Robust Stochastic Optimization

Zhi Chen
Department of Management Sciences, College of Business, City University of Hong Kong
zhi.chen@cityu.edu.hk

Melvyn Sim, Peng Xiong
Department of Analytics & Operations, NUS Business School, National University of Singapore
dcsimm@nus.edu.sg, bizxio@nus.edu.sg

We present a new distributionally robust optimization model called the robust stochastic optimization (RSO), which unifies both scenario-tree based stochastic linear optimization and distributionally robust optimization in a practicable framework that can be solved using the state-of-the-art commercial optimization solvers. The model of uncertainty incorporates both discrete and continuous random variables, typically assumed in scenario-tree based stochastic linear optimization and distributionally robust optimization respectively. To address the non-anticipativity of recourse decisions, we introduce the event-wise recourse adaptations, which integrate the scenario-tree adaptation originating from stochastic linear optimization and the affine adaptation popularized in distributionally robust optimization. Our proposed event-wise ambiguity set is rich enough to capture traditional statistic-based ambiguity sets with convex generalized moments, mixture distribution, ϕ-divergence, Wasserstein (Kantorovich-Rubinstein) metric, and also opens up novel machine-learning-based ones using techniques such as K-means clustering, and classification and regression trees (CART). We also provide several interesting RSO models, including optimizing over the Hurwicz criterion and two-stage problems over Wasserstein ambiguity sets. We develop a new algebraic modeling package, RSOME to facilitate the implementation of RSO models.

Key words: stochastic linear optimization, distributionally robust optimization, machine learning.

History: July 19, 2019.

1. Introduction

In the era of data analytics, the ubiquity of general purpose deterministic mathematical optimization frameworks such as linear, mixed-integer and conic optimization models, as well as their impact on improving management decision making, cannot be understated. Algebraic modeling packages and state-of-the-art optimization solvers have been developed on these successful frameworks to facilitate implementation of prescriptive analytics to address a wide variety of real-world problems. Comparatively, frameworks to support generic modeling and optimization under uncertainty, despite their importance, are relatively less established. These frameworks include stochastic linear optimization, robust optimization and more recently, distributionally robust optimization, each of them has its strengths and weaknesses.
Stochastic linear optimization extends the linear optimization framework to minimize the total average cost associated with the optimal here-and-now and wait-and-see (or recourse) decisions under a known probability distribution (Danzig 1955). For enormous or infinite number of scenarios, we can obtain approximate here-and-now solutions using the sample average approximation (SAA) (Kall and Wallace 1994, Birge and Louveaux 2011, Shapiro and Homem-de Mello 1998, Kleywegt et al. 2002). These approximate solutions are necessarily random and optimistically biased, i.e., the actual realized objectives are statistically worse off than those attained by using SAA. Stochastic linear optimization has the versatility of modeling different types of recourse decisions including those with discrete outcomes, albeit at the expense of greater computational effort.

In classical robust optimization (Soyster 1973, Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998, Bertsimas and Sim 2004, Ben-Tal et al. 2015), the solution is obtained by reformulating the model to a deterministic optimization problem that can be solved using available solvers. The underlying uncertainty is a distribution-free continuous random variable with support confined within a convex uncertainty set. The solution obtained via classical robust optimization hedges against the worst-case outcome within the uncertainty set and hence is pessimistically biased, i.e., the realized objective value by the robust solution would often be better than the objective value attained by solving the robust optimization problem. To reduce the conservativeness, distributionally robust optimization incorporates an ambiguity set of probability distributions and its solution hedges against the worst-case distribution within the ambiguity set (Dupaˇcov´a 1976, Shapiro and Kleywegt 2002, El Ghaoui et al. 2003, Shapiro and Ahmed 2004, Delage and Ye 2010, Wiesemann et al. 2014). Under an embedded linear optimization framework, both robust optimization and distributionally robust optimization have been extended to address problems with recourse decisions (see, for instance, Ben-Tal et al. 2004, Takriti and Ahmed 2004, Bertsimas et al. 2019). As these models are generally computationally intractable (Shapiro and Nemirovski 2005, Ben-Tal et al. 2004), approximate solutions are sought by restricting the recourse decisions to affine mappings of the uncertainty and by requiring the recourse decisions to remain feasible almost surely.

Our goal in this paper is to introduce a new distributionally robust optimization model that we call robust stochastic optimization (RSO), which unifies both scenario-tree based stochastic linear optimization and distributionally robust optimization in a tractable framework. The RSO framework incorporates both discrete and continuous random variables, and introduces the event-wise static and event-wise affine adaptations to address the non-anticipativity of recourse decisions. The equipped event-wise ambiguity set is rich to capture traditional statistic-based ambiguity sets and also opens up novel ones using machine learning techniques such as K-means clustering, and classification and regression trees. We showcase several interesting models that we can integrate in our framework. We develop the RSO framework with the mindset that it can be integrated in a
general purpose software that would be accessible to modelers. As a proof of concept, we develop an algebraic modeling package, RSOME (Robust Stochastic Optimization Made Easy) to facilitate modeling of problems under the RSO framework. With its intuitive syntax, RSOME provides a convenient way for modelers to specify RSO models and interfaces with state-of-the-art commercial solvers such as CPLEX, Gurobi, and MOSEK to obtain the optimal solutions to these models.

Our work significantly extends the frameworks of Wiesemann et al. (2014) and Bertsimas et al. (2019). In terms of modeling uncertainty, the ambiguity sets in these two models do not include discrete random variable and they are special cases of our event-wise ambiguity set. Moreover, their ambiguity sets are chiefly moment-based and do not naturally incorporate statistical-distance-based-information such as Wasserstein metric or $\phi$-divergence. We also would like to highlight the difficulties of developing an algebraic modeling software based on the Wiesemann et al. (2014) framework. Among other things, to ensure tractability, the model requires to impose a nesting condition on the confidence sets and a key challenge is to check whether the nesting condition holds. To resolve these issues collectively, we introduce a discrete random variable as part of the event-wise ambiguity set and associate its outcome with confidence sets. An immediate consequence is the assurance of tractability without having to impose additional conditions on the confidence sets such as nesting and disjoint. In terms of recourse adaptation, Wiesemann et al. (2014) do not consider recourse decisions, while the model of Bertsimas et al. (2019) is based on affine recourse adaptation, which is a special case of our event-wise affine adaptation.

**Notations.** Boldface uppercase and lowercase characters denote matrices and vectors, respectively. We denote by $[N] \triangleq \{1, 2, \ldots, N\}$ the set of positive running indices up to $N$. We use $\mathcal{P}_0(\mathbb{R}^t)$ to represent the set of all distributions on $\mathbb{R}^t$. A random variable, $\tilde{z}$ is denoted with a tilde sign and we use $\tilde{z} \sim \mathbb{P}, \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^t)$ to define $\tilde{z}$ as an $I_z$-dimensional random variable with distribution $\mathbb{P}$.

**2. Framework for Robust Stochastic Optimization**

We now introduce the robust stochastic optimization (RSO) model, which combines both scenario-tree based stochastic linear optimization and distributionally robust optimization in a unified framework. The uncertainty associated with the RSO model comprises both discrete and continuous random variables. Specifically, $\tilde{s}$ represents a discrete random scenario taking values in $[S]$, while $\tilde{z}$ represents a continuous random variable with outcomes in $\mathbb{R}^{I_z}$. Conditioning on the realization of a scenario $s \in [S]$, the support set of the random variable $\tilde{z}$ is tractable conic representable and is denoted by $Z_s$. The joint distribution of $(\tilde{z}, \tilde{s})$ is denoted by $\mathbb{P} \in \mathcal{F}$, where $\mathcal{F}$ is the ambiguity set of probability distributions that share some identical distributional information.

We denote by $w \in \mathbb{R}^{J_w}$ the here-and-now decision of the RSO model. The recourse decisions depend on the realization of the random variables $\tilde{z}$ and $\tilde{s}$. As it would become clearer, to obtain
a tractable reformulation, we introduce two types of recourse decisions, which are function maps respectively denoted by \( x(s) : [S] \mapsto \mathbb{R}^{J_x} \) and \( y(s, z) : [S] \times \mathbb{R}^{I_z} \mapsto \mathbb{R}^{J_y} \). Here, \( x(\cdot) \) adapts only to the outcome of the random scenario \( \tilde{s} \), while \( y(\cdot, \cdot) \) adapts to the outcomes of \( \tilde{s} \) and \( \tilde{z} \). Similar to most tractable adaptive robust optimization problems, the RSO model requires that for each given scenario \( s \in [S] \), the function map \( y(s, z) \) is affinely dependent on \( z \) as follows:

\[
y(s, z) \triangleq y^0(s) + \sum_{i \in [I_z]} y^i(s) z_i,
\]

where \( y^0(s), \ldots, y^{I_z}(s) \) account for the raw decision variables associated with \( y(\cdot, \cdot) \) at scenario \( s \).

To characterize the objective function (with index 0) and constraints (with indices \( m \in [M] \)), we first define the following random variable mappings for all \( m \in [M] \cup \{0\} \),

\[
\begin{align*}
    a_m(s, z) & \triangleq a^0_{ms} + \sum_{i \in [I_z]} a^i_{ms} z_i \\
    b_m(s, z) & \triangleq b^0_{ms} + \sum_{i \in [I_z]} b^i_{ms} z_i \\
    c_m(s) & \triangleq c_{ms} \\
    d_m(s, z) & \triangleq d^0_{ms} + \sum_{i \in [I_z]} d^i_{ms} z_i
\end{align*}
\]

for given parameters

\[
a^i_{ms} \in \mathbb{R}^{J_x}, b^i_{ms} \in \mathbb{R}^{J_x}, c_{ms} \in \mathbb{R}^{J_y}, d^i_{ms} \in \mathbb{R} \quad \forall i \in [I_z] \cup \{0\}, \quad s \in [S].
\]

The objective function of the RSO model to be minimized,

\[
\sup_{\tilde{v} \in \mathcal{F}} \mathbb{E}_{\tilde{z}} \left[ a^\top_0 (\tilde{s}, \tilde{z}) w + b^\top_0 (\tilde{s}, \tilde{z}) x(\tilde{s}) + c^\top_0 (\tilde{s}) y(\tilde{s}, \tilde{z}) + d_0(\tilde{s}, \tilde{z}) \right],
\]

reflects the ambiguity aversion of the decision maker against an ambiguity set \( \mathcal{F} \) that we will introduce subsequently. Note that the random variable \( \tilde{z} \triangleq (\tilde{u}, \tilde{v}) \) includes both the primary \( I_u \)-dimensional random variable \( \tilde{u} \) and the auxiliary (or lifted) \( I_v \)-dimensional random variable \( \tilde{v} \) associated with \( \tilde{u} \). As in the same spirit of linear optimization models, the provision of the auxiliary random variable \( \tilde{v} \) would greatly enhance the modeling power of the RSO model.

There are two types constraints: hard and soft ones, which are respectively associated with the partition \( \mathcal{M}_1, \mathcal{M}_2 \subseteq [M] \) of indices of constraints.

The hard constraints (with indices \( m \in \mathcal{M}_1 \)) of the RSO model, which must be satisfied almost surely, are given by the following set of semi-infinite constraints:

\[
a^\top_m (s, z) w + b^\top_m (s, z) x(s) + c^\top_m (s) y(s, z) + d_m(s, z) \leq 0 \quad \forall z \in Z_s, \quad s \in [S], \quad m \in \mathcal{M}_1.
\]

Observe that for each given scenario \( s \in [S] \), the semi-infinite constraint corresponds to the standard linear robust counterpart, which can be transformed to a modest sized constraint system that can
be handled by modern solvers. Scenario-tree based stochastic linear optimization is a special case of the RSO model when in the absence of the recourse decision \( y(\cdot, \cdot) \). Likewise, adaptive robust optimization is also a special case when \( S = 1 \).

To enhance the modeling, RSO also supports soft constraints (with indices \( m \in \mathcal{M}_2 \)), which must be satisfied in expectation over all distributions within the ambiguity set \( \mathcal{F} \):

\[
\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[ a_m(\tilde{s}, \tilde{z})w + b_m(\tilde{s}, \tilde{z})x(\tilde{s}) + c_m(\tilde{s})y(\tilde{s}, \tilde{z}) + d_m(\tilde{s}, \tilde{z}) \right] \leq 0 \quad \forall m \in \mathcal{M}_2.
\]

As in the objective function, soft constraints are evaluated in the expected sense, and hence, they capture the risk neutrality of the decision maker under ambiguity aversion. By introducing additional recourse decisions that are also embedded in the hard constraints, the RSO model is capable to capture risk-averse objective functions or safeguarding constraints; see Section 5.

Apart from hard and soft constraints, for a given scenario \( s \), we can impose additional constraints jointly on \( w, x(s) \) and \( y^0(s), \ldots, y^{I_z}(s) \). Specifically, we have

\[
r(s) \triangleq (w, x(s), y^0(s), \ldots, y^{I_z}(s)) \in \mathcal{X}_s \quad \forall s \in [S],
\]

where the feasible set \( \mathcal{X}_s \) may encompass nonlinear constraints such as conic and integral ones.

### 3. Event-Wise Recourse Adaptations

Stochastic linear optimization and distributionally robust optimization have different approaches for addressing dynamic decision making, where uncertainty is revealed in stages and recourse decisions should be non-anticipative to uncertainty realization. In distributionally robust optimization, this can be achieved by restricting the dependency of a recourse decision on only a subset of the uncertainty \( \tilde{z} \) that has been revealed (see an example in Figure 1). In stark contrast, dynamic modeling in scenario-tree based stochastic linear optimization is more involved and requires enumerating the complete sample paths from the beginning to the end of the decision horizon. In this regard, a scenario represents a sample path and a scenario tree is typically used to showcase sample paths as well as decisions (Høyland and Wallace 2001, Pflug 2001, Heitsch and Römisch 2009).

Figure 2 presents the scenario tree for a three-stage problem with five scenarios: in accordance of non-anticipativity, the first-stage decision \( w \) is independent of the scenarios; while the second-stage decision \( x_1(\cdot) \) shall be indifferent among scenarios 1, 2, and 3 and be indifferent between scenarios 4 and 5, and the third stage decision \( x_2(\cdot) \) can adapt to scenarios 1, 2, \ldots, 5.

To formally specify the event-wise adaptation of the recourse decision, \( x(\cdot) \), we first define an event \( \mathcal{E} \subseteq [S] \) by a subset of scenarios. A partition of scenarios then induces a collection \( \mathcal{C} \) of mutually exclusive and collectively exhaustive (MECE) events. Correspondingly, we define a
Figure 1  Timeline of a multi-stage problem. Uncertain parameters are revealed in stages as $z_1, \ldots, z_T$. The decision $w$, made before any uncertainty realizes, is non-adaptive to any uncertain parameters. The recourse decision $y_t$, made after observing the uncertainty realization $z_t$, is adaptive to all revealed $z_1, \ldots, z_t$.

![Timeline Diagram](image.png)

Figure 2  Scenario tree. The recourse decision $x_1(\cdot)$ satisfies $x_1(1) = x_1(2) = x_1(3)$ and $x_1(4) = x_1(5)$, while the recourse decision $x_2(\cdot)$ can take differently depending on the outcome from scenarios 1, 2, 3, 4, and 5. The second level of the scenario tree gives a collection of two MECE events $\{1, 2\}$ and $\{4, 5\}$, while the third level gives a collection of five singleton MECE events $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, and $\{5\}$. The collection of MECE events associated with a lower level (e.g., the third level) is nested in the collection associated with a higher level (e.g., the second level).

![Scenario Tree Diagram](image.png)

mapping $H_C : [S] \mapsto C$ such that $H_C(s) = \mathcal{E}$, for which $\mathcal{E}$ is the only event in $C$ that contains the scenario $s$. Given a collection $\mathcal{C}$ of MECE events, we define the *event-wise static adaptation*,

$$A(C) \triangleq \left\{ x : [S] \mapsto \mathbb{R} \mid x(s) = x^\mathcal{E}, \mathcal{E} = H_C(s) \text{ for some } x^\mathcal{E} \in \mathbb{R} \right\},$$

which follows along a similar vein as in the scenario-tree based stochastic linear optimization. Note that each level of the scenario tree naturally gives a partition of scenarios that induces a collection of MECE events, which in turn is used in the input of event-wise recourse adaptations for recourse decisions associated with that level (i.e., stage); see Figure 2. In Appendix E, we provide a financial planning example to illustrate the use of event-wise static adaptation to formulate a multi-stage stochastic linear optimization problem via scenario tree.

Similarly, for the recourse decision, $y(\cdot, \cdot)$, we define the *event-wise affine adaptation*,

$$\bar{A}(C, \mathcal{I}) \triangleq \left\{ y : [S] \times \mathbb{R}^{|I_z|} \mapsto \mathbb{R} \mid y(s, z) = y^0(s) + \sum_{i \in \mathcal{I}} y^i(s) z_i \text{ for some } y^0, y^i \in A(C), i \in \mathcal{I} \right\}$$

for a subset $\mathcal{I} \subseteq [I_z]$. The information set $\mathcal{I}$ tracks the indices of revealed uncertainties when deciding the recourse decision $y(\cdot, \cdot)$. Along with the collection $\mathcal{C}$ of MECE events, it captures the non-anticipativity of $y(\cdot, \cdot)$.
Armed with the event-wise recourse adaptations, we propose the following RSO framework:

$$\min \sup_{\mathcal{F}} \mathbb{E}[a_0^T(\bar{s}, \bar{z})w + b_0^T(\bar{s}, \bar{z})x(\bar{s}) + c_0^T(\bar{s})y(\bar{s}, \bar{z}) + d_0(\bar{s}, \bar{z})]$$

subject to

$$\mathbb{E}[a_m^T(s, z)w + b_m^T(s, z)x(s) + c_m^T(s)y(s, z) + d_m(s, z)] \leq 0 \quad \forall z \in \mathcal{Z}_s, s \in [S], m \in \mathcal{M}_1$$

$$\sup_{\mathcal{F}} \mathbb{E}[a_m^T(\bar{s}, \bar{z})w + b_m^T(\bar{s}, \bar{z})x(\bar{s}) + c_m^T(\bar{s})y(\bar{s}, \bar{z}) + d_m(\bar{s}, \bar{z})] \leq 0 \quad \forall m \in \mathcal{M}_2$$

$$w, x(s), y^0(s), \ldots, y^I(s) \in \mathcal{X}_s \quad \forall s \in [S]$$

$$x_j \in \mathcal{A}(\mathcal{C}_y^j, \mathcal{T}_y^j) \quad \forall j \in [J_y]$$

$$y_j \in \bar{\mathcal{A}}(\mathcal{C}_y^j, \mathcal{T}_y^j) \quad \forall j \in [J_y]$$

for given $\mathcal{C}_y^j, j \in [J_y]$ and $\mathcal{C}_y^j, j \in [J_y]$ of MECE events, and information index sets $\mathcal{T}_y^j, j \in [J_y]$.

For a given scenario $s \in [S]$, the objective function and constraints are bi-affine functions of the underlying decision variable $r(s) \in \mathbb{R}^{J_r}$. Hence, we can write

$$a_m^T(s, z)w + b_m^T(s, z)x(s) + c_m^T(s)y(s, z) + d_m(s, z) \triangleq r^T(s)G_m(s)z + h_m(s) \quad \forall m \in [M] \cup \{0\},$$

for parameters $G_m(s) \in \mathbb{R}^{J_r \times I_z}$ and $h_m(s) \in \mathbb{R}$. This relation would enable us to reformulate the hard constraints into deterministic constraint systems using standard robust optimization techniques.

The expansive RSO model includes both static and adaptive multi-stage problems in its presentation, and we can reformulate it as a deterministic optimization problem using our developed algebraic modeling toolbox, RSOME. Before we could do so, we will next introduce the ambiguity set for the objective function and constraints.

### 4. Event-Wise Ambiguity Set

We propose the event-wise ambiguity set, which is representable in the format

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z} \times [S]) \right\}$$

for given events $\mathcal{E}_k, k \in [K]$ and given closed and convex sets $\mathcal{Z}_s, s \in [S], \mathcal{Q}_k, k \in [K]$, and $\mathcal{P} \subseteq \{p \in \mathbb{R}^S \mid \sum_{s \in [S]} p_s = 1\}$. The random variable $\bar{s}$ indicates a set of random scenarios whose realization probabilities may be uncertain. For different scenarios, the support of the random variable $\bar{z}$ could be different, while conditioning on the event realization, the expectation of $\bar{z}$ can also differ. Quite notably, we can effectively determine the worst-case expectation over the event-wise ambiguity set $\mathcal{F}$ by solving a classical robust optimization problem.
Theorem 1. Assuming the Slater’s condition holds, the worst-case expectation
\[
\sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ r^T (\tilde{s}) G_m (\tilde{s}) \tilde{z} + h_m (\tilde{s}) \right]
\]
is equivalent to the optimal value of the following classical robust optimization problem:
\[
\inf_{\gamma} \gamma \geq \alpha^T p + \sum_{k \in [K]} \beta_k^T \mu_k \quad \forall p \in \mathcal{P}, \quad \sum_{s \in \mathcal{E}_k} p_s \in \mathcal{Q}_k, \quad k \in [K] \\
\alpha_s + \sum_{k \in \mathcal{K}_s} \beta_k^T z \geq r^T (s) G_m (s) z + h_m (s) \quad \forall z \in \mathcal{Z}_s, \quad s \in [S] \\
\gamma \in \mathbb{R}, \quad \alpha \in \mathbb{R}^S, \quad \beta_k \in \mathbb{R}^I_s \quad \forall k \in [K],
\]
where for each \( s \in [S] \), \( \mathcal{K}_s = \{ k \in [K] \mid s \in \mathcal{E}_k \} \).

It is well known that the tractability of the classical robust optimization problem depends on the uncertainty sets \( \mathcal{Z}_s, s \in [S], \mathcal{Q}_k, k \in [k] \) and \( \mathcal{P} \). For practicability, these sets are confined to tractable conic representable sets such as polyhedral or second-order conic representable ones. Hence, by using an algebraic modeling toolbox such as RSOME, the robust optimization problem is automatically transformed to a polynomial sized linear or second-order conic optimization problem, which can be solved by commercial solvers such as CPLEX, Gurobi and MOSEK.

The event-wise ambiguity encompasses a wide spectrum of existing ambiguity sets in its intuitive expression and also inspires new ones based on machine learning techniques.

Uncertain Discrete Distribution
The event-wise ambiguity set can naturally specify uncertain discrete distributions as follows:
\[
\mathcal{F} = \left\{ P \in \mathcal{P}_0 (\mathbb{R}^I_s \times [S]) \mid \begin{array}{l}
(\tilde{z}, \tilde{s}) \sim P \\
\mathbb{P} [ \tilde{z} \in \mathcal{Z}_s \mid \tilde{s} = s ] = 1 \quad \forall s \in [S] \\
\mathbb{P} [ \tilde{s} = s ] = p_s \quad \forall s \in [S] \\
\text{for some } P \in \mathcal{P} \end{array} \right\}
\]
where each \( \mathcal{Z}_s = \{ \tilde{z}_s \} \) is a singleton set. As proposed in Ben-Tal et al. (2013), we can use \( \phi \)-divergence to characterize the uncertainty set \( \mathcal{P} \) of discrete probability distributions. If the uncertainty set \( \mathcal{P} \) is a singleton set, \( \text{i.e.}, P \) is fixed, then the corresponding ambiguity set would also shrinkage to a singleton set containing only the known discrete distribution \( \frac{1}{S} \sum_{s \in [S]} p_s \delta_{\tilde{s}_s} \).

Generalized Moments Ambiguity Set
Given a convex function \( \phi : \mathbb{R}^I_u \mapsto \mathbb{R}^I_v \), the ambiguity set based on generalized moments
\[
\mathcal{G} = \left\{ P \in \mathcal{P}_0 (\mathbb{R}^I_u) \mid \begin{array}{l}
\tilde{u} \sim P \\
\mathbb{E}_P [ \tilde{u} ] \in \mathcal{Q} \\
\mathbb{E}_P [ \phi (\tilde{u}) ] \leq \sigma \\\n\mathbb{P} [ \tilde{u} \in \mathcal{U} ] = 1 \end{array} \right\}
\]
can be mapped into the following event-wise ambiguity set with only one scenario, i.e., $S = 1$.

$$
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{l_u + l_v} \times \{1\}) \right\},
$$

where $\mathcal{Z} = \{(\hat{u}, \bar{v}) \mid u \in \mathcal{U}, v \geq \phi(u)\}$. Based on the lifting and projection theorem by Wiesemann et al. (2014) (see Theorem 5 therein), we have $\Pi_\mathcal{F} = \mathcal{G}$. The generalized moments via the convex function $\phi$ can provide interesting and useful statistical characterizations of the uncertainty $\hat{u}$, including (co)-variance, absolute deviation, semi-variance, and expected utility, among others.

**Wasserstein Ambiguity Set**

We consider a data-driven setting as in Mohajerin Esfahani and Kuhn (2018) on the design of a Wasserstein ambiguity set centered around the empirical distribution $\hat{\mathbb{P}}$. Given a tractable distance metric $\rho : \mathbb{R}^{l_u} \times \mathbb{R}^{l_u} \mapsto [0, +\infty)$, the Wasserstein metric (a.k.a Kantorovich-Rubinstein metric) between any two distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$ is defined via an optimization problem:

$$
d_W(\mathbb{P}, \hat{\mathbb{P}}) \triangleq \inf_{Q \in \mathcal{Q}(\mathbb{P}, \hat{\mathbb{P}})} \mathbb{E}_Q [\rho(\hat{u}, \bar{u})],
$$

where $\hat{u} \sim \mathbb{P}$, $\bar{u} \sim \hat{\mathbb{P}}$, and $\mathcal{Q}(\mathbb{P}, \hat{\mathbb{P}})$ is the set of all joint probability distributions on $\mathbb{R}^{l_u} \times \mathbb{R}^{l_u}$ with marginals $\mathbb{P}$ and $\hat{\mathbb{P}}$. The Wasserstein ambiguity set is then defined by

$$
\mathcal{G}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{U}) \mid \hat{u} \sim \mathbb{P}, d_W(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta \right\},
$$

which is a ball of radius $\theta \geq 0$ around $\hat{\mathbb{P}}$. Interestingly, we can provide a new lifted representation of the Wasserstein ambiguity set in the format of an event-wise ambiguity set.

$$
\mathcal{F}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{l_u + 1} \times [S]) \mid \left( (\hat{u}, \bar{v}), \bar{s} \right) \sim \mathbb{P}, \mathbb{E}_P [\bar{v} \mid \bar{s} = 1] \leq \theta, \mathbb{P} [(\hat{u}, \bar{v}) \in Z_s \mid \bar{s} = 1] = 1 \forall s \in [S] \right\},
$$

where the primary random variable $\hat{u}$ and the auxiliary random variable $\bar{v}$ jointly reside in lifted support sets $Z_s = \{(u, v) \mid u \in \mathcal{U}, v \geq \rho(u, \hat{u})\}, s \in [S]$ for different scenarios.

**Theorem 2.** The Wasserstein ambiguity set $\mathcal{G}_W(\theta)$ is equivalent to the marginal distribution of $\hat{u}$ under $\mathbb{P}$, for all $\mathbb{P} \in \mathcal{F}_W(\theta)$. That is, for all $\theta \geq 0$, $\mathcal{G}_W(\theta) = \Pi_\mathcal{F} \mathcal{F}_W(\theta)$. 
In Appendix C, we extend the result to type-$p$ Wasserstein metric, $p \in [1, \infty]$, where the distance metric between two distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$ is given by

$$d_p^W(\mathbb{P}, \hat{\mathbb{P}}) \triangleq \begin{cases} \inf_{Q \in Q(\mathbb{P}, \hat{\mathbb{P}})} \left( \mathbb{E}_Q \left[ \rho^p(\hat{u}, \hat{u}^\dagger) \right] \right)^{\frac{1}{p}} & \text{if } p \in [1, \infty) \\ \inf_{Q \in Q(\mathbb{P}, \hat{\mathbb{P}})} \text{Q-ess sup}_{u \in \mathcal{U}} \rho(\hat{u}, \hat{u}^\dagger) & \text{if } p = \infty. \end{cases}$$

Here the essential supremum of the joint distribution $Q$ is defined as

$$\text{Q-ess sup}_{u \in \mathcal{U}} \rho(\hat{u}, \hat{u}^\dagger) = \inf \{ M : \mathbb{Q} \left[ \rho(\hat{u} - \hat{u}^\dagger) > M \right] = 0 \}.$$

### Mixture Distribution Ambiguity Set

We can use the event-wise ambiguity set to specify a mixture distribution as proposed in Hanasusanto et al. (2015), which is useful, for example, in modeling ambiguous multi-modal distributions.

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0 \left( \mathbb{R}^{l_u + l_v} \times [S] \right) \middle| \begin{array}{l} ((\hat{u}, \hat{v}), \tilde{s}) \sim \mathbb{P} \\ \mathbb{E}_\mathbb{P} [\hat{u} | \tilde{s} = s] \in Q_s, \forall s \in [S] \\ \mathbb{E}_\mathbb{P} [\hat{v} | \tilde{s} = s] \leq \sigma_s, \forall s \in [S] \\ \mathbb{P} [(\hat{u}, \hat{v}) \in Z_s | \tilde{s} = s] = 1, \forall s \in [S] \\ \mathbb{P} [\tilde{s} = s] = p_s, \forall s \in [S] \end{array} \right\}, \quad (8)$$

where for each $s \in [S]$, $Z_s = \{ (\hat{u}, \hat{v}) | u \in \mathcal{U}_s, v \geq \phi(u) \}$. Note that any distribution $\mathbb{P} \in \mathcal{F}$ can be written as $\mathbb{P} = \sum_{s \in [S]} p_s \mathbb{P}_s$, where each mixture component $\mathbb{P}_s$ is an ambiguous distribution with support $\mathcal{U}_s$ and moments $\mathbb{E}_{\mathbb{P}_s} [\hat{u}] \in Q_s$ and $\mathbb{E}_{\mathbb{P}_s} [\phi(\hat{u})] \leq \sigma_s$. Hanasusanto et al. (2015) have used the mixture distribution ambiguity set to model the uncertain demand in the textile apparel industry, which is known for the multimodality and ambiguity.

### K-means Ambiguity Set

We can incorporate clustering techniques in machine learning to construct event-wise ambiguity sets directly from data. Given $N$ historical observations $\hat{u}_1, \ldots, \hat{u}_N$ for the primary random variable $\hat{u}$, we can partition the support set $\mathcal{U}$ into $S$ clusters $\mathcal{U}_s, s \in [S]$ using the K-means clustering (MacQueen et al. 1967, Dubes and Jain 1988), which gives centroids $\hat{\mu}_s, s \in [S]$ of these clusters (see an illustration in Figure 3). Associated with each cluster, we can determine its support set by

$$\mathcal{U}_s = \{ u \in \mathcal{U} | 2u^\top (\hat{\mu}_r - \hat{\mu}_s) \leq \hat{\mu}_s^\top \hat{\mu}_r - \hat{\mu}_s^\top \hat{\mu}_s, \forall r \in [S] \},$$

where the hyperplane $2u^\top (\hat{\mu}_r - \hat{\mu}_s) = \hat{\mu}_r^\top \hat{\mu}_r - \hat{\mu}_s^\top \hat{\mu}_s$ corresponds to the perpendicular bisector of $\hat{u}_s$ and $\hat{u}_r$. Let $\mathbb{I}$ denote the indicator function. The weight of a cluster is

$$\hat{p}_s = \frac{1}{N} \sum_{n \in [N]} \mathbb{I}[\hat{u}_n \in \mathcal{U}_s]$$
Figure 3  K-means clustering. The distance from a point to its corresponding centroid is not larger than the distance from it to any other centroid. The boundary of a cluster is determined by the boundary of the support set (solid lines) and the perpendicular bisectors of its centroid and centroids of other clusters (dash lines).

and the parameters associated with the convex generalized moments are

\[ \hat{\sigma}_s = \frac{1}{\hat{p}_s N} \sum_{n \in [N]} I[\hat{u}_n \in \mathcal{U}_s] \phi(\hat{u}_n). \]

The corresponding K-means ambiguity set is a special mixture distribution ambiguity set in the form (8) with cluster-wise estimates $Q_s = \{\hat{\mu}_s\}, p_s = \hat{p}_s, \sigma_s = \hat{\sigma}_s, s \in [S]$. To account for uncertainty in these estimates, we can further specify uncertainty sets for them.

In Appendix F, we present a three-period portfolio management problem, where we construct a two-layer K-means ambiguity set directly from the historical returns. Note that the main issue with using stochastic programming approach to address the same problem is how to obtain the scenario tree directly from data. While we can observe the sample paths from the historical returns, since the returns are continuous random variables, unfolding the underlying scenario tree would require us to impose additional assumptions on the underlying stochastic process of the returns. We also refer interested readers to a recent work by Perakis et al. (2018) of using real data to construct the K-means ambiguity set to address a joint pricing and production problem.

**Ambiguity Set with Side Information**

Suppose that the uncertainty of the primary random variable $\tilde{u}$ is strongly associated with side information or covariate, which we represent by an auxiliary random variable $\tilde{v}$, and collectively, we have the data $(\tilde{u}_1, \tilde{v}_1), \ldots, (\tilde{u}_N, \tilde{v}_N)$. We can incorporate such side information in the ambiguity set where we group the primary random variable $\tilde{u}$ based on the side information $\tilde{v}$. In particular, a broad range of machine learning techniques, including classification and regression trees, can separate $\tilde{v}_i, i \in [N]$ into $S$ scenarios. Correspondingly, we would partition $\tilde{u}_i, i \in [N]$ into $S$ scenarios,
Figure 4  Realizations $\hat{v}_n \in \mathbb{R}^2, n \in [N]$ of the auxiliary random variable are separated into 2 groups (black and white). Correspondingly, realizations $\hat{u}_n \in \mathbb{R}^3, n \in [N]$ of the primary random variable are grouped into 2 scenarios (black and white). The empirical information of these two groups can be different.

where we can specify the empirical information of each scenario in the event-wise ambiguity set (see the example in Figure 4). To determine the demands of taxis, Hao et al. (2019) use weather information as the side information and apply multivariate regression tree to obtain the scenarios. By doing so, they achieve significant improvement in taxis allocation under demand uncertainty.

5. Modeling Examples
The RSO framework is expansive and encompasses scenario-tree based stochastic linear optimization and distributionally robust optimization models. Although it is based on expectations of bi-affine functions, it can also provide a tight characterization of the worst-case expectations of some classes of quadratic functions known in the literature, including the seminal works of Ben-Tal and Nemirovski (1998) and Tütüncü and Koenig (2004), and extend them to include discrete scenarios (Appendix B). We next provide several examples in our framework, including optimizing over the Hurwicz criterion and models which have both discrete and continuous recourse decisions and where the uncertainty is characterized using the Wasserstein ambiguity set as well as the K-means ambiguity set (Appendix F)—both are directly constructed from data. We show that an algebraic modeling toolbox such as RSOME could greatly facilitate the implementation without worrying about the tedious reformulation.

Hurwicz Criterion
Hurwicz (1951) is arguably first to propose a decision criterion that articulates the tradeoff between pessimistic and optimistic objectives, which under distributional ambiguity can be formulated as

$$
(1 - \varphi) \sup_{P \in \mathcal{F}} \mathbb{E}_P [f(w, \hat{u}, \hat{s})] + \varphi \inf_{P \in \mathcal{F}} \mathbb{E}_P [f(w, \hat{u}, \hat{s})],
$$

where $\varphi$ is the Hurwicz parameter. The empirical implementation of this criterion is straightforward: the uncertainty region is defined by a set of empirical scenarios, and the criterion is then applied to the empirical data.
where the cost function \( f(w, u, s) \) depends on the here-and-now decision \( w \), and it is typically convex in \( u \) for given \( w \in X \) and scenario \( s \in [S] \). Here \( \varphi \in [0, 1] \) is the level of optimism, with \( \varphi = 0 \) (\( \varphi = 1 \)) being the most pessimistic (optimistic) perception of the objective value. In order to obtain a computationally tractable model, we often consider the most pessimistic objective (i.e., \( \varphi = 0 \)) because the best-case expectation for the most optimistic objective (i.e., \( \varphi = 1 \)) is typically non-convex in its decision \( w \). Quite notably, there is a class of ambiguity sets for which the best-case expectation would also be tractable.

**Proposition 1.** Consider an event-wise ambiguity set \( F \) in (3) such that for any \( P \in F \), it satisfies \( \mathbb{E}_P[\tilde{u} | \tilde{s} = s] = \mu_s \) and \( P[\tilde{s} = s] = p_s \) with known \( \mu_s \) and \( p_s \) for all \( s \in [S] \). Then for any function \( g(u, s): \mathbb{R}^{I_u} \times [S] \mapsto \mathbb{R} \) that is convex in \( u \) for a given \( s \in [S] \), we have

\[
\inf_{P \in F} \mathbb{E}_P[g(\tilde{u}, \tilde{s})] = \sum_{s \in [S]} p_s g(\mu_s, s).
\]

**Proof.** The proposition follows immediately from Jensen's inequality. \( \square \)

The mixture distribution ambiguity set with singleton sets \( Q_s, s \in [S] \) and the K-means ambiguity set fit in this class, for which we can optimize over the Hurwicz criterion

\[
\min_{w \in X} \left\{ (1 - \varphi) \sup_{P \in F} \mathbb{E}_P[f(w, \tilde{u}, \tilde{s})] + \varphi \sum_{s \in [S]} p_s f(w, \mu_s, s) \right\}
\]

by formulating via the RSO framework. Note that it would also be possible to account for scenarios with uncertain probabilities vector \( p = (p_s)_{s \in [S]} \subseteq \mathcal{P} \), as long as the uncertainty set \( \mathcal{P} \) is a polytope with modest number of extreme points, \( p^e, e \in [E] \). In such circumstances, we would solve

\[
\min_{e \in [E]} \min_{w \in X} \left\{ (1 - \varphi) \sup_{P \in F} \mathbb{E}_P[f(w, \tilde{u}, \tilde{s})] + \varphi \sum_{s \in [S]} p^e_s f(w, \mu_s, s) \right\}.
\]

**Expectation of Convex and Piecewise Affine Functions**

Expectation of convex and piecewise affine functions are commonly encountered in modeling risk aversion based on the utility (Gilboa and Schmeidler 1989) or risk measure including the shortfall risk measure (Föllmer and Schied 2002) and the optimized certainty equivalent (Ben-Tal and Teboulle 2007). We show that by simply introducing a recourse decision \( y(\cdot, \cdot) \), we can achieve an equivalent formulation under the RSO framework.

**Theorem 3.** The worst-case expectation

\[
\sup_{\mathcal{P} \in \mathcal{F}} \mathbb{E}_P \left[ \max_{\ell \in [L]} \left\{ r^\top(\tilde{s}) G(\tilde{s}) \tilde{z} + h(\tilde{s}) \right\} \right]
\]
for a finite index set \([L]\), is equivalent to the following problem

\[
\begin{align*}
\min & \quad \sup_{P \in \mathcal{F}} \mathbb{E}_P [y(\hat{s}, \tilde{z})] \\
\text{s.t.} & \quad y(s, z) \geq r^\top(s)G(s)z + h(s) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \ell \in [L] \\
& \quad y \in \mathcal{A}(\mathcal{C}, [I_z]),
\end{align*}
\]

where the collection \(\mathcal{C} \triangleq \{s\} | s \in [S]\) consists of singleton MECE events.

### Expected Utility with Mean-Covariance Ambiguity Sets

Consider the mean-covariance ambiguity set that commonly appears in portfolio management

\[
\mathcal{G}(\mu, \Sigma) = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_u} \right) \mid \begin{array}{l}
\hat{u} \sim P \\
\mathbb{E}_P[\hat{u}] = \mu \\
\mathbb{E}_P[(\hat{u} - \mu)(\hat{u} - \mu)^\top] = \Sigma
\end{array} \right\}.
\]

Here the \(I_u\)-dimensional random variable \(\hat{u}\) can be a scalar or a vector and refers to the random return(s) of the risky asset(s). For any utility function \(U : \mathbb{R} \mapsto \mathbb{R}\), Popescu (2007) has shown that the robust expected utility of the random weighted sum \(w^\top \hat{u}\) satisfies:

\[
\inf_{P \in \mathcal{G}(\mu, \Sigma)} \mathbb{E}_P[U(w^\top \hat{u})] = \inf_{P \in \mathcal{G}(w^\top \mu, w^\top \Sigma w)} \mathbb{E}_P[U(\tilde{u})].
\]

This property enables Natarajan et al. (2010) to obtain an attractive computationally tractable second-order cone reformulation when \(U\) is concave piecewise affine, which is surprising as a direct duality approach would result in a positive semidefinite program that is much harder to solve. We next recover this result using the RSO framework to obtain the second-order cone reformulation.

**Theorem 4.** Given a concave piecewise affine utility function \(U(u) = \min_{\ell \in [L]} \{g_\ell u + h_\ell\}\), the robust expected utility \(\inf_{P \in \mathcal{G}(\mu, \Sigma)} \mathbb{E}_P[U(w^\top \hat{u})]\) is equivalent to

\[
\begin{align*}
\max & \quad \inf_{P \in \mathcal{F}} \mathbb{E}_P [y(\hat{u}, \tilde{v})] \\
\text{s.t.} & \quad y(u, v) \leq g_\ell (ru + w^\top \mu) + h_\ell \quad \forall (u, v) \in \mathcal{Z}, \ \ell \in [L] \\
& \quad r \geq \sqrt{w^\top \Sigma w} \\
& \quad r \in \mathbb{R}, \ y \in \mathcal{A}(\{1\}, \{1, 2\}),
\end{align*}
\]

where the ambiguity set

\[
\mathcal{F} = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^2 \right) \mid \begin{array}{l}
(\hat{u}, \tilde{v}) \sim P \\
\mathbb{E}_P[\hat{u}] = 0, \ \mathbb{E}_P[\tilde{v}] \leq 1 \\
\mathbb{P}[(\hat{u}, \tilde{v}) \in \mathcal{Z}] = 1
\end{array} \right\}
\]

has only one scenario (i.e., \(S = 1\)) and takes a support set \(\mathcal{Z} = \{(u, v) \in \mathbb{R}^2 \mid v \geq u^2\}\).
Expectation of Saddle Functions

The RSO model is primary based on a linear optimization framework, where the objective function and soft constraints are bilinear with respect to the underlying decision variable \( r(s) \) and the random variable \( \tilde{z} \). With auxiliary decisions and auxiliary random variables, we can also consider saddle functions that are convex with respect to the decision variables and concave with respect to the random variables (see, Ben-Tal et al. 2015). Observe that unlike earlier robust and distributionally robust optimization models, the random variable mappings in (1) include affine relations involving the auxiliary random variable, \( \tilde{v} \), which is embedded in \( \tilde{z} \). This generality allows us to extend the objective function and soft constraints to saddle functions under the RSO framework.

We consider a saddle function \( f(r(s), u, s) \) such that for a given scenario \( s \), it is jointly convex with respect to the decision \( r(s) \in X_s \) for any fixed \( u \in U_u \) and jointly concave with respect to \( u \in U_u \) for any fixed \( r(s) \in X_s \) as follows:

\[
f(r(s), u, s) \triangleq \xi^T(r(s), s) \zeta(u, s) = \sum_{\ell \in [L]} \xi_{\ell}(r(s), s) \zeta_{\ell}(u, s).
\]

Here for a given scenario \( s \) and the corresponding partition of indices \( [I_v] = L_{s1} \cup L_{s2} \cup L_{s3} \cup L_{s4} \):

- \( \xi_{\ell}(r(s), s) \) is nonnegative and convex in \( r(s) \) and \( \zeta_{\ell}(u, s) \) is nonnegative and concave in \( u \), for \( \ell \in L_{s1} \);
- \( \xi_{\ell}(r(s), s) \) is nonnegative and affine in \( r(s) \) and \( \zeta_{\ell}(u, s) \) is concave in \( u \), for \( \ell \in L_{s2} \);
- \( \xi_{\ell}(r(s), s) \) is convex in \( r(s) \) and \( \zeta_{\ell}(u, s) \) is nonnegative and affine in \( u \), for \( \ell \in L_{s3} \);
- \( \xi_{\ell}(r(s), s) \) is affine in \( r(s) \) and \( \zeta_{\ell}(u, s) \) is affine in \( u \), for \( \ell \in L_{s4} \).

**Theorem 5.** The worst-case expectation \( \sup_{p \in \mathcal{P}} \mathbb{E}_p[f(r(\tilde{s}), \bar{u}, \tilde{s})] \) for the ambiguity set

\[
\mathcal{G} = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_u} \times [S] \right) \left| \begin{array}{l}
(\tilde{u}, \tilde{s}) \sim \mathbb{P} \\
\mathbb{E}_p[\tilde{u} | \tilde{s} \in \mathcal{E}_k] \in \mathcal{Q}_k \quad \forall k \in [K] \\
\mathbb{P}[\tilde{u} \in U_u | \tilde{s} = s] = 1 \quad \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] = p_s \quad \forall s \in [S] \\
\text{for some } p \in \mathcal{P}
\end{array} \right. \right\},
\]

is the same as

\[
\min_{p \in \mathcal{P}} \mathbb{E}_p[\bar{r}^T(\tilde{s})\bar{v}] \\
\text{s.t. } \bar{r}_{\ell}(s) \geq \xi_{\ell}(r(s), s) \quad \forall \ell \in L_{s1} \cup L_{s3}, \ s \in [S] \\
\bar{r}_{\ell}(s) = \xi_{\ell}(r(s), s) \quad \forall \ell \in L_{s2} \cup L_{s4}, \ s \in [S] \\
\bar{r}(s) \in \mathbb{R}^{I_u} \quad \forall s \in [S],
\]

(13)
for the lifted event-wise ambiguity set

\[
\mathcal{F} = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_u+I_v} \times [S] \right) \middle| \begin{array}{l}
(\hat{u}, \hat{v}, \hat{s}) \sim \mathbb{P} \\
\mathbb{E}_P [\hat{u} | \hat{s} \in \mathcal{E}_k] \in \mathcal{Q}_k \quad \forall k \in [K] \\
P \left[ (\hat{u}, \hat{v}) \in \mathcal{Z}_s \mid \hat{s} = s \right] = 1 \quad \forall s \in [S] \\
P \left[ \hat{s} = s \right] = p_s \quad \forall s \in [S]
\end{array} \right. \right\}
\]

(14)

with lifted support sets

\[
\mathcal{Z}_s = \left\{ (u, v) \in \mathbb{R}^{I_u+I_v} \middle| \begin{array}{l}
u \in \mathcal{U}_s \\
v_t \leq \zeta_\ell(u, s) \quad \forall \ell \in \mathcal{L}_{s_1} \cup \mathcal{L}_{s_2} \\
v_t = \zeta_\ell(u, s) \quad \forall \ell \in \mathcal{L}_{s_3} \cup \mathcal{L}_{s_4}
\end{array} \right. \quad \forall s \in [S].
\]

Two-Stage Problem with Wasserstein Ambiguity Sets

Optimization models based on the Wasserstein ambiguity set have recently attracted considerable interests from both stochastic programming and robust optimization communities. While most of the existing models are static, dynamic models with the Wasserstein ambiguity set are scarce due to limited solution approaches. We next demonstrate that the RSO framework provides a tractable approximation for two-stage linear optimization problems with the Wasserstein ambiguity set (6), which has the potential to serve the modeling of multi-stage dynamic problems. We refer to Bertsimas et al. (2018a) for multi-stage linear optimization problem with the type-\(\infty\) Wasserstein ambiguity set which admits an equivalent lifted format of our proposed event-wise ambiguity set.

In particular, we consider the following second-stage problem given the here-and-now decision \(w\) and the realization \(u\) of the underlying primary random variable \(\tilde{u}\).

\[
f(w, u) = \min_{y} c_0^\top y \\
\text{s.t. } a_\ell^\top (u) w + c_\ell^\top y \geq d_\ell(u) \quad \forall \ell \in [L] \\
y \in \mathbb{R}^{J_v},
\]

(15)

where similar to the random variable mappings in (1), for each \(\ell\)th constraint, \(a_\ell\) and \(d_\ell\) are affine mappings of the realization of \(\tilde{u}\). For any here-and-now decision \(w\), we approximate its worst-case expected second-stage cost \(\sup_{p \in \mathcal{P}_0} \mathbb{E}_p[f(w, \tilde{u})]\) under the type-1 Wasserstein ambiguity set through

\[
\min_{\tilde{z}} \sup_{p \in \mathcal{P}_0} \mathbb{E}_p \left[ c_0^\top y(\tilde{s}, \tilde{z}) \right] \\
\text{s.t. } a_\ell^\top (u) w + c_\ell^\top y(s, z) \geq d_\ell(u) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \ell \in [L] \\
y_j \in \tilde{A}(\mathcal{C}_j, \mathcal{I}_j) \quad \forall j \in [J_y],
\]

(16)

where \(\tilde{z} = (\tilde{u}, \tilde{v})\) and where the collections \(\mathcal{C}_j, j \in [J_y]\) of MECE events and information index sets \(\mathcal{I}_j \subseteq [I_u + 1], j \in [J_y]\) jointly control how the recourse decision \(y(\cdot, \cdot)\) adapts to \((\tilde{u}, \tilde{v})\) and \(\tilde{s}\). The
optimal \( w \) can then be selected by minimizing the sum of the deterministic first-stage cost \( c^Tw \) and the worst-case expected second-stage cost (16). In Appendix D, we report the performance of this conservative approximation in comparison with (i) the computationally expensive exact approach and (ii) a state-of-the-art approximation scheme by Hanasusanto and Kuhn (2018).

We can extend the event-wise adaptation to address a two-stage problems with the type-\( \infty \) Wasserstein ambiguity set, which, interestingly, would coincide with the multi-policy approximation (MPA) proposed by Bertsimas et al. (2018b). Generalization of MPA to multi-stage problems has appeared in Bertsimas et al. (2018a), which we can also incorporate in our event-wise adaptation.

6. Conclusion

The RSO model unifies an important class of scenario-tree based stochastic linear optimization problems and a number of distributionally robust optimization models (based on \( \phi \)-divergence, convex generalized moments, Wasserstein metric, etc.) that have been considered in isolation to date. As we have demonstrated, the RSO model also opens up to new approaches including those inspired by machine learning techniques. Algebraic modeling package such as RSOME would help us in navigating and evaluating the plethora of approaches to address a wide variety of optimization problems under uncertainty in practice.

References


Perakis, Georgia, Melvyn Sim, Qinshen Tang, Peng Xiong. 2018. Joint pricing and production: a fusion of machine learning and robust optimization. *Available at SSRN*.


A. Proofs

Proof of Theorem 1. Let \( \mu = (\mu_k)_{k \in [K]} \) and \( Q = \{ \mu \mid \mu_k \in Q_k \ \forall k \in [K] \} \). We can re-express

\[
\lambda^* = \sup_{\mathcal{P} \in \mathcal{P}} \mathbb{E}_\mathcal{P} \left[ r^\top (\tilde{s}) G_m(\tilde{s}) \tilde{z} + h_m(\tilde{s}) \right]
\]

by \( \lambda^* = \sup_{(p,\mu) \in \mathcal{P} \times Q} \lambda(p, \mu) \), where given \((p, \mu) \in \mathcal{P} \times Q\), we define an ambiguity set

\[
\mathcal{F}(p, \mu) = \left\{ \mathcal{P} \in \mathcal{P}_0(\mathbb{R}^{I_x} \times [S]) \left| \begin{array}{l}
(\tilde{z}, \tilde{s}) \sim \mathcal{P} \\
\mathbb{E}_\mathcal{P} [\tilde{z} | \tilde{s} \in \mathcal{E}_k] = \mu_k \ \forall k \in [K] \\
\mathbb{P} [\tilde{z} \in Z_s | \tilde{s} = s] = 1 \ \forall s \in [S] \\
\mathbb{P} [\tilde{s} = s] = p_s \ \forall s \in [S]
\end{array} \right. \right\}
\]

and correspondingly the worst-case expectation

\[
\lambda(p, \mu) = \sup_{\mathcal{P} \in \mathcal{F}(p, \mu)} \mathbb{E}_\mathcal{P} \left[ r^\top (\tilde{s}) G_m(\tilde{s}) \tilde{z} + h_m(\tilde{s}) \right].
\]

Using the law of total probability, we can construct the joint distribution \( \mathbb{P} \) of \((\tilde{z}, \tilde{s})\) from the marginal distribution \( \mathbb{P} \tilde{s} \) supported on \([S]\) and the conditional distributions \( \mathbb{P}_s \) of \( \tilde{z} \) given \( \tilde{s} = s, s \in [S] \). In this way, we can reformulate \( \lambda(p, \mu) \) as

\[
\lambda(p, \mu) = \sup \sum_{s \in [S]} p_s \mathbb{E}_{\mathbb{P}_s} \left[ r^\top (\tilde{s}) G_m(\tilde{s}) \tilde{z} + h_m(\tilde{s}) \right]
\]

s.t. \( \sum_{s \in \mathcal{E}_k} p_s \mathbb{E}_{\mathbb{P}_s} [\tilde{z}] = q_k \mu_k \ \forall k \in [K] \)

\( \mathbb{P}_s [\tilde{z} \in Z_s] = 1 \ \forall s \in [S] \)

with \( q_k = \sum_{s \in \mathcal{E}_k} p_s, k \in [K] \). We can express the dual of \( \lambda(p, \mu) \) as

\[
\lambda_1(p, \mu) = \inf \sum_{s \in [S]} q_s \alpha_s + \sum_{k \in [K]} q_k \beta_k^\top \mu_k
\]

s.t. \( \alpha_s + p_s \sum_{k \in \mathcal{K}_s} \beta_k^\top z \geq p_s (r^\top (s) G_m(s) z + h_m(s)) \ \forall z \in Z_s, s \in [S] \)

\( \alpha \in \mathbb{R}^S, \beta_k \in \mathbb{R}^{I_z} \ \forall k \in [K] \)

\[
= \inf \alpha^\top p + \sum_{k \in [K]} q_k \beta_k^\top \mu_k
\]

s.t. \( \alpha_s + \sum_{k \in \mathcal{K}_s} \beta_k^\top z \geq r^\top (s) G_m(s) z + h_m(s) \ \forall z \in Z_s, s \in [S] \)

\( \alpha \in \mathbb{R}^S, \beta_k \in \mathbb{R}^{I_z} \ \forall k \in [K] \),

where the second equality follows from for all \( s \in [S] \), first changing variable from \( \alpha_s \) to \( p_s \alpha_s \) and then dividing both sides of the constraint by \( p_s \), which is allowed since \( p \in \mathcal{P} \) is strictly positive.

By weak duality, \( \lambda(p, \mu) \leq \lambda_1(p, \mu) \). By the general min-max theorem, we further observe that

\[
\lambda_1 = \sup_{(p, \mu) \in \mathcal{P} \times Q} \lambda_1(p, \mu) \leq \lambda^*_2,
\]
where
\[ \lambda_2^* = \inf \gamma \]
\[
\text{s.t. } \gamma \geq \alpha^T p + \sum_{k \in [K]} q_k \beta_k^T \mu_k \quad \forall p \in \mathcal{P}, \mu_k \in \mathcal{Q}_k, \ k \in [K] \]
\[
\alpha_s + \sum_{k \in K_s} \beta_k^T z \geq r^T(s)G_m(s)z + h_m(s) \quad \forall z \in \mathcal{Z}_s, \ s \in [S] \]
\[
\gamma \in \mathbb{R}, \ \alpha \in \mathbb{R}^S, \ \beta_k \in \mathbb{R}^{I_z} \quad \forall k \in [K]. \]

Due to the presence of products of uncertain variables (e.g., \( q_k \mu_k \)), problem (17) is nonconvex. Since \( p > 0 \) (and hence \( q_k > 0 \)), an equivalent convex representation can be obtained by changing variables in problem (17) from \( q_k \mu_k \) to \( \mu_k \) for all \( k \in [K] \), which turns out to be problem (4).

Assuming the conic representation of the following system
\[
\begin{align*}
\frac{\sum_{s \in E_k} \xi_s}{\sum_{s \in E_k} \tau_s} & \in \mathcal{Q}_k \quad \forall k \in [K] \\
\frac{\xi_s}{\tau_s} & \in \mathcal{Z}_s \quad \forall s \in [S] \\
\tau & \in \mathcal{P}
\end{align*}
\]

satisfies the Slater’s condition (see Theorem 1.4.2 in Ben-Tal and Nemirovski 2001), one can establish strong duality, i.e., \( \lambda^* = \lambda_1^* = \lambda_2^* \) and show that problem (4) is solvable (see Theorem 1 in Bertsimas et al. 2019).

**Proof of Theorem 2.** We consider an ambiguity set without the auxiliary random variable \( \tilde{v} \)
\[
\tilde{G}_W(\theta) = \left\{ p \in \mathcal{P}_0(\mathbb{R}^{I_u} \times [S]) \mid \begin{align*}
(\tilde{u}, \tilde{s}) & \sim \mathbb{P} \\
\mathbb{E}_p[\rho(\tilde{u}, \tilde{s}) | \tilde{s} \in [S]] & \leq \theta \\
\mathbb{P}[\tilde{u} = \mathcal{U} | \tilde{s} = s] & = 1 \quad \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] & = \frac{1}{S} \quad \forall s \in [S]
\end{align*} \right\}. \tag{19}
\]

Since this ambiguity set satisfies \( \Pi_{(\tilde{u}, \tilde{s})} F_W(\theta) = \tilde{G}_W(\theta) \) for all \( \theta \geq 0 \), thus it is sufficient to prove \( \Pi_a \tilde{G}_W(\theta) = G_W(\theta) \) for all \( \theta \geq 0 \).

To this end, we first prove \( G_W(\theta) \subseteq \Pi_a \tilde{G}_W(\theta) \). Consider \( \tilde{u} \sim \mathbb{P} \) for some \( \mathbb{P} \in \tilde{G}_W(\theta) \). By definition of the Wasserstein ambiguity set \( G_W(\theta) \), there exists a joint distribution \( \mathbb{Q} \in \mathcal{P}(\mathbb{P}, \mathbb{P}) \) of \( (\tilde{u}, \tilde{u}^l) \) such that \( \Pi_a \mathbb{Q} = \mathbb{P} \), \( \Pi_a^! \mathbb{Q} = \mathbb{P} \), and \( \mathbb{E}_\mathbb{Q}[\rho(\tilde{u}, \tilde{u}^l)] \leq \theta \). Since we can construct \( \mathbb{Q} \) from the marginal distribution \( \tilde{\mathbb{P}} \) of \( \tilde{u}^l \) supported on \( \{\tilde{u}_1, \ldots, \tilde{u}_S\} \) and the conditional distributions \( \mathbb{P}_s \) of \( \tilde{u} \), given the realization of \( \tilde{u}^l \) is \( \tilde{u}_s \), \( s \in [S] \), we have \( (\tilde{u}, \tilde{u}^l) \sim \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s \otimes \delta_{\tilde{u}_s} \). We can then construct a distribution \( \mathbb{Q}' \in \mathcal{P}_0(\mathbb{R}^{I_u} \times [S]) \) for the random variable \( (\tilde{u}, \tilde{s}) \sim \mathbb{Q}' \) via \( \mathbb{Q}' = \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s \otimes \delta_{\tilde{s}} \).

Observe that \( \mathbb{Q}' \in \tilde{G}_W(\theta) \), hence \( G_W(\theta) \subseteq \Pi_a \tilde{G}_W(\theta) \).
To prove $\Pi_a \tilde{G}_W(\theta) \subseteq G_W(\theta)$, we fix any $\mathbb{P} \in \tilde{G}_W(\theta)$ and we write its projection over $\tilde{u}$ as $\Pi_a \mathbb{P} = \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s$, where $\mathbb{P}_s$ is the conditional distribution of $\tilde{z}$ given the outcome of the random scenario $s$. We can then construct a joint distribution $\mathbb{Q} = \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s \otimes \delta_{s\tilde{u}}$ of $(\tilde{u}, \tilde{u}^\dagger)$ that satisfies

$$
\mathbb{E}_{\mathbb{Q}} [\rho(\tilde{u}, \tilde{u}^\dagger)] = \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{\mathbb{P}_s} [\rho(\tilde{u}, \tilde{u}_s)] = \mathbb{E}_{\mathbb{P}} [\rho(\tilde{u}, \tilde{u}_s) | s \in [S]] \leq \theta.
$$

Hence, $\Pi_a \mathbb{P} \in G_W(\theta)$, which gives $\Pi_a \tilde{G}_W(\theta) \subseteq G_W(\theta)$ to conclude $G_W(\theta) = \Pi_a \tilde{G}_W(\theta)$.

Proof of Theorem 3. Previous derivations in the proof of Theorem 1 implies that (i) the worst-case expectation (9) is equivalent to the following problem

$$
\inf \gamma \\
\text{s.t. } \gamma \geq \alpha^\top p + \sum_{k \in [K]} \beta_k^\top \mu_k \quad \forall p \in \mathcal{P}, \frac{\mu_k}{\sum_{s \in \mathcal{S}_k} p_s} \in \mathcal{Q}_k, \ k \in [K]
$$

$$
\alpha_s + \sum_{k \in \mathcal{S}_k} \beta_k^\top z \geq r^\top(s)G(s)z + h(s) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \ell \in [L]
$$

$$
\gamma \in \mathbb{R}, \alpha \in \mathbb{R}^S, \beta_k \in \mathbb{R}^L \quad \forall k \in [K];
$$

and (ii) problem (10) is equivalent to

$$
\inf \gamma \\
\text{s.t. } \gamma \geq \alpha^\top p + \sum_{k \in [K]} \beta_k^\top \mu_k \quad \forall p \in \mathcal{P}, \frac{\mu_k}{\sum_{s \in \mathcal{S}_k} p_s} \in \mathcal{Q}_k, \ k \in [K]
$$

$$
\alpha_s + \sum_{k \in \mathcal{S}_k} \beta_k^\top z \geq y(s, z) \quad \forall z \in \mathcal{Z}_s, \ s \in [S]
$$

$$
y(s, z) \geq r^\top(s)G(s)z + h(s) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \ell \in [L]
$$

$$
y \in \tilde{A}(\tilde{C}, [I_s])
$$

$$
\gamma \in \mathbb{R}, \alpha \in \mathbb{R}^S, \beta_k \in \mathbb{R}^L \quad \forall k \in [K].
$$

It is then sufficient to construct a feasible solution to problem (21) from a feasible solution to problem (20) such that the constructive solution yields the same objective. Indeed, given a feasible solution $(\gamma^\dagger, \alpha^\dagger, (\beta_k^\dagger)_{k \in [K]})$ to problem (20), we can construct such a desired solution via:

$$
\gamma = \gamma^\dagger, \ \alpha = \alpha^\dagger, \ \beta_k = \beta_k^\dagger \quad \forall k \in [K], \ \gamma(s, z) = \alpha_s^\dagger + \sum_{k \in \mathcal{S}_k} (\beta_k^\dagger)^\top z \quad \forall s \in [S],
$$

for which the recourse decision $y(\cdot, \cdot) \in \tilde{A}(\tilde{C}, [I_s])$.

Proof of Theorem 4. Using the result of Popescu (2007), we first show that

$$
\inf_{\mathbb{P} \in \tilde{G}(w, \mu, \Sigma w)} \mathbb{E}_{\mathbb{P}} [U(\tilde{u})] = \sup_r \left\{ \inf_{\mathbb{P} \in \mathbb{P}} \mathbb{E}_{\mathbb{P}} [U(r\tilde{u} + \mu)] \right\}.
$$

By duality, we have

$$
\inf_{\mathbb{P} \in \tilde{G}(w, \mu, \Sigma w)} \mathbb{E}_{\mathbb{P}} [U(\tilde{u})] = \inf_{\mathbb{P} \in \tilde{G}(0, \mu, \Sigma w)} \mathbb{E}_{\mathbb{P}} [U(\tilde{u} + w^\top \mu)]
$$

$$
= \sup_{\alpha, \beta_1, \beta_2} \left\{ \alpha + w^\top \Sigma w \cdot \beta_2 \mid \alpha + \beta_1 u + \beta_2 u^2 \leq U(u + w^\top \mu) \ \forall u \right\}.
$$
Note that it requires $\beta_2 \leq 0$ for the above problem to be feasible, as otherwise the constraint would be violated for some sufficiently large $u$. Hence, we can further rewrite this problem into

$$
\sup_{\alpha, \beta_1, \beta_2, r} \left\{ \alpha + r^2 \beta_2 \left| \begin{array}{c}
\alpha + \beta_1 u + \beta_2 u^2 \leq U(u + w^T \mu) \forall u \\
r \geq \sqrt{w^T \Sigma w}
\end{array} \right. \right\} = \sup_r \left\{ \inf_{P \in G(0, r^2)} E_P[U(\tilde{u} + w^T \mu)] \left| r \geq \sqrt{w^T \Sigma w} \right. \right\} \\
= \sup_r \left\{ \inf_{P \in G(0, 1)} E_P[U(r\tilde{u} + w^T \mu)] \left| r \geq \sqrt{w^T \Sigma w} \right. \right\} \\
= \sup_{\alpha, \beta_1, \beta_2, r} \left\{ \alpha + \beta_2 \left| \begin{array}{c}
\alpha + \beta_1 u + \beta_2 u^2 \leq U(ru + w^T \mu) \forall u \\
r \geq \sqrt{w^T \Sigma w}
\end{array} \right. \right\} \\
= \sup_{\alpha, \beta_1, \beta_2, r} \left\{ \alpha + \beta_2 \left| \begin{array}{c}
\alpha + \beta_1 u + \beta_2 u^2 \leq U(ru + w^T \mu) \forall (u, v) : v \geq u^2 \\
\beta_2 \leq 0, r \geq \sqrt{w^T \Sigma w}
\end{array} \right. \right\} \\
= \sup_r \left\{ \inf_{P \in F} E_P[U(r\tilde{u} + w^T \mu)] \left| r \geq \sqrt{w^T \Sigma w} \right. \right\}.
$$

The result then follows by applying Theorem 3. \qed

**Proof of Theorem 5.** Observe that for any feasible recourse decision $\bar{r}(\cdot)$ to problem (13), we have

$$
\sup_{P \in F} E_P[\bar{r}^T(\tilde{s}) \bar{v}] = \sup_{P \in G} E_P[\bar{r}^T(\tilde{s}) \zeta(\tilde{u}, \tilde{s})].
$$

In addition, the optimal $\bar{r}^*(\cdot)$ to problem (13) satisfies $\bar{r}^*_\ell(s) = \xi_\ell(r(s), s)$ for all $\ell \in [I_v]$ and $s \in [S]$. Therefore, our claim holds. \qed

### B. Worst-Case Expectation of Quadratic Functions

The RSO framework can be used to provide a tight characterization of the worst-case expectation of some quadratic functions that are known in the literature and extend them to include discrete scenarios. Let $S'$ be the space of symmetric matrices in $\mathbb{R}^{I \times I}$. Given $X, Y \in S'$, we denote by $X \succeq Y$ (resp., $X \succ Y$) to represent $X - Y$ is positive semidefinite (resp., definite), and denote by $X \cdot Y$ as the trace inner product of $X, Y$. Special matrices and vectors of the appropriate dimension include $O, I$, and $0$, which respectively correspond to the zero matrix, the identity matrix, and the zero vector.

#### Bi-Convex-Quadratic Function

We explore the following bi-convex-quadratic function as an extension of Ben-Tal and Nemirovski (1998) to include discrete scenarios:

$$
g(r(s), u, s) \triangleq u^T A^T(r(s), s) A(r(s), s) u + 2u^T b(r(s), s) + c(r(s), s),
$$
where given a scenario $s$, $A(r(s), s), b(r(s), s), c(r(s), s)$ are affine mappings of $r(s)$. The event-wise ambiguity set is given by

$$
\mathcal{G} = \left\{ P \in P_0(\mathbb{R}^{I_u} \times [S]) \mid \begin{array}{l}
(\tilde{u}, \tilde{s}) \sim P \\
\mathbb{E}_P[(\frac{1}{u}) (\frac{1}{u})^\top | \tilde{s} \in \mathcal{E}_k] \in Q_k \quad \forall k \in [K] \\
P[\tilde{s} = s] = p_s \quad \forall s \in [S]
\end{array} \right\}.
$$

The support set $\mathcal{U}_s$ is general enough to capture the ubiquitous uncertainty set $\{u \mid u^\top \Lambda_s u \leq 1\}$ parameterized by some $\Lambda_s \succ 0$, for which we only need to define

$$
\mathcal{U}_s = \left\{ U \in \mathbb{S}^{I_u+1} \mid U \cdot \begin{pmatrix}
-1 & 0^\top \\
0 & \Lambda_s
\end{pmatrix} \leq 0 \right\}. \quad (22)
$$

**Theorem 6.** The worst-case expectation

$$
\sup_{P \in \mathcal{G}} \mathbb{E}_P[g(r(\tilde{s}), \tilde{u}, \tilde{s})] \quad (23)
$$

is bounded from above by

$$
\min \sup_{P \in \mathcal{F}} \mathbb{E}_P[R(\tilde{s}) \cdot \tilde{Z}] \quad (24)
$$

where the lifted event-wise ambiguity set

$$
\mathcal{F} = \left\{ P \in P_0(\mathbb{S}^{I_u+1} \times [S]) \mid \begin{array}{l}
(\tilde{Z}, \tilde{s}) \sim P \\
\mathbb{E}_P[\tilde{Z} | \tilde{s} \in \mathcal{E}_k] \in Q_k \quad \forall k \in [K] \\
P[\tilde{Z} \in \mathcal{Z}_s | \tilde{s} = s] = 1 \quad \forall s \in [S] \\
P[\tilde{s} = s] = p_s \quad \forall s \in [S]
\end{array} \right\}
$$

takes lifted support sets $\mathcal{Z}_s = \{ Z \in \mathcal{U}_s \mid Z \succeq O, \ [Z]_{1,1} = 1 \}, s \in [S]$. Moreover, the bound is tight for ellipsoidal support sets defined in (22).

**Proof of Theorem 6.** We note that

$$
g(r(s), u, s) = \begin{pmatrix}
1 & b^\top(r(s), s) \\
b(r(s), s) & A^\top(r(s), s)A(r(s), s)
\end{pmatrix} \cdot \begin{pmatrix}
1 & 1 \\
u & u^\top
\end{pmatrix}.
$$
By Schur complement, each positive semidefinite constraint of problem (24) is equivalent to

\[ R(s) \succeq \begin{pmatrix}
1 & b^T(r(s), s) \\
(\mathbf{r}(s), s) A^T(r(s), s) A(r(s), s)
\end{pmatrix}. \]

Since \( Z \in \mathcal{Z}_s \) is positive semidefinite, an optimal \( R(s) \) would be

\[ R(s) = \begin{pmatrix}
1 & b^T(r(s), s) \\
(\mathbf{r}(s), s) A^T(r(s), s) A(r(s), s)
\end{pmatrix}. \]

Observe that the ambiguity set \( \mathcal{F} \) coincides with \( \mathcal{G} \) if every support set \( Z \in \mathcal{Z}_s \) is replaced by \( \bar{Z}_s = \{ Z \in U_s \mid Z \succeq O, [Z]_{1,1} = 1, \text{rank}(Z) = 1 \} \), which however, would lead to a harder problem to solve due to the rank constraint. Since \( \bar{Z}_s \subseteq \mathcal{Z}_s \), we obtain the conservative upper bound.

We next show that the bound is tight for ellipsoidal uncertainty sets defined in (22). After using Theorem 1 to reformulate problem (23), we need to deal with the following robust counterpart.

\[ \alpha_s \geq \Phi_s \bullet \left( \begin{pmatrix}
1 \\
\mathbf{u}
\end{pmatrix} \right)^\top \forall \mathbf{u}^\top \Lambda_s \mathbf{u} \leq 1 \]

for some \( \alpha_s \in \mathbb{R}, \Phi_s \in \mathbb{S}^{I_u+1} \), which by S-lemma, is equivalent to

\[ \begin{pmatrix}
\alpha_s & 0^\top \\
0 & 0
\end{pmatrix} + \delta_s \begin{pmatrix}
-1 & 0^\top \\
0 & \Lambda_s
\end{pmatrix} \succeq \Phi_s, \]

for some \( \delta_s \geq 0 \). On the other hand, the robust counterpart in the reformulation of problem (24)

\[ \alpha_s \geq \Phi_s \bullet Z \forall Z \bullet \begin{pmatrix}
-1 & 0^\top \\
0 & \Lambda_s
\end{pmatrix} \leq 0, \ [Z]_{1,1} = 1, \ Z \succeq O \]

is equivalent to

\[ \begin{pmatrix}
\tau_s & 0^\top \\
0 & 0
\end{pmatrix} + \delta_s \begin{pmatrix}
-1 & 0^\top \\
0 & \Lambda_s
\end{pmatrix} \succeq \Phi_s, \]

for some \( \tau_s \leq \alpha_s \) and \( \delta_s \geq 0 \), for which we can replace \( \tau_s \) with \( \alpha_s \) without affecting its feasibility. This establishes the desired tight bound for ellipsoidal uncertainty sets.

\[ \Box \]

**Bi-Conic-Quadratic Function**

We can also extend the bi-conic-quadratic function considered in Ben-Tal and Nemirovski (1998) to include discrete scenarios as follows:

\[ h(r(s), \mathbf{u}, s) \triangleq \| A(r(s), s) \mathbf{u} + b(r(s), s) \|_2, \]

where given \( s, A(r(s), s), b(r(s), s) \) are affine mappings of \( r(s) \). The event-wise ambiguity set takes

\[ \mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0 (\mathbb{R}^{I_u} \times [S]) \mid \begin{array}{l}
(\tilde{\mathbf{u}}, \tilde{s}) \sim \mathbb{P} \\
\mathbb{P} \left[ \begin{pmatrix} 0 \\ \mathbf{u} \end{pmatrix} = \mathbf{s} \right] = 1 \forall s \in [S] \\
\mathbb{P} \left[ \tilde{\mathbf{s}} = s \right] = p_s \forall s \in [S] \\
\text{for some } p \in \mathcal{P}
\end{array} \right\}. \]
Theorem 7. The worst-case expectation

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P [h(r(s), \tilde{u}, \tilde{s})]$$

is bounded from above by

$$\min \sup_{P \in \mathcal{F}} \mathbb{E}_P [x(\tilde{s})]$$

s.t. \( x(s) \geq R(s) \cdot Z \)

\[ \left( \begin{array}{c} R(s) \\ \begin{bmatrix} b(r(s), s) \\ A^\top(r(s), s) \end{bmatrix} \end{array} \right) \succeq O \quad \forall s \in [S] \]

\[ x \in \mathcal{A}(\mathcal{C}), \]

where \( \mathcal{C} \triangleq \{ \{ s \} | s \in [S] \} \) and where the lifted event-wise ambiguity set

\[ \mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0 \left( \mathbb{S}_+^{l+1} \times [S] \right) \right\} \]

\[ \mathbb{P}[Z \in \mathcal{Z}_s | \tilde{s} = s] = 1 \quad \forall s \in [S] \]

\[ \mathbb{P}[\tilde{s} = s] = p_s \quad \forall s \in [S] \]

takes lifted support sets \( \mathcal{Z}_s = \{ Z \in \mathcal{U}_s | Z \succeq O, [Z]_{1,1} = 1 \}, s \in [S] \). Moreover, the bound is tight for ellipsoidal uncertainty sets defined in (22).

Proof of Theorem 7. Since the ambiguity set does not contain any expectation constraint, we can obtain a tractable reformulation by replacing \( h(r(s), u, s) \) with a recourse variable \( x(s) \) and imposing the following constraint (see reformulation in Theorem 1):

$$x^2(s) \geq h^2(r(s), u, s) \quad \forall \left( \begin{array}{c} 1 \\ u \end{array} \right) \left( \begin{array}{c} 1 \\ u \end{array} \right)^\top \in \mathcal{U}_s, s \in [S].$$

We next discuss how such a constraint can be specified in problem (25). Observe that

$$h^2(r(s), u, s) = \left( \begin{array}{c} b^\top(r(s), s) \\ A^\top(r(s), s) \end{array} \right) \left( b(r(s), s) \ A(r(s), s) \right) \cdot \left( \begin{array}{c} 1 \\ u \end{array} \right) \left( \begin{array}{c} 1 \\ u \end{array} \right)^\top.$$
Affine-Quadratic Function

As an extension of Tüüncü and Koenig (2004), we consider a saddle function that is convex quadratic with respect to the decision variable and that is affine with respect to $z$:

$$g(r(s), z, s) \triangleq r^\top(s)H(s, z)r(s) + r^\top(s)G(s)z + h(s),$$

where given a scenario $s$, $H(s, z)$ is an affine mapping of $z$ and $Z_s \subseteq \{z \mid H(z, s) \succeq O\}$. Introducing auxiliary variables $R(s) \in \mathcal{S}_{Iv}, s \in [S]$ and using the Schur complement, the robust expectation

$$\min_{P \in \mathcal{F}} \sup_{P \in \mathcal{F}} \mathbb{E}_P [g(r(\tilde{s}), \tilde{z}, \tilde{s})]$$

s.t.

$$\begin{pmatrix} 1 & r^\top(s) \\ r(s) & R(s) \end{pmatrix} \succeq 0 \quad \forall s \in [S]$$

$$R(s) \in \mathcal{S}_{Iv} \quad \forall s \in [S],$$

which falls within the RSO framework.

C. Representation of Wasserstein Ambiguity Sets

For $p \in [1, \infty)$, the type-$p$ Wasserstein metric between two distributions $P$ and $\hat{P}$ for a given distance metric $\rho$ is defined as

$$d_W^p(P, \hat{P}) \triangleq \inf_{Q \in \mathcal{Q}(P, \hat{P})} \mathbb{E}_Q \left[ \rho^p(\tilde{u}, \tilde{u}^\top) \right]^{\frac{1}{p}}.$$

Correspondingly, the type-$p$ Wasserstein ambiguity set is defined by

$$G^p_W(\theta) = \left\{ P \in \mathcal{P}_0(\mathcal{U}) \mid \begin{array}{c} \hat{u} \sim P \\
\tilde{u} \sim \hat{P} \\
\rho^p(\tilde{u}, \tilde{u}^\top) \leq \theta \\
\end{array} \right\}.$$

Consider another distance metric $\rho^p$ and the corresponding type-1 Wasserstein metric $\tilde{d}_W(P, \hat{P})$ between $P$ and $\hat{P}$ which is determined by

$$\tilde{d}_W(P, \hat{P}) \triangleq \inf_{Q \in \mathcal{Q}(P, \hat{P})} \mathbb{E}_Q \left[ \rho^p(\tilde{u}, \tilde{u}^\top) \right].$$

We then have

$$G^p_W(\theta) = \left\{ P \in \mathcal{P}_0(\mathcal{U}) \mid \begin{array}{c} \hat{u} \sim P \\
\tilde{u} \sim \hat{P} \\
\tilde{d}_W(P, \hat{P}) \leq \theta^p \\
\end{array} \right\}.$$

Equivalently, for $p \in [1, \infty)$, the type-$p$ Wasserstein ambiguity set of radius $\theta$ can be re-interpreted as a type-1 Wasserstein ambiguity set of radius $\theta^p$ where the type-1 Wasserstein metric between $P$ and $\hat{P}$ is $\tilde{d}_W(P, \hat{P})$. From this perspective, we can directly use Theorem 2 to represent the type-$p$ Wasserstein ambiguity set $G^p_W(\theta)$ in the format of an event-wise ambiguity set.
For \( p = \infty \), the type-\( \infty \) Wasserstein metric between two distributions \( P \) and \( \hat{P} \) is defined as

\[
d^\infty_W(P, \hat{P}) \triangleq \inf_{Q \in \mathcal{Q}(P, \hat{P})} \text{Q-ess sup}_{U \times U} \rho(\tilde{u}, \hat{u}^\dagger),
\]

where the essential supremum of the joint distribution \( Q \) is defined by

\[
\text{Q-ess sup}_{U \times U} \rho(\tilde{u}, \hat{u}^\dagger) = \inf \{ M : Q[\rho(\tilde{u} - \hat{u}^\dagger) > M] = 0 \}.
\]

Bertsimas et al. (2018a) show that a distribution \( P \) in the type-\( \infty \) Wasserstein ambiguity set

\[
\mathcal{G}^\infty_W(\theta) = \left\{ P \in \mathcal{P}_0(U) \mid \tilde{u} \sim P, d^\infty_W(P, \hat{P}) \leq \theta \right\}
\]

is indeed a mixture distribution \( P = \frac{1}{S} \sum_{s \in [S]} P_s \) consisting of ambiguous components such that for every \( s \in [S] \), \( P_s \in \mathcal{P}_0(U) \) and \( P_s[\rho(\tilde{u}, \hat{u}_s) \leq \theta] = 1 \). Therefore, we can represent the type-\( \infty \) Wasserstein ambiguity set using the following mixture distribution ambiguity set

\[
\mathcal{F}^\infty_W(\theta) = \left\{ P \in \mathcal{P}_0(\mathbb{R}^I \times [S]) \mid \begin{array}{l}
(\tilde{u}, \tilde{s}) \sim P, \\
P[\tilde{u} \in U, \rho(\tilde{u}, \hat{u}_s) \leq \theta \mid \tilde{s} = s] = 1 \quad \forall s \in [S] \\
P[\tilde{s} = s] = \frac{1}{S} \quad \forall s \in [S]
\end{array} \right\},
\]

which is an event-wise ambiguity set satisfying \( \mathcal{G}^\infty_W(\theta) = \Pi_s \mathcal{F}^\infty_W(\theta) \) for all \( \theta \geq 0 \).

D. Computational Experiments with Wasserstein Ambiguity Sets

We focus on two-stage linear optimization problems with the data-driven Wasserstein ambiguity set of type-1 in the form (7), given some past observations \( \hat{u}_1, \ldots, \hat{u}_S \) of the uncertainty.

Multi-Item Newsvendor Problem

We consider a multi-item newsvendor problem with \( I_u \) different items. For each item \( i \ (i \in [I_u]) \), its unit selling price and ordering cost are denoted by \( p_i \) and \( c_i \), respectively. Under a total budget \( d \), the decision maker decides the ordering quantity \( w_i \) of each item before its random demand \( \tilde{u}_i \) is observed. Once the demand realizes, the selling quantity of each item is decided as \( \min \{ w_i, u_i \} \).

The decision maker maximizes the worst-case expected operating revenue by solving

\[
\max \inf_{P \in \mathcal{F}_W(\theta)} \mathbb{E}_P \left[ \sum_{i \in [I_u]} p_i \min \{ w_i, \tilde{u}_i \} \right]
\]

\[
\text{s.t. } c^\top w = d, \ w \geq 0,
\]

which can be recast as a minimization problem,

\[
\min -p^\top w + \sup_{P \in \mathcal{F}_W(\theta)} \mathbb{E}_P \left[ \sum_{i \in [I_u]} p_i (w_i - \tilde{u}_i)^+ \right]
\]

\[
\text{s.t. } c^\top w = d, \ w \geq 0.
\]
In the objective function, 
\[
\sum_{i \in [I_u]} p_i (w_i - u_i)^+ = \max_{\mathcal{J} \subseteq [I_u]} \sum_{j \in \mathcal{J}} p_j (w_j - u_j)
\]
is convex and piecewise affine involving \(2^{I_u}\) pieces. Thus by Theorem 3, problem (27) can be exactly solved by

\[
\begin{align*}
\lambda^* &= \min -p^T w + \sup_{\mathcal{F}_w(\theta)} E_{p} [y(\tilde{s}, \tilde{z})] \\
\text{s.t.} &\quad y(s, z) \geq \sum_{j \in \mathcal{J}} p_j (w_j - u_j) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \mathcal{J} \subseteq [I_u] \\
&\quad c^T w = d, \ \ w \geq 0 \\
&\quad y \in \bar{A}(\bar{C}, [I_u + 1]),
\end{align*}
\]

where we introduce a recourse variable \(y(\cdot, \cdot)\) following the event-wise affine adaptation with the collection \(\bar{C} \triangleq \{ \{s\} | s \in [S] \}\). Problem size of this exact approach however, increases exponentially in the number of items. Alternatively, we can obtain an upper bound by solving an RSO problem:

\[
\begin{align*}
\lambda &= \min -p^T x + \sup_{\mathcal{F}_w(\theta)} E_{p} [p^T y(\tilde{s}, \tilde{z})] \\
\text{s.t.} &\quad y(s, z) \geq 0 \quad \forall z \in \mathcal{Z}_s, \ s \in [S] \\
&\quad y(s, z) \geq w - u \quad \forall z \in \mathcal{Z}_s, \ s \in [S] \\
&\quad c^T w = d, \ \ w \geq 0 \\
&\quad y_i \in \bar{A} (\bar{C}, \mathcal{I}) \quad \forall i \in [I_u],
\end{align*}
\]

where we control how the recourse decision \(y(\cdot, \cdot)\) adapts to \((\tilde{u}, \tilde{v})\) and \(\tilde{s}\) through choosing the collection \(\mathcal{C}\) of MECE events and the information index set \(\mathcal{I}\).

We consider \(I_u \in \{5, 7\}\) and \(S \in \{5, 10, 20, 50\}\). The random demand belongs to a support set \(\mathcal{U} = [0, \bar{u}]\), and we use the Euclidean norm \(\| \cdot \|_2\) as the distance metric. In each instance, we randomly generate the upper bound \(\bar{u}\) from a uniform distribution on \([0, 100]^{I_u}\). Subsequently, past observations are randomly generated from the uniform distribution on \([0, \bar{u}]\). We set \(c_i = 1, \ i \in [I_u]\) and \(b = 50I_u\), and we generate \(p\) from a uniform distribution on \([0, 5]^{I_u}\). For different choices of \(\theta\), we run 100 random instances and compare the performance of different cases of event-wise recourse adaptation against the exact approach:

- case 1: \(y(\cdot, \cdot)\) adapts on \(\tilde{u}, \tilde{v}, \tilde{s}\), i.e., \(\mathcal{C} = \{\{s\} | s \in [S]\}\) and \(\mathcal{I} = [I_u + 1]\);
- case 2: \(y(\cdot, \cdot)\) adapts only on \(\tilde{u}, \tilde{v}\), i.e., \(\mathcal{C} = \{1, \ldots, S\}\) and \(\mathcal{I} = [I_u + 1]\);
- case 3: \(y(\cdot, \cdot)\) adapts only on \(\tilde{u}\), i.e., \(\mathcal{C} = \{1, \ldots, S\}\) and \(\mathcal{I} = [I_u]\).

Case 1 corresponds to the full event-wise affine adaptation. We turn off the event-wise adaptation in case 2, while in case 3 we further deprive the recourse decision \(y(\cdot, \cdot)\) of the affine adaptation on the auxiliary random variable \(\tilde{v}\). For each case, we consider the following relative gap between the objective value using the event-wise recourse adaptation and the exact optimal objective value:

\[
\frac{\lambda^* - \lambda}{\lambda^*} \times 100%.
\]
Table 1 5 items: 90-th percentile optimality gaps (in %) of case 1 (left), case 2 (middle), and case 3 (right).

<table>
<thead>
<tr>
<th>S</th>
<th>( \theta )</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1  2.0  8.7</td>
<td>&lt;0.1  3.4  8.8</td>
<td>0.1  5.4  9.2</td>
<td>0.2  7.0  9.9</td>
<td>0.5  9.4 10.5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1  4.5 11.6</td>
<td>&lt;0.1  6.4 11.9</td>
<td>0.2  8.6 12.3</td>
<td>0.3 11.0 12.7</td>
<td>1.0 12.6 13.0</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1  4.8  5.7</td>
<td>&lt;0.1  5.3  5.8</td>
<td>&lt;0.1  5.6  6.1</td>
<td>0.1  6.4  6.5</td>
<td>0.3  7.9  7.9</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>&lt;0.1  6.8  7.9</td>
<td>&lt;0.1  7.4  8.0</td>
<td>&lt;0.1  8.1  8.4</td>
<td>0.1  9.2  9.2</td>
<td>0.3 11.3 11.3</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 7 items: 90-th percentile optimality gaps (in %) of case 1 (left), case 2 (middle), and case 3 (right).

<table>
<thead>
<tr>
<th>S</th>
<th>( \theta )</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1  1.3  8.9</td>
<td>0.1  2.3  9.0</td>
<td>0.2  3.8  9.3</td>
<td>0.3  5.6  9.8</td>
<td>0.6  8.2 10.5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1  2.5  6.2</td>
<td>0.1  3.5  6.2</td>
<td>0.1  5.0  6.4</td>
<td>0.2  6.0  6.6</td>
<td>0.5  7.6  7.7</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1  4.8  6.7</td>
<td>0.1  5.7  6.8</td>
<td>0.1  6.7  6.9</td>
<td>0.2  7.3  7.3</td>
<td>0.6  7.9  7.9</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.1  5.5  6.1</td>
<td>0.1  5.8  6.1</td>
<td>0.2  6.3  6.3</td>
<td>0.2  6.3  6.6</td>
<td>0.5  7.6  7.6</td>
<td></td>
</tr>
</tbody>
</table>

Results for \( I_u = 5 \) and \( I_u = 7 \) are summarized in Table 1 and Table 2, respectively. With (i) the notion of event-wise adaptation and (ii) the inclusion of auxiliary random variable \( \tilde{v} \), the full event-wise affine adaptation could provide a reasonably good conservative approximation to the exact approach; while excluding either (i) or (ii) may lead to a more conservative approximation.

We evaluate the scalability of the full event-wise affine adaptation (case 1), the affine adaptation without event-wise dependence (case 2), and the exact approach, by comparing their computation times and limits for different pairs of problem sizes. For the exact approach, the computer runs out of memory when the number of items exceeds 10 and the number of samples exceeds 5 (see Table 3), which is not practically favorable. In contrast, we are able to obtain a conservative solution via the full event-wise affine adaptation with modest computational effort. Quite interestingly, the event-wise adaptation that plays the key role in delivering the less conservative approximation seems to require only a little extra computational effort (see Table 4).

The following code segment shows how to implement the full event-wise affine adaptation for the multi-item newsvendor problem with Wasserstein ambiguity sets in RSOME.

```plaintext
% I: number of items
% S: number of past observations
% \theta: radius
```
(S, IU)
(5, 5) (10, 5) (20, 5) (50, 5) (100, 5) (200, 5) (5, 10)
0.1 1.1 0.2 0.5 1.3 9.4 9.8

Table 3  Computation times of the exact approach. We report only those (S, IU) pairs for which the exact approach were solved.

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1</td>
<td>0.2</td>
<td>0.5</td>
<td>0.7</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1</td>
<td>0.3</td>
<td>0.6</td>
<td>2.4</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1</td>
<td>0.7</td>
<td>1.3</td>
<td>2.2</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>0.8</td>
<td>1.8</td>
<td>6.8</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.4</td>
<td>5.2</td>
<td>10.3</td>
<td>17.8</td>
<td>31.0</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.5</td>
<td>2.0</td>
<td>5.5</td>
<td>11.0</td>
<td>19.1</td>
<td>32.8</td>
</tr>
<tr>
<td>300</td>
<td>0.8</td>
<td>3.1</td>
<td>7.8</td>
<td>16.5</td>
<td>32.3</td>
<td>49.6</td>
</tr>
<tr>
<td>500</td>
<td>1.4</td>
<td>5.7</td>
<td>16.6</td>
<td>36.6</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>800</td>
<td>2.5</td>
<td>14.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.7</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1</td>
<td>0.3</td>
<td>0.6</td>
<td>1.3</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>0.8</td>
<td>1.8</td>
<td>6.8</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.2</td>
<td>1.1</td>
<td>2.7</td>
<td>5.1</td>
<td>9.2</td>
<td>10.5</td>
</tr>
<tr>
<td>200</td>
<td>0.4</td>
<td>1.7</td>
<td>5.2</td>
<td>15.8</td>
<td>30.7</td>
<td>25.3</td>
</tr>
<tr>
<td>300</td>
<td>0.8</td>
<td>3.2</td>
<td>8.6</td>
<td>14.2</td>
<td>36.1</td>
<td>48.4</td>
</tr>
<tr>
<td>500</td>
<td>1.1</td>
<td>5.3</td>
<td>13.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>800</td>
<td>4.3</td>
<td>13.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 4  Computation times and limits of the affine approximation without event-wise adaptation (left) and the full event-wise affine adaptation (right). The symbol “—” indicates “out of memory”.

```matlab
% Gamma: total budget
% cost (price): cost (price) parameters
% ubar: upper bound of demand
% U = (u_1, ..., u_S): past realizations

% Create RSOME model
model = rsome('newsvendor');

% Define random variables
u = model.random;              % random demand
v = model.random;              % auxiliary random variable
P  = model.ambiguity(S);       % create ambiguity set
% Define support sets for scenarios
for s = 1:S
    P(s).superset(0 <= u, u <= ubar, norm(u - U(:,s)) <= v);
end

% Define probabilities for scenarios
```
Experiment of Hanasusanto and Kuhn (2018)

We benchmark the RSO model against a state-of-art approximation scheme proposed by Hanasusanto and Kuhn (2018). Particularly, we repeat their experiment using the same set-ups.

Consider the second-stage problem of the form

$$f(u) = \min \{ e^\top y \mid y \succeq 0, y \geq Au - b \}, \quad \text{(30)}$$

where $e$ is a vector of ones. The problem does not have any here-and-now decision $w$ and assumes that the random variable $\tilde{u}$ resides in a box $\mathcal{U} = [0, 1]^I$. Under the distance metric $\rho(u, u^\dagger) = \|u - u^\dagger\|_2$, Hanasusanto and Kuhn (2018) have shown that the worst-case expectation

$$\sup_{P \in \mathcal{F}_W(\theta)} \mathbb{E}_P[f(\tilde{u})] \quad \text{(31)}$$
amounts exactly to the optimal value of the following copositive program.

$$\inf_{\alpha} \frac{1}{S} \sum_{s \in [S]} \left( \alpha_s + q^T \psi_s - \beta \|\tilde{u}_s\|_2^2 + \sum_{\ell \in [I_u + J_y]} \phi_\ell q_\ell^2 \right) + \beta \theta^2$$

$$\begin{align*}
\text{s.t.} & \quad \left( \begin{array}{c}
\beta I + Q^T \text{diag}(\phi_s Q) - \frac{1}{2} T^T - Q^T \text{diag}(\phi_s) W^T - \beta \tilde{u}_s - \frac{1}{2} Q^T \psi_s \\
-\frac{1}{2} T - W \text{diag}(\phi_s) Q - \frac{1}{2} W \text{diag}(\phi_s) W^T - \frac{1}{2} (W \phi_s - \tilde{h})^T
\end{array} \right) \succeq_{K_{\text{cop}}} O \quad \forall s \in [S] \\
\alpha & \in \mathbb{R}^S, \quad \beta \in \mathbb{R}_+ \quad \psi_s, \phi_s \in \mathbb{R}^{I_u + J_y} \\
\end{align*}$$

(32)

Here, $K_{\text{cop}} = \{ M \in \mathbb{S}^K \mid x^T M x \geq 0 \ \forall x \geq 0 \}$ is the copositive cone,

$$\begin{align*}
Q = \begin{pmatrix} O \\ I \end{pmatrix}, \quad q = \begin{pmatrix} e \\ -e \end{pmatrix}, \quad T = \begin{pmatrix} A \\ O \end{pmatrix}, \quad \tilde{h} = \begin{pmatrix} -b \\ 0 \end{pmatrix}, \quad W = \begin{pmatrix} W & O \\ O & -I \end{pmatrix} \quad \text{with} \quad W = \begin{pmatrix} I \\ I \end{pmatrix},
\end{align*}$$

and $O$, $I$, $0$ and $e_i$ respectively correspond to the zero matrix, the identity matrix, the zero vector and the $i$-th standard unit basis, all of which are of the appropriate dimension. Because the copositive program (32) is generally intractable, Hanasusanto and Kuhn (2018) adopt a conservative $K_0$-approximation by replacing the copositive cone $K_{\text{cop}}$ with

$$K_0 = \{ M \in \mathbb{S}^K \mid M = P + N, P \geq O, N \geq O \} \subseteq K_{\text{cop}},$$

which leads problem (32) to a semidefinite program.

We run numerical tests for different pairs of the uncertainty dimension $I_u$ and the sample size $S$, and for each pair, we use the same set-ups as in Hanasusanto and Kuhn (2018) to generate 100 random instances. The Wasserstein radius is set to $\theta = 1/ S$. The dimension $J_y$ of the recourse decision is sampled uniformly at random from $\{1, 2, \ldots, \lceil \log(I_u + 1) \rceil \}$, $A$ is sampled uniformly from $[0, 1]^{J_y \times I_u}$, and $b$ is sampled uniformly from $[0, e^T A_{1:1}] \times \cdots \times [0, e^T A_{1:J_y}]$. Here, $A_{1:1}$ stands for the first row of $A$ and so forth. Lastly, we generate independent training samples from the uniform distribution on $[0, 1]^{I_u}$. We evaluate the worst-case expectation (31) approximately by using (i) the $K_0$-approximation and (ii) the following full event-wise affine adaptation:

$$\min \sup_{\pi \in \mathcal{F}_W(\theta)} E_{\pi} \left[ e^T y(s, \tilde{z}) \right]$$

$$\begin{align*}
\text{s.t.} & \quad y(s, z) \geq 0 \quad \forall z \in Z_s, \ s \in [S] \\
y(s, z) \geq Au - b \quad \forall z \in Z_s, \ s \in [S] \\
y_j \in \mathcal{A}\left(\{1, \ldots, S\}, \lceil I_u + 1 \rceil \right) \quad \forall j \in [J_y],
\end{align*}$$

where for each $s \in [S]$, $Z_s = \{(u, v) \mid u \in [0, 1]^{I_u}, v \geq \|u - \tilde{u}_s\|_2^2 \}$.

Quite surprisingly, for all pairs of problem sizes, the solutions of both approximation approaches coincide for all 100 randomly generated instances. Unfortunately, we are not able to give a formal
proof for this observation. Nevertheless, this observation supports that our proposed event-wise affine adaptation delivers solutions with competitive approximation quality as the state-of-the-art approximation scheme by Hanasusanto and Kuhn (2018). We report in Table 5 the average computation times of both approaches. In terms of computation efficiency, the event-wise affine adaptation outperforms because it leads to a second-order cone approximation to problem (32).

We note that the $K_0$-approximation by Hanasusanto and Kuhn (2018) also works when the cost vector of the second-stage problem (30) is affinely affected by the uncertainty, while our event-wise affine adaptation does not. On the other hand, the event-wise affine adaptation works with more general distance metrics and more general support sets that are not necessarily polyhedral (in the current experiment, the support set is a box $[0,1]^4$), while the $K_0$-approximation does not.

### E. Multi-Stage Stochastic Financial Planning Problem

We adopt a financial planning problem from Birge and Louveaux (2011) to illustrate how to incorporate the scenario tree approach in the RSO framework.

At the beginning of the first stage in this multi-stage problem, the decision maker allocates the wealth $d$ into two possible investment types, stocks (S) and bonds (B). Eight possible scenarios may occur, which corresponds to independent and equal likelihoods of having high returns of 1.25 for stocks and 1.14 for bonds, or low returns of 1.06 for stocks and 1.12 for bonds over subsequent stages (see Figure 5). Hence we can construct the following singleton ambiguity set of the known

<table>
<thead>
<tr>
<th>$S$</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.3 &lt; 0.1</td>
<td>0.3 &lt; 0.1</td>
<td>0.3 &lt; 0.1</td>
<td>0.5 &lt; 0.1</td>
<td>2.3 &lt; 0.1</td>
</tr>
<tr>
<td>10</td>
<td>0.4 &lt; 0.1</td>
<td>0.4 &lt; 0.1</td>
<td>0.5 &lt; 0.1</td>
<td>0.9 &lt; 0.1</td>
<td>5.1 &lt; 0.1</td>
</tr>
<tr>
<td>20</td>
<td>0.6 &lt; 0.1</td>
<td>0.7 &lt; 0.1</td>
<td>0.8 &lt; 0.1</td>
<td>1.8 &lt; 0.1</td>
<td>11.6 &lt; 0.1</td>
</tr>
<tr>
<td>40</td>
<td>1.1 &lt; 0.1</td>
<td>1.3 &lt; 0.1</td>
<td>1.6 &lt; 0.1</td>
<td>3.5 &lt; 0.1</td>
<td>23.1 0.1</td>
</tr>
<tr>
<td>80</td>
<td>2.3 &lt; 0.1</td>
<td>2.5 &lt; 0.1</td>
<td>3.2 &lt; 0.1</td>
<td>7.8 &lt; 0.1</td>
<td>51.1 0.1</td>
</tr>
<tr>
<td>160</td>
<td>4.5 &lt; 0.1</td>
<td>5.1 &lt; 0.1</td>
<td>7.0 &lt; 0.1</td>
<td>18.0 0.2</td>
<td>118.0 0.3</td>
</tr>
<tr>
<td>320</td>
<td>9.2 &lt; 0.1</td>
<td>10.8 0.1</td>
<td>15.5 0.2</td>
<td>45.4 0.3</td>
<td>281.5 0.6</td>
</tr>
<tr>
<td>640</td>
<td>19.7 &lt; 0.1</td>
<td>26.9 0.2</td>
<td>43.9 0.3</td>
<td>141.5 1.0</td>
<td>684.3 2.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$I_u$</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3 &lt; 0.1</td>
<td>0.3 &lt; 0.1</td>
<td>0.3 &lt; 0.1</td>
<td>0.5 &lt; 0.1</td>
<td>2.3 &lt; 0.1</td>
</tr>
<tr>
<td>10</td>
<td>0.4 &lt; 0.1</td>
<td>0.4 &lt; 0.1</td>
<td>0.5 &lt; 0.1</td>
<td>0.9 &lt; 0.1</td>
<td>5.1 &lt; 0.1</td>
</tr>
<tr>
<td>20</td>
<td>0.6 &lt; 0.1</td>
<td>0.7 &lt; 0.1</td>
<td>0.8 &lt; 0.1</td>
<td>1.8 &lt; 0.1</td>
<td>11.6 &lt; 0.1</td>
</tr>
<tr>
<td>40</td>
<td>1.1 &lt; 0.1</td>
<td>1.3 &lt; 0.1</td>
<td>1.6 &lt; 0.1</td>
<td>3.5 &lt; 0.1</td>
<td>23.1 0.1</td>
</tr>
<tr>
<td>80</td>
<td>2.3 &lt; 0.1</td>
<td>2.5 &lt; 0.1</td>
<td>3.2 &lt; 0.1</td>
<td>7.8 &lt; 0.1</td>
<td>51.1 0.1</td>
</tr>
<tr>
<td>160</td>
<td>4.5 &lt; 0.1</td>
<td>5.1 &lt; 0.1</td>
<td>7.0 &lt; 0.1</td>
<td>18.0 0.2</td>
<td>118.0 0.3</td>
</tr>
<tr>
<td>320</td>
<td>9.2 &lt; 0.1</td>
<td>10.8 0.1</td>
<td>15.5 0.2</td>
<td>45.4 0.3</td>
<td>281.5 0.6</td>
</tr>
<tr>
<td>640</td>
<td>19.7 &lt; 0.1</td>
<td>26.9 0.2</td>
<td>43.9 0.3</td>
<td>141.5 1.0</td>
<td>684.3 2.3</td>
</tr>
</tbody>
</table>

**Table 5** Computation times (in seconds) of $K_0$-approximation (left) and event-wise affine adaptation (right).
discrete distribution of uncertain returns over all stages.

\[ \mathcal{F} = \left\{ \mathbf{P} \in \mathcal{P}_0 \left( \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \times [S] \right) \mid (\tilde{z}_1, \tilde{z}_2, \tilde{z}_3, \tilde{s}) \sim \mathbf{P}, \mathbf{P} \left[ (\tilde{z}_1, \tilde{z}_2, \tilde{z}_3) \in \mathcal{Z}_s \mid \tilde{s} = s \right] = 1 \quad \forall s \in [S] \right\}, \]

where \( S = 8 \) and the singleton support sets \( \mathcal{Z}_s = \{(a_{1s}, a_{2s}, a_{3s})\}, s \in [S] \) are determined by \( a_{1s} = (1.25, 1.14) \) for \( s \in \{1, 2, 3, 4\}, a_{2s} = (1.25, 1.14) \) for \( s \in \{1, 2, 5, 6\}, a_{3s} = (1.25, 1.14) \) for \( s \in \{1, 3, 5, 7\}, \) and \( a_{4s} = (1.06, 1.12) \) otherwise.

The decision maker evaluates the difference between the final return \( r \) and a prescribed target \( \tau \) based on a concave and piecewise affine utility function that takes \( U(r - \tau) = r - \tau \) if \( r \geq \tau \) and \( U(r - \tau) = 4(r - \tau) \) otherwise. The initial investment decisions \( w \), made before the first stage returns of stocks and bonds realize, must be indifferent among all eight scenarios. The rebalanced investment decision \( x_1 \), made after the first stage returns realize but before the second stage returns realize, shall be indifferent among scenarios \( \{1, 2, 3, 4\} \) and indifferent among scenarios \( \{5, 6, 7, 8\} \). Similarly, \( x_2 \) is indifferent between scenarios \( \{1, 2\} \) as well as between scenarios \( \{3, 4\}, \{5, 6\}, \) and \( \{7, 8\} \). Finally, the nonnegative auxiliary recourse decisions \( \bar{x} \) and \( x \), respectively standing for the excess above or shortfall below the target, are adaptive to revealed uncertainties and thus can be
different across the eight scenarios. In all, we can formulate the RSO model as follows:

\[
\max \inf_{p \in \mathcal{F}} \mathbb{E}_p [\mathcal{F}(\hat{s}) - 4\mathcal{F}(\hat{s})]
\]

s.t. \(w_1, w_2 \geq 0, \ w_1 + w_2 = d\)

\[
x_{11}(s) + x_{12}(s) - \tilde{z}_1^\top w = 0 & \quad \forall z \in z, s \in [S] \\
x_{21}(s) + x_{22}(s) - \tilde{z}_2^\top x(s) = 0 & \quad \forall z \in z, s \in [S] \\
z_i^\top x_i(s) - \mathcal{I}(s) + x(s) = \tau & \quad \forall z \in z, s \in [S] \\
x_{11}(s), x_{12}(s), x_{21}(s), x_{22}(s), \mathcal{I}(s), \mathcal{F}(s) \geq 0 & \quad \forall s \in [S] \\
x_{11}, x_{12} \in \mathcal{A}([\{1, 2, 3, 4\}, \{5, 6, 7, 8\}] ) \\
x_{21}, x_{22} \in \mathcal{A}([\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}] ) \\
\mathcal{I}, \mathcal{F} \in \mathcal{A}([\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}]).
\]

F. Portfolio Management with K-means Adaptive Rebalancing

We consider a three-period portfolio allocation and rebalancing problem to minimize the investment risk at the last period taking into account of transaction costs. At the beginning of the first period, we decide the number of shares \(w_i \geq 0\) of stock \(i \in [I]\) to invest at price \(a_i\), incurring a transaction cost \(b_iw_i\). The price of stock \(i\) in the second period is \(\tilde{a}_i^1 \triangleq a_i(\tilde{u}_i^1 + 1)\), where \(\tilde{u}_i^1\) is the corresponding return. Subsequently, for each stock \(i\), we rebalance its shares to \(x_i \geq 0\), which incurs a transaction cost \(b_i|x_i - w_i|\). In the last period, the price of stock \(i\) is \(\tilde{a}_i^2 \triangleq a_i(\tilde{u}_i^2 + 1)\), where \(\tilde{u}_i^2\) is the third period return with respect to the first period price. The effective portfolio return at the last period, taking into account of the total transaction costs, amounts to

\[
w^\top \tilde{a}^1 - w^\top a + x^\top \tilde{a}^2 - x^\top \tilde{a}^1 - b^\top (w + |x - w|) = w^\top A\tilde{u}^1 + x^\top A(\tilde{u}^2 - \tilde{u}^1) - b^\top (w + |x - w|),
\]

where \(A = \text{diag}(a)\) and the operator \(|\cdot|\) takes the absolute value component-wise.

Ideally, the rebalancing decision \(x\) should depend on the realization of \(\tilde{u}^1\). However, this would lead to an intractable problem. Instead, we propose a K-means adaptive approach, where the recourse decision \(x(\hat{s})\) depends on the random scenario \(\hat{s}\) that is associated with the realization of \(\tilde{u}^1\). In particular, using the available historical returns \(\{(\hat{u}_n^1, \hat{u}_n^2)\}_{n \in [N]}\), we construct a two-layer K-means ambiguity set by first partitioning \(\{\hat{u}_n^1\}_{n \in [N]}\) into \(K_1\) clusters, each of which we then further partition into \(K_2\) clusters based on a subset of \(\{\hat{u}_n^2\}_{n \in [N]}\) that are affiliated with this specific first-layer cluster. As a result, we obtain a total number of \(S = K_1K_2\) scenarios, each of which corresponds to a unique cluster determined by the first and second layers; see an illustration in Figure 2. For each of the first-layer cluster \(k \in [K_1]\), we denote by \(E_k \subseteq [S]\) the set of scenarios associated with that cluster. Correspondingly, we define \(\kappa(s) \in [K_1]\) as the specific first-layer cluster
that the scenario $s$ affiliates with. Observe that $C = \{E_k \mid k \in [K_1]\}$ is a collection of MECE events. In this way, we obtain the two-layer K-means ambiguity set

$$
\mathcal{F} = \left\{ P \in \mathcal{P}_0 (\mathbb{R}^{2I_u+2I_v} \times [S]) \mid \begin{array}{l}
((\hat{u}^1, \hat{u}^2, \hat{v}^1, \hat{v}^2), \hat{s}) \sim P \\
\mathbb{E}_P [\hat{u}^1 \mid \hat{s} \in E_k] = \hat{\mu}^1_k \quad \forall k \in [K_1] \\
\mathbb{E}_P [\hat{v}^1 \mid \hat{s} \in E_k] \leq \hat{\sigma}^1_k \quad \forall k \in [K_1] \\
\mathbb{E}_P [\hat{u}^2 \mid \hat{s} = s] = \hat{\mu}^2_s \quad \forall s \in [S] \\
\mathbb{E}_P [\hat{v}^2 \mid \hat{s} = s] \leq \hat{\sigma}^2_s \quad \forall s \in [S] \\
P [((\hat{u}^1, \hat{u}^2, \hat{v}^1, \hat{v}^2) \in Z_s \mid \hat{s} = s) = 1 \quad \forall s \in [S] \\
P [\hat{s} = s] = p_s \quad \forall s \in [S]
\end{array}\right\},
$$

where for each $s \in [S]$, the cluster-wise support set is determined by

$$
Z_s = \{(u^1, u^2, v^1, v^2) \mid u^1 \in \mathcal{U}^1_{\kappa(s)}, u^2 \in \mathcal{U}^2_s, v^1 \geq \phi(u^1), v^2 \geq \phi(u^2)\}.
$$

The objective function evaluates the worst-case conditional value-at-risk (CVaR) of the final return at a pre-specified risk threshold $\varepsilon \in (0, 1)$.

$$
\sup_{P \in \mathcal{F}} \mathbb{P} - \text{CVaR}_\varepsilon \left( w^\top A\hat{u}^1 + x^\top (\hat{s}) A(\hat{u}^2 - \hat{u}^1) - b^\top (w + |x(\hat{s}) - w|) \right),
$$

which using now standard techniques, can be rewritten as

$$
\min_{\delta} \delta + \frac{1}{\varepsilon} \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ (w^\top (\hat{s}) A (\hat{u}^1 - \hat{u}^2) - w^\top A\hat{u}^1 + b^\top (w + |x(\hat{s}) - w|) - \delta)^+ \right].
$$
Using Theorem 3, we formulate the RSO model for this portfolio optimization problem as follows:

\[
\begin{align*}
\min & \quad \delta + \frac{1}{\varepsilon} \sup_{\mathcal{F}} \mathbb{E}_\mathcal{F} [y(\mathbf{s}, \mathbf{z})] \\
\text{s.t.} & \quad y(\mathbf{s}, \mathbf{z}) \geq 0 \\
& \quad y(\mathbf{s}, \mathbf{z}) \geq \mathbf{x}^\top(s) A(\mathbf{u}^1 - \mathbf{u}^2) - \mathbf{w}^\top A \mathbf{u}^1 + b^\top (\mathbf{w} + \mathbf{x}(s)) - \delta \quad \forall \mathbf{z} \in \mathcal{Z}_s, s \in [S] \\
& \quad \bar{x}(s) \geq \mathbf{x}(s) - \mathbf{w} \quad \forall s \in [S] \\
& \quad \bar{x}(s) \geq \mathbf{w} - \mathbf{x}(s) \quad \forall s \in [S] \\
& \quad b^\top \bar{x}(s) \leq \eta \quad \forall s \in [S] \\
& \quad \mathbf{x}(s) \geq \mathbf{0} \quad \forall s \in [S] \\
& \quad \mathbf{a}^\top \mathbf{w} = d \\
& \quad \mathbf{w} \geq \mathbf{0} \\
& \quad x_i, \bar{x}_i \in \mathcal{A}(\mathcal{C}) \quad \forall i \in [I_u] \\
& \quad y \in \mathcal{A}(\bar{\mathcal{C}}, [2I_u + 2I_v])
\end{align*}
\]

where \( \mathbf{z} \triangleq (\hat{\mathbf{u}}^1, \hat{\mathbf{u}}^2, \hat{\mathbf{v}}^1, \hat{\mathbf{v}}^2) \) and \( \bar{\mathcal{C}} \triangleq \{ s \mid s \in [S] \} \) consists of singleton MECE events. For simplicity, we impose a limit \( \eta \) on the transaction cost to prohibit over rebalancing in the second period, and \( d \) is the initial allocation budget. We next provide the sample code in RSOME to elucidate the intuitive implementation of the RSO model via an algebraic modeling language.

**Sample Code for K-means Adaptive Rebalancing**

We assume \( \mathcal{E}_k = \{(k - 1)K_2 + 1, \ldots, kK_2\} \subseteq [S] \) for all \( k \in [K_1] \). Correspondingly, \( \kappa(s) = \lfloor \frac{s}{K_2} \rfloor \) for all \( s \in [S] \). We take a convex function \( \phi \) that specifies the mean absolute deviation of each random return within a particular cluster. Hence for each \( s \in [S] \), the cluster-wise support set is given by

\[ \mathcal{Z}_s = \{(u^1, u^2, v^1, v^2) \mid D^1_{\kappa(s)} \mathbf{u}^1 \leq f^1_s, D^2_{s} \mathbf{u}^2 \leq f^2_s, v^1 \geq |\mathbf{u}^1 - \hat{\mu}^1_{\kappa(s)}|, v^2 \geq |\mathbf{u}^2 - \hat{\mu}^2_s|\}, \]

where each cluster is in fact a polyhedron and where \( |\cdot| \) applies component-wise. The estimates \( \{D^1_k\}_{k \in [K_1]}, \{f^1_k\}_{k \in [K_1]}, \{\hat{\mu}^1_k\}_{k \in [K_1]}, \{\hat{\sigma}^1_k\}_{k \in [K_1]} \) are contained in MATLAB cells D1, f1, mu1, sigma1, and similarly, \( \{D^2_s\}_{s \in [S]}, \{f^2_s\}_{s \in [S]}, \{\mu^2_s\}_{s \in [S]}, \{\sigma^2_s\}_{s \in [S]} \) are contained in D2, f2, mu2, sigma2.

```matlab
% I: number of stocks
% K1: number of first-layer clusters
% K2: number of second-layer clusters
% ps: probabilities of clusters
% epsilon: risk threshold
% a,b,d,eta: parameters

% Create RSOME model
model = rsome('portfolio');

% Define random variables
```
\begin{verbatim}
% Define support sets for scenarios
for s = 1:K1*K2
    P(s).suppset(D1{ceil(s/K2)}*u(:,1) <= f1{ceil(s/K2)}, ...
                  D2{s}*u(:,2) <= f2{s}, ...
                  v(:,1) >= abs(u(:,1) - mu1{ceil(s/K2)}), ...
                  v(:,2) >= abs(u(:,2) - mu2{s}));
end
%
% Define probabilities for scenarios
pr = P.prob;
P.probset(pr == ps);
%
% Define event-wise expectation
for k = 1:K1
    P((k-1)*K2+1:k*K2).exptset(expect(u(:,1)) == mu1{k}, ...
                                expect(v(:,1)) <= sigma1{k});
end
for s = 1:K1*K2
    P(s).exptset(expect(u(:,2)) == mu2{s}, ...
                  expect(v(:,2)) <= sigma2{s});
end
%
% Declare K-means ambiguity set
model.with(P);
%
% Define decision variables
w = model.decision(I,1);
x = model.decision(I,1);
xbar = model.decision(I,1);
y = model.decision;
delta = model.decision;
%
% Define event-wise adaptation
for k = 1:K1
    x.evtadapt((k-1)*K2+1:k*K2);
    xbar.evtadapt((k-1)*K2+1:k*K2);
end
for s = 1:K1*K2
    y.evtadapt(s);
end
%
% Define affine adaptation
y.affadapt(u);
y.affadapt(v);
%
% Define objective function
model.min(delta + expect((1/epsilon)*y));
\end{verbatim}
% Define constraints
model.append(y >= 0);
model.append(y >= x' * diag(a)*(u(:, 1) - u(:, 2)) ... 
- w' * diag(a)*u(:, 1) + b'(w + xbar) - delta);
model.append(xbar >= abs(x - w));
model.append(b' * xbar <= eta);
model.append(x >= 0);
model.append(a' * w == d);
model.append(w >= 0);

% Solution
model.solve;

G. Endnote
All mathematical programs in numerical experiments are solved using MOSEK on an Intel Core (TM) @ 3.40 GHz with 8GB RAM. The semidefinite program related to the $K_0$-approximation is implemented using the CVX interface (Grant and Boyd 2008), while remaining models are implemented using our developed algebraic modeling package RSOME (available at https://www.rsomerso.com/).