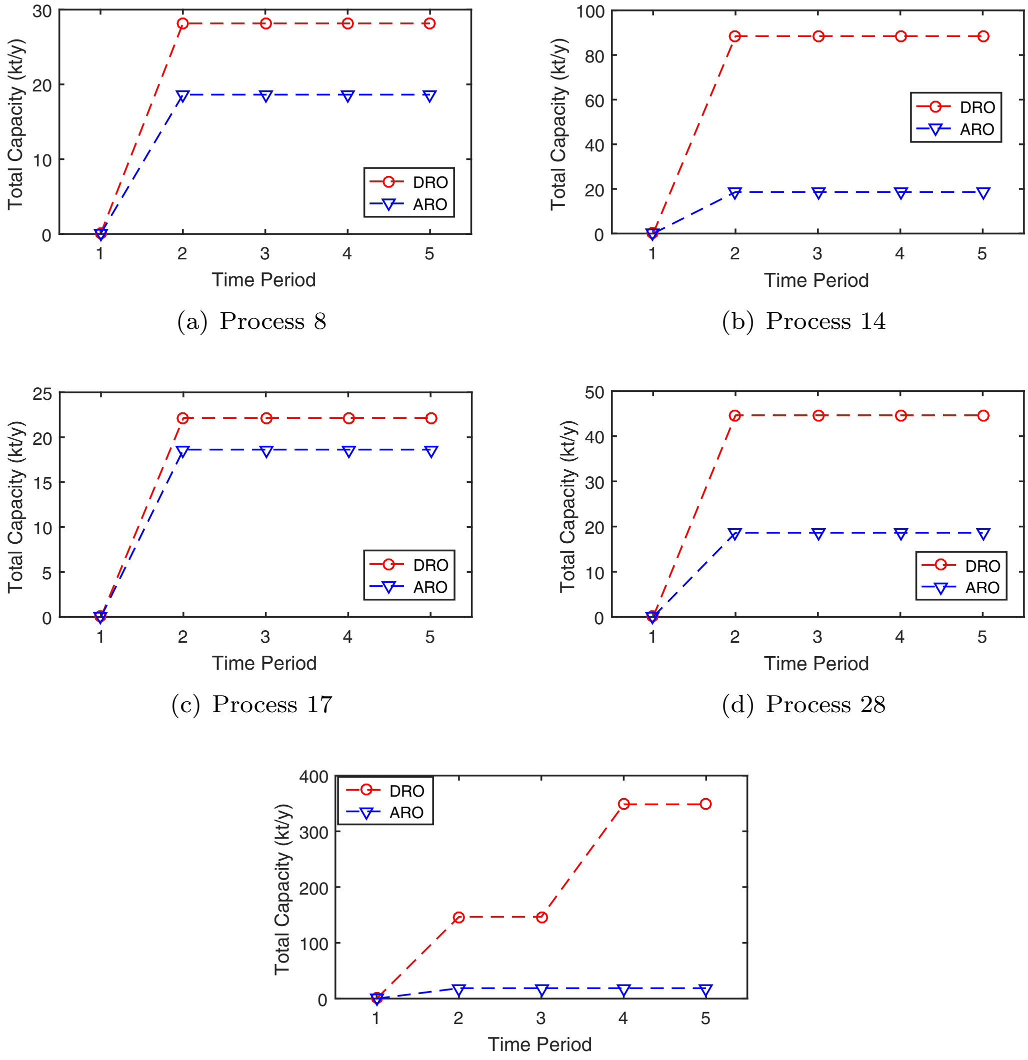
Fig. 9. Optimal design and planning decisions in time period 5 determined by the multi-stage ARO model.

which is made up of investment costs, operating costs, material purchase costs, and sales profits. Constraint (80) enforces the upper and lower limits of capacity expansions in each period, while (81) indicates the additive nature of expanded capacities. Inequalities (82) and (83) limit the budgets of capacity expansions in each time period. The equality (85) speaks about the mass balance. (86) and (87) enforce that the purchase amount and sales amount cannot exceed the limits of suppliers and markets. The non-negativity of continuous variables is ensured by (88). Binary variables $\{Y_{it}\}$ are used to indicate whether the capacity of process i will be expanded in time period t. It is assumed that Processes 12, 13, 16, and 38 have initial capacities of 19.9, 12.5, 150 and 100 kton/year, respectively. Other parameter configurations are detailed in Appendix C (See Supplementary).

In this case, we consider random demands $\{du_{jt}\}$ over five time periods, which constitute a 80-dimensional random vector $\boldsymbol{\xi}$, and one thousand uncertainty data are collected in total. Variables $\{Y_{it},\ QE_{it},\ Q_{it}\}$ pertaining to capacity expansions are specified as here-and-now decisions $\mathbf{x}$ that are implemented before uncertainty demands $\{du_{jt}\}$ are known, while the operating levels $\{W_{it}\}$, purchase amounts $\{P_{jt}\}$ and sales amounts $\{S_{jt}\}$ serve as recourse decisions $\mathbf{y}$. We denote by $\mathbf{d}_{1:\,t}$ the realizations of uncertain demands up to time period t, and assume that $\{W_{it}\}$, $\{P_{jt}\}$ and $\{S_{jt}\}$ are sequentially determined after demands $\mathbf{d}$ are unfolded stage by stage, which yields a multi-stage decision-making process. Such a dependence can be mathematically described as $\{W_{it}(\mathbf{d}_{1:\,t})\}$, $\{P_{jt}(\mathbf{d}_{1:\,t})\}$ and $\{S_{jt}(\mathbf{d}_{1:\,t})\}$. To model the sequential behavior of decision making and derive a tractable approximation, we adopt the ADR with non-anticipativity constraints detailed in Section 4.2.

Similar to the previous case study, we solve the planning case using the deterministic model, multi-stage DRO model and multi-stage ARO model. Parameters in the deterministic model are set as their nominal values. In the multi-stage DRO model, the number of truncation point is set as seven ($K = 3$). As with the multi-stage ARO model, the box uncertainty set is adopted with its size estimated with 1000 available samples. Different from the two-stage case, here approximations of the multi-stage ARO problem must be made by adopting the ADR because the decomposition algorithm no longer works in the multi-stage setting. Finally the multi-stage ARO problem is cast as an MILP. All problems are modelled in GAMS 24.7.4, in which all induced MILPs are solved using CPLEX 12.7.0 with optimality gaps set as 0.1%.

Table 6 showcases the performance comparisons of solving various optimization problems under demand uncertainties. The multi-stage DRO problem has more variables and constraints than the other two problems, and thus takes the most time to



(a) Process 8 (b) Process 14 (c) Process 17 (d) Process 28 (e) Process 32

Fig. 10. Optimal capacity expansion decisions over the entire planning horizon determined by multi-stage DRO and ARO.

Table 6
Optimization results in the process network planning case.

	Deterministic Planning	Multi-Stage DRO	Multi-Stage ARO
Cont. var.	851	753,732	189,012
Bin. var.	190	190	190
Constraints	1224	407,775	103,215
CPU time (s)	0.11	1264.16	100.69
Objective (M$)	7521	2741	2434

solve. In spite of this, the computational burden is still affordable in practice. In terms of objectives, solving the multi-stage DRO problem returns 12.61% higher NPV than solving the multi-stage ARO problem, primarily due to different risk measures adopted. The optimal objective of the deterministic problem is much higher than those of two multi-stage problems. Such a gap in objectives is also due to the conservatism induced by ADRs for solving two multi-stage problems. Nonetheless, multi-stage problems are still preferred in an uncertain environment, since the deterministic problem organically falls short of addressing demand uncertainties and reducing potential risks.

Next we investigate the conservatism of different approaches by taking a closer look at their decisions. The optimal process design and planning decisions in time period 5 derived by solving the deterministic problem, the multi-stage DRO and ARO problems are visualized in Figs. 7–9, respectively, in which total capacities are indicated under each operated process. We can see from Fig. 7 that Processes 8, 12, 13, 14, 16, 17, 28, 32, 34, 35, 36, 38 are chosen as operating processes by the deterministic planning approach. By contrast, Figs. 8 and 9 show that less processes are operated by multi-stage DRO and ARO since capacities of Processes 34, 35, and 36 have not been expanded all along. It bears rationality because less capacity expansions are determined in a "here-and-now" manner by two multi-stage approaches and some unnecessary costs can be avoided in face of extremal realizations of random demands. Specifically, although the same processes have been selected for capacity expansions by two multi-stage approaches, their exact amounts of expansions are different. We can observe from Figs. 8 and 9 that, multi-stage DRO typically decides much more capacities than multi-stage ARO does, especially for Processes 8, 14, 17, 28, and 32, indicating that a less conservative solution is obtained.

Fig. 10 further highlights the changes of capacities over the entire planning horizon determined by the two multi-stage approaches. Note that capacities of Processes 8, 14, 17, and 28 are expanded by all approaches at time period 2. As for Process 32, its capacity is expanded at time period 4 one more time by the multi-stage DRO approach. Therefore, it has a much larger capacity at the final stage, which is much less conservative than that obtained by multi-stage ARO, and could enable higher profits under demand randomness. It indicates that ARO only accounts for the worst-case realization within the support set, thereby resulting in conservative solutions. By contrast, the proposed DRO-based planning approach is able to utilize more meaningful distributional information from data by using the ambiguity set, thereby better hedging against uncertainties.

7. Concluding remarks

In this paper, a novel DRO approach is proposed for solving process planning and scheduling problems under uncertainties, and hedging against the inexactness of probability distributions of uncertainties. The most-used formulations of ambiguity sets in literature cannot be directly applicable in large-scale mixed-integer problems, and there are few approaches to the parameter tuning of general ambiguity sets, thereby posing significant challenges for applying DRO to process planning and scheduling. To address these research challenges, an effective data-driven approach is proposed for ambiguity set constructions based on PCA and first-order deviation functions, as an important contribution of this work. When the deterministic problem is an MILP, the proposed ambiguity set finally leads to MILP reformulations of the DRO counterpart with additional linear constraints, which can be conveniently handled by general purpose commercial grade solvers. In addition, the proposed data-based approach can help excavating rich statistical information from uncertainty data, including both second-order and high-order moment information. To account for the multi-stage and multi-level decision-making structure in process operations, we further develop multi-stage DRO models by including "wait-and-see" decisions made after uncertainty realizations, and adopt ADR approximations to derive tractable solutions. Applications in industrial process network planning and batch process scheduling demonstrate that, by leveraging data information, the proposed DRO-based approach can effectively hedge against distributional ambiguity, since the conservatism of solutions can be greatly reduced and more profits can be attained by utilizing inexact distributional information. Although we only report applications in planning and scheduling in this work, the proposed approach is substantially based upon a general form of MILP for the deterministic model, so it enjoys promising generality in dealing with various problems in process systems engineering, which are worth future investigations.

Acknowledgement

We acknowledge partial financial support from the National Science Foundation (NSF) CAREER Award (CBET-1643244).

Appendix A. Equivalence of ambiguity sets

Proposition 1. *The ambiguity set $\mathcal{D}$ in (5) is equivalent to the set including all marginal distribution of $\boldsymbol{\xi}$ under $\mathbb{Q} \in \bar{\mathcal{D}}$ in (6).*

Proof. Let $\prod_{\boldsymbol{\xi}} \bar{\mathcal{D}}$ stand for the set of all marginal distributions of $\boldsymbol{\xi}$ under any $\mathbb{Q} \in \bar{\mathcal{D}}$. We first show that $\prod_{\boldsymbol{\xi}} \bar{\mathcal{D}} \subseteq \mathcal{D}$. Given $\mathbb{Q} \in \bar{\mathcal{D}}$, we assume that the marginal distribution of $\boldsymbol{\xi}$ under $\mathbb{Q}$ is denoted as $\mathbb{P}$. It is then obvious that $\mathbb{P}\{\boldsymbol{\xi} \in \Xi\} = 1$. In addition, we have

$$\mathbb{E}_{\mathbb{P}}\{g_i(\boldsymbol{\xi})\} = \mathbb{E}_{\mathbb{Q}}\{g_i(\boldsymbol{\xi})\} \le \mathbb{E}_{\mathbb{Q}}\{\boldsymbol{\varphi}\} \le \boldsymbol{\gamma}, \tag{90}$$

where the first inequality is due to the fact that $\mathbb{P}\{(\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi}\} = 1$, and the second inequality arises from the definition of the lifted ambiguity set $\bar{\mathcal{D}}$ in (6). This indicates that $\mathbb{P} \in \mathcal{D}$ and hence we have $\prod_{\boldsymbol{\xi}} \bar{\mathcal{D}} \subseteq \mathcal{D}$.

Conversely, suppose $\mathbb{P} \in \mathcal{D}$. We then construct a joint probability distribution $\mathbb{Q}$ of random variables $(\boldsymbol{\xi}, \boldsymbol{\varphi})$, whose realization satisfies

$$(\boldsymbol{\xi}, \boldsymbol{\varphi}) = (\boldsymbol{\xi}, g(\boldsymbol{\xi})), \tag{91}$$

where $\boldsymbol{\xi}$ follows the distribution $\mathbb{P}$. In other words, the realization of $\boldsymbol{\varphi}$ is a deterministic function of $\boldsymbol{\xi}$. Then it can be readily verified that

$$\mathbb{P}\{(\boldsymbol{\xi}, \boldsymbol{\varphi}) \in \bar{\Xi}\} = 1, \tag{92}$$

and

$$\mathbb{E}_{\mathbb{Q}}\{\boldsymbol{\varphi}\} = \mathbb{E}_{\mathbb{Q}}\{g(\boldsymbol{\xi})\} = \mathbb{E}_{\mathbb{P}}\{g(\boldsymbol{\xi})\} \le \boldsymbol{\gamma}. \tag{93}$$

Therefore, we have $\mathbb{P} \in \bar{\mathcal{D}}$ and thus $\prod_{\boldsymbol{\xi}} \bar{\mathcal{D}} \supseteq \mathcal{D}$. This completes the proof. □

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.compchemeng.2017.12.002.

References

Acevedo, J., Pistikopoulos, E.N., 1998. Stochastic optimization based algorithms for process synthesis under uncertainty. Comput. Chem. Eng. 22, 647–671.
Ahmed, S., Sahinidis, N.V., 1998. Robust process planning under uncertainty. Ind. Eng. Chem. Res. 37, 1883–1892.
Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A., 2004. Adjustable robust solutions of uncertain linear programs. Math. Program. 99, 351–376.
Bertsimas, D., Doan, X.V., Natarajan, K., Teo, C.P., 2010. Models for minimax stochastic linear optimization problems with risk aversion. Math. Oper. Res. 35, 580–602.
Bertsimas, D., Gupta, V., Kallus, N., 2017a. Data-driven robust optimization. Math. Program. 1–58. doi:10.1007/s10107-017-1125-8.
Bertsimas, D., Sim, M., & Zhang, M. (2017b). A practically efficient approach for solving adaptive distributionally robust linear optimization problems. URL http://www.optimization-online.org/DB_FILE/2016/03/5353.pdf.
Castro, P.M., Barbosa-Póvoa, A.P., Matos, H.A., Novais, A.Q., 2004. Simple continuous–time formulation for short-term scheduling of batch and continuous processes. Ind. Eng. Chem. Res. 43, 105–118.
Chen, X., Sim, M., Sun, P., Zhang, J., 2008. A linear decision-based approximation approach to stochastic programming. Oper. Res. 56, 344–357.
Chu, Y., Wassick, J.M., You, F., 2013. Efficient scheduling method of complex batch processes with general network structure via agent-based modeling. AIChE J. 59, 2884–2906.
Chu, Y., You, F., 2015. Model-based integration of control and operations: overview, challenges, advances, and opportunities. Comput. Chem. Eng. 83, 2–20.
Delage, E., Ye, Y., 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. 58, 595–612.
Ghaoui, L.E., Oks, M., Oustry, F., 2003. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. Oper. Res. 51, 543–556.
Goh, J., Sim, M., 2010. Distributionally robust optimization and its tractable approximations. Oper. Res. 58, 902–917.
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.
Grossmann, I.E., Apap, R.M., Calfa, B.A., Garcia-Herreros, P., Zhang, Q., 2016. Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty. Comput. Chem. Eng. 91, 3–14.
Hanasusanto, G.A., Roitch, V., Kuhn, D., Wiesemann, W., 2017. Ambiguous joint chance constraints under mean and dispersion information. Oper. Res. 65, 751–767.
Jiao, Z., Ran, L., Guan, L., Wang, X., Chu, H., 2017. Fleet management for electric vehicles sharing system under uncertain demand. In: International Conference on Service Systems and Service Management (ICSSSM). IEEE, pp. 1–6.
Li, Z., Ierapetritou, M., 2008. Process scheduling under uncertainty: review and challenges. Comput. Chem. Eng. 32, 715–727.
Lin, X., Janak, S.L., Floudas, C.A., 2004. A new robust optimization approach for scheduling under uncertainty:: I. Bounded uncertainty. Comput. Chem. Eng. 28, 1069–1085.
Liu, M.L., Sahinidis, N.V., 1996. Optimization in process planning under uncertainty. Ind. Eng. Chem. Res. 35, 4154–4165.

Liu, Y., Xu, H., Yang, S.J.S., Zhang, J., 2017. Distributionally robust equilibrium for continuous games: Nash and Stackelberg models. Eur. J. Oper. Res.

Nakao, H., Shen, S., Chen, Z., 2017. Network design in scarce data environment using moment-based distributionally robust optimization. Comput. Oper. Res. 88, 44–57.

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., 2018. 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.

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

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

Shapiro, A., 2001. On duality theory of conic linear problems. In: Semi-Infinite Programming. Springer, pp. 135–165.

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.

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

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.

Wassick, J.M., Agarwal, A., Akiya, N., Ferrio, J., Bury, S., You, F., 2012. Addressing the operational challenges in the development, manufacture, and supply of advanced materials and performance products. Comput. Chem. Eng. 47, 157–169.

Wei, W., Liu, F., Mei, S., 2016. Distributionally robust co-optimization of energy and reserve dispatch. IEEE Trans. Sustain. Energy 7, 289–300.

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

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52.

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.

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

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, 1547–5905.

Yue, J., Chen, B., Wang, M.C., 2006. Expected value of distribution information for the newsvendor problem. Oper. Res. 54, 1128–1136.

Zhang, Q., Morari, M.F., Grossmann, I.E., Sundaramoorthy, A., Pinto, J.M., 2016. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. Comput. Chem. Eng. 86, 106–119.

Zhang, Y., Shen, S., Erdogan, S.A., 2017. Distributionally robust appointment scheduling with moment-based ambiguity set. Oper. Res. Lett. 45, 139–144.

Zhao, C., Jiang, R., 2017. Distributionally robust contingency-constrained unit commitment. IEEE Trans. Power Syst.

Zymler, S., Kuhn, D., Rustem, B., 2013. Distributionally robust joint chance constraints with second-order moment information. Math. Program. 1–32.