The Dao of Robustness
Achieving Robustness in Prescriptive Analytics

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

Melvyn Sim, Minglong Zhou
Department of Analytics & Operations (DAO), NUS Business School, National University of Singapore
melvynsim@gmail.com, minglong_zhou@u.nus.edu

We present a model for optimization under uncertainty called robustness optimization that favors solutions
for which the model’s constraint would be the most robust or least fragile under uncertainty. The decision
maker does not have to size the uncertainty set, but specifies an acceptable target, or loss of optimality
compared to the baseline model, as a tradeoff for the model’s ability to withstand greater uncertainty. We
axiomatize the decision criterion associated with robustness optimization, termed as the fragility measure,
which is a class of Brown and Sim (2009) satisficing measure, and it satisfies the properties of monotonicity,
positive homogeneity, subadditivity, pro-robustness, and anti-fragility. We provide a representation theorem
and connect it with known fragility measures including the decision criterion associated with the GRC-sum
of Ben-Tal et al. (2017) and the riskiness index of Aumann and Serrano (2008). We present a suite of
practicable robustness optimization models for prescriptive analytics including linear, adaptive linear, data-
driven adaptive linear, combinatorial and dynamic optimization problems. Similar to robust optimization, we
show that robustness optimization via minimizing the fragility measure can also be done in a tractable way.
We also provide numerical studies on static, adaptive, and data-driven adaptive problems and show that the
solutions to the robustness optimization models can withstand greater impact of uncertainty compared to the
Corresponding robust optimization models without increasing the cost or incurring additional computational
effort.

Key words: Robust optimization, robustness optimization, prescriptive analytics, fragility measure

History: June 30, 2020.

1. Introduction

“The Dao follows the way of nature (道法自然)” – Lao Tzu

Optimization under uncertainty, despite its importance and ubiquity in real-world problems,
has been a perennial difficulty in prescriptive analytics. The overarching challenge in optimization
under uncertainty is formulating an efficiently, or at least practicably, solvable model that would
sufficiently mitigate the adverse impact of uncertainty. A tractable optimization problem under uncertainty is often associated with a model of uncertainty, which restricts the nature of uncertainty. If nature were to be restricted in this way, then the solution to the model would also be optimal with respect to the model of uncertainty.

In a stochastic linear optimization problem, we often assume a model of uncertainty with a discrete distribution to avoid high dimensional integration and enable us to solve the problem using large-scale linear optimization techniques (see, e.g., Kall and Wallace 1994, Shapiro et al. 2009, Birge and Louveaux 2011). In a dynamic optimization problem, we would often assume an underlying Markov process to reduce the state-space of the problem in an attempt to circumvent the “curse of dimensionality”. In a robust optimization problem, the uncertainty model assumes an adversarial nature that selects the parameters from an uncertainty set that would have the worst impact on the model (see, e.g., Soyster 1973, Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998). The budgeted uncertainty set proposed by Bertsimas and Sim (2004) retains the linear optimization framework, which can be extended to solving discrete optimization problems. In particular, Bertsimas and Sim (2003) show that the same model of uncertainty also retains the computational complexity of combinatorial optimization problems. In distributionally robust optimization, which generalizes robust optimization, the adversarial nature selects the worst probability distributions over an ambiguity set of probability distributions (see, e.g., Delage and Ye 2010, Wiesemann et al. 2014, Bertsimas et al. 2019, Chen et al. 2020). Similar to robust optimization, the computational tractability of the distributionally robust optimization problem depends on the characterization of the ambiguity set.

The fact remains that we can only be certain about the uncertainty of nature, and this begs the question - what if we were wrong about the model of uncertainty? Indeed, it has well been known that solving an optimization model that assumes a deterministic state of nature could result in unacceptably fragile solutions (see, e.g., Ben-Tal et al. 2009), yet few decision makers would be willing to pay the price of a fully robust solution that could withstand the worst-case impact. For better performance in objective, most would be willing to tolerate some levels of model infeasibility when facing uncertain outcomes.

Robust optimization approach addresses this problem by fashioning an uncertainty set that allows the parameters to vary within a prescribed neighborhood from their deterministic nominal values. It is also well established in robust optimization that under some probabilistic assumptions, modest sized uncertainty sets could provide very high probabilistic protection of a linear constraint with affine perturbation against infeasibility (see, e.g., Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004). However, if we were wrong about the probabilistic assumptions, a robust optimization
model with a modest sized uncertainty set may not sufficiently protect the model against uncertain outcomes.

The same issues arise in stochastic and dynamic optimization where we assume an uncertainty model comprising random variables with known probability distributions. When using sample average approximation (SAA) as a tractable approximation to stochastic optimization problems, the out-of-sample performance can be rather inferior compared to the results reflected in the SAA model (Kleywegt et al. 2002). In the data-driven setting, due to the optimizer’s curse (Smith and Winkler 2006), if we solve a stochastic optimization problem using an empirical distribution from a dataset and test the out-of-sample outcome on another, we should also expect inferior results. Distributionally robust optimization models with probability-distance-based ambiguity sets (see, e.g., Ben-Tal et al. 2013, Mohajerin Esfahani and Kuhn 2018) can provide “regularized” solutions, which can effectively mitigate the optimizer’s curse. However, it may be difficult to determine the right probability distance parameter for the ambiguity set. Such parameter is usually chosen via cross-validation techniques (Mohajerin Esfahani and Kuhn 2018), provided there are sufficient data to do so effectively.

In this paper, we present a model for optimization under uncertainty called robustness optimization. Unlike robust optimization approaches, we allow nature to take its course, even rendering solutions infeasible, and favor those for which the model’s constraint would be the most robust or least fragile under uncertainty. In the robustness optimization model, the decision maker specifies a target, or an acceptable loss of optimality compared to the baseline model, as a tradeoff for the model’s ability to withstand greater uncertainty. The use of target is prevalent in management decision making (see, e.g, Simon 1959, Mao 1970, Chen and Tang 2019) and we believe that articulating preference for robustness from a target-driven perspective is also more interpretable for the decision maker.

We axiomatize the decision criterion associated with the robustness optimization, termed as the fragility measure, which relates to the maximum level of model infeasibility that may occur relative to the magnitude of deviation from the baseline uncertainty. A fragility measure satisfies the properties of monotonicity, positive homogeneity, subadditivity, pro-robustness, and anti-fragility and it belongs to a class of satisficing measures (Brown and Sim 2009, Brown et al. 2012). We provide a representation theorem and connect with fragility measures that are known in the literature including the riskiness index of Aumann and Serrano (2008).

Our work is also related to the stream of research along the lines of comprehensive robust optimization of Ben-Tal et al. (2006) or more recently termed as globalized robust optimization proposed by Ben-Tal et al. (2017), which permits the modeler to specify how the magnitude of constraint violation should be bounded when the realization of the uncertainty occurs outside the uncertainty
set. This has been extended to probabilistic enveloping constraints (Xu et al. 2012) and soft robust model (Ben-Tal et al. 2010), which has been shown to be related to convex measure of risks (Follmer and Schied 2004). Incidentally, to provide the parameters of the globalized robust optimization model, Ben-Tal et al. (2017) introduce a GRC-sum model (Section 5.2), which is a non-stochastic robustness optimization model for which the baseline problem is a robust optimization model.

We show that robustness optimization models for prescriptive analytics can be solved in a tractable way including linear, adaptive linear, data-driven adaptive linear, combinatorial, and dynamic optimization models. We provide computational studies on static, adaptive, and data-driven adaptive problems and show that for the same objective values attained, the solutions to the robustness optimization models can withstand greater impact of uncertainty compared to classical robust optimization models without incurring additional computational effort. We also illustrate in a data-driven adaptive network lot-sizing problem on how we can set targets in the robustness optimization model to improve solutions over the empirically optimized model, which is more intuitive and effective than specifying the hyper-parameter used in a statistical-distance-based robust optimization model.

**Notation.** We use boldface lowercase letters for vectors (e.g., $\mathbf{\theta}$), and calligraphic letters for sets (e.g., $\mathcal{X}$). We use $[N]$ to denote the running index $\{1, 2, 3, \ldots, N\}$ for $N$ a known integer. We adopt the convention that $\inf \emptyset = +\infty$, where $\emptyset$ is the empty set. A random variable $\hat{\nu}$ is denoted with a tilde sign such as $\hat{\nu} \sim \mathbb{P}, \mathbb{P} \in \mathcal{P}_0$, where $\mathcal{P}_0$ to represent the set of all possible distributions. For $\hat{\nu}_1, \hat{\nu}_2$, we use $\hat{\nu}_1 \succeq \hat{\nu}_2$ to imply $\hat{\nu}_1$ state-wise dominates $\hat{\nu}_2$. For multivariate random variable, we use $\mathcal{P}_0(\mathcal{Z})$ to represent the set of all distributions for the multivariate random variable that has support $\mathcal{Z} \subseteq \mathbb{R}^N$. Specifically, we use $\hat{\mathbf{Z}} \sim \mathbb{P}, \mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$ to define $\hat{\mathbf{Z}}$ as a multivariate random variable with support $\mathcal{Z}$ and distribution $\mathbb{P}$. For two distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$, we use $\mathbb{P} \ll \hat{\mathbb{P}}$ to denote that $\mathbb{P}$ is absolutely continuous with respect to $\hat{\mathbb{P}}$.

### 2. Motivating Robustness Optimization

To motivate the model which we call robustness optimization, we first consider the following deterministic baseline optimization problem,

$$
Z_0 = \max \mathbf{c}^\top \mathbf{x}
$$

$$
s.t. \quad f(\mathbf{x}, \hat{\mathbf{z}}) \leq 0
$$

$$
\mathbf{x} \in \mathcal{X},
$$

for a given function $f : \mathcal{X} \times \mathcal{Z} \mapsto \mathbb{R}$, where the input to the second argument is subject to uncertainty. The uncertain parameters of the problem are collectively denoted by the random variable $\hat{\mathbf{z}}$ over the support $\mathcal{Z} \subseteq \mathbb{R}^N$, which is assumed bounded. For the deterministic problem, only a chosen baseline value $\hat{\mathbf{z}} \in \mathcal{Z}$ is considered in the constraint.
To better protect the constraint against infeasibility, robust optimization solves the following problem,

\[
Z_r = \max c^\top x \\
\text{s.t. } f(x, z) \leq 0 \quad \forall z \in \mathcal{U}(r) \\
x \in \mathcal{X},
\]

(2)

where \( \mathcal{U}(r) \) is typically a norm-based adjustable uncertainty set,

\[
\mathcal{U}(r) = \{z \in \mathcal{Z} \mid ||z - \hat{z}|| \leq r\}.
\]

The absolutely robust model is given by

\[
Z_\infty = \max c^\top x \\
\text{s.t. } f(x, z) \leq 0 \quad \forall z \in \mathcal{Z} \\
x \in \mathcal{X},
\]

though this approach may be perceived as overly conservative, especially when \( Z_\infty \) is significantly lower than \( Z_0 \).

The most commonly adopted budgeted uncertainty set of Bertsimas and Sim (2004) is a norm-based adjustable uncertainty set with \( \mathcal{Z} = [-1, 1]^N \) and \( \ell_1 \)-norm for which the baseline value is \( \hat{z} = 0 \). The parameter \( r \) is intuitively interpreted as the maximum number of uncertain parameters that nature chooses to deviate from their baseline values.

The robust optimization problem necessitates the solution \( x \) to remain feasible for any realization of \( z \) in the uncertainty set. It is important to note that robust optimization model does not specify how the constraint should behave whenever the uncertain parameters deviate from the uncertainty set. To understand the potential issue, consider the following robust continuous knapsack problem under the budgeted uncertainty set,

\[
\max \sum_{i \in [50]} x_i \\
\text{s.t. } z^\top x \leq 50 \quad \forall z \in \mathcal{U}(50) \\
x \in [0, 1]^{100}.
\]

Observe that there would be multiple optimal solutions to Problem (3), with \( x_i^* = 1 \) for all \( i \in [50] \), though the one also with \( x_i = 0 \) for all \( i \in [N] \setminus [50] \) would be the most preferred optimal solution since it leads to feasibility of the first constraint for all \( z \in \mathcal{Z} \). The issue of having multiple optimal solutions in such robust optimization problems has also been observed in Iancu and Trichakis (2014). We could resolve this issue if the uncertain constraint is evaluated for all possible realizations of \( z \in \mathcal{Z} \), taking into account of how different realizations would directly impact on the feasibility of the constraint.
Consider the following optimization problem,

\[
\kappa_0 = \min k \\
\text{s.t. } f(x, z) \leq k\|z - \hat{z}\| \quad \forall z \in Z \\
c^\top x \geq \tau \\
x \in \mathcal{X}, k \geq 0,
\]

which we call a robustness optimization model. Unlike robust optimization, the robustness optimization model allows uncertainty to vary over the entire support, while controlling the level of infeasibly as much as possible whenever nature deviates from the baseline value. Observe that in robust optimization, the norm relating to the deviation of uncertain parameters from their baseline values is articulated in the uncertainty set and it is decoupled from the critical constraint that is impacted by uncertainty. Because not all uncertainty are adversarial, without associating with how badly the constraint might be violated, the robust optimization approach may not be as effective in mitigating constraint violation risks. In contrast, the same norm in robustness optimization is directly associated with the feasibility of the model and relates to how the magnitude of deviation may affect the degree of infeasibility of the problem. Specifically, the maximum allowable constraint violation is proportional to how much \( z \) would deviate from the baseline value \( \hat{z} \). The optimal proportionality factor \( \kappa_0 \) effectively describes the level of fragility of the model. As it decreases in value, the lower magnitude of infeasibility the model could occur, relative to how far the uncertain outcome deviates from the baseline value. Absolute robustness is ensured when the level of fragility is zero. Robustness of a model is measured by how much the level of fragility can be reduced for the model.

Observe that the constraint of Problem (1) is implied at \( z = \hat{z} \), hence, any solution that is infeasible in the deterministic baseline problem would also be so in the robustness optimization problem. Because the robustness optimization model does not restrict \( z \) to an uncertainty set, it can also eliminate the degeneracy issue we face in the robust knapsack problem (3). In particular, with \( \tau = 50 \), the absolutely robust solution, for which \( \kappa_0 = 0 \) is achieved uniquely when \( x_i = 1 \) for \( i \in [50] \) and zero otherwise.

Another important difference from the robust optimization model is that the decision maker specifies the parameter \( \tau \), as opposed to setting the size of the uncertainty set \( r \) for the robust optimization model. Without other assumptions on the underlying uncertainity, it is usually not intuitive for the decision maker to pin down the value of \( r \) for the robust optimization problem, though this may not be true in engineering related problems where the parameter \( r \) could be easy to specify. A common approach is to relate \( r \) to some probabilistic guarantees of feasibility, based on additional distributional assumptions of the underlying uncertainty. For instance, Bertsimas
and Sim (2004), Chen et al. (2007b) provide a probabilistic guarantee for a linear constraint under
affine perturbation. Under the assumptions that the random variable $\tilde{z}$ has support $[-1, 1]^N$ and its
components are independently distributed with means precisely at the origin, by choosing $r = c\sqrt{N}$, the
robust solution under the budgeted uncertainty set would guarantee a feasible probability of
at least $1 - \exp(-c^2/2)$. However, it is important to note that the bound can be rather weak and
may not apply if the function $f(x, z)$ is not biaffine. For instance, if the function is associated with
the objective function of an adaptive linear optimization problem given by

$$f(x, z) = \min_{y, v} v$$

s.t. $\sum_{i \in [N]} y_i \leq x + v$

$y \geq z$

$y \geq -z,$

so that $f(x, z) = ||z||_1 - x$, then we would have $\mathbb{P}[f(r, \tilde{z}) \leq 0] = \mathbb{P}[\tilde{z} \in \mathcal{U}(r)]$, which is the minimal
probability guarantee assured by the robust optimization solution. However, such assurance may
not be adequate for $r = c\sqrt{N}$ as demonstrated in the following result.

**Proposition 1.** Let $\tilde{z}_i$, $i \in [N]$, be independent random variables with support $Z = [-1, 1]^N$,
means at the origin, and total variance of at least $N\theta > 0$. For a given budgeted uncertainty set
$\mathcal{U}(r)$ and $c \in (0, \theta\sqrt{N})$, then

$$\mathbb{P}[\tilde{z} \in \mathcal{U}(c\sqrt{N})] \leq \exp\left(-\frac{(\theta\sqrt{N} - c)^2}{2\theta}\right).$$

(5)

Hence, the probability that the realization of uncertain parameter $\tilde{z}$ lies in the uncertainty set
$\mathcal{U}(c\sqrt{N})$ decreases exponentially to zero as the dimension $N$ increases. Therefore, one should exercise caution in using the budgeted uncertainty set beyond affinely perturbed constraints because its probability bound can become ineffective as $N$ increases.

In contrast, for the robustness optimization problem, the input parameter $\tau$, $\tau \in [Z_\infty, Z_0]$ is
directly related to the objective value of the model the decision maker is addressing, and it can be interpreted as the target objective she is willing to accept, relative to a reference – the baseline optimization objective, $Z_0$. The model will then determine the most robust solution which also achieves the target objective. Therefore, the robustness optimization benchmarks the baseline problem as a reference. While one may argue that target, $\tau$ is hard to specify, it should at least be more tangible and intuitive to specify compared to the parameter $r$ used in the robust optimization problem (2).
Despite the differences between the robust and robustness optimization problems, there is one important similarity - both approaches have the same computational complexity. In particular, the robust counterpart in the robust optimization problem is

$$f(x, z) \leq 0 \quad \forall z \in Z : \|z - \hat{z}\| \leq r,$$

while the robust counterpart for the robustness optimization problem can be expressed as

$$f(x, z) - ku \leq 0 \quad \forall z \in Z, u \in \mathbb{R} : \|z - \hat{z}\| \leq u,$$

which can be optimized efficiently depending on the function $f(x, z)$ and the support set $Z$. We refer interested readers to Ben-Tal and Nemirovski (1998), El Ghaoui et al. (1998), Ben-Tal et al. (2015) for how the robust counterparts for many interesting functions can be reformulated as a tractable optimization problem. A recent exciting work on approximating hard robust counterparts has also been proposed in Roos et al. (2018).

Hence, as in the case of robust optimization models, the choice of distance function depends on, among other things, how we could efficiently optimize the robust counterparts and, more importantly, how well the solutions would perform in empirical tests.

Note also that while we focus on uncertainty at the constraint, we can also consider cases where the uncertain parameters appear at the objective function as follows:

<table>
<thead>
<tr>
<th>Baseline</th>
<th>Robust</th>
<th>Robustness ($\tau \in [Z_0, Z_{\infty}]$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_0 = \min f(x, \hat{z})$</td>
<td>$Z_r = \min \max f(x, z)$</td>
<td>$\min k$</td>
</tr>
<tr>
<td>s.t. $x \in \mathcal{X}$</td>
<td>s.t. $x \in \mathcal{X}$</td>
<td>s.t. $f(x, z) - \tau \leq k |z - \hat{z}|$, $\forall z \in Z$</td>
</tr>
<tr>
<td>$x \in \mathcal{X}, k \geq 0$</td>
<td>$x \in \mathcal{X}, k \geq 0$</td>
<td></td>
</tr>
</tbody>
</table>

We emphasize that robustness optimization should not be misconstrued as a relaxation of robust optimization. For instance, the robust constraint in Problem (2) indicates

$$f(x, z) \leq 0 \quad \forall z \in \mathcal{U}(r)$$
$$f(x, z) \leq +\infty \quad \forall z \in Z \setminus \mathcal{U}(r),$$

while the robustness constraint in Problem (4) implies that

$$f(x, z) \leq \kappa_0 \|z - \hat{z}\| \quad \forall z \in \mathcal{U}(r)$$
$$f(x, \hat{z}) \leq \kappa_0 \|z - \hat{z}\| \quad \forall z \in Z \setminus \mathcal{U}(r).$$

Hence, while the robustness optimization model appears to be a relaxation of robust optimization for $z \in \mathcal{U}(r)$, it is more constrained than the robust optimization model when the realization is outside the uncertainty set. Note that $\mathcal{U}(r)$ is typically a small set compared to the support $Z$, and
hence we are actually tightening the robustness constraint for a much larger set \( \mathcal{Z} \setminus \mathcal{U}(r) \), which the robust optimization model simply ignores.

A more general robustness optimization model has been proposed by Ben-Tal et al. (2017), when they introduce the GRC-sum model, which for a single constraint has the following representation,

\[
\kappa_r = \min k \\
\text{s.t. } f(x, z) \leq k \min_{u \in \mathcal{U}(r)} \{ \|z - u\| \} \forall z \in \mathcal{Z} \\
c^\top x \geq \tau \\
x \in \mathcal{X}, k \geq 0.
\]  

Incidentally, the corresponding baseline optimization problem is a robust optimization model of Problem (2), which ensures that the constraint is feasible for all \( z \in \mathcal{U}(r) \). As the uncertainty set has the same norm, observe that

\[
\min_{u \in \mathcal{U}(r)} \{ \|z - u\| \} = \max\{ \|z - \hat{z}\| - r, 0 \},
\]

and the GRC-sum model is simplified to

\[
\min k \\
\text{s.t. } f(x, z) \leq k \max\{ \|z - \hat{z}\| - r, 0 \} \forall z \in \mathcal{Z} \\
c^\top x \geq \tau \\
x \in \mathcal{X}, k \geq 0,
\]

which is nearly computationally as attractive as the robust and robustness optimization models. The GRC-sum requires the modeler to specify the parameter \( r \geq 0 \), which is useful in cases, such as with physical limitations or design specifications, where \( r \) is mandated by requirement. Nevertheless, the impact to the feasibility of the uncertain constraint as \( r \) varies is mixed. Observe that \( \kappa_r \) is non-decreasing in \( r \), hence, while it will improve the feasibility of the problem for \( z \in \mathcal{U}(r) \), in may also compromise the feasibility of the constraint for \( z \in \mathcal{Z} \setminus \mathcal{U}(r) \). The tradeoffs can be evaluated empirically when the uncertainties are stochastic, which we have done so in our simulation studies.

3. Fragility-based Robustness Optimization

To generalize the robustness optimization model, we need to extend to stochastic uncertainty and we present the following baseline optimization problem,

\[
Z_0 = \max c^\top x \\
\text{s.t. } \mathbb{E}_\theta[f(x, \hat{z})] \leq 0 \\
x \in \mathcal{X},
\]
where $\tilde{z}$ is a random variable with a baseline probability distribution $\hat{P} \in \mathcal{P}_0(\mathcal{Z})$. In prescriptive analytics, $\hat{P} \in \mathcal{P}_0(\mathcal{Z})$ is often associated with the empirical distribution in a data-driven setting, which is different from the actual distribution that is unknown to the modeler. Hence, to avoid overfitting to $\hat{P}$, the distributionally robust formulation to the above baseline problem is

$$Z_r = \max c^\top x$$

s.t. $\mathbb{E}_\mathbb{P} [f(x, \tilde{z})] \leq 0 \quad \forall \mathbb{P} \in \mathcal{B}(r)$

$$x \in \mathcal{X},$$

where the probability-distance-based ambiguity set is given by

$$\mathcal{B}(r) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \mid \tilde{z} \sim \mathbb{P}, \ \Delta(\mathbb{P}, \hat{\mathbb{P}}) \leq r \right\},$$

in which $\Delta(\mathbb{P}, \hat{\mathbb{P}})$ represents the probability distance of distributions $\mathbb{P}$ from $\hat{\mathbb{P}}$ defined as follows:

**DEFINITION 1.** A probability distance function, $\Delta(\mathbb{P}, \hat{\mathbb{P}})$ is a nonnegative function on the domain of probability distributions such that $\Delta(\mathbb{P}, \hat{\mathbb{P}}) = 0$ if $\mathbb{P} = \hat{\mathbb{P}}$.

Under this definition, the probability-distance-based ambiguity sets would include $\varphi$-divergence (Pardo 2006, Ben-Tal et al. 2013, Gotoh et al. 2020, 2018) and Wasserstein metric, also known as Kantorovich–Rubinstein metric (Gao and Kleywegt 2016, Mohajerin Esfahani and Kuhn 2018).

We now present a more representative (distributionally) robustness optimization model,

$$\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \mathbb{E}_\mathbb{P} [f(x, \tilde{z})] \leq k \Delta(\mathbb{P}, \hat{\mathbb{P}}) \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\
& \quad c^\top x \geq \tau \\
& \quad x \in \mathcal{X}.
\end{align*}$$

To express non-stochastic robustness optimization models, we let $\hat{\mathbb{P}} : \hat{\mathbb{P}}[\tilde{z} = \hat{z}] = 1$ and $\Delta(\mathbb{P}, \hat{\mathbb{P}}) = \mathbb{E}_\mathbb{P} [\delta(\tilde{z}, \hat{z})]$, for some nonnegative distance function satisfying $\delta(\tilde{z}, \hat{z}) = 0$. Define $g(x, k, z) := f(x, z) - k \delta(z, \hat{z})$ so that the safeguarding constraint of Problem (8) can be expressed as

$$\mathbb{E}_\mathbb{P} [g(x, k, \tilde{z})] \leq 0 \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}).$$

Observe that

$$\sup_{z \in \mathcal{Z}} g(x, k, z) \leq \sup_{\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})} \mathbb{E}_\mathbb{P} [g(x, k, \tilde{z})] \leq \sup_{\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})} \text{ess sup}_{\mathbb{P}} [g(x, k, \tilde{z})] \leq \sup_{\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})} \sup_{z \in \mathcal{Z}} g(x, k, z),$$

which implies that the safeguarding constraint is equivalent to,

$$f(x, z) \leq k \delta(z, \hat{z}) \quad \forall z \in \mathcal{Z},$$
hence generalizing the earlier non-stochastic robustness optimization models that we have presented.

Interestingly, with the same probability distance metric, the distributionally robust optimization (7) is not necessarily equivalent to the corresponding robust optimization problem (2), unless the function $f(x, z)$ is concave in $z$, for which the result can easily be shown using Jensen’s inequality. Hence, there is a greater connection between distributionally robustness optimization and robustness optimization than the connection between distributionally robust optimization and robust optimization.

The decision criterion associated with the robustness optimization model (8) is unusual. Hence, it is important to justify the decision criterion, which we call the **fragility measure**, by laying out some reasonable properties or axioms the criterion is based on. As we will reveal, the fragility measure belongs to a class of satisfying measures proposed by Brown and Sim (2009). Let $(\Omega, \mathcal{F})$ be a measurable space with $\sigma$-algebra $\mathcal{F}$. The baseline probability distribution is given by $\hat{P}$, $\hat{P} \in \mathcal{P}_0$. Denote $\mathcal{L} = \{ \hat{v} | \hat{v} \in L^\infty(\Omega, \mathcal{F}, \hat{P}), \forall \hat{P} \in \mathcal{P}_0 \}$ as the space of bounded measurable real valued functions under all distributions in $\mathcal{P}_0$. Note that such an assumption is common in the literature to obtain a useful representation theorem, though it would also rule out the random variable $f(x, \hat{z})$, $\hat{z} \sim P$, $P \in \mathcal{P}_0(\mathcal{Z})$ for an unbounded support $\mathcal{Z}$ for which $\sup_{z \in \mathcal{Z}} f(x, z) = \infty$. Let $\hat{v} \in \mathcal{L}$ denote the random variable representing uncertain outcomes of a model’s constraint, with positive values being infeasible.

**Definition 2 (Fragility Measure).** The functional $\rho : \mathcal{L} \mapsto [0, +\infty]$ is a fragility measure associated with the baseline probability $\hat{P} \in \mathcal{P}_0$ if and only if it has the following representation

$$\rho(\hat{v}) = \min k$$

s.t. $E_{\hat{P}} [\hat{v}] \leq k \Delta(\hat{P}, \hat{P}) \quad \forall \hat{P} \in \mathcal{P}_0$

$$k \geq 0,$$

for some probability distance function $\Delta$.

With this definition, Problem (8) is equivalent to

$$\min \rho(f(x, \hat{z}))$$

s.t. $c^T x \geq \tau$

$$x \in \mathcal{X}.$$  \hspace{1cm} (10)

We next show that the fragility measure is associated with the salient properties that are consistent with coherent decision making for reducing fragility of the solution and attaining model’s robustness, while ensuring tractability of the problem when the criterion is to be minimized.
Theorem 1 (Axioms of Fragility Measure). The functional $\rho : \mathcal{L} \mapsto [0, +\infty]$ is a fragility measure associated with the baseline probability $\bar{P} \in \mathcal{P}_0$ if and only if it is lower semi-continuous and satisfies the following properties:

1. **Monotonicity:** If $\tilde{v}_1 \geq \tilde{v}_2$, then $\rho(\tilde{v}_1) \geq \rho(\tilde{v}_2)$.
2. **Positive homogeneity:** For any $\lambda \geq 0$, we have $\rho(\lambda \tilde{v}) = \lambda \rho(\tilde{v})$.
3. **Subadditivity:** $\rho(\tilde{v}_1 + \tilde{v}_2) \leq \rho(\tilde{v}_1) + \rho(\tilde{v}_2)$.
4. **Pro-robustness:** If $\tilde{v} \leq 0$, then $\rho(\tilde{v}) = 0$.
5. **Anti-fragility:** If $\mathbb{E}_\bar{P}[\tilde{v}] > 0$, then $\rho(\tilde{v}) = +\infty$.

Moreover, the probability distance $\Delta$ associated with $\rho$ is given by

$$\Delta(\bar{P}, \tilde{P}) = \sup_{\tilde{v} \in \mathcal{L}} \left\{ \mathbb{E}_\bar{P}[\tilde{v}] \mid \rho(\tilde{v}) \leq 1 \right\}.$$  \hspace{1cm} (11)

The first three properties of the fragility measures coincide with three out of the four axioms of coherent monetary risk measures proposed by Artzner et al. (1999). The monotonicity property requires that if the uncertain outcomes of the model’s constraint are never smaller in values in the direction of infeasibility compared to another random variable for all scenarios, then its fragility measure should not reflect a lower value than the other. Positive homogeneity property dictates that the fragility measure scales accordingly with the underlying uncertainty. Consequently, the fragility measure would have the same “units” as the constraint. Likewise, the property of subadditivity is synonymous with the preference for risk pooling, which is associated with uncertainty aversion. It implies that the collective fragility measure of the combined uncertainty in meeting one aggregate constraint should be smaller than the sum of the fragility measures if the feasibility of the uncertain constraints are considered separately. Incidentally, positive homogeneity and subadditivity imply convexity, which is also an important precursor to obtaining a tractable optimization model when the fragility measure is to be minimized.

The pro-robustness property asserts that if the model’s constraint is absolutely feasible, then the corresponding fragility measure should be the lowest value at zero. The anti-fragility property ensures that any solution with finite fragility measure would also be feasible in the baseline constraint, i.e., $\mathbb{E}_\bar{P}[\tilde{v}] \leq 0$. These two properties would rule out any monetary risk measure as a candidate for fragility measure, since they would violate the translation invariance property (see Definition 3).

Theorem 1 shows that any metric that satisfies the salient properties is a fragility measure, even though it may not have the same explicit probability-distance-based representation. Note that the reciprocal a fragility measure corresponds to a satisficing measures of Brown and Sim (2009). However, unlike the satisficing measure, the fragility measure does not include a target, since it
evaluates the fragility of the uncertain constraint \( f(x, \tilde{z}) \leq 0 \) for which the right hand side value is zero. As in Problem (10), the target of the robustness optimization model may not be articulated at the uncertain constraint but at the objective function associated with the baseline optimization problem. Similar to the satisficing measures, the fragility measure has a risk-based representation via normalized convex risk measure.

**Definition 3.** A normalized convex risk measure is a lower semi-continuous functional \( \mu : \mathcal{L} \mapsto \mathbb{R} \), that satisfies the following properties:

1. **Monotonicity:** If \( \tilde{v}_1 \geq \tilde{v}_2 \), then \( \mu(\tilde{v}_1) \geq \mu(\tilde{v}_2) \).
2. **Translation invariance:** For any \( a \in \mathbb{R} \), \( \mu(\tilde{v} + a) = \mu(\tilde{v}) + a \).
3. **Convexity:** For any \( \lambda \in [0, 1] \), \( \mu(\lambda \tilde{v}_1 + (1 - \lambda)\tilde{v}_2) \leq \lambda \mu(\tilde{v}_1) + (1 - \lambda)\mu(\tilde{v}_2) \).
4. **Normalization:** \( \mu(0) = 0 \).

**Proposition 2 (Risk-based Representation).** The functional \( \rho : \mathcal{L} \mapsto [0, +\infty] \) is a fragility measure associated with the baseline probability \( \mathbb{P} \in \mathcal{P}_0 \) if and only if there exists some normalized convex risk measure \( \mu : \mathcal{L} \mapsto \mathbb{R} \), satisfying

\[
\mu(\tilde{v}) \geq \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{v}] \quad \forall \tilde{v} \in \mathcal{L},
\]

such that

\[
\rho(\tilde{v}) = \inf \{ k > 0 \mid k\mu(\tilde{v})/k \leq 0 \}.
\]

Note that the property associated with the inequality (12) is implied for convex risk measures that are law-invariant under \( \hat{\mathbb{P}} \) (see Follmer and Schied 2004). Proposition 2 implies that apart from its probability-distance-based representation, we can also construct a fragility measure using convex risk measure, which is well studied in the literature (see, e.g., Follmer and Schied 2002, 2004). Some specific examples of fragility measure include the riskiness index of Aumann and Serrano (2008) and the essential riskiness index of Zhang et al. (2019).

Having now defined the fragility measure and presented its representations, we can now formally define a class of robustness optimization models.

**Definition 4.** A *fragility-based* robustness optimization model is an optimization problem that finds a solution for which a set of fragility measures evaluated on the model’s uncertain constraints are Pareto minimal.

With this definition, we note that several different fragility-based robustness optimization models have already been introduced in the literature, including, among others, the Ben-Tal et al. (2017) GRC-sum model for multiple constraints, Chen et al. (2015), where a weighted sum of riskiness indices are minimized, and Zhang et al. (2019), where a sum of essential riskiness indices is minimized. Given the focus of the paper is solely on fragility-based robustness optimization models, unless otherwise stated, all references to robustness optimization in the paper are fragility-based. As a contrast, we mention a different plausible robustness optimization model.
3.1. Maximal-set-based robustness optimization model

Apart from the fragility-based robustness optimization models, there are potentially other types of robustness optimization models that can be motivated for which the decision criteria are not associated with fragility measures. A closely related model is the following one that maximizes the size of the ambiguity set while keeping the feasibility of constraints in check,

$$
\begin{align*}
\text{max} & \quad r \\
\text{s.t.} & \quad \mathbb{E}_P[f(x, \tilde{z})] \leq 0 \quad \forall \tilde{P} \in \mathcal{B}(r) \\
& \quad c^\top x \geq \tau \\
& \quad x \in \mathcal{X}, r \geq 0,
\end{align*}
$$

which we refer to as the maximal-set-based robustness optimization model. The decision criterion is a scale-invariant quasi-concave satisficing measure of Brown and Sim (2009) and such robustness optimization models have appeared in Chen and Sim (2009), Lim and Wang (2017), among others.

Similar to robust optimization, this approach does not protect the level of constraint violation whenever uncertainty occurs outside the uncertainty set, though it would eradicate the multiple optimal solutions that we see in Problem (3). The optimal solution of the maximal-set-based robustness optimization model can be obtained by solving a sequence of robust optimization problems until the uncertain constraint would remain feasible for the largest possible uncertainty set. Hence, this model is generally more difficult to optimize. Unlike the robustness optimization, which is based on a convex fragility measure, we cannot easily maximize over the sum of the sizes of different uncertainty sets. For tractability in problems with multiple uncertain constraints, one may restrict to using one uncertainty set, but this would lead to the issues of having multiple solutions when the size of the uncertainty set is limited by one constraint. For example,

$$
\begin{align*}
\text{max} & \quad r \\
\text{s.t.} & \quad z^\top x \leq 50 \quad \forall z \in \mathcal{U}(r) \\
& \quad z^\top 1 \leq 50 \quad \forall z \in \mathcal{U}(r) \\
& \quad \sum_{j \in [50]} x_j \geq 50 \\
& \quad x \in [0, 1]^{100},
\end{align*}
$$

where $\mathcal{U}(r) = \{z \in [-1, 1]^{100} \mid ||z||_1 \leq r\}$, would yield an objective of $r^* = 50$, with issues of multiple optimal solutions that we see in Problem (3). Hence, lexicographic maximization would be essential to ensure Pareto optimality for multiple constraint, which would increase the computational burden of the problem. In contrast, minimizing the sum of two fragility measures, each for an uncertain constraint, would resolve the issue tractably.
4. Robustness Optimization for Prescriptive Analytics

We present a suite of useful and practicable robustness optimization models for prescriptive analytics that are ubiquitous in management decision making, including linear, adaptive linear, data-driven adaptive linear, combinatorial and dynamic optimization problems. We focus on comparing the formulations and computational complexities with the corresponding robust optimization models. In these models, we conclude that robustness optimization models are at least as easy to solve than the corresponding robust optimization models.

4.1. Linear optimization problem

We start from a linear optimization problem, which is arguably the most important optimization tool for management decision making. We consider the following baseline linear optimization problem in the context of production planning,

\[
Z_0 = \max e^T x \\
\text{s.t. } a_i(z)^T x \leq b_i(z) \quad \forall i \in [I] \\
x \geq 0,
\]

(14)

to determine the optimal profit maximizing production decision \(x\). Specifically, the uncertain parameter \(a_{ij}(\hat{z})\) denotes the number of units of \(i\)th resource needed to produce a unit of the \(j\)th item and \(b_i(\hat{z})\) represents the quantity of the \(i\)th resource available for production. These parameters affinely depend on some exogenous uncertainty \(z \in \mathcal{Z}, \mathcal{Z} = [-1, 1]^N:\)

\[
a_i(z) = a_{i,0} + \sum_{n \in [N]} a_{i,n} z_n \\
b_i(z) = b_{i,0} + \sum_{n \in [N]} b_{i,n} z_n \quad \forall i \in [I],
\]

with baseline parameter \(\hat{z} = 0\) used in Problem (14).

Bertsimas and Sim (2004) propose the following robust optimization model:

\[
\max c^T x \\
\text{s.t. } a_i(z)^T x \leq b_i(z) \quad \forall i \in [I], \forall z \in \mathcal{U}(r_i) \\
x \geq 0
\]

(15)

where \(\mathcal{U}(r_i) = \{z \in \mathcal{Z} \mid \|z\|_1 \leq r_i\}, i \in [I]\) are the budgeted uncertainty sets. Under an assumed ambiguity set for the distribution of \(\hat{z}\), the parameter \(r_i\) can be related to the probability of feasibility of the \(i\)th resource constraint, which allows the modeler to weigh in and specify \(r_i\).
for individual constraints. Problem (15) can be reformulated as the following linear optimization problem:

$$\begin{align*}
\text{max } & \mathbf{c}^\top \mathbf{x} \\
\text{s.t. } & \mathbf{a}^\top_{i,0} \mathbf{x} + \sum_{n \in [N]} s_{i,n} + r_i \kappa_i \leq b_{i,0} \quad \forall i \in [I] \\
& \kappa_i + s_{i,n} \geq \mathbf{a}^\top_{i,n} \mathbf{x} - b_{i,n} \quad \forall i \in [I], n \in [N] \\
& \kappa_i + s_{i,n} \geq -\mathbf{a}^\top_{i,n} \mathbf{x} + b_{i,n} \quad \forall i \in [I], n \in [N] \\
& \mathbf{k} \geq 0, \quad \mathbf{s} \geq 0 \\
& \mathbf{x} \geq 0,
\end{align*}$$

which is computationally attractive and immensely popular.

The robustness linear optimization problem analogous the Bertsimas and Sim (2004) robust optimization model is as follows

$$\begin{align*}
\text{min } & \mathbf{w}^\top \mathbf{k} \\
\text{s.t. } & \mathbf{a}_i(z)^\top \mathbf{x} - b_i(z) \leq \kappa_i \|z\|_1 \quad \forall i \in [I], \forall z \in \mathcal{Z} \\
& \mathbf{c}^\top \mathbf{x} \geq \tau \\
& \mathbf{x} \geq 0, \quad \mathbf{k} \geq 0,
\end{align*}$$

(16)

where \(w_i, \quad i \in [I]\) are weights to reflect the relative importance of individual constraints. We can choose \(w_i\) to be the cost needed to replenish one unit of the \(i\)th the resource, which is more natural to specify than the size parameters in robust optimization approach, to reflect the relative importance of different resource constraints. By reformulating the robust counterpart to its dual formulation, Problem (16) admits the following linear formulation:

$$\begin{align*}
\text{min } & \mathbf{w}^\top \mathbf{k} \\
\text{s.t. } & \mathbf{a}^\top_{i,0} \mathbf{x} + \sum_{n \in [N]} s_{i,n} \leq b_{i,0} \quad \forall i \in [I] \\
& \kappa_i + s_{i,n} \geq \mathbf{a}^\top_{i,n} \mathbf{x} - b_{i,n} \quad \forall i \in [I], n \in [N] \\
& \kappa_i + s_{i,n} \geq -\mathbf{a}^\top_{i,n} \mathbf{x} + b_{i,n} \quad \forall i \in [I], n \in [N] \\
& \mathbf{k} \geq 0, \quad \mathbf{s} \geq 0 \\
& \mathbf{c}^\top \mathbf{x} \geq \tau \\
& \mathbf{x} \geq 0,
\end{align*}$$

which is computationally as attractive the Bertsimas and Sim (2004) robust optimization model. The robustness optimization model is also less prone to the issues of having multiple optimal solutions because the objective seeks to minimize a weighted sum of the fragility measures associated with each of the resource constraint.
4.2. Adaptive linear optimization problem

Adaptive linear optimization, where recourse decisions are delayed after the realization of uncertainty, is very common in problems arising in operations management including, inter alia, deciding inventories at various locations to meet uncertain demands in a distribution network. We focus on a two-stage adaptive linear optimization model. In the first stage, we set the values of here-and-now variables \( x \in \mathcal{X} \) before the realization of the uncertain parameters \( z \in \mathcal{Z} \subseteq \mathbb{R}^N \), where the support set \( \mathcal{Z} \) contains \( 0 \) in its interior. The classical adaptive robust linear optimization model can be written as

\[
Z_r = \min \sup_{z \in \mathcal{U}(r)} \{ c(z)^T x + Q(x, z) \} \tag{17}
\]

subject to \( x \in \mathcal{X} \), where \( \mathcal{U}(r) := \{ z \in \mathcal{Z} \mid \|z\|_1 \leq r \} \) is the budgeted uncertainty set,

\[
Q(x, z) = \min d^Ty
\]

subject to \( A(z)x + By \geq b(z) \)

\[
y \in \mathbb{R}^P,
\]

is the second stage objective function with

\[
A(z) := A_0 + \sum_{i \in [N]} A_i z_i, \quad b(z) := b_0 + \sum_{i \in [N]} b_i z_i, \quad c(z) := c_0 + \sum_{i \in [N]} c_i z_i,
\]

being affine mappings of \( z \). The goal is to determine the optimal here-and-now decision \( x \in \mathcal{X} \), and after the realization of uncertain parameters, \( z \), the optimal continuous wait-and-see decisions \( y \in \mathbb{R}^P \) is determined by solving Problem (18). We can express Problem (17) as

\[
Z_r = \min \sup_{z \in \mathcal{U}(r)} \{ c(z)^T x + d^Ty(z) \}
\]

subject to \( A(z)x + By(z) \geq b(z) \) for all \( z \in \mathcal{U}(r) \)

\[
y \in \mathbb{R}^{N,P},
\]

where the family of recourse functions is defined as

\[
\mathcal{R}^{N,P} := \{ y \mid y(z) : \mathbb{R}^N \mapsto \mathbb{R}^P, y \text{ is a measurable function} \}.
\]

Note that apart from \( Z_0 \), which solves a linear optimization problem, Problem (19) is generally intractable because the recourse function \( y(z) \) is unrestricted and akin to having infinite number of decision variables. A popular method to tractably solve Problem (19) is to use affine recourse adaptation, where the recourse function \( y(z) \) is restricted to an affine function of \( z \) (see, e.g., Ben-Tal et al. 2004) as follows:

\[
\mathcal{L}^{N,P} := \left\{ y \in \mathcal{R}^{N,P} \mid y(z) = y_0 + \sum_{i \in N} y_i z_i \text{ for some } y_i \in \mathbb{R}^P, i \in [N] \cup \{0\} \right\}.
\]
Borrowing the terminology from stochastic programming, the second stage optimization problem is said to have \textit{relatively complete recourse} if for any \( x \in \mathcal{X} \) and \( z \in \mathcal{Z} \), there exists some \( y \in \mathbb{R}^p \) such that \( A(z)x + By \geq b(z) \). \textit{Complete recourse} is an important class that is associated with the recourse matrix \( B \). Specifically, for any right hand side vector, \( t \in \mathbb{R}^M \), there exists a feasible recourse \( w \in \mathbb{R}^p \) such that \( Bw \geq t \).

Given a fixed budget \( \tau \geq Z_0 \), we present the adaptive linear robustness optimization model as follows:

\[
\begin{align*}
\min \ k \\
\text{s.t.} \quad & c(z)^T x + Q(x, z) - \tau \leq k\|z\|_1 \quad \forall z \in \mathcal{Z} \\
& x \in \mathcal{X}, k \geq 0,
\end{align*}
\]

or equivalently

\[
\begin{align*}
\min \ k \\
\text{s.t.} \quad & c(z)^T x + d^T y(z) - \tau \leq k\|z\|_1 \quad \forall z \in \mathcal{Z} \\
& A(z)x + By(z) \geq b(z) \quad \forall z \in \mathcal{Z} \\
& y \in \mathcal{R}^{N,P} \\
& x \in \mathcal{X}, k \geq 0,
\end{align*}
\]

(20)

which is also generally an intractable problem. The following example illustrates that the standard affine recourse adaptation may not necessarily guarantee feasibility, even if the problem has complete recourse:

\[
\begin{align*}
\min \ k \\
\text{s.t.} \quad & y(z) - 0 \leq k|z| \quad \forall z \in \mathbb{R} \\
& y(z) \geq z \quad \forall z \in \mathbb{R} \\
& y(z) \geq -z \quad \forall z \in \mathbb{R} \\
& y \in \mathcal{R}^{1,1}, k \geq 0.
\end{align*}
\]

(21)

Observe that the optimal \( k^* = 1 \) and recourse function \( y^*(z) = |z| \). If we restrict to affine recourse adaptation so that \( y(z) = y_0 + y_1 z \) for some \( y_0, y_1 \in \mathbb{R} \), then the semi-infinite constraints, \( y_0 + y_1 z \geq z, \forall z \in \mathbb{R} \) implies \( y_1 = 1 \), while \( y_0 + y_1 z \geq -z, \forall z \in \mathbb{R} \) would imply \( y_1 = -1 \), which is infeasible. To resolve this issue, we will subsequently introduce the lifted affine recourse adaptation. We first present an alternate formulation of Problem (20) based on a lifted support set, so that the constraints, apart from the recourse functions, are affine functions of the uncertain parameters.

**Proposition 3.** Problem (20) can be equivalently written as:

\[
\begin{align*}
\min \ k \\
\text{s.t.} \quad & c(z)^T x + d^T y(z, u) - \tau \leq ku \quad \forall (z, u) \in \tilde{Z} \\
& A(z)x + By(z, u) \geq b(z) \quad \forall (z, u) \in \tilde{Z} \\
& y \in \mathcal{R}^{N+1,P} \\
& x \in \mathcal{X}, k \geq 0,
\end{align*}
\]

(22)
where $\tilde{Z}$ is a lifted support set defined as:

$$\tilde{Z} = \{(z, u) \mid z \in Z, \|z\|_1 \leq u\}.$$

Hence, the lifted affine recourse adaptation of Problem (20) can be written explicitly as the following robust optimization problem:

$$\min k$$

$$\text{s.t. } c(z)^T x + d^T y_0 + \sum_{i \in [N]} d^T y_i z_i + d^T y_{N+1} u - \tau \leq ku \quad \forall (z, u) \in \tilde{Z}$$

$$A(z) x + B y_0 + \sum_{i \in [N]} B y_i z_i + B y_{N+1} u \geq b(z) \quad \forall (z, u) \in \tilde{Z}$$

$$y_i \in \mathbb{R}^P$$

$$x \in \mathcal{X}, k \geq 0.$$

**Theorem 2.** Suppose the second stage problem (18) has complete recourse, then for any given $\tau \geq Z_0$, the lifted affine recourse adaptation is feasible in Problem (23) and the objective is finite. Moreover, for $P = 1$, the lifted affine recourse adaptation would yield the optimal solution in Problem (22).

This result is quite surprising and important because, despite the adaptive model being a difficult problem to solve exactly, under complete recourse, the lifted affine recourse adaptation approximation does not limit the choice of targets $\tau \geq Z_0$ for the decision maker. Regarding Problem (21), we note that based on Theorem 2, the lifted affine recourse adaptation with $y(z, u) = u$ would yield the optimal solution. Incidentally, the assumption of complete recourse is necessary here, and we provide a counter example in Appendix B. We also remark that the lifted affine recourse adaptation can be applied universally to improve the solution of the adaptive robust linear optimization of Problem (19). Other more computationally intensive approaches can be used to improve the recourse adaptation include piecewise affine recourse adaptation techniques (see, e.g., Goh and Sim 2010, Chen et al. 2008) and Fourier-Motzkin elimination of recourse variables introduced by Zhen et al. (2018).

**4.3. Data-driven adaptive linear optimization problem**

As a generalization of the adaptive linear optimization problem, we now consider the data-driven linear optimization model. The baseline data-driven linear optimization problem is given by

$$Z_0 = \min \mathbb{E}_p \left[ c(\tilde{z})^T x + Q(x, \tilde{z}) \right]$$

$$\text{s.t. } x \in \mathcal{X},$$
where \( \hat{P} \in \mathcal{P}_0(\mathcal{Z}) \) is an empirical baseline distribution with \( \hat{P}[\hat{z} = \hat{z}_s] = 1/S \) and \( \hat{z}_s \) representing a realized data record under scenario \( s \in [S] \). Equivalently, we have

\[
Z_0 = \min \frac{1}{S} \sum_{s \in [S]} (c(\hat{z}_s)^\top x + d^\top y_s)
\]

\[
s.t. \quad A(\hat{z}_s)x + By_s \geq b(\hat{z}_s) \quad \forall s \in [S]
\]

\[
y_s \in \mathbb{R}^p \quad \forall s \in [S]
\]

\[
x \in \mathcal{X}.
\]

To avoid having poor solutions due to overfitting to the empirical distribution, the data-driven adaptive distributionally robust optimization model has recently been proposed by Gao and Kleywegt (2016), Gao et al. (2017), Mohajerin Esfahani and Kuhn (2018) among others, for which the ambiguity set is characterized by a family of distributions whose Wasserstein distances from the empirical distribution are limited by a specified parameter \( r \). To obtain highly tractable models, we focus on the \( \ell_1 \)-norm Wasserstein metric (of type-1) as follows:

\[
W(P, \hat{P}) := \inf_{Q \in \mathcal{P}_0(\mathcal{Z} \times \mathcal{Z})} \left\{ \mathbb{E}_Q[\|\hat{z} - \hat{v}\|_1] \mid (\hat{z}, \hat{v}) \sim Q, \hat{z} \sim P, \hat{v} \sim \hat{P} \right\}.
\]

Using the same model as the adaptive robust linear optimization problems, the data-driven adaptive distributionally robust optimization model solves the following problem:

\[
Z_r = \min_{\mathcal{P} \in B(r)} \sup_{P \in \mathcal{P}(r)} \mathbb{E}_P [c(\hat{z})^\top x + Q(x, \hat{z})]
\]

\[
s.t. \quad x \in \mathcal{X},
\]

for a given Wasserstein-based ambiguity set,

\[
B(r) := \left\{ \mathcal{P} \in \mathcal{P}_0(\mathcal{Z}) \mid \hat{z} \sim \mathcal{P}, W(\mathcal{P}, \hat{P}) \leq r \right\}.
\]

Problem (25) is generally intractable for \( r > 0 \) and Chen et al. (2020) propose a tractable scenario-wise lifted affine recourse adaptation to solve the problem approximately that performs almost as well as the exact model for a multi-item newsvendor problem. We propose the data-driven adaptive linear robustness optimization model as follows:

\[
\min k
\]

\[
s.t. \quad \mathbb{E}_P [c(\hat{z})^\top x + Q(x, \hat{z})] - \tau \leq kW(\mathcal{P}, \hat{P}) \quad \forall \mathcal{P} \in \mathcal{P}_0(\mathcal{Z})
\]

\[
x \in \mathcal{X}, \quad k \geq 0,
\]

for given \( \tau \geq Z_0 \).
**Theorem 3.** The data-driven adaptive linear robustness optimization can be equivalently written as:

\[
\begin{align*}
\min \; & k \\
\text{s.t.} \; & \frac{1}{S} \sum_{s \in [S]} \sup_{(z,u) \in \tilde{Z}_s} \left\{ c(z)^T x + d^T y_s(z,u) - ku \right\} \leq \tau \\
& A(z)x + B y_s(z,u) \geq b(z) \quad \forall (z,u) \in \tilde{Z}_s, s \in [S] \\
& y_s \in \mathbb{R}^{N+1,P} \quad \forall s \in [S] \\
& x \in X, \; k \geq 0,
\end{align*}
\]

(26)

where the lifted support set associated with each empirical scenario \(s \in [S]\) is defined as

\[
\tilde{Z}_s := \{(z,u) \in \mathbb{Z} \times \mathbb{R} \mid \|z - \hat{z}_s\|_1 \leq u\}.
\]

(27)

**Corollary 1.** The adaptive linear robustness optimization of Problem (22) is a special case of the data-driven adaptive linear robustness optimization when \(S = 1\) and \(\hat{z}_1 = 0\).

This connection is interesting and worth noting because, in general, the adaptive robust linear optimization is not a special case of the data-driven adaptive robust linear optimization model.

Similar to the adaptive distributionally robust optimization, Problem (26) is generally intractable. To obtain the here-and-now solution \(x \in X\) approximately, we consider a scenario-wise lifted affine recourse adaptation of Problem (26) as follows:

\[
\begin{align*}
\min \; & k \\
\text{s.t.} \; & \frac{1}{S} \sum_{s \in [S]} \sup_{(z,u) \in \tilde{Z}_s} \left\{ c(z)^T x + d^T \left( y_{s,0} + \sum_{i \in [N]} y_{s,i} z_i + y_{s,N+1} u \right) - ku \right\} \leq \tau \\
& A(z)x + B \left( y_{s,0} + \sum_{i \in [N]} y_{s,i} z_i + y_{s,N+1} u \right) \geq b(z) \quad \forall (z,u) \in \tilde{Z}_s, s \in [S] \\
& y_{s,i} \in \mathbb{R}^P \quad \forall i \in [N+1] \cup \{0\}, s \in [S] \\
& x \in X, \; k \geq 0.
\end{align*}
\]

(28)

**Theorem 4.** Suppose the second stage problem (18) has complete recourse, then for any given \(\tau \geq Z_0\), the scenario-wise lifted affine recourse adaptation is feasible and the objective is finite. Moreover, when \(P = 1\), the scenario-wise lifted affine recourse adaptation would yield the optimal solution in Problem (26).

**A portfolio selection problem.** As an application, we consider a portfolio selection problem under a data-driven setting, which is of practical interests in financial management, and derive an explicit robustness optimization formulation. For the baseline problem, we consider a risk averse investor investing in \(N\) risky assets using the empirical distribution \(\hat{P}\) constructed from historical
data \( (\hat{z}_1, \ldots, \hat{z}_s) \) on the returns of these risky assets. The investor maximizes her expected utility based on the empirical distribution in the following optimization problem,

\[
Z_0 = \max \mathbb{E}_\hat{p} \left[ u(\mathbf{x}^\top \hat{z}) \right] \\
\text{s.t. } \mathbf{1}^\top \mathbf{x} = 1 \\
\mathbf{x} \in \mathbb{R}^N,
\]

for a concave piecewise affine increasing utility, \( u(v) = \min\{a_i v + b_i\} \), with \( a_i \geq 0 \), which can be formulated as a data-driven adaptive linear optimization with complete recourse that is one-dimensional. Assuming an unbounded support, \( \mathcal{Z} = \mathbb{R}^N \), we formulate the corresponding robustness portfolio optimization problem for a given target utility \( \tau \leq Z_0 \) as follows:

\[
\begin{align*}
\min k \\
\text{s.t. } \mathbb{E}_p \left[ \tau - u(\mathbf{x}^\top \hat{z}) \right] &\leq k W(p, \hat{p}) \\
\mathbf{1}^\top \mathbf{x} &= 1 \\
\mathbf{x} &\in \mathbb{R}^N, k \geq 0.
\end{align*}
\]

(29)

**Theorem 5.** The optimal portfolio in Problem (29) can be obtained by solving the following optimization problem:

\[
\begin{align*}
\min &\quad \|\mathbf{x}\|_\infty \\
\text{s.t. } &\mathbb{E}_p \left[ u(\mathbf{x}^\top \hat{z}) \right] \geq \tau \\
\mathbf{x}^\top \mathbf{1} &= 1 \\
\mathbf{x} &\in \mathbb{R}^N.
\end{align*}
\]

(30)

At the high level, the robustness optimization approach improves the robustness by diversifying as much as possible, while trying to attain the target level of expected utility, \( \tau \) over the empirical distribution. It diversifies the portfolio more aggressively than the classical mean-variance portfolio selection of (Markowitz 1952).

**Corollary 2.** When \( \tau \leq \mathbb{E}_p [u(\mathbf{1}^\top \hat{z}/N)] \), the optimal portfolio is the equal-weighted \( 1/N \) portfolio.

Hence, whenever the target utility of the portfolio return \( \tau \) is less than the empirical utility of the equal weighted portfolio, then the optimum portfolio would be the equal-weighted portfolio. As the target utility increases, the decision maker commits to less ambiguity averse portfolio and diversify less in order to achieve a more ambitious target utility.

Despite the simplicity, it has been well-known from the empirical study of DeMiguel et al. (2009) that equal-weighted portfolio outperforms many risk models in practice. In particular, when they compare the equal-weighted portfolio against 14 models across seven empirical datasets, none of the model is consistently better than the equal-weighted portfolio in out-of-sample performance. When
there are 25 assets, they show that mean-variance strategy and its extensions require an estimation window of over 3000 months in order to outperform the equal-weighted portfolio, alluding to their impracticality in addressing actual portfolio selection problems.

The optimality of equal weighted portfolios has also been established in distributionally robust portfolio optimization models (see, e.g., Pflug et al. 2012, Mohajerin Esfahani and Kuhn 2018) as the Wasserstein ball increases in size within the ambiguity set. In the robustness optimization model, the optimal portfolio depends on the target of the decision maker, which we believe is more interpretable for the decision maker to specify than the radius of the Wasserstein ball.

### 4.4. Combinatorial optimization problem

Combinatorial optimization problems with uncertainty in objectives are applicable in transportation, network design and capital budgeting problems. Following the seminal work of Bertsimas and Sim (2003), we consider the following baseline combinatorial optimization problem:

\[
Z_0 = \min_{\mathbf{x} \in \mathcal{X}} \mathbf{c}(\hat{z})^T \mathbf{x}
\]

where \( \mathcal{X} \subseteq \{0,1\}^N \) and the cost parameter \( \mathbf{c}(z) \) depends on some exogenous parameter \( z \in \mathcal{Z} \), \( \mathcal{Z} = [0,1]^N \). More specifically, for any \( n \in [N] \):

\[
c_n(z) = c_n + d_n z_n,
\]

with \( \mathbf{c}, \mathbf{d} \geq \mathbf{0} \) and baseline parameter \( \hat{z} = \mathbf{0} \). Bertsimas and Sim (2003) present the following robust optimization model:

\[
\min \mathbf{c}^T \mathbf{x} + \max_{\mathbf{z} \in \mathcal{U}(r)} \sum_{n \in [N]} d_n x_n z_n \\
\text{s.t. } \mathbf{x} \in \mathcal{X},
\]

where the budgeted uncertainty set, \( \mathcal{U}(r) \), is defined as:

\[
\mathcal{U}(r) = \{ z \in \mathcal{Z} \mid \|z\|_1 \leq r \}.
\]

They show that the optimal solution can be obtained by solving \( N + 1 \) of the following combinatorial optimization problems with different linear objective functions:

\[
\min_{\mathbf{x} \in \mathcal{X}} \left\{ \mathbf{c}^T \mathbf{x} + \sum_{n \in [N]} (d_n - k)^+ x_n \right\},
\]

for \( k \in \{0,d_1,\ldots,d_N\} \). Hence, the robust model (31) is polynomial-time solvable if the baseline combinatorial problem is also polynomial-time solvable (Bertsimas and Sim 2003, Theorem 3).
We present the robustness optimization model analogous to Bertsimas and Sim (2003) robust optimization model as follows:

$$\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad c^\top x + \sum_{n \in [N]} d_n x_n - \tau \leq k\|z\|_1 \quad \forall z \in Z \\
& \quad x \in \mathcal{X}, k \geq 0,
\end{align*}$$

(32)

for some cost budget $\tau \geq Z_0$.

**Theorem 6.** The combinatorial robustness optimization problem (32) is equivalent to the following combinatorial optimization problem:

$$\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \min_{x \in \mathcal{X}} \left\{ c^\top x + \sum_{n \in [N]} (d_n - k)^+ x_n \right\} \leq \tau \\
& \quad k \geq 0.
\end{align*}$$

By performing bisectional search search on $k$ we can obtain the optimal solution up to $\epsilon > 0$ accuracy by solving at most $\lceil \log_2(\|d\|_\infty/\epsilon) \rceil$, combinatorial optimization problems with different linear objective functions.

Similar to Bertsimas and Sim (2003) robust optimization model, Theorem 6 implies that the robustness optimization model (32) is polynomial-time solvable if the baseline combinatorial problem is also polynomial-time solvable, though the number of baseline problems to solve does not grow with $N$. As a simple comparison, consider a shortest path problem with $N = 1000$ arcs and the largest deviation of travel time along an arc is $\tilde{d} = 10$. The robust model requires solving $N + 1 = 1001$ baseline combinatorial optimization problems, while the robustness model only requires solving 20 baseline problems to achieve an accuracy of $\epsilon = 10^{-5}$.

### 4.5. Dynamic optimization problem

Dynamic optimization is often used in operations management to elicit the optimal policies in a dynamic decision making environment where information is revealed over time. Robust dynamic optimization problems have been well studied in the literature (see, e.g., Hansen and Sargent 2001, Iyengar 2005, Lim and Shanthikumar 2007, Xu and Mannor 2012, Wiesemann et al. 2013). Here we revisit this from the perspective of robustness optimization with the aim of preserving the same policy structure and computational complexity as the baseline model. We focus on the discrete-time setting and consider the dynamic system

$$s_{t+1} = f_t(s_t, a_t, z_t) \quad \forall t \in [T],$$
where \( s_t \in S \subseteq \mathbb{R}^{N_s}, a_t \in \mathbb{R}^{N_a}, z_t \in \mathbb{R}^{N_z} \) are the state of the system, the control, and the realization of uncertainty \( z_t \) in period \( t \), respectively, for all \( t \in [T] \). The control \( a_t \) is constrained to be in a set \( A_t(s_t) \) which is specified by the system state \( s_t \). For tractability, we assume that the baseline distribution \( \hat{P} \) is discrete and \( z_1, \ldots, z_T \) are independently distributed. We denote \( h_t = (s_j, a_j, z_j)_{j \in \{t - 1\}} \in \mathcal{H}_t \) as the information history at time \( t \) before \( s_t \) is realized. Given an initial state \( s_1 \), we want to find a policy \( \pi = \{x_t, 1, \ldots, x_T, \} \), where \( x_t : \mathcal{H}_t \times S_t \to \mathbb{R}^{N_a} \) maps the history at time \( t \), including its current state to a control. Denote \( \Pi \) as the set of all admissible policies, i.e., the set of all \( \pi = \{x_t, 1, \ldots, x_T, \} \) with \( x_t(h_t, s_t) \in A_t(s_t) \) for all \( h_t \in \mathcal{H}_t, s_t \in S_t, t \in [T] \). The problem of minimizing expected loss can be formulated as

\[
Z_0 = \min_{\pi \in \Pi} \mathbb{E}_{\mathbb{P}} \left[ \sum_{t \in [T]} l_t(s_t, x_t(h_t, s_t), z_t) \right],
\]

where the function \( l_t \) maps to the loss in period \( t \).

This problem is usually solved via Bellman equations. Specifically, there exists an optimal policy \( \pi^* = \{x^*_1, \ldots, x^*_T, \} \), \( x_t : S_t \to \mathbb{R}^{N_a} \) that depends only on \( s_t \) at time \( t \), and it is determined recursively by backward induction,

\[
x^*_t(s_t) = \arg \min_{a_t \in A_t(s_t)} \mathbb{E}_{\mathbb{P}} \left[ l_t(s_t, a_t, z_t) + J_{t+1}(f_t(s_t, a_t, z_t)) \right] \quad \forall s_t \in S_t, t \in [T],
\]

where

\[
J_t(s_t) := \begin{cases} 
\min_{a_t \in A_t(s_t)} \mathbb{E}_{\mathbb{P}} \left[ l_t(s_t, a_t, z_t) + J_{t+1}(f_t(s_t, a_t, z_t)) \right] & \text{if } t \in [T] \\
0 & \text{if } t = T + 1.
\end{cases}
\]

It is important to highlight the caveat that to obtain an optimal policy for the dynamic optimization problem, which depends only on the current state without the historical information baggage, the assumption of independence is critical. In a data-driven setting, the assumption of \( z_1, \ldots, z_T \) being independently distributed is unrealistic. Specifically for \( \hat{P} \) to represent an empirical distribution for which the independence assumption holds, one would require the data size to be \( K^T \), where \( K \) is the cardinality of \( \{z_t(\omega) \mid \omega \in \Omega \} \).

From the distributionally robust optimization perspective, an ambiguity set can be used to avoid overfitting to the distribution \( \hat{P} \). However, the challenge is to find a suitable one for which the robust dynamic optimization problem would also retain the state-dependent policy as the baseline problem. To achieve this, we present the ambiguity set constructed based on the Kullback-Leibler divergence, \( D_{KL} \), a particular \( \phi \)-divergence (e.g., Pardo 2006, Liese and Vajda 2006) and defined as:

\[
D_{KL}(\mathbb{P}||\hat{\mathbb{P}}) := \begin{cases} 
\mathbb{E}_{\mathbb{P}} \left[ \log \left( \frac{d\mathbb{P}}{d\hat{\mathbb{P}}} \right) \right] & \text{if } \mathbb{P} \ll \hat{\mathbb{P}} \\
\infty & \text{otherwise}.
\end{cases}
\]
In this case, the distributionally robust formulation is given by

$$Z_r = \min_{\pi \in \Pi} \sup_{P \in \mathcal{B}_{KL}(r)} \mathbb{E}_{P} \left[ \sum_{t \in [T]} l_t(s_t, x_t(h_t, s_t), \tilde{z}_t) \right],$$

(35)

where $\mathcal{B}_{KL}(r) = \left\{ P \in \mathcal{P}_0 \mid \tilde{z} \sim P, D_{KL}(P, \hat{P}) \leq r \right\}$. The worst-case expectation associated with the ambiguity set $\mathcal{B}_{KL}(r)$ has been established by Ahmadi-Javid (2012), who shows for any given $\tilde{v} \in \mathcal{L}$,

$$\sup_{P \in \mathcal{B}_{KL}(r)} \mathbb{E}_P[\tilde{v}] = \inf_{k > 0} \left\{ k \log \mathbb{E}_\hat{P} \left[ \exp \left( \frac{\tilde{v}}{k} \right) \right] + kr \right\}.$$  

(36)

We then have the following results.

**Theorem 7.** For a given initial state $s_1$, the optimal policy of the dynamic distributionally robust optimization problem (35), $\pi^* = \{x_1^*, \ldots, x_T^*\}$, $x_t : \mathcal{S}_t \to \mathbb{R}^{N_a}$ is determined recursively by backward induction,

$$x_t^*(s_t) = \arg \min_{a_t \in \mathcal{A}_t(s_t)} k^* \log \mathbb{E}_\hat{P} \left[ \exp \left( \frac{l_t(s_t, a_t, \tilde{z}_t) + G_{t+1}(k^*, f_t(s_t, a_t, \tilde{z}_t))}{k^*} \right) \right] \quad \forall s_t \in \mathcal{S}_t, \ t \in [T],$$

(37)

where

$$G_t(k, s_t) := \begin{cases} \min_{a_t \in \mathcal{A}_t(s_t)} k \log \mathbb{E}_\hat{P} \left[ \exp \left( \frac{l_t(s_t, a_t, \tilde{z}_t) + G_{t+1}(k, f_t(s_t, a_t, \tilde{z}_t))}{k} \right) \right] & \text{if } t \in [T] \\ 0 & \text{if } t = T + 1, \end{cases}$$

(38)

and

$$k^* = \arg \inf_{k > 0} \{ G_1(k, s_1) + kr \}.$$  

(39)

Hence, similar to the baseline problem (33), there exists an optimum policy for which the optimal decision in period $t$ depends solely on the state of the system, $s_t \in \mathcal{S}_t$, $t \in [T]$. Nevertheless, we generally do not know how to optimize over $k$ in Problem (39) almost as efficiently as the baseline problem, unless, for instance, $G_1(k, s_1)$ is a convex function over $k > 0$. We also do not know to specify $r$ in a meaningful way. While the parameter has a relation with Value-at-Risk (Ahmadi-Javid 2012) under distribution $\hat{P}$, the actual distribution would differ from $\hat{P}$.

Interestingly, the dynamic optimization problem under the robustness optimization paradigm can be efficiently solved. In particular, with the distance function being Kullback-Leibler divergence, the dynamic robustness optimization problem is

$$k^* = \min k$$

s.t. $\mathbb{E}_\hat{P} \left[ \sum_{t \in [T]} l_t(s_t, x_t(h_t, s_t), \tilde{z}_t) \right] - \tau \leq kD_{KL}(P, \hat{P}) \quad \forall P \in \mathcal{P}_0,$

$$\pi \in \Pi, \ k \geq 0,$$

(40)
for given $\tau > Z_0$.

The dual presentation of entropic convex risk measure and KL-divergence (Follmer and Schied 2002) states that

$$\sup_{\mathbb{P} \in \mathcal{P}_0} \left\{ \mathbb{E}_\mathbb{P} [\tilde{\nu}] - k D_{KL}(\mathbb{P} \| \tilde{\mathbb{P}}) \right\} = k \log \mathbb{E}_\mathbb{P} [\exp (\tilde{\nu}/k)] ,$$  \hspace{1cm} (41)

where the right-hand-side is a non-increasing function of $k$. Hence, the fragility measure in Problem (40) is closely related to the riskiness index in Aumann and Serrano (2008), and similar measures has been applied to other problems (e.g., Chen et al. 2015, Hall et al. 2015, Zhou et al. 2019).

**Theorem 8.** For a given target $\tau > Z_0$ and initial state $s_1$, the optimal policy of the dynamic robustness optimization problem (40), $\pi^* = \{x_1^*, \ldots, x_T^*\}$, $x_t : S_t \rightarrow R_{+}N_t$ is determined recursively by backward induction (37) and (38) with

$$k^* = \inf \ k \quad \text{s.t.} \quad G_1(k, s_1) \leq \tau \quad \text{and} \quad k > 0 .$$ \hspace{1cm} (42)

As $G_1(k, s_1)$ is non-increasing in $k$, $k^*$ can be determined efficiently by bisection search.

While both the robustness optimization model and the distributionally robust optimization model preserve the system state dependency, the former has a computational advantage over the latter, since we can efficiently optimize $k$ in Problem (42).

This dynamic robustness optimization method is useful in prescriptive analytics to obtain an implementable policy that safeguards against distributional ambiguity as much as possible to achieve a desirable target performance. Theorem 8 implies that due to the dynamic recursion Equation (38), the optimal policy from the robustness optimization model would usually preserve the same structure as the model with given distribution. For example, consider the classical multi-period inventory control and pricing problem. With the recursion (38) and the technique in Chen et al. (2007a), we can show that the optimal policy has the structure of $(s, S, A, p)$, preserving the same structure as the baseline dynamic optimization problem of Chen and Simchi-Levi (2004).

5. **Simulation Studies**

We conduct three simulation studies comparing the empirical solution performances of the robust, the robustness, as well as the GRC-sum optimization models. The first study focuses on a static knapsack optimization problem, while the second study considers an adaptive network lot-sizing problem. In the last study, we extend the adaptive network lot-sizing problem to a data-driven setting and compare the robustness optimization model against the distributionally robust optimization model that uses the Wasserstein-based ambiguity set.
5.1. Revisiting the price of robustness

We consider the following baseline knapsack problem proposed in Bertsimas and Sim (2004),

$$\begin{align*}
\max & \sum_{i \in [N]} c_i x_i \\
\text{s.t.} & \sum_{i \in [N]} w_i(\tilde{z}_i) x_i \leq b \\
& x_i \in \{0, 1\} \quad \forall i \in [N],
\end{align*}$$

where the uncertain weights are affinely dependent on the random variable, $z \in \mathcal{Z}$, $\mathcal{Z} = [-1, 1]^N$:

$$w_i(z_i) = \hat{w}_i + \delta_i z_i \quad \forall i \in [I],$$

with the baseline parameter $\tilde{z} = 0$.

In the experiments, we have $N = 50$, $b = 2000$, each unit profit $c_i$ taken randomly from \{10, 12, 14, 16, 18\}, each baseline weight $\hat{w}_i$ drawn randomly from \{20, 22, 24, \ldots, 80\} and $\delta = 0.2 \hat{w}$. Hence, the uncertain weights will take values in $[0.8 \hat{w}, 1.2 \hat{w}]$, though the decision maker does not know the actual distribution. We compare the robustness optimization (RnO) model to the robust optimization (RO) model of Bertsimas and Sim (2004). Specifically, we vary the budget of uncertainty $r$ and for each $r$, we solve the corresponding robust optimization problem to obtain the optimal robust solution at the profit of $Z_r$. Subsequently, we obtain the solutions to the robustness optimization with the target profit at $\tau = Z_r$. Hence, all comparisons of solutions between the two approaches are made on the same attainable profit. In this instance, the baseline profit is $Z_0 = 592$, and the worst-case profit is $Z_{\infty} = 534$.

In the first simulation experiment, we assume that $\tilde{z}$ is uniformly distributed on the support $\mathcal{Z}$, and we generate 10,000 samples beforehand to compare the average performance of the two models. We plot in Figure 1(a) the probability of infeasibility of the robust optimization as the solutions vary with $r$. We also compare this with the solutions of the robustness optimization, setting the target profit to be the same as profit attainable by the robust optimization problem. The probability bound plotted in the figure is the theoretical bound for the robust optimization, derived in Bertsimas and Sim (2004). We can see that this theoretical bound is a loose one; therefore, using this bound as a guidance to select parameter $r$ is not practical. In Figure 1(b), we replicate the same plot against the target profit, $\tau$. As we can see, robustness optimization solutions constitute a Pareto efficient frontier. For any profit achieved, the robustness optimization solutions are more robust to capacity violation. Similarly, in Figure 1(c) and 1(d), we compare the average violations of the capacity constraints for the solutions of the robust optimization and the robustness optimization problems, as they vary with $r$ and $\tau = Z_r$, respectively.
In the second simulation experiment, we assume that \( \tilde{z} \) is now uniformly distributed in \([-0.5, 1]^N\) instead of the full support \( Z \). Other setups remain the same and we summarize the comparison in Figure 2. Notice that in this case, the probability bound derived in Bertsimas and Sim (2004) does not hold any more (as plotted in Figure 2(a)); therefore, it is even more difficult to select parameter \( r \) in practice. We have also tried other random instances and using skewed distributions to generate the test data, and we observe similar results. We present additional results for an example using beta distributions in Appendix C.

We observe from these figures that for the same attainable profit, the solutions obtained by robustness optimization would generally dominate those obtained by robust optimization. While we observe monotonicity in the performance of solutions as we vary \( \tau \) in the robustness optimization model, this is not the case as we vary \( r \) in the robust optimization problem. We also observe that for a modest reduction of profit from the baseline knapsack problem, we can significantly improve the robustness of the solutions. We note that it is generally difficult to determine \( r \) to match the desired performance for the robust optimization model. Even if the assumption of ambiguous distributions is correct, the probability bound of Bertsimas and Sim (2004) may also be too weak to be useful as a guideline for the choice of \( r \). In sum, for the same price of robustness, robustness optimization consistently yields more robust solutions than those obtained by robust optimization without incurring more computational efforts.

We have also tried GRC-sum model in this example, which gives similar solution performances as the robustness optimization solutions for which \( r = 0 \). We note that the inferiority of the robust optimization model for the knapsack problem could well be due to multiple optimal solutions, and we expect other techniques like the Pareto optimization of Iancu and Trichakis (2014) and the maximal-set-based robustness optimization model (13) to also weed out the inferior dominated solutions.

5.2. An adaptive network lot-sizing problem
We consider a similar network lot-sizing problem as in Bertsimas and de Ruiter (2016). The decision maker sells a single item at \( N \) different stores. We must determine the initial stock allocation \( x_i \in [0, \delta_i] \) at a unit ordering cost \( c_i \), for different stores \( i \in [N] \), prior to the realization of random demands \( \tilde{z}_i \) for \( i \in [N] \). After observing the demands, we can transport stock \( y_{ij} \) from store \( i \) to store \( j \) at a unit transportation cost \( t_{ij} \) to better satisfy demands. We use \( w_i \) to represents the emergency orders at store \( i \), at a unit cost \( l_i > c_i \). We randomly generate the \( N \) stores in a \( 10 \times 10 \) grid. The transportation cost \( t_{ij} \) is set to be proportional to the euclidean distance, \( D_{ij} \), between stores \( i \) and \( j \).
We consider the following benchmarking robust optimization model under the lifted affine recourse adaptation

$$Z_r = \min_{(z,u) \in \mathcal{U}(r)} \max \left\{ c^\top x + \sum_{i \in [N]} d_i^\top y_i(z,u) + l^\top w(z) \right\}$$

s.t. \[ x_i + w_i(z,u) + \sum_{j \in [N]} y_{ij}(z,u) - \sum_{j \in [N]} y_{ij}(z,u) - z_i \geq 0 \quad \forall (z,u) \in \mathcal{U}(r), \forall i \in [N] \]

\[ y(z,u) \geq 0, \quad w(z,u) \geq 0 \quad \forall (z,u) \in \mathcal{U}(r) \]

\[ 0 \leq x \leq \bar{x} \]

\[ y \in \mathcal{L}^{N+1,N \times N}, \quad w \in \mathcal{L}^{N+1,N}, \]
where $\mathcal{U}(r) := \{(z, u) \in \mathcal{Z} \times \mathbb{R} \mid \|z\|_1 \leq u, u \leq r\}$ is the lifted budgeted uncertainty set. Our affine adaptive robustness optimization problem is as follows:

$$\begin{align*}
\min \ k \\
n\text{s.t. } & \ c^T x + \sum_{i \in [N]} d_i^T y_i(z, u) + l^T w(z, u) - \tau \leq ku \quad \forall (z, u) \in \mathcal{Z} \\
& x_i + w_i(z, u) + \sum_{j \in [N]} y_{ji}(z, u) - \sum_{j \in [N]} y_{ij}(z, u) - z_i \geq 0 \quad \forall (z, u) \in \mathcal{Z}, \forall i \in [N] \\
& y(z, u) \geq 0, \ w(z, u) \geq 0 \quad \forall (z, u) \in \mathcal{Z} \\
& 0 \leq x \leq \delta \\
& y \in \mathcal{L}^{N+1,N \times N}, \ w \in \mathcal{L}^{N+1,N}.
\end{align*}$$

Note that with the provision of the emergency orders in the model would make this a complete recourse problem. In this study, we set $N = 20$ and $\mathcal{Z} = [0,20]^N$. For all $i \in [N], j \in [N]$, we set $c_i = 10, \ l_i = 30, \ \delta_i = 20, \ t_{ij} = D_{ij}$. To enforce some correlations in demands, we let the total demand in the network be $D_T = 30\sqrt{N}$. To test the solutions, we generate 10,000 samples as the testing data. In each sample, demand at each stores $i$ is uniformly generated from $[0,20]$ and then
normalized to reflect demand correlation (e.g., total demand in the network is always normalized to $D_r$).

We solve the robust optimization model with different values of $r$ and the robustness optimization model with different values of $\tau$. Both models order more stocks in the first stage as their respective parameters increase. Then, we compare the out-of-sample average total costs and total costs at the 95th percentile of the two models with respect to the first-stage ordering cost. The result is summarized in Figure 3, and the standard deviation of any sample mean is less than one. As we can see, the efficient frontier of the RnO model dominates that of the RO model, i.e., RnO model gives a better initial allocation of stocks. This illustrates that RnO is an effective decision making model to start with. It also implies that for any given solution of the RO model, we can use the RnO model to further improve the performance by adjusting $\tau$ in a binary search so that their first stage ordering costs would coincide.

![Figure 3](image)

**Figure 3** Comparison between RO and RnO: Average cost (left) and cost at the 95th percentile (right).

Next, we compare the robustness optimization with the following GRC-sum-based model:

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \mathbf{c}^\top \mathbf{x} + \sum_{i \in [N]} \mathbf{d}_i^\top \mathbf{y}_i(z,u) + I^\top \mathbf{w}(z,u) - \tau \leq ku & \forall (z,u) \in \mathcal{Z}(r) \\
& \quad x_i + w_i(z,u) + \sum_{j \in [N]} y_{ij}(z,u) - \sum_{j \in [N]} y_{ij}(z,u) - z_i \geq 0 & \forall (z,u) \in \mathcal{Z}(r), \forall i \in [N] \\
& \quad \mathbf{y}(z,u) \geq 0, \quad \mathbf{w}(z,u) \geq 0 & \forall (z,u) \in \mathcal{Z}(r) \\
& \quad 0 \leq \mathbf{x} \leq \mathbf{\bar{x}} \\
& \quad \mathbf{y} \in \mathcal{L}^{N+1,N \times N}, \quad \mathbf{w} \in \mathcal{L}^{N+1,N}, 
\end{align*}
\]
where \( \tilde{Z}(r) \) is the lifted GRC-sum uncertainty set,

\[
\tilde{Z}(r) := \{(z, u) \in Z \times \mathbb{R} \mid \|z\|_1 \leq u + r, u \geq 0\}.
\]

The result is summarized in Figure 4. Similar to Figure 3, the performance comparison is plotted against the first-stage ordering cost to illustrate the quality of first-stage stock allocation. Notice that the robustness optimization is equivalent to having \( r = 0 \). As we can see from Figure 4, the efficient frontier of robustness optimization model \((r = 0)\) dominates those of GRC-sum-based model with \( r = 20, 30, 60 \). We have also tried \( r = 5, 10, 15 \) in this experiment, the solutions coincide with the robustness optimization model. The result is not surprising. The GRC-sum-based model is a hybrid of the robustness optimization and the robust optimization model. As the performance of the robustness optimization model already dominates the corresponding robust optimization model in this experiment, we do not expect the GRC-sum-based model to lead to improvement in solutions. We have tried generating the testing data using other distributions as well as varying \( D_T \) values, and we observe similar results. We present additional results for an example using beta distributions in Appendix C.

Figure 4  Summary of performance: GRC-sum-based model with \( r = 0, 20, 30, 60 \) at different first-stage costs
5.3. A data-driven adaptive network lot-sizing problem

The data-driven network lot-sizing problem, which can be written as the following data-driven adaptive robustness optimization with scenario-wise affine recourse adaptation as follows:

\[
\min k \\
\text{s.t. } c^T x - \tau + \frac{1}{S} \sum_{s \in [S]} v_s \leq 0 \\
\sum_{j \in [N]} d_{ij}^T y_{ij}^{(s)}(z, u) + l^T w^{(s)}(z, u) - ku \leq v_s \quad \forall (z, u) \in \tilde{Z}_s, \; \forall s \in [S] \\
x_i + w_i^{(s)}(z, u) + \sum_{j \in [N]} y_{ij}^{(s)}(z, u) - \sum_{j \in [N]} y_{ij}^{(s)}(z, u) - z_i \geq 0 \quad \forall (z, u) \in \tilde{Z}_s, \; \forall s \in [S], \; \forall i \in [N] \\
y^{(s)}(z, u) \geq 0, \; w^{(s)}(z, u) \geq 0 \quad \forall (z, u) \in \tilde{Z}, \; s \in [S] \\
0 \leq x \leq \delta \\
y^{(s)} \in \mathcal{L}^{N+1,N \times N}, \; w^{(s)} \in \mathcal{L}^{N+1,N} \quad \forall s \in [S],
\]

where the lifted uncertainty set associated with each scenario \( s \in [S] \) is as defined in Equation (27).

The general setting is the same as described in the previous study. In this study, we set \( N = 20 \) and \( Z = [0, 40]^N \). For all \( i \in [N], j \in [N] \), we set \( c_i = 10, \; l_i = 30, \; \delta_i = 40, \; t_{ij} = 2D_{ij} \). We let \( S = 5 \), i.e., the empirical distribution consists of five historical samples. To test the solutions, we generate 10,000 samples as the testing data. In each sample, demands at different locations are independently generated. We use a normal distribution, \( \mathcal{N}(20, 10^2) \), to generate these demands, and the demands are truncated to ensure they lie in the support set \( \tilde{Z} \).

Similar to the previous study, we solve the data-driven distributionally robust optimization (DRO) model with different values of Wasserstein radius \( r \) and the data-driven robustness optimization (DRnO) model with different values of target \( \tau \). Then, we compare the out-of-sample costs with respect to the first-stage ordering cost. We also benchmark the empirical optimization model, e.g., Problem (24), which solves the data-driven optimization problem using the empirical distribution, and we refer to this as the SAA model and use \( Z_{SAA} \) to denote its objective value. We present the performance comparison in Figure 5. The empirical optimization only gives a single solution, and it performs poorly because it overfits to the historical data. In many practical settings, the historical data is not rich enough to depict the true distribution; therefore, the performance of empirical optimization is inferior in these cases. With different input parameters, both DRO and DRnO models can improve cost metrics significantly, especially cost at higher quantiles. Furthermore, DRnO model outperforms DRO model in terms of both average cost and costs at different quantiles. For the average total cost, the standard deviations of sample mean of the robustness optimization solutions under different targets range from one to five, which are small enough to conclude the improvement is statistically significant. We have tried varying the standard deviation.
of the underlying normal distribution and generating demands from some uniform distributions and Poisson distributions, and we observe similar results. We present results for an example where demands are generated from a normal distribution $N(20, 12^2)$ in Appendix C. In Appendix D, we present a sample code on how to model DRO and DRnO in this problem using RSOME (Chen et al. 2020).

Another benefit of robustness optimization is that it is much easier to select a target parameter, as opposed to selecting a radius parameter for the Wasserstein ball. The target is a management-related parameter while the Wasserstein radius is an abstract tuning parameter. We present how the out-of-sample cost metrics changes with respect to the two model parameters in Figure 6. We plot both the average cost and the cost at the 90th percentile. The curve for the 95th percentile is similar to the 90th percentile (see Figure 5) and is therefore omitted. The parameter $r$ is not interpretable operationally, and the out-of-sample performance is sensitive to the choice of this parameter. As we can see, the average cost and 90th percentile change rapidly within a small range of radius $r$. In other words, a small change in $r$ can lead to a very different solution. On the other hand, the change in out-of-sample performance is much more smooth with respect to the change in target in DRnO. Hence, a small change in the target parameter will not lead to a significant change in solution.

In addition, the “nice” range of target $\tau$ is relatively stable with respect to changes in other model parameters, e.g., support set. However, the “nice” range of radius $r$ can change significantly as other parameters change. We provide an illustration in Figure 7, where we expand the support set from $[0, 40]^N$ to $[0, 400]^N$. We plot the out-of-sample cost metrics with respect to model parameters and overlay with the curves in Figure 6 to present the difference. The cost metrics of each instance are normalized by the respective SAA out-of-sample average cost so that they are on a similar scale. Clearly, the robustness optimization approach is far less sensitive to the specification of support set compared to the robust optimization model. As we observe, the “nice” range of radius $r$ changes significantly while the performance with respect to the normalized target $\tau/Z_{SAA}$ is very stable to this change. Hence, in addressing data-driven problems, there are remarkable challenges faced by robust optimization models to determine to right hyper-parameter for the ambiguity set, which can be highly sensitive to how the support set is being specified. In contrast, besides the target parameter being more interpretable and intuitive to specify in management decision problems, it is also far less sensitive to the specification of the support set of the random parameters.
Figure 5  Summary of performance: DRnO and DRO models at different first-stage costs

Figure 6  Out-of-sample cost metrics w.r.t. Wasserstein radius $r$ in DRO (left) and target $\tau$ in DRnO (right).

Figure 7  Change of the “nice” range of radius in DRO (left) and normalized target in DRnO (right).
References


A. Proof of Results

Proof of Proposition 1. According to the Chernoff bound, for all $t > 0$, we have

$$
P \left[ \sum_{i \in [N]} |\tilde{z}_i| \leq (1 - \epsilon)\mu \right] \leq \exp(t(1 - \epsilon)\mu) \prod_{i \in [N]} \mathbb{E}_p[\exp(-t|\tilde{z}_i|)] ,
$$

where

$$
\mu = \sum_{i \in [N]} \mathbb{E}_p[|\tilde{z}_i|] \geq \sum_{i \in [N]} \mathbb{E}_p[\tilde{z}_i^2] \geq N\theta.
$$

Notice that $|\tilde{z}_i|$ has a support set $[0, 1]$. Let $\tilde{v}_i \sim \text{Bernoulli}(p_i)$ be a Bernoulli random variable with the same mean as $|\tilde{z}_i|$. Then

$$
\mathbb{E}_p[\exp(-t|\tilde{z}_i|)] \leq \mathbb{E}_p[\exp(-t\tilde{v}_i)] .
$$

(43)

We first prove the above inequality. For any random variable $\tilde{v}$ with mean $\hat{v}$ and a support set $[0, 1]$, we consider the following distributionally robust optimization problem:

$$
Z(t) = \sup_{P \in \mathcal{Q}} \mathbb{E}_p[\exp(-t\tilde{v})] ,
$$

where the ambiguity set is $\mathcal{Q} := \{P \in \mathcal{P}_0([0, 1]^N) \mid \tilde{v} \sim P, \mathbb{E}_p[\tilde{v}] = \hat{v}\}$. By weak duality, we have $Z(t) \leq Z_d(t)$, where

$$
Z_d(t) = \min r_0 + r_1 \hat{v}
\text{ s.t. } r_0 \geq \exp(-tv) - r_1v \quad \forall v \in [0, 1].
$$

Because $\exp(-tv) - r_1v$ is convex in $v$, the maximum occurs at the boundary $v = 0$ or $v = 1$. Therefore, the above problem is equivalent to

$$
Z_d(t) = \min r_0 + r_1 \hat{v}
\text{ s.t. } r_0 \geq \exp(-t) - r_1
r_0 \geq 1.
$$

This is a linear optimization problem and the optimal value is $Z_d(t) = 1 - \hat{v} + \hat{v}\exp(-t)$, which is the moment generating function of a Bernoulli random variable with mean $\hat{v}$. Therefore, $Z(t) \leq Z_d(t)$ leads to Inequality (43).

By Inequality (43), we can write

$$
P \left[ \sum_{i \in [N]} |\tilde{z}_i| \leq (1 - \epsilon)\mu \right] \leq \inf_{t > 0} \exp(t(1 - \epsilon)\mu) \prod_{i \in [N]} \mathbb{E}_p[\exp(-t|\tilde{z}_i|)]
\leq \inf_{t > 0} \exp(t(1 - \epsilon)\mu) \prod_{i \in [N]} \mathbb{E}_p[\exp(-t\tilde{v}_i)]
\leq \exp \left( -\frac{\mu\epsilon^2}{2} \right) .
$$
where the last inequality follows from the multiplicative Chernoff bound for the sum of independent Bernoulli random variables, which is a known result.

Finally, the proposition follows because

\[
P \left[ \tilde{z} \in \mathcal{U}(c\sqrt{N}) \right] = P \left[ \sum_{i \in [N]} |\tilde{z}_i| \leq c\sqrt{N} \right] \\
= P \left[ \sum_{i \in [N]} |\tilde{z}_i| \leq \left( 1 - \left( 1 - \frac{c\sqrt{N}}{\mu} \right) \right) \mu \right] \\
\leq \exp \left\{ - \frac{\mu \left( 1 - \frac{c\sqrt{N}}{\mu} \right)^2}{2} \right\} \\
\leq \exp \left( - \frac{N\theta \left( 1 - \frac{c\sqrt{N}}{\theta} \right)^2}{2} \right) \\
= \exp \left( - \frac{(\theta\sqrt{N} - c)^2}{2\theta} \right),
\]

where the last inequality follow because \( \mu \geq N\theta \) and \( c \in (0, \theta\sqrt{N}) \). \( \square \)

**Proof of Theorem 1.** We first prove the “only if” direction. Given a fragility measure \( \rho \) defined by Equation (9), we show that it has all the five properties in this theorem and is lower semi-continuous, i.e., \( \{ \tilde{v} \mid \rho(\tilde{v}) \leq a \} \) is a closed set for any \( a \geq 0 \). For convenience, we define

\[
\mathcal{K}(\tilde{v}) := \left\{ k \geq 0 \mid \mathbb{E}_P [\tilde{v}] \leq k \Delta(P, \hat{P}), \ \forall P \in \mathcal{P}_0 \right\},
\]

and hence \( \rho(\tilde{v}) = \inf \mathcal{K}(\tilde{v}) \).

1. **Monotonicity.** If \( \tilde{v}_1 \geq \tilde{v}_2 \), then \( \mathbb{E}_P [\tilde{v}_1] \geq \mathbb{E}_P [\tilde{v}_2] \) for any \( P \in \mathcal{P}_0 \). That is, for any \( k \in \mathcal{K}(\tilde{v}_1) \), we must have \( k \in \mathcal{K}(\tilde{v}_2) \). Therefore, \( \mathcal{K}(\tilde{v}_1) \subseteq \mathcal{K}(\tilde{v}_2) \). Taking the infimum gives \( \rho(\tilde{v}_1) \geq \rho(\tilde{v}_2) \).

2. **Positive homogeneity.** The case of \( \lambda = 0 \) is trivial. Consider any \( \lambda > 0 \). Notice that

\[
\rho(\lambda \tilde{v}) = \inf \left\{ k > 0 \mid \mathbb{E}_P [\lambda \tilde{v}] \leq k \Delta(P, \hat{P}), \ \forall P \in \mathcal{P}_0 \right\} \\
= \inf \left\{ k > 0 \mid \mathbb{E}_P [\tilde{v}] \leq \lambda k \Delta(P, \hat{P}), \ \forall P \in \mathcal{P}_0 \right\} \\
= \lambda \inf \left\{ \beta > 0 \mid \mathbb{E}_P [\tilde{v}] \leq \beta \Delta(P, \hat{P}), \ \forall P \in \mathcal{P}_0 \right\} \\
= \lambda \rho(\tilde{v}).
\]

3. **Subadditivity.** Suppose \( k_1 \in \mathcal{K}(\tilde{v}_1) \) and \( k_2 \in \mathcal{K}(\tilde{v}_2) \). It is not hard to see that

\[
\mathbb{E}_P [\tilde{v}_1 + \tilde{v}_2] \leq (k_1 + k_2) \Delta(P, \hat{P}), \ \forall P \in \mathcal{P}_0,
\]

which indicates \( (k_1 + k_2) \in \mathcal{K}(\tilde{v}_1 + \tilde{v}_2) \). The subadditivity then follows by taking the infimum.
4. **Pro-robustness.** If \( \tilde{v} \leq 0 \), then for all \( \mathbb{P} \in \mathcal{P}_0 \) and \( k > 0 \) we have \( \mathbb{E}_\mathbb{P}[\tilde{v}] \leq 0 \leq k\Delta(\mathbb{P}, \mathbb{P}) \). That implies \( \rho(\tilde{v}) = 0 \).

5. **Anti-fragility.** Suppose \( \mathbb{E}_\mathbb{P}[\tilde{v}] > 0 \), then \( \mathcal{K}(\tilde{v}) = \emptyset \) because \( \Delta(\mathbb{P}, \mathbb{P}) = 0 \). Therefore, \( \rho(\tilde{v}) = \inf \mathcal{K}(\tilde{v}) = \inf \emptyset = \infty \).

The lower semi-continuity of \( \rho \) can be shown as follows. Consider any converging sequence of random variable \( \tilde{v}_1, \ldots, \tilde{v}_n \) such that \( \tilde{v}_n \to \tilde{v} \) as \( n \to +\infty \). For any fixed value \( a \geq 0 \), we need to show that \( \rho(\tilde{v}) \leq a \) if \( \rho(\tilde{v}_n) \leq a \) for all \( n > 0 \). This is true because the expectation is a continuous measure, \( i.e., \lim_{n \to +\infty} \mathbb{E}_\mathbb{P}[\tilde{v}_n] = \mathbb{E}_\mathbb{P}[\tilde{v}] \).

We now prove the “if” direction. Consider any given lower semi-continuous function \( \rho \) which satisfies all the five properties in this theorem. Note that \( \rho \) is also convex, because positive homogeneity and subadditivity imply convexity. We first show that the function \( \Delta \) defined by Equation (11) is a probability distance function. Then, we show \( \inf \mathcal{K}(\tilde{v}) = \rho(\tilde{v}) \), \( i.e., \) the dual representation (11) is valid.

We note that since \( \rho(0) = 0 < 1 \) by Pro-robustness, \( \Delta(\mathbb{P}, \mathbb{P}) \geq \mathbb{E}_\mathbb{P}[0] = 0 \) for all \( \mathbb{P} \in \mathcal{P}_0 \). Moreover, due to the property of Anti-Fragility, \( \rho(\tilde{v}) = \infty > 1 \) for all \( \tilde{v} \) with \( \mathbb{E}_\mathbb{P}[\tilde{v}] > 0 \). Therefore, by Equation (11),

\[
\Delta(\mathbb{P}, \mathbb{P}) = \sup_{\tilde{v} \in \mathcal{L}} \{ \mathbb{E}_\mathbb{P}[\tilde{v}] \mid \rho(\tilde{v}) \leq 1, \mathbb{E}_\mathbb{P}[\tilde{v}] \leq 0 \} \leq 0;
\]

together with \( \Delta(\mathbb{P}, \mathbb{P}) \geq 0 \) for all \( \mathbb{P} \in \mathcal{P}_0 \) we know \( \Delta(\mathbb{P}, \mathbb{P}) = 0 \). Hence, \( \Delta \) is a probability distance function.

Now, it remains to show \( \inf \mathcal{K}(\tilde{v}) = \rho(\tilde{v}) \). With \( \Delta \) defined as in Equation (11), given any \( \tilde{v} \in \mathcal{L} \), we define the set \( \mathcal{K}(\tilde{v}) \) as in Equation (44). To prove the result, we start from the case where \( \rho(\tilde{v}) \in (0, \infty) \).

We first show \( \inf \mathcal{K}(\tilde{v}) \leq \rho(\tilde{v}) \). Consider any \( k > \rho(\tilde{v}) \). By Positive homogeneity, \( \rho(\tilde{v}/k) = \rho(\tilde{v})/k \leq 1 \). Given any \( \mathbb{P} \in \mathcal{P}_0 \), \( \Delta(\mathbb{P}, \mathbb{P}) = \sup_{\tilde{v} \in \mathcal{L}} \{ \mathbb{E}_\mathbb{P}[\tilde{v}] \mid \rho(\tilde{v}) \leq 1 \} \geq \mathbb{E}_\mathbb{P}[\tilde{v}/k] \), which implies \( \mathbb{E}_\mathbb{P}[\tilde{v}] \leq k\Delta(\mathbb{P}, \mathbb{P}) \). Hence, \( k \in \mathcal{K}(\tilde{v}) \). This indicates \( \inf \mathcal{K}(\tilde{v}) \leq \rho(\tilde{v}) \).

We then show \( \inf \mathcal{K}(\tilde{v}) \geq \rho(\tilde{v}) \). Consider any \( 0 < k < \rho(\tilde{v}) \), and hence \( \rho(\tilde{v}/k) = \rho(\tilde{v})/k > 1 \). Denote a set \( \mathcal{W} = \{ \tilde{w} \in \mathcal{L} \mid \rho(\tilde{w}) \leq 1 \} \). Then by convexity and lower semi-continuity of \( \rho \), \( \mathcal{W} \) is a closed convex set and \( \tilde{v}/k \notin \mathcal{W} \). Therefore, by Hahn-Banach separation theorem, there exists a linear functional \( l \) with

\[
\infty > l(\tilde{v}/k) > \beta > l(\tilde{w}), \quad \forall \tilde{w} \in \mathcal{W}
\]

for some \( \beta \in \mathbb{R} \). Consider any \( \tilde{w} \leq \varepsilon \) with a \( \varepsilon < 0 \), then for all \( \lambda > 0 \), \( \lambda\tilde{w} \leq \lambda\varepsilon < 0 \) and hence \( \lambda\tilde{w} \in \mathcal{W} \) by Pro-robustness. Therefore, \( \beta > l(\lambda\tilde{w}) = \lambda l(\tilde{w}) \), where the equality holds since \( l \) is a linear functional. As it is true for all \( \lambda > 0 \), we know \( l(\tilde{w}) \leq 0 \). It further implies that \( l \) is a positive linear functional. WLOG, we can normalize \( l \) such that \( l(1) = 1 \). In this case there exists \( \mathbb{P} \in \mathcal{P}_0 \).
such that \( l(\tilde{w}) = E_P[\tilde{w}] \) for all \( \tilde{w} \in \mathcal{L} \). With this particular \( P \), \( E_P[\tilde{v}/k] > \beta \geq \sup_{\tilde{w} \in \mathcal{W}} E_P[\tilde{w}] = \Delta(P, \tilde{P}) \). This indicates \( E_P[\tilde{v}] > k\Delta(P, \tilde{P}) \) and \( k \not\in K(\tilde{v}) \). Therefore, \( \inf K(\tilde{v}) \geq \rho(\tilde{v}) \).

We hence conclude \( \inf K(\tilde{v}) = \rho(\tilde{v}) \) whenever \( \rho(\tilde{v}) \in (0, \infty) \). For the case of \( \rho(\tilde{v}) = 0 \), we just need the above proof of \( \inf K(\tilde{v}) \leq \rho(\tilde{v}) \) to conclude \( \inf K(\tilde{v}) = 0 \). For the case of \( \rho(\tilde{v}) = \infty \), we just need the above proof of \( \inf K(\tilde{v}) \geq \rho(\tilde{v}) \) to conclude \( \inf K(\tilde{v}) = \infty \). □

**Proof of Proposition 2.** The proof follows largely from existing proofs, see, e.g., Hall et al. (2015). For the purpose of completeness, we include the proof here nevertheless.

First, we prove the “if” direction. Suppose there exists some normalized convex risk measure \( \mu \) as stated in the proposition, and let \( \rho(\tilde{v}) = \inf \{ k > 0 \mid k\mu(\tilde{v}/k) \leq 0 \} \). For convenience, define \( K_\mu(\tilde{v}) := \{ k > 0 \mid k\mu(\tilde{v}/k) \leq 0 \} \). We show that \( \rho \) has the desired properties as follows.

1. **Monotonicity.** If \( \tilde{v}_1 \geq \tilde{v}_2 \), then \( \forall k > 0 \), we have \( \tilde{v}_1/k \geq \tilde{v}_2/k \), \( k\mu(\tilde{v}_1/k) \geq k\mu(\tilde{v}_2/k) \). Hence, for any \( k \in K_\mu(\tilde{v}_1) \), we must have \( k \in K_\mu(\tilde{v}_2) \). Taking the infimum gives us \( \rho(\tilde{v}_1) \geq \rho(\tilde{v}_2) \).

2. **Positive homogeneity.** The case of \( \lambda = 0 \) is trivial and now we consider any \( \lambda > 0 \). \( \rho(\lambda \tilde{v}) = \inf \{ k > 0 \mid k\mu(\lambda \tilde{v}/k) \leq 0 \} = \inf \{ \lambda k > 0 \mid \lambda k\mu(\lambda \tilde{v}/\lambda k) \leq 0 \} = \lambda \inf \{ k > 0 \mid k\mu(\tilde{v}/k) \leq 0 \} = \lambda \rho(\tilde{v}) \).

3. **Subadditivity.** Consider any \( \tilde{v}_1, \tilde{v}_2 \in \mathcal{L} \) and \( k_1 \in K_\mu(\tilde{v}_1), k_2 \in K_\mu(\tilde{v}_2) \). Then, by the convexity of \( \mu \), we have \( (k_1 + k_2)\mu((k_1\tilde{v}_1/k_1 + k_2\tilde{v}_2/k_2)/(k_1 + k_2)) \leq k_1\mu(\tilde{v}_1/k_1) + k_2\mu(\tilde{v}_2/k_2) \leq 0 \). Therefore, we have \( (k_1 + k_2) \in K_\mu(\tilde{v}_1 + \tilde{v}_2) \). Taking the infimum gives us \( \rho(\tilde{v}_1 + \tilde{v}_2) \leq \rho(\tilde{v}_1) + \rho(\tilde{v}_2) \).

4. **Pro-robustness.** Because \( \tilde{v} \leq 0 \), we have \( \mu(\tilde{v}/k) \leq \mu(0) = 0 \) for any \( k > 0 \). Therefore, \( k\mu(\tilde{v}/k) \leq 0 \) for all \( k > 0 \), which indicates \( \rho(\tilde{v}) = 0 \).

5. **Anti-Fragility.** Suppose \( E_P[\tilde{v}] > 0 \) and consider any \( k > 0 \). Then \( E_P[\tilde{v}/k] > 0 \), and hence \( \mu(\tilde{v}/k) \geq E_P[\tilde{v}/k] > 0 \). Therefore, \( K_\mu(\tilde{v}) \) is an empty set. Taking the infimum gives us \( \rho(\tilde{v}) = +\infty \).

The lower semi-continuity of \( \rho \) can be shown as follows. Consider any converging sequence of random variable \( \tilde{v}_1, \ldots, \tilde{v}_n \) such that \( \tilde{v}_n \to \tilde{v} \) as \( n \to +\infty \). For any fixed value \( a \geq 0 \), we need to show that \( \rho(\tilde{v}) \leq a \) if \( \rho(\tilde{v}_n) \leq a \) for all \( n > 0 \). This is true because \( \mu \) is a lower semi-continuous function, i.e., \( \lim_{n \to +\infty} \mu(\tilde{v}_n) = \mu(\tilde{v}) \). Specifically, if \( \rho(\tilde{v}_n) \leq a \) for all \( n > 0 \), then \( \mu(\tilde{v}_n/a) \leq 0 \) for all \( n > 0 \), indicating \( \mu(\tilde{v}/a) \leq 0 \). Therefore, we have \( \rho(\tilde{v}) \leq a \).

We now prove the “only if” direction. Consider any fragility measure \( \rho \). Then we need to prove that there exists a normalized convex risk measure \( \mu \) satisfying \( \mu(\tilde{v}) \geq E_P[\tilde{v}] \) and \( \rho(\tilde{v}) = \inf \{ k > 0 \mid k\mu(\tilde{v}/k) \leq 0 \} \forall \tilde{v} \in \mathcal{L} \). To this end, we consider the \( \mu \) defined as follows,

\[
\mu(\tilde{v}) = \inf \{ a \mid \rho(\tilde{v} - a) \leq 1 \}.
\]

We now show that such \( \mu \) satisfies the requirement. First, we prove that this \( \mu \) is a convex risk measure. Define \( K_\mu(\tilde{v}) := \{ a \mid \rho(\tilde{v} - a) \leq 1 \} \).
1. Monotonicity. For any \( \tilde{v}_1 \geq \tilde{v}_2 \), we have \( \rho(\tilde{v}_1 - a) \geq \rho(\tilde{v}_2 - a) \) for all \( a \in \mathbb{R} \). Then, for any \( a \in K(\tilde{v}_1) \), we must have \( a \in K(\tilde{v}_2) \). Taking the infimum gives us \( \mu(\tilde{v}_1) \geq \mu(\tilde{v}_2) \).

2. Translation invariance. For any \( a' \in \mathbb{R} \), we have \( \mu(\tilde{v} + a') = \inf \{ a \mid \rho(\tilde{v} - (a - a')) \leq 1 \} = a' + \inf \{ a - a' \mid \rho(\tilde{v} - (a - a')) \leq 1 \} = a' + \mu(\tilde{v}) \).

3. Convexity. Consider any \( \tilde{v}_1, \tilde{v}_2 \in \mathcal{L}, \lambda \in [0, 1], a_1 \in K(\tilde{v}_1) \) and \( a_2 \in K(\tilde{v}_2) \). Then, \( \rho(\tilde{v}_1 - a_1) \leq 1 \) and \( \rho(\tilde{v}_2 - a_2) \leq 1 \). Therefore, \( \rho(\lambda \tilde{v}_1 + (1 - \lambda) \tilde{v}_2 - (\lambda a_1 + (1 - \lambda) a_2)) = \rho(\lambda(\tilde{v}_1 - a_1) + (1 - \lambda)(\tilde{v}_2 - a_2)) \leq \lambda \rho(\tilde{v}_1 - a_1) + (1 - \lambda) \rho(\tilde{v}_2 - a_2) \leq 1 \), where the first inequality is due to Subadditivity and Positive homogeneity of \( \rho \). It indicates \( \lambda a_1 + (1 - \lambda) a_2 \in K(\lambda \tilde{v}_1 + (1 - \lambda) \tilde{v}_2) \). Taking the infimum gives us the convexity.

4. Normalization. Because of anti-fragility, we have \( \rho(a') = +\infty \) for all \( a' > 0 \). Therefore, \( \mu(0) = 0 \).

To show \( \mu(\tilde{v}) \geq \mathbb{E}_\rho \bar{v} \), notice that \( \forall a < \mathbb{E}_\rho \bar{v}, \mathbb{E}_\rho [\bar{v} - a] > 0 \) and hence \( \rho(\bar{v} - a) = \infty \), \( a \not\in K(\bar{v}) \). Therefore, \( \mu(\tilde{v}) = \inf K(\bar{v}) \geq \mathbb{E}_\rho \bar{v} \).

To show \( \rho(\tilde{v}) = \inf \{ k > 0 \mid k \mu(\tilde{v} / k) \leq 0 \} \), we observe \( \inf \{ k > 0 \mid k \mu(\tilde{v} / k) \leq 0 \} = \inf \{ k > 0 \mid \exists a \leq 0 : \rho(\bar{v} / k - a) \leq 1 \} = \inf \{ k > 0 \mid \rho(\bar{v} / k) \leq 1 \} = \inf \{ k > 0 \mid \rho(\bar{v}) \leq k \} = \rho(\tilde{v}) \). □

**Proof of Proposition 3.** We will show that for any optimal solution in (20) we can find a feasible solution in (22) that has the same objective value and *vice versa*.

Consider any optimal solution \( \tilde{k}, \tilde{x}, \tilde{y} \) in Problem (20). We can define \( \tilde{y}(z, u) := \tilde{y}(z) \) for all \( (z, u) \in \bar{Z} \). In addition, \( \Pi_z \bar{Z} = \bar{Z} \). Therefore, the following constraints are satisfied in Problem (22):

\[
A(z)\tilde{x} + B\tilde{y}(z, u) \succeq b(z), \quad \forall (z, u) \in \bar{Z},
\]

\[
\tilde{y} \in \mathcal{R}^{N+1,P}.
\]

In addition, we have

\[
\begin{align*}
\mathbf{c}(z)^T \tilde{x} + d^T \tilde{y}(z, u) &- \tau \leq \tilde{k} \|z\|_1, \quad \forall z \in \mathcal{Z},
\end{align*}
\]

which implies

\[
\begin{align*}
\mathbf{c}(z)^T \tilde{x} + d^T \tilde{y}(z, u) &- \tau \leq \tilde{k} u, \quad \forall (z, u) \in \bar{Z}.
\end{align*}
\]

Therefore, \( \tilde{k}, \tilde{x}, \tilde{y} \) are feasible in Problem (22) and lead to the same objective value.

Now, consider any optimal solution \( \tilde{k}, \tilde{x}, \tilde{y} \) in Problem (22). Because \( A(z)\tilde{x} + B\tilde{y}(z, u) \succeq b(z) \) for all \( (z, u) \in \bar{Z} \), we must have \( A(z)\tilde{x} + B\tilde{y}(z, \|z\|_1) \succeq b(z) \) for all \( z \in \mathcal{Z} \). Let us define \( \tilde{y}(z) := \tilde{y}(z, \|z\|_1) \), which satisfies

\[
A(z)\tilde{x} + B\tilde{y}(z) \succeq b(z), \quad \forall z \in \mathcal{Z},
\]

\[
\tilde{y} \in \mathcal{R}^{N+P}.
\]

Similarly, because \( \mathbf{c}(z)^T \tilde{x} + d^T \tilde{y}(z, u) - \tau \leq \tilde{k} u \) for all \( (z, u) \in \bar{Z} \), we must have \( \mathbf{c}(z)^T \tilde{x} + d^T \tilde{y}(z, \|z\|_1) - \tau \leq \tilde{k} \|z\|_1 \) for all \( z \in \mathcal{Z} \). Then, the above defined function \( \tilde{y} \) will also satisfy

\[
\mathbf{c}(z)^T \tilde{x} + d^T \tilde{y}(z) - \tau \leq \tilde{k} \|z\|_1, \quad \forall z \in \mathcal{Z}.
\]
Therefore, \( \bar{k}, \bar{x}, \bar{y} \) are also feasible in Problem (20) and lead to the same objective value.

**Proof of Theorem 2.** We first prove the feasibility under complete recourse. Because \( \tau \geq Z_0 \), there exists some \( \bar{x} \in \mathcal{X} \) and \( \bar{y}_0 \) such that

\[
A_0 \bar{x} + B \bar{y}_0 \geq b_0,
\]

\[
c_0^\top \bar{x} + d^\top \bar{y}_0 - \tau \leq 0.
\]

As the problem has complete recourse, for any given \( \bar{y}_1, \ldots, \bar{y}_N \), there exists there exists some \( \bar{y}_{N+1} \) such that

\[
B \bar{y}_{N+1} \geq \begin{pmatrix}
\max_{i \in [N]} \{|[b_i - A_i \bar{x} - B_i \bar{y}_i]|\} \\
\vdots \\
\max_{i \in [N]} \{|[b_i - A_i \bar{x} - B_i \bar{y}_i]|\}
\end{pmatrix},
\]

where we use \([v]_m\) to denote the \( m \)-th element in a vector \( v \). Therefore, for any \( (z, u) \in \bar{Z} \), we have

\[
[B \bar{y}_{N+1}]_m u \geq \sum_{i \in [N]} [B \bar{y}_{N+1}]_m |z_i| \geq \sum_{i \in [N]} |[b_i - A_i \bar{x} - B_i \bar{y}_i]| |z_i| \geq \sum_{i \in [N]} [b_i - A_i \bar{x} - B_i \bar{y}_i]_m z_i, \quad \forall m \in [M],
\]

which leads to

\[
B \bar{y}_{N+1} u \geq \sum_{i \in [N]} (b_i - A_i \bar{x} - B_i \bar{y}_i) z_i, \quad \forall (z, u) \in \bar{Z}.
\]

Recall that we also have \( A_0 \bar{x} + B \bar{y}_0 \geq b_0 \); therefore, we have

\[
A(z) \bar{x} + B \bar{y}_0 + \sum_{i \in [N]} B_i \bar{y}_i z_i + B \bar{y}_{N+1} u \geq b(z), \quad \forall (z, u) \in \bar{Z}.
\]

Choose

\[
\bar{k} = \max \{0, \max_{i \in [N]} \{|d^\top \bar{y}_i + c_i^\top \bar{x}|\} + d^\top \bar{y}_{N+1}\}.
\]

Then, a set of feasible solutions is given by \( \bar{x}, \bar{k}, \bar{y}_0, \ldots, \bar{y}_{N+1} \), would satisfy

\[
c(z)^\top \bar{x} + d^\top \bar{y}_0 + \sum_{i \in [N]} d^\top \bar{y}_i z_i + d^\top \bar{y}_{N+1} u - \bar{k} u \leq \bar{k} u, \quad \forall (z, u) \in \bar{Z},
\]

\[
A(z) \bar{x} + B \bar{y}_0 + \sum_{i \in [N]} B_i \bar{y}_i z_i + B \bar{y}_{N+1} u \geq b(z), \quad \forall (z, u) \in \bar{Z},
\]

\[
\bar{x} \in \mathcal{X}, \bar{k} \geq 0.
\]

We now prove the optimality of of the affine recourse adaptation model (23) when \( P = 1 \). In this case, matrix \( B \) has dimension \( M \times 1 \), i.e., \( B \in \mathbb{R}^{M,1} \) and it is either strictly positive or negative due to the assumption of complete recourse. When \( dB \leq 0 \), the solutions to both Problem (22) and Problem (23) are trivial and their objective values coincide. For non-trivial cases, we can, without loss of generality, focus on \( d > 0 \) and \( B > 0 \). Consequently, any feasible solution in (22) must satisfy

\[
\frac{k u + \tau - c(z)^\top x}{d} \geq y(z, u) \geq \max_{m \in [M]} \left\{ \frac{|b(z) - A(z)x|}{B_m} \right\}, \quad \forall (z, u) \in \bar{Z}.
\]
Therefore, an optimal recourse function would be
\[ y(z, u) := \frac{k u + \tau - c(z)^\top x}{d}, \]
which is an affine function of $z$ and $u$. Therefore, there exists an optimal lifted affine function for Problem (23) that achieves the same optimal solution as Problem (22).

\[
\text{Proof of Theorem 3.} \quad \text{Based on the definition of Wasserstein metric, we can rewrite the stochastic linear robustness optimization model as}
\]
\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \mathbb{E}_Q [c(\tilde{z})^\top x + d^\top y(\tilde{z}) - k \|\tilde{z} - \bar{v}\|_1] - \tau \leq 0, \quad \forall Q \in \mathcal{Q}, \\
 & \quad A(z)x + By(z) \geq b(z), \quad \forall z \in Z, \\
 & \quad y \in \mathcal{R}^{N,P}, \quad x \in \mathcal{X}, \quad k \geq 0
\end{align*}
\]
where the ambiguity set $\mathcal{Q}$ is defined as
\[
\mathcal{Q} := \left\{ Q \in \mathcal{P}_0(Z \times Z) \mid (\tilde{z}, \bar{v}) \sim Q, \tilde{z} \sim P, \bar{v} \sim \hat{P} \right\}.
\]

Given the structure of $\hat{P}$, we can rewrite the above problem as
\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{Q \in \mathcal{P}_0(Z \times Z)} [c(\tilde{z})^\top x + d^\top y(\tilde{z}) - k \|\tilde{z} - \hat{z}_s\|_1] \leq \tau, \quad \forall Q \in \mathcal{P}_0(Z), s \in [S], \\
 & \quad A(z)x + By(z) \geq b(z), \quad \forall z \in Z, \\
 & \quad y \in \mathcal{R}^{N,P}, \quad x \in \mathcal{X}, \quad k \geq 0,
\end{align*}
\]
where $Q_s$ can be seen as the conditional probability distribution of $\tilde{z}$ given $\bar{v} = \hat{z}_s$. The worst case distribution $Q_s^* \in \mathcal{P}_0(Z), s \in [S]$, is a one-point distribution. Therefore, the above problem is equivalent to
\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \frac{1}{S} \sum_{s \in [S]} \sup_{\tilde{z} \in Z} \left\{ c(\tilde{z})^\top x + d^\top y(\tilde{z}) - k \|\tilde{z} - \hat{z}_s\|_1 \right\} \leq \tau, \\
 & \quad A(z)x + By(z) \geq b(z), \quad \forall z \in Z, \\
 & \quad y \in \mathcal{R}^{N,P}, \quad x \in \mathcal{X}, \quad k \geq 0,
\end{align*}
\]

We will show that for any optimal solution in (45) we can find a feasible solution in (26) with the same objective value and vice versa.

Consider any optimal solution $(\tilde{k}, \tilde{x}, \tilde{y})$ in Problem (45). We define $\tilde{y}_s(z, u) := \tilde{y}(z)$ for all $(z, u) \in \tilde{Z}_s, s \in [S]$. Because $A(z)x + B\tilde{y}(z) \geq b(z)$ for all $z \in \tilde{Z}$, we also have $A(z)x + B\tilde{y}_s(z, u) \geq b(z)$ for all $(z, u) \in \tilde{Z}_s, s \in [S]$.\]
In addition,
\[
\frac{1}{S} \sum_{s \in [S]} \sup_{z \in Z} \left\{ c(z)^\top \bar{x} + d^\top \bar{y}(\hat{z}) - \bar{k} \|z - \hat{z}_s\|_1 \right\} \leq \tau
\]
indicates
\[
\frac{1}{S} \sum_{s \in [S]} \sup_{(z, u) \in \hat{Z}_s} \left\{ c(z)^\top \bar{x} + d^\top \bar{y}_s(z, u) - \bar{k} u \right\} \leq \tau
\]
because of the definition of \( \bar{y} \) and \( \hat{Z}_s, s \in [S] \). Therefore, \((\bar{k}, \bar{x}, \bar{y}_1, \ldots, \bar{y}_S)\) is feasible in Problem (26) and leads to the same objective value.

Now, consider any optimal solution \((\bar{k}, \bar{x}, \bar{y}_1, \ldots, \bar{y}_S)\) in Problem (26). We define \( \bar{y}(z) := \bar{y}_{s^*(z)}(z, \|z - \hat{z}_{s^*(z)}\|_1) \) for all \( z \in Z \), where
\[
s^*(z) := \arg \min_{s \in [S]} \{ d^\top \bar{y}_s(z, \|z - \hat{z}_s\|_1) \}.
\]
Because \( A(z)\bar{x} + B\bar{y}_s(z, u) \geq b(z) \) for all \((z, u) \in \hat{Z}_s, s \in [S]\), we must have
\[
A(z)\bar{x} + B\bar{y}_s(z, \|z - \hat{z}_s\|_1) \geq b(z), \quad \forall z \in Z, s \in [S].
\]
By the definition of \( \bar{y} \), we have
\[
A(z)\bar{x} + B\bar{y}(z) \geq b(z), \quad \forall z \in Z.
\]
Notice that
\[
\frac{1}{S} \sum_{s \in [S]} \sup_{(z, u) \in \hat{Z}_s} \left\{ c(z)^\top \bar{x} + d^\top \bar{y}_s(z, u) - \bar{k} u \right\} \leq \tau,
\]
which indicates
\[
\frac{1}{S} \sum_{s \in [S]} \sup_{z \in Z} \left\{ c(z)^\top \bar{x} + d^\top \bar{y}_s(z, \|z - \hat{z}_s\|_1) - \bar{k} \|z - \hat{z}_s\|_1 \right\} \leq \tau.
\]
Then, by the definition of \( \bar{y} \), we have
\[
\frac{1}{S} \sum_{s \in [S]} \sup_{z \in Z} \left\{ c(z)^\top \bar{x} + d^\top \bar{y}(z) - \bar{k} \|z - \hat{z}_s\|_1 \right\} \leq \tau.
\]
The above follows from the definition of \( \bar{y} \); specifically,
\[
d^\top \bar{y}(z) \leq d^\top \bar{y}_s(z, \|z - \hat{z}_s\|_1), \quad \forall z \in Z, s \in [S].
\]
Therefore, from \((\bar{k}, \bar{x}, \bar{y})\), we can get a feasible solution \((\bar{k}, \bar{x}, \bar{y})\) for Problem (45) that leads to the same objective value.
\[\square\]
Proof of Theorem 4. Because \( \tau \geq Z_0 \), there exists some \( \bar{x} \in \mathcal{X} \) and \( \bar{y}_s \in \mathbb{R}^p, s \in [S] \) such that
\[
\frac{1}{S} \sum_{s \in [S]} (c(\hat{z}_s)^\top \bar{x} + d^\top \bar{y}_s) \leq \tau, \\
A(\hat{z}_s)\bar{x} + B\bar{y}_s \geq b(\hat{z}_s), \quad \forall s \in [S].
\]
Because matrix \( B \) has complete recourse, we can find a \( \hat{y}_{s,N+1} \) such that:
\[
B\hat{y}_{s,N+1} \geq \left( \begin{array}{c}
\max_{i \in [N]} \{|b_i - A_i \bar{x}| \} \\
\vdots \\
\max_{i \in [N]} \{|b_i - A_i \bar{x}|_M \}
\end{array} \right).
\]
Therefore, for any \((z, u) \in \tilde{Z}_s, s \in [S]\) and \( m \in [M] \), we have
\[
[B\hat{y}_{N+1}]_m u \geq \sum_{i \in [N]} |[B\hat{y}_{N+1}]_m| \sum_{i \in [N]} |[b_i - A_i \bar{x}]_m| |z_i - \hat{z}_{si}| \geq \sum_{i \in [N]} |[b_i - A_i \bar{x}]_m| (z_i - \hat{z}_{si}),
\]
which indicates
\[
B\hat{y}_{N+1} u \geq \sum_{i \in [N]} (b_i - A_i \bar{x})(z_i - \hat{z}_{si}), \quad \forall (z, u) \in \tilde{Z}_s, \quad \forall s \in [S].
\]
Notice that we can rewrite \( A(z) x \) and \( b(z) \) as:
\[
A(z) x = A(z - \hat{z}_s + \hat{z}_s) x = A(\hat{z}_s) x + \sum_{i \in [N]} A_i x (z_i - \hat{z}_{si}), \quad \forall s \in [S], \forall z \in Z,
\]
\[
b(z) = b(z - \hat{z}_s + \hat{z}_s) = b(\hat{z}_s) + \sum_{i \in [N]} b_i (z_i - \hat{z}_{si}), \quad \forall s \in [S], \forall z \in Z.
\]
Let us define \( \hat{y}_{s,0} := \bar{y}_s \) and \( \hat{y}_{s,i} := 0 \) for \( i \in [N] \). By above inequality and the alternative form of \( A(z)x \) and \( b(z) \), we have
\[
A(z) \bar{x} + B(\hat{y}_{s,0} + \hat{y}_{s,N+1} u + \sum_{i \in [N]} \hat{y}_{s,i} z_i) \geq b(z), \quad \forall (z, u) \in \tilde{Z}_s, \quad \forall s \in [S].
\]
Next, we define
\[
\hat{k} := \max \left\{ 0, \max_{i \in [N], s \in [S]} \left\{ |c_i^\top \bar{x}| + d^\top \hat{y}_{s,N+1} \right\} \right\},
\]
which satisfies
\[
\sum_{i \in [N]} c_i^\top (z_i - \hat{z}_{si}) + d^\top \hat{y}_{s,N+1} u - \hat{k} u \leq 0, \quad \forall (z, u) \in \tilde{Z}_s, \quad \forall s \in [S].
\]
Notice that we can rewrite \( c(z)^\top x \) as
\[
c(z)^\top x = c(z - \hat{z}_s + \hat{z}_s)^\top x = c(\hat{z}_s)^\top x + \sum_{i \in [N]} c_i^\top (z_i - \hat{z}_{si}), \quad \forall s \in [S], \forall z \in Z.
\]
By above inequality and alternative form of $c(z)\top x$, we have

$$\frac{1}{S} \sum_{s \in [S]} \sup_{(z,u) \in \bar{Z}_s} \left\{ c(z)\top x + d\top (\hat{y}_{s,0} + \sum_{i \in [N]} \hat{y}_{s,i} z_i + \hat{y}_{s,N+1} u) - k u \right\}$$

$$\leq \frac{1}{S} \sum_{s \in [S]} (c(\hat{z}_s)\top x + d\top \hat{y}_s)$$

$$\leq \tau.$$

Therefore, $(\hat{k}, \hat{x}, \hat{y}_{s,0}, \ldots, \hat{y}_{s,N+1})$ constitutes a feasible solution to Problem (28).

Now, we show that the scenario-wise lifted affine recourse adaptation would yield the exact objective value as Problem (26) when $P = 1$. In this case, matrix $B$ has dimension $M \times 1$, i.e., $B \in \mathbb{R}^{M,1}$ and it is either strictly positive or negative. When $dB \leq 0$, the solutions to both Problem (28) and Problem (26) are trivial and their objective values coincide. For non-trivial cases, we can, without loss of generality, focus on $d > 0$ and $B > 0$.

In Problem (26), the recourse function $y_s(z)$ must satisfy

$$y_s(z,u) \geq \max_{m \in [M]} \left\{ \frac{[b(z) - A(z)x]_m}{[B]_m} \right\}, \quad \forall (z,u) \in \bar{Z}_s, \forall s \in [S].$$

In addition,

$$\frac{1}{S} \sum_{s \in [S]} \sup_{(z,u) \in \bar{Z}_s} \left\{ c(z)\top x + dy_s(z,u) - ku \right\} \leq \tau$$

can be equivalently written as

$$\frac{1}{S} \sum_{s \in [S]} v_s \leq \tau$$

$$v_s \geq c(z)\top x + dy_s(z,u) - ku, \quad \forall (z,u) \in \bar{Z}_s, \forall s \in [S].$$

The above indicates that the recourse function $y_s(z)$ must satisfy

$$y_s(z,u) \leq \frac{v_s - c(z)\top x + ku}{d}, \quad \forall (z,u) \in \bar{Z}_s, \forall s \in [S].$$

Therefore, an optimal recourse function for any $s \in [S]$ would be

$$y_s(z,u) := \frac{v_s - c(z)\top x + ku}{d},$$

which is an affine function of $z$ and $u$ for a given scenario $s$. Therefore, there exists an optimal scenario-wise lifted affine function for Problem (28) that achieves the same optimal solution as Problem (26).
Proof of Theorem 5. By Theorem 3 and the definition of utility function \( u \), the robustness constraint in Problem (29) can be written as:

\[
\begin{align*}
\min_k & \quad \tau + \frac{1}{S} \sum_{s \in [S]} v_s \\
\text{s.t.} & \quad v_s \geq \sup_{(z,w) \in \tilde{Z}_s} \{-a_i x^\top z - b_i - kw\}, \quad \forall i \in [I], \\
& \quad 1^\top x = 1, \\
& \quad x \in \mathbb{R}^N, \quad k \geq 0,
\end{align*}
\]

where we define the lifted uncertainty set \( \tilde{Z}_s := \{(z,w) \mid z \in \mathbb{R}^N, \ w \geq \|z - \hat{z}_s\|_1\} \).

Now, we focus on the inner maximization. By strong duality, for any \( i \in [I] \) and \( s \in [S] \), the inner maximization problem \( Z_{is} = \sup_{(z,w) \in \tilde{Z}_s} \{-a_i x^\top z - b_i - kw\} \) can be reformulated as:

\[
Z_{is} = \inf_{n \in [N]} \sum_{s \in [S]} (p_{2n}^{(is)} - p_{1n}^{(is)}) \hat{z}_{sn} - b_i
\]

s.t. \( p_{2n}^{(is)} - p_{1n}^{(is)} = -a_i x_n, \)

\( p_{1n}^{(is)} + p_{2n}^{(is)} = -k, \)

\( p_i^{(is)} \leq 0, \quad p_{2n}^{(is)} \leq 0. \)

The objective function in above problem is effectively \(-a_i x^\top \hat{z}_s - b_i\). In addition, by \( p_{2n}^{(is)} - p_{1n}^{(is)} = -a_i x_n, p_{1n}^{(is)} + p_{2n}^{(is)} = -k, p_{1n}^{(is)} \leq 0 \), and \( p_{2n}^{(is)} \leq 0 \), we must have:

\[
k \geq a_i x_n,
\]

\[
k \geq -a_i x_n.
\]

Notice that \( a_i \geq 0 \), for all \( i \in [I] \). The above must hold for all \( i \in [I] \) and \( n \in [N] \); therefore, we have

\[
k \geq \bar{a}\|x\|_\infty,
\]

where we define \( \bar{a} := \max_{i \in [I]} \{a_i\} \).

Therefore, the full problem can be written as

\[
\begin{align*}
\min_k & \quad \tau + \frac{1}{S} \sum_{s \in [S]} v_s \\
\text{s.t.} & \quad v_s \geq -a_i x^\top \hat{z}_s - b_i, \quad \forall i \in [I], \\
& \quad k \geq \bar{a}\|x\|_\infty, \\
& \quad 1^\top x = 1, \\
& \quad x \in \mathbb{R}^N,
\end{align*}
\]
which is equivalent to

\[
\begin{align*}
\min & \quad k \\
\text{s.t.} & \quad \tau + \frac{1}{S} \sum_{s \in [S]} v_s \leq 0, \\
& \quad v_s \geq -u(x^T \hat{z}_s), \\
& \quad k \geq \hat{a}\|x\|_{\infty}, \\
& \quad 1^T x = 1, \\
& \quad x \in \mathbb{R}^N.
\end{align*}
\]

Notice that \( \hat{a} \) is merely a scaling constant in constraint \( k \geq \hat{a}\|x\|_{\infty} \), the above model has the same solution as:

\[
\begin{align*}
\min & \quad \|x\|_{\infty} \\
\text{s.t.} & \quad \tau + \frac{1}{S} \sum_{s \in [S]} v_s \leq 0, \\
& \quad v_s \geq -u(x^T \hat{z}_s), \\
& \quad 1^T x = 1, \\
& \quad x \in \mathbb{R}^N.
\end{align*}
\]

The final reformulation in the Theorem then follows directly by replacing the \( v_s \) variable in the first constraint with \( -u(x^T \hat{z}_s) \).

\[\Box\]

**Proof of Theorem 6.** First, the robustness counterpart can be written as:

\[
\begin{align*}
\max_{n=1}^{N} d_n x_n z_n - k \|z\|_1 \quad & \iff \max_{n=1}^{N} (d_n x_n - k) z_n \\
\text{s.t.} & \quad z \in Z \quad & \text{s.t.} & \quad z \geq 0, \; z \leq 1.
\end{align*}
\]

It follows from the strong duality that the formulation on the right is equivalent to:

\[
Y = \min_{n=1}^{N} y_n
\]

\[
\text{s.t. } \quad y_n \geq d_n x_n - k, \; \forall n \in [N] \\
& \quad y_n \geq 0, \quad \forall n \in [N].
\]

Because \( x_n \) is binary for any \( n \in [N] \) and \( k \geq 0 \), the above optimization problem has a closed-form solution:

\[
Y = \sum_{n=1}^{N} \max\{d_n - k, 0\} x_n
\]

Therefore, the original robustness optimization problem (32) can be written as:

\[
\begin{align*}
k^* = & \min k \\
\text{s.t.} & \quad c^T x + \sum_{n=1}^{N} \max\{d_n - k, 0\} x_n \leq \tau \\
& \quad x \in \mathcal{X}, \; k \geq 0.
\end{align*}
\]
For the second part, since the objective function is $k$, we can do a bisection search on $k$. Note that $k^* \leq \bar{d}$ because we choose $\tau \geq Z_0$, for practicality. For any fixed $k \leq \bar{d}$, we solve a feasibility problem that has the same complexity as the baseline combinatorial optimization problem. Starting from a feasible region, $[0, \bar{d}]$, each bisection search iteration reduces the searching region by half. Therefore, to achieve any given accuracy $\epsilon$, we solve at most $\lceil \log_2(\bar{d}/\epsilon) \rceil$ baseline problems with different linear objective functions.

\[ \text{Proof of Theorem 7} \] By Equations (35) and (36),

\[ Z_r = \min_{\pi \in \Pi} \inf_{k > 0} \left\{ k \log \mathbb{E}_{\tilde{\pi}} \left[ \sum_{t \in [T]} l_t(s_t, x_t(h_t, s_t), \tilde{z}_t) \right] + kr \right\} = \inf_{k > 0} \{ G(k) + kr \}, \]

where the second equality is due to switching the minimization and infimum, and

\[ G(k) = \min_{\pi \in \Pi} k \log \mathbb{E}_{\tilde{\pi}} \left[ \exp \left( \frac{\sum_{t \in [T]} l_t(s_t, x_t(h_t, s_t), \tilde{z}_t)}{k} \right) \right]. \] \hspace{1cm} (46)

Problem (46), which optimizes the certainty equivalent for exponential disutility, has been discussed by Jacobson (1973). Specifically, the optimal policy can be obtained via the recursion in Equation (38) with $G_{T+1}(k, s) := 0$. We then have $G(k) = G_1(k, s_1)$.

\[ \text{Proof of Theorem 8.} \] For $k > 0$, the first constraint in Problem (40) can be equivalently reformulated as

\[ k \log \mathbb{E}_{\tilde{\pi}} \left[ \exp \left( \frac{\sum_{t \in [T]} l_t(s_t, x_t(h_t, s_t), \tilde{z}_t)}{k} \right) \right] \leq \tau. \]

Combining it with $\pi \in \Pi$, we can equivalently have the first constraint in Problem (42). Since the left hand side function is non-increasing in $k$, we can perform bisection search on $k$ to determine $k^*$ with high degree of accuracy by solving a modest number of subproblems. At each subproblem (with a given $k$), we obtain $G_1(k, s_1)$ and corresponding optimal policy in the same way as the dynamic recursion in Theorem 7.
B. Infeasibility of Lifted Affine Recourse Adaptation under Relative Complete Recourse

We adopt the example used in Bertsimas et al. (2019). Consider the case of unbounded support $Z = [-1,1]^2$ and

$$\min k$$

s.t. $0 \times y(z) - 0 \leq k \|z\|_1, \quad \forall z \in Z$

$$y(z) \geq z_1 - z_2 \quad \forall z \in Z$$

$$y(z) \geq z_2 - z_1 \quad \forall z \in Z$$

$$y(z) \leq z_1 + z_2 + 2 \quad \forall z \in Z$$

$$y(z) \leq -z_1 - z_2 + 2 \quad \forall z \in Z$$

$$y \in \mathbb{R}^{3,1}, k \geq 0,$$

for which a feasible recourse function would be $y(z) = |z_1 - z_2|$. Hence, this is a relatively complete recourse problem. Under the lifted affine recourse adaptation, we solve the following problem:

$$\min k$$

s.t. $0 \leq ku, \quad \forall (z,u) : u \geq \|z\|_1, z \in Z$

$$y_0 + y_1 z_1 + y_2 z_2 + y_3 u \geq z_1 - z_2 \quad \forall (z,u) : u \geq \|z\|_1, z \in Z$$

$$y_0 + y_1 z_1 + y_2 z_2 + y_3 u \geq z_2 - z_1 \quad \forall (z,u) : u \geq \|z\|_1, z \in Z$$

$$y_0 + y_1 z_1 + y_2 z_2 + y_3 u \leq z_1 + z_2 + 2 \quad \forall (z,u) : u \geq \|z\|_1, z \in Z$$

$$y_0 + y_1 z_1 + y_2 z_2 + y_3 u \leq -z_1 - z_2 + 2 \forall (z,u) : u \geq \|z\|_1, z \in Z$$

$$y_0, y_1, y_2, y_3 \in \mathbb{R}, k \geq 0.$$

For $z = 0$, the pair of semi-infinite constraints $y_0 + y_3 u \geq 0$ and $y_0 + y_3 u \leq 2$ for all $u \geq 0$, implies that $y_3 = 0$. With $z_1 = z_2 = 1$, we have $y_0 + y_1 + y_2 = 0$, and with $z_1 = z_2 = -1$, we have $y_0 - y_1 - y_2 = 0$, both implying $y_0 = 0$. Moreover, with $z_1 = 1, z_2 = -1$, we would require $y_1 - y_2 = 2$. However, infeasibility occurs when $z_1 = -1, z_2 = 1$, which mandates $-y_1 + y_2 = 2$. Hence, no lifted affine recourse function would be feasible in the above problem.
C. Additional Simulation Results

The knapsack problem In the main text, we used uniform distribution to generate the test data. Here, we use a skewed distribution, Beta(3, 2.5). Specifically, $\bar{z} = 2\bar{v} - 1$, where $\bar{v} \sim $ Beta(3, 2). The results are summarized in Figure 8.

![Figure 8](image_url)

(a) Prob. violation w.r.t budget of uncertainty $r$

(b) Prob. violation w.r.t target $\tau$

(c) Avg. violation w.r.t budget of uncertainty $r$

(d) Avg. violation w.r.t target $\tau$

The adaptive network lot-sizing problem In the main text, we used uniform distribution to generate the test data. Here, we use a skewed distribution, Beta(2, 1). Specifically, In each sample, demand at each stores $i$ is first generated by $\tilde{z}_i = \delta_i\tilde{v}_i$, $\tilde{v}_i \sim $ Beta(2, 1), and then normalized to reflect demand correlation. The results are summarized in Figure 9 and Figure 10.

The data-driven adaptive network lot-sizing problem In the main text, we used a normal distribution, $N(20, 10^2)$, to generate the test data. Here, we use a normal distribution $N(20, 12^2)$. The results are summarized in Figure 11, Figure 12, and Figure 13.
Figure 9  Comparison between RO and RnO: Average cost (left) and cost at the 95th percentile (right).

Figure 10  Summary of performance: GRC-sum-based model with $r = 0, 20, 30, 60$ at different first-stage costs

Figure 11  Summary of performance: DRnO and DRO models at different first-stage costs
(a) Cost metrics w.r.t. Wasserstein radius $r$

(b) Cost metrics w.r.t. target $\tau$

Figure 12  Out-of-sample cost metrics w.r.t. Wasserstein radius $r$ in DRO (left) and target $\tau$ in DRnO (right).

(a) Cost metrics w.r.t. Wasserstein radius $r$

(b) Cost metrics w.r.t. target $\tau$

Figure 13  Change of the “nice” range of radius in DRO (left) and normalized target in DRnO (right).
D. A Sample Code for DRnO and DRO using RSOME

RSOME is a Matlab modeling package for distributionally robust optimization (Chen et al. 2020), which can be used to solve robustness optimization and data-driven robustness optimization. Here, we provide a sample code of the data-driven adaptive network lot-sizing problem. The sample code for the DRO model is as follows:

```matlab
%%%Parameters%%%
% S: number of samples
% N: number of stores
% D: maximum possible demand
% c: cost of ordering
% l: cost of emergency ordering
% t: transportation cost
% d: historical samples of demands
% r: radius of the Wasserstein ball

%%%Variables%%%%
% x: initial stock allocation
% w: second-stage emergency ordering
% y: second-stage transhipment

%%%Model%%%%
model = rsome('DRO');

% Define random variables
z = model.random(N,1);  %demand
u = model.random;       %lifted variable

% Define scenarios and the lifted joint ambiguity set
P = model.ambiguity(S);
for s = 1:S
    P(s).suppset(0 <= z, z <= D, norm(z - d(s,:)) <= u);
end
pr = P.prob;
P.probset(pr == 1/S);

% Define event-wise expectation
P.exptset(expect(u) <= r);

% Declare Wasserstein ambiguity set
model.with(P);

% Define decision variables
x = model.decision(N,1);
y = model.decision(N,N);
w = model.decision(N,1);

% Define scenario-wise adaptation
```
for s = 1:S
    y.evtadapt(s);
    w.evtadapt(s);
end
% Define affine adaptation
y.affadapt(z);
y.affadapt(u);
w.affadapt(z);
w.affadapt(u);
% Define objective function
model.min(c'*x+expect(sum(sum(t.*y) + c2'*w));

% Define constraints
for i = 1:N
    model.append(z(i)-x(i)-w(i)+sum(y(i,:))-sum(y(:,i)) <= 0);
end

model.append(y >= 0);
model.append(w >= 0);
model.append(x >= 0);
model.append(x <= D);

% Solution
model.solve;

The sample code for the DRnO model is as follows:

%%%Parameters%%%
% S: number of samples
% N: number of stores
% D: maximum possible demand
% c: cost of ordering
% l: cost of emergency ordering
% t: transportation cost
% d: historical samples of demands
% T: target

%%%Variables%%%%
% x: initial stock allocation
% w: second-stage emergency ordering
% y: second-stage transhipment

%%%Model%%%%
model = rsome('DRnO');

% Define random variables
z = model.random(N,1);   %demand
u = model.random;         %lifted variable
% Define scenarios and the lifted joint ambiguity set
P = model.ambiguity(S);
for s = 1:S
    P(s).suppset(0 <= z, z <= D, norm(z - d(s,:)',1) <= u);
end

pr = P.prob;
P.probsupset(pr == 1/S);
% Declare ambiguity set
model.with(P);

% Define decision variables
k = model.decision();
x = model.decision(N,1);
y = model.decision(N,N);
w = model.decision(N,1);

% Define scenario-wise adaptation
for s = 1:S
    y.evtadapt(s);
    w.evtadapt(s);
end
% Define affine adaptation
y.affadapt(z);
y.affadapt(u);
w.affadapt(z);
w.affadapt(u);
% Define objective function
model.min(k);

% Define constraints
model.append(c'x + expect(sum(sum(t.*y)) + c2*w - k*u) <= T)
for i = 1:N
    model.append(z(i)-x(i)-w(i)+sum(y(i,:))-sum(y(:,i)) <= 0);
end
model.append(y >= 0);
model.append(w >= 0);
model.append(x >= 0);
model.append(x <= D);
% Solution
model.solve;