Supermodularity in Two-Stage Distributionally Robust Optimization

Daniel Zhuoyu Long  
Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Shatin, Hong Kong  
zylong@se.cuhk.edu.hk,

Jin Qi  
Department of Industrial Engineering and Decision Analytics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong jinqi@ust.hk,

Aiqi Zhang  
Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Shatin, Hong Kong  
aqzhang@se.cuhk.edu.hk,

In this paper, we solve a class of two-stage distributionally robust optimization problems which have the property of supermodularity. We exploit the explicit upper bounds on the expectation of supermodular functions and derive the worst-case distribution for the robust counterpart. This enables us to develop an efficient method to derive an exact optimal solution of these two-stage problems. Further, we provide a necessary and sufficient condition to check whether any given two-stage optimization problem has supermodularity. We apply this framework to classic problems, including the multi-item newsvendor problem, the appointment scheduling problem and general assemble-to-order (ATO) systems. While these problems are typically computationally challenging, they can be solved efficiently using our approach.

Key words: distributionally robust optimization; two-stage problems; supermodularity

1. Introduction

Many real-world optimization problems with uncertainties can be formulated as two-stage optimization models. In such problems, we first make a “here-and-now” decision. In the second stage, after the uncertainties are realized, we choose the optimal action, which we call the “wait-and-see” decision.

This two-stage optimization formulation has drawn extensive attention from both the operations management and optimization communities as it can model a wide range of operational problems. For instance, in an assemble-to-order (ATO) system, the here-and-now decision is the ordering quantities of the components while the wait-and-see decision is the assembly plan which represents the amount of each type of component to be used in assembling each type of product when demand occurs. In appointment scheduling problems, the here-and-now decision is the scheduled appoint-
ment time while there is no wait-and-see decision. Other operational examples include multi-item newsvendor, facility location, unit commitment problems, etc.

A conventional way to deal with two-stage optimization problems is stochastic programming (Shapiro et al. 2009, Birge and Louveaux 2011), in which uncertainties are assumed to follow some given probability distributions. However, since two-stage decision problems typically involve wait-and-see decisions that constitute functional optimization, they tend to be of infinite dimension and in general computationally intractable (Shapiro and Nemirovski 2005). A widely used approximation method is the sampling average approximation (SAA) approach. However, this approach may lead to poor performance if the sample size is small, while incurring a computational burden if the sample size is large (Shapiro et al. 2009).

Alternatively, two-stage optimization problems can be addressed using robust optimization, which was introduced by Soyster (1973) and promoted by Ben-Tal and Nemirovski (1998), El Ghaoui et al. (1998), Bertsimas and Sim (2004). Using robust optimization, instead of optimizing the expectation of objective functions, we seek solutions that are immune to a distribution-free uncertainty set. However, this type of problem is still hard to solve in general because of its two-stage nature. Some approximation methods have been proposed to address the intractable nature of the problem, such as the linear decision rule (Ben-Tal et al. 2004), and more complicated methods including the polynomial (Bertsimas et al. 2011), segregated linear (Chen and Zhang 2009) and piecewise linear (Ben-Tal et al. 2009) decision rules. These approaches restrict solutions to specific functions of the uncertainty realizations (such as affine functions). The functions are parameterized by a finite number of coefficients and lead to computational tractability.

In addition, if the problems have some special structures, the approximated solutions can be proved to be near-optimal or even optimal. Bertsimas and Goyal (2010) show that for a two-stage stochastic problem, the static solutions derived from the corresponding robust version give a 2-approximation to the original stochastic problem if both the uncertainty set and the probability measure are symmetric. For the linear decision rule, Bertsimas et al. (2010b) prove its optimality in multi-period robust optimization problems when the problem is one-dimensional with convex costs. Bertsimas and Goyal (2012) further give the result that linear decision rules can be optimal in a two-stage setting if the uncertainty set is a simplex. Kuhn et al. (2011) apply the linear decision rule approximation to the primal and dual problems separately, in both stochastic programming and robust optimization problems, where the gap between the two approximated values is used to estimate the loss of optimality. The numerical example shows that in the specific setting they adopt, the relative gap between the bounds can be consistently low.

However, since classic robust optimization does not use any distributional information except the support, the solution can be overly conservative and therefore too extreme for practical applications.
To overcome this, by incorporating an ambiguity set $\mathcal{F}$ of probability distributions, distributionally robust optimization (DRO) has been introduced to seek solutions which protect against the worst-case distribution over all admissible ones (Delage and Ye 2010, Goh and Sim 2010, Wiesemann et al. 2014). The distributional ambiguity set containing all possible probability distributions is characterized by certain distributional information. These sets are often based on moment information (Delage and Ye 2010, Zymler et al. 2013a,b, Mehrotra and Zhang 2014) or statistical measures, including the $\phi$-divergence (Ben-Tal et al. 2013) and Wasserstein distance (Gao and Kleywegt 2016, Esfahani and Kuhn 2018, Hanasusanto and Kuhn 2018). Chen et al. (2019) recently propose a scenario-based distributional ambiguity set, which can model a broader class of uncertainty sets, e.g., uncertainty sets with both moment and Wasserstein distance information.

For the two-stage distributionally robust optimization, while solutions can be derived by many parametric decision rules as in robust optimization, the approximation gap is still significant. Further, very few studies have been conducted to examine the general equivalent reformulations and tractability conditions required to solve for exact analytical solutions. Bertsimas et al. (2010a) investigate the cases with ambiguity sets constructed by first and second moments and objective functions being nondecreasing convex piecewise linear disutility functions of the second-stage costs. They show that, if uncertainties only appear in the objective function of the second stage, then the original problems can be equivalently reformulated as semidefinite programs. Bansal et al. (2018) propose decomposition algorithms for two-stage distributionally robust linear problems with discrete distributions, as well as conditions under which the algorithms are finitely convergent. Hanasusanto and Kuhn (2018) show that for problems with complete recourse and the ambiguity sets being 2-Wasserstein balls centered at a discrete distribution, if additionally the uncertainty appears only in constraints of the second-stage problem, then there exists a co-positive cone reformulation.

We extend the previous literature by exploiting the property of supermodularity for a broad class of two-stage distributionally robust optimization problems. Hence, besides the distributionally robust optimization, another stream of studies that are closely related to our work is that on supermodularity. The concept of supermodularity has proved its importance in the area of economics and operations research. In particular, it has economic implications in terms of complementarity between resources. Consequently, scholars are also interested in exploring the supermodularity in their parametric optimization problems in order to derive certain monotonic comparative statistics. However, the results are rather scattered and the proof is usually problem-specific. For the general case, Topkis (1998) first introduces lattice conditions on the feasible set to derive the property of supermodularity. While the lattice condition is quite restrictive, Chen et al. (2013) extend it and study the sufficient condition for a class of two-dimensional parametric optimization problems. A
recent work by Chen et al. (2018b) has provided a systematic study of the conditions both necessary and sufficient to identify the property of supermodularity. Because of the essential implication of complementarity, in a few studies, supermodularity is incorporated within robust optimization to analyze the worst-case performance. Specifically, Agrawal et al. (2012) prove that when the marginal distributions are of two-point distribution and the cost function is convex and supermodular, there exists a polynomial-time algorithm for the optimization problem under uncertainties. In multi-stage robust optimization, Iancu et al. (2013) show that the linear decision rule gives an optimal solution when the objective function is supermodular and the uncertainty set is of a certain special structure.

In this paper, we solve a class of two-stage distributionally robust optimization problems in which the second-stage optimal value is supermodular in the realization of uncertainty. Under the setting of scenario-based ambiguity sets with supports, means and mean absolute deviations (MADs), we exploit the explicit upper bounds on the expectation of supermodular functions and derive the worst-case distribution in the robust counterpart precisely. This can make the two-stage distributionally robust optimization problem tractable. Further, we provide a necessary and sufficient condition to check whether any given two-stage optimization problem has the property of supermodularity. We then identify a class of two-stage optimization problems with supermodularity. These include several classic problems, e.g., multi-item newsvendor problems, general ATO systems and the appointment scheduling problem. While these problems are typically computationally challenging, they can be solved efficiently using our approach.

Our key contributions are summarized as follows.

1. We show that for a specific distributional uncertainty set with moment information, the second-stage problem has an explicit common worst-case distribution whenever it has the property of supermodularity. By inserting this worst-case distribution, the original two-stage problem can be reduced to a deterministic optimization problem of polynomial size.

2. When the second-stage problem is given by a linear programming formulation, we provide a necessary and sufficient condition to check its supermodularity. A simple algorithm is proposed to determine whether the condition is satisfied.

3. We apply our results to several important operational problems, including multi-item newsvendor problems, appointment scheduling problems and ATO systems. For the first two problems, the supermodularity property holds and we can reduce them to tractable formulations. For ATO systems, we provide several special structures in which supermodularity holds.

The rest of this paper is organized as follows. In Section 2, we define the model and illustrate the supermodular requirement for tractability. Then in Section 3, we demonstrate the equivalent conditions for the supermodularity of the objective function in the second stage. Some applications
in practical operations management problems are given in Section 4. We finally conclude the paper in Section 5.

**Notation and convention:** We represent vectors and matrices by lower- and upper-case boldface characters, respectively. We denote by \( x_i \) the \( i \)-th element of the vector \( x \). For any integer \( K \geq 1 \), we define \( [K] = \{1, \ldots, K\} \), which is the set of positive running indices to \( K \). Given any matrix \( A = (a_{ij})_{i \in [m], j \in [n]} \in \mathbb{R}^{m \times n} \), we let \( a_i^T \) and \( A_j \) be its \( i \)-th row vector and \( j \)-th column vector, respectively. Further, we use \( A_I \) to represent its submatrix \( (a_{ij})_{i \in I, j \in [n]} \in \mathbb{R}^{\mid I \mid \times n} \) for any \( I \subseteq [m] \), and we use \( \mid \cdot \mid \) to represent the cardinality of a set. We denote \( \text{span}(A) \) to be the column space of \( A \). We also define two operations join ("\( \vee \)") and meet ("\( \wedge \)") such that \( x' \lor x'' = (\max\{x'_i, x''_i\})_{i=1, \ldots, n} \) and \( x' \land x'' = (\min\{x'_i, x''_i\})_{i=1, \ldots, n} \) for any vectors \( x', x'' \in \mathbb{R}^n \). We let \( e_i \) be the vector with only the \( i \)-th entry being 1 and all others being 0, and \( 1 \) be the vector with all the entries being 1. Random variables are represented by characters with the tilde sign such as \( \tilde{z} \) with \( z \) being its realization.

## 2. Tractability of Two-stage Problems with Supermodularity

In this section, we explore computational tractability in a special class of two-stage distributionally robust linear optimization problems which exhibit the property of supermodularity.

**Model**

The decision maker faces a two-stage problem. In the first stage, the decision maker has to make the *here-and-now* decisions \( x \in \mathbb{R}^d \) before the uncertainty \( \tilde{z} \), an \( n \)-dimensional random vector, is realized. After that, the uncertainty is revealed and observed by the decision maker, who then moves to the second stage and makes the *wait-and-see* decisions \( y \in \mathbb{R}^m \). For a given first-stage decision \( x \) and an uncertainty realization \( z \), we denote the second-stage cost by \( g(x, z) \). It can be evaluated by the following linear programming problem,

\[
g(x, z) = \min_b \quad b^T y
\]

s.t. \( Wx + Uy \leq Vz + v^0 \),

where \( b \in \mathbb{R}^m \), \( W \in \mathbb{R}^{r \times l} \), \( U \in \mathbb{R}^{r \times m} \), \( V \in \mathbb{R}^{r \times n} \) and \( v^0 \in \mathbb{R}^r \) are given constants. For tractability, we only consider the right-hand-side uncertainty in the second-stage problem here, so that the uncertainty realization only affects the right-hand-side coefficients of the optimization problem (1). By contrast, the coefficients related to \( x \) and \( y \) are fixed. We let \( g(x, z) = \infty \) if Problem (1) is infeasible.

We consider the distributionally robust setting such that the true distribution of \( \tilde{z} \) is only known to belong to an ambiguity set \( \mathcal{F} \). Therefore, for a given first-stage decision \( x \), the expected second-stage cost is evaluated under the worst-case distribution and hence is

\[
\sup_{P \in \mathcal{F}} \mathbb{E}_P[g(x, \tilde{z})].
\]
By choosing the first-stage decision $\mathbf{x}$, the decision maker aims to minimize the sum of the deterministic first-stage cost and the worst-case expected second-stage cost. It can be formulated as

$$
\min_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbf{a}^T \mathbf{x} + \sup_{\mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[ g(\mathbf{x}, \mathbf{z}) \right] \right\},
$$

(2)

where $\mathbf{a} \in \mathbb{R}^l$ is a given constant vector, $\mathcal{X} \subseteq \mathbb{R}^l$ is the set of all feasible first-stage decisions.

In order to capture the distributional information of $\mathbf{z}$, we adopt the scenario-wise ambiguity set which was recently proposed by Chen et al. (2019). Specifically, we assume

$$
\mathcal{F} = \left\{ \mathbb{P} \left| \begin{array}{ll}
\mathbb{E}_{\mathbb{P}}[\mathbf{z} | \hat{k} = k] = \mathbf{\mu}^k, & \forall k \in [K] \\
\mathbb{E}_{\mathbb{P}}[\mathbf{z}_i | \hat{k} = k] \leq \delta_i^k, & \forall k \in [K], \forall i \in [n] \\
\mathbb{P}(\mathbf{z}^k \leq \mathbf{z} \leq \mathbf{z}^k | \hat{k} = k) = 1, & \forall k \in [K] \\
\mathbb{P}(\hat{k} = k) = q_k, & \forall k \in [K] \\
\mathbf{q} \in \mathcal{Q} 
\end{array} \right. \right\}.
$$

(3)

Here a random scenario $\hat{k}$ is introduced and its realization affects the distributional information of $\mathbf{z}$. In particular, if the random scenario is realized as $k \in [K]$, we have corresponding distributional information for $\mathbf{z}$: mean being $\mathbf{\mu}^k$, MAD of $\mathbf{z}_i$ being bounded by $\delta_i^k$ for all $i \in [n]$, and support being $[\mathbf{z}^k, \mathbf{z}^k]$. The probability that $\hat{k}$ realizes as $k$ is $q_k$. We also allow ambiguity in $\mathbf{q} = (q_k)_{k \in [K]}$ and we only know that $\mathbf{q}$ is in a given polyhedron $\mathcal{Q} = \{ \mathbf{q} | R\mathbf{q} \leq \mathbf{v}, \mathbf{q} \geq \mathbf{0} \}$. Since $\mathbf{q}$ represents the probability mass, we assume $\mathcal{Q} \subseteq \{ \mathbf{q} \in \mathbb{R}^K_+ \mid \mathbf{1}^T\mathbf{q} = 1 \}$.

We first consider the case of $K = 1$. The distributional ambiguity set $\mathcal{F}$ is reduced to the conventional one with means, supports and MADs information, which has been studied in the literature. When the uncertain variable is one-dimensional, the worst-case expectation has a decision independent expression if the objective function is convex (Ben-Tal and Hochman 1972). Postek et al. (2019) and Den Hertog and van Leeuwaarden (2019) use MAD information to focus on a special case where all the components of the random variables are independent. In practice, the MAD information is also easy to estimate (Postek et al. 2018).

The incorporation of random scenarios brings modeling flexibility and can capture a broad class of information in a more intuitive way, e.g., multi-modal distribution or covariate information, and can result in less conservative solutions than in the case with a fixed scenario. When the set $\mathcal{Q}$ is a singleton and $\delta_i^k = 0$ for any $k \in [K], i \in [n]$, the information set $\mathcal{F}$ reduces to the case with a known discrete distribution.

To explore the solvability of Problem (2), we will first investigate the worst-case distribution of $\mathbf{z}$ conditioning on a given scenario. After that, we provide a computationally tractable reformulation for Problem (2) with a random scenario, i.e., with $\mathcal{F}$ defined in Equation (3).
The case with a fixed scenario

When the scenario $\tilde{k}$ is realized as $k$ for some $k \in [K]$, we define $\mathcal{F}^k$ to be a set of probability distributions in this specific scenario. That is,

$$
\mathcal{F}^k = \left\{ \mathbb{P}^k \left| \begin{array}{l}
\mathbb{E}_{\mathbb{P}^k}[\tilde{z}] = \mu^k, \\
\mathbb{E}_{\mathbb{P}^k}[|\tilde{z} - \mu^k|] \leq \delta_i^k, \\
\mathbb{P}^k(\tilde{z}^k \leq \tilde{z} \leq \tilde{z}^k) = 1
\end{array} \right. \right\}.
$$

(4)

To obtain the worst-case distribution of $\tilde{z}$ in this case, we first show the convexity of $g$.

**Lemma 1** $g(x, z)$ is convex in $z$ for any given $x$.

Proof: For completeness, we show the proof here although it is straightforward. Consider any given $\alpha \in [0, 1], x, z', z''$. Let $y', y''$ be any feasible solutions in Problem (1) when $z = z', z''$, respectively. Then we have $Wx + Uy' \leq Vz' + v^0, Wx + Uy'' \leq Vz'' + v^0$. This implies $Wx + Uy^\alpha \leq Vz^\alpha + v^0$, where $y^\alpha = \alpha y' + (1 - \alpha)y''$. Hence, $y^\alpha$ is feasible in Problem (1) when $z = z^\alpha$, and $g(x, z^\alpha) \leq b^T y^\alpha = \alpha b^Ty' + (1 - \alpha)b^Ty''$. Taking minimal over all feasible $y', y''$, we have $g(x, z^\alpha) \leq \alpha g(x, z') + (1 - \alpha)g(x, z'')$. \hfill $\square$

With the convexity of $g$, the worst-case distribution in the case of $\tilde{k} = k$ has the following characteristics.

**Proposition 1** For any $x$, there exists $\mathbb{P}^{k*} \in \text{arg sup}_{\mathbb{P}^k \in \mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(x, \tilde{z})]$ such that for all $i \in [n]$, the marginal distribution is independent of $x$ and can be calculated as

$$
\mathbb{P}^{k*}(\tilde{z}^k_i = w) = \begin{cases} 
\frac{\delta^k_i}{2(\mu^k_i - \tilde{z}^k_i)} & \text{if } w = \tilde{z}^k_i \\
1 - \frac{\delta^k_i (\mu^k_i - \tilde{z}^k_i)}{2(\tilde{z}^k_i - \mu^k_i)(\mu^k_i - \tilde{z}^k_i)} & \text{if } w = \mu^k_i \\
\frac{\delta^k_i}{2(\tilde{z}^k_i - \mu^k_i)} & \text{if } w = \tilde{z}^k_i \\
0 & \text{otherwise.}
\end{cases}
$$

(5)

Proof: When the mean absolute deviation is known to be an exact value, Ben-Tal and Hochman (1972) show that there is a worst-case distribution with the structure of Equation (5). Here $\mathcal{F}^k$ has an upper bound, rather than the exact value, of the mean absolute deviation. It is not difficult to show that the worst-case distribution remains the same. Alternatively, the above result can also be derived from the standard duality approach of robust optimization. \hfill $\square$

According to Proposition 1, there exists a worst-case distribution such that at each dimension $i$, $i \in [n]$, the marginal distribution of $\tilde{z}_i$ has non-zero probability mass at only three points: the lower bound, mean and upper bound. Therefore, to evaluate $\sup_{\mathbb{P}^k \in \mathcal{F}^k} \mathbb{E}_{\mathbb{P}^k}[g(x, \tilde{z})]$, it suffices to focus on
the distributions with support \( \{ z \mid z_i \in \{ z^k_i, \mu^k_i, z_{-i}^k \}, \ i \in [n] \} \). Unfortunately, the number of points in this set is exponentially large in \( n \). It essentially renders the two-stage problem computationally challenging to solve. However, we next show that if the function \( g(x, z) \) is supermodular in \( z \), the computational burden can be eased.

Recall that given any function \( f : \mathbb{R}^n \to \mathbb{R} \), it is supermodular if

\[
f(w') + f(w'') \leq f(w' \land w'') + f(w' \lor w'')
\]

for any unordered \( w', w'' \in \mathbb{R}^n \). When the function has two dimensions and we know only the marginal distributions but not the joint distribution, it has been shown that the supermodularity leads to an explicit form of worst-case distribution.

**Lemma 2** (Rachev and Rüschendorf 1998) Consider any supermodular function \( f : \mathbb{R}^2 \to \mathbb{R} \), and any two-dimensional random vector \( \tilde{w} \) with the marginal cumulative distribution function for \( \tilde{w}_1, \tilde{w}_2 \) being \( F_1, F_2 \), respectively. Let \( \mathcal{P} = \{ \mathbb{P} \mid \mathbb{P}(\tilde{w}_1 \leq x) = F_i(x) \ \forall x \in \mathbb{R}, i = 1, 2 \} \) be the set of all possible distributions for \( \tilde{w} \). Then

\[
\mathbb{E}_{\mathbb{P}} [ f(\tilde{w}_1, \tilde{w}_2) ] \leq \int_0^1 f(F_1^{-1}(u), F_2^{-1}(u)) du \quad \forall \mathbb{P} \in \mathcal{P}.
\]

Clearly, the upper bound in Lemma 2 is achieved when \( (\tilde{w}_1, \tilde{w}_2) \overset{d}{=} (F_1^{-1}(\tilde{u}), F_2^{-1}(\tilde{u})) \) with \( \tilde{u} \) being a random variable following the uniform distribution \( U(0, 1) \). In this worst-case distribution, considering any two realizations of \( \tilde{w} \) and denote them by \( w', w'' \), we then have \( u', u'' \in [0, 1] \) such that \( w' = (F_1^{-1}(u'), F_2^{-1}(u')) \) and \( w'' = (F_1^{-1}(u''), F_2^{-1}(u'')) \). This implies \( w', w'' \), and hence all pairs of realizations are ordered. Intuitively, this is because we can move the probability mass of any unordered pair to their join and meet, such that the marginal distribution is unchanged and the expectation of \( f(\tilde{w}) \) increases due to the supermodularity of \( f \). Interestingly, this result can be extended to the case with general dimensions and significantly reduces the number of possible realizations for the worst-case distribution.

**Lemma 3** Consider any supermodular function \( f : \mathbb{R}^n \to \mathbb{R} \), and any \( n \)-dimensional discrete random vector \( \tilde{w} \). Let \( \mathcal{P} = \{ \mathbb{P} \mid \mathbb{P}(\tilde{w}_i = x_{ij}) = p_{ij} \ \forall j \in [m_i], i \in [n] \} \) be the ambiguity set based on the marginal distributions on all dimensions and characterized by \( m_i, i \in [n] \) and \( p_{ij} > 0, x_{ij}, j \in [m_i], i \in [n] \) with \( \sum_{j \in [m_i]} p_{ij} = 1, i \in [n] \). Then there exists \( \mathbb{P}^* \in \arg\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_\mathbb{P} [ f(\tilde{w}) ] \) such that the set \( \{ w \in \mathbb{R}^n \mid \mathbb{P}^*(\tilde{w} = w) > 0 \} \) forms a chain.
Proof: Consider any $\mathbb{P} \in \mathcal{P}$ such that there exists an unordered pair $\mathbf{w}', \mathbf{w}''$ with $p' = \mathbb{P}(\tilde{\mathbf{w}} = \mathbf{w}') > 0$, $p'' = \mathbb{P}(\tilde{\mathbf{w}} = \mathbf{w}'') > 0$. WLOG, assume $p' \leq p''$. We construct a new probability distribution $\mathbb{P}^o$, such that

$$
\mathbb{P}^o(\tilde{\mathbf{w}} = \mathbf{w}) = \begin{cases} 
0 & \text{if } \mathbf{w} = \mathbf{w}' \\
 p'' - p' & \text{if } \mathbf{w} = \mathbf{w}'' \\
 \mathbb{P}(\tilde{\mathbf{w}} = \mathbf{w}' \land \mathbf{w}'') + p' & \text{if } \mathbf{w} = \mathbf{w}' \land \mathbf{w}'' \\
 \mathbb{P}(\tilde{\mathbf{w}} = \mathbf{w}' \lor \mathbf{w}'') + p' & \text{if } \mathbf{w} = \mathbf{w}' \lor \mathbf{w}'' \\
 \mathbb{P}(\tilde{\mathbf{w}} = \mathbf{w}) & \text{otherwise.}
\end{cases}
$$

In particular, based on $\mathbb{P}$, we move the probability mass $p'$ from the realization of $\mathbf{w}'$, $\mathbf{w}''$ to $\mathbf{w}' \land \mathbf{w}'', \mathbf{w}' \lor \mathbf{w}''$. That does not change the marginal distribution and hence $\mathbb{P}^o \in \mathcal{P}$. Moreover, compared with the support of $\mathbb{P}$, that of $\mathbb{P}^o$ has one less unordered pair. We also observe that

$$
\mathbb{E}_{\mathbb{P}^o} [f(\tilde{\mathbf{w}})] - \mathbb{E}_{\mathbb{P}} [f(\tilde{\mathbf{w}})] = p' \left( f(\mathbf{w}' \land \mathbf{w}'') + f(\mathbf{w}' \lor \mathbf{w}'') - f(\mathbf{w}') - f(\mathbf{w}'') \right) \geq 0,
$$

where the last inequality is due to the supermodularity of $f$. Therefore, we can always reduce the number of unordered pairs (if there is any) in the support while the value of expectation on $f(\tilde{\mathbf{w}})$ either increases or remains unchanged. Finally we obtain $\mathbb{P}^* \in \arg \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [f(\tilde{\mathbf{w}})]$ such that the support of $\mathbb{P}^*$ has no unordered pair.

We notice that a recent work by Chen et al. (2018a) also develop a similar result as Lemma 3. Intuitively, when moving the same amount of probability mass from any two points $\mathbf{w}', \mathbf{w}''$ to $\mathbf{w}' \land \mathbf{w}'', \mathbf{w}' \lor \mathbf{w}''$, the marginal distribution does not change but the expectation of $f(\tilde{\mathbf{w}})$ is higher because of the supermodularity of $f$. Hence, a worst-case distribution is to move all probability mass from the unordered pair to their join and meet. This seemingly leads to a worst-case distribution that is highly positively correlated and hence not always realistic in some applications. However, we later will show that our adoption of the scenario-wise ambiguity set would address this issue and the worst-case distribution in our model can be correlated in any way.

Recall that according to Proposition 1, the worst-case distribution for $\sup_{p \in \mathcal{F}_k} \mathbb{E}_{\mathcal{F}_k} [g(\mathbf{x}, \mathbf{z})]$ has an explicit three-point distribution in each dimension. We are now ready to use Proposition 1 and Lemma 3 to explore this worst-case distribution.

**Proposition 2** If $g(\mathbf{x}, \mathbf{z})$ is supermodular in $\mathbf{z}$ for any $\mathbf{x}$, we have $\sup_{p \in \mathcal{F}_k} \mathbb{E} [g(\mathbf{x}, \mathbf{z})] = \sum_{i \in [2n+1]} p_i g(\mathbf{x}, z^i)$ for any $\mathbf{x}$, where $p, z^i, i \in [2n+1]$ are output by Algorithm 1 when the input is $\mathcal{F}_k$.

**Proof**: For notational simplicity, we drop the superscript $k$ which represents the scenario $k$; we also assume $\mathbf{z}_i = -1, \mu_i = 0, \tilde{z}_i = 1$ for all $i \in [n]$, the general case can be proved in the same way.

During the progress of this algorithm, for each $j \in [2n+1]$, we define $\mathbf{mp}^{j,i}$, which stands for the remaining marginal probability for iteration $j$ at dimension $i$, as

$$
\mathbf{mp}^{j,i} = \begin{cases} 
q_{i}^{j} & \text{if } z_{i}^{j} = 1 \quad (\mathbf{mp}^{j,i} \in \mathbb{R} \text{ in this case}) \\
(q_{i}^{j}, \mathbb{P}^*(\tilde{z}_i = 1)) & \text{if } z_{i}^{j} = 0 \quad (\mathbf{mp}^{j,i} \in \mathbb{R}^2 \text{ in this case}) \\
(q_{i}^{j}, \mathbb{P}^*(\tilde{z}_i = 0), \mathbb{P}^*(\tilde{z}_i = 1)) & \text{if } z_{i}^{j} = -1 \quad (\mathbf{mp}^{j,i} \in \mathbb{R}^3 \text{ in this case})
\end{cases}
$$
Algorithm 1 algorithm for worst case distribution
1: **Input:** $\mathcal{F}^k$ in Equation (4) with given $\mu^k$, $\delta^k$, $\tilde{z}^k$, $\tilde{x}^k$
2: **Initialization:**
   - denote $\mathbb{P}^{k^*}$ as the worst-case distribution in Proposition 1 and calculate $\mathbb{P}^k(\tilde{z}_i = w)$ for $w \in \{\tilde{z}_i, \mu^k_i, \tilde{z}^k\}$, $i \in [n]$ using Equation (5)
   - $z^i = \tilde{z}^k$, $q^i = (\mathbb{P}^k(\tilde{z}_1^k = \tilde{z}^k), \mathbb{P}^k(\tilde{z}_2^k = \tilde{z}^k), \ldots, \mathbb{P}^k(\tilde{z}_n^k = \tilde{z}^k))$, $p_1 = \min\{q_1^i, \ldots, q_n^i\}$ and $j = 1$
3: while $j \leq 2n$
   4: choose $r_j$ as the minimal index in $[n]$ such that $q_{r_j}^j = p_j$
   5: $z^{j+1} = z^j$, $q^{j+1} = q^j - p_j 1$
   6: update $z_{r_j}^{j+1} = \mu^k_{r_j}$ if its existing value is $z^k_{r_j}$, and $z_{r_j}^{j+1} = \tilde{z}^k_{r_j}$ if its existing value is $\mu^k_{r_j}$
   7: update $q_{r_j}^{j+1} = \mathbb{P}^{k^*}(\tilde{z}_{r_j} = z_{r_j}^{j+1})$
   8: $p_{j+1} = \min\{q_1^{j+1}, q_2^{j+1}, \ldots, q_n^{j+1}\}$
   9: update $j = j + 1$
10: **return** $z^1, z^2, \ldots, z^{2n+1}$ and $p = (p_1, p_2, \ldots, p_{2n+1})$

We also define $c_j = 1^T \mathbf{m} p^{j,i}$ which represents the remaining total probability mass. Correspondingly, we denote the set of information $\mathcal{I}^i = \{z^i, \mathbf{m} p^{j,i}, \ldots, \mathbf{m} p^{j,n}, c_j\}$.

Given a set of information $\mathcal{I}^i$, we say it is valid if it satisfies the following four conditions: 1) $z^i \in \{-1,0,1\}^n$; 2) $\mathbf{m} p^{j,i} \in [0, 1]^{2-s^i} \forall i \in [n]$; 3) $\mathbf{m} p^{j,i} > 0 \forall i \in [n]$ where we denote $\mathbf{m} p^{j,i}$ as the last element of the vector $\mathbf{m} p^{j,i}$; and 4) $1^T \mathbf{m} p^{j,i} = c_j \forall i \in [n]$.

By induction, we now show that $\mathcal{I}^i$ is valid for all $j \in [2n+1]$. First, when $j = 1$, the conditions 1), 2) and 3) are obviously satisfied. The condition 4) is also satisfied since $1^T \mathbf{m} p^{j,i} = \mathbb{P}^*(\tilde{z}_i = -1) + \mathbb{P}^*(\tilde{z}_i = 0) + \mathbb{P}^*(\tilde{z}_i = 1) = 1$ for all $i \in [n]$, and $c_1 = 1$.

Suppose $\mathcal{I}^j$ is valid for some $j \in [2n]$. Based on the algorithm, the elements in $\mathcal{I}^{j+1}$ are obtained as follows. First, $p_j = \min\{\mathbf{m} p^{j,i}, \ldots, \mathbf{m} p^{j,n}\}$, $r_j = \min\{i \in [n] : \mathbf{m} p^{j,i} = p_j\}$. After that, $z^{j+1} = z^j + e_{r_j}$.

We now prove that $z^{j+1}$ is not 1 by contradiction. Assume to the contrary, i.e., $z^{j+1} = 1$, then $\mathbf{m} p^{j,r_j} \in \mathbb{R}$, we have $c_j = 1^T \mathbf{m} p^{j,r_j} = \mathbf{m} p^{j,r_j} = p_j$. For any $i \in [n] \setminus \{r_j\}$, we observe i) $\mathbf{m} p^{j,i} \geq p_j = c_j$ (the inequality is because of our choice of $p_j$); ii) $\mathbf{m} p^{j,i} > 0$; and iii) $1^T \mathbf{m} p^{j,i} = c_j$ and $\mathbf{m} p^{j,i} \geq 0$.

The last two observations are because $\mathcal{I}^j$ is valid and hence satisfies conditions 2), 3) and 4).

Hence, we have $\mathbf{m} p^{j,i} \in \mathbb{R}$ and then $z^{j+1} = z^j + e_{r_j}$. That implies $z^j = 1$. We notice that for any $t \in [j-1]$, $z^{t+1} = z^t + e_{r_j}$ for some $i \in [n]$. So moving from $z^1 = -1$ to $z^j = 1$ requires $2n$ steps, i.e., $j = 2n + 1$, which contradicts $j \in [2n]$. Hence, $z^{j+1} = 1$ is false, and we must have $z^{j+1} \in \{-1, 0\}$. We can conclude that $z^{j+1} = z^j + e_{r_j} \in \{-1, 0, 1\}^n$, the condition 1) is satisfied for $\mathcal{I}^{j+1}$. As a result, condition 2) is obviously satisfied by the way $\mathbf{m} p^{j,i}$ is calculated.
With the algorithm, we know $\mathbf{m}^{j+1,r_j}$ can be obtained from the vector of $\mathbf{m}^{j,r_j}$ by removing the first component. Therefore, $\mathbf{m}_\text{end}^{j+1,r_j} = \mathbf{m}_\text{end}^{j,r_j} > 0$, the condition 3) is satisfied when $i = r_j$. Moreover, $1^T \mathbf{m}^{j+1,r_j} = 1^T \mathbf{m}_1^{j,r_j} - \mathbf{m}_1^{j,r_j} = c_j - p_j$. We also observe $c_j - p_j = 1^T \mathbf{m}^{j+1,r_j} \geq \mathbf{m}_\text{end}^{j+1,r_j} > 0$ and hence $c_j > p_j$.

For any $i \in [n] \setminus \{r_j\}$, since $z_i^{j+1} = z_i^j$, $\mathbf{m}^{j+1,i}$ and $\mathbf{m}^{j,i}$ are both of dimension $(2 - z_i^{j+1,i})$, they differ only at the first dimension; in particular,

$$
\mathbf{m}_s^{j+1,i} = \begin{cases}
\mathbf{m}_s^{j,i} - p_j & \text{if } s = 1 \\
\mathbf{m}_s^{j,i} & \text{if } z_i^{j+1} \in \{-1, 0\} \text{ and } s \neq 1
\end{cases}.
$$

We note that if $z_i^{j+1} = z_i^j = 1$, then $\mathbf{m}^{j,i}, \mathbf{m}^{j+1,i} \in \mathbb{R}_+$, and $\mathbf{m}_1^{j,i} = 1^T \mathbf{m}_1^{j,i} = c_j > p_j$, $\mathbf{m}_1^{j+1,i} = \mathbf{m}_1^{j,i} - p_j > 0$. If $z_i^{j+1} = z_i^j \in \{-1, 0\}$, obviously $\mathbf{m}_1^{j+1,i} = \mathbf{m}_1^{j,i} > 0$. Therefore, condition 3) is satisfied for $i$. Moreover, by Equation (6) we also know $1^T \mathbf{m}^{j+1,i} = 1^T \mathbf{m}^{j,i} - p_j = c_j - p_j$. Since we have previously obtained $1^T \mathbf{m}^{j+1,r_j} = c_j - p_j$, condition 4) is also satisfied. We conclude $\mathcal{I}^{j+1}$ is also valid and it finishes the induction, i.e., $\mathcal{I}^j$ is valid for all $j \in [2n + 1]$.

Now, for any $j \in [2n + 1]$, we define $\mathcal{Q}^j$ as the set of all mass functions with the marginal mass given by $\mathbf{m}^{i,1}, \ldots, \mathbf{m}^{i,n}$ and the possible realizations forming a chain. More specifically, define $\mathbf{w}^{j,i} \in \{-1, 0, 1\}^{2 - z_i^j}$ by

$$
\mathbf{w}^{j,i} = \begin{cases}
(-1, 0, 1) & \text{if } z_i^j = -1 \\
(0, 1) & \text{if } z_i^j = 0 \\
1 & \text{if } z_i^j = 1
\end{cases},
$$

which is the vector of all possible realization at dimension $i$, $i \in [n]$, and $\mathcal{W}_i^j = \{z \mid z_i^j \leq z \leq 1\} \cap \{-1, 0, 1\}^n$ which is the set of all possible realization of vector $z$; then

$$
\mathcal{Q}^j = \left\{ f^j : \mathcal{W}_i^j \rightarrow [0, 1] \mid \sum_{z \in \mathcal{W}_i^j : z_i = z_i^j} f^j(z) = \mathbf{m}_1^{j,i}, \{z \mid f^j(z) > 0\} \text{ forms a chain} \right\}.
$$

Noticing that $\mathcal{W}_i^{j+1} = \{z \in \mathcal{W}_i^j : z_{r_j} \neq z_{r_j}^j \}$, we define another set $\hat{\mathcal{Q}}^j$ by

$$
\hat{\mathcal{Q}}^j = \left\{ f^j : \mathcal{W}_i^j \rightarrow [0, 1] \mid \begin{array}{l}
f^j(z_i^j) = p_j \\
f^j(z) = 0 & \forall z \in \mathcal{W}_i^j \text{ with } z_{r_j} = z_{r_j}^j \text{ and } z \neq z_i^j \\
f^j(z) = f^j(z) & \forall z \in \mathcal{W}_i^{j+1} \\
f^j+1 \in \mathcal{Q}^{j+1}
\end{array} \right\}.
$$

We next prove $\mathcal{Q}^j = \hat{\mathcal{Q}}^j$.

First, consider any $f^j \in \mathcal{Q}^j$. Suppose $\exists z^o \in \mathcal{W}_i^j$ with $z_{r_j}^o = z_{r_j}^j$ and $z^o \neq z_i^j$ such that $f^j(z^o) > 0$. That implies the existence of $s \in [n] \setminus \{r_j\}$ such that $z_s^o > z_s^j$. Hence,

$$
\sum_{z \in \mathcal{W}_i^j : z_{r_j} = z_{r_j}^j, z_s = z_s^j} f^j(z) = \sum_{z \in \mathcal{W}_i^j : z_{r_j} = z_{r_j}^j} f^j(z) - \sum_{z \in \mathcal{W}_i^j : z_{r_j} = z_{r_j}^j, z_s \neq z_s^j} f^j(z) = p_j - \sum_{z \in \mathcal{W}_i^j : z_{r_j} = z_{r_j}^j, z_s \neq z_s^j} f^j(z) \leq p_j - f^j(z^o) < p_j,
$$
\[
\sum_{z \in \mathcal{W}^j: z_r = z_j^*} f^i(z) = \sum_{z \in \mathcal{W}^j: z_r = z_j^*} f^i(z) - \sum_{z \in \mathcal{W}^j: z_r = z_j^*} f^i(z) \geq p_j - \sum_{z \in \mathcal{W}^j: z_r = z_j^*} f^i(z) > 0.
\]

Therefore, we have \( z^* \in \mathcal{W}^j \) such that \( z_r^* > z_{r_j}^* = z_j^* \), \( z_r^* = z_j^* < z_0^* \) and \( f^j(z^*) > 0 \). This contradicts \( \{z \mid f^j(z) > 0\} \) forming a chain. Therefore, \( f^j(z) = 0 \ \forall z \in \mathcal{W}^j \) with \( z_r = z_j^* \) and \( z \neq z^j \), and \( f^j(z^j) = m_{p_1}^{r_j} - \sum_{z \in \mathcal{W}^j: z_r = z_j^*} f^j(z) = p_j - 0 = p_j \). Therefore, \( f^j \) satisfies the first two conditions in \( \hat{Q}^j \). The corresponding \( f^{j+1} \) is in \( Q^{j+1} \) can be easily verified by showing the chain structure and checking the equality constraints on the marginal mass. Hence, we have \( f^j \in \hat{Q}^j \).

We now prove the reverse. Consider any \( f^j \in \hat{Q}^j \) and we check whether it satisfies the two conditions in \( Q^j \). The first condition, which is on the marginal mass, can be verified by standard algebra. The second condition, which is on the chain structure, is straightforward. Therefore, we have \( f^j \in Q^j \). We can conclude that \( Q^j = \hat{Q}^j \) for all \( j \in [2n + 1] \).

Finally, by representing \( Q^j \) in the form of \( \hat{Q}^j \), with recursion we can easily get

\[
Q^1 = \left\{ f: \mathcal{W}^j \to [0, 1] \mid \begin{align*}
&f(z^i) = p_i, \quad i \in [2n] \\
&f(z) = 0, \quad \forall z \in \mathcal{W}^1 \setminus \{z^i, i \in [2n]\} \setminus \mathcal{W}^{2n+1} \\
&\hat{f} \in Q^{2n+1}
\end{align*} \right\} \quad (7)
\]

We note that since \( z^j \in \{-1, 0, 1\}^n \), \( z^1 = -1 \), and any time the movement from \( z^j \) to \( z^{j+1} \) is to increase one dimension by 1 while maintaining other dimensions unchanged, and hence we have \( z^{2n+1} = 1 \). Therefore, \( \mathcal{W}^{2n+1} = \{z^{2n+1}\} \). Then by Equation (7), we have

\[
Q^1 = \left\{ f: \mathcal{W}^j \to [0, 1] \mid \begin{align*}
&f(z^i) = p_i, \quad i \in [2n+1] \\
&f(z) = 0, \quad \forall z \in \mathcal{W}^1 \setminus \{z^i, i \in [2n+1]\}
\end{align*} \right\}
\]

Hence, the result is proved. \( \square \)

Specifically, Algorithm 1 returns the worst-case distribution, which has support on only 2\( n \) + 1 points. Moreover, this worst-case distribution depends only on the information set \( \mathcal{F}^k \) rather than the first-stage decision \( \mathbf{x} \). Hence, given the scenario \( \tilde{k} \) realizing as \( k \), the second-stage cost can be obtained.

**Incorporating the uncertain scenario**

In solving the general two-stage optimization problem (2), Proposition 2 shows how to evaluate the second-stage expected cost efficiently under the worst-case distribution when the uncertain scenario realizes as \( k \). We now incorporate the uncertainty in the scenario \( \tilde{k} \).

Based on the definition of \( \mathcal{F} \) and \( \mathcal{F}^k \) in Equations (3) and (4), we have that, for any \( \mathcal{P} \in \mathcal{F} \), there exists \( q \in Q, \mathcal{P}^k \in \mathcal{F}^k \), \( k \in [K] \) such that

\[
\mathcal{P} (\tilde{z} \leq z) = \sum_{k \in [K]} \mathcal{P} (\tilde{k} = k) \cdot \mathcal{P} (\tilde{z} \leq z | \tilde{k} = k) = \sum_{k \in [K]} q_k \mathcal{P}^k (\tilde{z} \leq z) \quad \forall z \in \mathbb{R}^n.
\]
Therefore,

$$\sup_{\mathcal{P} \in \mathcal{P}} \mathbb{E}_\mathcal{P}[g(x, \tilde{z})] = \max_{q \in \mathcal{Q}} \mathbb{E}_\mathcal{Q} \left[ \sum_{k \in [K]} q_k \mathbb{E}_{\mathcal{P}^k}[g(x, \tilde{z})] \right] = \max_{q \in \mathcal{Q}} \sum_{k \in [K]} q_k \sup_{\mathcal{P}^k \in \mathcal{P}^k} \mathbb{E}_{\mathcal{P}^k}[g(x, \tilde{z})]. \quad (8)$$

We denote $z^{k,1}, \ldots, z^{k,2n+1}, p^k$ as the output of Algorithm 1 with input $\mathcal{F}^k$, $k \in [K]$. Observing that $\mathcal{Q}$ is a polyhedron, we then have the following reformulation.

**Theorem 1** If $g(x, z)$ is supermodular in $z$ for any $x$, Problem (2) is equivalent to the following linear programming problem,

$$\begin{align*}
\min & \quad a^T x + \nu^T l \\
\text{s.t.} & \quad R_k^T l \geq \sum_{i \in [2n+1]} p_i^k b^T y^{k,i}, \quad k \in [K] \\
& \quad W x + U y^{k,i} \leq V z^{k,i} + v^0, \quad k \in [K], i \in [2n+1] \\
& \quad l \geq 0, \\
& \quad x \in \mathcal{X}.
\end{align*} \quad (9)$$

**Proof:** Within the objective function in Problem (2), the second-stage cost, $\sup_{\mathcal{P} \in \mathcal{P}} \mathbb{E}_\mathcal{P}[g(x, \tilde{z})]$, can be reformulated as follows.

$$\begin{align*}
\sup_{\mathcal{P} \in \mathcal{P}} \mathbb{E}_\mathcal{P}[g(x, \tilde{z})] &= \max_{q \in \mathcal{Q}} \sum_{k \in [K]} q_k \sup_{\mathcal{P}^k \in \mathcal{P}^k} \mathbb{E}_{\mathcal{P}^k}[g(x, \tilde{z})] \\
&= \min \nu^T l \\
& \quad \text{s.t.} \quad R_k^T l \geq \sup_{\mathcal{P}^k \in \mathcal{P}^k} \mathbb{E}_{\mathcal{P}^k}[g(x, \tilde{z})], \quad k \in [K] \\
& \quad l \geq 0 \\
&= \min \nu^T l \\
& \quad \text{s.t.} \quad R_k^T l \geq \sum_{i \in [2n+1]} p_i^k g(x, z^{k,i}), \quad k \in [K] \\
& \quad l \geq 0 \\
&= \min \nu^T l \\
& \quad \text{s.t.} \quad R_k^T l \geq \sum_{i \in [2n+1]} p_i^k b^T y^{k,i}, \quad k \in [K] \\
& \quad W x + U y^{k,i} \leq V z^{k,i} + v^0, \quad k \in [K], i \in [2n+1] \\
& \quad l \geq 0
\end{align*}$$

Here the first equality follows from Equation (8) and the assumption on $\mathcal{Q}$, the second equality is due to strong duality, the third holds by Proposition 2 and the assumption that $g(x, z)$ is supermodular in $z$, and we have the last equality by inserting the definition of $g(x, z)$. With this reformulation, we can transform Problem (2) to Problem (9). \hfill \square

Intuitively, the reformulation in Theorem 1 incorporates all possible scenarios in the worst-case distribution, and assigns one corresponding second-stage decision to each of those scenarios.
Therefore, the two-stage problem can be formulated as a static linear programming problem. Nevertheless, the classic approach using this idea has to handle an exponential number of scenarios, leading to computational intractability. Here, by exploring the potential property of supermodularity in the uncertainties, we reduce the number of scenarios to $K(2^n + 1)$, which is of polynomial size and hence the problem becomes tractable.

We also remark that as mentioned before, given a scenario realization $k$, the worst-case distribution has positive correlation since we move probability mass to ordered pairs in general. However, by incorporating the random scenario, the correlation between any pair of uncertain factors can have an arbitrary value in $[-1, 1]$. Consider the following example for illustration. Suppose there are only two scenarios (i.e., $K = 2$), the probability for each scenario to happen is known to be half (i.e., $Q = \{(1/2, 1/2)\}$), $\mathbf{z}^1 = (0, 100)$, $\mathbf{z}^2 = (100, 0)$, $\mathbf{z}^2 = (101, 1)$. Then the worst-case distribution must be the one with half of the chance $(0, 100) \leq \mathbf{z} \leq (1, 101)$, and the other half of the chance $(100, 0) \leq \mathbf{z} \leq (101, 1)$. Hence, the correlation coefficient is close to $-1$.

3. Conditions for supermodularity of the second-stage problems

If the second-stage cost, $g(\mathbf{x}, \mathbf{z})$, is supermodular in $\mathbf{z}$, Section 2 has shown that a tractable formulation can be achieved. Unfortunately, unlike convexity, supermodularity in $\mathbf{z}$ is not a feature embedded in all two-stage problems. It depends on the structure of the two-stage problem. In this section, we aim to identify a broad class of two-stage problems such that their second-stage cost is supermodular in the uncertain factor.

We reformulate the second-stage cost, $g(\mathbf{x}, \mathbf{z})$ as

$$
g(\mathbf{x}, \mathbf{z}) = \min \ b^T y
$$

s.t. $Uy - Vz \leq -Wx + v^0. \tag{10}$

It is the optimal value of a parametric optimization problem which is parametrized by $\mathbf{z}$, and we need to explore the supermodularity in this parameter. Note that here we do not consider $\mathbf{x}$ as a parameter to parameterize the optimization problem since we focus on the supermodularity on $\mathbf{z}$ but not on $\mathbf{x}$, and hence we put it on the right-hand-side of the constraint. In the parametric optimization problem, the supermodularity of the optimal value in parameters has been studied when the optimization is a maximization. Nevertheless, in Equation (10), we have a minimization problem, which leads to the essential difference from previous studies. Typically, the lattice structure of the feasible set is a key for supermodularity in the parametric maximization problem. By contrast, to investigate the parametric minimization problem, we need to introduce the following concept called the inverse additive lattice.
Definition 1 Given two positive integers \( m, n \), a set \( S \subseteq \mathbb{R}^m \times \mathbb{R}^n \) is an inverse additive lattice if for any \( p, q \in \mathbb{R}^n \), \( z', z'' \in \mathbb{R}^n \) with \( (p, z' \land z'') \), \( (q, z' \lor z'') \) \( \in S \), there exist \( y', y'' \in \mathbb{R}^m \) such that \( (y', z') \), \( (y'', z'') \) \( \in S \) and \( y' + y'' = p + q \).

We now show that the inverse additive lattice is a necessary and sufficient condition for supermodularity in the parametric minimization problem. Given any first-stage decision \( x \), we denote the set of all feasible pairs of \( (y, z) \) as \( S(x) \), i.e., \( S(x) = \{(y, z) \mid Uy - Vz \leq -Wx + v^0\} \).

Theorem 2 Given any \( x \), \( g(x, z) \) is supermodular in \( z \) for any \( b \) if and only if \( S(x) \) is an inverse additive lattice.

Proof: We first prove the “if” part. Suppose \( S(x) \) is an inverse additive lattice, then given any \( z', z'', p, q \) with \( (p, z' \land z'') \), \( (q, z' \lor z'') \) \( \in S(x) \), there exist \( y', y'' \) such that \( (y', z') \), \( (y'', z'') \) \( \in S(x) \) and \( y' + y'' = p + q \). We then have
\[
g(x, z') + g(x, z'') \leq b^T y' + b^T y'' = b^T p + b^T q.
\]

Taking the minimum on the right-hand-side over all \( p, q \) with \( (p, z' \land z'') \), \( (q, z' \lor z'') \) \( \in S(x) \), we obtain \( g(x, z') + g(x, z'') \leq g(x, z' \land z'') + g(x, z' \lor z'') \).

Next we prove the “only if” part by contradiction. Suppose \( S(x) \) is not an inverse additive lattice, then there exist \( z', z'', p, q \) with \( (p, z' \land z'') \), \( (q, z' \lor z'') \) \( \in S(x) \) but \( p + q \notin W \), where the set \( W \) is defined as \( W = \{ r + s \mid (r, z'), (s, z'') \in S(x) \} \). According to the definition of \( S(x) \), we can easily see that \( W \) is convex and closed. By the Hyperplane Separation Theorem, there exist a vector \( \eta \) and a real number \( \lambda \) such that,
\[
\eta^T (p + q) < \lambda < \eta^T w \quad \forall w \in W.
\]

Consider the second-stage cost function \( g(x, z) \) (defined in Equation (1)) with coefficient \( b = \eta \). We have
\[
g(x, z') + g(x, z'') = \min \{ \eta^T y \mid (y, z') \in S(x) \} + \min \{ \eta^T y \mid (y, z'') \in S(x) \}
= \min \{ \eta^T (r + s) \mid (r, z'), (s, z'') \in S(x) \}
= \min \{ \eta^T w \mid w \in W \} > \lambda,
\]
\[
g(x, z' \land z'') + g(x, z' \lor z'') = \min \{ \eta^T y \mid (y, z' \land z'') \in S(x) \} + \min \{ \eta^T y \mid (y, z' \lor z'') \in S(x) \}
\leq \eta^T (p + q) < \lambda.
\]

Therefore, \( g(x, z') + g(x, z'') > g(x, z' \land z'') + g(x, z' \lor z'') \), which contradicts the supermodularity. The “only if” part is completed. \( \square \)
Theorem 2 presents a necessary and sufficient condition for the second-stage cost being supermodular in the uncertainty realization \( z \) for a given first-stage decision \( x \). Now it remains to characterize the structure of the second-stage problem such that the condition can always be satisfied for any \( x \).

**Theorem 3** \( g(x, z) \) is supermodular in \( z \) for any \( x, b \) and \( v^0 \) if and only if \( U \in \mathbb{R}^{r \times m} \) and \( V \in \mathbb{R}^{r \times n} \) satisfy one of the following conditions:

1) \( \text{rank}(U) = r \),

2) for all \( I \subseteq [r] \), \( \beta \in \mathbb{R}_+^d \) satisfying \( |I| = \text{rank}(U) + 1 \), \( \text{rank}(U_I) = \text{rank}(U) \) and \( V_I \beta \in \text{span}(U_I) \), we must have \( \beta_i(V_I), e \in \text{span}(U_I) \) holds for every \( i \in [n] \).

**Proof**: Based on Theorem 2, the above theorem is equivalent to this statement: \( S(x) \) is an additive inverse lattice for all \( x \) and \( v^0 \) if and only if \( U \) and \( V \) satisfy one of the two conditions in the above theorem. We prove the equivalent statement as follows.

First we prove the “if” direction by contradiction. Suppose there exist \( x \) and \( v^0 \) such that \( S(x) \) is not an additive inverse lattice, i.e., we have \( z', z'', p, q \) with \( z^\land = z' \land z'' \), \( z^\lor = z' \lor z'' \) and \( (p, z^\land), (q, z^\lor) \in S(x) \), such that \( y' + y'' \neq p + q \) holds for all \( y', y'' \) with \( (y', z'), (y'', z'') \in S(x) \).

We denote \( c = -Wx + v^0 \), \( t^1 = Up - Vz^\land \leq c \), \( t^2 = Uq - Vz^\lor \leq c \). Here the two inequalities are due to \( (p, z^\land), (q, z^\lor) \in S(x) \) and the definition of \( S(x) \). We define a set \( W \) as

\[
W = \{ y \in \mathbb{R}^m \mid (t^1 \land t^2) + Vz' \leq Uy \leq (t^1 \lor t^2) + Vz' \}.
\]

Note that \( W \) should be an empty set, otherwise there exists a \( y^0 \in W \) and hence

\[
Uy^0 - Vz' \leq (t^1 \lor t^2) \leq c,
\]

\[
U(p + q - y^0) - Vz'' = Up - Vz^\lor + Uq - Vz^\land - (Uy^0 - Vz') \leq t^1 + t^2 - (t^1 \land t^2) = t^1 \lor t^2 \leq c,
\]

which implies both \( (y^0, z'), (p + q - y^0, z'') \in S(x) \), and contradicts the previous statement on \( y', y'' \) resulting from the assumption.

We now show that the first part of the condition in our theorem is not true. If \( \text{rank}(U) = r \), we can solve \( y \) with \( Uy = (t^1 \land t^2) + Vz' \leq (t^1 \lor t^2) + Vz' \), which contradicts the emptiness of \( W \). Therefore, \( \text{rank}(U) < r \).

We then focus on the second part of the condition in our theorem. The emptiness of \( W \) leads to the infeasibility of the following optimization problem:

\[
\begin{align*}
\max & \quad 0 \\
\text{s.t.} & \quad \begin{bmatrix} U & -U \end{bmatrix} y \leq \begin{bmatrix} (t^1 \lor t^2) + Vz' \\ -(t^1 \land t^2) - Vz' \end{bmatrix}.
\end{align*}
\]
Furthermore, by Lemma 6 we know that there exists \( \mathcal{I} \subseteq [r] \), \( |\mathcal{I}| = \text{rank}(U) + 1 \) with \( \text{rank}(U_\mathcal{I}) = \text{rank}(U) \) such that the problem

\[
\begin{align*}
\text{max} & \quad 0 \\
\text{s.t.} & \quad \begin{bmatrix} U_\mathcal{I} \\ -U_\mathcal{I} \end{bmatrix} y \leq \begin{bmatrix} (t_1^2 \lor t_2^2) + V_\mathcal{I} z' \\ -(t_1^1 \land t_2^2) - V_\mathcal{I} z' \end{bmatrix}
\end{align*}
\]  

(11)

is also infeasible. We write the dual of (11) as follows,

\[
\begin{align*}
\min & \quad r^T ((t_1^2 \lor t_2^2) + V_\mathcal{I} z') - s^T ((t_1^1 \land t_2^2) + V_\mathcal{I} z') \\
\text{s.t.} & \quad U_\mathcal{I}^T (r - s) = 0 \\
& \quad r, s \geq 0.
\end{align*}
\]

(12)

Observing that \( r = s = 0 \) gives a feasible solution of (12), the infeasibility of the primal problem implies the unboundedness of the above dual problem. Therefore, there exist \( r, s \geq 0 \) with \( U_\mathcal{I}^T (r - s) = 0 \) such that the following inequalities holds,

\[
0 > r^T ((t_1^2 \lor t_2^2) + V_\mathcal{I} z') - s^T ((t_1^1 \land t_2^2) + V_\mathcal{I} z')
\]

\[
= r^T ((t_1^2 \lor t_2^2) + V_\mathcal{I} z' - U_\mathcal{I} q) - s^T ((t_1^1 \land t_2^2) + V_\mathcal{I} z' - U_\mathcal{I} q)
\]

\[
\geq r^T (t_1^2 + V_\mathcal{I} z' - U_\mathcal{I} q) - s^T (t_2^2 + V_\mathcal{I} z' - U_\mathcal{I} q)
\]

\[
= (r - s)^T V_\mathcal{I} (z' - z^\lor),
\]

where the first inequality is obtained from the unboundedness of (12), the first equality is due to \( U_\mathcal{I}^T (r - s) = 0 \), the second inequality follows from \( t_1^1 \land t_2^2 \leq t_1^1 \lor t_2^2 \), and the second equality comes from \( t_1^2 = U_\mathcal{I} q - V_\mathcal{I} z^\lor \). We remark that in the above equation, if we use \( U_\mathcal{I} p \) instead of \( U_\mathcal{I} q \) in the first equality, and \( t_1^1 \) instead of \( t_2^2 \) in the second inequality, then \( 0 > (r - s)^T V_\mathcal{I} (z' - z^\land) \) can be obtained similarly.

We define \( \Delta_1 = (r - s)^T V_\mathcal{I} (z' - z^\lor) \), \( \Delta_2 = (r - s)^T V_\mathcal{I} (z' - z^\land) \), and \( \beta = \frac{z' - z^\lor}{\Delta_1} - \frac{z' - z^\land}{\Delta_2} \).

We have three observations on \( \beta \). First, \( \beta \geq 0 \) since \( \Delta_1, \Delta_2 < 0 \) and \( z^\land \leq z' \leq z^\lor \).

Second, \( V_\mathcal{I} \beta \in \text{span}(U_\mathcal{I}) \). To see this, recall that for any matrix, its column space is the orthogonal complement of the null space of its transpose; therefore, we can equivalently show that null \( (U_\mathcal{I}^T)^\perp \subseteq \text{null} (V_\mathcal{I} \beta)^T \), where null \( (\cdot) \) is the null space of a given matrix. Since \( |\mathcal{I}| = \text{rank}(U) + 1 = \text{rank} (U_\mathcal{I}) + 1 = \text{rank} (U_\mathcal{I}^T) + 1 \), null \( (U_\mathcal{I}^T) \) is of dimension 1. That implies \( \forall w \in \text{null}(U_\mathcal{I}^T) \), we have \( w = k(r - s) \) for some \( k \in \mathbb{R} \). Therefore,

\[
(V_\mathcal{I} \beta)^T w = k(r - s)^T V_\mathcal{I} \beta = k \left( \frac{(r - s)^T V_\mathcal{I} (z' - z^\lor)}{\Delta_1} - \frac{(r - s)^T V_\mathcal{I} (z' - z^\land)}{\Delta_2} \right) = k(1 - 1) = 0.
\]

That is, \( w \in \text{null} ((V_\mathcal{I} \beta)^T) \). Hence, null \( (U_\mathcal{I}^T) \subseteq \text{null} ((V_\mathcal{I} \beta)^T) \) and then we have \( V_\mathcal{I} \beta \in \text{span}(U_\mathcal{I}) \).
The third observation is that there exists some \( i \in [n] \) such that \((V_I)_{\beta_i} \notin \text{span}(U_I)\). To show this, we denote \( \mathcal{H} = \{ i \in [n] \mid z'_i \leq z''_i \} \). We then have for every \( i \in \mathcal{H} \), \( z'^i = z'_i, z''_i \) and hence \( \beta_i = \frac{z'_i - z''_i}{\Delta_1} \). In addition, since for every \( i \in [n] \setminus \mathcal{H} \), \( z'_i > z''_i \), \( \frac{z'_i - z''_i}{\Delta_1} = \frac{z'_i - z'_i}{\Delta_1} = 0 \), we have

\[
(r - s)^T \sum_{i \in \mathcal{H}} \beta_i (V_I)_i = (r - s)^T \left( \sum_{i \in \mathcal{H}} \beta_i (V_I)_i + \sum_{i \in [n] \setminus \mathcal{H}} 0 \cdot (V_I)_i \right) = (r - s)^T V_I \frac{z'_i - z''_i}{\Delta_1} = 1.
\]

Hence, \((r - s) \notin \text{null} \left( \left( \sum_{i \in \mathcal{H}} \beta_i (V_I)_i \right)^T \right)\), which implies that \( \text{null} (U_I^T) \) is not a subset of \( \text{null} \left( \left( \sum_{i \in \mathcal{H}} \beta_i (V_I)_i \right)^T \right) \). Consequently we have \( \sum_{i \in \mathcal{H}} \beta_i (V_I)_i \notin \text{span}(U_I) \), implying that there exists some \( i \in \mathcal{H} \) such that \((V_I)_i, \beta_i \notin \text{span}(U_I)\).

With the three observations, we have a contradiction of the second condition in Theorem 3.

We next prove the “only if” direction by contradiction. Assume the condition in the theorem is not satisfied. That is, \( \text{rank}(U) < r \) and there exist some \( I \subseteq [r] \), \( \beta \in \mathbb{R}_+^r \) satisfying \( |I| = \text{rank}(U) + 1 \), \( \text{rank}(U_I) = \text{rank}(U) \) and \( V_I \beta \notin \text{span}(U_I) \), such that \( \beta_i (V_I)_i \notin \text{span}(U_I) \) for some \( i \in [n] \). Note that in this case, we can find a vector \( \alpha \in \mathbb{R}^m \) such that \( U_I \alpha = V_I \beta \).

We arbitrarily choose \( z^\wedge \in \mathbb{R}^m, p \in \mathbb{R}^m \) and let \( z^\vee = z^\wedge + \beta \geq z^\wedge, q = p - \alpha \), then \( U_I p - V_I z^\wedge = U_I (q + \alpha) - V_I (z^\vee - \beta) = U_I q - V_I z^\vee \). We also arbitrarily choose \( x \), and then choose \( v^0 \) such that \( c = -W x + v^0 \) is with \( c_I = U_I p - V_I z^\wedge \) and \( c_j \) being large enough for all \( j \notin I \). Then we have \((p, z^\wedge), (q, z^\vee) \in S(x) \). We further define \( z' = z^\wedge + \beta_i e_i, z'' = z^\vee - \beta_i e_i \) so that \( z'^\wedge \wedge z'' = z^\wedge, z'^\vee \vee z'' = z^\vee \). Then we have

\[
c_I + V_I z' = c_I + V_I (z^\wedge + \beta_i e_i) = U_I p - V_I z^\wedge + V_I (z^\wedge + \beta_i e_i) = U_I p + \beta_i (V_I)_i \notin \text{span}(U_I),
\]

where the last relationship holds since \( U_I p \in \text{span}(U_I) \) but \( \beta_i (V_I)_i \notin \text{span}(U_I) \).

Hence, \( \{ y \in \mathbb{R}^m \mid U_I y = c_I + V_I z' \} = \emptyset \), i.e. for any \( y' \) satisfying \( U_I y' - V_I z' \leq c \), there exists \( j \in I \) such that \( u_j^T y' - v_j^T z' < c_j \). If there exists some \( y'' \) with \( U_I y'' - V_I z'' \leq c \) satisfies \( y' + y'' = p + q \),

\[
U_I y'' - V_I z'' = U_I (p + q - y') - V_I (z^\wedge + z^\vee - z')
\]

\[
= U_I p - V_I z^\wedge + U_I q - V_I z^\vee - U_I y' - V_I z'
\]

\[
= 2c_I - (U_I y' - V_I z'),
\]

then we should have \( 2c_j - (u_j^T y' - v_j^T z') > c_j \) for the above mentioned \( j \), which contradicts the assumption \((y'', z'') \in S(x) \). Hence we prove the necessity of the conditions on \( U, V \).

For any given matrices \( U \in \mathbb{R}^{r \times m}, V \in \mathbb{R}^{r \times n} \), we provide the following algorithm to check explicitly whether the condition in Theorem 3 is met.
Algorithm 2 checking algorithm

1: Input: $U \in \mathbb{R}^{r \times m}, V \in \mathbb{R}^{r \times n}$

2: Initialization: $r_0 = \text{rank}(U), s = 1$

3: if $r_0 < r$ then

4: arbitrarily remove columns in $U$, if any, until $U$ has only $r_0$ linearly independent columns

5: for all $I \subseteq [r]$ with $|I| = r_0$ and $U_I$ invertible, do

6: for $i \in [r] \setminus I$ do

7: $d_i^T = v_i^T - u_i^T U_I^{-1} V_I$

8: if there exist components $d_{ia}, d_{ib}$ such that $d_{ia}d_{ib} < 0$ then

9: $s = 0$, go to line 10

10: return $s$

Theorem 4 The condition in Theorem 3 is satisfied if and only if Algorithm 2 returns $s = 1$.

Proof: The case for $	ext{rank}(U) = r$ is straightforward, so we only consider the case where $	ext{rank}(U) < r$. In that case, we only need to verify whether $U, V$ satisfy the second part of the condition in Theorem 3, which depends solely on the relationship between $V$ and span$(U)$. Thus, removing the dependent columns in $U$ does not change the satisfaction or violation of the conditions. Therefore, the procedure in line 4 of the algorithm does not change the result and WLOG, we can assume $U$ has $r_0$ columns, i.e., $m = r_0$.

First we look into the case where Algorithm 2 returns $s = 0$. This implies that there exists an index set $I \subseteq [r]$ and indices $i \in [r] \setminus I, a, b \in [r_0]$ with $|I| = r_0$, $U_I$ invertible and $d_{ia}d_{ib} < 0$. WLOG, we let $d_{ia} > 0, d_{ib} < 0$.

Denote $\beta = \frac{e_a}{d_{ia}} - \frac{e_b}{d_{ib}} \geq 0, \alpha = U_I^{-1} \left( \frac{(V_I)_a}{d_{ia}} - \frac{(V_I)_b}{d_{ib}} \right)$, then

$$
\begin{bmatrix}
V_I \\
v_i^T
\end{bmatrix} \beta = \begin{bmatrix}
(V_I)_a \\
v_i^T
\end{bmatrix} \begin{bmatrix}
d_{ia} & d_{ib} \\
v_i^T - \frac{(V_I)_b}{d_{ib}} & \frac{(V_I)_a}{d_{ia}}
\end{bmatrix}
\begin{bmatrix}
(V_I)_a \\
v_i^T
\end{bmatrix} \begin{bmatrix}
\frac{(V_I)_a}{d_{ia}} - \frac{(V_I)_b}{d_{ib}} \\
\frac{(V_I)_a}{d_{ia}} - \frac{(V_I)_b}{d_{ib}}
\end{bmatrix}
= U_I^T \begin{bmatrix}
\frac{(V_I)_a}{d_{ia}} - \frac{(V_I)_b}{d_{ib}} \\
\frac{(V_I)_a}{d_{ia}} - \frac{(V_I)_b}{d_{ib}}
\end{bmatrix}
= U_I^T \alpha,
$$

We let $\tilde{I} = I \cup \{i\}$. The above equality implies $V_{\tilde{I}} \beta = U_{\tilde{I}} \alpha \in \text{span}(U_{\tilde{I}})$. On the other hand, for $\beta_a(V_{\tilde{I}})_a$ we have

$$
\beta_a \begin{bmatrix}
(V_{\tilde{I}})_a \\
v_{ia}
\end{bmatrix} = \begin{bmatrix}
\frac{(V_{\tilde{I}})_a}{d_{ia}} \\
\frac{(V_{\tilde{I}})_a}{d_{ia}} + \frac{u_i^T U_{\tilde{I}}^{-1} (V_{\tilde{I}})_a}{d_{ia}}
\end{bmatrix}
= U_{\tilde{I}} \begin{bmatrix}
U_{\tilde{I}}^{-1}(V_{\tilde{I}})_a \\
u_i^T
\end{bmatrix}
\begin{bmatrix}
0 \\
1
\end{bmatrix}
= U_{\tilde{I}} \frac{(V_{\tilde{I}})_a}{d_{ia}} + \begin{bmatrix}
0 \\
1
\end{bmatrix}.
$$

Since $U_{\tilde{I}}$ is invertible, there is no $\gamma \in \mathbb{R}^{r_0}$ such that $U_{\tilde{I}} \gamma = \begin{bmatrix}
U_{\tilde{I}} \\
u_i^T
\end{bmatrix} \gamma = \begin{bmatrix}
0 \\
1
\end{bmatrix}$. Hence $\beta_a(V_{\tilde{I}})_a \notin \text{span}(U_{\tilde{I}})$ and the second part of the condition in Theorem 3 is violated.
We now investigate the case where the second part of the condition in Theorem 3 is violated. That means, there exist \( \tilde{I} \subseteq [r] \), \( \beta \geq 0 \) and \( a \in [r_0] \) such that \( |\tilde{I}| = r_0 + 1 \), rank(\( U_\tilde{I} \)) = \( r_0 \), \( V_\tilde{I} \beta \in \text{span}(U_\tilde{I}) \) but \( \beta_a(V_\tilde{I})_a \notin \text{span}(U_\tilde{I}) \). We choose \( I \subseteq \tilde{I} \) such that \( |I| = r_0 \) and \( U_I^{-1} \) invertible, and denote \( \tilde{i} \) as the unique index in \( \tilde{I} \backslash I \). Then we have the following equations,

\[
V_\tilde{I} \beta = \begin{bmatrix} V_I & 0 \\ 0 & V_I \end{bmatrix} \begin{bmatrix} \beta \\ -u_i^T U_I^{-1} V_I \end{bmatrix} + \begin{bmatrix} U_I \\ u_i^T \end{bmatrix} U_I^{-1} V_\tilde{I} \beta = \begin{bmatrix} 0 \\ d_i^T \end{bmatrix} \beta + U_I U_I^{-1} V_\tilde{I} \beta,
\]

\[
\beta_a(V_\tilde{I})_a = \beta_a \left( \begin{bmatrix} (V_I)_a \\ v_i \\ u_i \end{bmatrix} \right) = \beta_a \left( \begin{bmatrix} 0 \\ v_i - u_i^T U_I^{-1} (V_I)_a \end{bmatrix} + \begin{bmatrix} U_I \\ u_i^T \end{bmatrix} U_I^{-1} (V_I)_a \right) = \begin{bmatrix} 0 \\ \beta_a d_i \end{bmatrix} + \beta_a U_I U_I^{-1} (V_I)_a.
\]

Since \( V_\tilde{I} \beta, U_I U_I^{-1} V_\tilde{I} \beta, \beta_a U_I U_I^{-1} (V_I)_a \in \text{span}(U_\tilde{I}) \) and \( \beta_a(V_\tilde{I})_a \notin \text{span}(U_\tilde{I}) \), the above equations imply \( \begin{bmatrix} 0 \\ d_i^T \end{bmatrix} \beta \in \text{span}(U_\tilde{I}) \) and \( \beta_a d_i \notin \text{span}(U_\tilde{I}) \). According to \( \begin{bmatrix} 0 \\ d_i^T \end{bmatrix} \beta \in \text{span}(U_\tilde{I}) \), there exists \( \alpha \in \mathbb{R}^{r_0} \) with \( U_\tilde{I} \alpha = 0, u_i^T \alpha = d_i^T \beta \). Since \( U_I \) is invertible, \( \alpha = 0 \) and hence \( d_i^T \beta = u_i^T \alpha = 0 \).

According to \( \beta_a d_i \notin \text{span}(U_\tilde{I}) \), we obtain \( \beta_a d_i \neq 0 \). As \( \beta \geq 0 \), \( \beta_a d_i \neq 0 \) and \( d_i^T \beta = 0 \), we must have an index \( b \in [r_0] \) such that \( d_i, d_b \) are of different signs. Hence the algorithm returns \( s = 0 \).

By now, given any two-stage optimization problem (2), we can use Algorithm 2 to verify whether the second-stage cost function is supermodular in \( z \). If the answer is positive, we can use the result in Theorem 1 and reformulate it to be equivalent to Problem (9), allowing us to derive the optimal solution efficiently.

4. Applications

In this section we apply the above theoretical results to three classic operational problems, which are hard to solve in general. The first application is a single-period multi-item newsvendor problem, where the objective is to maximize the agent’s utility. The second is an appointment scheduling problem where we minimize the expected sum of waiting time and overtime. The third is ATO systems, and we identify a class of systems which are tractable using our method. In this section, some common notations may be used in different applications with different meanings.

4.1. Multi-item newsvendor

Multi-item newsvendor problems seek the optimal inventory levels of multiple goods with fixed prices and uncertain demands (Hadley and Whitin 1963). Since these items correlate with each other either through some budget constraint or by a particular utility function, the problem may become much harder to solve. In the distributionally robust setting, Hanasusanto et al. (2015) assume a risk-averse decision maker who minimizes a linear combination of conditional value-at-risk and expectation of profit function. The demand distribution is multi-modal. They show that the resulting problem is NP-hard and by using the quadratic decision rule, they solve it approximately.
with a semidefinite program. Natarajan et al. (2017) use semi-variance to capture the asymmetry of demand distributions. They also develop a semidefinite program to derive the lower bound for the original problem. We next use the results in the sections above to show that the multi-item newsvendor problem can be solved efficiently within our setting.

Consider a single-period multi-item newsvendor problem with \( n \) different items. The selling price and ordering cost of item \( i \) are denoted by \( r_i \) and \( c_i \), respectively. Before the random demand \( \tilde{z} \) is resolved, we need to decide the ordering quantity \( x \), which is subject to a budget \( \Gamma \). The goal is to maximize the worst-case expected utility of revenue, where the utility function \( \hat{u} \) is concave increasing. We hence have the following optimization problem

\[
\max \inf \mathbb{E}_x \left[ \hat{u} \left( \sum_{i \in [n]} r_i \min\{x_i, \tilde{z}_i\} \right) \right],
\]

where we denote \( \mathcal{X}_{\text{news}} = \{x \mid c^T x \leq \Gamma, x \geq 0\} \). To use the theoretical results in Section 2, we reformulate it as an equivalent minimization problem

\[
\min \sup \mathbb{E}_x \left[ u \left( -r^T x + \sum_{i \in [n]} r_i (x_i - \tilde{z}_i)^+ \right) \right],
\]

where \( u(w) = -\hat{u}(-w) \) for all \( w \in \mathbb{R} \) and can be considered as the disutility function. We now establish the supermodularity of the objective function in \( z \).

**Lemma 4** Given any convex and non-decreasing function \( \varphi : \mathbb{R} \to \mathbb{R} \) and monotone supermodular function \( f : \mathbb{R}^n \to \mathbb{R} \), the function \( \phi \) defined as \( \phi(z) = \varphi(f(z)) \forall z \in \mathbb{R}^n \) is supermodular.

**Proof:** For any \( z', z'' \in \mathbb{R}^n \), we denote \( a = f(z' \wedge z''), b = f(z'), c = f(z''), d = f(z' \vee z'') \) and \( d_0 = b + c - a \). From the supermodularity of \( f \) we know \( b + c \leq a + d \); hence, \( d_0 \leq d \). We then have

\[
\phi(z') + \phi(z'') = \varphi(b) + \varphi(c) \leq \varphi(a) + \varphi(d_0) \leq \varphi(a) + \varphi(d) = \phi(z' \wedge z'') + \phi(z' \vee z''),
\]

where the second inequality arises because \( \varphi \) is non-decreasing. To demonstrate the first inequality, we notice that either \( a \leq \min\{b, c\} \leq \max\{b, c\} \leq d_0 \) (if \( f \) is increasing) or \( a \geq \max\{b, c\} \geq \min\{b, c\} \geq d_0 \) (if \( f \) is decreasing) holds; since \( a + d_0 = b + c \) and \( \varphi \) is convex, we then have the first inequality in Equation (15). This proves the supermodularity of \( \phi \). \( \square \)

**Corollary 1** The function \( u \left( -r^T x + \sum_{i \in [n]} r_i (x_i - \tilde{z}_i)^+ \right) \) is supermodular in \( z \) for any given \( x \).

**Proof:** Obviously, function \( -r^T x + \sum_{i \in [n]} r_i (x_i - \tilde{z}_i)^+ \) is decreasing and supermodular in \( z \). As \( u \) is convex increasing, by Lemma 4 we know that \( u \left( -r^T x + \sum_{i \in [n]} r_i (x_i - \tilde{z}_i)^+ \right) \) is supermodular in \( z \). \( \square \)
As Corollary 1 has established the supermodularity of the objective function in Problem (14), we can now reformulate Problem (14) into a tractable one using Theorem 1. Recall that $z^{k,i}, p^k_i$ for $i \in [2n+1]$ are output from Algorithm 1 with input $\mathcal{F}^k$, $k \in [K]$, and characterize the worst-case distribution. Hence, Problem (14) has an equivalent reformulation as follows,

$$\min \ \nu^T l$$

s.t. $R^k_i l \geq \sum_{i=1}^{2n+1} p^k_i u \left( -r^T x + \sum_{i \in [n]} r_i (x_i - z^{k,i}_i)^+ \right), \quad k \in [K]$\hspace{1cm} (16)

$$l \geq 0,$$

$$x \in \mathcal{X}_{\text{news}}.$$

Specifically, when the disutility $u$ is a piecewise linear function, since $p^k_i \geq 0$ for all $i \in [2n+1], k \in [K]$, Problem (14) can be reformulated as a linear optimization problem.

When minimizing the Conditional Value-at-Risk (CVaR), as Hanasusanto et al. (2015) do, the problem can be represented as

$$\min_{x \in \mathcal{X}_{\text{news}}} \sup_{\nu \in \mathcal{F}} \text{CVaR}_\rho \left( -r^T x + \sum_{i \in [n]} r_i (x_i - z_i)^+ \right). \hspace{1cm} (17)$$

We next show that this problem under our distributional ambiguity set can be solved as a linear optimization problem.

**Corollary 2** Given the feasible set $\mathcal{X}_{\text{news}}$ and the distributional ambiguity set $\mathcal{F}$ denoted by the formulation (3), Problem (17) is equivalent to the following linear optimization problem

$$\min \ \theta + \frac{1}{\rho} \nu^T l$$

s.t. $R^k_i l \geq \sum_{s \in [i]} p_s^k (r^T (y^{k,s} - x) - \theta), \quad k \in [K], i \in [2n+1]$\hspace{1cm} (18)

$$y^{k,i} \geq x - z^{k,i}, \quad k \in [K], i \in [2n+1]$$

$$y^{k,i} \geq 0, \quad k \in [K], i \in [2n+1]$$

$$R^k_i l \geq 0, \quad k \in [K]$$

$$l \geq 0,$$

$$x \in \mathcal{X}_{\text{news}}.$$

$p^k_i, z^{k,i} i \in [2n+1]$ in the formulation are obtained as the output of Algorithm 1 given the ambiguity sets $\mathcal{F}^k, k \in [K]$ defined by (4).

**Proof:** The reformulation can be implemented as follows,

$$\min_{x \in \mathcal{X}_{\text{news}}} \sup_{\nu \in \mathcal{F}} \text{CVaR}_\rho \left( -r^T x + \sum_{i \in [n]} r_i (x_i - z_i)^+ \right)$$
\[
\begin{align*}
&= \min_{x \in \mathcal{X}_{\text{new}}} \theta + \frac{1}{\rho} \sup_{\mathbb{P} \in \mathcal{F}} \left[ -r^T x + r^T (x - z)^+ - \theta \right]^+ \\
&= \min \theta + \frac{1}{\rho} \nu^T l \\
&\quad \text{s.t. } R_i l \geq \sum_{i \in [2n+1]} p_i^k \left( -r^T x + r^T (x - z^{k,i})^+ - \theta \right)^+, \quad k \in [K] \\
&\quad l \geq 0, \quad x \in \mathcal{X}_{\text{new}}.
\end{align*}
\]

The first equality follows from the definition of CVaR. For any fixed \( x, r, \theta \), function \(-r^T x + r^T (x - z)^+ - \theta \) is non-increasing in \( z \). Given any \( k \in [K] \), since \( z^{k,i} \leq z^{k,i+1} \) for all \( i \in [2n] \) based on Proposition 2, we then have

\[-r^T x + r^T (x - z^{k,i})^+ - \theta \geq -r^T x + r^T (x - z^{k,i+1})^+ - \theta, \quad \forall i \in [2n].\]

Therefore,

\[
\sum_{i \in [2n+1]} p_i^k \left( -r^T x + r^T (x - z^{k,i})^+ - \theta \right)^+ = \max \left\{ 0, \max_{i \in [2n+1]} \left\{ \sum_{s \in [i]} p_s^k \left( -r^T x + r^T (x - z^{k,s})^+ - \theta \right) \right\} \right\}.
\]

By introducing auxiliary variables \( y^{k,s} \) such that \( y^{k,s} \geq x - z^{k,s} \) and \( y^{k,s} \geq 0 \), we complete the proof. \qed

Given the objective of minimizing CVaR and the multi-modal demand assumption, we notice that our work differs from Hanasusanto et al. (2015) in the scenario-based distributional information. While their work considers the first two moments and derive an approximate solution by solving a semidefinite programming problem, here we focus on partial marginal information and obtain an exact linear programming reformulation of the original problem.

### 4.2. Appointment scheduling problems

The appointment scheduling problem, which schedules the arrival times of customers, has wide applications in service delivery systems (Cayirli and Veral 2003, Gupta and Denton 2008). Due to uncertain service times, distributionally robust optimization becomes one of the main approaches (Kong et al. 2013, Mak et al. 2014, Qi 2016). We next show that in our setting, the problem can be reduced to a linear optimization problem, which is a simplification compared with all previous studies.

Consider an operating theatre which needs to schedule \( n \) surgeries within a given time period \( L \). For each surgery \( i, i \in [n] \), its actual duration is a random variable denoted by \( \tilde{z}_i \). We need to decide the scheduled duration \( x_i \) for each surgery \( i \) at the beginning of the planning horizon. The
waiting time of the $i$-th patient is denoted by $w_i$, $i \in [n]$. In addition, we denote the overtime of the operating theatre by $w_{n+1}$. We then have

$$w_1 = 0,$$
$$w_2 = \max\{\tilde{z}_1 - x_1, 0\},$$
$$w_3 = \max\{w_2 + \tilde{z}_2 - x_2, 0\} = \max\{\tilde{z}_1 - x_1 + \tilde{z}_2 - x_2, 0\},$$
$$\vdots$$
$$w_{n+1} = \max\{w_n + \tilde{z}_n - x_n, 0\} = \max\left\{\sum_{s \in [n]} (\tilde{z}_s - x_s), \sum_{s=2}^{n} (\tilde{z}_s - x_s), \ldots, \tilde{z}_n - x_n, 0\right\}.

The goal is to find the best scheduling decision, i.e., $x$, such that we can minimize the worst-case expectation of the sum of all patients’ waiting time and the operating theatre’s overtime. We formulate it with a two-stage optimization structure,

$$\min \sup_{x \in \mathcal{X}_{\text{supp}}} \mathbb{E}_p \left[ \sum_{t \in [n]} \max \left\{ \max_{j \in t} \sum_{s = j}^t (\tilde{z}_s - x_s), 0 \right\} \right] = \min \sup_{x \in \mathcal{X}_{\text{supp}}} \mathbb{E}_p [g(x, \tilde{z})],$$

where the feasible set is defined as $\mathcal{X}_{\text{supp}} = \{x \mid 1^T x \leq L, x \geq 0\}$, and the second-stage problem can be written as

$$g(x, z) = \min_{y_1, \ldots, y_n} \sum_{t \in [n]} y_t$$
$$\text{s.t. } y_t \geq \sum_{s = j}^t (z_s - x_s), j \in [t], t \in [n]$$
$$y_t \geq 0, \quad t \in [n].$$

We can use the condition in Theorem 3 to verify the supermodularity of the function $g(x, z)$.

**Theorem 5** The function $g(x, z)$ defined in Problem (19) is supermodular in $z$ for any given $x$. Hence Problem (18) is equivalent to the following linear optimization problem

$$\min \nu^T l$$
$$\text{s.t. } R_k^T l \geq \sum_{i \in [2n+1]} p_i^k 1^T y^{k,i}, \quad k \in [K]$$
$$y^{k,i} \geq 0, \quad k \in [K], i \in [2n+1]$$

$$y^{k,i} \geq 0, \quad k \in [K], i \in [2n+1]$$

$$l \geq 0, \quad x \in \mathcal{X}_{\text{supp}},$$

where $p_i^k, z^{k,i}, k \in [K], i \in [2n+1]$ are the output of Algorithm 1 given the ambiguity sets $\mathcal{F}^k, k \in [K]$ defined by (4).

**Proof:** See Appendix.
Therefore, given the information set $\mathcal{F}$ as in Equation (3), we can reformulate Problem (18) in a computationally tractable way using Theorem 1. Our linear programming reformulation provides an exact solution while preserving the original structure of the problem. Moreover, the computational complexity can be reduced significantly compared to the semidefinite programming reformulations in the literature.

4.3. Assemble-to-order systems

The ATO system is an important operational problem. We refer interested readers to Song and Zipkin (2003) and Atan et al. (2017) for a comprehensive review. Although this problem has attracted substantial attention, it is still not clear how to derive the optimal decision in general. We now apply our theoretical result to the ATO problem and identify a class of systems where the optimal decision can be obtained efficiently.

Here we formally describe the problem using the formulation of Song and Zipkin (2003). Specifically, we first need to decide $x_i$, which is the order-up-to inventory level of the component $i, i \in [I]$. After that, the uncertain demand $\tilde{z}_j$ for end product $j$ is realized and we can observe its realization, $z_j, j \in [N]$. We then need to decide the second-stage decision $y_j$, which is the quantity of product $j$ to be assembled, $j \in [N]$. With the goal of minimizing the worst-case expected cost, we have the following formulation,

$$
\begin{align*}
\min & \quad c^T (x - x_{int}) + \sup_{\mathcal{F}} \mathbb{E}_p \left[ g(x, \tilde{z}) \right] \\
\text{s.t.} & \quad x \geq x_{int},
\end{align*}
$$

(20)

where

$$
\begin{align*}
g(x, z) &= \min \quad h^T (x - Ay) + p^T (z - y) - r^T y \\
\text{s.t.} & \quad Ay \leq x, \\
& \quad y \leq z, \\
& \quad y \geq 0.
\end{align*}
$$

(21)

Here $c$ and $x_{int}$ are the per-unit ordering cost and initial inventory level of the components, respectively; $g$ is the function which maps the order-up-to inventory level of components and demand realization of end products to the minimal second-stage cost; $h$ is the per-unit inventory holding cost of the leftover components; $p$ and $r$ are the per-unit penalty cost of the shortage and per-unit selling price of the end products, respectively. The elements in matrix $A$, i.e., $a_{ij} \geq 0$, represent the number of units of component $i$ required to assemble one unit of end product $j$. Different ATO systems are characterized by different matrices $A \in \mathbb{R}^{I \times N}$. We provide the equivalent conditions on $A$ such that the function $g$ is supermodular.

With our theoretical results, we identify a class of ATO systems which can lead to the supermodularity of the function $g$ and hence we can solve the optimal order-up-to level for components efficiently. The first result is for the supermodularity on any pair of components.
Lemma 5 Consider any pair of distinct indices $i, j \in [n]$. The function $g(x, z)$ is supermodular in $(z_i, z_j)$ for any given $z_k, k \in [n] \setminus \{i, j\}, x, h, p, r$ if the following condition is true. For any $s \in \{1, 2, \ldots, \min\{l-1, n-2\}\}$, $\beta_1, \beta_2 > 0$, row index set $I_0$ and column index set $J_0$ such that 1) $|I_0| = s + 1$ and $|J_0| = s$, 2) $i, j \notin J_0$, 3) the submatrix $A_{I_0, J_0}$ is full column rank, 4) $\beta_1(A_{I_0})_i + \beta_2(A_{I_0})_j \in \text{span}(A_{I_0, J_0})$, we have $(A_{I_0})_i, (A_{I_0})_j \in \text{span}(A_{I_0, J_0})$. Here $A_{I_0, J_0}$ is the submatrix of $A$ obtained by removing all rows with index not in $I_0$ and columns with index not in $J_0$.

Proof: See Appendix. □

We are now ready to show the condition on $A$ such that the function $g(x, z)$ is supermodular in the whole vector $z$.

Theorem 6 The function $g(x, z)$ is supermodular in $z$ for any $x, h, p, r$ if and only if every $2 \times 3$ submatrix of the matrix $A$ contains at least two column vectors which are linearly dependent.

Proof: We first prove the “if” direction. Suppose the condition in this theorem is satisfied. We prove the supermodularity of $g(x, z)$ in $z$ by showing that the condition in Lemma 5 is satisfied for all distinct pairs of $i, j \in [n]$, hence $g(x, z)$ is supermodular in every distinct pair $(z_i, z_j)$. To this end, consider any $s \in \{1, 2, \ldots, \min\{l-1, n-2\}\}$, $\beta_1, \beta_2 > 0$, row index set $I_0$ and column index set $J_0$ such that 1) $|I_0| = s + 1$ and $|J_0| = s$, 2) $i, j \notin J_0$, 3) the submatrix $A_{I_0, J_0}$ defined as in Lemma 5 is full column rank, 4) $\beta_1(A_{I_0})_i + \beta_2(A_{I_0})_j \in \text{span}(A_{I_0, J_0})$. We need to prove that $(A_{I_0})_i, (A_{I_0})_j \in \text{span}(A_{I_0, J_0})$. We show this by contradiction.

Assuming the contrary, i.e., at least one of $(A_{I_0})_i$, and $(A_{I_0})_j$ is not in span($A_{I_0, J_0}$). WLOG, we let $(A_{I_0})_i \notin \text{span}(A_{I_0, J_0})$. Consider a submatrix $Q \in \mathbb{R}^{(s+1) \times (s+2)}$ defined by $Q = [(A_{I_0})_i \ (A_{I_0})_j \ A_{I_0, J_0}]$. Since $(A_{I_0})_i \notin \text{span}(A_{I_0, J_0})$, we have rank($Q$) = $s + 1$. By Lemma 7, at least two columns in $Q$ are linearly dependent. Since $A_{I_0, J_0}$ is full column rank, there cannot be two columns in $A_{I_0, J_0}$ which are linearly dependent. In addition, since $(A_{I_0})_i \notin \text{span}(A_{I_0, J_0})$, it is impossible that $(A_{I_0})_i$ and one column in $A_{I_0, J_0}$ are linearly dependent. We note that $(A_{I_0})_j$ and $(A_{I_0})_i$ are linearly independent, otherwise $\beta_1(A_{I_0})_i + \beta_2(A_{I_0})_j \in \text{span}(A_{I_0, J_0})$ contradicts $(A_{I_0})_i \notin \text{span}(A_{I_0, J_0})$ and $\beta_1, \beta_2 > 0$. Hence, $(A_{I_0})_j$ and one column in $A_{I_0, J_0}$ are linearly dependent. In this case, $(A_{I_0})_i \notin \text{span}(A_{I_0, J_0})$ contradicts $\beta_1(A_{I_0})_i + \beta_2(A_{I_0})_j \in \text{span}(A_{I_0, J_0})$ and $\beta_1 > 0$. Therefore, $(A_{I_0})_i, (A_{I_0})_j \in \text{span}(A_{I_0, J_0})$.

We now prove the “only if” direction by contradiction. We first consider the case that $A \in \mathbb{R}_+^{2 \times 3}$. Assume the contrary, i.e., every two columns in $A$ are linearly independent. Then we must have one column in $A$ being a conical combination of the other two columns. WLOG, let $A_3$ be a conical combination of $A_1, A_2$. We remark that multiplying any strictly positive constant by a row/column in $A$, or switching rows, or switching columns does not affect whether the function $g$
is supermodular. Therefore, we can make the following simplification on \( A \). Since \( A_1, A_2 \) are linearly independent, WLOG, we can let \( A = \begin{bmatrix} 1 & a & c \\ b & 1 & d \end{bmatrix} \) with \( ab < 1 \). Since \( A_3 \) is a conical combination of \( A_1, A_2 \), we have \( cd > 0 \); WLOG, we can let \( d = 1 \), i.e., \( A = \begin{bmatrix} 1 & a & c \\ b & 1 & 1 \end{bmatrix} \). Multiplying the first row by 1/c, and then multiplying the first column by c, we have \( A = \begin{bmatrix} 1 & a/c & 1 \\ bc & 1 & 1 \end{bmatrix} \). Let \( a/c, bc \) be the new \( a, b \), we have \( A = \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix} \) with \( ab < 1 \). Again, since \( A_3 \) is a conical combination of \( A_1, A_2 \), we have either \( a, b < 1 \) or \( a, b > 1 \). Together with \( ab < 1 \), we know \( a, b < 1 \). In summary, WLOG, we let \( A = \begin{bmatrix} 1 \\ b \end{bmatrix} \) with \( a, b \in [0, 1) \).

We define \( \bar{g}(x, z) = g(x, z) - p^T z \), then it is equivalent to prove that \( \bar{g}(x, z) \) is not supermodular in \( z \). We now construct such a counterexample. Let \( h = 0, r = 0, p = (1, 1, \epsilon) \) with any \( \epsilon \in (0, 1) \). We choose \( x = (1 - ab)1, z' = (1 - a, 0, 1 - ab), z'' = (0, 1 - b, 1 - ab) \). Denote \( z^\wedge = z' \wedge z'', z^\vee = z' \vee z'' \), we have \( z^\wedge = (0, 0, 1 - ab), z^\vee = (1 - a, 1 - b, 1 - ab) \). We notice that

\[
\bar{g}(x, z) = \min \left\{ -p^T y \mid Ay \leq x, 0 \leq y \leq z \right\}
\]

\[
= \min \begin{cases}
-y_1 - y_2 - \epsilon y_3 \\
\text{s.t. } y_1 + ay_2 + y_3 \leq 1 - ab \\
by_1 + y_2 + y_3 \leq 1 - ab \\
(0, 0, 0) \leq (y_1, y_2, y_3) \leq (z_1, z_2, z_3)
\end{cases}
\]  

(22)

Hence,

\[
\bar{g}(x, z') = \min \begin{cases}
-y_1 - \epsilon y_3 \\
\text{s.t. } y_1 + y_3 \leq 1 - ab, \\
y_1 \leq 1 - a, \ y_2 = 0, \ y_3 \geq 0
\end{cases},
\]

\[
\bar{g}(x, z'') = \min \begin{cases}
-y_2 - \epsilon y_3 \\
\text{s.t. } y_2 + y_3 \leq 1 - ab, \\
y_1 = 0, \ 0 \leq y_2 \leq 1 - b, \ y_3 \geq 0
\end{cases},
\]

\[
\bar{g}(x, z^\wedge) = \min \begin{cases}
-\epsilon y_3 \\
\text{s.t. } y_3 \leq 1 - ab, \\
y_1 = y_2 = 0, \ y_3 \geq 0
\end{cases},
\]

\[
\bar{g}(x, z^\vee) = \min \begin{cases}
-y_1 - y_2 - \epsilon y_3 \\
\text{s.t. } y_1 + ay_2 + y_3 \leq 1 - ab, \\
y_1 \geq 0, \ y_2 \geq 0, \ 0 \leq y_1 \leq 1 - a, \ 0 \leq y_2 \leq 1 - b, \ y_3 \geq 0
\end{cases}.
\]

Since \( 0 < \epsilon < 1 \), in the optimization problem for \( \bar{g}(x, z') \), the optimal solution should be that \( y_1 \) goes to the upper bound, i.e. \( y_1 = 1 - a, y_2 = 0 \) and \( y_3 = (1 - ab) - (1 - a) = a(1 - b) \). Similarly, in the optimization problem for \( \bar{g}(x, z'') \), the optimal \( y = (0, 1 - b, b(1 - a)) \); in that for \( \bar{g}(x, z^\wedge) \), the optimal \( y = (0, 0, 1 - ab) \); in that for \( \bar{g}(x, z^\vee) \), the optimal \( y = (1 - a, 1 - b, 0) \). We then have

\[
\bar{g}(x, z') + \bar{g}(x, z'') - \bar{g}(x, z^\wedge) - \bar{g}(x, z^\vee)
\]

\[
= -(1 - a + \epsilon a(1 - b)) + (1 - b + \epsilon b(1 - a)) - \epsilon(1 - ab) - (1 - a + 1 - b)
\]

\[
= \epsilon(1 - a - b + ab)
\]

\[
= \epsilon(1 - a)(1 - b) > 0,
\]
where the last equality holds since $0 < a, b < 1$. Therefore, $ar{g}(x, z^\vee) + ar{g}(x, z') < ar{g}(x, z') + ar{g}(x, z'')$, this function $\bar{g}$ is not supermodular.

For the general case of $A \in \mathbb{R}^{n \times n}_+$, we can prove the result by the same contradiction. WLOG, we assume the $2 \times 3$ submatrix of $A$, which is obtained by deleting all rows except the first two and all columns except the first three, is such that each two columns in it are linearly independent. We can then let $z_i' = z_i'' = 0$ for $i \in \{4, 5, \ldots, n\}$ and $x$, be sufficiently large for $i \in \{3, 4, \ldots, l\}$ such that it would not affect the feasible region of $y$. We then have $\bar{g}(x, z)$ with exactly the same expression in Equation (22). Therefore, we still have $\bar{g}(x, z^\vee) + \bar{g}(x, z') < \bar{g}(x, z') + \bar{g}(x, z'')$.

Given any ATO system characterized by a matrix $A$, according to Theorem 6, we can verify whether the function $g(x, z)$ is supermodular in $z$ or not. If the answer is positive, we then adopt Theorem 1 to obtain the optimal order-up-to level for the ATO problem efficiently.

The condition provided by Theorem 6 is not restrictive. It is satisfied by a number of systems whose significance is well recognized in the literature. We apply it to some examples as follows.

- **W-System**: $A = \begin{bmatrix} D \\ c \end{bmatrix} \in \mathbb{R}^{(n+1) \times n}_+$ where $D = (d_{ij})_{i \in [n], j \in [n]} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with only $d_{ii} > 0$, $\forall i \in [n]$ being nonzero. This system has $(n+1)$ components and $n$ products. The last component is a common component and used in all products; for all other components, each is specific to a single product. The condition in Theorem 6 is satisfied, $g(x, z)$ is supermodular in $z$.

- **M-System**: $A = \begin{bmatrix} D \\ c \end{bmatrix} \in \mathbb{R}^{n \times (n+1)}_+$ where $D = (d_{ij})_{i \in [n], j \in [n]} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with only $d_{ii} > 0$, $\forall i \in [n]$ being nonzero. This system has $n$ components and $(n+1)$ products. The last product uses all components; for all other products, each is specific to a single component. The condition in Theorem 6 is not satisfied, and hence $g(x, z)$ is not supermodular in $z$ in general.

- **Tree system**: This system was first studied by Zipkin (2016). For any $i \in [l]$, denote $S_i = \{j : a_{ij} \neq 0\}$ being the index set of products which require component $i$. Tree systems are those such that for any two components $i, i'$ with $S_i \cap S_{i'} \neq \emptyset$, we have either $S_i \subseteq S_{i'}$ or $S_{i'} \subseteq S_i$. While the general tree system does not guarantee supermodularity of $g(x, z)$, the following two special cases do.

  — **Binary tree system**: all elements in $A$ are either 0 or 1.

  — **Proportional tree system**: for any two components $i, i'$ with the set of common products $C = S_i \cap S_{i'} \neq \emptyset$, $\frac{a_{ij}}{a_{i'j}}$ takes the same value for all $j \in C$.

- **$A \in \mathbb{R}^{l \times 2}_+$**, i.e. there are only two products in the system. $A$ satisfies the condition in Theorem 6, i.e. $g(x, z)$ is supermodular in $z$.

5. **Conclusion**

This paper identifies a tractable class of two-stage distributionally robust optimization problems and derives the precise optimal solutions. We consider the problem with a scenario-based ambiguity
set. Given any realization of the uncertain scenario, we know the information of supports, means and MADs for the underlying uncertainties. Our results show that any two-stage problem has a computationally tractable reformulation whenever its second-stage cost function is supermodular in the uncertainty realization. This reformulation relies on the common worst-case distribution, which is independent of the first-stage decision and can be pre-calculated via an efficient algorithm. As a result, our reformulation preserves the original structure of the problem and keeps the computational complexity at the same level as the nominal problem. While the reformulation is based on the requirement for supermodularity in the second-stage problem, we provide a necessary and sufficient condition to check whether this requirement is met for any given two-stage problem.

Subsequently, it can be verified that a wide range of practical problems fit within our framework of two-stage distributionally robust optimization with supermodularity. Instances include multi-item newsvendor problem, appointment scheduling problem and general ATO systems. While these problems are considered to be computationally challenging in general, with our approach, they can be solved exactly and efficiently.

References


Mak, Ho-Yin, Ying Rong, Jiawei Zhang. 2014. Appointment scheduling with limited distributional information. Management Science 61(2) 316–334.


Appendix

Proof of Theorem 5

The matrix form of (19) can be written as

$$
g(x, z) = \min \ 1^T y$$

s.t. \( U y - V z \leq -W x, \) \hspace{1cm} (23)

where \( U = -\tilde{U}, \ V = W = -\tilde{V}, \ \tilde{U} = \begin{bmatrix} \tilde{U}^1 \\ \vdots \\ \tilde{U}^n \\ \tilde{U}^{n+1} \end{bmatrix} \in \mathbb{R}^{2+3n \times n}, \ \tilde{V} = \begin{bmatrix} \tilde{V}^1 \\ \vdots \\ \tilde{V}^n \\ \tilde{V}^{n+1} \end{bmatrix} \in \mathbb{R}^{2+3n \times n} \) such that

\[
\tilde{U}^t \in \mathbb{R}^{t \times n} \text{ are with elements of } \tilde{u}_{js}^t = \begin{cases} 1 & \text{if } t \leq s, \ \text{for } j \in [t], \ s, t \in [n], \\ 0 & \text{otherwise} \end{cases}
\]

\[
\tilde{V}^t \in \mathbb{R}^{t \times n} \text{ are with elements of } \tilde{v}_{js}^t = \begin{cases} 1 & \text{if } j \leq s \leq t, \ \text{for } j \in [t], \ s, t \in [n], \\ 0 & \text{otherwise} \end{cases}
\]

\[
\tilde{V}^{n+1} = I_{n \times n}, \ \tilde{V}^{n+1} = 0_{n \times n}.
\]

We prove the theorem by showing that \( U, V \) satisfy the second part of the condition in Theorem 3. Note that it is equivalent to prove that \( \tilde{U}, \tilde{V} \) satisfy the condition. To this end, we consider any \( I \subseteq \left[ \frac{n^2 + 3n}{2} \right], \beta \in \mathbb{R}^n \) with \( |I| = n + 1, \ \text{rank}(\tilde{U}_I) = n, \ \text{and} \ \tilde{V}_I \beta \in \text{span}(\tilde{U}_I). \)

Note that \( \text{rank}(\tilde{U}) = n, \) and each row of \( \tilde{U}_I \in \mathbb{R}^{(n+1) \times n} \) has only one nonzero element which takes a value of 1. Therefore, there exists \( \omega \in [n] \) such that \( \tilde{U} \) has two row vectors being \( e_\omega^T, \) and exactly one row vector being \( e_i^T \) for each \( i \in [n] \setminus \{\omega\}. \) We let \( RI_1, \ldots, RI_{n+1} \) be the distinct row indices such that \( I = \{RI_1, \ldots, RI_{n+1}\}, \ \tilde{u}_{RI_i} = e_i^T, \ i \in [n] \) and \( \tilde{u}_{RI_{n+1}} = e_\omega^T. \) WLOG, we can let \( RI_\omega < RI_{n+1}. \)

Moreover, for notational brevity, when constructing \( \tilde{U}_I, \) we let its \( i \)-th row vector be \( \tilde{u}_{RI_i}^T, \) and hence \( \tilde{U}_I = \begin{bmatrix} I \\ e_\omega \end{bmatrix}; \ \tilde{V}_I \) is constructed correspondingly. In this case, for any \( \alpha \in \mathbb{R}^n, \ \tilde{U}_I \alpha = \begin{bmatrix} \alpha \\ \alpha_\omega \end{bmatrix}. \)

This implies that, given any \( \gamma \in \mathbb{R}^{n+1}, \) we have \( \gamma \in \text{span}(\tilde{U}) \) if and only if \( \gamma_\omega = \gamma_{n+1}. \) Therefore, by \( \tilde{V}_I \beta \in \text{span}(\tilde{U}_I) \) we know \( \tilde{v}_{RI_\omega, \beta} = \tilde{v}_{RI_{n+1}, \beta}. \) Our objective is to show \( \beta_i(\tilde{V}_I), \in \text{span}(\tilde{U}_I), \) which is equivalent to \( \beta_i \tilde{v}_{RI_i} = \beta \tilde{v}_{RI_{n+1}} \), \( \forall i \in [n]. \) To demonstrate this, we consider two cases.

- Case 1: both \( \tilde{u}_{RI_\omega}^T \) and \( \tilde{u}_{RI_{n+1}}^T \) are extracted from \( \tilde{U}^\omega, \) i.e., \( RI_\omega, RI_{n+1} \in \left\{ \frac{\omega(\omega-1)}{2} + 1, \ldots, \frac{\omega(\omega-1)}{2} + \omega \right\}. \) We denote \( j_\omega = RI_\omega - \frac{\omega(\omega-1)}{2} \) and \( j_{n+1} = RI_{n+1} - \frac{\omega(\omega-1)}{2}, \) i.e., \( \tilde{u}_{RI_\omega}^T \) and \( \tilde{u}_{RI_{n+1}}^T \) are the \( j_\omega \)-th and \( j_{n+1} \)-th rows in \( \tilde{U}^\omega, \) respectively. By the structure of \( \tilde{V}^\omega, \) we know for all \( s \in [n], \)

\[
\tilde{v}_{RI_\omega, s} = \tilde{v}_{j_\omega, s} = \begin{cases} 1 & s = j_\omega, \ldots, \omega \\ 0 & s = 1, \ldots, j_\omega - 1 \ or \ s = \omega + 1, \ldots, n, \end{cases}
\]

\[
\tilde{v}_{RI_{n+1}, s} = \tilde{v}_{j_{n+1}, s} = \begin{cases} 1 & s = j_{n+1}, \ldots, \omega \\ 0 & s = 1, \ldots, j_{n+1} - 1 \ or \ s = \omega + 1, \ldots, n. \end{cases}
\]
In this case, \( \tilde{v}_{RL,i}^T \beta = \tilde{v}_{RL,i+1}^T \beta \) implies \( \sum_{j=j_\omega}^{j_n} \beta_j = \sum_{j=j_{n+1}}^{j_n} \beta_j \); and hence \( \beta_j = 0, \forall j \in \{ j_\omega, \ldots, j_{n+1} - 1 \} \) since \( \beta \geq 0 \). Now for any arbitrary \( i \in [n] \), the equation \( \beta_i \tilde{v}_{RL,i} = \beta_i \tilde{v}_{RL,i+1} \) always holds since 1) \( \tilde{v}_{RL,i} = \tilde{v}_{RL,i+1} = 0 \) when \( i = 1, \ldots, j_\omega - 1 \) or \( i = \omega + 1, \ldots, n \); 2) \( \beta_i = 0 \) when \( i = j_\omega, \ldots, j_{n+1} - 1 \); 3) \( \tilde{v}_{RL,i} = \tilde{v}_{RL,i+1} = 1 \) when \( i = j_{n+1}, \ldots, \omega \).

- Case 2: \( \tilde{u}_{RL,i}^T \) is extracted from \( \tilde{U}^\omega \) while \( \tilde{u}_{RL,i+1}^T \) is extracted from \( \tilde{U}^{n+1} \). The submatrix \( \tilde{V}^{n+1} = \tilde{U}_n \times n \) implies in this case \( \tilde{v}_{RL,i}^T = 0 \). Hence, \( \tilde{v}_{RL,i}^T \beta = \tilde{v}_{RL,i+1}^T \beta \) implies \( 0 = \tilde{v}_{RL,i}^T \beta = \sum_{i \in [n]} \beta_i \tilde{v}_{RL,i} \). Since \( \tilde{v}_{RL,i} \geq 0 \) and \( \beta \geq 0 \), we then have \( \beta_i \tilde{v}_{RL,i} = 0 = \beta_i \tilde{v}_{RL,i+1} \) for all \( i \in [n] \).

Given that function \( g(x, z) \) is supermodular in \( z \) for any feasible \( x \), the linear programming reformulation can be easily obtained following Theorem 1. \( \square \)

**Proof of Lemma 5**

WLOG, we let \( i = 1, j = 2 \), i.e. we prove the condition on the supermodularity with respect to \( z_1, z_2 \). To make use of Theorem 3, based on the formulation (21), we represent the second-stage cost as a function of \( z_1, z_2 \), which is defined by

\[
g_{12}(z_1, z_2) = h^T x + p^T z + \begin{pmatrix} \min & - (A^T h + p + r)^T y \end{pmatrix}
\text{s.t. } U y - V \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \leq -W \begin{bmatrix} z_3 \\ \vdots \\ z_n \end{bmatrix},
\]  

(24)

where

\[
U = \begin{bmatrix} \mathbf{I}_{n \times n} \\ -\mathbf{1}_{n \times n} \\ \mathbf{A} \end{bmatrix} \in \mathbb{R}^{(2n+l) \times n}, 
V = \begin{bmatrix} \mathbf{I}_{2 \times 2} \\ \mathbf{0}_{(2n+l-2) \times 2} \end{bmatrix} \in \mathbb{R}^{(2n+l) \times 2}, 
W = \begin{bmatrix} \mathbf{0}_{2 \times (n-2)} \\ -\mathbf{I}_{(n-2) \times (n-2)} \mathbf{0}_{(n-2) \times 1} \mathbf{0}_{n \times (n-2)} \mathbf{0}_{l \times (n-2)} \mathbf{-I}_{l \times l} \end{bmatrix} \in \mathbb{R}^{(2n+l) \times (n-2+l)}.
\]

Suppose the condition in this lemma is satisfied, then we will show that the second part of the condition in Theorem 3 is also satisfied and hence \( g_{12}(z_1, z_2) \) is supermodular for all corresponding coefficients.

Consider any index set \( I \subseteq [2n+l], \beta \in \mathbb{R}^{2 \times n}_+, \alpha \in \mathbb{R}^n \) such that \( |I| = n + 1, \text{rank}(U_I) = n \) and \( V_I \beta = U_{12} \alpha \). It is not difficult to verify that if \( I \not\subseteq I \setminus_{12} \), we have either \( \beta_i \in \text{span}(U_I) \) or \( \beta_i \in \text{span}(U_{I \setminus_{12}}) \), \( i \in [1, 2] \), and hence the condition is satisfied. Therefore, it remains to consider the case with \( \{1, 2\} \subseteq I \).

Let us define row index sets \( I_1 = I \cap [2n], I_2 = I \cap [2n+1, \ldots, 2n+l] \). Therefore, \( I_1 \cap I_2 = \emptyset \) and \( I_1 \cup I_2 = I \); moreover, \( U_I = \begin{bmatrix} U_{I_1} \\ U_{I_2} \end{bmatrix} \) with \( U_{I_1} \) containing the rows extracted from \( \begin{bmatrix} \mathbf{I}_{n \times n} \\ -\mathbf{I}_{n \times n} \end{bmatrix} \) and \( U_{I_2} \) from \( A \).

We also define column index sets \( J_1 = \{ j \in [n]: j \in I_1 \text{ or } j + n \in I_1 \} \) and \( J_2 = [n] \setminus J_1 \). Since \( \{1, 2\} \subseteq I \), we know \( \{1, 2\} \subseteq I_1 \) and hence \( \{1, 2\} \subseteq J_1 \). Therefore, WLOG, we can assume \( J_1 = \{1, 2\} \subseteq J_1 \).
\{1, \ldots, |J_1|\} and \mathcal{J}_0 = \{|J_1| + 1, \ldots, n\}. We then have \(U_{X_1} = [U_{X_1, J_1} U_{X_1, J_0}]\). Since \(u_i = e_i\) if \(i \in [n]\) and \(u_i = -e_{i-n}\) if \(i \in \{n + 1, \ldots, 2n\}\), we have \(U_{X_1, J_0} = 0\). We further denote \(I_0 = \{i : i + 2n \in I_2\}\), then \(U_{X_2} = A_{x_0} = [A_{x_0, J_1} A_{x_0, J_0}]\). The matrix \(U_{X}\) can then be divided into four submatrix blocks as follows,

\[
U_{X} = \begin{bmatrix}
U_{X_1} \\
U_{X_2}
\end{bmatrix} = \begin{bmatrix}
U_{X_1, J_1} & 0 \\
A_{x_0, J_1} & A_{x_0, J_0}
\end{bmatrix}.
\]

We also remark that the first two rows of \(U_{X_1}\) are \(e_1^T, e_2^T\).

For notational convenience we denote \(k = |I_1| - 1\); hence, \(|I_1| = k + 1, |I_2| = n - k, k \in [n]\). Observing that 1) \(|I_2| = n - k\) yields \(\text{rank}(U_{X_2}) \leq n - k\) and 2) \(\text{rank}(U_{X}) = n\), we have \(\text{rank}(U_{X_1}) \geq k\). Moreover, \(\text{rank}(U_{X_1}) \leq |I_1| = k + 1\). Therefore, \(\text{rank}(U_{X_1}) \in \{k, k + 1\}\). We also observe that \(|J_1| = \text{rank}(U_{X_1})\).

- Case 1: \(|J_1| = \text{rank}(U_{X_1}) = k\). Since each row of \(U_{X}\) contains exactly one nonzero component, we conclude that there exists a unique \(i \in [k]\) such that both \(e_1^T, -e_1^T\) are row vectors in \(U_{X_1}\).

  We first show that \(t \notin \{1, 2\}\) by contradiction. Assume the contrary. WLOG, let \(t = 1\). In this case, both \(e_1^T\) and \(-e_1^T\) are in \(U_{X_1}\). We have concluded above that \(e_1^T, e_1^T\) must be the first two rows of the \(U_{X, J_1}\), which implies that \(-e_1^T\) is not in the first two rows. Therefore, \(U_{X} \alpha = V_{X} \beta = (\beta_1, \beta_2, 0, \ldots, 0)\) requires \(\alpha_1 = \beta_1\) with \(-\alpha_1 = 0\), which contradicts with \(\beta_1 > 0\).

  Now we focus on the case where \(t \in [k]\}\{1, 2\}. For any \(i \in \{1, 2\}\), let \(\gamma_i = \beta_i, \gamma_j = 0 \forall j \in J_1 \{i\}\), and \(\gamma_{J_0} := (\gamma_j)_{j \in J_0}\) is the solution of

\[
A_{x_0, J_0} \gamma_{J_0} = -\beta_i(A_{x_0})_i.
\]

Note that \(U_{X_1, J_0} = 0\) and \(\text{rank}(U_{X}) = n\) imply \(\text{rank}(A_{x_0, J_0}) = n - k\). Therefore, \(A_{x_0, J_0} \in \mathbb{R}^{(n-k) \times (n-k)}\) is invertible, the above equation has a unique solution of \(\gamma_{J_0}\). With \(\gamma \in \mathbb{R}^n\) chosen in this way, we have

\[
\begin{cases}
u_i^T \gamma = e_1^T \gamma = \gamma_i = \beta_i \\
u_j^T \gamma \in \{e_2^T \gamma, -e_2^T \gamma\} = \{0\} \forall j \in I_1 \{i\} \\
U_{X_1} \gamma = A_{x_0} \gamma = \beta_i(A_{x_0})_i + A_{x_0, J_0} \gamma_{J_0} = 0.
\end{cases}
\]

Therefore, we have \(\beta_i(V_{X})_i = U_{X} \gamma \in \text{span}(U_{X})\).

- Case 2: \(|J_1| = \text{rank}(U_{X_1}) = k + 1\). In this case, \(\text{rank}(A_{x_0}) = \text{rank}(A_{x_0, J_0}) = n - k - 1\). For any \(i \in \{1, 2\}\), we choose \(\gamma\) in the same way as in the previous case, with the only difference being that now \(A_{x_0, J_0} \in \mathbb{R}^{(n-k) \times (n-k-1)}\) is full column rank instead of invertible, which does not affect the existence of the solution \(\gamma_{J_0}\) given that the condition 4) in this lemma is satisfied. Similarly, we still have \(\beta_i(V_{X})_i = U_{X} \gamma \in \text{span}(U_{X})\).

  Hence, the condition in Theorem 3 is always satisfied and the proof is completed. \qed
Lemma 6 (Chen et al. 2018b) Consider any matrix $U \in \mathbb{R}^{r \times m}$ with $\text{rank}(U) < r$. Suppose that system \[ \begin{align*}
U x &\leq \overline{c} \\
-U x &\leq -\underline{c}
\end{align*}\] is infeasible. Then there exists $\mathcal{I} \subseteq [r]$ with $|\mathcal{I}| = \text{rank}(U) + 1$ and $\text{rank}(U_{\mathcal{I}}) = \text{rank}(U)$ such that system \[ \begin{align*}
U_{\mathcal{I}} x &\leq \overline{c}_{\mathcal{I}} \\
-U_{\mathcal{I}} x &\leq -\underline{c}_{\mathcal{I}}
\end{align*}\] is also infeasible.

Lemma 7 (Chen et al. 2018b) Consider any matrix $Q \in \mathbb{R}^{s \times (s+1)}$ with $\text{rank}(Q) = s$, $s \geq 2$. If every $2 \times 3$ submatrix of $Q$ contains at least two column vectors which are linearly dependent, then $Q$ has at least two column vectors which are linearly dependent.