

# Robust Conic Satisficing

Arjun Ramachandra

Engineering Systems & Design, Singapore University of Technology and Design

arjun.ramachandra@gmail.com

Napat Rujeerapaiboon

Department of Industrial Systems Engineering & Management, National University of Singapore

napat.rujeerapaiboon@nus.edu.sg

Melvyn Sim

Department of Analytics & Operations, National University of Singapore

melvynsim@gmail.com

In practical optimization problems, we typically model uncertainty as a random variable though its true probability distribution is unobservable to the decision maker. Historical data provides some information of this distribution that we can use to approximately quantify the risk of an evaluation function that depends on both our decision and the uncertainty. This empirical optimization approach is vulnerable to the issues of overfitting, which could be overcome by several data-driven robust optimization techniques. To tackle overfitting, Long et al. (2022) propose a *robust satisficing* model, which is specified by a performance target and a penalty function that measures the deviation of the uncertainty from its nominal value, and it yields solutions with superior out-of-sample performance. We generalize the robust satisficing framework to conic optimization problems with recourse, which has broader applications in predictive and prescriptive analytics including risk management, statistical supervised learning, among others. We derive an exact semidefinite optimization formulation for biconvex quadratic evaluation function, with quadratic penalty and ellipsoidal support set. More importantly, under complete and bounded recourse, and reasonably chosen polyhedral support set and penalty function, we propose safe approximations that do not lead to infeasible problems for any reasonably chosen target. We demonstrate that the assumption of complete and bounded recourse is however not unimpeachable, and then introduce a novel *perspective casting* technique to derive an equivalent conic optimization problem that satisfies the stated assumptions. Finally, we showcase the computational study on data-driven portfolio optimization that maximizes the expected exponential utility of the portfolio returns. We demonstrate that the solutions obtained via robust satisficing can significantly improve over the solutions obtained by stochastic optimization models, including the celebrated Markowitz model, which is prone to overfitting.

*Key words:* robust optimization, robust satisficing, conic optimization, affine recourse adaptation

---

## 1. Introduction

नासतो विद्यते भावो नाभावो विद्यते सतः ।

“Of the impermanent there is no certainty” - Gita (2.16)

Data is an integral part of predictive and prescriptive analytics that we would use for evaluating approximately the potential risks of decisions that might arise in future. Although historical data provide information of the model’s underlying uncertainty, the true probability distribution of the random variable remains unobservable to the decision maker. In stochastic optimization, we evaluate the model’s risk-based objective function using an explicit probability distribution with parameters estimated from historical data (Shapiro et al. 2014, Birge and Louveaux 2011). In particular, an empirical optimization problem is a stochastic optimization problem when we use the empirical distribution for risk evaluation. However, as articulated in Smith and Winkler (2006) *Optimizer’s curse*, stochastic optimization models are prone to overfitting and would likely yield inferior solutions when the risks are evaluated on the true distributions. This is exemplified in the study of DeMiguel et al. (2009) that demonstrates the inferiority of optimized portfolios when evaluated on the actual data.

To tackle overfitting, Mohajerin Esfahani and Kuhn (2018) propose the seminal framework for data-driven robust optimization, by incorporating an ambiguity set of probability distributions that are proximal to the empirical distribution with respect to the Wasserstein metric. The approach alleviates the optimizer’s curse by avoiding overfitting to the empirical distribution. When appropriately sized, the ambiguity set captures the true distribution with high level of confidence, and the robust solutions could yield superior out-of-sample performance compared to solutions obtained from empirical optimization models. Shafieezadeh-Abadeh et al. (2019) extend the models to solving supervised learning problems and establish the relation with regularization for Lipschitz continuous loss function and unbounded support for the underlying random variable. As in the case of regularization, the size of the ambiguity set is a hyper-parameter that can be determined by cross-validation techniques.

Long et al. (2022) recently propose an alternative approach for improving the solutions of data-driven optimization problems known as robust satisficing. For one data sample, this model corresponds to a case of the GRC-sum model of Ben-Tal et al. (2017). Instead of sizing the ambiguity set, the satisficing model delivers solutions that aim to achieve the specified risk-based target to a feasible extent, notwithstanding the potential mismatch between the true distribution and the empirical distribution. Sim et al. (2021) extend the robust satisficing approach to solving supervised learning problems and show in studies on well-known data sets that the robust satisficing approach would generally outperform the robust optimization approach, when the respective target and size parameters are determined via cross-validation. In this paper, we generalize the robust satisficing

framework to conic optimization problems with recourse, which has broader applications in predictive and prescriptive analytics including, *inter alia*, risk management and statistical supervised learning.

As in the case of robust optimization models, it can be computationally challenging to solve robust satisficing models exactly. However, since the inception of robust optimization, we now have an arsenal of tools to address and solve either exactly, or by providing tractable safe approximations for various kinds of robust optimization models (see, for instance, Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004, Ben-Tal et al. 2004, Kuhn et al. 2011, Bertsimas et al. 2010, Delage and Ye 2010, Wiesemann et al. 2014, Ben-Tal et al. 2015, Bertsimas et al. 2019, Chen et al. 2020, among others). We wish to highlight Bertsimas and de Ruiter (2016) for proposing the dualizing technique and applying affine dual recourse adaptation to address an adaptive robust linear optimization problem. When the uncertainty set is polyhedral, this approach can also be used to provide safe approximations for robust optimization models with biconvex constraint functions including those with recourse adaptation (de Ruiter et al. 2018, Roos et al. 2020). We use this approach to obtain solutions to our proposed robust conic satisficing models. We also highlight the key challenge that in providing safe approximations to the robust satisficing models; it is *sine qua non* to ensure that the approximations do not lead to infeasible problems if the chosen target is above the optimum objective when the empirical optimization problem is minimized. This necessitates careful analysis and provides conditions for which safe approximations can be provided for a broad class of robust satisficing models. We summarize the contributions of this paper below:

1. We provide a unifying robust conic satisficing framework to tackle overfitting in data-driven optimization that can be applied to predictive and prescriptive analytics. We solve the robust conic satisficing model either exactly when possible, or provide a *tractable safe approximation* that ensures its feasibility for any chosen target that is achievable by the empirical optimization problem.
2. For a biconvex quadratic evaluation function with a quadratic penalty and an ellipsoidal support set, we derive an exact semidefinite optimization formulation of the robust satisficing problem. For a more general conic representable evaluation function, we derive a tractable safe approximation using *affine dual recourse adaptation* under assumptions of complete and bounded recourse along with a polyhedral support set and polyhedral penalty function.
3. We show that the perspectification approach to model nonlinear convex function as a conic optimization problem does not necessarily have the property of complete and bounded recourse. Hence, we introduce a novel *perspective casting* approach to model such an evaluation function as an equivalent conic optimization problem with complete and bounded recourse over a *featured domain*.

4. For linear optimization with recourse, we show that the tractable safe approximation via affine dual recourse adaptation dominates the primal approach in terms of the approximation quality and computational performance.
5. We showcase the performance benefits of the robust satisficing framework with a portfolio selection problem that maximizes expected exponential utility compared to the solutions generated from empirical optimization and stochastic optimization models, which are prone to overfitting.

**Notation:** We use  $\mathbb{R}$  to denote the space of reals while  $\mathbb{R}_+$  and  $\mathbb{R}_{++}$  denote the sets of non-negative and strictly positive real numbers, respectively. We use boldface small-case letters (*e.g.*  $\mathbf{x}$ ) to denote vectors, capitals (*e.g.*  $\mathbf{A}$ ) to denote matrices and capital caligraphic letters to denote sets (*e.g.*  $\mathcal{Z}$ ) including cones (*e.g.*  $\mathcal{K}$ ).  $\mathcal{R}^{m,n}$  and  $\mathcal{L}^{m,n}$  are used to denote the set of all functions and its sub-class of affine functions, respectively, from  $\mathbb{R}^m$  to  $\mathbb{R}^n$ . The transpose of a vector (matrix) is denoted by  $\mathbf{x}^\top$  ( $\mathbf{A}^\top$ ). We use  $[n]$  to denote the running index set  $\{1, 2, \dots, n\}$ , and the superscript indexing (*e.g.*  $\mathbf{w}^i$ ) is used to denote the  $i^{\text{th}}$  vector (matrix) among a countable set of vectors  $\{\mathbf{w}^i\}$  (matrices) and subscript indexing (*e.g.*  $\mathbf{A}_i$ ) to denote the  $i^{\text{th}}$  row of a matrix  $\mathbf{A}$ . A random vector is denoted with a tilde sign (*e.g.*  $\tilde{\mathbf{z}}$ ), and its probability distribution is denoted by  $\tilde{\mathbf{z}} \sim \mathbb{P}$ ,  $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$ , where  $\mathcal{P}_0(\mathcal{Z})$  represents the set of all possible distributions with support  $\mathcal{Z} \subseteq \mathbb{R}^n$ . For a composite random variable,  $f(\tilde{\mathbf{z}})$ ,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , we use  $\mathbb{E}_{\mathbb{P}}[f(\tilde{\mathbf{z}})]$  to denote expectation of the composite random variable over the distribution  $\tilde{\mathbf{z}} \sim \mathbb{P}$ . Finally,  $\mathbf{0}$  ( $\mathbf{1}$ ) denotes the vector of all zeros (ones) and its dimension should be clear from the context, while the identity matrix of order  $n$  is denoted by  $\mathbf{I}_n$ .

## 2. Mitigating ambiguity in data-driven optimization

We consider a data-driven risk-based minimization problem explored in Mohajerin Esfahani and Kuhn (2018), where we denote by  $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$  a vector of decision variables and  $g(\mathbf{x}, \mathbf{z})$  is the *evaluation function* associated with a vector of input parameters  $\mathbf{z} \in \mathcal{Z} \subseteq \mathbb{R}^{n_z}$  that is subject to uncertainty. We denote  $\tilde{\mathbf{z}}$  as the corresponding random variable with an underlying probability distribution  $\mathbb{P}^* \in \mathcal{P}_0(\mathcal{Z})$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}^*$ . Hence, for a given decision  $\mathbf{x} \in \mathcal{X}$ , the evaluation function  $g(\mathbf{x}, \tilde{\mathbf{z}})$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}^*$ , is a composite random variable. Ideally, we would like to determine the optimal solution to the following optimization problem,

$$Z^* = \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}^*} [g(\mathbf{x}, \tilde{\mathbf{z}})], \quad (1)$$

where the objective function is the expectation of the evaluation function. However, the underlying distribution  $\mathbb{P}^*$  is not observable. Instead, we have access to  $\Omega$  samples and for each  $\omega \in [\Omega]$ ,  $\hat{\mathbf{z}}^\omega$  denotes an independent realization of the random variable  $\tilde{\mathbf{z}}$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}^*$ . In situations of limited data availability, it is well possible to have  $\Omega = 1$ . In which case, using the terminology in robust optimization,  $\hat{\mathbf{z}}^1$  would represent the nominal values of the input parameters. We also denote the empirical distribution by  $\hat{\mathbb{P}} \in \mathcal{P}_0(\mathcal{Z})$ ,  $\tilde{\mathbf{z}} \sim \hat{\mathbb{P}}$ , such that  $\hat{\mathbb{P}}[\tilde{\mathbf{z}} = \hat{\mathbf{z}}^\omega] = 1/\Omega$ ,  $\omega \in [\Omega]$ .

To set our work apart from Mohajerin Esfahani and Kuhn (2018), we focus on an evaluation function that is non-piecewise-linear and bi-convex in  $(\mathbf{x}, \mathbf{z})$ . As we will show, our framework is a generalization and could address a broader class of problems such as those arising in risk management, for which the evaluation function is associated with a risk-adverse decision maker that can be expressed as

$$g(\mathbf{x}, \mathbf{z}) = \ell(f(\mathbf{x}, \mathbf{z})),$$

where  $f(\mathbf{x}, \mathbf{z})$  is the underlying cost function and  $\ell: \mathbb{R} \rightarrow \mathbb{R}$  is a convex increasing disutility function. For instance, the exponential disutility, which is far more commonly adopted than other types of disutility functions combined (cor 1995), has the following representation

$$\ell_a(w) = \begin{cases} (\exp(aw) - 1)/a & \text{if } a > 0, \\ w & \text{if } a = 0 \end{cases}$$

where  $a$  is known as the Arrow-Pratt measure of risk-aversion. Note that the disutility function is normalized such that  $\ell_a(0) = 0$  and  $1 \in \partial\ell_a(0)$ .

The same framework can also be applied to solving linear regression and classification problem (see, for instance, Shafieezadeh-Abadeh et al. 2019), where we predict the scalar response variable  $\tilde{z}_{n+1}$  using an affine function of  $n$  explanatory variables  $\tilde{z}_1, \dots, \tilde{z}_n$ . The evaluation function relates to the prediction loss, *i.e.*,

$$g(\mathbf{x}, \mathbf{z}) = \ell \left( \left( x_{n+1} + \sum_{i \in [n]} x_i z_i \right) - z_{n+1} \right),$$

where  $\ell: \mathbb{R} \rightarrow \mathbb{R}_+$  is a non-piecewise-linear and convex loss function such as

1. Quadratic loss:  $\ell(w) = w^2$ .
2. Huber loss with parameter  $\delta > 0$ :  $\ell(w) = \frac{w^2}{2}$  if  $|w| \leq \delta$ ; otherwise,  $\ell(w) = \delta(|w| - \frac{\delta}{2})$ .
3. Squared hinge loss:  $\ell(w) = (\max\{w, 0\})^2$ .
4. Logexp loss:  $\ell(w) = \log(1 + \exp(w))$ .

## 2.1. Empirical and stochastic optimization

Since  $\mathbb{P}^*$  is unobservable, a reasonable approach is to solve the *empirical optimization* model as follows

$$Z_0 = \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\hat{\mathbb{P}}} [g(\mathbf{x}, \tilde{\mathbf{z}})] = \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{\Omega} \sum_{\omega \in [\Omega]} g(\mathbf{x}, \hat{\mathbf{z}}^\omega). \quad (2)$$

ASSUMPTION 1. *We assume that the empirical optimization problem (1) is solvable, i.e., there exists  $\hat{\mathbf{x}} \in \mathcal{X}$  such that*

$$Z_0 = \frac{1}{\Omega} \sum_{\omega \in [\Omega]} g(\hat{\mathbf{x}}, \hat{\mathbf{z}}^\omega).$$

More generally, in stochastic optimization, we would assume a distribution  $\mathbb{P}^\dagger \in \mathcal{P}_0(\mathbb{R}_z^n)$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}^\dagger$  in which the parameters to characterize the distribution  $\mathbb{P}^\dagger$  is determined from the empirical distribution,  $\hat{\mathbb{P}}$ . For instance, it is common to assume multivariate normal distribution, *i.e.*,  $\mathbb{P}^\dagger = \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  where  $\hat{\boldsymbol{\mu}} = \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\mathbf{z}}]$  and  $\hat{\boldsymbol{\Sigma}} = \mathbb{E}_{\hat{\mathbb{P}}}[(\tilde{\mathbf{z}} - \hat{\boldsymbol{\mu}})(\tilde{\mathbf{z}} - \hat{\boldsymbol{\mu}})^\top]$ . Subsequently, stochastic optimization solves the following optimization problem,

$$Z^\dagger = \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbb{P}^\dagger} [g(\mathbf{x}, \tilde{\mathbf{z}})], \quad (3)$$

which has an objective function that is much harder to evaluate compare to the empirical optimization problem. Nevertheless, *sample average approximation* (SAA) can be used to approximate the objective function of Problem (3) by solving the following optimization problem

$$Z^{\dagger, S} = \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{S} \sum_{s \in [S]} g(\mathbf{x}, \mathbf{z}^{\dagger, s}), \quad (4)$$

where  $\mathbf{z}^{\dagger, s}$ ,  $s \in [S]$  are  $S$  samples independently generated from the distribution  $\mathbb{P}^\dagger$ . We stress that SAA differs from empirical optimization because in the latter, we only have access to the available empirical samples, but not the underlying probability distribution that would enable us to generate an arbitrary number of independent samples. Moreover, it should be noted that with sufficient computational resources, SAA could at best obtain the optimum solution to Problem (3), which is not the same as the actual optimum solution of Problem (1). For instance, if  $g(\mathbf{x}, \mathbf{z})$  is affine in the uncertain parameters, then

$$\mathbb{E}_{\mathbb{P}^\dagger} [g(\mathbf{x}, \tilde{\mathbf{z}})] = g(\mathbf{x}, \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\mathbf{z}}]) = \mathbb{E}_{\hat{\mathbb{P}}} [g(\mathbf{x}, \tilde{\mathbf{z}})],$$

and, in this case, solving the empirical optimization problem would yield the same solution as the exact stochastic optimization problem.

Smith and Winkler (2006) observe that if we solve a stochastic optimization problem using parameters estimated from the empirical distribution, due to the *Optimizer's curse*, we should also expect inferior out-of-sample results. In the context of regression and classification, solving the the empirical optimization problem is similar to solving the problem without regularization, which can lead to poor out-of-sample performance due to overfitting.

## 2.2. Data-driven robust optimization

To mitigate the *Optimizer's curse*, Mohajerin Esfahani and Kuhn (2018) propose the seminal data driven robust optimization model, which solves the following optimization problem

$$Z_r = \min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{F}(r)} \mathbb{E}_{\mathbb{P}} [g(\mathbf{x}, \tilde{\mathbf{z}})], \quad (5)$$

for a given ambiguity set that considers a family of probability distributions within the vicinity of  $\hat{\mathbb{P}}$  given by

$$\mathcal{F}(r) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \left| \begin{array}{l} \tilde{\mathbf{z}} \sim \mathbb{P} \\ \Delta_q(\mathbb{P}, \hat{\mathbb{P}}) \leq r \end{array} \right. \right\},$$

where

$$\Delta_q(\mathbb{P}, \hat{\mathbb{P}}) := \inf_{\mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2)} \left\{ (\mathbb{E}_{\mathbb{Q}} [\|\tilde{\mathbf{z}} - \tilde{\mathbf{v}}\|^q])^{1/q} \mid (\tilde{\mathbf{z}}, \tilde{\mathbf{v}}) \sim \mathbb{Q}, \tilde{\mathbf{z}} \sim \mathbb{P}, \tilde{\mathbf{v}} \sim \hat{\mathbb{P}} \right\},$$

is the type- $q$  Wasserstein distance metric, for  $q \geq 1$ . The consideration of the Wasserstein distance metric is inspired by the following statistical property of how likely the unobservable true distribution  $\mathbb{P}^*$  is within the proximity of  $\hat{\mathbb{P}}$  as follows:

**THEOREM 1. (Theorem 2 of Fournier and Guillin 2015).** *Let  $\mathbb{P}^\Omega$  denote the distribution that governs the distribution of the independent samples  $\tilde{\mathbf{z}}^1, \dots, \tilde{\mathbf{z}}^\Omega$  drawn from  $\mathbb{P}^*$  for which the empirical distribution  $\hat{\mathbb{P}}$  is constructed. When the true data-generating distribution  $\mathbb{P}^*$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}^*$  is a light-tailed distribution such that  $\mathbb{E}_{\mathbb{P}^*}[\exp(\|\tilde{\mathbf{z}}\|^\alpha)] < \infty$  for some  $\alpha > 1$ , then for all  $r \geq 0$ ,*

$$\mathbb{P}^\Omega \left[ \Delta_q(\mathbb{P}^*, \hat{\mathbb{P}}) > r \right] \leq \varepsilon_q(r)$$

for some function  $\varepsilon_q(r) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  that decreases to zero at an exponential rate in  $r > 0$ .

Mohajerin Esfahani and Kuhn (2018) derive an equivalent minimax reformulation of Problem (5) when  $q = 1$ . Such a reformulation can be readily extended to other  $q \geq 1$ , *i.e.*,

$$\begin{aligned} Z_r = \min & \frac{1}{\Omega} \sum_{\omega \in [\Omega]} \sup_{\mathbf{z} \in \mathcal{Z}} \{g(\mathbf{x}, \mathbf{z}) - k(\|\mathbf{z} - \hat{\mathbf{z}}^\omega\|^q - r^q)\} \\ \text{s.t. } & \mathbf{x} \in \mathcal{X}, k \geq 0. \end{aligned} \tag{6}$$

Shafieezadeh-Abadeh et al. (2019) have applied the data-driven robust optimization framework to improve the solutions of linear regression and classification problems, where they have established the connections with regularization if the loss function is Lipschitz continuous and having an unbounded support  $\mathcal{Z} = \mathbb{R}^{n_z}$ . More recently, Sim et al. (2021) extend the results to loss functions derived from Lipschitz continuous functions raised to the power of  $p$ ,  $p \geq 1$ . They propose tractable type- $p$  robust optimization and satisficing models and using type- $p$  Wasserstein distance, and show their relations with  $p$ -root regularization problems when the support is unbounded. However, apart from these special cases, when the support is bounded and the underlying evaluation function is nonlinear, it remains computational challenging to solve the data-driven robust optimization problem exactly.

### 2.3. On classical robust optimization

It is interesting to note that for the extreme case of  $\Omega = 1$ , the data-driven robust optimization problem becomes

$$\begin{aligned} Z_r &= \min \sup_{\mathbf{z} \in \mathcal{Z}} \{g(\mathbf{x}, \mathbf{z}) - k(\|\mathbf{z} - \hat{\mathbf{z}}^1\|^q - r^q)\} \\ \text{s.t. } &\mathbf{x} \in \mathcal{X}, k \geq 0, \end{aligned} \quad (7)$$

which has semblance to the classical stochastic-free robust optimization problem,

$$\underline{Z}_r = \min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{z} \in \mathcal{U}_r} \{g(\mathbf{x}, \mathbf{z})\} = \min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{z} \in \mathcal{Z}} \min_{k \geq 0} \{g(\mathbf{x}, \mathbf{z}) - k(\|\mathbf{z} - \hat{\mathbf{z}}^1\|^q - r^q)\} \quad (8)$$

where  $\hat{\mathbf{z}}^1$  is the nominal value and  $\mathcal{U}_r = \{\mathbf{z} \in \mathcal{Z} \mid \|\mathbf{z} - \hat{\mathbf{z}}^1\| \leq r\}$  is the norm-based uncertainty set. Hence, by weak duality  $\underline{Z}_r \leq Z_r$ , and by strong duality the two coincide when  $g(\mathbf{x}, \mathbf{z})$  is also concave in  $\mathbf{z}$ .

To illustrate the distinction between the data-driven robust optimization with  $\Omega = 1$  sample and classical robust optimization, we consider an uncertainty quantification problem with a univariate  $z$ . Supposing that  $q = 2$ ,  $g(z) = (z - 1)^2$  and  $\mathcal{Z} = \mathbb{R}_+$ , we have

$$\underline{Z}_1 = \max_{z \in [0,1]} \{(z - 1)^2\} = 1,$$

whereas

$$\begin{aligned} Z_1 &= \min_{k \geq 0} \max_{z \geq 0} \{(z - 1)^2 - k(z^2 - 1)\} \\ &= \min_{k \geq 0, t} \{t \mid t \geq (z - 1)^2 - k(z^2 - 1) \forall z \geq 0\} \\ &= \min_{k \geq 0, t} \{t \mid (k - 1)z^2 + 2z + (t - 1 - k) \geq 0 \forall z \geq 0\}. \end{aligned}$$

For this optimization problem characterizing  $Z_1$ , it necessarily holds that  $k - 1$  and  $t - 1 - k$  are both non-negative, and thus  $Z_1 \geq 2 > \underline{Z}_1$ .

While classical robust optimization models are well justified in engineering problems such as trust topology design Ben-Tal and Nemirovski (1998), it may not necessarily be the best approach in mitigating risks in management inspired optimization problems. In the context of risk management, where  $g(\mathbf{x}, \mathbf{z}) = \ell(f(\mathbf{x}, \mathbf{z}))$ , we note that since  $\ell$  is an increasing function, we have

$$\underline{Z}_r = \ell \left( \min_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{z} \in \mathcal{U}_r} \{f(\mathbf{x}, \mathbf{z})\} \right),$$

which, unlike Problem (7), the optimal robust solution of Problem (8) would not be dependent on the disutility function at all. *We emphasize that, as far as we know, the statistical property of the robust solution has only been established for when the evaluation function is affine (Bertsimas and Sim 2004, Ben-Tal and Nemirovski 2000) or concave (Bertsimas et al. 2018) in its uncertain parameters.* For instance, under the assumption that the random variable  $\tilde{\mathbf{z}}$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}$ ,  $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$ , with  $\mathcal{Z} = [-1, 1]^{n_z}$ ,



$\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \mathbf{0}$  its components are independently distributed, Bertsimas and Sim (2004) show that if  $g(\mathbf{x}, \mathbf{z})$  is affine in  $\mathbf{z}$ , by choosing  $r = c\sqrt{n_z}$ , then

$$\mathbb{P}\left[g(\mathbf{x}, \tilde{\mathbf{z}}) \leq \max_{\mathbf{z} \in \mathcal{U}_r} g(\mathbf{x}, \mathbf{z})\right] > 1 - \exp(-c^2/2).$$

However, it is important to note that the bound can be rather weak and may not apply if the evaluation function  $g(\mathbf{x}, \mathbf{z})$  is not affine in the uncertain parameters. In the extreme case, if the function is  $g(\mathbf{x}, \mathbf{z}) = \|\mathbf{z}\|_1$ , then we would have

$$\mathbb{P}\left[g(\mathbf{x}, \tilde{\mathbf{z}}) \leq \max_{\mathbf{z} \in \mathcal{U}_r} g(\mathbf{x}, \mathbf{z})\right] = \mathbb{P}\left[\|\tilde{\mathbf{z}}\|_1 \leq \max_{\mathbf{z} \in \mathcal{Z}: \|\mathbf{z}\|_1 \leq r} \|\mathbf{z}\|_1\right] = \mathbb{P}[\tilde{\mathbf{z}} \in \mathcal{U}_r],$$

which is the minimal probability guarantee assured by the robust optimization solution. However, such assurance may not be adequate for  $r = c\sqrt{n_z}$  as demonstrated in the following result.

**PROPOSITION 1.** *For each  $i \in [n_z]$ , let  $\tilde{z}_i$  be an independently distributed random variable with support  $[-1, 1]$ , zero mean and variance  $\theta > 0$ . For a given budgeted uncertainty set  $\mathcal{U}_r$ ,  $r = c\sqrt{n_z}$  and  $c \in (0, \theta\sqrt{n_z})$ , then*

$$\mathbb{P}[\tilde{\mathbf{z}} \in \mathcal{U}_r] \leq \exp\left(-\frac{(\theta\sqrt{n_z} - c)^2}{2\theta}\right). \quad (9)$$

*Proof.* The proof is relegated to Appendix A.

Hence, the probability that the realization of uncertain parameter  $\tilde{\mathbf{z}}$  lies in the uncertainty set  $\mathcal{U}_r$ ,  $r = c\sqrt{n_z}$  decreases exponentially to zero as the dimension  $n_z$  increases.

Therefore, one should exercise caution in using the budgeted uncertainty set beyond evaluation functions that are affine in the uncertain parameters because its probability bound can become ineffective as  $n_z$  increases. In contrast, the finite sample probabilistic guarantee that motivates the data-driven robust optimization model does not depend on the nature of the evaluation function.

#### 2.4. Data-driven robust satisficing

A target based data driven robust satisficing model has recently been proposed by Long et al. (2022), which we also generalize to type- $q$  Wasserstein distance metric. For a given target  $\tau \geq Z_0$ , we solve the following optimization problem

$$\begin{aligned} & \text{minimize} && k \\ & \text{subject to} && \mathbb{E}_{\mathbb{P}}[g(\mathbf{x}, \tilde{\mathbf{z}})] \leq \tau + k(\Delta_q(\mathbb{P}, \hat{\mathbb{P}}))^q \quad \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\ & && \mathbf{x} \in \mathcal{X}, k \geq 0, \end{aligned} \quad (10)$$

As a consequence of Theorem 1, the feasible solution of the robust satisficing problem would satisfy

$$\mathbb{P}^\Omega [\mathbb{E}_{\mathbb{P}^*} [g(\mathbf{x}, \tilde{\mathbf{z}})] > \tau + kr^q] \leq \mathbb{P}^\Omega [\Delta_q(\mathbb{P}^*, \hat{\mathbb{P}}) > r] \leq \varepsilon_q(r),$$

for all  $r \geq 0$ , justifying the lower value of  $k$ , the higher the level of confidence that the actual objective function evaluated on the true distribution is within  $kr^q + \tau$ , such that the probability of exceeding the confidence range decreases to zero at an exponential rate in  $r$ , accordingly to the measure concentration result of Fournier and Guillin (2015); see *e.g.* Theorem 1.

Observe that

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}} [g(\mathbf{x}, \tilde{\mathbf{z}})] \leq \tau + k(\Delta_q(\mathbb{P}, \hat{\mathbb{P}}))^q && \forall \mathbb{P} \in \mathcal{P}_0(\mathcal{Z}) \\ \Leftrightarrow & \mathbb{E}_{\mathbb{Q}} [g(\mathbf{x}, \tilde{\mathbf{z}})] \leq \tau + k\mathbb{E}_{\mathbb{Q}} [\|\tilde{\mathbf{z}} - \tilde{\mathbf{v}}\|^q] && \forall \mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2) : (\tilde{\mathbf{z}}, \tilde{\mathbf{v}}) \sim \mathbb{Q}, \tilde{\mathbf{v}} \sim \hat{\mathbb{P}} \\ \Leftrightarrow & \mathbb{E}_{\mathbb{Q}} [g(\mathbf{x}, \tilde{\mathbf{z}}) - k\|\tilde{\mathbf{z}} - \tilde{\mathbf{v}}\|^q] \leq \tau && \forall \mathbb{Q} \in \mathcal{P}_0(\mathcal{Z}^2) : (\tilde{\mathbf{z}}, \tilde{\mathbf{v}}) \sim \mathbb{Q}, \tilde{\mathbf{v}} \sim \hat{\mathbb{P}} \\ \Leftrightarrow & \frac{1}{\Omega} \sum_{\omega \in [\Omega]} \sup_{\mathbb{P}' \in \mathcal{P}_0(\mathcal{Z})} \mathbb{E}_{\mathbb{P}'} [g(\mathbf{x}, \tilde{\mathbf{z}}) - k\|\tilde{\mathbf{z}} - \hat{\mathbf{z}}^\omega\|^q] \leq \tau \\ \Leftrightarrow & \frac{1}{\Omega} \sum_{\omega \in [\Omega]} \sup_{\mathbf{z} \in \mathcal{Z}} \{g(\mathbf{x}, \mathbf{z}) - k\|\mathbf{z} - \hat{\mathbf{z}}^\omega\|^q\} \leq \tau. \end{aligned}$$

Hence, the data-driven robust satisficing problem can be expressed as the following robust optimization problem,

$$\begin{aligned} & \text{minimize} && k \\ & \text{subject to} && \frac{1}{\Omega} \sum_{\omega \in [\Omega]} \sup_{\mathbf{z} \in \mathcal{Z}} \{g(\mathbf{x}, \mathbf{z}) - k\|\mathbf{z} - \hat{\mathbf{z}}^\omega\|^q\} \leq \tau \\ & && \mathbf{x} \in \mathcal{X}, k \geq 0, \end{aligned} \tag{11}$$

which is computationally equivalent to the data-driven robust optimization model. In the numerical studies of Long et al. (2022), the family of solutions generated by the robust satisficing model would dominate of those obtained by solving the data-driven the robust optimization models. On solving supervised learning problems, Sim et al. (2021) demonstrate in studies on well-known data sets that the robust satisficing approach would generally outperform the robust optimization approach, when the respective target and size parameters are determined via cross-validation. Hence, we will focus on solving robust satisficing model. Observe that we can express Problem (11) as

$$\begin{aligned} & \text{minimize} && k \\ & \text{subject to} && \frac{1}{\Omega} \sum_{\omega \in [\Omega]} v_\omega \leq \tau \\ & && (\mathbf{x}, v_\omega, k) \in \mathcal{Y}^\omega \quad \forall \omega \in [\Omega] \\ & && \mathbf{x} \in \mathcal{X}, \mathbf{v} \in \mathbb{R}^\Omega, k \in \mathbb{R}_+, \end{aligned} \tag{12}$$

where

$$\mathcal{Y}^\omega \triangleq \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \mid \sup_{\mathbf{z} \in \mathcal{Z}^\omega} \{g^\omega(\mathbf{x}, \mathbf{z}) - kp(\mathbf{z})\} \leq v \right\},$$

$\mathcal{Z}^\omega \triangleq \{\mathbf{z} \in \mathbb{R}^{n_z} \mid \mathbf{z} + \hat{\mathbf{z}}^\omega \in \mathcal{Z}\}$ ,  $g^\omega(\mathbf{x}, \mathbf{z}) \triangleq g(\mathbf{x}, \mathbf{z} + \hat{\mathbf{z}}^\omega)$  and  $p(\mathbf{z}) \triangleq \|\mathbf{z}\|^q$ . Note that under this representation, we have  $\mathbf{0} \in \mathcal{Z}_\omega$  for all  $\omega \in [\Omega]$  and  $p(\mathbf{0}) = 0$ . The set  $\mathcal{Y}^\omega$  is typically computationally intractable to characterize exactly. Long et al. (2022) focus on piecewise-linear and bi-convex evaluation functions and their proposed approach to solve the robust satisficing model does not extend beyond linearity.

To extend the scope of robust satisficing, we will replace  $\mathcal{Y}^\omega$ , with a computationally tractable convex set,  $\bar{\mathcal{Y}}^\omega \subseteq \mathcal{Y}^\omega$ , for all  $\omega \in [\Omega]$ , in the following optimization problem,

$$\begin{aligned} & \text{minimize} && k \\ & \text{subject to} && \frac{1}{\Omega} \sum_{\omega \in [\Omega]} v_\omega \leq \tau \\ & && (\mathbf{x}, v_\omega, k) \in \bar{\mathcal{Y}}^\omega \quad \forall \omega \in [\Omega] \\ & && \mathbf{x} \in \mathcal{X}, \mathbf{v} \in \mathbb{R}^\Omega, k \in \mathbb{R}_+. \end{aligned} \tag{13}$$

As Problem (13) is a conservative approximation of Problem (10), a key challenge in approximating the robust satisficing model is to choose  $\bar{\mathcal{Y}}^\omega$ ,  $w \in [\Omega]$  such that Problem (13) remains feasible for any appropriately chosen target  $\tau > Z_0$ , where we recall that  $Z_0$  denotes the optimal objective value of the empirical optimization model (1).

DEFINITION 1. We say Problem (13) is a *tractable safe approximation* of the data-driven robust satisficing problem (12) if for all  $\omega \in [\Omega]$  the following properties of  $\bar{\mathcal{Y}}^\omega$  hold:

1.  $\bar{\mathcal{Y}}^\omega$  is a computationally tractable convex set.
2.  $\bar{\mathcal{Y}}^\omega \subseteq \mathcal{Y}^\omega$ .
3. The set  $\{\mathbf{x} \in \mathcal{X} \mid (\mathbf{x}, v, k) \in \bar{\mathcal{Y}}^\omega\}$  is nondecreasing in  $v$  and  $k$ .
4. For any  $\mathbf{x} \in \mathcal{X}$  and  $v > g^\omega(\mathbf{x}, \mathbf{0})$ , there exists a finite  $\bar{k} \geq 0$  such that  $(\mathbf{x}, v, \bar{k}) \in \bar{\mathcal{Y}}^\omega$ .

PROPOSITION 2. Under Assumption 1, if Problem (13) is a tractable safe approximation of Problem (12), then Problem (13) is feasible for all  $\tau > Z_0$ .

*Proof.* The proof is relegated to Appendix A.

### 3. Tractable formulations and safe approximations

In this section, we focus on solving the data-driven robust satisficing problem (11), either exactly or approximately, while ensuring that the problem would remain feasible for all appropriate chosen targets  $\tau > Z_0$ . As the consequence of Proposition 2, we will derive either exact or tractable safe approximation of the the data-driven robust satisficing problem (11) by focusing on the set

$$\mathcal{Y} \triangleq \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \mid \sup_{\mathbf{z} \in \mathcal{Z}} \{g(\mathbf{x}, \mathbf{z}) - kp(\mathbf{z})\} \leq v \right\}, \tag{14}$$

for a given convex support set  $\mathcal{Z}$  that contains the origin,  $\mathbf{0} \in \mathcal{Z}$ , a non-piecewise-linear and biconvex evaluation function  $g(\mathbf{x}, \mathbf{z})$  and a convex non-negative penalty function that satisfies  $p(\mathbf{z}) = 0$  iff  $\mathbf{z} = \mathbf{0}$ . Note that  $p(\mathbf{z})$  is not necessarily a power norm and that we have removed the superscript  $\omega$  to simply our exposition. Indeed, how we can construct the tractable convex set  $\bar{\mathcal{Y}} \subseteq \mathcal{Y}$  that satisfies the properties of Definition 1, would depend on the functions  $g$  and  $p$  as well the support set  $\mathcal{Z}$ . While robust counterparts and their tractable approximations have been well studied in the literature, the key challenge is to ensure that for any  $\mathbf{x} \in \mathcal{X}$  and  $v > g(\mathbf{x}, \mathbf{0})$ , there exists a finite  $k \in \mathbb{R}_+$  such that  $(\mathbf{x}, v, k) \in \bar{\mathcal{Y}}$ .

For a biconvex quadratic evaluation function, a quadratic penalty function and an ellipsoidal uncertainty set, we derive the exact reformulation in the form of a tractable semidefinite optimization problem. Then, for a more general conic representable function with polyhedral support and penalty, we focus on obtaining a tractable safe approximation using *affine dual recourse adaptation* technique.

### 3.1. Quadratic evaluation and penalty functions

This problem is motivated from the classical robust optimization proposed in Ben-Tal and Nemirovski (1998), which involves the following quadratic biconvex evaluation function,

$$g(\mathbf{x}, \mathbf{z}) = \|\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{a}(\mathbf{x})\|_2^2 + \mathbf{b}(\mathbf{x})^\top \mathbf{z} + c(\mathbf{x}),$$

where  $\mathbf{a} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_a}$ ,  $\mathbf{A} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_a \times n_z}$ ,  $\mathbf{b} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_z}$ ,  $c : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$  are affine mappings of the decision variables  $\mathbf{x}$ . This is perhaps the only meaningful nonlinear example in which the robust satisficing problem is tractable and can be formulated exactly as a convex optimization problem without the need for safe approximation. To do so, the penalty function has to be quadratic, and we focus on a squared Euclidean norm,  $p(\mathbf{z}) = \|\mathbf{z}\|_2^2 = \mathbf{z}^\top \mathbf{z}$ . We can also consider an ellipsoidal support set  $\mathcal{E}_r = \{\mathbf{z} \in \mathbb{R}^{n_z} \mid \|\mathbf{z}\|_2 \leq r\}$ , for  $r > 0$ .

PROPOSITION 3. *The set*

$$\mathcal{Y} = \{(\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \mid \|\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{a}(\mathbf{x})\|_2^2 + \mathbf{b}(\mathbf{x})^\top \mathbf{z} + c(\mathbf{x}) \leq k\mathbf{z}^\top \mathbf{z} + v, \forall \mathbf{z} \in \mathcal{E}_r\}$$

is equivalent to the following

$$\left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \mid \begin{array}{l} \exists \lambda \geq 0 : \\ \begin{bmatrix} \mathbf{I}_{n_a} & \mathbf{a}(\mathbf{x}) & \mathbf{A}(\mathbf{x}) \\ \mathbf{a}(\mathbf{x})^\top & -c(\mathbf{x}) + v - \lambda r^2 & -\frac{1}{2}\mathbf{b}(\mathbf{x})^\top \\ \mathbf{A}(\mathbf{x})^\top & -\frac{1}{2}\mathbf{b}(\mathbf{x}) & (k + \lambda)\mathbf{I}_{n_z} \end{bmatrix} \succeq \mathbf{0} \end{array} \right\}.$$

Moreover the set is feasible for any  $\mathbf{x} \in \mathcal{X}$ ,  $v > g(\mathbf{x}, \mathbf{0}) = \|\mathbf{a}(\mathbf{x})\|_2^2 + c(\mathbf{x})$ .

*Proof.* The proof follows from Ben-Tal and Nemirovski (1998) and is relegated to Appendix A.

For a more general robust satisficing problem, exact tractable reformulations may not exist, and thus approximations would be needed. We next introduce a class of conic representable evaluation functions for which we can derive a tractable safe approximation.

### 3.2. Conic representable evaluation function

To represent the non-piecewise-linear and biconvex objective function, we consider a rather general conic representable evaluation function  $g$  of the form

$$\begin{aligned} g(\mathbf{x}, \mathbf{z}) = \text{minimize} \quad & \mathbf{d}^\top \mathbf{y} \\ \text{subject to} \quad & \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \\ & \mathbf{y} \in \mathbb{R}^{n_y}, \end{aligned} \tag{15}$$

where  $\mathbf{f}: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_f}$ ,  $\mathbf{F}: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_f \times n_z}$  are affine mappings in  $\mathbf{x}$ , and  $\mathcal{K}$  is a proper cone, *i.e.*, a full dimensional, closed and pointed convex cone. Since the decision variable  $\mathbf{y}$  in Problem (15) is made after observing  $\mathbf{x}$  and  $\mathbf{z}$ , we will call  $\mathbf{y}$  the *recourse* variable. Indeed, if the cone  $\mathcal{K}$  is the non-negative orthant, then the  $g$  function would represent the second stage optimization problem of a standard two-stage stochastic optimization model. The function  $g$  is sufficiently broad enough to represent many types of functions considered in the robust optimization literature, including biconvex linear or quadratic constraint functions (Ben-Tal and Nemirovski 1998), non-linear biconvex constraint functions (Roos et al. 2020, Zhen et al. 2017), adaptive linear optimization model (Ben-Tal et al. 2004, Bertsimas and de Ruiter 2016) and adaptive convex optimization model (de Ruiter et al. 2018).

**ASSUMPTION 2 (Complete and bounded recourse).** *We assume that for any  $\mathbf{v} \in \mathbb{R}^{n_f}$ , there exists a  $\mathbf{y} \in \mathbb{R}^{n_y}$  such that  $\mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{v}$ . Moreover, there does not exist  $\mathbf{y} \in \mathbb{R}^{n_y}$  such that  $\mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{0}$  and  $\mathbf{d}^\top \mathbf{y} < 0$ . These conditions ensure that Problem (15) is always finite and strictly feasible.*

We remark that the term *complete recourse* is an extension of the same term used in stochastic linear optimization (see, *e.g.*, Birge and Louveaux 2011) to our model where the second stage problem is a conic optimization problem. It ensures that Problem (15) is strictly feasible. To see the strict feasibility, let  $\mathbf{v} \in \mathbb{R}^{n_f}$  such that  $\mathbf{v} \succ_{\mathcal{K}} \mathbf{0}$ , and under complete recourse, there exists a  $\mathbf{y} \in \mathbb{R}^{n_y}$  such that  $\mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} + \mathbf{v} \succ_{\mathcal{K}} \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z}$ . The bounded recourse condition ensures that the  $g$  function of Problem (15) is always finite.

As we will later show, we require Assumption 2 to obtain a tractable safe approximation of the robust satisficing problem. However, it is not necessarily clear how we can transform a nonlinear evaluation function to satisfy the assumption.

PROPOSITION 4. Consider a differentiable convex loss function,  $\ell : \mathbb{R}^{n_w} \rightarrow \mathbb{R}$  whose epigraph  $\text{epi}(\ell) \triangleq \{(\mathbf{w}, v) \in \mathbb{R}^{n_w} \times \mathbb{R} \mid \ell(\mathbf{w}) \leq v\}$  does not contain a line. The perspective cone of  $\ell$ , defined as

$$\mathcal{K}_\ell \triangleq \{(\mathbf{w}, v_1, v_2) \in \mathbb{R}^{n_w} \times \mathbb{R} \times \mathbb{R}_+ \mid v_2 \ell(\mathbf{w}/v_2) \leq v_1\},$$

where

$$0\ell(\mathbf{w}/0) \triangleq \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla \ell(\mathbf{u})^\top \mathbf{w}\}.$$

is a proper cone, i.e., closed, convex, solid and pointed.

*Proof.* It is not difficult to see that  $\mathcal{K}_\ell$  is a cone that is closed and solid. We will next show that the cone  $\mathcal{K}_\ell$  is also convex, and pointed. Since  $\ell$  is convex and differentiable, then for any  $\mathbf{u}, \mathbf{w} \in \mathbb{R}^{n_w}$ , we have

$$\ell(\mathbf{w}) \geq \nabla \ell(\mathbf{u})^\top (\mathbf{w} - \mathbf{u}) + \ell(\mathbf{u})$$

so that

$$\ell(\mathbf{w}) = \max_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla \ell(\mathbf{u})^\top (\mathbf{w} - \mathbf{u}) + \ell(\mathbf{u})\},$$

which allows us to define

$$y_2 \ell(\mathbf{w}/y_2) \triangleq \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla \ell(\mathbf{u})^\top (\mathbf{w} - y_2 \mathbf{u}) + y_2 \ell(\mathbf{u})\}, \quad (16)$$

for  $y_2 \geq 0$ . As a result, the cone can also be expressed as

$$\mathcal{K}_\ell = \{(\mathbf{w}, v_1, v_2) \in \mathbb{R}^{n_w} \times \mathbb{R} \times \mathbb{R}_+ \mid \nabla \ell(\mathbf{u})^\top (\mathbf{w} - v_2 \mathbf{u}) + v_2 \ell(\mathbf{u}) \leq v_1 \quad \forall \mathbf{u} \in \mathbb{R}^{n_w}\},$$

which constitutes an intersection of infinitely many half spaces. The cone  $\mathcal{K}_\ell$  is thus manifestly convex. We will show by contradiction that the cone is also pointed. Suppose otherwise, there would exist  $(\mathbf{w}, v_1, v_2) \neq \mathbf{0}$  such that  $(\mathbf{w}, v_1, v_2) \in \mathcal{K}_\ell$  and  $-(\mathbf{w}, v_1, v_2) \in \mathcal{K}_\ell$ . Hence, we must have  $v_2 = 0$ , and  $(\mathbf{w}, v_1) \neq \mathbf{0}$  such that

$$\begin{aligned} \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla \ell(\mathbf{u})^\top \mathbf{w}\} &\leq v_1 \\ \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{-\nabla \ell(\mathbf{u})^\top \mathbf{w}\} &\leq -v_1, \end{aligned}$$

or equivalently

$$\nabla \ell(\mathbf{u})^\top \mathbf{w} = v_1 \quad \forall \mathbf{u} \in \mathbb{R}^{n_w}.$$

Hence, for all  $k \in \mathbb{R}$ ,

$$\ell(k\mathbf{w}) = \max_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla \ell(\mathbf{u})^\top (k\mathbf{w} - \mathbf{u}) + \ell(\mathbf{u})\} = kv_1 + \max_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla \ell(\mathbf{u})^\top (-\mathbf{u}) + \ell(\mathbf{u})\} = kv_1 + \ell(\mathbf{0}),$$

which contradicts that the epigraph of  $\ell$  does not contain a line because  $(k\mathbf{w}, kv_1 + \ell(\mathbf{0})) \in \text{epi}(\ell)$  for all  $k \in \mathbb{R}$ .  $\square$

Using the perspective of convex function technique, we can express the function as a conic minimization problem as follows

$$\begin{aligned} \ell(\mathbf{w}) = \text{minimize } & y \\ \text{subject to } & \begin{bmatrix} \mathbf{0} \\ 1 \\ 0 \end{bmatrix} y \succeq_{\mathcal{K}_\ell} \begin{bmatrix} -\mathbf{w} \\ 0 \\ -1 \end{bmatrix}, \\ & y \in \mathbb{R}. \end{aligned} \tag{17}$$

We observe that given affine mappings  $\mathbf{w} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_w}$ ,  $\mathbf{W} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_w \times n_z}$ , the evaluation function  $g(\mathbf{x}, \mathbf{z}) = \ell(\mathbf{w}(\mathbf{x}) + \mathbf{W}(\mathbf{x})\mathbf{z})$  would be conic representable. However, the condition of complete recourse required in Assumption 2 does not hold because there does not exist  $y \in \mathbb{R}$  such that

$$\begin{bmatrix} \mathbf{0} \\ 1 \\ 0 \end{bmatrix} y \succeq_{\mathcal{K}_\ell} \begin{bmatrix} \mathbf{0} \\ 0 \\ 1 \end{bmatrix}.$$

In our computation study, we have verified that the direct application of the perspective of convex function approach does not lead to a feasible tractable safe approximation problem.

### 3.3. Perspective casting

To resolve the feasibility issue, we introduce a novel *perspective casting* approach to provide a conic optimization model that would satisfy Assumption 2, and that its objective would match  $\ell(\mathbf{w})$  when evaluated over an *featured domain*,  $\mathbf{w} \in \mathcal{W} \subseteq \mathbb{R}^{n_w}$ .

**DEFINITION 2.** The *perspective depth* of a differentiable convex function  $\ell : \mathbb{R}^{n_w} \rightarrow \mathbb{R}$  on a *featured domain*,  $\mathcal{W} \subseteq \mathbb{R}^{n_w}$  is defined as

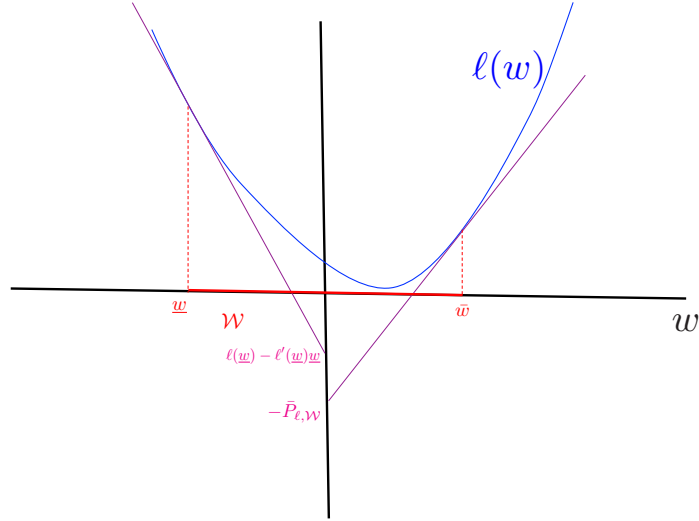
$$\bar{P}_{\ell, \mathcal{W}} \triangleq \sup_{\mathbf{u} \in \mathcal{W}} \{ \nabla \ell(\mathbf{u})^\top \mathbf{u} - \ell(\mathbf{u}) \}.$$

**PROPOSITION 5.** For a univariate differentiable convex loss function  $\ell : \mathbb{R} \rightarrow \mathbb{R}$  and a featured domain  $\mathcal{W} = [\underline{w}, \bar{w}]$ ,

$$\bar{P}_{\ell, \mathcal{W}} = \max_{v \in \{\underline{w}, \bar{w}\}} \{ \ell'(v)v - \ell(v) \}.$$

*Proof.* The proof is relegated to Appendix A.

Specking intuitively, the perspective depth refers to the distance of the closest illuminating point directly below the origin, so that the illuminating point would be visible from everywhere on the surface of the opaque epigraph of  $\ell$  that can be mapped from the featured domain,  $\mathcal{W}$  (see Figure 1).



**Figure 1** An illustration of perspective depth on  $\mathcal{W} = [w, \bar{w}]$ .

**THEOREM 2 (Perspective casting).** Consider a differentiable convex loss function  $\ell : \mathbb{R}^{n_w} \rightarrow \mathbb{R}$  whose epigraph does not contain a line, and a featured domain  $\mathcal{W} \subseteq \mathbb{R}^{n_w}$ . For a given  $P \geq \bar{P}_{\ell, \mathcal{W}}$ , let

$$\begin{aligned} \hat{\ell}(\mathbf{w}) = \text{minimize } & y_1 + P y_2 \\ \text{subject to } & \underbrace{\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_B \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \succeq_{\mathcal{K}} \begin{bmatrix} -\mathbf{w} \\ -P \\ 0 \\ 1 \end{bmatrix} \end{aligned} \quad (18)$$

$\mathbf{y} \in \mathbb{R}^2$ ,

where  $\mathcal{K} = \mathcal{K}_\ell \times \mathbb{R}_+$ . Then

$$\begin{aligned} \hat{\ell}(\mathbf{w}) &\leq \ell(\mathbf{w}) & \forall \mathbf{w} \in \mathbb{R}^{n_w}, \\ \hat{\ell}(\mathbf{w}) &= \ell(\mathbf{w}) & \forall \mathbf{w} \in \mathcal{W}. \end{aligned}$$

Moreover, Problem (18) has a complete and bounded recourse.

*Proof.* Observe that Problem (18) is equivalent to the following optimization problem

$$\hat{\ell}(\mathbf{w}) = \min_{y_2 \geq 1} \{P y_2 + y_2 \ell(\mathbf{w}/y_2) - P\},$$

hence

$$\hat{\ell}(\mathbf{w}) \leq \ell(\mathbf{w}) \quad \forall \mathbf{w} \in \mathbb{R}^{n_w}.$$



For any  $\mathbf{w} \in \mathcal{W}$ , we have

$$\begin{aligned}
\hat{\ell}(\mathbf{w}) &= \min_{y_2 \geq 1} \{Py_2 + y_2\ell(\mathbf{w}/y_2) - P\} \\
&= \min_{y_2 \geq 1} \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{Py_2 + \nabla\ell(\mathbf{u})^\top(\mathbf{w} - y_2\mathbf{u}) + y_2\ell(\mathbf{u}) - P\} \\
&\geq \min_{y_2 \geq 1} \sup_{\mathbf{u} \in \mathcal{W}} \{Py_2 + \nabla\ell(\mathbf{u})^\top(\mathbf{w} - y_2\mathbf{u}) + y_2\ell(\mathbf{u}) - P\} \\
&= \min_{y_2 \geq 1} \sup_{\mathbf{u} \in \mathcal{W}} \{y_2(P - \nabla\ell(\mathbf{u})^\top\mathbf{u} + \ell(\mathbf{u})) + \nabla\ell(\mathbf{u})^\top\mathbf{w} - P\} \\
&= \sup_{\mathbf{u} \in \mathcal{W}} \{\nabla\ell(\mathbf{u})^\top(\mathbf{w} - \mathbf{u}) + \ell(\mathbf{u})\} \\
&= \ell(\mathbf{w}),
\end{aligned}$$

where the first inequality is due to (16) and the second-to-last equality is due to having  $y_2 = 1$  at optimality because

$$P \geq \nabla\ell(\mathbf{u})^\top\mathbf{u} - \ell(\mathbf{u})$$

for all  $\mathbf{u} \in \mathcal{W}$ . Finally, to show that Problem (18) has a complete recourse, consider any  $\mathbf{w} \in \mathbb{R}^{n_w}$ ,  $\mathbf{v} \in \mathbb{R}^3$  and observe that

$$\mathbf{B} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ y_1 \\ y_2 \\ y_2 \end{bmatrix} \succeq_{\mathcal{K}} \begin{bmatrix} \mathbf{w} \\ v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

is feasible for

$$\begin{aligned}
y_1 &= v_1 + [(v_3 - v_2)^+ + 1] \ell\left(-\frac{\mathbf{w}}{(v_3 - v_2)^+ + 1}\right) \\
y_2 &= \max\{v_2, v_3\} + 1.
\end{aligned}$$

To show that the problem has a bounded recourse, observe that for any  $\mathbf{y} \in \mathbb{R}^2$ ,  $\mathbf{B}\mathbf{y} \in \mathcal{K}$ , we have  $y_2 \geq 0$  and

$$\begin{aligned}
Py_2 + y_1 &\geq Py_2 + y_2\ell(\mathbf{0}/y_2) \\
&= Py_2 + \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{\nabla\ell(\mathbf{u})^\top(\mathbf{0} - y_2\mathbf{u}) + y_2\ell(\mathbf{u})\} \\
&= Py_2 + y_2 \sup_{\mathbf{u} \in \mathbb{R}^{n_w}} \{-\nabla\ell(\mathbf{u})^\top\mathbf{u} + \ell(\mathbf{u})\} \\
&\geq Py_2 + y_2 \inf_{\mathbf{u} \in \mathcal{W}} \{-\nabla\ell(\mathbf{u})^\top\mathbf{u} + \ell(\mathbf{u})\} \\
&= \left(P - \sup_{\mathbf{u} \in \mathcal{W}} \{\nabla\ell(\mathbf{u})^\top\mathbf{u} - \ell(\mathbf{u})\}\right) y_2 \geq 0,
\end{aligned}$$

where the first equality again follows from (16). Hence, the proof is completed.  $\square$

We note that the Huber loss function with parameter  $\delta > 0$  has a finite perspective depth of  $\delta^2/2$  on its entire domain, *i.e.*,  $\mathcal{W} = \mathbb{R}$ . However, although the Huber function is Lipschitz continuous, it does not imply that any differentiable convex Lipschitz continuous function would have finite perspective depth on  $\mathcal{W} = \mathbb{R}$ , as exemplified in the following function:

$$\ell(w) = |w| - \log(|w| + 1),$$

under which  $\bar{P}_{\ell, \mathbb{R}} \geq \lim_{w \uparrow \infty} \ell'(w)w - \ell(w) = \infty$ .

EXAMPLE 1. The log-sum-exponential function commonly used in logistic regression,

$$\ell(\mathbf{w}) = \log \left( \sum_{i \in [n_w]} \exp(w_i) \right),$$

has a finite perspective depth of zero on its entire domain  $\mathcal{W} = \mathbb{R}^{n_w}$ . To see this, let

$$r(t) = \ell(\mathbf{u}t)$$

for some  $\mathbf{u} \in \mathbb{R}^{n_w}$ . Observe that

$$r'(1) - r(1) = \nabla \ell(\mathbf{u})^\top \mathbf{u} - \ell(\mathbf{u}).$$

Since  $r$  is convex and differentiable in  $t$ , from Proposition 5, we have

$$r'(1) - r(1) \leq \max \left\{ r'(0) \times 0 - r(0), \lim_{t \uparrow \infty} \{r'(t)t - r(t)\} \right\} = \max \left\{ -\log(n_w), \lim_{t \uparrow \infty} \{r'(t)t - r(t)\} \right\}.$$

However,

$$\begin{aligned} \lim_{t \uparrow \infty} \{r'(t)t - r(t)\} &= \lim_{t \uparrow \infty} \log \left( \frac{\exp \left( \frac{\sum_{i \in [n_w]} \exp(u_i t) u_i t}{\sum_{i \in [n_w]} \exp(u_i t)} \right)}{\sum_{i \in [n_w]} \exp(u_i t)} \right) \\ &= \lim_{t \uparrow \infty} \log \left( \frac{\exp \left( \frac{\sum_{i \in [n_w]} \exp((u_i - \bar{u})t) u_i t}{\sum_{i \in [n_w]} \exp((u_i - \bar{u})t)} \right)}{\exp(\bar{u}t) \sum_{i \in [n_w]} \exp((u_i - \bar{u})t)} \right) \\ &= \lim_{t \uparrow \infty} \log \left( \frac{\exp \left( \frac{\bar{n} \bar{u} t}{\bar{n}} \right)}{\bar{n} \exp(\bar{u}t)} \right) = -\log(\bar{n}), \end{aligned}$$

where  $\bar{u} = \max_{i \in [n_w]} \{u_i\}$  and  $\bar{n} = |\{i \in [n_w] | u_i = \bar{u}\}|$ .

Hence, for all  $\mathbf{u} \in \mathbb{R}^{n_w}$

$$\nabla \ell(\mathbf{u})^\top \mathbf{u} - \ell(\mathbf{u}) \leq \max\{-\log(n_w), -\log(\bar{n})\} \leq 0,$$

and that the bound is tight when  $\bar{n} = 1$ , *i.e.*,  $\mathbf{u}$  has a unique maximum.

EXAMPLE 2 (DISUTILITY FUNCTION). We can consider the disutility function of the cost associated with a two-stage linear optimization problem with complete recourse as follows

$$\begin{aligned} g(\mathbf{x}, \mathbf{z}) &= \text{minimize} && \ell(\mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y}) \\ &\text{subject to} && \mathbf{D}\mathbf{y} \geq \mathbf{a}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\mathbf{z} \\ &&& \mathbf{y} \in \mathbb{R}^{n_y}, \end{aligned}$$

for some differentiable convex and increasing function  $\ell: \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ , representing disutility, some matrix  $\mathbf{D} \in \mathbb{R}^{n_a \times n_y}$  and affine mappings  $\mathbf{a}: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_a}$ ,  $\mathbf{A}: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_a \times n_z}$ . Hence, for any  $\mathbf{v} \in \mathbb{R}^{n_a}$  there exists  $\mathbf{y} \in \mathbb{R}^{n_y}$  such that  $\mathbf{D}\mathbf{y} \geq \mathbf{v}$ . Suppose we can find  $\underline{w}$  and  $\bar{w}$  such that

$$\underline{w} \leq \ell^{-1}(g(\mathbf{x}, \mathbf{z})) \leq \bar{w} \quad \forall \mathbf{x} \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}.$$

From Proposition 5, we choose  $P \in \mathbb{R}$  such that

$$P \geq \max\{\ell'(\underline{w})\underline{w} - \ell(\underline{w}), \ell'(\bar{w})\bar{w} - \ell(\bar{w})\}.$$

Then, according to Theorem 2, we would have the following conic representation.

$$\begin{aligned} g(\mathbf{x}, \mathbf{z}) = & \text{minimize } v_1 + Pv_2 \\ & \text{subject to } \begin{bmatrix} \mathbf{d}^\top & 0 & 0 \\ \mathbf{0}^\top & 1 & 0 \\ \mathbf{0}^\top & 0 & 1 \\ \mathbf{0}^\top & 0 & 1 \\ \mathbf{D} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ v_1 \\ v_2 \end{bmatrix} \succeq_{\mathcal{K}} \begin{bmatrix} -\mathbf{c}^\top \mathbf{x} \\ -P \\ 0 \\ 1 \\ \mathbf{a}(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} \mathbf{0}^\top \\ \mathbf{0}^\top \\ \mathbf{0}^\top \\ \mathbf{0}^\top \\ \mathbf{A}(\mathbf{x}) \end{bmatrix} \mathbf{z} \\ & \mathbf{y} \in \mathbb{R}^{n_y}, \mathbf{v} \in \mathbb{R}^2 \end{aligned}$$

where  $\mathcal{K} = \mathcal{K}_\ell \times \mathbb{R}_+ \times \mathbb{R}_+^{n_a}$ .

### 3.4. Polyhedral penalty function and support

Under the conic representable evaluation function, we can express the feasible set

$$\mathcal{Y} = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \mathbf{y} \in \mathcal{R}^{n_z, n_y} : \\ \mathbf{d}^\top \mathbf{y}(\mathbf{z}) - kp(\mathbf{z}) \leq v \quad \forall \mathbf{z} \in \mathcal{Z} \\ \mathbf{B}\mathbf{y}(\mathbf{z}) \succeq_{\mathcal{K}} \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \quad \forall \mathbf{z} \in \mathcal{Z} \end{array} \right. \right\} \quad (19)$$

where  $\mathcal{R}^{m,n}$  denotes the family of all functions from  $\mathbb{R}^m$  to  $\mathbb{R}^n$ , *i.e.*,

$$\mathcal{R}^{m,n} = \{\mathbf{y} \mid \mathbf{y}: \mathbb{R}^m \rightarrow \mathbb{R}^n\}.$$

The feasible set remains intractable even if we restrict the recourse  $\mathbf{y}$  to a static function that does not depend on  $\mathbf{z}$ . As we will reveal, we overcome this challenge by using a technique of dualizing twice similar to that in Roos et al. (2020), first over the recourse variables  $\mathbf{y}$  and then over the uncertain parameters  $\mathbf{z}$  to absorb the conic nature of the original problem into a new uncertainty set while the polyhedral support and penalty show up as linear constraints of the resultant formulation. We can thus use familiar approximation methods such as affine recourse adaptation to obtain tractable feasible set  $\bar{\mathcal{Y}} \subseteq \mathcal{Y}$ . Although feasibility is not guaranteed with such approximations even under

assumptions of complete recourse (see, *e.g.*, Bertsimas et al. 2019), our results however show that whenever  $v > g(\mathbf{x}, \mathbf{0})$ , the *affine dual recourse adaptation* would always yield a feasible solution,  $(\mathbf{x}, v, k) \in \bar{\mathcal{Y}}$  for some  $k \in \mathbb{R}_+$ . To do so, we require the following assumptions on the penalty function and support,

**ASSUMPTION 3 (Polyhedral support and penalty).** *We assume the following:*

(i) *The support set  $\mathcal{Z}$  is a nonempty polyhedron*

$$\mathcal{Z} = \{\mathbf{z} \in \mathbb{R}^{n_z} \mid \mathbf{H}\mathbf{z} \leq \mathbf{h}\},$$

for some  $\mathbf{H} \in \mathbb{R}^{n_h \times n_z}$  and  $\mathbf{h} \in \mathbb{R}_+^{n_h}$ .

(ii) *The polyhedral penalty function  $p(\zeta) : \mathbb{R}^{n_z} \rightarrow \mathbb{R}_+$  can be expressed as*

$$p(\zeta) = \max_{(\lambda, \eta) \in \mathcal{V}} \{\lambda^\top \zeta - \eta\},$$

where  $\mathcal{V}$  is a polyhedral

$$\mathcal{V} = \{(\lambda, \eta) \in \mathbb{R}^{n_z} \times \mathbb{R}_+ \mid \exists \mu \in \mathbb{R}^{n_\mu} : \mathbf{M}\lambda + \mathbf{N}\mu + \mathbf{s}\eta \leq \mathbf{t}\},$$

for some  $\mathbf{M} \in \mathbb{R}^{n_m \times n_z}$ ,  $\mathbf{N} \in \mathbb{R}^{n_m \times n_\mu}$ , and  $\mathbf{s}, \mathbf{t} \in \mathbb{R}^{n_m}$ . In addition, (i)  $\mathcal{V}$  contains the origin so that  $p(\zeta) \geq 0$  and  $p(\mathbf{0}) = 0$ , (ii) there exists  $\hat{\mu} \in \mathbb{R}^{n_\mu}$  such that  $\mathbf{N}\hat{\mu} < \mathbf{t}$ , which ensures  $p(\zeta) > 0$  if  $\zeta \neq \mathbf{0}$ , and (iii)  $\mathcal{V}$  is bounded, which ensures  $p(\zeta) < \infty$ .

It is common to choose a polyhedral norm as the penalty function, which has a similar representation.

**PROPOSITION 6 (Polyhedral norm).** *Under Assumption 3, if the penalty function  $p$  is a norm, then it has the representation*

$$p(\zeta) = \max_{\lambda \in \mathbb{R}^{n_z}, \mu \in \mathbb{R}^{n_\mu}} \{\lambda^\top \zeta \mid \mathbf{M}\lambda + \mathbf{N}\mu \leq \mathbf{t}\},$$

for which

$$\{\lambda \mid \mathbf{M}\lambda + \mathbf{N}\mu \leq \mathbf{t}\} = \{\lambda \mid -\mathbf{M}\lambda + \mathbf{N}\mu \leq \mathbf{t}\}.$$

Its dual norm is given by

$$p^*(\zeta) = \min_{\mu \in \mathbb{R}^{n_\mu}, \delta \in \mathbb{R}_+} \{\delta \mid \mathbf{M}\zeta + \mathbf{N}\mu \leq \delta \mathbf{t}\}.$$

*Proof.* The proof is relegated to Appendix A.

**EXAMPLE 3 (BUDGETED NORM).** We illustrate the modeling potential of the norm-based penalty defined in Proposition 6 with a budgeted norm which computes the sum of the  $\Gamma \in \{1, \dots, n_z\}$  largest absolute components of an  $n_z$ -dimensional vector, *i.e.*,

$$p_\Gamma(\zeta) = \max_{\mathcal{S} \subseteq [n_z], |\mathcal{S}| = \Gamma} \sum_{i \in \mathcal{S}} |\zeta_i|$$

so that  $p_1(\zeta) = \|\zeta\|_\infty$  and  $p_{n_z}(\zeta) = \|\zeta\|_1$ . This can be represented as the following linear optimization problem

$$\begin{aligned} p_\Gamma(\zeta) &= \max_{\lambda \in \mathbb{R}^{n_z}} \left\{ \lambda^\top \zeta \mid \sum_{i \in [n_z]} |\lambda_i| \leq \Gamma, |\lambda_i| \leq 1, \forall i \in [n_z] \right\} \\ &= \max_{\lambda, \mu \in \mathbb{R}^{n_z}} \left\{ \lambda^\top \zeta \mid \begin{array}{l} \sum_{i \in [n_z]} \mu_i \leq \Gamma \\ \lambda_i - \mu_i \leq 0 \quad \forall i \in [n_z] \\ -\lambda_i - \mu_i \leq 0 \quad \forall i \in [n_z] \\ \mu_i \leq 1 \quad \forall i \in [n_z] \end{array} \right\}, \end{aligned}$$

which satisfies the properties of the polyhedral penalty in Assumption 3.

We also remark that since second-order conic constraints can be approximated accurately via a modest sized polyhedron (Ben-Tal and Nemirovski 2001), the representation of polyhedral penalty is quite broad and can be used to approximate many different types of convex nonlinear penalty functions such as, *inter alia*, convex polynomials and  $\ell_p$ -norms, for  $p \geq 1$ .

### 3.5. Affine dual recourse adaptation

Under Assumptions 2 and 3, we will show that the set  $\mathcal{Y}$  which is defined in (19) through a conic inequality and a polyhedral uncertainty set admits an equivalent reformulation as a set of a similar nature but with linear constraints and a conic uncertainty set. The following result is a precursor for obtaining the alternate formulation of  $\mathcal{Y}$ , which would eventually enable us to obtain a safe approximation of the robust satisficing problem via affine dual recourse adaptation.

**PROPOSITION 7.** *Under Assumption 3, for any  $\mathbf{a} \in \mathbb{R}^{n_z}$  and  $k \geq 0$ , we have*

$$\max_{z \in \mathcal{Z}} \{\mathbf{a}^\top z - kp(z)\} = \min_{\eta \in \mathbb{R}_+, \beta \in \mathbb{R}_+^{n_h}} \{\beta^\top \mathbf{h} + \eta \mid (\mathbf{a} - \mathbf{H}^\top \beta, \eta, k) \in \bar{\mathcal{V}}\}$$

where  $\bar{\mathcal{V}}$  is the perspective cone of  $\mathcal{V}$  given by

$$\bar{\mathcal{V}} = \{(\lambda, \eta, k) \in \mathbb{R}^{n_z} \times \mathbb{R}_+^2 \mid \exists \mu \in \mathbb{R}^{n_\mu} : \mathbf{M}\lambda + \mathbf{N}\mu + s\eta \leq tk\}.$$

If the penalty function  $p$  is a norm, then we have

$$\max_{z \in \mathcal{Z}} \{\mathbf{a}^\top z - kp(z)\} = \min_{\beta \in \mathbb{R}_+^{n_h}} \{\beta^\top \mathbf{h} \mid p^*(\mathbf{a} - \mathbf{H}^\top \beta) \leq k\}.$$

*Proof.* The proof is relegated to Appendix A. □

**THEOREM 3.** *Under Assumptions 2 & 3, for any  $\mathbf{x} \in \mathbb{R}^{n_x}$  and  $k \geq 0$ , we have*

$$\mathcal{Y} = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \mid \begin{array}{l} \exists \beta \in \mathcal{R}^{n_f, n_h}, \mu \in \mathcal{R}^{n_f, n_\mu}, \eta \in \mathcal{R}^{n_f, 1} : \\ \rho^\top \mathbf{f}(\mathbf{x}) + \beta(\rho)^\top \mathbf{h} + \eta(\rho) \leq v \quad \forall \rho \in \mathcal{P} \\ \mathbf{M}(\mathbf{F}(\mathbf{x})^\top \rho - \mathbf{H}^\top \beta(\rho)) + \mathbf{N}\mu(\rho) + s\eta(\rho) \leq tk \quad \forall \rho \in \mathcal{P} \\ \beta(\rho) \geq \mathbf{0}, \eta(\rho) \geq 0 \quad \forall \rho \in \mathcal{P} \end{array} \right\} \quad (20)$$

where  $\mathcal{P} = \{\boldsymbol{\rho} \in \mathcal{K}^* \mid \mathbf{B}^\top \boldsymbol{\rho} = \mathbf{d}\}$ .

*Proof.* First, it follows from (19) that

$$(\mathbf{x}, v, k) \in \mathcal{Y} \iff \max_{\mathbf{z} \in \mathcal{Z}} \min_{\mathbf{y}} \{ \mathbf{d}^\top \mathbf{y} - kp(\mathbf{z}) \mid \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \} \leq v.$$

Under complete and bounded recourse, observe that the inner minimization (over  $\mathbf{y}$ ) is strictly feasible and its objective is finite. Thus, we can transform it into a maximization problem via conic duality, that is,

$$\begin{aligned} (\mathbf{x}, v, k) \in \mathcal{Y} &\iff \max_{\mathbf{z} \in \mathcal{Z}} \max_{\boldsymbol{\rho} \in \mathcal{P}} \{ \boldsymbol{\rho}^\top (\mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z}) - kp(\mathbf{z}) \} \leq v \\ &\iff \max_{\boldsymbol{\rho} \in \mathcal{P}} \left\{ \boldsymbol{\rho}^\top \mathbf{f}(\mathbf{x}) + \max_{\mathbf{z} \in \mathcal{Z}} \{ \boldsymbol{\rho}^\top \mathbf{F}(\mathbf{x})\mathbf{z} - kp(\mathbf{z}) \} \right\} \leq v. \end{aligned}$$

Invoking Proposition 7 to transform the inner maximization (over  $\mathbf{z}$ ) to a minimization problem (over  $\boldsymbol{\beta}$ ,  $\boldsymbol{\mu}$  and  $\eta$ ) completes the proof.  $\square$

In line with the recourse variable  $\mathbf{y}$  and the uncertainty  $\mathbf{z}$  in the original definition (19) of  $\mathcal{Y}$ , we may interpret  $(\boldsymbol{\beta}, \boldsymbol{\mu}, \eta)$  and  $\boldsymbol{\rho}$  as a recourse variable and as an uncertainty, respectively. Inspired by how (20) is obtained, we shall refer to  $(\boldsymbol{\beta}, \boldsymbol{\mu}, \eta)$  as the dual recourse and to  $\mathcal{P}$  as the dual uncertainty set. Under Assumption 2, the feasible set  $\mathcal{Y}$  as defined in (19) is guaranteed to be non-empty, but it is not necessary easy to identify a primal recourse  $\mathbf{y}$  that is compatible with a feasible  $(\mathbf{x}, v, k)$ . On the other hand, we will show below that there always exists a feasible dual recourse  $(\boldsymbol{\beta}, \boldsymbol{\mu}, \eta)$  that is easy to construct. This observation will then be used to construct the following tractable subset  $\bar{\mathcal{Y}}$  of  $\mathcal{Y}$ :

$$\bar{\mathcal{Y}} = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \boldsymbol{\beta} \in \mathcal{L}^{n_f, n_h}, \boldsymbol{\mu} \in \mathcal{L}^{n_f, n_\mu}, \eta \in \mathcal{L}^{n_f, 1} : \\ \boldsymbol{\rho}^\top \mathbf{f}(\mathbf{x}) + \boldsymbol{\beta}(\boldsymbol{\rho})^\top \mathbf{h} + \eta(\boldsymbol{\rho}) \leq v \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\ \mathbf{M}(\mathbf{F}(\mathbf{x})^\top \boldsymbol{\rho} - \mathbf{H}^\top \boldsymbol{\beta}(\boldsymbol{\rho})) + \mathbf{N}\boldsymbol{\mu}(\boldsymbol{\rho}) + s\eta(\boldsymbol{\rho}) \leq tk \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\ \boldsymbol{\beta}(\boldsymbol{\rho}) \geq \mathbf{0}, \eta(\boldsymbol{\rho}) \geq 0 \quad \forall \boldsymbol{\rho} \in \mathcal{P} \end{array} \right. \right\} \quad (21)$$

where  $\mathcal{L}^{m,n}$  denotes the sub-class of functions in  $\mathcal{R}^{m,n}$  that are affinely dependent on the inputs as follows:

$$\mathcal{L}^{m,n} = \left\{ \mathbf{y} \in \mathcal{R}^{m,n} \mid \exists \boldsymbol{\pi} \in \mathbb{R}^m, \boldsymbol{\Pi} \in \mathbb{R}^{m \times n} : \mathbf{y}(\mathbf{z}) = \boldsymbol{\pi} + \boldsymbol{\Pi}\mathbf{z} \right\}.$$

**THEOREM 4.** *Under Assumptions 2 & 3,  $\bar{\mathcal{Y}}$  is a tractable safe approximation of  $\mathcal{Y}$ .*

*Proof.* For any  $\mathbf{x} \in \mathcal{X}$  and  $v > g(\mathbf{x}, \mathbf{0})$ , we first show that the solution  $\boldsymbol{\beta}(\boldsymbol{\rho}) = \mathbf{0}$ , and  $\eta(\boldsymbol{\rho}) = 0$  robustly satisfies the requirements in (21). With vanishing  $\boldsymbol{\beta}$  and  $\eta$ , we first observe that

$$\max_{\boldsymbol{\rho} \in \mathcal{P}} \{ \boldsymbol{\rho}^\top \mathbf{f}(\mathbf{x}) + \boldsymbol{\beta}(\boldsymbol{\rho})^\top \mathbf{h} + \eta(\boldsymbol{\rho}) \} = \min_{\mathbf{y}} \{ \mathbf{d}^\top \mathbf{y} \mid \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{f}(\mathbf{x}) \} = g(\mathbf{x}, \mathbf{0}) < v,$$

where the first equality is due to the strong duality (as the minimization problem is strictly feasible).

We next show that the remaining constraints can be satisfied because there exists  $\hat{k} > 0$  such that

$$\mathbf{M}\mathbf{F}(\mathbf{x})^\top \boldsymbol{\rho} + \mathbf{N}\hat{\boldsymbol{\mu}}\hat{k} \leq \hat{\mathbf{t}} \quad \forall \boldsymbol{\rho} \in \mathcal{P}. \quad (22)$$

Indeed, since  $\mathbf{N}\hat{\boldsymbol{\mu}} < \hat{\mathbf{t}}$ , it implies that the set  $\{\boldsymbol{\lambda} \mid \mathbf{M}\boldsymbol{\lambda} + \mathbf{N}\hat{\boldsymbol{\mu}} \leq \hat{\mathbf{t}}\}$  must contain the origin in its interior. Hence, there exists a norm  $\|\cdot\|$  such that

$$\{\boldsymbol{\lambda} \mid \|\boldsymbol{\lambda}\| \leq 1\} \subseteq \{\boldsymbol{\lambda} \mid \mathbf{M}\boldsymbol{\lambda} + \mathbf{N}\hat{\boldsymbol{\mu}} \leq \hat{\mathbf{t}}\}.$$

Therefore, it suffices to show that there exists a finite  $\hat{k} > 0$  such that

$$\max_{\boldsymbol{\rho} \in \mathcal{P}} \|\mathbf{F}(\mathbf{x})^\top \boldsymbol{\rho}\| \leq \hat{k}.$$

Suppose that the dual uncertainty set  $\mathcal{P}$  is unbounded for the sake of a contradiction. Then, there exists a vector  $\mathbf{v} \in \mathbb{R}^{n_f}$  such that  $\max_{\boldsymbol{\rho} \in \mathcal{P}} \boldsymbol{\rho}^\top \mathbf{v}$  is unbounded, and therefore its corresponding dual  $\min_{\mathbf{y}} \{\mathbf{d}^\top \mathbf{y} \mid \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{v}\}$  must be infeasible, which contradicts with Assumption 2. Hence,  $\mathcal{P}$  is a bounded set, and a finite and positive  $\hat{k}$  exists, which in turns implies that  $(\mathbf{x}, v, \hat{k}) \in \bar{\mathcal{Y}}$ .

Next, to show that  $\bar{\mathcal{Y}}$  is a tractable set, observe that since the dual recourse variables  $(\boldsymbol{\beta}, \boldsymbol{\mu}, \eta)$  are restricted to affine functions, we can express the requirements in (21) more compactly as

$$\exists \mathbf{L} \in \mathbb{R}^{(n_h + n_\mu + 1) \times (n_f + 1)} \quad \forall \boldsymbol{\rho} \in \mathcal{P}: \quad \boldsymbol{\gamma}(\mathbf{x}, v, k, \mathbf{L}) + \boldsymbol{\Gamma}(\mathbf{x}, v, k, \mathbf{L})\boldsymbol{\rho} \leq \mathbf{0}$$

where  $\mathbf{L}$  gathers the affine dual recourse adaption coefficients of  $(\boldsymbol{\beta}, \boldsymbol{\mu}, \eta)$  and  $\boldsymbol{\gamma}, \boldsymbol{\Gamma}$  are appropriate affine mappings. Under Assumption 2, for any  $\boldsymbol{\gamma} \in \mathbb{R}^{n_\gamma}$  and  $\boldsymbol{\Gamma} \in \mathbb{R}^{n_\gamma \times n_f}$ , the robust counterpart of  $\boldsymbol{\gamma} + \boldsymbol{\Gamma}\boldsymbol{\rho} \leq \mathbf{0}, \forall \boldsymbol{\rho} \in \mathcal{P}$  is given by the following linear conic constraint

$$\exists \mathbf{V} \in \mathbb{R}^{n_\gamma \times n_y} : \begin{cases} \boldsymbol{\gamma} + \mathbf{V}\mathbf{d} \leq \mathbf{0} \\ \mathbf{B}(\mathbf{V}_i)^\top \succeq_{\mathcal{K}} \boldsymbol{\Gamma}_i^\top \quad \forall i \in [n_\gamma]. \end{cases}$$

Indeed, observe that the given robust constraint can be written down as

$$\max_{\boldsymbol{\rho} \in \mathcal{P}} \boldsymbol{\gamma}_i + \boldsymbol{\Gamma}_i \boldsymbol{\rho} \leq 0 \quad \forall i \in [n_\gamma] \iff \min_{\mathbf{v}^i \in \mathbb{R}^{n_y}} \{\boldsymbol{\gamma}_i + \mathbf{d}^\top \mathbf{v}^i \mid \mathbf{B}\mathbf{v}^i \succeq_{\mathcal{K}} \boldsymbol{\Gamma}_i^\top\} \leq 0 \quad \forall i \in [n_\gamma],$$

where the equivalence holds because the minimization problem is convex and strictly feasible. We then denote  $[\mathbf{v}^1, \dots, \mathbf{v}^{n_\gamma}]^\top$  by  $\mathbf{V}$ . Since  $\bar{\mathcal{Y}}$  constitutes a tractable set and is increasingly larger with  $k$  and  $v$ , we conclude that it is indeed a safe tractable approximation of  $\mathcal{Y}$  as desired.  $\square$

We now consider a special, but an important case where the tractable safe approximation is exact.

**THEOREM 5.** *Suppose that the evaluation function  $g$  can be expressed as*

$$g(\mathbf{x}, \mathbf{z}) = \min_{y \in \mathbb{R}} \{y \mid \mathbf{1}y \geq \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z}\}, \quad (23)$$

then  $\bar{\mathcal{Y}} = \mathcal{Y}$ .

*Proof.* Observe that the dual uncertainty set is now a simplex,  $\mathcal{P} = \{\boldsymbol{\rho} \in \mathbb{R}_+^{n_f} \mid \mathbf{1}^\top \boldsymbol{\rho} = 1\}$ . It has well been shown (see, *e.g.*, Zhen et al. 2018) that if the uncertainty set is a simplex, then approximation via affine recourse adaption is exact.  $\square$

We can now present the tractable safe approximation of the data-driven robust satisficing problem (10), noting that for each  $\omega \in [\Omega]$

$$\mathcal{Z}^\omega = \{\boldsymbol{\zeta} \in \mathbb{R}^{n_z} \mid \mathbf{H}\boldsymbol{\zeta} \leq \mathbf{h}^\omega\}$$

where  $\mathbf{h}^\omega = \mathbf{h} - \mathbf{H}\hat{\mathbf{z}}^\omega$ , and the evaluation function  $g^\omega : \mathcal{X} \times \mathcal{Z}_\omega \rightarrow \mathbb{R}$ ,

$$\begin{aligned} g^\omega(\mathbf{x}, \mathbf{z}) = & \text{minimize} && \mathbf{d}^\top \mathbf{y} \\ & \text{subject to} && \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{f}^\omega(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \\ & && \mathbf{y} \in \mathbb{R}^{n_y}, \end{aligned} \quad (24)$$

where  $\mathbf{f}^\omega(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\hat{\mathbf{z}}^\omega$ .

Using affine dual recourse adaptation that we have derived earlier, the tractable safe approximation of the data-driven robust satisficing in the form of Problem (13) is given by

$$\begin{aligned} \min & k \\ \text{s.t.} & \frac{1}{\Omega} \sum_{\omega \in [\Omega]} v_\omega \leq \tau \\ & \boldsymbol{\rho}^\top \mathbf{f}^\omega(\mathbf{x}) + (\mathbf{h}^\omega)^\top \boldsymbol{\beta}^\omega(\boldsymbol{\rho}) + \eta^\omega(\boldsymbol{\rho}) \leq v_\omega && \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \mathbf{M}(\mathbf{F}(\mathbf{x})^\top \boldsymbol{\rho} - \mathbf{H}^\top \boldsymbol{\beta}^\omega(\boldsymbol{\rho})) + \mathbf{N}\boldsymbol{\mu}^\omega(\boldsymbol{\rho}) + s\eta^\omega(\boldsymbol{\rho}) \leq t\mathbf{k} && \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \boldsymbol{\beta}^\omega(\boldsymbol{\rho}) \geq \mathbf{0}, \eta^\omega(\boldsymbol{\rho}) \geq 0 && \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+, \mathbf{v} \in \mathbb{R}^\Omega, \boldsymbol{\beta}^1, \dots, \boldsymbol{\beta}^\Omega \in \mathcal{L}^{n_f, n_h}, \boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^\Omega \in \mathcal{L}^{n_f, n_\mu}, \eta^1, \dots, \eta^\Omega \in \mathcal{L}^{n_f, 1} \end{aligned} \quad (25)$$

where  $\mathcal{P} = \{\boldsymbol{\rho} \in \mathcal{K}^* \mid \mathbf{B}^\top \boldsymbol{\rho} = \mathbf{d}\}$ .

This result is computationally significant since, despite the difficulty to solve it exactly, if  $\tau > Z_0$ , thanks to Proposition 2 we can still obtain a feasible solution of Problem (10) by solving a modest sized conic optimization problem.

### 3.6. On linear optimization with recourse

We revisit the data-driven linear optimization problem of Long et al. (2022) and explore how the affine dual recourse adaptation can be used to provide a better tractable safe approximation, and with superior computational performance. Specifically, we consider the  $\ell_1$ -norm penalty function,  $p(\boldsymbol{\zeta}) = \|\boldsymbol{\zeta}\|_1$ , and the evaluation function is a linear optimization problem given by

$$\begin{aligned} g(\mathbf{x}, \mathbf{z}) = & \text{minimize} && \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} \\ & \text{subject to} && \mathbf{B}\mathbf{y} \geq \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \\ & && \mathbf{y} \in \mathbb{R}^{n_y}, \end{aligned} \quad (26)$$



and we can express the feasible set from (14) as

$$\mathcal{Y} = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \mathbf{y} \in \mathcal{R}^{n_z, n_y} : \\ \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y}(\mathbf{z}) - k \|\mathbf{z}\|_1 \leq v \quad \forall \mathbf{z} \in \mathcal{Z} \\ \mathbf{B}\mathbf{y}(\mathbf{z}) \geq \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \quad \forall \mathbf{z} \in \mathcal{Z} \end{array} \right. \right\}. \quad (27)$$

Generally, two-stage robust linear optimization problems are not tractable and are typically solved approximately via affine primal recourse adaptation, but this would not necessarily lead to a feasible solution (Long et al. 2022). As a result, we consider a more flexible, non-affine primal recourse adaptation extension with extra coefficients  $\mathbf{q}^\dagger \in \mathbb{R}^{n_y}$ :

$$\mathbf{y}(\mathbf{z}) = \mathbf{q} + \mathbf{Q}\mathbf{z} + \mathbf{q}^\dagger \|\mathbf{z}\|_1, \quad (28)$$

which results in the following tractable safe approximation

$$\bar{\mathcal{Y}}_P = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \mathbf{q} \in \mathbb{R}^{n_y}, \mathbf{Q} \in \mathbb{R}^{n_y \times n_z}, \mathbf{q}^\dagger \in \mathbb{R}^{n_y} : \\ \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top (\mathbf{q} + \mathbf{Q}\mathbf{z} + \mathbf{q}^\dagger \|\mathbf{z}\|_1) - k \|\mathbf{z}\|_1 \leq v \quad \forall \mathbf{z} \in \mathcal{Z} \\ \mathbf{B}(\mathbf{q} + \mathbf{Q}\mathbf{z} + \mathbf{q}^\dagger \|\mathbf{z}\|_1) \geq \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z} \quad \forall \mathbf{z} \in \mathcal{Z} \\ k - \mathbf{d}^\top \mathbf{q}^\dagger \geq 0, \mathbf{B}\mathbf{q}^\dagger \geq \mathbf{0} \end{array} \right. \right\}, \quad (29)$$

where the two linear constraints  $k - \mathbf{d}^\top \mathbf{q}^\dagger \geq 0$  and  $\mathbf{B}\mathbf{q}^\dagger \geq \mathbf{0}$  are added to ensure that the two robust constraints are concave in the uncertain  $\mathbf{z}$  and consequently tractability of the safe approximation.

**PROPOSITION 8.** *Under Assumption 2,  $\bar{\mathcal{Y}}_P$  is a tractable safe approximation of  $\mathcal{Y}$ , and has the following explicit representation*

$$\bar{\mathcal{Y}}_P = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \mathbf{q} \in \mathbb{R}^{n_y}, \mathbf{Q} \in \mathbb{R}^{n_y \times n_z}, \mathbf{q}^\dagger \in \mathbb{R}^{n_y}, \mathbf{w}^0, \dots, \mathbf{w}^{n_f} \in \mathbb{R}_+^{n_h} : \\ \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{q} + \mathbf{h}^\top \mathbf{w}^0 \leq v \\ -(k - \mathbf{d}^\top \mathbf{q}^\dagger) \mathbf{1} \leq \mathbf{H}^\top \mathbf{w}^0 - \mathbf{Q}^\top \mathbf{d} \leq (k - \mathbf{d}^\top \mathbf{q}^\dagger) \mathbf{1} \\ \mathbf{f}_i(\mathbf{x}) + \mathbf{h}^\top \mathbf{w}^i \leq \mathbf{B}_i \mathbf{q}^\dagger \quad \forall i \in [n_f] \\ -\mathbf{B}_i \mathbf{q}^\dagger \mathbf{1} \leq \mathbf{H}^\top \mathbf{w}^i - \mathbf{F}_i^\top(\mathbf{x}) + \mathbf{Q}^\top \mathbf{B}_i^\top \leq \mathbf{B}_i \mathbf{q}^\dagger \mathbf{1} \quad \forall i \in [n_f] \end{array} \right. \right\}. \quad (30)$$

*Proof.* Similar results have been shown in (Long et al. 2022), and for completeness we provide an alternate proof in Appendix A.

Next, we will similarly analyze the dual robust satisficing by considering the following equivalent dual tractable safe approximation from (21) as follows

$$\bar{\mathcal{Y}}_D = \left\{ (\mathbf{x}, v, k) \in \mathcal{X} \times \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \boldsymbol{\pi} \in \mathbb{R}^{n_h}, \boldsymbol{\Pi} \in \mathbb{R}^{n_h \times n_f} : \\ \mathbf{c}^\top \mathbf{x} + \boldsymbol{\rho}^\top \mathbf{f}(\mathbf{x}) + (\boldsymbol{\pi} + \boldsymbol{\Pi}\boldsymbol{\rho})^\top \mathbf{h} \leq v \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\ -k\mathbf{1} \leq \mathbf{H}^\top (\boldsymbol{\pi} + \boldsymbol{\Pi}\boldsymbol{\rho}) - \mathbf{F}(\mathbf{x})^\top \boldsymbol{\rho} \leq k\mathbf{1} \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\ \boldsymbol{\pi} + \boldsymbol{\Pi}\boldsymbol{\rho} \geq \mathbf{0} \quad \forall \boldsymbol{\rho} \in \mathcal{P} \end{array} \right. \right\}, \quad (31)$$

which we have earlier shown is feasible due to complete recourse, for dual uncertainty set  $\mathcal{P}$  that is  $\{\boldsymbol{\rho} \in \mathbb{R}_+^{n_f} \mid \mathbf{B}^\top \boldsymbol{\rho} = \mathbf{d}\}$  and for  $\boldsymbol{\pi}$  and  $\mathbf{\Pi}$  that are the intercept and the gradient of  $\boldsymbol{\beta}(\boldsymbol{\rho})$  with respect to  $\boldsymbol{\rho}$ . Our objective here is therefore to show that, despite the restricted linearity of  $\boldsymbol{\beta}(\boldsymbol{\rho})$ , the dual tractable safe approximation is a tighter approximation than the primal safe approximation. This observation provides a fresh perspective, which goes in a direction opposite to Bertsimas and de Ruiter (2016), that the primal and the dual affine adaptations are not necessarily equivalent approximations.

**THEOREM 6.** *Under Assumption 2,*

$$\bar{\mathcal{Y}}_P \subseteq \bar{\mathcal{Y}}_D.$$

*Proof.* First of all, we compare the feasible sets (29) and (31) when  $\mathbf{x}$  is fixed to  $\mathbf{x}' \in \mathcal{X}$ . We can then abbreviate  $\mathbf{f}(\mathbf{x}')$  and  $\mathbf{F}(\mathbf{x}')$  to  $\mathbf{f}'$  and  $\mathbf{F}'$ , respectively. Besides, we let  $v'$  denote the value of  $v - \mathbf{c}^\top \mathbf{x}'$ . Equivalently, it suffices to show that

$$\bar{\mathcal{Y}}'_D = \left\{ (v, k) \in \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \boldsymbol{\pi} \in \mathbb{R}^{n_h}, \mathbf{\Pi} \in \mathbb{R}^{n_h \times n_f} : \\ \boldsymbol{\rho}^\top \mathbf{f}' + (\boldsymbol{\pi} + \mathbf{\Pi} \boldsymbol{\rho})^\top \mathbf{h} \leq v' \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\ -k \mathbf{1} \leq \mathbf{H}^\top (\boldsymbol{\pi} + \mathbf{\Pi} \boldsymbol{\rho}) - (\mathbf{F}')^\top \boldsymbol{\rho} \leq k \mathbf{1} \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\ \boldsymbol{\pi} + \mathbf{\Pi} \boldsymbol{\rho} \geq \mathbf{0} \quad \forall \boldsymbol{\rho} \in \mathcal{P} \end{array} \right. \right\}, \quad (32)$$

is a superset of

$$\bar{\mathcal{Y}}'_P = \left\{ (v, k) \in \mathbb{R} \times \mathbb{R}_+ \left| \begin{array}{l} \exists \mathbf{q} \in \mathbb{R}^{n_y}, \mathbf{Q} \in \mathbb{R}^{n_y \times n_z}, \mathbf{q}^\dagger \in \mathbb{R}^{n_y} : \\ \mathbf{d}^\top (\mathbf{q} + \mathbf{Q} \mathbf{z} + \mathbf{q}^\dagger \|\mathbf{z}\|_1) - k \|\mathbf{z}\|_1 \leq v' \quad \forall \mathbf{z} \in \mathcal{Z} \\ \mathbf{B} (\mathbf{q} + \mathbf{Q} \mathbf{z} + \mathbf{q}^\dagger \|\mathbf{z}\|_1) \geq \mathbf{f}' + \mathbf{F}' \mathbf{z} \quad \forall \mathbf{z} \in \mathcal{Z} \\ k - \mathbf{d}^\top \mathbf{q}^\dagger \geq 0, \mathbf{B} \mathbf{q}^\dagger \geq \mathbf{0} \end{array} \right. \right\}. \quad (33)$$

For any  $(v, k) \in \bar{\mathcal{Y}}'_P$ , there exists  $(\mathbf{q}, \mathbf{Q}, \mathbf{q}^\dagger)$  that satisfies all of the inequality requirements in (33). It then suffices to show that for the same  $(v, k)$ , we can construct  $\boldsymbol{\pi}$  and  $\mathbf{\Pi}$  such that  $(\boldsymbol{\pi}, \mathbf{\Pi})$  satisfies the inequality requirements in (32). To achieve this, we specifically consider  $\boldsymbol{\pi} = \mathbf{w}^0 \in \mathbb{R}_+^{n_h}$  and  $\mathbf{\Pi} = [\mathbf{w}^1, \dots, \mathbf{w}^{n_f}] \in \mathbb{R}_+^{n_h \times n_f}$ , where  $\{\mathbf{w}^i\}_{i=0}^{n_f}$  satisfy the following conditions.

$$\mathbf{d}^\top \mathbf{q} + \mathbf{h}^\top \mathbf{w}^0 \leq v' \quad (34a)$$

$$-(k - \mathbf{d}^\top \mathbf{q}^\dagger) \mathbf{1} \leq \mathbf{H}^\top \mathbf{w}^0 - \mathbf{Q}^\top \mathbf{d} \leq (k - \mathbf{d}^\top \mathbf{q}^\dagger) \mathbf{1} \quad (34b)$$

$$\mathbf{f}'_i + \mathbf{h}^\top \mathbf{w}^i \leq \mathbf{B}_i \mathbf{q} \quad \forall i \in [n_f] \quad (34c)$$

$$-\mathbf{B}_i \mathbf{q}^\dagger \mathbf{1} \leq \mathbf{H}^\top \mathbf{w}^i - (\mathbf{F}')^\top_i + \mathbf{Q}^\top \mathbf{B}_i^\top \leq \mathbf{B}_i \mathbf{q}^\dagger \mathbf{1} \quad \forall i \in [n_f] \quad (34d)$$

Note that the existence  $\{\mathbf{w}^i\}_{i=0}^{n_h}$  is guaranteed by Proposition 8, which holds when  $\mathcal{X} = \{\mathbf{x}'\}$ .

We will now show that this choice of  $(\boldsymbol{\pi}, \mathbf{\Pi})$  satisfies all constraints in Problem (32). By a usual duality argument, the first inequality in (32) is robustly satisfied for all  $\boldsymbol{\rho} \in \mathcal{P}$  if and only if there exists a vector  $\boldsymbol{\nu} \in \mathbb{R}^{n_y}$  such that

$$\boldsymbol{\pi}^\top \mathbf{h} + \boldsymbol{\nu}^\top \mathbf{d} \leq v' \quad \text{and} \quad \mathbf{B}\boldsymbol{\nu} \geq \mathbf{f}' + \mathbf{\Pi}^\top \mathbf{h}.$$

Thanks to the inequalities (34a) and (34c),  $\boldsymbol{\nu}$  can be simply chosen as  $\mathbf{q}$ .

Similarly, the two subsequent inequalities of (32), namely  $-k\mathbf{1} \leq \mathbf{H}^\top(\boldsymbol{\pi} + \mathbf{\Pi}\boldsymbol{\rho}) - (\mathbf{F}')^\top \boldsymbol{\rho} \leq k\mathbf{1}$ , are robustly satisfied for all  $\boldsymbol{\rho} \in \mathcal{P}$  if and only if there exist matrices  $\mathbf{\Phi}, \mathbf{\Psi} \in \mathbb{R}^{n_y \times n_z}$  such that

$$\left. \begin{array}{l} \mathbf{H}^\top \boldsymbol{\pi} + \mathbf{\Phi}^\top \mathbf{d} \leq k\mathbf{1} \\ \mathbf{B}\mathbf{\Phi} \geq \mathbf{\Pi}^\top \mathbf{H} - \mathbf{F}' \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{l} -\mathbf{H}^\top \boldsymbol{\pi} + \mathbf{\Psi}^\top \mathbf{d} \leq k\mathbf{1} \\ \mathbf{B}\mathbf{\Psi} \geq \mathbf{F}' - \mathbf{\Pi}^\top \mathbf{H}. \end{array} \right.$$

In this case, we can choose,  $\mathbf{\Phi} = \mathbf{q}^\dagger \mathbf{1}^\top - \mathbf{Q}$  and  $\mathbf{\Psi} = \mathbf{q}^\dagger \mathbf{1}_{n_z}^\top + \mathbf{Q}$ . Observe that

$$\begin{aligned} \mathbf{H}^\top \boldsymbol{\pi} + \mathbf{\Phi}^\top \mathbf{d} &= \mathbf{H}^\top \mathbf{w}^0 + \mathbf{1}(\mathbf{q}^\dagger)^\top \mathbf{d} - \mathbf{Q}^\top \mathbf{d} \leq k\mathbf{1} \\ \mathbf{B}\mathbf{\Phi} - \mathbf{\Pi}^\top \mathbf{H} + \mathbf{F}' &= [\mathbf{B}_i \mathbf{q}^\dagger \mathbf{1}^\top - \mathbf{B}_i \mathbf{Q} - (\mathbf{w}^i)^\top \mathbf{H} + \mathbf{F}'_i]_{i=1}^{n_f} \geq \mathbf{0}, \end{aligned}$$

where the inequalities holds due to the inequalities (34b) and (34d), as desired. Similarly,

$$\begin{aligned} -\mathbf{H}^\top \boldsymbol{\pi} + \mathbf{\Psi}^\top \mathbf{d} &= -\mathbf{H}^\top \mathbf{w}^0 + \mathbf{1}(\mathbf{q}^\dagger)^\top \mathbf{d} + \mathbf{Q}^\top \mathbf{d} \leq k\mathbf{1} \\ \mathbf{B}\mathbf{\Psi} + \mathbf{\Pi}^\top \mathbf{H} - \mathbf{F}' &= [\mathbf{B}_i \mathbf{q}^\dagger \mathbf{1}^\top + \mathbf{B}_i \mathbf{Q} + (\mathbf{w}^i)^\top \mathbf{H} - \mathbf{F}'_i]_{i=1}^{n_f} \geq \mathbf{0}. \end{aligned}$$

Finally, the last requirement in (32) is trivially satisfied because of the non-negativity of  $\boldsymbol{\pi}$ ,  $\mathbf{\Pi}$  and  $\boldsymbol{\rho}$ , by construction. Hence,  $\bar{\mathcal{Y}}_P \subseteq \bar{\mathcal{Y}}_D$ , and the theorem follows.  $\square$

Finally, to illustrate the distinction between the primal non-affine and the dual affine adaptations of a two-stage robust satisficing linear optimization problem, we consider the following specific instance

$$\begin{aligned} &\text{minimize} && k \\ &\text{subject to} && \mathbf{d}^\top \mathbf{y}(\mathbf{z}) \leq \tau + k\|\mathbf{z}\|_1 \quad \forall \mathbf{z} : \mathbf{H}\mathbf{z} \leq \mathbf{h} \\ &&& \mathbf{B}\mathbf{y}(\mathbf{z}) \geq \mathbf{f} + \mathbf{F}\mathbf{z} \quad \forall \mathbf{z} : \mathbf{H}\mathbf{z} \leq \mathbf{h} \\ &&& \mathbf{y} \in \mathcal{R}^{n_y, n_z}, k \in \mathbb{R}_+ \end{aligned} \tag{35}$$

for the parameters listed in Appendix B. The chosen recourse matrix  $\mathbf{B}$  is non-negative and numerically satisfies  $\min_{\mathbf{y}} \{\mathbf{d}^\top \mathbf{y} \mid \mathbf{B}\mathbf{y} \geq \mathbf{0}\} = 0$  and Assumption 2 is thus satisfied. For different values of the target objective  $\tau$  that are sufficiently large, Figure 2 suggests that the dual affine adaptation is an approximation that is less conservative compared to its primal non-affine counterpart.

In Appendix C, we present an additional computational study of solving various robust network lot-sizing problems. Focusing on evaluating the computational speed, we show that the dual affine adaptation method again significantly outperforms the primal method in solving both the robust optimization and satisficing models.

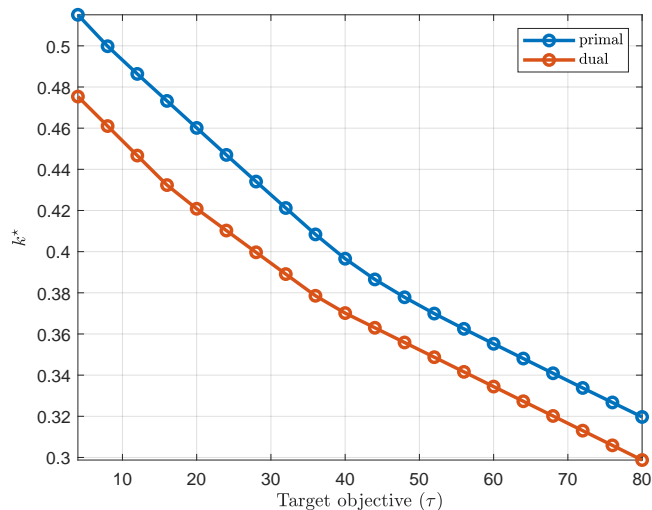


Figure 2 Robust satisficing objectives obtained from the primal non-affine and the dual affine adaptations.

#### 4. Portfolio optimization via expected utility maximization

In our numerical study, we consider a data-driven portfolio optimization problem where  $\tilde{\mathbf{z}}$  denotes a random vector of asset returns, with distribution  $\tilde{\mathbf{z}} \sim \mathbb{P}^*$ ,  $\mathbb{P}^* \in \mathcal{P}_0(\mathcal{Z})$ . Ideally, the portfolio manager would like to maximize the certainty equivalent of the portfolio returns, *i.e.*,

$$\begin{aligned} & \text{maximize} && u_a^{-1} \left( \mathbb{E}_{\mathbb{P}^*} [u_a(\mathbf{x}^\top \tilde{\mathbf{z}})] \right) \\ & \text{subject to} && \mathbf{x} \in \mathcal{X}, \end{aligned}$$

over the feasible set  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{1}^\top \mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}$ , and based on an exponential utility with risk aversion  $a > 0$ , *i.e.*,  $u_a(r) = (1 - \exp(-ar))/a$ . However, the actual distribution  $\mathbb{P}^*$  is unobservable. Instead, we are only given  $\Omega$  historical returns given by  $\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^\Omega$ .

For the empirical optimization model, we minimize the expected disutility of the portfolio loss as follows:

$$Z_0 = \min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\hat{\mathbb{P}}} [\ell_a(-\mathbf{x}^\top \tilde{\mathbf{z}})] = \min_{\mathbf{x} \in \mathcal{X}} \frac{1}{\Omega} \sum_{\omega \in [\Omega]} \ell_a(-\mathbf{x}^\top \hat{\mathbf{z}}^\omega), \quad (36)$$

where  $\ell_a(w) = -u_a(-w) = (\exp(aw) - 1)/a$  is the disutility defined on loss  $w$ .

For the stochastic optimization model, we assume a multivariate normal distribution,  $\mathbb{P}^\dagger = \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ . The assumption of normal distribution implies that these parameters can be estimated by maximizing the likelihood function so that  $\hat{\boldsymbol{\mu}} = \mathbb{E}_{\hat{\mathbb{P}}}[\tilde{\mathbf{r}}]$  and  $\hat{\boldsymbol{\Sigma}} = \mathbb{E}_{\hat{\mathbb{P}}}[(\tilde{\mathbf{r}} - \hat{\boldsymbol{\mu}})(\tilde{\mathbf{r}} - \hat{\boldsymbol{\mu}})^\top]$ , which are computed over the  $\Omega$  empirical samples. However, asset returns frequently exhibit non-normal behavior. This is particularly true of certain types of assets, such as hedge funds, emerging markets, or assets with credit risk, and is even more evident during financial crises. Under normality, maximizing the

certainly equivalent of the portfolio returns under exponential utility is equivalent to solving the celebrated Markowitz mean-variance portfolio optimization problem as follows:

$$\begin{aligned} & \text{maximize} && \mathbf{x}^\top \hat{\boldsymbol{\mu}} - a \frac{\mathbf{x}^\top \hat{\boldsymbol{\Sigma}} \mathbf{x}}{2} \\ & \text{subject to} && \mathbf{x} \in \mathcal{X}. \end{aligned} \quad (37)$$

Comparing to the empirical optimization problem, which is an exponential conic optimization problem that depends on the data size, the Markowitz portfolio optimization problem is a fixed and modest sized quadratic optimization problem. For the robust satisficing model, we assume that the support set is a box  $\mathcal{Z} = \{\mathbf{z} | \underline{\mathbf{z}} \leq \mathbf{z} \leq \bar{\mathbf{z}}\}$ , where  $\underline{\mathbf{z}}$  and  $\bar{\mathbf{z}}$  are determined from data. The corresponding robust satisficing investment problem can be formulated as

$$\begin{aligned} & \text{minimize} && k \\ & \text{subject to} && \frac{1}{\Omega} \sum_{\omega \in [\Omega]} \sup_{\mathbf{z} \in \mathcal{Z}^\omega} (g^\omega(\mathbf{x}, \mathbf{z}) - k p_\Gamma(\mathbf{z})) \leq \tau \\ & && \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+. \end{aligned} \quad (38)$$

where  $p_\Gamma(\mathbf{z})$  is the budgeted norm (see Example 3),  $\mathcal{Z}^\omega = [\underline{\mathbf{z}} - \hat{\mathbf{z}}^\omega, \bar{\mathbf{z}} - \hat{\mathbf{z}}^\omega]$ , and  $g^\omega(\mathbf{x}, \mathbf{z})$  is the conic representable evaluation function satisfying  $g^\omega(\mathbf{x}, \mathbf{z}) = \ell_a(-\mathbf{x}^\top(\hat{\mathbf{z}}^\omega + \mathbf{z}))$  for all  $\mathbf{x} \in \mathcal{X}$ ,  $\mathbf{z} \in \mathcal{Z}^\omega$ . Specifically, from Theorem 2, we express the evaluation function as

$$\begin{aligned} g^\omega(\mathbf{x}, \mathbf{z}) = & \text{minimize} && y_1 + P y_2 \\ & \text{subject to} && \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \succeq_{\mathcal{K}} \begin{bmatrix} \mathbf{x}^\top \hat{\mathbf{z}}^\omega \\ -P \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \mathbf{x}^\top \\ \mathbf{0}^\top \\ \mathbf{0}^\top \\ \mathbf{0}^\top \end{bmatrix} \mathbf{z} \\ & && \mathbf{y} \in \mathbb{R}^2, \end{aligned}$$

where  $\mathcal{K} = \mathcal{K}_{\ell_a} \times \mathbb{R}_+$ . Note that since we are evaluating  $\ell_a(-\mathbf{x}^\top(\hat{\mathbf{z}}^\omega + \mathbf{z}))$  for  $\mathbf{x} \in \mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{1}^\top \mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}$  and  $\mathbf{z} \in \mathcal{Z}^\omega$ , it suffices to choose the featured domain  $\mathcal{W} = [\underline{\mathbf{w}}, \bar{\mathbf{w}}]$  such that

$$\bar{w} \geq \max_{\mathbf{z} \in \mathcal{Z}^\omega, \mathbf{x} \in \mathcal{X}} \{-\mathbf{x}^\top(\hat{\mathbf{z}}^\omega + \mathbf{z})\} = \max_{\mathbf{z} \in [\underline{\mathbf{z}}, \bar{\mathbf{z}}], \mathbf{x} \in \mathcal{X}} \{-\mathbf{x}^\top \mathbf{z}\} = -\min_{i \in [n]} \{\underline{z}_i\}$$

and

$$\underline{w} \leq \min_{\mathbf{z} \in \mathcal{Z}^\omega, \mathbf{x} \in \mathcal{X}} \{-\mathbf{x}^\top(\hat{\mathbf{z}}^\omega + \mathbf{z})\} = \min_{\mathbf{z} \in [\underline{\mathbf{z}}, \bar{\mathbf{z}}], \mathbf{x} \in \mathcal{X}} \{-\mathbf{x}^\top \mathbf{z}\} = -\max_{i \in [n]} \{\bar{z}_i\}.$$

From Proposition 5, we choose

$$P \geq \max_{v \in \{\underline{w}, \bar{w}\}} \{\ell'_a(v)v - \ell_a(v)\} = \max_{v \in \{\underline{w}, \bar{w}\}} \{\exp(av)(v - 1/a) + 1/a\},$$

so that

$$g^\omega(\mathbf{x}, \mathbf{z}) = \ell_a(-\mathbf{x}^\top(\hat{\mathbf{z}}^\omega + \mathbf{z})) \quad \forall \mathbf{x} \in \mathcal{X}, \mathbf{z} \in \mathcal{Z}^\omega.$$

PROPOSITION 9. *The tractable safe approximation of Problem (38) is given by*

$$\begin{aligned}
& \text{minimize } k \\
& \text{subject to } \frac{1}{\Omega} \sum_{\omega \in [\Omega]} v_\omega \leq \tau \\
& \mathbf{x}^\top \hat{\mathbf{z}}^\omega \rho_1 - P \rho_2 + \rho_4 + (\bar{\mathbf{z}}^\omega)^\top \bar{\boldsymbol{\beta}}^{\omega,0} - (\underline{\mathbf{z}}^\omega)^\top \underline{\boldsymbol{\beta}}^{\omega,0} \\
& \quad + \sum_{j \in [4]} ((\bar{\mathbf{z}}^\omega)^\top \bar{\boldsymbol{\beta}}^{\omega,j} - (\underline{\mathbf{z}}^\omega)^\top \underline{\boldsymbol{\beta}}^{\omega,j}) \rho_j \leq v_\omega \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\
& \mathbf{1}^\top \boldsymbol{\mu}^{\omega,0} + \sum_{j \in [4]} \mathbf{1}^\top \boldsymbol{\mu}^{\omega,j} \rho_j \leq \Gamma k \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\
& x_i \rho_1 - \bar{\beta}_i^{\omega,0} + \underline{\beta}_i^{\omega,0} - \mu_i^{\omega,0} + \sum_{j \in [4]} (-\bar{\beta}_i^{\omega,j} + \underline{\beta}_i^{\omega,j} - \mu_i^{\omega,j}) \rho_j \leq 0 \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega], i \in [n] \\
& -x_i \rho_1 + \bar{\beta}_i^{\omega,0} - \underline{\beta}_i^{\omega,0} - \mu_i^{\omega,0} + \sum_{j \in [4]} (\bar{\beta}_i^{\omega,j} - \underline{\beta}_i^{\omega,j} - \mu_i^{\omega,j}) \rho_j \leq 0 \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega], i \in [n] \\
& \mu_i^{\omega,0} + \sum_{j \in [4]} \mu_i^{\omega,j} \rho_j \leq k \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega], i \in [n] \\
& \bar{\beta}_i^{\omega,0} + \sum_{j \in [4]} \bar{\beta}_i^{\omega,j} \rho_j \geq 0 \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega], i \in [n] \\
& \underline{\beta}_i^{\omega,0} + \sum_{j \in [4]} \underline{\beta}_i^{\omega,j} \rho_j \geq 0 \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega], i \in [n] \\
& \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+, \mathbf{v} \in \mathbb{R}^\Omega, \bar{\boldsymbol{\beta}}^{\omega,j}, \underline{\boldsymbol{\beta}}^{\omega,j}, \boldsymbol{\mu}^{\omega,j} \in \mathbb{R}^n \quad \forall \omega \in [\Omega], j \in \{0, \dots, 4\},
\end{aligned} \tag{39}$$

where  $\bar{\mathbf{z}}^\omega = \bar{\mathbf{z}} - \hat{\mathbf{z}}^\omega$  and  $\underline{\mathbf{z}}^\omega = \underline{\mathbf{z}} - \hat{\mathbf{z}}^\omega$ . Every constraint in Problem (39) is linear in the dual uncertainty  $\boldsymbol{\rho}$  and its robust counterpart could be derived from the following equivalence:

$$\forall \boldsymbol{\rho} \in \mathcal{P} : \mathbf{v}^\top \boldsymbol{\rho} \leq w \iff \exists y \in \mathbb{R} : \begin{cases} (y - v_3)(\exp(-av_1/(y - v_3)) - 1)/a + v_2 + Py \leq w, \\ y \geq v_3, \\ y \geq v_4. \end{cases}$$

*Proof.* The proof is relegated to Appendix A.  $\square$

To evaluate the performance of the portfolio, we consider the distribution  $\mathbb{P}^*$ ,  $\tilde{\mathbf{z}} \sim \mathbb{P}^*$ , such that each  $\tilde{z}_i$  is an independently distributed two-point random variable,

$$\mathbb{P}^*[\tilde{z}_i = z] = \begin{cases} \beta_i & \text{if } z = \mu_i + \sigma_i \frac{\sqrt{\beta_i(1-\beta_i)}}{\beta_i}, \\ 1 - \beta_i & \text{if } z = \mu_i - \sigma_i \frac{\sqrt{\beta_i(1-\beta_i)}}{1-\beta_i}. \end{cases}$$

Observe that  $\mathbb{E}_{\mathbb{P}^*}[\tilde{z}_i] = \mu_i$  and  $\mathbb{E}_{\mathbb{P}^*}[(\tilde{z}_i - \mu_i)^2] = \sigma_i^2$  represent the mean and the variance of the  $i^{\text{th}}$  asset. The skewness of the  $i^{\text{th}}$  asset would depend on  $\beta_i$ ; higher values imply more negative skewness and result in large losses and small upside gains. We consider  $n = 8$ ,

$$\begin{aligned}
\boldsymbol{\mu} &= [0.12, 0.16, 0.14, 0.13, 0.15, 0.12, 0.14, 0.15]^\top, \\
\boldsymbol{\sigma} &= [0.18, 0.22, 0.20, 0.16, 0.14, 0.10, 0.14, 0.19]^\top,
\end{aligned}$$

and  $\beta_i$ ,  $i \in [n]$ , is chosen as  $\beta_i = \frac{1}{2} \left(1 + \frac{i}{n+1}\right)$ . Hence, the return distributions for stocks with high index numbers in the portfolio are more negatively skewed than those with low index numbers. For a given  $\mathbf{x} \in \mathcal{X}$ , we can evaluate the actual certainty equivalent of the portfolio return as follows

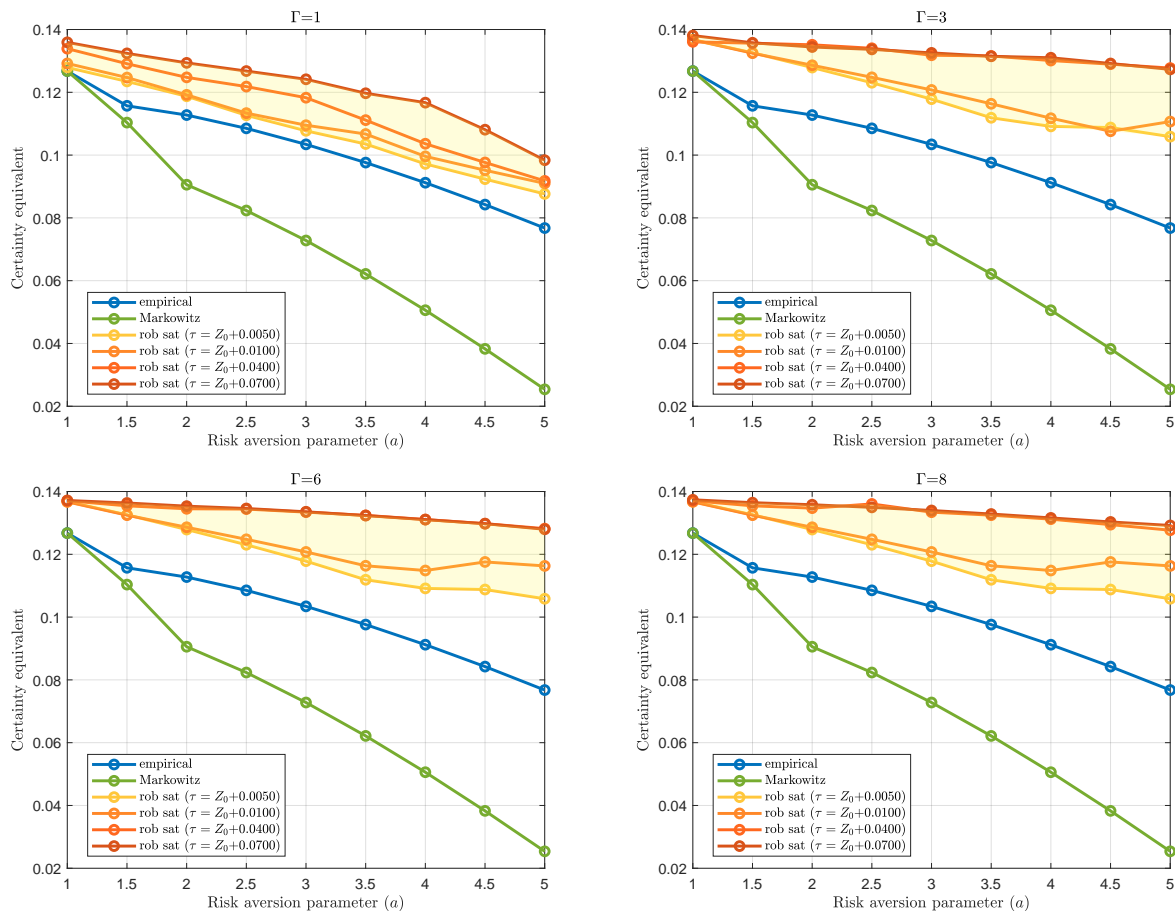
$$\begin{aligned} & u_a^{-1} \left( \mathbb{E}_{\mathbb{P}^*} [u_a(\mathbf{x}^\top \tilde{\mathbf{z}})] \right) \\ &= -\frac{1}{a} \log \left( \mathbb{E}_{\mathbb{P}^*} [(\exp(-a\mathbf{x}^\top \tilde{\mathbf{z}}))] \right) \\ &= \mathbf{x}^\top \boldsymbol{\mu} - \frac{1}{a} \sum_{i \in [n]} \log \left( \beta_i \exp \left( \frac{-ax_i \sigma_i \sqrt{\beta_i(1-\beta_i)}}{\beta_i} \right) + (1-\beta_i) \exp \left( \frac{ax_i \sigma_i \sqrt{\beta_i(1-\beta_i)}}{1-\beta_i} \right) \right). \end{aligned} \quad (40)$$

For benchmarking purposes, we first generate 100 independent realizations of  $\tilde{\mathbf{z}}$  from the two-point distribution as described above. For a given  $a \geq 0$ , we denote the optimal solutions (computed using the generated samples) of the empirical and stochastic optimization problems in (36) and (37) by  $\mathbf{x}_a^{\text{emp}}$  and  $\mathbf{x}_a^{\text{stoc}}$  (also referred to as the Markowitz solution), respectively. We then solve the robust satisficing problem using the safe approximation from (39), by conservatively choosing the parameters of the support set as

$$\underline{z}_j = \min_{i \in [n], \omega \in [\Omega]} \{ \hat{z}_i^\omega \} - 0.1 \quad \text{and} \quad \bar{z}_j = \max_{i \in [n], \omega \in [\Omega]} \{ \hat{z}_i^\omega \} + 0.1.$$

for all  $j \in [n]$ , *i.e.*, all  $z_j$ 's share the same support. Even though this support is unnecessarily large, the robust satisficing solutions, as we will see, are not overly conservative. Then, we assign  $\underline{w} = -\bar{z}_1$ ,  $\bar{w} = -\underline{z}_1$  and choose the budgeted norm parameter  $\Gamma \in \{1, 3, 6, 8\}$ . For each such  $\Gamma$ , we select four target risk values above the optimal empirical objective  $Z_0$ , *i.e.*,  $\tau = Z_0 + \lambda$  for  $\lambda = \{0.005, 0.01, 0.04, 0.07\}$ . Denote the respective optimal satisficing solutions by  $\mathbf{x}_{a,\Gamma,\tau}^{\text{st}}$ . Note that while we provide a range of values of  $\gamma$  and  $\lambda$  in our numerical studies, in a data-driven environment, these hyper-parameters should be determined using cross-validation techniques (see, *e.g.*, Mohajerin Esfahani and Kuhn 2018, Shafieezadeh-Abadeh et al. 2019, among others).

All optimization problems are solved by using Mosek 9.2.38 together with YALMIP modeling language (Löfberg 2004). After obtaining  $\mathbf{x}_{a,\Gamma,\tau}^{\text{st}}$ ,  $\mathbf{x}^{\text{emp}}$  and  $\mathbf{x}^{\text{stoc}}$  for varying  $a, \Gamma, \tau$ , the certainty equivalent of each investment strategy is then computed using (40). The obtained results are shown in Figure 3. For  $a = 1$ , it can be observed that the certainty equivalents for the empirical and stochastic models coincide. Indeed, when the investor is mildly risk-averse, both resulting portfolios invest all wealth in the asset with the highest empirical mean return. The satisficing model, on the other hand, opts for a more diversified investment and yields a superior certainty equivalent for all values of  $\Gamma$  and  $\tau$ . The performance gaps between the satisficing model and the empirical and the stochastic models appear to increase with  $a$ . Amongst these three different approaches, the stochastic model yields the most disappointing certainty equivalents because it suffers from trusting the samples too much and from imposing an incorrect shape for the asset return distribution.



**Figure 3** Certainty equivalent comparison between the empirical, stochastic (Markowitz) and robust satisfying investment models.

*Acknowledgements:* We are grateful to Jianzhe Zhen for sharing the source code used in Roos et al. (2020) and Peng Xiong for the valuable discussions regarding RSOME.

## References

1995. Characteristics of decisions in decision analysis practice. *Journal of the Operational Research Society* **46**(3) 304–314.
- Ben-Tal, A., R. Brekelmans, D. den Hertog, J. Vial. 2017. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing* **29**(2) 350–366.
- Ben-Tal, A., D. den Hertog, J. Vial. 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming* **149**(1) 265–299.
- Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical programming* **99**(2) 351–376.
- Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* **23**(4) 769–805.



- Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming* **88**(3) 411–424.
- Ben-Tal, A., A. Nemirovski. 2001. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research* **26**(2) 193–205.
- Bertsimas, D., F. de Ruiter. 2016. Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing* **28**(3) 500–511.
- Bertsimas, D., V. Gupta, N. Kallus. 2018. Data-driven robust optimization. *Mathematical Programming* **167**(2) 235–292.
- Bertsimas, D., D. Iancu, P. Parrilo. 2010. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research* **35**(2) 363–394.
- Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.
- Bertsimas, D., M. Sim, M. Zhang. 2019. Adaptive distributionally robust optimization. *Management Science* **65**(2) 604–618.
- Birge, J., F. Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.
- Chen, Z., M. Sim, P. Xiong. 2020. Robust stochastic optimization made easy with RSOME. *Management Science* .
- de Ruiter, F., J. Zhen, D. den Hertog. 2018. Dual approach for two-stage robust nonlinear optimization. Available on Optimization-Online.
- Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* **58**(3) 595–612.
- DeMiguel, V., L. Garlappi, R. Uppal. 2009. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *The Review of Financial Studies* **22**(5) 1915–1953.
- Fournier, N., A. Guillin. 2015. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162**(3) 707–738.
- Kuhn, D., W. Wiesemann, A. Georghiou. 2011. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* **130**(1) 177–209.
- Löfberg, J. 2004. YALMIP: A toolbox for modeling and optimization in MATLAB. *In Proceedings of the CACSD Conference*. Taipei, Taiwan.
- Long, Z., M. Sim, M. Zhou. 2022. Robust satisficing. *Operations Research* **In Press**.
- Mohajerin Esfahani, P., D. Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* **171**(1) 115–166.
- Roos, E., D. den Hertog, A. Ben-Tal, F. de Ruiter, J. Zhen. 2020. Tractable approximation of hard uncertain optimization problems. Available on Optimization-Online.

- 
- Shafieezadeh-Abadeh, S., D. Kuhn, P. Mohajerin Esfahani. 2019. Regularization via mass transportation. *Journal of Machine Learning Research* **20**(103) 1–68.
- Shapiro, A., D. Dentcheva, A. Ruszczyński. 2014. *Lectures on stochastic programming: modeling and theory*. SIAM.
- Sim, M., L. Zhao, M. Zhou. 2021. Tractable robust supervised learning models. Available at SSRN 3981205.
- Smith, J., R. Winkler. 2006. The optimizer’s curse: Suhnepticism and postdecision surprise in decision analysis. *Management Science* **52**(3) 311–322.
- Wiesemann, W., D. Kuhn, M. Sim. 2014. Distributionally robust convex optimization. *Operations Research* **62**(6) 1358–1376.
- Zhen, J., F. de Ruiter, D. den Hertog. 2017. Robust optimization for models with uncertain SOC and SDP constraints. Available on Optimization-Online.
- Zhen, J., D. Den Hertog, M. Sim. 2018. Adjustable robust optimization via Fourier–Motzkin elimination. *Operations Research* **66**(4) 1086–1100.

## Appendix A: Proofs of Results

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

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

where

$$\mu = \sum_{i \in [n_z]} \mathbb{E}_{\mathbb{P}} [|\tilde{z}_i|] \geq \sum_{i \in [n_z]} \mathbb{E}_{\mathbb{P}} [z_i^2] = n_z \theta.$$

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,

$$\mathbb{E}_{\mathbb{P}} [\exp(-t|\tilde{z}_i|)] \leq \mathbb{E}_{\mathbb{P}} [\exp(-t\tilde{v}_i)]. \quad (41)$$

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:

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

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

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

Because  $\exp(-tv) - r_1 v$  is convex in  $v$ , the maximum occurs at the boundary  $v = 0$  or  $v = 1$ . Therefore, the above problem is equivalent to

$$\begin{aligned} Z_d(t) &= \min r_0 + r_1 \hat{v} \\ \text{s.t. } &r_0 \geq \exp(-t) - r_1 \\ &r_0 \geq 1 \\ &r_0, r_1 \in \mathbb{R}. \end{aligned}$$

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 (41).

By Inequality (41), we can write

$$\begin{aligned} \mathbb{P} \left[ \sum_{i \in [n_z]} |\tilde{z}_i| \leq (1 - \epsilon)\mu \right] &\leq \inf_{t > 0} \exp(t(1 - \epsilon)\mu) \prod_{i \in [n_z]} \mathbb{E}_{\mathbb{P}} [\exp(-t|\tilde{z}_i|)] \\ &\leq \inf_{t > 0} \exp(t(1 - \epsilon)\mu) \prod_{i \in [n_z]} \mathbb{E}_{\mathbb{P}} [\exp(-t\tilde{v}_i)] \\ &\leq \exp\left(-\frac{\mu\epsilon^2}{2}\right) \end{aligned}$$

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

$$\begin{aligned}
\mathbb{P} [\tilde{\mathbf{z}} \in \mathcal{U}_{c\sqrt{n_z}}] &= \mathbb{P} \left[ \sum_{i \in [n_z]} |\tilde{z}_i| \leq c\sqrt{n_z} \right] \\
&= \mathbb{P} \left[ \sum_{i \in [n_z]} |\tilde{z}_i| \leq \left( 1 - \left( 1 - \frac{c\sqrt{n_z}}{\mu} \right) \right) \mu \right] \\
&\leq \exp \left( -\frac{\mu \left( 1 - \frac{c\sqrt{n_z}}{\mu} \right)^2}{2} \right) \\
&\leq \exp \left( -\frac{n_z \theta \left( 1 - \frac{c\sqrt{n_z}}{n_z \theta} \right)^2}{2} \right) \\
&= \exp \left( -\frac{(\theta\sqrt{n_z} - c)^2}{2\theta} \right),
\end{aligned}$$

where the last inequality follows from  $\mu \geq n_z \theta$  and  $c \in (0, \theta\sqrt{n_z})$ .  $\square$

*Proof of Proposition 2.* Since the empirical optimization problem is solvable, there exists  $\hat{\mathbf{x}} \in \mathcal{X}$  such that

$$Z_0 = \frac{1}{\Omega} \sum_{\omega \in [\Omega]} g(\hat{\mathbf{x}}, \hat{\mathbf{z}}^\omega) = \frac{1}{\Omega} \sum_{\omega \in [\Omega]} g^\omega(\hat{\mathbf{x}}, \mathbf{0}).$$

Hence, for any  $\tau > Z_0$ , let  $v_\omega = g^\omega(\hat{\mathbf{x}}, \mathbf{0}) + (\tau - Z_0)$ . Observe that  $\frac{1}{\Omega} \sum_{\omega \in [\Omega]} v_\omega = \tau$  and  $v_\omega > g^\omega(\hat{\mathbf{x}}, \mathbf{0})$ . From the forth property of Definition 1, there exists  $\bar{k}^\omega \in \mathbb{R}_+$ ,  $\omega \in [\Omega]$ , such that

$$(\hat{\mathbf{x}}, v_\omega, \bar{k}^\omega) \in \bar{\mathcal{Y}}^\omega \quad \forall \omega \in [\Omega].$$

Likewise, from the third property of Definition 1, a feasibility solution of Problem (13) is to choose  $k = \max_{\omega \in [\Omega]} \{\bar{k}^\omega\}$ , together with  $\mathbf{x} = \hat{\mathbf{x}}$  and the above  $v_\omega$ ,  $\omega \in [\Omega]$ .  $\square$

*Proof of Proposition 3.* To avoid clutter, we first drop the dependency of  $\mathbf{A}, \mathbf{b}, \mathbf{c}$  on  $\mathbf{x}$ , and then we expand the squared Euclidean norm in the robust constraint  $\|\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{a}(\mathbf{x})\|_2^2 + \mathbf{b}(\mathbf{x})^\top \mathbf{z} + c(\mathbf{x}) \leq k\mathbf{z}^\top \mathbf{z} + v$ ,  $\forall \mathbf{z} \in \mathcal{E}_r$  to equivalently express it as

$$\begin{bmatrix} 1 \\ \mathbf{z} \end{bmatrix}^\top \begin{bmatrix} -c + v - \mathbf{a}^\top \mathbf{a} - (\mathbf{A}^\top \mathbf{a} + \frac{1}{2} \mathbf{b})^\top \\ -(\mathbf{A}^\top \mathbf{a} + \frac{1}{2} \mathbf{b}) \quad k\mathbf{I}_{n_z} - \mathbf{A}^\top \mathbf{A} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{z} \end{bmatrix} \geq 0 \quad \forall \mathbf{z} : \begin{bmatrix} 1 \\ \mathbf{z} \end{bmatrix}^\top \begin{bmatrix} r^2 & \mathbf{0}^\top \\ \mathbf{0} & -\mathbf{I}_{n_z} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{z} \end{bmatrix} \geq 0.$$

Applying  $\mathcal{S}$ -lemma leads to another equivalent representation of our quadratic constraint which is

$$\begin{aligned}
&\exists \lambda \geq 0 : \begin{bmatrix} -c + v + \mathbf{a}^\top \mathbf{a} - (\mathbf{A}^\top \mathbf{a} + \frac{1}{2} \mathbf{b})^\top \\ -(\mathbf{A}^\top \mathbf{a} + \frac{1}{2} \mathbf{b}) \quad k\mathbf{I}_{n_z} - \mathbf{A}^\top \mathbf{A} \end{bmatrix} - \lambda \begin{bmatrix} r^2 & \mathbf{0}^\top \\ \mathbf{0} & -\mathbf{I}_{n_z} \end{bmatrix} \succeq \mathbf{0} \\
&\iff \exists \lambda \geq 0 : \begin{bmatrix} -c + v + \lambda r^2 & -\frac{1}{2} \mathbf{b}^\top \\ -\frac{1}{2} \mathbf{b} & (k + \lambda) \mathbf{I}_{n_z} \end{bmatrix} - \begin{bmatrix} \mathbf{a}^\top \\ \mathbf{A}^\top \end{bmatrix} \mathbf{I}_{n_a} \begin{bmatrix} \mathbf{a}^\top \\ \mathbf{A}^\top \end{bmatrix}^\top \succeq \mathbf{0}.
\end{aligned}$$

Finally, using Schur's complement and recovering the dependency of  $\mathbf{A}, \mathbf{b}, \mathbf{c}$  on  $\mathbf{x}$  yields the desired result.

For the second half of the proof, as there exists  $\hat{k} > 0$  such that

$$\hat{k} \begin{bmatrix} \mathbf{I}_{n_a} & \mathbf{a}(\mathbf{x}) \\ \mathbf{a}(\mathbf{x})^\top & v - c(\mathbf{x}) \end{bmatrix} \succeq \begin{bmatrix} \mathbf{A}(\mathbf{x}) \\ -\frac{1}{2} \mathbf{b}(\mathbf{x})^\top \end{bmatrix} \begin{bmatrix} \mathbf{A}(\mathbf{x}) \\ -\frac{1}{2} \mathbf{b}(\mathbf{x})^\top \end{bmatrix}^\top.$$

By the virtue of Schur's complement, choosing a vanishing  $\lambda$  and  $k = \hat{k}$  satisfies the semidefinite constraint presented in the proposition. The proof is now complete.  $\square$

*Proof of Proposition 5.* We first consider  $0 \leq v \leq \bar{v}$ . Since  $\ell$  is a convex function, we have

$$\begin{aligned}
& \ell'(\bar{v})\bar{v} - \ell(\bar{v}) - (\ell'(v)v - \ell(v)) \\
&= \ell'(\bar{v})\bar{v} - \ell'(v)v - (\ell(\bar{v}) - \ell(v)) \\
&\geq \ell'(\bar{v})\bar{v} - \ell'(\bar{v})v - (\ell(\bar{v}) - \ell(v)) : \text{since } \ell'(v) \leq \ell'(\bar{v}) \text{ and } v \geq 0 \\
&= \ell'(\bar{v})(\bar{v} - v) - (\ell(\bar{v}) - \ell(v)) \\
&\geq (\ell(\bar{v}) - \ell(v)) - (\ell(\bar{v}) - \ell(v)) : \text{since } \ell \text{ is convex} \\
&= 0.
\end{aligned}$$

Hence, if  $0 \leq \underline{w} \leq \bar{w}$ , we have

$$\max_{v \in [\underline{w}, \bar{w}]} \{\ell'(v)v - \ell(v)\} = \ell'(\bar{w})\bar{w} - \ell(\bar{w}) = \max_{v \in \{\underline{w}, \bar{w}\}} \{\ell'(v)v - \ell(v)\}.$$

Likewise, consider  $0 \geq v \geq \underline{v}$  as follows,

$$\begin{aligned}
& \ell'(\underline{v})\underline{v} - \ell(\underline{v}) - (\ell'(v)v - \ell(v)) \\
&= \ell'(\underline{v})\underline{v} - \ell'(v)v - (\ell(\underline{v}) - \ell(v)) \\
&\geq \ell'(\underline{v})\underline{v} - \ell'(\underline{v})v - (\ell(\underline{v}) - \ell(v)) : \text{since } \ell'(v) \geq \ell'(\underline{v}) \text{ and } v \leq 0 \\
&= \ell'(\underline{v})(\underline{v} - v) - (\ell(\underline{v}) - \ell(v)) \\
&\geq (\ell(\underline{v}) - \ell(v)) - (\ell(\underline{v}) - \ell(v)) : \text{since } \ell \text{ is convex} \\
&= 0.
\end{aligned}$$

Hence, if  $\underline{w} \leq \bar{w} \leq 0$ , we have

$$\max_{v \in [\underline{w}, \bar{w}]} \{\ell'(v)v - \ell(v)\} = \ell'(\underline{w})\underline{w} - \ell(\underline{w}) = \max_{v \in \{\underline{w}, \bar{w}\}} \{\ell'(v)v - \ell(v)\}.$$

Finally, if  $\underline{w} \leq 0 \leq \bar{w}$ , we would also have

$$\max_{v \in [\underline{w}, \bar{w}]} \{\ell'(v)v - \ell(v)\} = \max_{v \in \{\underline{w}, \bar{w}\}} \{\ell'(v)v - \ell(v)\}.$$

□

*Proof of Proposition 6* Suppose that  $p$  is a norm function and define

$$\mathcal{J} = \bigcup_{\zeta \in \mathbb{R}^{n_z}} \arg \max_{(\lambda, \eta) \in \mathcal{V}} \{\lambda^\top \zeta - \eta\}.$$

We will first show that  $(\lambda^*, \eta^*) \in \mathcal{J}$  implies  $\eta^* = 0$ , that is,  $\eta$  is superfluous in the optimization problem underlying the definition of  $p$ . Suppose otherwise for the sake of a contradiction that there exists  $(\zeta^*, \lambda^*, \eta^*) \in \mathbb{R}^{n_z} \times \mathbb{R}^{n_z} \times \mathbb{R}_{++}$  such that  $(\lambda^*, \eta^*) \in \arg \max_{(\lambda, \eta) \in \mathcal{V}} \{\lambda^\top \zeta^* - \eta\}$ . It follows that  $p(\zeta^*) = (\lambda^*)^\top \zeta^* - \eta^*$  and, as  $p$  is a norm,  $p(2\zeta^*) = 2(\lambda^*)^\top \zeta^* - 2\eta^*$ . Note also that

$$p(2\zeta^*) = \max_{(\lambda, \eta) \in \mathcal{V}} \{\lambda^\top (2\zeta^*) - \eta\} \geq (\lambda^*)^\top (2\zeta^*) - \eta^*.$$

By comparing  $p(2\zeta^*)$  with its lower bound, we find  $\eta^* \leq 0$ , reaching hence a contradiction. In conclusion,  $\eta$  always vanish at optimality, and we can assume that  $\mathbf{s} = \mathbf{0}$  without any loss of generality:

$$p(\zeta) = \max_{\lambda \in \mathbb{R}^{n_z}, \mu \in \mathbb{R}^{n_\mu}} \{\lambda^\top \zeta \mid M\lambda + N\mu \leq \mathbf{t}\}.$$

Since  $p(\zeta) = p(-\zeta)$  for all  $\zeta \in \mathbb{R}^{nz}$ , the polytope  $\mathcal{Q}_1 = \{\lambda \mid M\lambda + N\mu \leq t\}$  must be identical to the polytope  $\mathcal{Q}_2 = \{\lambda \mid -M\lambda + N\mu \leq t\}$ . Otherwise, suppose  $\lambda^* \in \mathcal{Q}_1$  but  $\lambda^* \notin \mathcal{Q}_2$ , then by a separating hyperplane argument there would exist a vector  $\zeta^* \in \mathbb{R}^{nz}$  such that

$$p(\zeta^*) = \max_{\lambda \in \mathcal{Q}_1} \{\lambda^\top \zeta^*\} \geq \lambda^{*\top} \zeta^* > \max_{\lambda \in \mathcal{Q}_2} \{\lambda^\top \zeta^*\} = p(-\zeta^*),$$

which is a contradiction. Likewise, similar contradiction can be established if  $\lambda^* \in \mathcal{Q}_2$  but  $\lambda^* \notin \mathcal{Q}_1$ . Hence,  $\mathcal{Q}_1 = \mathcal{Q}_2$ . Next, we derive the dual norm  $p^*$ :

$$\begin{aligned} p^*(\zeta) &= \max_{\omega \in \mathbb{R}^{nz}} \{\omega^\top \zeta \mid p(\omega) \leq 1\} \\ &= \max_{\omega \in \mathbb{R}^{nz}} \{\omega^\top \zeta \mid \lambda^\top \omega \leq 1, \forall (\lambda, \mu) : M\lambda + N\mu \leq t\} \\ &= \max_{\omega \in \mathbb{R}^{nz}, \alpha \in \mathbb{R}_+^{nm}} \{\omega^\top \zeta \mid \alpha^\top t \leq 1, M^\top \alpha = \omega, N^\top \alpha = \mathbf{0}\} \\ &= \max_{\alpha \in \mathbb{R}_+^{nm}} \{\alpha^\top M\zeta \mid \alpha^\top t \leq 1, N^\top \alpha = \mathbf{0}\} \\ &= \min_{\mu \in \mathbb{R}^{nm}, \delta \in \mathbb{R}_+} \{\delta \mid M\zeta + N\mu \leq \delta t\}, \end{aligned}$$

where the third maximization problem constitutes a robust counterpart of the second and the fifth equation holds due to the standard linear optimization duality argument.  $\square$

*Proof of Proposition 7.* We first consider  $k > 0$ . From the definition of  $p$  in Assumption 3, we find

$$\max_{z \in \mathcal{Z}} \{\mathbf{a}^\top z - kp(z)\} = \max_{z \in \mathcal{Z}} \min_{(\lambda, \eta) \in \mathcal{V}} \{(\mathbf{a} - k\lambda)^\top z + k\eta\}. \quad (42)$$

Since this is linear optimization problem, by the standard linear optimization duality argument, this latter maximization problem can be expressed as

$$\begin{aligned} &\max_{z \in \mathcal{Z}} \min_{(\lambda, \eta) \in \mathcal{V}} \{(\mathbf{a} - k\lambda)^\top z + k\eta\} \\ &= \min_{(\lambda, \eta) \in \mathcal{V}} \max_{z \in \mathcal{Z}} \{(\mathbf{a} - k\lambda)^\top z + k\eta\} \\ &= \min_{(\lambda, \eta) \in \mathcal{V}, \beta} \{\beta^\top \mathbf{h} + k\eta \mid \beta \geq \mathbf{0}, \mathbf{H}^\top \beta = \mathbf{a} - k\lambda\} \\ &= \min_{\beta \geq \mathbf{0}, \eta \geq 0} \{\beta^\top \mathbf{h} + k\eta \mid (\mathbf{a} - \mathbf{H}^\top \beta, k\eta, k) \in \bar{\mathcal{V}}\} \\ &= \min_{\beta \geq \mathbf{0}, \eta \geq 0} \{\beta^\top \mathbf{h} + \eta \mid (\mathbf{a} - \mathbf{H}^\top \beta, \eta, k) \in \bar{\mathcal{V}}\}, \end{aligned}$$

where the first interchange between minimization and maximization is justified because  $\mathcal{V}$  is bounded, and this concludes the desired equivalence. Next, we consider the case when  $k = 0$ . Since  $\mathcal{V}$  is bounded, we must have  $\{(\lambda, \eta) \mid (\lambda, \eta, 0) \in \bar{\mathcal{V}}\} = \{(\mathbf{0}, 0)\}$ . Hence,  $(\mathbf{a} - \mathbf{H}^\top \beta, \eta, 0) \in \bar{\mathcal{V}}$  is equivalent to  $\eta = 0$  and  $\mathbf{H}^\top \beta = \mathbf{a}$ . By noting that  $\max_{z \in \mathcal{Z}} \mathbf{a}^\top z = \min_{\beta \in \mathbb{R}_+^{nh}} \{\beta^\top \mathbf{h} \mid \mathbf{H}^\top \beta = \mathbf{a}\}$ , the first half of the proposition follows.

If the penalty function  $p$  is a norm, from the derivation of the dual norm in Proposition 6, we have

$$\begin{aligned} &\min_{\beta \in \mathbb{R}_+^{nh}} \{\beta^\top \mathbf{h} \mid p^*(\mathbf{a} - \mathbf{H}^\top \beta) \leq k\} \\ &= \min_{\beta \in \mathbb{R}_+^{nh}, \mu \in \mathbb{R}^{nm}, \delta \in \mathbb{R}_+} \{\beta^\top \mathbf{h} \mid \delta \leq k, M(\mathbf{a} - \mathbf{H}^\top \beta) + N\mu \leq \delta t\} \\ &= \min_{\beta \in \mathbb{R}_+^{nh}, \mu \in \mathbb{R}^{nm}} \{\beta^\top \mathbf{h} \mid M(\mathbf{a} - \mathbf{H}^\top \beta) + N\mu \leq kt\}, \end{aligned}$$

where the second equality follows because, to ensure that  $(\mathbf{0}, 0) \in \mathcal{V}$ ,  $\mathbf{t}$  must be non-negative making the constraint  $\delta \leq k$  binding at optimality. Noting from the previous half of this proposition and the proof of Proposition 6 that, as  $\mathbf{s}$  vanishes when the penalty function is a norm, the latter minimization problem is equivalent to  $\max_{\mathbf{z} \in \mathcal{Z}} \{\mathbf{a}^\top \mathbf{z} - kp(\mathbf{z})\}$  completes the proof.  $\square$

*Proof of Proposition 8.* It suffices to show that there a feasible solution in  $\bar{\mathcal{Y}}_P$ , for any  $\mathbf{x} \in \mathcal{X}$  and  $v > g(\mathbf{x}, \mathbf{0})$ . Let  $\mathbf{q}$  be the optimal  $\mathbf{y}$  solution to Problem (26) for  $\mathbf{z} = \mathbf{0}$  that yields the objective  $g(\mathbf{x}, \mathbf{0})$ , and  $k = \mathbf{d}^\top \hat{\mathbf{q}}^\dagger$  where  $\hat{\mathbf{q}}^\dagger \in \mathbb{R}^{n_y}$  satisfies  $\mathbf{B}\hat{\mathbf{q}}^\dagger \geq \mathbf{0}$  and  $\mathbf{B}\hat{\mathbf{q}}^\dagger \geq \max_{i \in [n_f], j \in [n_z]} |\mathbf{F}_{ij}(\hat{\mathbf{x}})| \cdot \mathbf{1}$ . Note that  $\hat{\mathbf{q}}^\dagger$  always exists because of complete recourse, which we assume. The suggested solution robustly satisfies first inequality requirement in (29) because

$$\mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top (\mathbf{q} + \hat{\mathbf{q}}^\dagger \|\mathbf{z}\|_1) = \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{q} + k \|\mathbf{z}\|_1 = g(\mathbf{x}, \mathbf{0}) + k \|\mathbf{z}\|_1 \leq v + k \|\mathbf{z}\|_1.$$

Moreover, the second inequality requirement is also robustly satisfied as

$$\begin{aligned} \mathbf{B}(\mathbf{q} + \hat{\mathbf{q}}^\dagger \|\mathbf{z}\|_1) &\geq \mathbf{f}(\mathbf{x}) + \mathbf{B}\hat{\mathbf{q}}^\dagger \|\mathbf{z}\|_1 \\ &\geq \mathbf{f}(\mathbf{x}) + \|\mathbf{z}\|_1 \left\{ \max_{i \in [n_f], j \in [n_z]} |\mathbf{F}_{ij}(\mathbf{x})| \right\} \cdot \mathbf{1} \geq \mathbf{f}(\mathbf{x}) + \mathbf{F}(\mathbf{x})\mathbf{z}, \quad \forall \mathbf{z} \in \mathcal{Z} \end{aligned}$$

where the first two inequalities follow from the feasibility of  $\mathbf{y} = \mathbf{q}$  in Problem (26) for  $\mathbf{z} = \mathbf{0}$ , and the construction of  $\hat{\mathbf{q}}^\dagger$ , respectively. Consequently,  $(\mathbf{x}, v, k) \in \bar{\mathcal{Y}}_P$ . Since  $\bar{\mathcal{Y}}_P$  is increasing in  $v$  and  $k$ , if it is a tractable set, which we show next, then it must be a tractable safe approximation of  $\mathcal{Y}$ .

Regarding the explicit tractable formulation, we note that the first robust constraint of the feasible set (29) can be expressed as

$$\max_{\mathbf{z}} \{ \mathbf{d}^\top \mathbf{Q}\mathbf{z} - (k - \mathbf{d}^\top \hat{\mathbf{q}}^\dagger) \|\mathbf{z}\|_1 \mid \mathbf{H}\mathbf{z} \leq \mathbf{h} \} \leq \tau - \mathbf{d}^\top \mathbf{q} - \mathbf{c}^\top \mathbf{x}.$$

By Proposition 7, we can replace the maximization problem on the left-hand side of by a minimization problem:

$$\min_{\mathbf{w}^0} \{ \mathbf{h}^\top \mathbf{w}^0 \mid \|\mathbf{H}^\top \mathbf{w}^0 - \mathbf{Q}^\top \mathbf{d}\|_\infty \leq k - \mathbf{d}^\top \hat{\mathbf{q}}^\dagger, \mathbf{w}^0 \geq \mathbf{0} \}.$$

Similarly, the second robust constraint of the feasible set (29) can be written down as

$$\max_{\mathbf{z}} \{ (\mathbf{F}_i(\mathbf{x}) - \mathbf{B}_i \mathbf{Q})\mathbf{z} - \mathbf{B}_i \hat{\mathbf{q}}^\dagger \|\mathbf{z}\|_1 \mid \mathbf{H}\mathbf{z} \leq \mathbf{h} \} \leq \mathbf{B}_i \mathbf{q} - \mathbf{f}_i(\mathbf{x}) \quad \forall i \in [n_f],$$

whose left-hand side maximization problem can be replaced by

$$\min_{\mathbf{w}^i} \{ \mathbf{h}^\top \mathbf{w}^i \mid \|\mathbf{H}^\top \mathbf{w}^i - \mathbf{F}_i^\top(\mathbf{x}) + \mathbf{Q}^\top \mathbf{B}_i^\top\|_\infty \leq \mathbf{B}_i \hat{\mathbf{q}}^\dagger, \mathbf{w}^i \geq \mathbf{0} \} \quad \forall i \in [n_f],$$

Finally, the two deterministic linear constraints (namely,  $k - \mathbf{d}^\top \hat{\mathbf{q}}^\dagger \geq 0$  and  $\mathbf{B}\hat{\mathbf{q}}^\dagger \geq \mathbf{0}$ ) of the feasible set (29) are redundant in view of the feasible set (30) and can therefore be safely omitted. The proof is thus completed.  $\square$

*Proof of Proposition 9.* We first express the budgeted norm in Example 3 as well as the support sets  $\mathcal{Z}^\omega = [\underline{\mathbf{z}} - \hat{\mathbf{z}}^\omega, \bar{\mathbf{z}} - \hat{\mathbf{z}}^\omega]$ ,  $\omega \in [\Omega]$ , as

$$p_\Gamma(\zeta) = \max_{\lambda, \mu \in \mathbb{R}^n} \left\{ \lambda^\top \zeta \left( \underbrace{\begin{bmatrix} \mathbf{0}^\top \\ \mathbf{I} \\ -\mathbf{I} \\ \mathbf{0} \end{bmatrix}}_M \lambda + \underbrace{\begin{bmatrix} \mathbf{1}^\top \\ -\mathbf{I} \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}}_N \mu \leq \underbrace{\begin{bmatrix} \Gamma \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{1} \end{bmatrix}}_t \right) \right\} \quad \text{and} \quad \mathcal{Z}^\omega = \left\{ \mathbf{z} \in \mathbb{R}^{n_x} \left( \underbrace{\begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \end{bmatrix}}_H \mathbf{z} \leq \underbrace{\begin{bmatrix} \bar{\mathbf{z}}^\omega \\ -\underline{\mathbf{z}}^\omega \end{bmatrix}}_{h^\omega} \right) \right\},$$

where  $\bar{\mathbf{z}}^\omega = \bar{\mathbf{z}} - \hat{\mathbf{z}}^\omega$  and  $\underline{\mathbf{z}}^\omega = \underline{\mathbf{z}} - \hat{\mathbf{z}}^\omega$ . Then, we express the evaluation function  $g^\omega(\mathbf{x}, \mathbf{z})$ ,  $\omega \in [\Omega]$ , as

$$\begin{aligned} g^\omega(\mathbf{x}, \mathbf{z}) = \text{minimize} \quad & y_1 + Py_2 \\ \text{subject to} \quad & \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}}_{\mathbf{y}} \succeq_{\mathcal{K}} \underbrace{\begin{bmatrix} \mathbf{x}^\top \hat{\mathbf{z}}^\omega \\ -P \\ 0 \\ 1 \end{bmatrix}}_{f^\omega(\mathbf{x})} + \underbrace{\begin{bmatrix} \mathbf{x}^\top \\ \mathbf{0}^\top \\ \mathbf{0}^\top \\ \mathbf{0}^\top \end{bmatrix}}_{\mathbf{F}(\mathbf{x})} \mathbf{z} \\ & \mathbf{y} \in \mathbb{R}^2, \end{aligned}$$

where  $\mathcal{K} = \mathcal{K}_{\ell_a} \times \mathbb{R}_+$ . The safe approximation of Problem (38), which is due to (25), reads

$$\begin{aligned} & \text{minimize} \quad k \\ & \text{subject to} \quad \frac{1}{\Omega} \sum_{\omega \in [\Omega]} v_\omega \leq \tau \\ & \mathbf{x}^\top \hat{\mathbf{z}}^\omega \rho_1 - P\rho_2 + \rho_4 + (\bar{\mathbf{z}}^\omega)^\top \bar{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) - (\underline{\mathbf{z}}^\omega)^\top \underline{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) \leq v_\omega \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \mathbf{1}^\top \boldsymbol{\mu}^\omega(\boldsymbol{\rho}) \leq \Gamma k \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \mathbf{x}\rho_1 - \bar{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) + \underline{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) - \boldsymbol{\mu}^\omega(\boldsymbol{\rho}) \leq \mathbf{0} \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & -\mathbf{x}\rho_1 + \bar{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) - \underline{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) - \boldsymbol{\mu}^\omega(\boldsymbol{\rho}) \leq \mathbf{0} \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \boldsymbol{\mu}^\omega(\boldsymbol{\rho}) \leq \mathbf{1}k \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \bar{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) \geq \mathbf{0}, \underline{\boldsymbol{\beta}}^\omega(\boldsymbol{\rho}) \geq \mathbf{0} \quad \forall \boldsymbol{\rho} \in \mathcal{P}, \omega \in [\Omega] \\ & \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+, \mathbf{v} \in \mathbb{R}^\Omega, \bar{\boldsymbol{\beta}}^1, \dots, \bar{\boldsymbol{\beta}}^\Omega \in \mathcal{L}^{4,n}, \underline{\boldsymbol{\beta}}^1, \dots, \underline{\boldsymbol{\beta}}^\Omega \in \mathcal{L}^{4,n}, \boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^\Omega \in \mathcal{L}^{4,n}. \end{aligned} \tag{43}$$

where the dual uncertainty set  $\mathcal{P}$  is

$$\mathcal{P} = \left\{ \boldsymbol{\rho} \in \mathcal{K}^* \mid \mathbf{B}^\top \boldsymbol{\rho} = \begin{bmatrix} 1 \\ P \end{bmatrix} \right\}.$$

Problem (39) is then obtained by characterizing each  $\bar{\boldsymbol{\beta}}^\omega \in \mathcal{L}^{4,n}$  ( $\underline{\boldsymbol{\beta}}^\omega \in \mathcal{L}^{4,n}$ ) as  $\bar{\boldsymbol{\beta}}^{\omega,0} + \sum_{j \in [4]} \bar{\boldsymbol{\beta}}^{\omega,j} \rho_j$  for some  $\bar{\boldsymbol{\beta}}^{\omega,0}, \dots, \bar{\boldsymbol{\beta}}^{\omega,4} \in \mathbb{R}^n$  ( $\underline{\boldsymbol{\beta}}^{\omega,0} + \sum_{j \in [4]} \underline{\boldsymbol{\beta}}^{\omega,j} \rho_j$  for some  $\underline{\boldsymbol{\beta}}^{\omega,0}, \dots, \underline{\boldsymbol{\beta}}^{\omega,4} \in \mathbb{R}^n$ ). Analogous treatment is applied for each  $\boldsymbol{\mu}^\omega \in \mathcal{L}^{4,n}$ .

To complete the proof, we finally note that

$$\begin{aligned} \forall \boldsymbol{\rho} \in \mathcal{P} : \mathbf{v}^\top \boldsymbol{\rho} \leq w & \iff \max_{\boldsymbol{\rho} \in \mathcal{P}} \mathbf{v}^\top \boldsymbol{\rho} \leq w \\ & \iff \min_{\mathbf{y} \in \mathcal{K}} \{y_1 + Py_2 : \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{v}\} \leq w \\ & \iff \exists \mathbf{y} \in \mathbb{R}^2 : \mathbf{B}\mathbf{y} \succeq_{\mathcal{K}} \mathbf{v}, y_1 + Py_2 \leq w \\ & \iff \exists \mathbf{y} \in \mathbb{R}^2 : \begin{cases} y_1 + Py_2 \leq w \\ (y_2 - v_3)(\exp(-av_1/(y_2 - v_3)) - 1)/a \leq (y_1 - v_2) \\ y_2 - v_3 \geq 0 \\ y_2 - v_4 \geq 0 \end{cases} \\ & \iff \exists y \in \mathbb{R} : \begin{cases} (y - v_3)(\exp(-av_1/(y - v_3)) - 1)/a + v_2 + Py \leq w \\ y \geq v_3 \\ y \geq v_4, \end{cases} \end{aligned}$$

where the last equivalence is due to the elimination of  $y_1$  and renaming  $y_2$  as  $y$ .  $\square$



**Appendix B: Parameters used in Problem (35)**

$$\begin{aligned}
\mathbf{H} &= \begin{bmatrix} 0.2220, 0.6117, 0.0807, 0.2741, 0.5999, 0.1442, 0.0243, 0.5777, 0.2591 \\ 0.8707, 0.7659, 0.7384, 0.4142, 0.2658, 0.1656, 0.2046, 0.0016, 0.8025 \\ 0.2067, 0.5184, 0.4413, 0.2961, 0.2847, 0.9639, 0.6998, 0.5155, 0.8705 \\ 0.9186, 0.2968, 0.1583, 0.6288, 0.2536, 0.9602, 0.7795, 0.6398, 0.9227 \\ 0.4884, 0.1877, 0.8799, 0.5798, 0.3276, 0.1884, 0.0229, 0.9856, 0.0022 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \mathbf{h} = \begin{bmatrix} 0.4695 \\ 0.9815 \\ 0.3989 \\ 0.8137 \\ 0.5465 \\ 100 \\ 100 \\ 100 \\ 100 \\ 100 \\ 100 \\ 100 \\ 100 \\ 100 \end{bmatrix}, \\
\mathbf{F} &= \begin{bmatrix} 0.8248, 0.5464, 0.3655, 0.6389, 0.9435, 0.1008, 0.3715, 0.4783, 0.8005 \\ 0.0942, 0.7961, 0.2443, 0.4934, 0.1117, 0.3834, 0.0124, 0.8500, 0.0204 \\ 0.3610, 0.0511, 0.7951, 0.5835, 0.8436, 0.5104, 0.8597, 0.5147, 0.5726 \\ 0.0355, 0.1887, 0.3521, 0.9393, 0.3460, 0.9611, 0.1111, 0.4466, 0.4114 \end{bmatrix}, \\
\mathbf{d} &= \begin{bmatrix} 0.8167 \\ 0.5661 \end{bmatrix}^\top, \quad \mathbf{f} = \begin{bmatrix} 0.6354 \\ 0.8119 \\ 0.9267 \\ 0.9126 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.7709, 0.1115 \\ 0.4849, 0.2512 \\ 0.0291, 0.9649 \\ 0.0865, 0.6318 \end{bmatrix}.
\end{aligned}$$

**Appendix C: Additional study on the adaptive network lot-sizing problem**

Next to demonstrate the handling of two-stage linear optimization under uncertainty using a satisficing approach, we consider a network lot-sizing example similar to Bertsimas and de Ruiter (2016) and de Ruiter et al. (2018). Suppose that there are  $n$  nodes, each of which faces a random demand  $z_i$ ,  $i \in [n]$ . Throughout, the support set of  $\mathbf{z}$  is assumed to be a hyperrectangle, *i.e.*,  $\mathcal{Z} = \{\mathbf{z} \in \mathbb{R}_+^{n_z} \mid \mathbf{z} \leq \bar{\mathbf{z}}\}$ . The initial stock  $x_i \geq 0$  at each node is to be determined prior to the realization of the random demands. Similarly, we impose that  $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}_+^{n_x} \mid \mathbf{x} \leq \bar{\mathbf{x}}\}$ . After observing the demand, we can transport stock  $y_{ij} \geq 0$  from node  $i$  to node  $j$ . To ensure that the demands can always be fulfilled, we allow for an emergency order  $w_i \geq 0$  to be made at each node. Initial and emergency orders are purchased at the unit costs  $c_i \geq 0$  and  $\ell_i \geq c_i$ , respectively, while the unit transportation costs are denoted by  $t_{ij} \geq 0$ . It is assumed that  $t_{ij}$  is equal to the distance between the two nodes, and hence  $t_{ij} = t_{ji}$ .

First, we present the robust variant of the lot-sizing problem

$$\begin{aligned}
Z_r = \text{minimize} \quad & \mathbf{c}^\top \mathbf{x} + x_0 \\
\text{subject to} \quad & \mathbf{x} + \mathbf{Y}(\mathbf{z})^\top \mathbf{1} - \mathbf{Y}(\mathbf{z})\mathbf{1} + \mathbf{w}(\mathbf{z}) \geq \mathbf{z} & \forall \mathbf{z} \in \mathcal{U}_r \\
& \langle \mathbf{T}, \mathbf{Y}(\mathbf{z}) \rangle + \boldsymbol{\ell}^\top \mathbf{w}(\mathbf{z}) \leq x_0 & \forall \mathbf{z} \in \mathcal{U}_r \\
& \mathbf{Y}(\mathbf{z}) \geq \mathbf{0}, \mathbf{w}(\mathbf{z}) \geq \mathbf{0} & \forall \mathbf{z} \in \mathcal{U}_r \\
& \mathbf{x} \in \mathcal{X}, x_0 \in \mathbb{R}, \mathbf{Y} \in \mathcal{R}^{n,n \times n}, \mathbf{w} \in \mathcal{R}^{n,n},
\end{aligned} \tag{44}$$

where  $\mathcal{U}_r = \{\mathbf{z} \in \mathbb{R}_+^n \mid \mathbf{z} \leq \bar{\mathbf{z}}, \mathbf{1}^\top \mathbf{z} \leq r\}$  denotes the uncertainty set and the epigraph variable  $x_0$  captures the worst-case transportation and emergency purchase costs. Similar to the proof of Theorem 3, we can dualize Problem (44) twice (the first time over the second-stage decisions  $\mathbf{Y}, \mathbf{w}$  and the second time over the uncertain demands  $\mathbf{z}$ ) to obtain an alternative formulation.

PROPOSITION 10. *Problem (44) is equivalent to*

$$\begin{aligned}
Z_r = \text{minimize} \quad & \mathbf{c}^\top \mathbf{x} + x_0 \\
\text{subject to} \quad & \boldsymbol{\beta}(\boldsymbol{\rho})^\top \bar{\mathbf{z}} + \hat{\boldsymbol{\beta}}(\boldsymbol{\rho})r \leq \boldsymbol{\rho}^\top \mathbf{x} + x_0 & \forall \boldsymbol{\rho} \in \mathcal{P} \\
& \boldsymbol{\beta}(\boldsymbol{\rho}) + \hat{\boldsymbol{\beta}}(\boldsymbol{\rho})\mathbf{1} \geq \boldsymbol{\rho} & \forall \boldsymbol{\rho} \in \mathcal{P} \\
& \boldsymbol{\beta}(\boldsymbol{\rho}) \geq \mathbf{0}, \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}) \geq 0 & \forall \boldsymbol{\rho} \in \mathcal{P} \\
& \mathbf{x} \in \mathcal{X}, x_0 \in \mathbb{R}, \boldsymbol{\beta} \in \mathcal{R}^{n,n}, \hat{\boldsymbol{\beta}} \in \mathcal{R}^{n,1},
\end{aligned} \tag{45}$$

where the dual uncertainty set  $\mathcal{P}$  is given by  $\{\boldsymbol{\rho} \in \mathbb{R}_+^n : \boldsymbol{\rho} \leq \boldsymbol{\ell}, \boldsymbol{\rho}\mathbf{1}^\top - \mathbf{1}\boldsymbol{\rho}^\top \leq \mathbf{T}\}$ .

*Proof.* We could express the robust constraints of Problem (44) as

$$\begin{aligned}
& \max_{\mathbf{z} \in \mathcal{U}_r} \min_{\mathbf{Y} \geq \mathbf{0}, \mathbf{w} \geq \mathbf{0}} \{ \langle \mathbf{T}, \mathbf{Y} \rangle + \boldsymbol{\ell}^\top \mathbf{w} \mid \mathbf{x} + \mathbf{w} + \mathbf{Y}^\top \mathbf{1} - \mathbf{Y}\mathbf{1} \geq \mathbf{z} \} \leq x_0 \\
& \iff \max_{\mathbf{z} \in \mathcal{U}_r} \min_{\mathbf{Y} \geq \mathbf{0}, \mathbf{w} \geq \mathbf{0}} \max_{\boldsymbol{\rho} \geq \mathbf{0}} \{ \langle \mathbf{T}, \mathbf{Y} \rangle + \boldsymbol{\ell}^\top \mathbf{w} + \boldsymbol{\rho}^\top (\mathbf{z} - \mathbf{x} - \mathbf{w} + \mathbf{Y}\mathbf{1} - \mathbf{Y}^\top \mathbf{1}) \} \leq x_0 \\
& \iff \max_{\mathbf{z} \in \mathcal{U}_r} \max_{\boldsymbol{\rho} \geq \mathbf{0}} \left\{ \boldsymbol{\rho}^\top (\mathbf{z} - \mathbf{x}) + \min_{\mathbf{Y} \geq \mathbf{0}, \mathbf{w} \geq \mathbf{0}} \{ \langle \mathbf{Y}, \mathbf{T} + \boldsymbol{\rho}\mathbf{1}^\top - \mathbf{1}\boldsymbol{\rho}^\top \rangle + (\boldsymbol{\ell} - \boldsymbol{\rho})^\top \mathbf{w} \} \right\} \leq x_0 \\
& \iff \max_{\mathbf{z} \in \mathcal{U}_r} \max_{\boldsymbol{\rho} \in \mathcal{P}} \{ \boldsymbol{\rho}^\top (\mathbf{z} - \mathbf{x}) \} \leq x_0 \\
& \iff \max_{\boldsymbol{\rho} \in \mathcal{P}} \min_{\boldsymbol{\beta} \geq \mathbf{0}, \hat{\boldsymbol{\beta}} \geq 0} \{ \boldsymbol{\beta}^\top \bar{\mathbf{z}} + \hat{\boldsymbol{\beta}}r - \boldsymbol{\rho}^\top \mathbf{x} \mid \boldsymbol{\beta} + \hat{\boldsymbol{\beta}}\mathbf{1} \geq \boldsymbol{\rho} \} \leq x_0.
\end{aligned}$$

Treating  $\boldsymbol{\rho}$  as uncertainty as well as  $\boldsymbol{\beta}$  and  $\hat{\boldsymbol{\beta}}$  as dual recourse variables finally completes the proof.  $\square$

For the robust satisficing variant of the problem, we consider

$$\begin{aligned}
& \text{minimize} \quad k \\
& \text{subject to} \quad \mathbf{x} + \mathbf{Y}(\mathbf{z})^\top \mathbf{1} - \mathbf{Y}(\mathbf{z})\mathbf{1} + \mathbf{w}(\mathbf{z}) \geq \mathbf{z} & \forall \mathbf{z} \in \mathcal{Z} \\
& \quad \mathbf{c}^\top \mathbf{x} + \langle \mathbf{T}, \mathbf{Y}(\mathbf{z}) \rangle + \boldsymbol{\ell}^\top \mathbf{w}(\mathbf{z}) \leq \tau + k\|\mathbf{z}\|_1 & \forall \mathbf{z} \in \mathcal{Z} \\
& \quad \mathbf{Y}(\mathbf{z}) \geq \mathbf{0}, \mathbf{w}(\mathbf{z}) \geq \mathbf{0} & \forall \mathbf{z} \in \mathcal{Z} \\
& \quad \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+, \mathbf{Y} \in \mathcal{R}^{n,n \times n}, \mathbf{w} \in \mathcal{R}^{n,n},
\end{aligned} \tag{46}$$

where  $\tau \geq Z_0 = 0$  is the prescribed target objective. The dualized formulation (27) corresponding to this problem is explicitly given below.

PROPOSITION 11. *Problem (46) is equivalent to*

$$\begin{aligned}
& \text{minimize} && k \\
& \text{subject to} && (\mathbf{c} - \boldsymbol{\rho})^\top \mathbf{x} + \boldsymbol{\beta}(\boldsymbol{\rho})^\top \bar{\mathbf{z}} \leq \tau \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\
& && \boldsymbol{\beta}(\boldsymbol{\rho}) + k\mathbf{1} \geq \boldsymbol{\rho} \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\
& && \boldsymbol{\beta}(\boldsymbol{\rho}) \geq \mathbf{0} \quad \forall \boldsymbol{\rho} \in \mathcal{P} \\
& && \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+, \boldsymbol{\beta} \in \mathcal{R}^{n,n},
\end{aligned} \tag{47}$$

where the dual uncertainty set  $\mathcal{P}$  is given by  $\{\boldsymbol{\rho} \in \mathbb{R}_+^n : \boldsymbol{\rho} \leq \boldsymbol{\ell}, \boldsymbol{\rho}\mathbf{1}^\top - \mathbf{1}\boldsymbol{\rho}^\top \leq \mathbf{T}\}$ .

*Proof.* The proof widely parallels to that of Proposition 10 and is thus omitted.  $\square$

Due to the linearity of the lot-sizing objective and constraint functions, we can solve Problem (44) approximately using an affine adaptation on either the primal or the dual recourse. These approximations turn out to be equivalent; see Proposition 12 below and also Bertsimas and de Ruiter (2016) who first establish a similar equivalence for a variant of the lot-sizing problem without emergency orders. For the robust satisficing problem (46), Long et al. (2022) explore how it could be solved using a primal affine adaptation, whereas here we demonstrate how the same problem could be solved using a dual affine adaptation. Perhaps unsurprisingly, we leverage Theorems 1 and 2 of Bertsimas and de Ruiter (2016) to establish in Proposition 13 that these two approximations are also equivalent. Though as earlier shown, we emphasize that for a broader class of two-stage robust satisficing linear optimization problems with complete recourse, the dual affine adaptation can provide a *strictly* less conservative approximation compared to the primal affine adaptation.

PROPOSITION 12. *Problems (44) and (45) attain the same objective value when solved approximately using affine recourse adaptation.*

*Proof.* In their Theorem 1, Bertsimas and de Ruiter (2016) show an equivalent reformulation of Problem (44) which is

$$\begin{aligned}
& \text{minimize} && \mathbf{c}^\top \mathbf{x} + x_0 \\
& \text{subject to} && \boldsymbol{\beta}(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}})^\top \bar{\mathbf{z}} + \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}})r \leq \boldsymbol{\rho}^\top \mathbf{x} + \hat{\rho}x_0 \quad \forall (\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) \in \hat{\mathcal{P}} \\
& && \boldsymbol{\beta}(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) + \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}})\mathbf{1} \geq \boldsymbol{\rho} \quad \forall (\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) \in \hat{\mathcal{P}} \\
& && \boldsymbol{\beta}(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) \geq \mathbf{0}, \hat{\boldsymbol{\beta}}(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) \geq 0 \quad \forall (\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) \in \hat{\mathcal{P}} \\
& && \mathbf{x} \in \mathcal{X}, x_0 \in \mathbb{R}, \boldsymbol{\beta} \in \mathcal{R}^{n+1,n}, \hat{\boldsymbol{\beta}} \in \mathcal{R}^{n+1,1},
\end{aligned} \tag{48}$$

where  $\hat{\mathcal{P}} = \{(\boldsymbol{\rho}, \hat{\boldsymbol{\rho}}) \in \mathbb{R}_+^n \times \mathbb{R}_+ \mid \boldsymbol{\rho} \leq \hat{\rho}\boldsymbol{\ell}, \boldsymbol{\rho}\mathbf{1}^\top - \mathbf{1}\boldsymbol{\rho}^\top \leq \hat{\rho}\mathbf{T}, \mathbf{1}^\top \boldsymbol{\rho} + \hat{\rho} = 1\}$ . They also argue in their Theorem 2 that the respective affine recourse approximations of Problems (44) and (48) are equivalent. As a result, it suffices to show that the affine recourse approximations of Problems (45) and (48) are equivalent.

First, we write down the affine recourse approximation of Problem (45) by restricting  $\beta(\rho)$  and  $\hat{\beta}(\rho)$  to  $\beta^i + \beta^s \rho$  and  $\hat{\beta}^i + \hat{\beta}^s \rho$  (where ‘i’ and ‘s’ indicate the intercept and slope of the affine decision rules), respectively, and obtain

$$\begin{aligned}
& \text{minimize} && \mathbf{c}^\top \mathbf{x} + x_0 \\
& \text{subject to} && \bar{\mathbf{z}}^\top \beta^i + r\hat{\beta}^i + (\bar{\mathbf{z}}^\top \beta^s + r\hat{\beta}^s) \rho \leq \rho^\top \mathbf{x} + x_0 && \forall \rho \in \mathcal{P} \\
& && \beta^i + \mathbf{1}\hat{\beta}^i + (\beta^s + \mathbf{1}\hat{\beta}^s) \rho \geq \rho && \forall \rho \in \mathcal{P} \\
& && \beta^i + \beta^s \rho \geq \mathbf{0}, \quad \hat{\beta}^i + \hat{\beta}^s \rho \geq 0 && \forall \rho \in \mathcal{P} \\
& && \mathbf{x} \in \mathcal{X}, x_0 \in \mathbb{R}, \beta^i \in \mathbb{R}^n, \beta^s \in \mathbb{R}^{n \times n}, \hat{\beta}^i \in \mathbb{R}, \hat{\beta}^s \in \mathbb{R}^{1 \times n}.
\end{aligned} \tag{49}$$

Next, for Problem (48), we first observe that the uncertainty set  $\hat{\mathcal{P}}$  requires  $\hat{\rho}$  to be linearly dependent on  $\rho$ . Hence, we can simply ignore the additional uncertain parameter  $\hat{\rho}$  and work with the projection of  $\hat{\mathcal{P}}$  on  $\rho$ . By a slight abuse of notation, we will denote this projection by  $\hat{\mathcal{P}}$  and note that

$$\hat{\mathcal{P}} = \{\rho \in \mathbb{R}_+^n \mid \rho \leq (1 - \mathbf{1}^\top \rho) \ell, \rho \mathbf{1}^\top - \mathbf{1} \rho^\top \leq (1 - \mathbf{1}^\top \rho) \mathbf{T}\}.$$

Note that as  $\hat{\mathcal{P}} \subset \mathbb{R}_+^n$  and as  $\ell \geq \mathbf{0}$ , it is a necessity that  $\mathbf{1}^\top \rho < 1$  for all  $\rho \in \hat{\mathcal{P}}$ . We are now ready to present the explicit affine recourse approximation of Problem (48), which is

$$\begin{aligned}
& \text{minimize} && \mathbf{c}^\top \mathbf{x} + x_0 \\
& \text{subject to} && \bar{\mathbf{z}}^\top \beta^i + r\hat{\beta}^i + (\bar{\mathbf{z}}^\top \beta^s + r\hat{\beta}^s) \rho \leq \rho^\top \mathbf{x} + (1 - \mathbf{1}^\top \rho) x_0 && \forall \rho \in \hat{\mathcal{P}} \\
& && \beta^i + \mathbf{1}\hat{\beta}^i + (\beta^s + \mathbf{1}\hat{\beta}^s) \rho \geq \rho && \forall \rho \in \hat{\mathcal{P}} \\
& && \beta^i + \beta^s \rho \geq \mathbf{0}, \quad \hat{\beta}^i + \hat{\beta}^s \rho \geq 0 && \forall \rho \in \hat{\mathcal{P}} \\
& && \mathbf{x} \in \mathcal{X}, x_0 \in \mathbb{R}, \beta^i \in \mathbb{R}^n, \beta^s \in \mathbb{R}^{n \times n}, \hat{\beta}^i \in \mathbb{R}, \hat{\beta}^s \in \mathbb{R}^{1 \times n}.
\end{aligned} \tag{50}$$

It remains to show that Problems (49) and (50) are equivalent. First, we will show that Problem (50) is a relaxation of Problem (49). To this end, for any feasible solution  $\mathbf{X} = (\mathbf{x}, x_0, \beta^i, \beta^s, \hat{\beta}^i, \hat{\beta}^s)$  of Problem (49), we will show that  $\mathbf{X}' = (\mathbf{x}, x_0, \beta^i, \beta^s - \beta^i \mathbf{1}^\top, \hat{\beta}^i, \hat{\beta}^s - \hat{\beta}^i \mathbf{1}^\top)$  is feasible in Problem (50). Note that, as both solutions share the same  $\mathbf{x}$  and the same  $x_0$ , they attain the same objective value in their respective problem.

For any  $\rho \in \hat{\mathcal{P}}$ , it is readily seen that  $\rho / (1 - \mathbf{1}^\top \rho) \in \mathcal{P}$ . As a result, the feasibility of  $\mathbf{X}$  in view of Problem (49) implies, for all  $\rho \in \hat{\mathcal{P}}$ , that

$$\begin{aligned}
& (1 - \mathbf{1}^\top \rho) (\bar{\mathbf{z}}^\top \beta^i + r\hat{\beta}^i) + (\bar{\mathbf{z}}^\top \beta^s + \hat{\beta}^s) \rho \leq \rho^\top \mathbf{x} + (1 - \mathbf{1}^\top \rho) x_0, \\
& (1 - \mathbf{1}^\top \rho) (\beta^i + \mathbf{1}\hat{\beta}^i) + (\beta^s + \mathbf{1}\hat{\beta}^s) \rho \geq \rho, \\
& (1 - \mathbf{1}^\top \rho) \beta^i + \beta^s \rho \geq \mathbf{0}, \\
& (1 - \mathbf{1}^\top \rho) \hat{\beta}^i + \hat{\beta}^s \rho \geq 0.
\end{aligned}$$

Rearranging terms in the above four inequalities yields

$$\begin{aligned}
& \bar{\mathbf{z}}^\top \beta^i + r\hat{\beta}^i + \left( \bar{\mathbf{z}}^\top (\beta^s - \beta^i \mathbf{1}^\top) + r(\hat{\beta}^s - \hat{\beta}^i \mathbf{1}^\top) \right) \rho \leq \rho^\top \mathbf{x} + (1 - \mathbf{1}^\top \rho) x_0, \\
& \beta^i + \mathbf{1}\hat{\beta}^i + \left( (\beta^s - \beta^i \mathbf{1}^\top) + \mathbf{1}(\hat{\beta}^s - \hat{\beta}^i \mathbf{1}^\top) \right) \rho \geq \rho, \\
& \beta^i + (\beta^s - \beta^i \mathbf{1}^\top) \rho \geq \mathbf{0}, \\
& \hat{\beta}^i + (\hat{\beta}^s - \hat{\beta}^i \mathbf{1}^\top) \rho \geq 0,
\end{aligned}$$

for all  $\boldsymbol{\rho} \in \hat{\mathcal{P}}$ , which in turn implies that  $\mathbf{X}'$  is indeed feasible in Problem (50).

Conversely, for any feasible solution  $\mathbf{X} = (\mathbf{x}, x_0, \boldsymbol{\beta}^i, \boldsymbol{\beta}^s, \hat{\boldsymbol{\beta}}^i, \hat{\boldsymbol{\beta}}^s)$  of Problem (50), one can similarly show that  $\mathbf{X}' = (\mathbf{x}, x_0, \boldsymbol{\beta}^i, \boldsymbol{\beta}^s + \boldsymbol{\beta}^i \mathbf{1}^\top, \hat{\boldsymbol{\beta}}^i, \hat{\boldsymbol{\beta}}^s + \hat{\boldsymbol{\beta}}^i \mathbf{1}^\top)$  is feasible in Problem (49) to conclude that Problem (49) is a relaxation of Problem (50). To see this, observe that  $\boldsymbol{\rho}/(1 + \mathbf{1}^\top \boldsymbol{\rho}) \in \hat{\mathcal{P}}$  for any  $\boldsymbol{\rho} \in \mathcal{P}$ . As a result, the feasibility of  $\mathbf{X}$  in view of Problem (50) implies, for all  $\boldsymbol{\rho} \in \mathcal{P}$ , that

$$\begin{aligned} (1 + \mathbf{1}^\top \boldsymbol{\rho}) (\bar{\mathbf{z}}^\top \boldsymbol{\beta}^i + r \hat{\boldsymbol{\beta}}^i) + (\bar{\mathbf{z}}^\top \boldsymbol{\beta}^s + r \hat{\boldsymbol{\beta}}^s) \boldsymbol{\rho} &\leq \boldsymbol{\rho}^\top \mathbf{x} + x_0, \\ (1 + \mathbf{1}^\top \boldsymbol{\rho}) (\boldsymbol{\beta}^i + \mathbf{1} \hat{\boldsymbol{\beta}}^i) + (\boldsymbol{\beta}^s + \mathbf{1} \hat{\boldsymbol{\beta}}^s) \boldsymbol{\rho} &\geq \boldsymbol{\rho}, \\ (1 + \mathbf{1}^\top \boldsymbol{\rho}) \boldsymbol{\beta}^i + \boldsymbol{\beta}^s \boldsymbol{\rho} &\geq \mathbf{0}, \\ (1 + \mathbf{1}^\top \boldsymbol{\rho}) \hat{\boldsymbol{\beta}}^i + \hat{\boldsymbol{\beta}}^s \boldsymbol{\rho} &\geq \mathbf{0}. \end{aligned}$$

Rearranging the terms in the above inequalities shows that  $\mathbf{X}'$  is indeed feasible in Problem (49) as desired. Therefore, the optimal objective value of Problem (49) constitutes both a lower and an upper bound of that of Problem (50). The proof is hence completed.  $\square$

**PROPOSITION 13.** *Problems (46) and (47) attain the same objective value when solved approximately using affine recourse adaptation.*

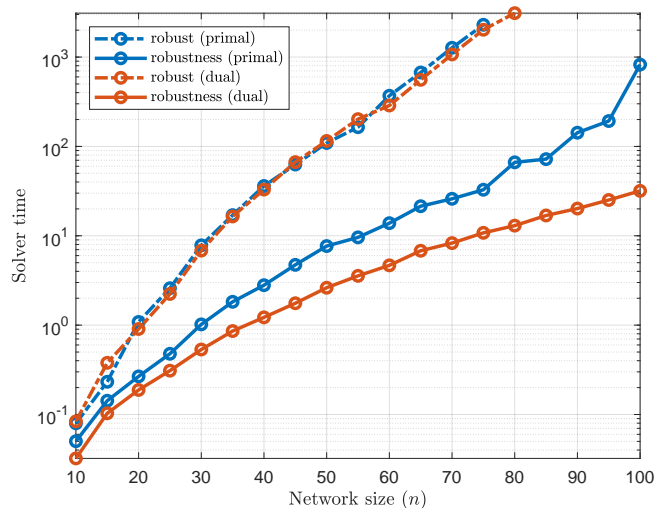
*Proof.* First, we can use Bertsimas and de Ruiter (2016, Theorem 1) to argue that Problem (46) is equivalent to

$$\begin{aligned} &\text{minimize} && k \\ &\text{subject to} && (\hat{\rho} \mathbf{c} - \boldsymbol{\rho})^\top \mathbf{x} + \bar{\mathbf{z}}^\top \boldsymbol{\beta}(\boldsymbol{\rho}, \hat{\rho}) \leq \hat{\rho} \tau \quad \forall (\boldsymbol{\rho}, \hat{\rho}) \in \hat{\mathcal{P}} \\ & && \boldsymbol{\beta}(\boldsymbol{\rho}, \hat{\rho}) + k \hat{\rho} \mathbf{1} \geq \boldsymbol{\rho} \quad \forall (\boldsymbol{\rho}, \hat{\rho}) \in \hat{\mathcal{P}} \\ & && \boldsymbol{\beta}(\boldsymbol{\rho}, \hat{\rho}) \geq \mathbf{0} \quad \forall (\boldsymbol{\rho}, \hat{\rho}) \in \hat{\mathcal{P}} \\ & && \mathbf{x} \in \mathcal{X}, k \in \mathbb{R}_+, \boldsymbol{\beta} \in \mathcal{R}^{n+1, n}, \end{aligned}$$

where  $\hat{\mathcal{P}}$  is the same as that in the proof of Proposition 12. The proof largely follows that of Proposition 12, *i.e.*, one can show that the affine adaptation approximation of the above problem is equivalent to that of (47), which in turn implies that the affine adaptation approximations of (46) and (47) are equivalent (Bertsimas and de Ruiter 2016, Theorem 2). Details are omitted for brevity.  $\square$

Our goal here is to show that the robust satisficing approach (especially when it is approximately solved using the proposed affine dual adaptation from Section 3) can be a relatively computationally friendly alternative to the classical robust optimization approach. We consider a network of a varying number of nodes  $n \in \{10, 15, \dots, 100\}$  with  $\mathcal{X} = \mathcal{Z} = [0, 20]^n$ . To ensure suitable variability among the nodes, the components of the initial ordering cost  $\mathbf{c}$  and emergency cost  $\boldsymbol{\ell}$  are generated uniformly from  $[8, 10]$  and  $[18, 20]$  respectively. We then select the node locations randomly from  $[0, 10]^2$  and accordingly compute the (Euclidean) distance matrix  $\mathbf{T}$ . Figure 4 compares the computational times required by Gurobi & RSOME in logarithmic scale of the primal robust problem (44), dual robust problem (45), primal robust satisficing problem (46) and dual robust satisficing problem (47). Using Gurobi 9.1.1 with RSOME (Robust Stochastic Optimization Made Easy) modeling language (Chen et al. 2020) and Python 3.7.7, the primal models (44) and (46) cannot be solved within a time limit of one hour (on an Intel Core i7 2.7GHz MacBook with 16GB of RAM) when the

number of nodes exceed 80. When  $n = 100$ , the dual robust model (45) takes about 15 minutes, whereas the dual robust satisficing model (47) takes only about 30 seconds. The better efficiency of the dual approaches was first hypothesized and observed in Bertsimas and de Ruiter (2016), whereas the better efficiency of the robust satisficing approach is, to our knowledge, established here for the first time.



**Figure 4** Computational times (in seconds) of the robust optimization (rob opt) and the robust satisficing (rob sat) lot-sizing solutions.