A practicable framework for distributionally robust linear optimization

Dimitris Bertsimas  
Massachusetts Institute of Technology  
Melvyn Sim, Meilin Zhang  
NUS Business School, National University of Singapore

We developed a modular framework to obtain exact and approximate solutions to a class of linear optimization problems with recourse with the goal to minimize the worst-case expected objective over an ambiguity set of distributions. The ambiguity set is specified by linear and conic quadratic representable expectation constraints and the support set is also linear and conic quadratic representable. We propose an approach to lift the original ambiguity set to an extended one by introducing additional auxiliary random variables. We show that by replacing the recourse decision functions with generalized linear decision rules that have affine dependency on the uncertain parameters and the auxiliary random variables, we can obtain good and sometimes tight approximations to a two-stage optimization problem. This approach extends to a multistage problem and improves upon existing variants of linear decision rules. We demonstrate the practicability of our framework by developing a new algebraic modeling package named ROC, a C++ library that implements the techniques developed in this paper.

Key words: distributionally robust optimization, risk, ambiguity

1. Introduction

Real world optimization problems are often confounded by the difficulties of addressing the issues of uncertainty. In characterizing uncertainty, Knight (1921) is among the first to establish the distinction of risk, where the probability distribution of the uncertainty is known, and ambiguity, where it is not. Ambiguity exists in practice because it is often difficult or impossible to obtain the true probability distribution due to the possibly lack of available or “good enough” empirical records associated with the uncertain parameters. However, in normative decision making, ambiguity is often ignored in favor of risk preferences over subjective probabilities. Notably, Ellsberg (1961) demonstrates that choice under the presence of ambiguity cannot be reconciled by subjective risk preferences and his findings are corroborated in later studies including the groundbreaking research of Hsu et al. (2005).

In classical stochastic optimization models, uncertainties are represented as random variables with probability distributions and the decision makers optimize the solutions according to their
risk preferences (see, for instance, Birge and Louveaux (1997), Ruszczynski and Shapiro (2003)). In particular, risk neutral decision makers prefer solutions that yield optimal expected or average objectives, which are evaluated based on the given probability distributions that characterize the uncertain parameters of the models. Hence, classical stochastic optimization models do not account for ambiguity and subjective probability distributions are used in these models whenever the true distributions are unavailable.

In recent years, research on ambiguity has garnered considerable research interest in various fields including economics, mathematical finance and operations research. Under the pretext of ambiguity aversion, robust optimization is a relatively new approach that deals with ambiguity in mathematical optimization problems. In classical robust optimization, full ambiguity is assumed and uncertainty is distribution free described by only the support set, which is typically in the form of a conic representable bounded convex set (see Ghaoui and Lebret (1997), Ghaoui et al. (1998), Ben-Tal and Nemirovski (1998, 1999, 2000), Bertsimas and Sim (2004)). Hence, in contrast to a stochastic optimization model, a robust optimization model does not directly account for risk preferences.

Both risk and ambiguity should be taken into account in modeling an optimization problem under uncertainty. From the decision theoretic perspective, Gilboa and Schmeidler (1989) propose to rank preferences based on the worst-case expected utility or disutility over an ambiguity set of distributions. Scarf (1958) is arguably the first to conjure such an optimization model when he studies a single-product newsvendor problem in which the precise demand distribution is unknown but is only characterized by its mean and variance. Indeed, such models have been discussed in the context of minimax stochastic optimization models (see Žáčková (1966), Dupacova (1987), Breton and El Hachem (1995), Shapiro and Kleywegt (2002), Shapiro and Ahmed (2004)), and recently in the context of distributionally robust optimization models (see Chen et al. (2007), Chen and Sim (2009), Popescu (2007), Delage and Ye (2010), Xu and Mannor (2012), Wiesemann et al. (2013)).

Many optimization problems involve dynamic decision makings in an environment where uncertainties are progressively unfolded in stages. Unfortunately, such problems often suffer from the “curse of dimensionality” and are typically computationally intractable (see Shapiro and Nemirovski (2005), Dyer and Stougie (2006), Ben-Tal et al. (2004)). One approach to circumvent the intractability is to restrict the dynamic or recourse decisions to being affinely dependent of the uncertain parameters, an approach known as linear decision rule. Linear decision rules appear in early literatures of stochastic optimization models but are abandoned due to their lack of optimality (see Garstka and Wets (1974)). The interest in linear decision rules is rekindled by Ben-Tal et al. (2004) in their seminal work that extends classical robust optimization to encompass recourse decisions. To further motivate linear decision rules, Bertsimas et al. (2010) establish the optimality of
linear decision rules in some important classes of dynamic optimization problems under full ambiguity. In more general classes of problems, Chen and Zhang (2009) improve the optimality of linear decision rules by extending linear decision rules to encompass affine dependency on the auxiliary parameters that are used to characterize the support set. Chen et al. (2007) also use linear decision rules to provide tractable solutions to a class of distributionally robust optimization problems with recourse. Henceforth, variants of linear and piecewise-linear decision rules have been proposed to improve the performance of more general classes of distributional robust optimization problems while maintaining the tractability of these problems. Such approaches include the deflected and segregated linear decision rules of Chen et al. (2008), the truncated linear decision rules of See and Sim (2009), and the bideflected and (generalized) segregated linear decision rules of Goh and Sim (2010). Interestingly, there is also a revival in decision rules for addressing stochastic optimization problems. Specifically, Kuhn et al. (2011) propose primal and dual linear decision rules techniques to solve multistage stochastic optimization problems that would also quantify the potential loss of optimality as the result of such approximations.

Despite the importance of addressing uncertainty in optimization problems, it is often ignored in practice due to the elevated complexity of modeling these problems compared to their deterministic counterparts. A practicable framework for optimization under uncertainty should also translate to viable software solutions that are potentially intuitive to the users and would enable them to focus on modeling issues and relieve them from the burden of algorithm tweaking and code troubleshooting. Software that facilitates robust optimization modeling have begun to surface in recent years. Existing toolboxes for robust optimization include YALMIP\textsuperscript{1}, AIMMS\textsuperscript{2} and ROME\textsuperscript{3}. Of which, ROME and AIMMS have provisions for decision rules and hence, they are capable of addressing dynamic optimization problems under uncertainty. AIMMS is a commercial software package that adopts the classical robust linear optimization framework where uncertainty is only characterized by the support set without distributional information. ROME is an algebraic modeling toolbox built in the MATLAB environment that implements the distributionally robust linear optimization framework of Goh and Sim (2010). Despite the polynomial tractability, the reformulation approach of Goh and Sim (2010) can be rather demanding, which could limit the scalability potential needed for addressing larger sized problems.

In this paper, we develop a new modular framework to obtain exact and approximate solutions to a class of linear optimization problems with recourse with the goal to minimize the worst-case expected objective over an ambiguity set of distributions. Our contributions to this paper are as follows:

1. We propose to focus on a standard ambiguity set where the family of distributions are characterized by linear and conic quadratic representable expectation constraints and the support set
is also linear and conic quadratic representable. As we will show, the standard ambiguity set has important ramifications on the tractability of the problem.

2. We adopt the approach of Wiesemann et al. (2013) to lift the original ambiguity set to an extended one by introducing additional auxiliary random variables. We show that by replacing the recourse decision functions with generalized linear decision rules that have affine dependency on the uncertain parameters and the auxiliary random variables, we can obtain good and sometimes tight approximations to a two-stage optimization problem. This approach is easy to compute, extends to a multistage problem and improves upon existing variants of linear decision rules developed in Chen and Zhang (2009), Chen et al. (2008), See and Sim (2009), Goh and Sim (2010).

3. We demonstrate the practicability of our framework by developing a new algebraic modeling package named ROC, a C++ library that implements the techniques developed in this paper.

**Notations.** Given a $N \in \mathbb{N}$, we use $[N]$ to denote the set of running indices, $\{1, \ldots, N\}$. We generally use bold faced characters such as $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$ to represent vectors and matrices. We use $[\mathbf{x}]_i$ or $x_i$ to denote the $i$ element of the vector $\mathbf{x}$. We use $(x)^+$ to denote $\max\{x, 0\}$. Special vectors include $\mathbf{0}$, $\mathbf{1}$ and $\mathbf{e}_i$, which are respectively the vector of zeros, the vector of ones and the standard unit basis vector. Given $N, M \in \mathbb{N}$, we denote $\mathcal{R}^{N,M}$ as the space of all measurable functions from $\mathbb{R}^N$ to $\mathbb{R}^M$ that are bounded on compact sets. For a proper cone $K \subseteq \mathbb{R}^L$ (i.e., a closed, convex and pointed cone with nonempty interior), we use the relations $\mathbf{x} \preceq_K \mathbf{y}$ or $\mathbf{y} \succeq_K \mathbf{x}$ to indicate that $\mathbf{y} - \mathbf{x} \in K$. Similarly, the relations $\mathbf{x} \prec_K \mathbf{y}$ or $\mathbf{y} \succ_K \mathbf{x}$ imply that $\mathbf{y} - \mathbf{x} \in \text{int}K$, where $\text{int}K$ represents the interior of the cone $K$. Meanwhile, $K^*$ is the dual cone of $K$ with $K^* = \{ \mathbf{y} : \mathbf{y}^T \mathbf{x} \geq 0, \mathbf{x} \in K \}$.

We use tilde to denote an uncertain or random parameter such as $\tilde{\mathbf{z}} \in \mathbb{R}^I$ without associating it with a particular probability distribution. We denote $\mathcal{P}_0(\mathbb{R}^I)$ as the set of all probability distributions on $\mathbb{R}^I$. Given a random vector $\tilde{\mathbf{z}} \in \mathbb{R}^I$ with probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I)$ and function $g \in \mathcal{R}^{I,P}$, we denote $\mathbb{E}_\mathbb{P}(g(\tilde{\mathbf{z}}))$ as the expectation of the random variable, $g(\tilde{\mathbf{z}})$ over the probability distribution $\mathbb{P}$. Similarly, for a set $\mathcal{W} \subseteq \mathbb{R}^I$, $\mathbb{P}(\tilde{\mathbf{z}} \in \mathcal{W})$ represents the probability of $\tilde{\mathbf{z}}$ being in the set $\mathcal{W}$ evaluated on the distribution $\mathbb{P}$. Suppose $\mathcal{Q} \in \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^L)$ is a joint probability distribution of two random vectors $\tilde{\mathbf{z}} \in \mathbb{R}^I$ and $\tilde{\mathbf{u}} \in \mathbb{R}^L$, then $\prod_{\tilde{\mathbf{z}}} \mathcal{Q} \in \mathcal{P}_0(\mathbb{R}^I)$ denotes the marginal distribution of $\tilde{\mathbf{z}}$ under $\mathcal{Q}$. Likewise, for a family of distributions, $\mathcal{G} \subseteq \mathcal{P}_0(\mathbb{R}^I \times \mathbb{R}^L)$, $\prod_{\tilde{\mathbf{z}}} \mathcal{G}$ represents the set of marginal distributions of $\tilde{\mathbf{z}}$ under all $\mathcal{Q} \in \mathcal{G}$, i.e. $\prod_{\tilde{\mathbf{z}}} \mathcal{G} = \{ \prod_{\tilde{\mathbf{z}}} \mathcal{Q} : \mathcal{Q} \in \mathcal{G} \}$.

2. **A two stage distributionally robust optimization problem**

In this section, we focus on a two-stage optimization problem where the first stage or here-and-now decision is a vector $\mathbf{x} \in \mathbb{R}^{N_1}$ chosen over the feasible set $X_1$. The cost incurred during the first stage in association with the decision $\mathbf{x}$ is deterministic and given by $c'\mathbf{x}$, $c \in \mathbb{R}^{N_1}$. In progressing
to the next stage, a vector of uncertain parameters $\tilde{z} \in \mathcal{W} \subseteq \mathbb{R}^{I_1}$ is realized; thereafter, we could determine the cost incurred at the second stage. Similar to a typical stochastic programming model, for a given decision vector, $x$ and a realization of the uncertain parameters, $z \in \mathcal{W}$, we evaluate the second stage cost via the following linear optimization problem,

$$Q(x, z) = \min \ {d'y}$$

s.t.  
$$A(z)x + By \geq b(z)$$  
$y \in \mathbb{R}^{N_2}$  

(1)

Here, $A \in \mathbb{R}^{I_1 \times M \times N_1}$, $b \in \mathbb{R}^{I_1 \times M}$ are functions that maps from the vector $z \in \mathcal{W}$ to the input parameters of the linear optimization problem. Adopting the common assumptions in robust optimization literatures, these functions are affinely dependent on $z \in \mathbb{R}^{I_1}$ and are given by,

$$A(z) = A^0 + \sum_{k \in [I_1]} A^k z_k,$$

$$b(z) = b^0 + \sum_{k \in [I_1]} b^k z_k,$$

with $A^0, A^1, ..., A^{I_1} \in \mathbb{R}^{M \times N_1}$ and $b^0, b^1, ..., b^{I_1} \in \mathbb{R}^{M}$. The matrix $B \in \mathbb{R}^{M \times N_2}$ and the vector $d \in \mathbb{R}^{N_2}$ are unaffected by the uncertainties, which corresponds to the case of fixed-recourse as defined in stochastic programming literatures.

The second stage decision (wait-and-see) is represented by the vector $y \in \mathbb{R}^{N_2}$, which is easily determined by solving a linear optimization problem after the uncertainty is realized. However, whenever the second stage problem is infeasible, we have $Q(x, z) = \infty$, and the first stage solution, $x$ would be rendered meaningless. As in the case of a standard stochastic programming model, $x$ has to be feasible in $X_1 \cap X_2$, where

$$X_2 = \{x \in \mathbb{R}^{N_1}: Q(x, z) < \infty \ \forall z \in \mathcal{W}\}.$$  

Unfortunately, checking the feasibility of $X_2$ is already NP-complete (see Ben-Tal et al. (2004)), hence, for simplicity, we focus on problems with relatively complete recourse, i.e.,

**Assumption 1.**

$$X_1 \subseteq X_2.$$  

In the context of stochastic programming, complete recourse refers to the characteristics of the recourse matrix, $B$ such that for any $t \in \mathbb{R}^M$, there exists $y \in \mathbb{R}^{N_2}$ such that $By \geq t$. Therefore, under complete recourse we have $X_2 = \mathbb{R}^{N_1}$.

**Model of uncertainty**

We assume that the probability distribution of $\tilde{z}$ belongs to an *ambiguity set*, $\mathcal{F}$ where the family of distributions are characterized in the following standardized framework,

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0 (\mathbb{R}^{I_1}) : \begin{align*}
\tilde{z} & \in \mathbb{R}^{I_1} \\
\mathbb{E}_\mathbb{P}(G\tilde{z}) & = \mu \\
\mathbb{E}_\mathbb{P}(g(\tilde{z})) & \leq \sigma \\
\mathbb{P}(\tilde{z} \in \mathcal{W}) & = 1
\end{align*} \right\}$$

(2)
with $G \in \mathbb{R}^{L_1 \times I_1}$, $\mu \in \mathbb{R}^{L_1}$, $\sigma \in \mathbb{R}^{L_2}$, $g \in \mathbb{R}^{I_1 \times L_2}$ is a convex conic quadratic representable function and $W$ is the support set, which is also is conic quadratic representable (we refer readers to Lobo et al. (1998), Alizadeh and Goldfarb (2003), Ben-Tal and Nemirovski (2001a) for references on second order conic optimization problems and conic quadratic representable functions and sets).

Hence, the epigraph of $g$ together with the support set $W$, i.e.,

$$\bar{W} = \left\{ (z, u) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} : g(z) \leq u \right\},$$

(3)

can be expressed as a projection from a possibly higher dimensional conic representable set. Here, the vector $u$ is associated with the additional dimensions needed to represent the epigraph of $g$. Observe that for all $z \in W$, $(z, g(z)) \in \bar{W}$. In particular, the explicit formulation of $\bar{W}$ is given by

$$\bar{W} = \left\{ (z, u, v) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} : \exists v \in \mathbb{R}^{I_3}, (z, u, v) \in \hat{W} \right\},$$

(4)

where we define $\hat{W}$ as the extended support set,

$$\hat{W} = \left\{ (z, u, v) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3} : Cz + Du + Ev \preceq_K h \right\},$$

with $C \in \mathbb{R}^{I_3 \times I_1}$, $D \in \mathbb{R}^{I_3 \times I_2}$, $E \in \mathbb{R}^{I_3 \times I_3}$, $h \in \mathbb{R}^{I_3}$ and $K \subseteq \mathbb{R}^{I_3}$ being a Cartesian product of second order cones and nonnegative orthants. The vector $v$ is the auxiliary variables associated with the conic reformulation of the epigraph of $g$ as well as the representation of the support set $W$. Hence, we can partition $[I_3]$ into two disjoint subsets $\mathcal{I}_3$, $\bar{\mathcal{I}}_3$, $\mathcal{I}_3 \cup \bar{\mathcal{I}}_3 = [I_3]$ such that $v_i$, $i \in \mathcal{I}_3$ are the auxiliary variables associated with epigraph of $g$ while $v_i$, $i \in \bar{\mathcal{I}}_3$ are those associated with the support set $W$. Note that for all $z \in W$ there exists $v \in \mathbb{R}^{I_3}$ such that $(z, g(z), v) \in \hat{W}$. Correspondingly, there also exists a function, $\nu \in \mathcal{R}^{I_1 \times I_3}$ that satisfies $(z, g(z), \nu(z)) \in \hat{W}$ for all $z \in W$. The formulation of the extended set is an important concept to obtain a tractable computational format in the form linear or second order conic optimization problems that can be solved efficiently using state-of-the-art commercial solvers such as CPLEX and Gurobi. We provide an explicit example using standard second order conic reformulation techniques as follows:

**Example 1.** The extended support set for

$$W = \left\{ z \in \mathbb{R}^{I_1} : \|z\|_1 \leq \Gamma, \|z\|_\infty \leq 1 \right\}$$

$$g_1(z) = |a_1^T z|$$

$$g_2(z) = |a_2^T z|^2$$

$$g_3(z) = ((a_3^T z)^3)$$

$$g_4(z) = \min\{w^T v : Hv \succeq_K f + Fz\}$$
is given by

\[ \mathcal{W} = \left\{ \left( z, u, \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \right) \in \mathbb{R}^{I_1} \times \mathbb{R}^4 \times \mathbb{R}^{I_3} : \right. \]

\[
\begin{align*}
& a_1 z \leq u_1, -a_1 z \leq u_1 \\
& \sqrt{\left( a_2' z \right)^2 + \left( \frac{u_2 - 1}{2} \right)^2} \leq \frac{u_2 + 1}{2} \\
& v_1 \geq 0, v_1 \geq a_3' z \\
& \sqrt{v_2^2 + \left( \frac{v_3 - 1}{2} \right)^2} \leq \frac{v_3 + 1}{2} \\
& \sqrt{v_3^2 + \left( \frac{u_4 - u_3}{2} \right)^2} \leq \frac{u_4 + u_3}{2} \\
& w' v_3 \leq u_4 \\
& Hv_3 \geq \kappa f + F z \\
& v_4 \geq z, v_4 \geq -z \\
& 1' v_4 \leq \Gamma, -1 \leq z \leq 1
\end{align*}
\]

Given \( z \in \mathcal{W} \), we can verify that \( u_1 = |a_1' z| \), \( u_2 = (a_2' z)^2 \), \( u_3 = (a_3' z)^3 \), \( u_4 = \min \{ d' v : H v \geq f + F z \} \), \( v_1 = (a_3' z)^+ \), \( v_2 = (a_3' z)^+ \), \( v_3 = \arg \min \{ d' v : H v \geq f + F z \} \), \( v_4 = (|z_1|, \ldots, |z_{I_1}|)' \) would be feasible in the extended support set \( \hat{W} \). Moreover, \( v_1, v_2, v_3 \) are those associated with the epigraphs of \( g_3 \) and \( g_4 \), while \( v_4 \) is related to the support set, \( W \).

For computational reasons, we impose the following Slater’s like conditions:

**Assumption 2.** There exists \( (z^\dagger, u^\dagger, v^\dagger) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3} \) such that

\[
G z^\dagger = \mu \\
\langle u^\dagger, \sigma \rangle < 0 \\
C z^\dagger + D u^\dagger + E v^\dagger \prec_K h.
\]

Hence, \( (z^\dagger, u^\dagger, v^\dagger) \in \text{int}\mathcal{W} \).

The standardized ambiguity set framework might appear to be rather restrictive, for instance, it is incapable of characterizing higher moment information such as covariance, i.e., \( \mathbb{E}_\rho(z_i z_j) = \sigma_{ij} \), \( i, j \in [I_1] \). While it is possible to extend the ambiguity set to include semidefinite constants to capture covariance information (see Wiesemann et al. (2013)), we avoid such considerations due to the practical difficulties of solving semidefinite optimization problems. Despite the limitations, the standardized ambiguity set is computationally attractive and has the potential of characterizing the dispersion of \( \tilde{z} \) such as location and spread among others. The location of the \( \tilde{z} \) can be specified using the affine expectation constraint while the dispersion can be characterized by bounding the expectation of multivariate convex functions over \( \tilde{z} \). If the location is uncertain and bounded within a range, such as \( \mathbb{E}_\rho(G \tilde{z}) \in [\mu, \mu] \), we can easily enforce this by bounding the expectation of appropriate affine functions over \( \tilde{z} \). There are various ways of providing information on spread. For a given location \( \mu \) and a vector \( a \in \mathbb{R}^{I_1} \) representing the direction of spread, one may impose \( \mathbb{E}_\rho(U(a' (\tilde{z} - \mu))) \leq \sigma \), where \( U \in \mathbb{R}^{1, I_1} \) is a convex conic quadratic representable function that penalizes nonzero values such as \( U(z) = |z|^{a/b}, a, b \in \mathbb{N}, a \geq b \geq 1 \). We could also impose constraints on the expectations of distances from the location \( \mu \) measured using norm functions. For instance,
given a conic quadratic representable norm function \( \| \cdot \| \), we can impose \( \mathbb{E}_P (U(\|\tilde{z} - \mu\|)) \leq \sigma \) to limit the spread of \( \tilde{z} \) away from \( \mu \). More generally, the shape of the dispersion may also be characterized through the function \( g \), which may be inspired from statistics or possibly from the structure of the optimization problem. We will leave these explorations to future research as the purpose of this paper is to provide the optimization framework as well as the software that we could use to facilitate future studies.

Given the ambiguity set, \( F \), we assume that the decision maker is ambiguity averse and the second stage cost is evaluated based on the worst case expectation over the ambiguity set given by

\[
\beta(x) = \sup_{P \in F} \mathbb{E}_P (Q(x, \tilde{z})).
\]  

Correspondingly, the here-and-now decision is determined by minimizing the sum of the deterministic first stage cost and the worst-case expected second stage cost over the ambiguity set as follows:

\[
\min c'x + \beta(x) \\
\text{s.t. } x \in X_1.
\]  

More generally, the second stage can involve a collection of \( K \) attributes \( \beta_k(x), k \in [K] \), each having similar structure as \( \beta(x) \) and the generalized model we solve is as follows:

\[
Z^* = \min c'x + \beta(x) \\
\text{s.t. } c_k'x + \beta_k(x) \leq \rho_k \forall k \in [K] \\
x \in X_1,
\]  

with \( c_k \in \mathbb{R}^{N_1}, k \in [K] \) and \( \rho \in \mathbb{R}^K \). For simplicity, we will focus on deriving the exact reformulation of \( \beta(x) \), which could then be integrated in Problem (6) to obtain the optimum here-and-now decision, \( x \in X_1 \). Naturally, similar reformulations can be extended to derive the epigraphs of \( \beta_k(x), k \in [K] \), which could be incorporated into Problem (7) to obtain a tractable optimization problem.

Observe that Problem (5) involves optimization of probability measures over a family of distributions and hence, it is not a finite dimensional optimization problem. Motivated from Wiesemann et al. (2013), we define the extended ambiguity set, \( G \) which involves auxiliary random variables over the extended support set \( \hat{W} \) as follows:

\[
G = \left\{ Q \in \mathcal{P}_0 \left( \mathbb{R}^{l_1} \times \mathbb{R}^{l_2} \times \mathbb{R}^{l_3} \right) : \begin{align*}
(\tilde{z}, \tilde{u}, \tilde{v}) &\in \mathbb{R}^{l_1} \times \mathbb{R}^{l_2} \times \mathbb{R}^{l_3} \\
\mathbb{E}_Q(G\tilde{z}) &= \mu \\
\mathbb{E}_Q(\tilde{u}) &\leq \sigma \\
Q((\tilde{z}, \tilde{u}, \tilde{v}) \in \hat{W}) &= 1 \end{align*} \right\}.
\]  

**Proposition 1.** The ambiguity set \( F \) is equivalent to the set of marginal distributions of \( \tilde{z} \) under \( Q \), for all \( Q \in G \), i.e.,

\[
F = \prod_i G.
\]
In particular, for a function \( \nu \in \mathbb{R}^{I_1,I_2} \) satisfying \((z,g(z),\nu(z)) \in \hat{W}\) for all \( z \in \mathcal{W} \) and \( P \in \mathcal{P} \), the probability distribution \( Q \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}) \) associated with the random variable \((\tilde{z},\tilde{u},\tilde{v}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}\) such that

\[
(\tilde{z},\tilde{u},\tilde{v}) = (\tilde{z},g(\tilde{z}),\nu(\tilde{z})) \quad P\text{-a.s.}
\]

also lies in \( \mathcal{G} \).

**Proof.** The proof is rather straightforward and a variant is presented in Wiesemann et al. (2013). We first show that \( \prod_I G \subseteq P \). Indeed, for any \( Q \in \mathcal{G} \), and \( P = \prod_I Q \), we have \( E_P(G \tilde{z}) = E_Q(G \tilde{z}) = \mu \). Moreover, since \( Q((\tilde{z},\tilde{u},\tilde{v}) \in \hat{W}) = 1 \), we have \( Q(\tilde{z} \in \mathcal{W}) = 1 \) and \( Q(g(\tilde{z}) \leq \tilde{u}) = 1 \). Hence, \( P(\tilde{z} \in \mathcal{W}) = 1 \) and

\[
E_P(g(\tilde{z})) = E_Q(g(\tilde{z})) \leq E_Q(\tilde{u}) \leq \sigma.
\]

Conversely, suppose \( P \in \mathcal{P} \), we observe that \( P((\tilde{z},g(\tilde{z})) \in \hat{W}) = 1 \). Since \((z,g(z),\nu(z)) \in \hat{W}\) for all \( z \in \mathcal{W} \), we can then construct a probability distribution \( Q \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}) \) associated with the random variable \((\tilde{z},\tilde{u},\tilde{v}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}\) so that

\[
(\tilde{z},\tilde{u},\tilde{v}) = (\tilde{z},g(\tilde{z}),\nu(\tilde{z})) \quad P\text{-a.s.}
\]

Observe that

\[
E_Q(\tilde{u}) = E_P(g(\tilde{z})) \leq \sigma
\]

and

\[
Q((\tilde{z},\tilde{u},\tilde{v}) \in \hat{W}) = 1.
\]

Hence, \( P \subseteq \prod_I G \). \( \Box \)

**Exact reformulation**

Before we derive an exact reformulation for evaluating \( \beta(x) \), \( x \in X_1 \), we first present the result to compute the worse case expectation of a piecewise linear convex function.

**Theorem 1.** Let \( U \in \mathbb{R}^{I_1,1} \) be a piecewise linear convex function given by

\[
U(z) = \max_{p \in [P]} \{ \zeta_p' \tilde{z} + c_p^0 \}
\]

for some \( \zeta_p \in \mathbb{R}^{I_1}, c_p^0 \in \mathbb{R} \), \( p \in [P] \). Suppose

\[
\beta^* = \sup_{P \in \mathcal{P}} E_P(U(\tilde{z}))
\]

is finite, then it can expressed as a standard robust counter problem

\[
\beta^* = \min r + s' \mu + t' \sigma
\]

s.t. \( r + s'(Gz) + t'u \geq U(z) \quad \forall (z,u,v) \in \hat{W} \)

\[
t \geq 0
\]

\[
r \in \mathbb{R}, s \in \mathbb{R}^{I_1}, t \in \mathbb{R}^{I_2}
\]

(9)
or equivalently

\[
\beta^* = \min r + s^\prime \mu + t^\prime \sigma \\
\text{s.t. } r \geq \pi_p^\prime h + \zeta_p^0 \quad \forall p \in [P] \\
C' \pi_p = \zeta_p - G' s \\
D' \pi_p = -t \\
E' \pi_p = 0, \\
\pi_p \succeq_\mathcal{K} 0 \\
t \geq 0
\]

(10)

Moreover, the optimization problem is also solvable.

**Proof.** From Proposition 1, we have equivalently

\[
\beta^* = \sup_{\mathcal{P} \in \mathcal{F}} \mathbb{E}_p \left( \max_{p \in [P]} \{ \zeta_p^{' \prime} \tilde{z} + \zeta_p^0 \} \right) = \sup_{\mathcal{Q} \in \mathcal{G}} \mathbb{E}_p \left( \max_{p \in [P]} \{ \zeta_p^{' \prime} \tilde{z} + \zeta_p^0 \} \right).
\]

By weak duality (referring to Isii (1962)), we have the following semi-infinite optimization problem

\[
\beta^* \leq \beta_1^* = \inf r + s^\prime \mu + t^\prime \sigma \\
\text{s.t. } r \geq \sup_{(z, u, v) \in \hat{\mathcal{W}}} \{(\zeta_p - G' s)' z - t' u + \zeta_p^0 \} \forall p \in [P] \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{l_1}, t \in \mathbb{R}^{l_2},
\]

where \( r, s \in \mathbb{R}^{l_1}, t \in \mathbb{R}^{l_2} \) are the dual variables corresponding to the expectation constraints of \( \mathcal{G} \). This is also equivalent to

\[
\beta_1^* = \inf r + s^\prime \mu + t^\prime \sigma \\
\text{s.t. } r \geq \sup_{(z, u, v) \in \hat{\mathcal{W}}} \{(\zeta_p - G' s)' z - t' u + \zeta_p^0 \} \forall p \in [P] \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{l_1}, t \in \mathbb{R}^{l_2},
\]

(11)

By weak conic duality (see, for instance, Ben-Tal and Nemirovski (2001a)), we have for all \( p \in [P] \),

\[
\sup_{(z, u, v) \in \hat{\mathcal{W}}} \{(\zeta_p - G' s)' z - t' u + \zeta_p^0 \} \leq \inf \pi_p^\prime h + \zeta_p^0 \\
\text{s.t. } C' \pi_p = \zeta_p - G' s \\
D' \pi_p = -t \\
E' \pi_p = 0 \\
\pi_p \succeq_\mathcal{K} 0 \\
\pi_p \in \mathbb{R}^{l_3} \quad p \in [P],
\]

where \( \pi_p \in \mathbb{R}^{l_3}, \forall p \in [P] \) are the dual variables associated with the conic constants in \( \hat{\mathcal{W}} \). Hence, using standard robust counterpart techniques, we substitute the dual formulations in Problem (11)
to yield the following compact conic optimization problem

\[
\beta_3^* = \inf \ r + s'\mu + t'\sigma \\
\text{s.t. } r \geq \pi_p'h + c_p^0 \quad \forall p \in \mathcal{P} \\
C'\pi_p = \zeta_p' - G's \quad \forall p \in \mathcal{P} \\
D'\pi_p = -t \quad \forall p \in \mathcal{P} \\
E'\pi_p = 0 \quad \forall p \in \mathcal{P} \\
\pi_p \succeq \chi*0 \quad \forall p \in \mathcal{P} \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2} \\
\pi_p \in \mathbb{R}^{L_3} \quad \forall p \in \mathcal{P}.
\]  

(12)

Observe that \(\beta^* \leq \beta_1^* \leq \beta_2^*\), and our goal is to establish strong duality by showing \(\beta_2^* \leq \beta^*\). Then we will next approach Problem (12) by taking the dual, which is

\[
\beta_3^* = \sup \sum_{p \in \mathcal{P}} (\zeta_p^0, \pi_p) \quad \forall p \in \mathcal{P} \\
\text{s.t. } \sum_{p \in \mathcal{P}} \alpha_p = 1 \\
\alpha_p \geq 0 \quad \forall p \in \mathcal{P} \\
\sum_{p \in \mathcal{P}} G\pi_p = \mu \\
\sum_{p \in \mathcal{P}} \bar{u}_p \leq \sigma \\
C\pi_p + D\bar{u}_p + E\bar{v}_p \preceq \chi \alpha_p h \quad \forall p \in \mathcal{P} \\
\alpha_p \in \mathbb{R}, \pi_p \in \mathbb{R}^{L_1}, \quad \forall p \in \mathcal{P} \\
\bar{u}_p \in \mathbb{R}^{L_2}, \bar{v}_p \in \mathbb{R}^{L_3} \quad \forall p \in \mathcal{P}.
\]  

(13)

Suppose \((z^1, u^1, v^1) \in \mathbb{R}^{L_1} \times \mathbb{R}^{L_2} \times \mathbb{R}^{L_3}\) satisfy the conditions in Assumption 2, then we can construct a strictly feasible solution

\[
\alpha_p = \frac{1}{P}, \bar{z}_p = \frac{z^1}{P}, \bar{u}_p = \frac{u^1}{P}, \bar{v}_p = \frac{v^1}{P},
\]

for all \(\forall p \in \mathcal{P}\). Hence, since Problem (13) is strictly feasible and, as we will subsequently show, is also bounded from above, strong duality holds and \(\beta_2^* = \beta_3^*\). Moreover, there exists a sequence of strictly feasible or interior solutions

\[
\left\{ (\alpha_p^k, \bar{z}_p^k, \bar{u}_p^k, \bar{v}_p^k)_{p \in \mathcal{P}} \right\}_{k \geq 0}
\]

such that

\[
\lim_{k \to \infty} \sum_{p \in \mathcal{P}} (\zeta_p^0 \alpha_p^k + \zeta_p'^{k} \bar{z}_p^k) = \beta_3^*.
\]

Observe that for all \(k\), \(\alpha_p^k > 0\), \(\sum_{p \in \mathcal{P}} \alpha_p^k = 1\) and we can construct a sequence of discrete probability distributions \(\{Q_k \in \mathcal{P}_0(\mathbb{R}^{L_1} \times \mathbb{R}^{L_2} \times \mathbb{R}^{L_3})\}_{k \geq 0}\) on random variable \((\bar{z}, \bar{u}, \bar{v}) \in \mathbb{R}^{L_1} \times \mathbb{R}^{L_2} \times \mathbb{R}^{L_3}\) such that

\[
Q_k \left( \frac{\bar{z}_p^k}{\alpha_p^k}, \frac{\bar{u}_p^k}{\alpha_p^k}, \frac{\bar{v}_p^k}{\alpha_p^k} \right) = \alpha_p^k \quad \forall p \in \mathcal{P}.
\]
Observe that,
\[ E_Q(k(G \tilde{z})) = \mu, E_Q(k(\tilde{u})) \leq \sigma, \mathcal{Q}_k((\tilde{z}, \tilde{u}, \tilde{v}) \in \hat{V}) = 1, \]
and hence \( \mathcal{Q}_k \in \mathcal{G} \) for all \( k \). Moreover,
\[
\beta^*_1 = \lim_{k \to \infty} \sum_{p \in [P]} (c^0_p \alpha_p^k + \zeta_p^k \tilde{z}^k_p)
\]
\[
= \lim_{k \to \infty} \sum_{p \in [P]} \alpha_p^k \left( \zeta_p^0 + \zeta_p^k \tilde{z}^k_p \right)
\]
\[
\leq \lim_{k \to \infty} \sum_{p \in [P]} \alpha_p^k \max_{q \in [P]} \left\{ \zeta_q^0 + \zeta_q^k \tilde{z}^k_p / \alpha_p^k \right\}
\]
\[
= \lim_{k \to \infty} E_{Q_k} \left( \max_{q \in [P]} \left\{ \zeta_q^0 + \zeta_q^k \tilde{z}^k \right\} \right)
\]
\[
\leq \sup_{Q \in \mathcal{G}} E_{Q}(U(\tilde{z}))
\]
\[
= \beta^*.
\]
Hence, \( \beta^* \leq \beta^*_1 \leq \beta^*_2 = \beta^*_3 \leq \beta^* \), and strong duality holds. Since \( \beta^* \) is finite, Problem (13) is bounded from above and hence, Problem (12) also solvable. Since the cone \( \mathcal{K} \) is a Cartesian product of second order cones and nonnegative orthants, it is also self dual, i.e. \( \mathcal{K}^* = \mathcal{K} \).

Noting that \( Q(x, z) \), \( x \in X_1 \) is also a piecewise linear convex function of \( z \), we can easily extend Theorem 1 so that the function \( \beta(x) \) can be evaluated and integrated in epigraphical form to solve Problem (7) as a standard optimization problem.

**Theorem 2.** Let \( \{p^1, ..., p^P\} \) be the set of all extreme points of the polyhedra
\[
\mathcal{P} = \left\{ p \in \mathbb{R}^M : B'p = d, p \geq 0 \right\}.
\]
For a given subset of extreme points indices, \( S \subseteq [P] \), we define
\[
\beta_S(x) = \min_{r + s'\mu + t'\sigma} \quad \text{s.t.} \quad r \geq \pi_i'h + p_i'b^0 - p_i'A^0x \quad \forall i \in S
\]
\[
C'\pi_i = \begin{bmatrix} \vdots \\ p'_i(b^1 - A^1x) \end{bmatrix} - G's \quad \forall i \in S
\]
\[
D'\pi_i = -t \quad \forall i \in S
\]
\[
E'\pi_i = 0 \quad \forall i \in S
\]
\[
\pi_i \geq \kappa 0 \quad \forall i \in S
\]
\[
t \geq 0
\]
\[
r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2}
\]
\[
\pi_i \in \mathbb{R}^{L_3} \quad \forall i \in S.
\]
If \( \beta(x) \), \( x \in X_1 \) is finite, then
\[
\beta_S(x) \leq \beta_{[P]}(x) = \beta(x).
\]
Proof. From strong linear optimization duality, we can express Problem (1) as

\[
Q(x, z) = \max_{p \in P} p'(b(z) - A(z)x)
\]

s.t. \[ p \in P. \tag{15} \]

Since \(Q(x, z)\) is finite for all \(x \in X_1\) (Assumption 1), Problem (15) has an extreme point optimum solution for all \(x \in X_1\). Hence, we can express Problem (15) explicitly as a piecewise linear convex function of \(z\) as follows:

\[
Q(x, z) = \max_{i \in [P]} \{ p_i'(b(z) - A(z)x) \},
\]

for all \(x \in X_1\). Since \(\beta(x)\) is finite, we can use the result of Theorem 1 to derive the exact reformulation for \(S = [P]\), to achieve \(\beta(x) = \beta_{[P]}(x)\). It is trivial to see that if \(S_1 \subseteq S_2 \subseteq [P]\), then

\[
\beta_{S_1}(x) \leq \beta_{S_2}(x) \leq \beta_{[P]}(x).
\]

\(\Box\)

Remark: Note that although a similar result has been derived in Wiesemann et al. (2013), the condition for strong duality is not presented as we have done so in Theorem 1.

Theorem (2) suggests an approach to compute the exact value of \(\beta(x)\), which may not be a polynomial sized problem due to possibly exponential number of extreme points. Unfortunately, the ”separation problem” associated with finding the extreme point involves solving the following bilinear optimization problem,

\[
\max_{p \in F} \left\{ \sup_{(x, u) \in \mathcal{W}} \left( \begin{bmatrix} p'(b^1 - A^1x) \\ \vdots \\ p'(b^I - A^Ix) \end{bmatrix} - G's \right)' z + p'(b^0 - A^0x) - t'u \right\}
\]

which is generally intractable. Nevertheless, Theorem (2) provides an approach to determine the lower bound of \(\beta(x)\), which might be useful to determine the quality of the solution. We will next show how we can tractably compute the upper bound of \(\beta(x)\) via linear decision rule approximations.

3. Generalized linear decision rules

Observe that any function, \(y \in \mathcal{R}^{f_1 \cdot N_2}\) satisfying

\[
A(z)x + By(z) \geq b(z) \quad \forall z \in \mathcal{W}
\]

would be an upper bound of \(\beta(x)\), \(x \in X_1\), i.e.,

\[
\beta(x) \leq \sup_{p \in F} \mathbb{E}_p (d'y(\tilde{z})).
\]
Moreover, equality is achieved if
\[ y(z) \in \arg\min \{ d'y : A(z)x + By \geq b(z) \} \]
for all \( z \in W \). Hence, we can express \( \beta(x) \) as a minimization problem over all measurable functions as follows:
\[
\beta(x) = \min \sup_{P \in F} \mathbb{E}_P(d'y(\tilde{z}))
\]
\[
s.t. \quad A(z)x + By(\tilde{z}) \geq b(\tilde{z}) \quad \forall \tilde{z} \in W
\]
for all \( x \in X \). Hence, we can express \((x), x \in X \) as a minimization problem over all measurable functions as follows:
\[
(x) = \min \sup_{P \in F} \mathbb{E}_P(d'y(\tilde{z}))\]
\[
s.t. \quad A(\tilde{z})x + By(\tilde{z}) \geq b(\tilde{z}) \quad \forall \tilde{z} \in W
\]
Unfortunately, Problem (16) is generally an intractable optimization problem as there could potentially be infinite number of constraints and variables. An upper bound of \((x) \) could be computed tractably by restricting \( y \) to a smaller class of measurable functions that can be characterized by a polynomial number of decision variables such as those that are affinely dependent on \( z \) or so called linear decision rules as follows:
\[
y(z) = y^0 + \sum_{j \in [I_1]} y_j z_j,
\]
for some \( y^0, y_j \in \mathbb{R}^{N_2}, j \in [I_1] \). However, the following example shows that linear decision rule may even be infeasible in problems with complete recourse.

**Example 2.** Consider the following complete recourse problem,
\[
\beta = \min \sup_{P \in F} \mathbb{E}_P(y(\tilde{z}))
\]
\[
s.t. \quad y(z) \geq z \quad \forall z \in \mathbb{R}
\]
\[
y(z) \geq -z \quad \forall z \in \mathbb{R}
\]
\[
y(z) \in \mathbb{R}^{1,1}
\]
where
\[
F = \{ P \in \mathcal{P}_0(\mathbb{R}) : \mathbb{E}_P(|\tilde{z}|) \leq 1 \}.
\]
Clearly, \( y(z) = |z| \) is the optimal decision rule that yields \( \beta = 1 \). However, under linear decision here (i.e., \( y(z) = y_0 + y_1 z \) for some \( y_0, y_1 \in \mathbb{R} \), we would encounter the following infeasibility issue
\[
y_0 + y_1 z \geq z \quad \forall z \in \mathbb{R}
\]
\[
y_0 + y_1 z \geq -z \quad \forall z \in \mathbb{R}
\]
Using the extended ambiguity set \( \mathcal{G} \), we propose the following generalized linear decision rule to encompass the auxiliary random variables \( \tilde{u} \) and \( \tilde{v} \) as well. For given subsets \( S_1 \subseteq [I_1], S_2 \subseteq [I_2], S_3 \subseteq [I_3] \), we define the following space of affine functions,
\[
\mathcal{L}^N(S_1, S_2, S_3) = \left\{ y : \mathbb{R}^{l_1} \times \mathbb{R}^{l_2} \times \mathbb{R}^{l_1} \to \mathbb{R}^N \mid \exists y^0, y_1^j, y_2^k, y_3^l \in \mathbb{R}^N, \forall i \in S_1, j \in S_2, k \in S_3 : y(z, u, v) = y^0 + \sum_{i \in S_1} y_1^i z_i + \sum_{j \in S_2} y_2^j u_j + \sum_{k \in S_3} y_3^k v_k \right\}.
\]
This decision rule generalizes the traditional linear decision rules that depends only on the underlying uncertainty, \( \tilde{z} \), in which case, we have \( S_2 = S_3 = \emptyset \). The segregated and extended linear decision
rules found in Chen and Zhang (2009), Chen et al. (2008), Goh and Sim (2010) are special cases of having \( S_3 \subseteq \tilde{I}_3 \), which incorporate auxiliary variables of the support set in the generalized linear decision rule. Based in the generalized linear decision rules, we obtain an upper bound of \( \beta(x) \), \( x \in X_1 \) as follows:

\[
\bar{\beta}(s_1, s_2, s_3)(x) = \min \sup_{E_{\mathcal{G}}} \mathbb{E}(d' y(\tilde{z}, \tilde{u}, \tilde{v})) \\
\text{s.t.} A(z)x + B y(z, u, v) \geq b(z) \quad \forall (z, u, v) \in \mathcal{W} \\
y \in \mathcal{L}^{N_z}(S_1, S_2, S_3). \tag{19}
\]

As the linear decision rule incorporates more auxiliary random variables, the quality of the bound improves, albeit at the expense of increased model size.

**Proposition 2.** Given \( x \in X_1 \), and \( S_1 \subseteq \tilde{S}_1 \subseteq [I_1] \), \( S_2 \subseteq \tilde{S}_2 \subseteq [I_2] \), and \( S_3 \subseteq \tilde{S}_3 \subseteq [I_3] \), we have

\[
\beta(x) \leq \bar{\beta}([I_1], [I_2], [I_3])(x) \leq \bar{\beta}(s_1, s_2, s_3)(x) \leq \bar{\beta}(s_1, s_2, s_3)(x). 
\]

**Proof.** The proof is trivial and hence omitted. \( \square \)

**Proposition 3.** For \( x \in X_1 \), Problem (19) is equivalent to the following robust counterpart problem,

\[
\bar{\beta}(s_1, s_2, s_3)(x) = \min r + s' \mu + t' \sigma \\
\text{s.t.} r + s'(G z) + t'u \geq d' y(z, u, v) \quad \forall (z, u, v) \in \mathcal{W} \\
A(z)x + B y(z, u, v) \geq b(z) \quad \forall (z, u, v) \in \mathcal{W} \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{l_1}, t \in \mathbb{R}^{l_2} \\
y \in \mathcal{L}^{N_z}(S_1, S_2, S_3). \tag{20}
\]

or explicitly as

\[
\bar{\beta}(s_1, s_2, s_3)(x) = \min r + s' \mu + t' \sigma \\
\text{s.t.} r - d' y^0 \geq \tau' h \\
\begin{align*}
[C' \pi]_i &= d' y^1_i - [G' s]_i & \forall i \in S_1 \\
[C' \pi]_i &= -[G' s]_i & \forall i \in [I_1] \setminus S_1 \\
[D' \pi]_j &= d' y^2_j - [t]_j & \forall j \in S_2 \\
[D' \pi]_j &= -[t]_j & \forall j \in [I_2] \setminus S_2 \\
[E' \pi]_k &= d' y^3_k & \forall k \in S_3 \\
[E' \pi]_k &= 0 & \forall k \in [I_3] \setminus S_3 \\
\begin{bmatrix} A^0 x + B y^0 - b^0 \end{bmatrix} & \geq \tau' h & \forall l \in [M] \\
[C' \tau]_l &= [b^0 - A^0 x - B y^1]_l & \forall l \in [M], \forall i \in S_1 \\
[C' \tau]_l &= [b^0 - A^0 x]_l & \forall l \in [M], \forall i \in [I_1] \setminus S_1 \\
[D' \tau]_l &= [B y^2]_l & \forall l \in [M], \forall j \in S_2 \\
[D' \tau]_l &= 0 & \forall l \in [M], \forall j \in [I_2] \setminus S_2 \\
[E' \tau]_k &= [B y^3]_l & \forall l \in [M], \forall k \in S_3 \\
[E' \tau]_k &= 0 & \forall l \in [M], \forall k \in [I_3] \setminus S_3 \\
\pi & \succeq K 0 \\
\tau_l & \succeq K 0 & \forall l \in [M] \\
r \in \mathbb{R}, s \in \mathbb{R}^{l_1}, t \in \mathbb{R}^{l_2} \\
\pi, \tau_l \in \mathbb{R}^{l_2}, \forall l \in [M]. \tag{21}
\end{align*}
\]
Proof. The proof follows from Theorem 1 and hence omitted. □

In Example (2), we show that a linear decision rule that depends solely on \( \tilde{z} \) may become infeasible if the support is unbounded. Suppose, the absolute deviations of \( \tilde{z} \) are bounded, we show that there exists a generalized linear decision rule involving the axillary random variable \( \tilde{u} \) that could resolve the infeasibility issue.

**Theorem 3.** Suppose Problem (19) has complete recourse, then exists a generalized linear decision rule

\[
y \in \mathcal{L}^N(\emptyset, [I_1], \emptyset),
\]

that is feasible in Problem (20) for the following family of distributions with bounded absolute deviations

\[
F_1 = \left\{ P \in \mathcal{P}_0(\mathbb{R}^{I_1}) : \mathbb{E}_P(|z_i|) \leq \sigma, \forall i \in [I_1] \right\}, \quad \sigma > 0.
\]

**Proof.** The extended ambiguity set associate with \( F_1 \) is

\[
G_1 = \left\{ Q \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_1}) : \mathbb{E}_Q(\tilde{u}) \leq \sigma, Q((\tilde{z}, \tilde{u}) \in \tilde{W}) = 1 \right\},
\]

in which the extended support set is \( \tilde{W} = \{(z, u) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_1} : u \geq z, u \geq -z \} \). The linear decision rule \( y \in \mathcal{L}^N(\emptyset, [I_1], \emptyset) \) is given by

\[
y(u) = y^0 + \sum_{i \in [I_1]} y^2_i u_i.
\]

Using these parameters, we need to show that the linear decision rule \( y(u) \) is feasible in the following problem,

\[
\begin{align*}
\min & \quad r + t' \sigma \\
\text{s.t.} & \quad r + t'u \geq d'y^0 + \sum_{i \in I_1} d^i y^2_i u_i, \forall (z, u) \in \tilde{W} \\
& \quad By^0 + \sum_{i \in I_1} B y^2_i u_i \geq b^0 - A^0 x + \sum_{i \in I_1} (b^i - A^i x) z_i, \forall (z, u) \in \tilde{W} \\
& \quad t \geq 0 \\
& \quad r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2} \\
& \quad y^0 \geq 0 \\
& \quad y^0, y^2_i \in \mathbb{R}^{N_2} \quad i \in [I_1].
\end{align*}
\]

(22)

Since \( B \) is complete recourse matrix, there exists \( \tilde{y}^0, \tilde{y}^2_i \) \( i \in [I_1] \), such that

\[
By^0 \geq b^0 - A^0 x,
\]

\[
By^2_i \geq (b^i - A^i x), \quad By^2_i \geq -(b^i - A^i x) \quad \forall i \in [I_1].
\]

Observe that given any \( a \in \mathbb{R}, b \in \mathbb{R}^{I_1} \)

\[
x + y'u \geq a + b'z \quad \forall (z, u) \in \tilde{W}
\]
if \( x \geq a \), and \( y_i \geq |b_i|, i \in [I_1] \). Hence, a feasible solution for Problem (22) would be

\[
\begin{align*}
  r &= d' \bar{y}^0 \\
  t_i &= \max\{d' \bar{y}^j, 0\} \; \forall i \in [I_1] \\
  y^j &= \bar{y}^j \\
  \forall j \in \{0\} \cup [I_1]
\end{align*}
\]

\( \square \)

The generalized linear decision rule achieves exact the value of \( \beta(x) \) for the following instance.

**Theorem 4.** For a complete recourse problem with \( N_2 = 1 \) and finite \( \beta(x) \), we have

\[
\beta(x) = \tilde{\beta}_{(I_1, [I_2], \emptyset)}(x).
\]

**Proof.** For \( N_2 = 1 \), the complete recourse matrix \( B \in \mathbb{R}^{M \times 1} \) must satisfy either \( B > 0 \) or \( B < 0 \).

Observe that the problem

\[
Q(x, z) = \min dy \\
\text{s.t.} \quad A(z)x + By \geq b(z) \; \forall (z, u, v) \in \hat{W} \\
y \in \mathbb{R},
\]

is unbounded below whenever \( dB < 0 \). Since \( \beta(x) \) is finite and the second stage decision variable \( y \) is unconstrained, we can assume without loss of generality that \( B > 0 \) and \( d \geq 0 \). In which case,

\[
Q(x, z) = d \max_{i \in [M]} \left\{ \frac{|b(z) - A(z)x|}{|B|} \right\}.
\]

Hence, applying Theorem 1, we have

\[
\beta(x) = \min d(r + s' + t' \sigma) \\
\text{s.t.} \quad r + s'(Gz) + t'u \geq \frac{|b(z) - A(z)x|}{|B|} \; \forall i \in [M], \forall (z, u, v) \in \hat{W} \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2}.
\]

The solution derived under generalized linear decision rule is

\[
\tilde{\beta}_{(I_1, [I_2], \emptyset)}(x) = \min r + s' + t' \sigma \\
\text{s.t.} \quad r + s'(Gz) + t'u \geq d(y^0 + y^1'z + y^2'u) \; \forall (z, u, v) \in \hat{W} \\
A(z)x + By(z, u) \geq b(z) \; \forall (z, u, v) \in \hat{W} \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2} \\
y \in \mathcal{L}([I_1], [I_2], \emptyset),
\]

or equivalently

\[
\tilde{\beta}_{(I_1, [I_2], \emptyset)}(x) = \min r + s' + t' \sigma \\
\text{s.t.} \quad r + s'(Gz) + t'u \geq d(y^0 + y^1'z + y^2'u) \; \forall (z, u, v) \in \hat{W} \\
(y^0 + y^1'z + y^2'u) \geq \frac{|b(z) - A(z)x|}{|B|} \; \forall i \in [M], \forall (z, u, v) \in \hat{W} \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2} \\
y^0 \in \mathbb{R}, y^1, y^2 \in \mathbb{R}^{L_1}, y^3 \in \mathbb{R}^{L_2}.
\]
Let \((r^\dagger, s^\dagger, t^\dagger)\) be a feasible solution of Problem (23). We can construct a feasible solution \((r, s, t, y_0^1, y_1^1, y_2^1)\) to Problem (24) by letting
\[
y_0^1 = r, \quad y_1^1 = G' s, \quad y_2^1 = t, \quad r = dr^\dagger, \quad s = ds^\dagger, \quad t = dt^\dagger,
\]
which yields the same objective as Problem (23). Hence, \(\bar{\beta}(\mathbf{x}) \leq \beta(\mathbf{x})\) and equality is achieved from Proposition 2. \(\square\)

**Improvement over deflected linear decision rules**

Chen et al. (2008), Goh and Sim (2010) propose a class of piecewise linear decision rules known as deflected linear decision rules which can also circumvent the issues of infeasibility in complete recourse problems. The approach requires to solve a set of subproblems given by
\[
f_i^* = \min_{d' y} \quad \text{s.t. } B y = q \\
q \geq e_i \quad y \in \mathbb{R}^{N_2}, q \in \mathbb{R}^M,
\]
for all \(i \in [M]\), which are not necessarily feasible optimization problems. Let \(\mathcal{M} \subseteq [M]\) denote the subset of indices in which their corresponding subproblems are feasible, i.e., \(\mathcal{M} = \{i \in [M] : f_i^* < \infty\}\), and \(\bar{\mathcal{M}} = [M] \setminus \mathcal{M}\). Correspondingly, let \((\bar{y}_i, \bar{q}_i)\) be the optimal solution of Problem (25) for all \(i \in \mathcal{M}\). Here, \(f_i^* = d' \bar{y}_i \geq 0, \ i \in \mathcal{M}\) is assumed or otherwise, \(Q(x, z)\) would be unbounded from below. The solution to deflected linear decision is obtained by solving the following optimization problem,
\[
\bar{\beta}_{DLDR}(\mathbf{x}) = \min \sup_{P \in \mathcal{F}} \mathbb{E}_P(d' y_{DLDR}(\tilde{z}))) + \sum_{i \in \mathcal{M}} \bar{f}_i^* \sup_{P \in \mathcal{F}} \mathbb{E}_{P}((-q_i(\tilde{z}))^+ ) \\
\text{s.t. } A(z)x + B y(z) = b(z) + q(z) \quad \forall z \in \mathcal{W} \\
q_i(z) \geq 0 \quad q_i \in \mathcal{L}^M([I_1], \emptyset, \emptyset) \quad \forall i \in \bar{\mathcal{M}}, \forall z \in \mathcal{W} \quad (26)
\]
Suppose \((y^*, q^*)\) is the optimal solution of Problem (26), the corresponding deflected linear decision rule is given by
\[
y_{DLDR}(z) = y^*(z) + \sum_{i \in \mathcal{M}} \bar{y}_i((-q_i^*(z))^+).
\]
Chen et al. (2008), Goh and Sim (2010) show that \(y_{DLDR}(\tilde{z})\) is a feasible solution of Problem (16). Moreover,
\[
\sup_{P \in \mathcal{F}} \mathbb{E}_P(d' y_{DLDR}(\tilde{z}))) \leq \bar{\beta}_{DLDR}(\mathbf{x}) \leq \bar{\beta}(\mathbf{x}) \leq \bar{\beta}_{(I_1, [0, \emptyset])}(\mathbf{x}).
\]
Our next result shows that the generalized linear decision rule can potentially improve the bound provided by deflected linear decision rule.
\begin{proposition}
\[ \tilde{\beta}_{(I_1, I_2, \emptyset)}(x) \leq \tilde{\beta}_{DLDR}(x). \]
\end{proposition}

\textbf{Proof.} From Theorem 1, we have the equivalent form of \( \tilde{\beta}_{DLDR}(x) \) as follows:

\[
\tilde{\beta}_{DLDR}(x) = \min r_0 + s_0' \mu + t_0' \sigma + \sum_{i \in \mathcal{M}} f_i^*(r_i + s_i' \mu + t_i' \sigma) \\
\text{s.t. } r_0 + s_0' (Gz) + t_0' u \geq d' y(z) \quad \forall (z, u, v) \in \hat{W} \\
q_i(z), \quad \forall (z, u, v) \in \hat{W} \\
t_i \geq 0 \quad \forall (z, u, v) \in \hat{W} \\
A(z)x + By(z) = b(z) + q(z) \quad \forall (z, u, v) \in \hat{W} \\
0 \quad \forall (z, u, v) \in \hat{W} \\
r_i \in \mathbb{R}, s_i \in \mathbb{R}^{L_1}, t_i \in \mathbb{R}^{L_2} \\
y \in \mathcal{L}^{N_2}([I_1], [0], \emptyset) \\
q \in \mathcal{L}^{M}([I_1], [0], \emptyset). 
\tag{27}
\]

Similarly, we have the equivalent form of \( \tilde{\beta}_{(I_1, I_2, \emptyset)}(x) \) as follows:

\[
\tilde{\beta}_{(I_1, I_2, \emptyset)}(x) = \min r + s' \mu + t' \sigma \\
\text{s.t. } r + s' (Gz) + t' u \geq d' y(z, u) \quad \forall (z, u, v) \in \hat{W} \\
A(z)x + By(z, u) = b(z) \quad \forall (z, u, v) \in \hat{W} \\
t \geq 0 \\
r \in \mathbb{R}, s \in \mathbb{R}^{L_1}, t \in \mathbb{R}^{L_2} \\
y \in \mathcal{L}^{N_2}([I_1], [I_2], \emptyset). 
\tag{28}
\]

Let \( y^i, q^i, r^i, s^i, t^i, i \in \{0\} \cup \mathcal{M} \) be a feasible solution of Problem (27). We will show that there exists a corresponding feasible solution for Problem (28) with the same objective value. Let

\[
\begin{align*}
    r &= r_0^i + \sum_{i \in \mathcal{M}} d_y r_i^i \\
    s &= s_0^i + \sum_{i \in \mathcal{M}} d_y s_i^i \\
    t &= t_0^i + \sum_{i \in \mathcal{M}} d_y t_i^i, \\
    y(z, u) &= y^i(z) + \sum_{i \in \mathcal{M}} (r_i^i + s_i^i (Gz) + t_i^i u) y_i.
\end{align*}
\]

Observe that the objective value of Problem (28) becomes

\[
\begin{align*}
    r + s' \mu + t' \sigma &= r_0^i + s_0^i (Gz) + t_0^i (Gz) + \sum_{i \in \mathcal{M}} (r_i^i + s_i^i (Gz) + t_i^i u) y_i \\
    &= r_0^i + s_0^i (Gz) + t_0^i (Gz) + \sum_{i \in \mathcal{M}} f_i^* (r_i^i + s_i^i (Gz) + t_i^i u).
\end{align*}
\]
We next check the feasibility of the solution in Problem (28). Note that \( t \geq 0 \) and for all \((z, u, v) \in \mathcal{W}\),

\[
r + s'(Gz) + t'u = r_0^i + \sum_{i \in M} d'y_i r_i^i + \left( s_0^i + \sum_{i \in M} d'y_i s_i^i \right)'(Gz) + \left( t_0^i + \sum_{i \in M} d'y_i t_i^i \right)'u
\]

\[
= r_0^i + s_0^i(Gz) + t_0^i u + \sum_{i \in M} \left( r_i^i + s_i^i(Gz) + t_i^i u \right) d'y_i
\]

\[
\geq d'y^i(z) + \sum_{i \in M} \left( r_i^i + s_i^i(Gz) + t_i^i u \right) d'y_i
\]

\[
= d'y(z, u),
\]

where the inequality follows from the first robust counterpart constraint in Problem (27). We now show the feasibility of second robust robust counterpart constraint in Problem (28). Observe that for all \((z, u, v) \in \mathcal{W}\),

\[
A(z)x + B y(z, u) = A(z)x + B y^i(z) + \sum_{i \in M} \left( r_i^i + s_i^i(Gz) + t_i^i u \right) B y_i
\]

\[
= b(z) + q^i(z) + \sum_{i \in M} \left( r_i^i + s_i^i(Gz) + t_i^i u \right) q_i
\]

\[
= b(z) + \sum_{i \in M} q_i^i(z) e_i + \sum_{i \in M} q_i^j(z) e_j + \sum_{i \in M} \left( r_i^i + s_i^i(Gz) + t_i^i u \right) q_i
\]

\[
geq b(z) + \sum_{i \in M} q_i^i(z) e_i + \sum_{i \in M} q_i^j(z) e_j + \sum_{i \in M} \left( r_i^i + s_i^i(Gz) + t_i^i u \right) e_i
\]

\[
= b(z) + \sum_{j \in M} q_j^i(z) e_j + \sum_{i \in M} \left( q_i^i(z) + r_i^i + s_i^i(Gz) + t_i^i u \right) e_i
\]

\[
\geq b(z).
\]

The first inequality holds because \( q_i \geq e_i \) and \( r_i^i + s_i^i(Gz) + t_i^i u \geq 0 \) for all \( i \in M \), \((z, u, v) \in \mathcal{W}\). The second inequality is due to \( r_i^i + s_i^i(Gz) + t_i^i u \geq -q_i^i(z) \) for all \( i \in M \), \((z, u, v) \in \mathcal{W}\) and \( q_i^i(z) \geq 0 \) for all \( i \in M \), \((z, u, v) \in \mathcal{W}\). This concludes our proof. □

**On the use and misuse of linear decision rules**

We introduce linear decision rules with the goal to obtain tractable formulations of so that the optimal here-and-now decision \( x \in X_1 \) can be determined and implemented. For a given \( x \in X_1 \), let \( y^* \) be the optimal function of Problem (16), and \( y_{GLDR}^* \) be the optimal generalized linear decision rule of Problem (19). For given a function, \( \nu \in \mathcal{R}^{J^1,J^2} \) satisfying \((z, g(z), \nu(z)) \in \mathcal{W}\) for all \( z \in \mathcal{W}\), the function \( y_{GLDR}^* \in \mathcal{R}^{J_1,J_2} \),

\[
y_{GLDR}^*(z) = y_{GLDR}^*(z, g(z), \nu(z))
\]

is a feasible solution to Problem (16). Moreover, the objective satisfies

\[
\sup_{p \in P} E_p \left( d'y_{GLDR}^*(\tilde{z}) \right) = \sup_{p \in P} E_p \left( d'y_{GLDR}^*(\tilde{z}, g(\tilde{z}), \nu(\tilde{z})) \right) \leq \sup_{q \in Q} E_q \left( d'y_{GLDR}^*(\tilde{z}, \tilde{u}, \tilde{v}) \right) = \tilde{\beta}(s_1, s_2, s_3)(x),
\]
where the inequality is due to Proposition (1). Suppose
\[ \beta(x) = \sup_{P \in \mathcal{F}} \mathbb{E}_P (d'y^*(\tilde{z})) = \sup_{P \in \mathcal{F}} \mathbb{E}_P (d'\hat{y}_{GLDR}(\tilde{z})) = \bar{\beta}(S_1, S_2, S_3)(x), \]
which is the case for complete recourse problems and \( N_2 = 1 \), there is a tendency to infer the optimality of \( \hat{y}_{GLDR}(z) \), such that
\[ d'\hat{y}_{GLDR}(z) = d'y^*(z) \quad \forall z \in \mathcal{W}. \]
However, this is not the case and we will demonstrate this fallacy in the following simple example.

**Example 3.** Consider the following complete recourse problem,
\[
\begin{align*}
\beta = \min & \sup_{P \in \mathcal{F}} \mathbb{E}_P (y(\tilde{z})) \\
\text{s.t.} & \quad y(z) \geq z \quad \forall z \in \mathbb{R} \\
& \quad y(z) \geq -z \quad \forall z \in \mathbb{R} \\
& \quad y \in \mathcal{R}^{1,1},
\end{align*}
\]
where
\[ \mathcal{F} = \{ P \in \mathcal{P}_0(\mathbb{R}) : \mathbb{E}_P (\tilde{z}) = 0, \mathbb{E}_P (\tilde{z}^2) \leq 1 \}. \]
Clearly, \( y^*(z) = |z| \) is the optimal solution and it is also the optimal objective value for all \( z \in \mathbb{R} \). However, under the generalized linear decision rule, we obtain \( \hat{y}_{GLDR}(z) = \frac{1+z^2}{2} \), which is almost always greater than \( y^*(z) \) except at \( z = 1 \) and \( z = -1 \). Incidentally, the worst case distribution \( P \in \mathcal{F} \) corresponds to the two point distributions with \( P(\tilde{z} = 1) = P(\tilde{z} = 1) = 1/2 \). Hence, this explains why the worse case expectations are the same.

Hence, from the above example, even if a generalized linear decision rule were to provide a close approximation to \( \beta(x) \), \( x \in X_1 \), the solution generated by the decision rule could be a far cry from the optimal function, \( y^* \). Therefore, we advise against using the generalized decision rule as a policy guide for future actions when uncertainty is realized. Instead, the second stage decision should be determined by solving a linear optimization problem after the uncertainty is resolved.

Another important feature of linear decisions rule is the ability to easily enforce non-anticipative conditions, which are necessary to capture the nature of multistage decisions where information is revealed in stages. For given subsets \( S_i \subseteq [I_1] \), that reflects information dependency of recourse decisions, \( y_i, i \in [N_2] \), we can consider the generalization of Problem (16) as follows:
\[
\gamma^*(x) = \min \sup_{P \in \mathcal{F}} \mathbb{E}_P (d'\hat{y}(\tilde{z}, \tilde{u}, \tilde{v}))
\]
\[
\begin{align*}
\text{s.t.} & \quad A(z)x + By(z, u, v) \geq b(z) \quad \forall (z, u, v) \in \mathcal{W} \\
& \quad y_i \in \mathcal{R}^{I_1,1}(S_i) \quad \forall i \in [N_2],
\end{align*}
\]
where we define the space of restricted measurable functions as
\[ \mathcal{R}^{I,N}(\mathcal{S}) = \{ y \in \mathcal{R}^{I,N} : y(v) = y(w) \quad \forall v, w \in \mathbb{R}^I : v_j = w_j, j \in \mathcal{S} \}. \]
Problem (30) solves for the optimum measurable function \( y \in \mathcal{R}^{I_1 \times N_2} \) that minimizes the worst case expected objective taking into account of the information dependency requirement. Clearly, this problem would be much harder to solve and we are not aware of a viable approach to compute the exact solution. Yet, despite the difficulty, it is relatively simple to use generalized linear decision rules to obtain an upper bound as follows:

\[
\tilde{\gamma}(x) = \min_{Q \in G} \sup_{d \in \mathcal{D}} \mathbb{E}_Q(d', y(z, u, v)) \\
\text{s.t. } A(z)x + By(z, u, v) \geq b(z) \quad \forall (z, u, v) \in \hat{W} \\
y_i \in \mathcal{L}^i(S_i^1, S_i^2, S_i^3) \quad \forall i \in [N_2],
\]

where the subsets \( S_i^2 \subseteq [I_2], S_i^3 \subseteq [I_3], \) are appropriately selected to abide by the information restriction imposed by \( S_i^1 \subseteq [I_1], i \in [N_2]. \) Again, we use the generalized linear decision rules to enable us to obtain a reasonably good here-and-now decision, \( x \in X_1 \) that accounts for how decisions might be adjusted as uncertainty unfolds over the stages. Similar to the standard adjustable robust optimization technique, we propose the rolling or folding horizon implementation where we solve for the new here-and-now decision using the latest available information as we proceed to the next stage.

In the next section, we will briefly describe a new algebraic modeling package named ROC and show how it could be used to facilitate modeling of distributionally robust linear optimization problems.

4. ROC: Robust Optimization C++ package

We developed ROC as a proof of concept to provide an intuitive environment for modeling and solving distributionally robust linear optimization problems that will free the user from dealing directly with the laborious and error-prone reformulations. ROC is developed in the C++ programming language, which is fast, highly portable and well suited for deployment of robust optimization technologies in decision support system. We will briefly discuss the key aspects of ROC and provide simple examples to illustrate the algebraic modeling package. Most algebraic modeling packages for optimization are geared towards modeling deterministic optimization problems. While a robust optimization problem may be formulated as a deterministic optimization problem, it is would be rather difficult for the modeler to explicitly code, say Problem (31) using these algebraic modeling packages.

A typical algebraic modeling package provides the standardized format for declaration of decision variables, transcription of constraints and objective functions, and interface with external solvers. ROC has additional features including declaration of uncertain parameters and linear decision rules, transcriptions of ambiguity sets and automatic reformulation of standard and distributionally robust counterparts using the techniques described in this paper. The current version of ROC solver is integrated with CPLEX and will be expanded to include other solvers. We refer readers to http://www.meilinzhang.com/software for more information on ROC.
Declaration of decisions, uncertain parameters and expressions

Code Segment 1 provides an example on how we define decision variables, uncertain parameters and linear decision rules in ROC. The code illustrates how the following deterministic decision variables are declared:

\[
\begin{aligned}
  x_1 &\in \mathbb{R}, x_2 \in [5, \infty), x_3 \in \{0, \ldots, 100\}, x_4 \in \{0, 1\}, s \in \mathbb{R}^6, t \in \mathbb{R}^{5 \times 8}.
\end{aligned}
\]

By C++ convention, an array of sized \( N \) is defined on indices 0, \ldots, \( N-1 \). The variable \( x_2 \) is also associated with the name “X2”, which would be useful in output display of the model. Note that \( \tilde{z}_1, \tilde{z}_2, \tilde{z}_3 \) are uncertain parameters in \( \mathbb{R} \) and \( \tilde{u} \) is an array of uncertain parameters in \( \mathbb{R}^6 \). The linear decision rules \( y_1, y_2, y_3 \) are declared. The user can selectively include the linear dependency using the addDR function. In this case, \( y_1 \) is affinely dependent on \( \tilde{z}_1 \), \( y_2 \) is affinely dependent on \( \tilde{z}_2 + \tilde{z}_3 \), and \( y_3 \) has the same dependency as \( y_1 \), i.e.,

\[
\begin{aligned}
  y_1(\tilde{z}_1) &= y_1^0 + y_1^1 \tilde{z}_1, \\
  y_2(\tilde{z}_2, \tilde{z}_3) &= y_2^0 + y_2^1 (\tilde{z}_2 + \tilde{z}_3), \\
  y_3(\tilde{z}_3) &= y_3^0 + y_3^1 \tilde{z}_3.
\end{aligned}
\]

where \( y_1^0, y_1^1, y_2^0, y_2^1, y_3^0 \) and \( y_3^1 \) are embedded decision variables that are declared in association with the linear decision rules.

```c++
// Declaration of decisions and uncertain parameters
ROVar x1, x2(5, ROInfinity, "X2"); // x1, x2 continuous decision variables
ROIntVar x3(0, 10); // x3 Integer variable
ROBinVar x4; // x4 binary variables
ROVarArray s(6); // an array of 6 decision variables
ROUn z1, z2, z3; // three uncertain parameters
ROUnArray u(6); // an array of 6 uncertain parameters

// Define a 2D array of 5 by 8 decision variables
ROVar2DArray t(5);
for (int i; i < 5; i++)
  t[i] = ROVarArray(8);

ROVar y1, y2, y3; // two linear decision rules
y1.addDR(z1); // add dependency on uncertain parameters to linear decision rule
y2.addDR(z2 + z3); // add dependency on uncertain parameters to linear decision rule
y3.clone(y1);
```

Code Segment 1: Declaration of decisions and uncertain parameters ROC.

We can also declare an expression, which is an object to contain either a quadratic function of decision variable or a biaffine function of the decision variables and uncertain parameters. An expression permits linear operations on constants, decision variables, uncertain parameters, linear
decision rules and other expressions and it is useful as temporary storage. Code Segment 2 shows some examples of expressions. Here, we have

\[
\begin{align*}
\text{expr1:} & \quad x_1^2 + x_2 \\
\text{expr2:} & \quad s_1 + 2s_4 - x_1\tilde{z}_1 \\
\text{expr3:} & \quad s_1 + 2s_4 - x_1\tilde{z}_1 + y_0^2 + y_2^2(\tilde{z}_2 + \tilde{z}_3) \\
\text{expr4:} & \quad s_1\tilde{u}_1 + 2s_2\tilde{u}_2 + 3s_3\tilde{u}_3 + 4s_4\tilde{u}_4 + 5s_5\tilde{u}_5 + 6s_6\tilde{u}_6.
\end{align*}
\]

Code Segment 2: The use of expressions in ROC.

Modeling Ambiguity sets

The ability to comprehensively model distributionally ambiguity sets ROC apart from other algebraic modeling packages. The \texttt{ambiguitySet} defined in Code Segment 3 describes the following ambiguity set

\[
G = \left\{ \mathcal{Q} \in \mathcal{P}_0\left(\mathbb{R}^3 \times \mathbb{R}^6\right) : E_\mathcal{Q}(\tilde{\mathbf{u}}_1) = 1 \right\},
\]

where

\[
\hat{\mathcal{W}} = \left\{ (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^3 \times \mathbb{R}^6 : \begin{array}{l}
|\mathbf{u}|_\infty \leq 5 \\
|\mathbf{u}|_1 \leq 4 \\
|\mathbf{u}|_2 \leq 3
\end{array} \right\}.
\]

Code Segment 3: Definition of an ambiguity set in ROC.

Note that the statement \texttt{ROSq(z1) <= z3} calls upon the function \texttt{ROSq}, which returns a newly declared uncertain parameter, say \( \tilde{v} \) so that \( \tilde{v} \leq \tilde{z}_3 \). Internally within the function, the epigraph of \( \tilde{z}_1^2 \leq \tilde{v} \) is automatically converted to a second order cone constraint, \( \sqrt{(\tilde{z}_1^2 - \frac{1}{2})^2 + \tilde{z}_1^2} \leq \frac{\tilde{v} + 1}{2} \). Hence, the user should be disciplined in convex representation of constraints and avoid statements such
as $\text{ROSq}(z_1)\geq 7$. Likewise, the functions $\text{RONorm1}$, $\text{RONorm2}$ and $\text{RONormInf}$ are provided within ROC for modeling convenience. These functions return newly declared uncertain parameters and internally represent the epigraphs of these functions using linear and second order conic constraints. The functions (such as $\text{RONorm1}$) may declare other uncertain parameters that are hidden from the user. Using this approach, we can also declare other common conic quadratic representable functions within ROC including higher powers and approximations of exponential functions, among others. We have also provided functions that linearly approximates second order cones as propose in Ben-Tal and Nemirovski (2001b), which may be useful if linearity of the model is desired. Note that decision variables are not permitted in the description of ambiguity set and that the user has the freedom to define multiple ambiguity sets.

**Declaration of a model, adding constraints and the objective function**

A ROC model consists of objects that represents a problem including constraints and the objective function. Deterministic constraints can be added in the model as shown in Code Segment 4, which models the following set of constraints

\[
\begin{align*}
    x_1^2 + x_2 & \leq 7, \\
    x_1^2 - 2x_1x_2 + x_2^2 & \leq 7, \\
    |x_1 - x_3| & \leq 7, \\
    (x_2 - x_3)^+ & \leq x_1, \\
    |s|_1 & \leq 4t^2, \\
    |s|_2 & \leq 6(x_1 + 2x_2), \\
    |s|_\infty & \leq -x_2^2, \\
    1's & \leq 10.
\end{align*}
\]

Similar to the descriptions of ambiguity sets, the functions return newly declared decision variables and internally represent the epigraphs of these functions using linear and second order conic constraints.

```plaintext
ROModel model;  // define a robust optimization model
model.add(expr1 <= t[1][4]);
model.add(x1*x1-2*x1*x2+ x2*x2 <= 7);
model.add(ROAbs(x1-x3) <= 7);
model.add(ROPos(x2-x3) <= x1);
model.add(RONorm1(s) <= 4*t^2);
model.add(RONorm2(s) <= 6*(x1 +2*x2));
model.add(RONormInf(s) <= -x2^2);
model.add(ROSum(s) <= 10);
```

Code Segment 4: Model declaration with deterministic constraints in ROC.

More interestingly, ROC is able to model robust counterpart constraint such as,

```plaintext
model.add(ROCConstraint(expr4 <= x1, ambiguitySet));
```
which automatically reformulates the following robust counterpart,

\[ s_1 u_1 + 2s_2 u_2 + 3s_3 u_3 + 4s_4 u_4 + 5s_5 u_5 + 6s_6 u_6 \leq x_1 \quad \forall (z, u) \in \hat{W}, \]

into a set of deterministic constraints. In the process, new decision variables may be declared that are hidden away from the user. Note the ambiguity set must be specified in the robust counterpart constraint, so that ROC an extract the underlying uncertainty set \( \hat{W} \). Hence, different ambiguity sets can be defined for use in different robust counterpart constraints. More interestingly, a distributionally robust counterpart over the worst case expectation such as,

\[ 1 \text{model.add(ROConstraint(ROExpect(expr2) >= x3 + x1), ambiguitySet));} \]

which corresponds to

\[ E_Q(s_1 + 2s_4 - x_1 \tilde{z}_1 + y_2^0 + y_2^1(\tilde{z}_2 + \tilde{z}_3)) \geq x_1 + x_2 \quad \forall Q \in G, \]

or equivalently as

\[ \sup_{Q \in G} E_Q(-(s_1 + 2s_4 - x_1 \tilde{z}_1 + y_2^0 + y_2^1(\tilde{z}_2 + \tilde{z}_3))) \leq -x_3 - x_2, \]

will be transformed to a set of deterministic constraints using Theorem 1.

The model should finally include an objective, which reflects either a minimization or maximization problem. If the objective expression contains uncertain parameters, then it must also incorporate the corresponding ambiguity set so the worst case objective can be evaluated. The following code segment illustrates an objective function that minimizes the worst case expectation of expr2 over the ambiguity set, \( G \).

\[ 1 \text{model.add(ROMinimize(ROExpect(expr2), ambiguitySet));} \]

5. Computation Experiment

In our experiment, we consider a multiproduct newsvendor problem with \( N \) different types of products, indexed by \( i \). For product \( i, i \in [N] \), its selling price and order cost are denoted by \( p_i \) and \( c_i \) respectively. Manager needs to decide each product’s order quantity \( x_i, i \in [N] \) before the demand \( \tilde{z} = (\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_N) \) is observed. Meanwhile, the total budget for purchasing all products is \( \Gamma \). After the demand becomes known, the selling quantity is decided as \( \min\{x_i, z_i\}, i \in [N] \). In order to maximize the expected operating revenue, the problem could be formulated as

\[
\Pi^* = \max \inf_{P \in F} E_P \left( \sum_{i \in [N]} p_i \min\{x_i, \hat{z}_i\} \right) \\
\text{s.t.} \quad c'x \leq \Gamma \\
x \geq 0 \\
x \in \mathbb{R}^N.
\]
To be consistent with the earlier framework, we formulate this as the following minimization problem

\[ Z^* = -\Pi^* = \min \quad -p'x + \sup_{P \in \mathcal{P}} \mathbb{E}_p \left( \sum_{i \in [n]} p_i ((x_i - \tilde{z}_i)^+) \right) \]

\[ \text{s.t. } c'x \leq \Gamma \]
\[ x \geq 0 \]
\[ x \in \mathbb{R}^N. \]

\[ (32) \]

To demonstrate the modeling power of the standardized framework for characterizing distributional ambiguity, we present the following unusual but interesting ambiguity set that is inspired by the structure of the optimization problem.

\[ F = \left\{ P \in \mathcal{P}_0(\mathbb{R}^N) : \begin{align*}
\mathbb{E}_P(\tilde{z}) &= \mu \\
\mathbb{E}_P(\tilde{z}_i^2) &\leq \mu_i^2 + \sigma_i^2 \\
\mathbb{E}_P \left( \sum_{i \in [N]} p_i (\mu_i - \tilde{z}_i)^+ \right) &\leq \psi \\
P(\tilde{z} \in \mathcal{W}) &= 1
\end{align*} \quad \forall i \in [N] \right\}, \]

\[ (33) \]

where \( \mathcal{W} = \{ z \in \mathbb{R}^n : 0 \leq \tilde{z} \leq \bar{z} \} \).

Correspondingly, the extended ambiguity set of \( F \) is given by

\[ G = \left\{ Q \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^{N+1} \times \mathbb{R}^N) : \begin{align*}
\mathbb{E}_Q(\tilde{z}) &= \mu \\
\mathbb{E}_Q(\tilde{u}_i) &\leq \mu_i^2 + \sigma_i^2 \\
\mathbb{E}_Q(\tilde{u}_{N+1}) &\leq \psi \\
Q(\tilde{z}, \tilde{u}, \bar{v}) &\in \mathcal{W} \quad (34) \end{align*} \right\}, \]

where \( \mathcal{W} = \{ z \in \mathbb{R}^n \times \mathbb{R}^N \times \mathbb{R}^{N+1} \times \mathbb{R}^N : \begin{align*}
0 &\leq \tilde{z} \leq \bar{z} \\
z_i^2 &\leq u_i \\
u_{N+1} &\geq p'v \\
v &\geq \mu - z \\
v &\geq 0 \quad \forall i \in [N] \right\} \),

Using the generalized linear decision rule, we solve the following two-stage distributionally robust optimization problem,

\[ \tilde{Z}^*(S_1, S_2, S_3) = \min \quad -p'x + \sup_{Q \in G} \mathbb{E}_Q(p'y(\tilde{z}, \tilde{u}, \bar{v})) \]

\[ \text{s.t. } c'x \leq \Gamma \]
\[ y(z, u, v) \geq 0 \quad \forall (z, u, v) \in \mathcal{W} \]
\[ y(z, u, v) \geq x - z \quad \forall (z, u, v) \in \mathcal{W} \]
\[ x \in \mathbb{R}^N \]
\[ y \in L^N(S_1, S_2, S_3). \]

\[ (35) \]
Formulating in ROC

Instead of deriving the explicit mathematical model of Problem (35), we present the formulation in ROC, which will automatically transform the problem and call upon a standard solver package such as CPLEX to obtain the solution. We first define the decision variables, \( x \in \mathbb{R}^N \), uncertain parameters, \( \tilde{z} \in \mathbb{R}^N \), \( \tilde{u} \in \mathbb{R}^{N+1} \), \( \tilde{v} \in \mathbb{R}^N \) and the linear decision rule \( y \in \mathcal{R}^N \) as shown in Code Segment 5.

```cpp
// Define Decision variables, decision rules and uncertain parameters
ROVarArray x(N, 0, ROInfinity, "X");
ROVarDRArray y(N);
ROUnArray z(N), u(N+1), v(N);
```

Code Segment 5: Defining decision variables, uncertain parameters and linear decision rule.

We next show how to characterize the dependency of the decision rule \( y \). Code Segment 6 presents an example where the decision rule \( y \) is defined in \( \mathcal{L}^N([N],[N+1],[N]) \), and hence it is fully dependent on all the uncertainty parameters including the auxiliary ones.

```cpp
// Adding dependency to linear decision rules
for(int i = 0; i < N; i++)
{
    for(int j = 0; j < N; j++)
    {
        y[i].addDR(z[j]);
        y[i].addDR(u[j]);
        y[i].addDR(v[j]);
    }
    y[i].addDR(u[N]);
}
```

Code Segment 6: Defining generalized linear decision rule in \( \mathcal{L}^N([N],[N+1],[N]) \)

Next, we specify the ambiguity set \( \mathcal{G} \) as shown in Code Segment 7.

```cpp
// Construct the Ambiguity Set
ROConstraintSet ambiguitySet;
ROExpr unExpr;
for(int i = 0; i < N; i++)
{
    ambiguitySet.add(ROExpect(z[i]) == mu[i]);
    ambiguitySet.add(ROExpect(u[i]) <= mu[i]*mu[i] + sigma[i]*sigma[i]);
    ambiguitySet.add(z[i] >= 0);
    ambiguitySet.add(z[i] <= barZ[i]);
    ambiguitySet.add(ROSq(z[i]) <= u[i]);
    ambiguitySet.add(v[i] >= 0);
    ambiguitySet.add(v[i] >= mu[i] - z[i]);
    unExpr += price[i] * v[i];
}
ambiguitySet.add(ROExpect(u[N]) <= psi);
ambiguitySet.add(u[N] >= unExpr);
```
Code Segment 7: Constructing the ambiguity set $G$. 

Finally, Code Segment 8 show how the we model Problem (35) in ROC.

```
ROModel model; // define a robust optimization model engine

// Adding constraints to Model
ROExpr expr1;
for(int i = 0; i < N; i++)
{
    expr1 += cost[i] * x[i];
    model.add(ROConstraint(y[i] >= 0, ambiguitySet));
    model.add(ROConstraint(y[i] >= x[i] - z[i], ambiguitySet));
}
model.add(expr1 <= budget);

// Adding objective expression
ROExpr objExpr1, objExpr2;
for(int i = 0; i < N; i++)
{
    objExpr1 -= price[i] * x[i];
    objExpr2 += price[i] * y[i];
}
model.add(ROMinimize(objExpr1 + ROExpect(objExpr2, ambiguitySet)));
model.solve();
```

Code Segment 8: Create the robust optimization model.

**Performance of the decision rules**

For the purpose of comparison, we next formulate the model to evaluate Problem (32) exactly. By observing that $\sum_{i \in [N]} (a_i)^+ = \max_{S: S \subseteq [N]} \left( \sum_{i \in S} a_i \right)$, we can transform Problem (32) to the following problem

$$Z^* = \min \ -p'x + \sup_{\psi \in \mathcal{F}} \mathbb{E}_{\psi} \left( \max_{S: S \subseteq [N]} \left( \sum_{i \in S} p_i(x_i - \tilde{z}_i) \right) \right)$$

\[ s.t. \ c'x \leq \Gamma \]
\[ x \geq 0 \]
\[ x \in \mathbb{R}^N \]

Noting that the number of subsets of $[N]$ equals to $2^N$, we will study a small problem so that it would be computationally variable to compare the quality of solutions obtained by linear decision rules. Hence, we restrict to $N = 10$. We solve for a particular instance with $\psi = 100$, $\Gamma = 500$ and the parameters associated with the products are shown in Table 1.
Table 1 Input parameters of multiproduct newsvendor problem

<table>
<thead>
<tr>
<th>Product ID</th>
<th>price[i]</th>
<th>cost[i]</th>
<th>mu[i]</th>
<th>sigma[i]</th>
<th>z_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.00</td>
<td>2.00</td>
<td>30.00</td>
<td>30.00</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>11.00</td>
<td>2.71</td>
<td>35.00</td>
<td>28.50</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>11.41</td>
<td>3.00</td>
<td>40.00</td>
<td>27.00</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>11.73</td>
<td>3.23</td>
<td>45.00</td>
<td>25.50</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>12.00</td>
<td>3.41</td>
<td>50.00</td>
<td>24.00</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>12.24</td>
<td>3.58</td>
<td>55.00</td>
<td>22.50</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>12.45</td>
<td>3.73</td>
<td>60.00</td>
<td>21.00</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>12.65</td>
<td>3.87</td>
<td>65.00</td>
<td>19.50</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>12.83</td>
<td>4.00</td>
<td>70.00</td>
<td>18.00</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>13.00</td>
<td>4.12</td>
<td>75.00</td>
<td>16.50</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2 shows the objective values of

\[ \Pi_1^* = -\bar{Z}^*(\emptyset, \emptyset, \emptyset) \]
\[ \Pi_2^* = -\bar{Z}^*([N], \emptyset, \emptyset) \]
\[ \Pi_3^* = -\bar{Z}^*([N],[N+1],\emptyset) \]
\[ \Pi_4^* = -\bar{Z}^*([N],[N+1],[N]) \]
\[ \Pi^* = -\bar{Z}^* \]

and also presents the corresponding optimal solutions. We observe that the improvement in objective values as the decision rule has dependent on greater subsets of uncertain parameters. In particular, for the case of full dependency, we have \( \Pi_4^* \) achieving the optimal objective value \( \Pi^* \), underscoring the potential and benefits of the generalized linear decision rule in addressing distributionally robust linear optimization problems.

**Endnotes**


**References**


Knight, F. H. (1921) Risk, uncertainty and profit. *Hart, Schaffner and Marx*.


