

# Globalized Adversarial Regret Optimization

## Robust Decisions with Uncalibrated Predictions

Jannis Kurtz <sup>\*1</sup> and Bart P.G. van Parys <sup>†2</sup>

<sup>1</sup>University of Amsterdam

<sup>2</sup>CWI Amsterdam

March 30, 2026

Optimization problems routinely depend on uncertain parameters that must be predicted before a decision is made. Classical robust and regret formulations are designed to handle erroneous predictions and can provide statistical error bounds in simple settings. However, when predictions lack rigorous error bounds (as is typical of modern machine learning methods) classical robust models often yield vacuous guarantees, while regret formulations can paradoxically produce decisions that are more optimistic than even a nominal solution. We introduce Globalized Adversarial Regret Optimization (GARO), a decision framework that controls adversarial regret, defined as the gap between the worst-case cost and the oracle robust cost, uniformly across all possible uncertainty set sizes. By design, GARO delivers absolute or relative performance guarantees against an oracle with full knowledge of the prediction error, without requiring any probabilistic calibration of the uncertainty set. We show that GARO equipped with a relative rate function generalizes the classical adaptation method of Lepskii (1993) to downstream decision problems. We derive exact tractable reformulations for problems with affine worst-case cost functions and polyhedral norm uncertainty sets, and provide a discretization and a constraint-generation algorithm with convergence guarantees for general settings. Finally, experiments demonstrate that GARO yields solutions with a more favorable trade-off between worst-case and mean out-of-sample performance, as well as stronger global performance guarantees.

---

\*j.kurtz@uva.nl

†bart.van.parys@cwi.nl

# 1. Introduction

Consider an optimization problem

$$\min_{x \in X} f(x, p^*) \tag{1}$$

which seeks a decision  $x \in X$  minimizing a cost function  $f : X \times P \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  given a problem parameter  $p^* \in P$ . The field of mathematical optimization, which began in earnest with Dantzig’s simplex method in 1947, has since produced powerful tools to find optimal decisions in Equation (1) algorithmically. However, before any such tools can be applied, practitioners must face the fact that the problem parameters  $p^*$  are typically subject to uncertainty. Before proceeding, a decision maker often first makes a prediction  $p_0$  of the problem parameters, and then makes a decision  $x_{nom}$  by solving a nominal “predict-then-optimize” formulation

$$\min_{x \in X} f(x, p_0). \tag{2}$$

## 1.1. From Classical to Wild Predictions

The use of predictions to facilitate downstream decision-making is an idea at least as old as mathematical optimization itself. Nevertheless, the nature of predictions has changed remarkably. We argue that downstream decision tools have largely failed to reflect these changes.

**Classical Predictions.** Predictions have classically been constructed with the help of statistical estimators. As a case in point, problem (1) may represent a vehicle routing problem (Toth and Vigo 2002) with an unknown cost vector  $p^* = \mathbb{E}[p]$  of expected edge costs. The decision-maker estimates  $p^*$  as the sample average  $p_0 = \frac{1}{N} \sum_{i=1}^N p_i$  of  $N$  historical realizations. If the decision-maker has access to *curated* data, it is common to assume that each historical observation is an independent sample sharing the same distribution as future random edge costs. In the context of such curated data, predictions often come with desirable concentration inequalities and associated error margin or confidence bounds. We use the problem of univariate mean estimation as a running example. After imposing sub-Gaussian tails  $\mathbb{E}[\exp(\lambda(p_i - p^*))] \leq \exp(\lambda^2 \sigma^2 / 2)$  for some given finite  $\sigma^2$ , it is well known that the classical empirical average predictor discussed earlier satisfies the concentration inequality  $\text{Prob}[|p_0 - p^*| > \gamma] \leq 2 \exp(-N\gamma^2 / (2\sigma^2))$ . Hence, we have  $\text{Prob}[|p_0 - p^*| \leq \gamma_N(\delta)] \geq 1 - \delta$  with  $d(p_0, p^*) = |p_0 - p^*|$  and  $\gamma_N(\delta) = \mathcal{O}(\sigma \sqrt{\log(2/\delta)/N})$ . We denote a prediction as *classical* when it enjoys an error margin guarantee

$$\text{Prob}[d(p_0, p^*) \leq \gamma_N(\delta)] \geq 1 - \delta \tag{3}$$

for any  $\delta \in (0, 1)$  and an appropriate<sup>1</sup> distance function  $d : P \times P \rightarrow \mathbb{R}_+ \cup \{\infty\}$  for all  $N \geq 1$ . Classical robust or regret optimization formulations are designed to exploit such classical predictions and are able to return decisions with rigorous performance guarantees.

**Adaptive Predictions.** It has however long been recognized that practical data is more challenging than what classical predictions can account for. We give two examples.

Many parameters related to natural phenomena or economic indicators are regularly varying, i.e.,  $\text{Prob}[p \geq x] = x^{-\nu}L(x)$  and  $L(x)$  a slowly varying function. This means that  $\mathbb{E}[\exp(\lambda(p_i - p^*))] = \infty$  for any  $\lambda > 0$  and  $\nu > 1$ , and that  $\mathbb{E}[|p_i - p^*|^a] = \infty$  for  $a > \nu$  while  $\mathbb{E}[|p_i - p^*|^a] < \infty$  for  $a < \nu$  (Nair et al. 2022). In particular, the empirical mean is a fragile estimator in this setting. Robust mean estimators which satisfy

$$\frac{1}{N} \sum_{i=1}^N \psi(\lambda_N(p_i - p_0)) = 0 \quad (4)$$

for influence function  $\psi$  and scaling parameter  $\lambda_N$  can be considered instead. The empirical mean corresponds to the case where the influence function is the identity. Assuming  $(\mathbb{E}[|p_i - p^*|^a])^{1/a} \leq \sigma_a$  for some finite  $\sigma_a < \infty$  and moment order  $a > 1$ , Bhatt et al. (2022) proposes a classical estimator  $p_0(\sigma_a)$  satisfying (3) of minimax optimal radius  $\gamma_N(\delta, \sigma_a) = \mathcal{O}(\sigma_a (\log(2/\delta)/N)^{(a-1)/a})$ . This is achieved with influence function  $\psi(\Delta) = \text{sign}(\Delta)\mathcal{O}(\log(|\Delta|^a))$  and scaling parameter  $\lambda_N(\sigma_a) = \mathcal{O}(\sigma_a^{-1}(\log(2/\delta)/N)^{1/a})$ .

Many data sets contain a small fraction  $\epsilon \in (0, 1/2)$  of the data points which are corrupted either due to malice or simple error. As was already the case in the context of heavy tail data, the empirical mean is well known to be fragile. This spurred on the development of robust alternatives. Huber (1964, 1981) advances a prediction  $p_0(\epsilon)$  based on Equation (4) with clipped influence function  $\psi(x) = \max(\min(x, 1), -1)$  which achieves (3) for minimax optimal radius  $\gamma_N(\delta, \epsilon) = \mathcal{O}(\sigma(\sqrt{\epsilon} + \sqrt{\log(2/\delta)/N}))$  with  $\lambda_N = \sigma^{-1}\sqrt{\epsilon + \log(2/\delta)/N}$  assuming only that  $\mathbb{E}[|p_i - p^*|^2] \leq \sigma^2 < \infty$ .

In the previously discussed examples the estimate  $p_0(\kappa)$  satisfies

$$\text{Prob}_\kappa [d(p_0(\kappa), p^*) \leq \gamma_N(\delta, \kappa)] \geq 1 - \delta \quad (5)$$

where  $\text{Prob}_\kappa$  denotes the probability law of the data when the structural parameter takes value  $\kappa$ . This guarantee depends critically on an auxiliary parameter  $\kappa \in K \subseteq [\kappa_{\min}, \kappa_{\max}]$  ( $\sigma_a$  and  $\epsilon$  in the case of heavy-tail or corrupt data) which is unknown in practice. As is the case in the discussed examples, we assume here that larger auxiliary parameters correspond to harder prediction problems, resulting in a nondecreasing margin  $\kappa \mapsto \gamma_N(\delta, \kappa)$  for any  $\delta, N$ . A conservative prediction  $p_0 = p_0(\kappa_{\max})$  corresponding to the worst-case parameter choice could be considered. However, adaptive guarantees

---

<sup>1</sup>We will assume throughout that the distance function is lower semicontinuous and satisfies  $d(p_0, p_0) = 0$  for all  $p_0 \in P$ .

can be obtained without resorting to the worst-case parameter. For instance, let  $\kappa_j = \kappa_{\min}\beta^j$  for  $j = 0, \dots, J$  with  $J = \lceil \log_\beta(\kappa_{\max}/\kappa_{\min}) \rceil$  and  $\beta > 1$ . Define the adaptive estimate  $p_0 = p_0(\kappa^*)$  where  $\kappa^*$  is the smallest grid point  $\kappa_j$  for which  $\{p \in P : d(p_0(\kappa_i), p) \leq \gamma_N(\delta, \kappa_i) \ \forall i \geq j\}$  is nonempty. Lepskii (1993) shows the adaptive guarantee

$$\text{Prob}_\kappa [d(p_0, p^*) \leq 2\gamma_N(\delta/(J+1), \beta\kappa)] \geq 1 - \delta \quad (6)$$

indicating performance on par with the predictor  $p_0(\kappa)$  which does have access to the unknown parameter when the distance function satisfies the triangle inequality. More generally, we denote a prediction  $p_0$  which does not depend on  $\kappa$  as *adaptive* when it enjoys an error margin guarantee

$$\text{Prob}_\kappa [d(p_0, p^*) \leq \gamma_N(\delta, \kappa)] \geq 1 - \delta. \quad (7)$$

for all  $N \geq 1$ . Although the adaptive estimates are not as conservative, their guarantees are still relative to an unknown parameter which prevents the use of downstream robust or regret optimization formulations with absolute performance guarantees.

**Wild Predictions.** Increasingly, modern machine learning methods (including deep neural networks (LeCun et al. 2015)), gradient boosted trees (Chen and Guestrin 2016), and large-scale foundation models (Bommasani et al. 2021)) produce predictions with remarkable empirical accuracy yet come with no rigorous performance guarantees (Goodfellow et al. 2015). We denote a prediction as *wild* when it is frequently empirically accurate ( $\gamma^* := d(p_0, p^*) \leq \gamma_{\text{tp}}$ ) but rigorous claims only establish  $\gamma^* \leq \gamma_{\text{wc}}$  where the ratio  $\gamma_{\text{wc}}/\gamma_{\text{tp}} \gg 1$  is very large. The formal worst-case bound  $\gamma_{\text{wc}}$  may stem, for instance, from a Lipschitz argument (Bartlett et al. 2017), a generalization bound (Bartlett and Mendelson 2002), or a smoothness assumption (Tsybakov 2009), but the resulting uncertainty set  $\{p : d(p_0, p) \leq \gamma_{\text{wc}}\}$  is so large as to render any robust formulation practically useless. Wild predictors are fundamentally different from adaptive ones: while adaptive predictions preserve a formal relationship between the radius  $\gamma_N(\delta, \kappa)$  and a structural parameter  $\kappa$ , wild predictions offer no such structure. The downstream decision-making challenge is therefore not to exploit a formal guarantee, but rather to design decisions that perform well when the prediction is accurate ( $\gamma^* \leq \gamma_{\text{tp}}$ ) while degrading gracefully relative to an appropriate benchmark when it is not ( $\gamma^* \leq \gamma_{\text{wc}}$ ).

## 1.2. Literature Review

The robust optimization literature lists several strategies to help reduce the conservatism inherent to worst-case formulations. Bertsimas and Sim (2004) show that budget uncertainty sets can to some extent help manage the conservatism. Fischetti and Monaci (2009) relax hard robust constraints by

allowing small, penalized violations, obtaining solutions that are nearly robust yet substantially less conservative. A comprehensive treatment of these ideas appears in Ben-Tal et al. (2009). Another model to reduce the conservatism of the classical robust model is robust regret or min-max regret (Savage 1951, Kouvelis and Yu 2013, Aissi et al. 2009) where the difference to the optimal value in each of the uncertain scenarios is minimized. All previously mentioned models have in common that the considered performance criteria can only be guaranteed on a (usually bounded) uncertainty set. Globalized robust methods attempt to control the performance of the calculated solution also outside of the uncertainty set Long et al. (2023), Ben-Tal et al. (2017). Our framework combines and complements the latter lines of work by measuring conservatism through a globalized adversarial regret operator and by providing a guarantee that degrades smoothly with the prediction error level.

A parallel development in computer science is the growing literature on *algorithms with predictions* (Mitzenmacher and Vassilvitskii 2022). These algorithms receive a (possibly erroneous) prediction and seek competitive ratios that interpolate between the optimal ratio when the prediction is exact and a worst-case ratio when it is adversarial. Foundational contributions include Lykouris and Vassilvitskii (2021) for caching and Purohit et al. (2018) for ski rental and scheduling. Our framework shares the same high-level aspiration: exploit accurate predictions while remaining robust to inaccurate ones. The key distinction is that algorithms with predictions operate in an online or combinatorial setting with discrete competitive ratios, whereas our framework targets continuous optimization under uncertainty with a smooth, tunable performance guarantee.

The bicriteria perspective on robustness (Chassein and Goerigk 2016) is also related: it treats nominal and worst-case cost as two competing objectives and traces the Pareto frontier between them. Our framework imposes a richer structure by controlling adversarial regret uniformly across all uncertainty levels  $\gamma \in \Gamma$ , rather than balancing only two extreme scenarios. A detailed comparison with robust, regret, and satisficing formulations is deferred to Section 2.

### 1.3. Contributions

The main contributions of this paper are as follows.

- We show that standard robust formulations yield vacuous guarantees under wild predictions, that classical regret formulations can produce decisions more optimistic than even a nominal formulation, and that satisficing formulations are under typical convexity assumptions merely robust formulations in disguise.
- We introduce *adversarial regret* — the gap between a decision’s worst-case cost and the oracle robust cost — as a novel performance target, and propose Globalized Adversarial Regret Optimization (GARO), which controls this adversarial regret uniformly over all prediction error levels.

- GARO can deliver either an *absolute* (23) or a *relative* (24) performance guarantee against the oracle robust benchmark. With a relative rate function, GARO generalizes the classical adaptation method of Lepskii (1993) from statistical estimation to downstream decision problems (Theorem 2).
- We derive exact tractable reformulations for affine and linear-polyhedral settings, and provide a discretization scheme and a constraint generation algorithm with convergence guarantees for general problems.

## 2. Uncertain Optimization

Decision-making in the face of uncertainty has been an active topic of research over the last fifty years, but the resulting tools were, by and large, developed with classical predictions in mind. We discuss here the benefits and downsides of classical uncertain optimization formulations in the face of wild predictions.

### 2.1. Robust Optimization

It has long been understood that blindly trusting the prediction  $p_0$  as a substitute for  $p^*$  in the nominal formulation (2) does result in rather gullible decisions. Ben-Tal and Nemirovski (2000) were perhaps the first to point out that even if  $d(p_0, p^*)$  is very small, the actual cost  $f(x_{nom}, p^*)$  can far exceed the anticipated cost  $f(x_{nom}, p_0)$ ; a phenomenon known as the “curse of optimization”. For a given uncertainty set  $P_{\gamma_0}$  robust optimization

$$\min_{x \in X} \max_{p \in P_{\gamma_0}} f(x, p) \tag{8}$$

has since been widely adopted to avoid this curse through a worst-case perspective. A notable property that explains the popularity of the worst-case perspective is that it preserves the convexity of the nominal problem. That is, if the objective function  $f$  is convex-concave and the constraint and parameter sets  $X$  and  $P_{\gamma_0}$  are convex, then the robust formulation (8) reduces to a convex-concave saddle point problem which, under rather mild technical assumptions, admits tractable algorithms (Ben-Tal and Nemirovski 2002).

The robust set  $P_{\gamma_0}$  characterizes all scenarios against which the decision maker wishes to anticipate. In the context of classical predictions, a natural choice is simply to consider a set

$$P_{\gamma_0} = \{p \in P : d(p_0, p) \leq \gamma_0\} \tag{9}$$

where the probability of the event  $p^* \in P_{\gamma_0}$  occurring is controlled explicitly by the classical guarantee

in Equation (3) through the selection of an appropriate robustness parameter  $\gamma_0$ . Robust optimization relies on the implication

$$p^* \in P_{\gamma_0} \implies f(x, p^*) \leq v_{wc}(x, \gamma_0) := \max_{p \in P_{\gamma_0}} f(x, p)$$

and the observation that in the face of the curse of optimization it typically holds that

$$v_{wc}(x_{rob}(\gamma_0), \gamma_0) := \min_{x \in X} v_{wc}(x, \gamma_0) \ll v_{wc}(x_{nom}, \gamma_0),$$

where  $X_{rob}(\gamma_0)$  denotes the set of optimal solutions of (8) with uncertainty set  $P_{\gamma_0}$  and  $x_{rob}(\gamma_0) \in X_{rob}(\gamma_0)$  is a selection. A hurdle to the practical adoption of robust formulations outside settings where extreme caution is indeed warranted is their conservatism (Roos and den Hertog 2020). By anticipating the worst-case scenario, performance may be poor in case the worst-case fails to realize. Although its associated worst-case performance guarantee

$$p^* \in P_{\gamma_0} \implies f(x_{rob}(\gamma_0), p^*) \leq v_{wc}^*(\gamma_0) := \min_{x \in X} v_{wc}(x, \gamma_0) \quad (10)$$

is best possible by construction, its strength nevertheless hinges on how large the optimal worst-case cost  $v_{wc}^*(\gamma_0)$  precisely is. When the minimal worst-case cost is very large the performance guarantee (10) becomes vacuous. The following example makes this discussion concrete.

**Example 1.** We want to find a hub location  $\mu$  to serve customers distributed at locations following  $\mathbb{P}^*$  within a support set  $\Xi$  where  $\|\xi - \mu\|^2$  corresponds to the frustration experienced by a customer at location  $\xi$  when served from a hub at location  $\mu$ , i.e.,

$$\min_{\mu} \int \|\xi - \mu\|^2 d\mathbb{P}^*(\xi). \quad (11)$$

Here the customers (used perhaps to instant gratification) experience superlinear frustration as service time increases. For our facility problem to be well posed it is necessary that  $\mathbb{P}^* \in \mathcal{P} := \{\mathbb{P} : \int \|\xi\|^2 d\mathbb{P}(\xi) < \infty\}$ . We remark that the unique optimal location in (11) is simply the mean  $\mu^* = \mu(\mathbb{P}^*) := \int \xi d\mathbb{P}^*(\xi)$  which achieves cost equal to the variance  $v^* = v(\mathbb{P}^*) := \int (\xi - \mu(\mathbb{P}^*))^2 d\mathbb{P}^*(\xi)$ .

The distribution of future customers  $\mathbb{P}^*$  may however not be known precisely to the decision-maker. Instead, consider a robust facility location formulation

$$\min_{\mu} \sup_{\mathbb{P} \in \mathcal{P}_{\gamma_0}} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi). \quad (12)$$

with respect to the Wasserstein ball  $\mathcal{P}_{\gamma_0} := \{\mathbb{P} \in \mathcal{P} : W(\mathbb{P}, \mathbb{P}_0) \leq \gamma_0\}$  for a judiciously chosen robustness parameter  $\gamma_0$ . Denote by  $\mu_{rob}(\gamma_0)$  the minimizer in the robust facility location problem (12). Depending on the value of  $\gamma_0$ , two extreme cases can be identified. Clearly, if  $\gamma_0 = 0$  we recover the unique nominal location  $\mu_{rob}(\gamma_0) = \mu_{nom} = \int \xi d\mathbb{P}_0(\xi)$  whereas at the other extreme for  $\gamma_0 \geq 2 \|\Xi\|$ , the set  $\mathcal{P}_{\gamma_0}$  contains all distributions supported on  $\Xi$  and hence the robust optimal location moves to the unique center<sup>2</sup> of  $\Xi$ . In the latter case,

$$\mu_{rob}(\gamma_0) = \arg \min_{\mu} \max_{\xi \in \Xi} \|\xi - \mu\|^2 = \text{ctr}(\Xi). \quad (13)$$

In Figure 1 the path  $\{\mu_{rob}(\gamma) : \gamma \geq 0\}$  between the discussed extreme cases is depicted as a gray line.

The robust facility location formulation promises if  $\mathbb{P}^* \in \mathcal{P}_{\gamma_0}$  that

$$\int \|\xi - \mu_{rob}(\gamma_0)\|^2 d\mathbb{P}^*(\xi) \leq v_{wc}^*(\gamma_0) := \min_{\mu} v_{wc}(\mu, \gamma_0) = \sup_{\mathbb{P} \in \mathcal{P}_{\gamma_0}} \int \|\xi - \mu_{rob}(\gamma_0)\|^2 d\mathbb{P}(\xi). \quad (14)$$

The strength of a worst-case guarantee hinges on whether the minimum worst-case cost is reasonably small or at least bounded. Clearly, for large  $\gamma_0 \geq 2 \|\Xi\|$  we get  $v_{wc}^*(\gamma_0) = \|\Xi\|^2$  and the worst-case guarantee (14) is vacuous when  $\Xi$  is large. Unfortunately, from Lemma 3 in the Appendix it follows that also for small value of  $\gamma_0 > 0$  we have  $v_{wc}^*(\gamma_0) \geq \min(\frac{\gamma_0}{2}, \|\Xi\|) \|\Xi\|$  limiting the robust formulation in (12) to bounded sets  $\Xi$ . The precise problem instance and Julia code to reproduce Figures 1–2 are available at <https://gitlab.com/vanparys/garo-example>.

## 2.2. Regret Optimization

Savage (1951) introduced the absolute regret formulation

$$\min_{x \in X} \max_{p \in P_{\gamma_0}} R(x, p) \quad (15)$$

in response to the apparent conservatism of the worst-case perspective advocated by Wald (1945) where the regret is defined absolutely as  $R(x, p) := f(x, p) - \min_{x' \in X} f(x', p)$ . This approach was later studied for discrete problem structures in Kouvelis and Yu (2013). Intuitively, the regret formulation (15) prefers decisions whose performance in hindsight can only be improved by a minimal amount.

By construction a regret optimal decision enjoys the performance bound

$$p^* \in P_{\gamma_0} \implies f(x_{reg}(\gamma_0), p^*) \leq \min_{x' \in X} f(x', p^*) + R^*$$

<sup>2</sup>For any bounded set  $S$  we write  $\|S\| := \min_{s \in \text{conv}(S)} \max_{s' \in S} \|s' - s\|$  for its circumradius and  $\text{ctr}(S)$  for its corresponding unique circumcenter.

where  $x_{reg}(\gamma_0)$  and  $R^*$  denote a minimizer and minimum in Equation (15), respectively. As performance is measured here relative to an oracle decision with access to the unknown parameter  $p^*$ , regret optimal decisions come with performance guarantees beyond merely the worst-case. In some sense, whereas robust formulations emphasize worst-case performance, regret formulations care about performance in all scenarios equally. By insisting on a minimal regret uniformly over all scenarios, regret formulations are inherently optimistic as the regret of a decision  $x$  balances equally its worst-case performance  $\max_{p \in \mathcal{P}_{\gamma_0}} f(x, p)$  and its best-case performance  $\min_{p \in \mathcal{P}_{\gamma_0}} f(x, p)$ .

Indeed, the optimal regret decision may offer even less protection against a worst-case scenario than a simple nominal decision, i.e.,

$$v_{wc}(x_{nom}, \gamma) < v_{wc}(x_{reg}(\gamma_0), \gamma)$$

for any  $\gamma \geq 0$ . A regret optimal decision may indeed forgo potential cost reductions in the predicted and worst-case scenarios to chase performance improvements in best-case scenarios, which seems to run counter to prudent decision making. Furthermore, outside a few toy problems (Agarwal and Zhang 2022, Perakis and Roels 2008), notably the newsvendor problem, regret formulations become computationally intractable already when the nominal problem (2) is a simple linear optimization problem (Averbakh and Lebedev 2005). We illustrate the optimism of regret formulations in our facility location example.

**Example 2.** Denote the regret of hub location  $\mu$  for a distribution of customers  $\mathbb{P}$  as

$$R(\mu, \mathbb{P}) := \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) - \min_{\mu'} \int \|\xi - \mu'\|^2 d\mathbb{P}(\xi) = \|\int \xi d\mathbb{P}(\xi) - \mu\|^2$$

where the final equality follows from the bias variance decomposition. It follows that the regret optimal facility location  $\mu_{reg}(\gamma_0) = \arg \min_{\mu} \max_{\mathbb{P} \in \mathcal{P}_{\gamma_0}} R(\mu, \mathbb{P})$  is simply the (unique) center of the convex set  $M(\gamma_0) := \{\int \xi d\mathbb{P}(\xi) : \mathbb{P} \in \mathcal{P}_{\gamma_0}\}$  collecting the means of all distributions in  $\mathcal{P}_{\gamma_0}$ . In the two extreme regimes discussed earlier, the robust and regret optimal formulations return the same optimal hub locations. Trivially, if  $\gamma_0 = 0$  we have  $\mu_{reg}(\gamma_0) = \mu_{nom} := \int \xi d\mathbb{P}_0(\xi)$  whereas for a large enough radius  $\gamma_0 \geq 2\|\Xi\|$ , the set  $\mathcal{P}_{\gamma_0}$  contains all distributions supported on  $\Xi$  and hence

$$\mu_{reg}(\gamma_0) = \text{ctr}(\text{conv}(\Xi)) = \text{ctr}(\Xi)$$

for any bounded set  $\Xi$ . In Figure 1 the path  $\{\mu_{reg}(\gamma) : \gamma \geq 0\}$  between the discussed extreme cases is depicted with a blue line.

From Lemma 4 in the Appendix we have  $M(\gamma_0) \subseteq B[\mu_{nom}, \gamma_0]$  (where  $B[c, r]$  denotes the

closed ball with center  $c$  and radius  $r$ ) and hence the worst-case regret remains well defined even for unbounded  $\Xi$  and remains bounded by  $\gamma_0^2$  in stark contrast to the worst-case cost which is unbounded. However, the regret optimal Weber point can be severely optimistic. Kantorovich-Rubinstein duality ensures that if  $\Xi$  is unconstrained that  $M(\gamma_0) = B[\mu_{nom}, \gamma_0]$  and hence the regret optimal location  $\mu_{reg}(\gamma_0) = \mu_{nom}$  coincides with the nominal location offering no additional protection. In case  $\Xi$  is bounded the situation can become even more dire. We probe the sensitivity of a location  $\mu$  by considering the difference

$$\begin{aligned} A(\mu, \gamma) &:= v_{wc}(\mu, \gamma) - v_{wc}^*(\gamma) \\ &= \sup_{W(\mathbb{P}, \mathbb{P}_0) \leq \gamma} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) - \min_{\mu'} \sup_{W(\mathbb{P}', \mathbb{P}_0) \leq \gamma} \int \|\xi - \mu'\|^2 d\mathbb{P}'(\xi). \end{aligned}$$

That is, the amount by which our worst-case cost is larger when the demand distribution can deviate from the predicted distribution by an amount  $\gamma$  compared to an oracle robust solution. We denote  $A$  as the adversarial regret as it measures our loss of performance when the distributional parameter  $\mathbb{P}$  is the action of an adversarial player.

To make the discussion tangible consider the problem setup depicted in Figure 1. We see in Figure 2 that the nominal facility location  $\mu_{nom} = \mu^*(\mathbb{P}_0)$  achieves the oracle cost at  $\gamma = 0$  but quickly deteriorates as  $\gamma$  increases (essentially depicting the curse of optimization). The robust facility location  $\mu_{rob}(\gamma_0)$  achieves the oracle cost at  $\gamma_0 = 1$  but even for smaller values  $\gamma \geq 0.25$  it achieves the oracle cost as  $\mu_{rob}(\gamma_0) = \text{ctr}(\Xi)$  for any  $\gamma_0 \geq 0.25$ . Observe however that

$$A(\mu_{nom}, \gamma) < A(\mu_{reg}(\gamma_0), \gamma) \quad \forall \gamma \geq 0 \tag{16}$$

indicating that the regret optimal facility location manages to suffer both larger nominal as well as larger worst-case costs than the nominal facility location.

### 2.3. Satisficing

In both the robust and the regret formulation discussed earlier, the set  $P_{\gamma_0}$  characterizes those parameters against which the decision maker wishes to anticipate. When the decision-maker has access to classical predictions satisfying the guarantee (3) the choice suggested in Equation (9) is both natural and adequate. With only access to adaptive or wild predictions, the set  $P_{\gamma_0}$  must be taken as the set of all possible parameter values  $P$ , rendering the absolute performance guarantees that robust and regret formulations aspire to simply unattainable.

We now describe a decision formulation introduced by Long et al. (2023) that enjoys performance guarantees relative to the performance of the prediction itself. Consider indeed the satisficing formu-

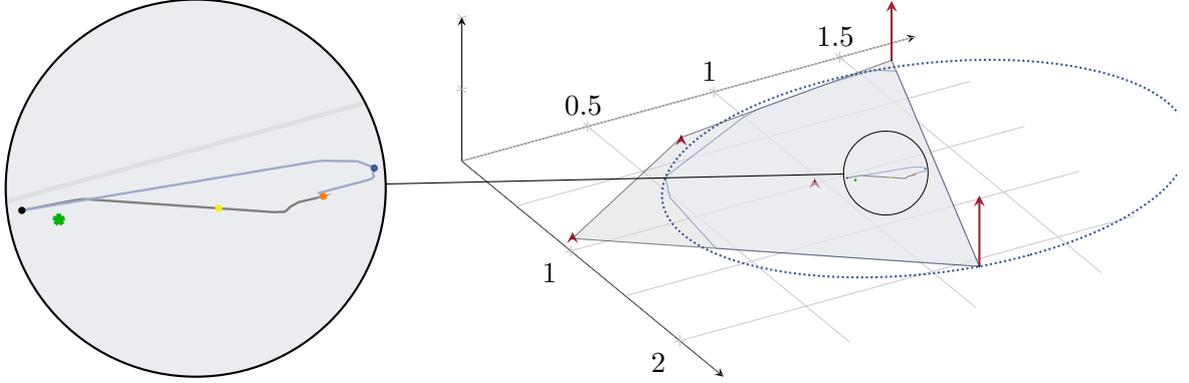


Figure 1: Facility location instance: customer locations  $\Xi$  (red marks) with predicted distribution  $\mathbb{P}_0$  (red arrows). The path of robust hub locations  $\{\mu_{rob}(\gamma)\}_{\gamma \geq 0}$  (gray) moves from the nominal location  $\mu_{nom}$  (orange) to  $\text{ctr}(\Xi)$  (black) as  $\gamma_0$  grows. The set  $M(\gamma_0)$  of attainable means under  $\mathcal{P}_{\gamma_0}$  (blue shaded polytope) and the regret-optimal location  $\mu_{reg} = \text{ctr}(M(\gamma_0))$  (blue dot) are also shown, together with the satisficing location  $\mu_{sat}$  (yellow). The GARO solution  $\mu_{garo}$  (green star) is discussed in Section 3.

lation

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & f(x, p) \leq f_0 + \alpha d(p_0, p) \quad \forall p \in P. \end{aligned} \quad (17)$$

Here  $f_0$  denotes a target cost level which the decision-maker deems *satisficing*. Let here now  $P_\infty = \{p \in P : d(p_0, p) < \infty\}$ . For the satisficing formulation (17) to be feasible and nontrivial we require that

$$v_{wc}^*(0) = \min_{x \in X} f(x, p_0) \leq f_0 \leq \min_{x \in X} \max_{p \in P_\infty} f(x, p) = v_{wc}^*(\max_{p \in P_\infty} d(p_0, p)) \quad (18)$$

as we assume here that  $p_0 \in P$ . We remark that for all  $p \notin P_\infty$  the satisficing constraint in (17) is void, since  $d(p_0, p) = \infty$  renders the right-hand side infinite. This distinction arises because  $d$  is extended-valued; when  $d$  takes only finite values,  $P_\infty = P$  and the remark is moot.

Let  $\alpha_{sat}$  be the optimal value,  $X_{sat}$  the set of optimal solutions, and  $x_{sat} \in X_{sat}$  a selection from Equation (17). Intuitively, we are looking for decisions whose performance  $f(x_{sat}, p^*)$  meets the target  $f_0$  if the prediction was indeed correct, i.e., in the event  $p_0 = p^*$ , and degrades at a minimal rate  $\alpha_{sat}$  relative to how wrong the prediction was in hindsight as measured by  $d(p_0, p^*)$ . Satisficing formulations deliver guarantees of the form

$$f(x_{sat}, p^*) \leq f_0 + \alpha_{sat} \gamma^* \quad (19)$$

which allow wild predictors to offer cost guarantees relative to the actual prediction error  $d(p_0, p^*) = \gamma^*$

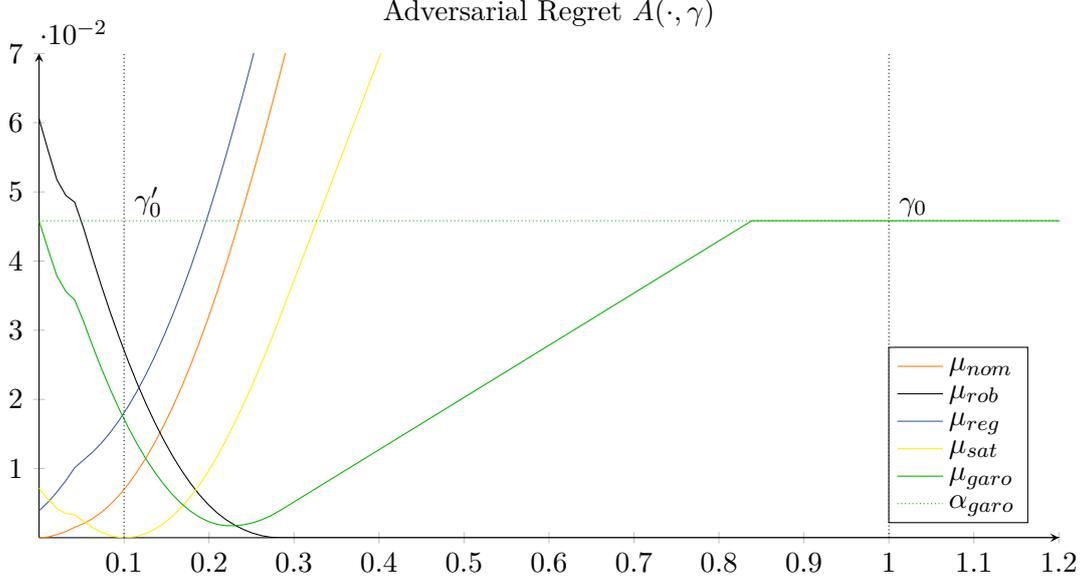


Figure 2: Adversarial regret  $A(\cdot, \gamma)$  as a function of the Wasserstein perturbation level  $\gamma$  for the five hub locations shown in Figure 1. The regret-optimal location  $\mu_{reg}$  (blue) is so optimistic that its adversarial regret exceeds that of the nominal location  $\mu_{nom}$  (orange) for all  $\gamma \geq 0$ , confirming (16). The robust location  $\mu_{rob}$  (black) achieves zero adversarial regret at  $\gamma_0$ . The GARO location  $\mu_{garo}$  (green) maintains uniformly small adversarial regret across all  $\gamma$ , bounded by the guarantee  $\alpha_{garo}$  (dotted green line).

rather than a worst-case radius  $\gamma_0$  as in robust formulations. Additionally, they work well with adaptive predictions which satisfy (7) and combined with (19) result in adaptive guarantee of the form

$$\text{Prob} [f(x_{sat}, p^*) \leq f_0 + \alpha_{sat} \gamma_N(\delta; \kappa)] \geq 1 - \delta.$$

Furthermore, if the objective function  $f$  is convex-concave and the constraint and parameter sets  $X$  and  $P_\infty$  are convex, satisficing formulations are under mild technical conditions amenable to tractable algorithms (Long et al. 2023).

**Example 3.** In the context of our facility location problem the satisficing formulation in Equation (17) reduces to

$$\begin{aligned} \min_{\mu, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) \leq f_0 + \alpha W(\mathbb{P}_0, \mathbb{P}) \quad \forall \mathbb{P} \in \mathcal{P} \end{aligned} \quad (20)$$

and is recognized as the satisficing formulation proposed by Long et al. (2023). Here, we choose

the satisfaction level  $v_{wc}^*(0) < f_0 := 1.04v_{wc}^*(0) < v_{wc}^*(\sup_{\mathbb{P} \in \mathcal{P}} W(\mathbb{P}_0, \mathbb{P}))$  to be 4% suboptimal in case  $\mathbb{P}^* = \mathbb{P}_0$  occurs. We mark the satisficing hub location  $\mu_{sat}$  with a yellow mark in Figure 1 and depict its adversarial regret in Figure 2. Proposition 2 implies that  $\mu_{sat} = \mu_{rob}(\gamma'_0)$  with  $\gamma'_0 = 0.1$ .

Unfortunately, a straightforward consequence of the lower bound in Lemma 3 in the Appendix is that we need  $\alpha \geq (\|\Xi\|^2 - f_0)/(2\|\Xi\|)$  for the satisficing formulation to be feasible. Hence, for  $\Xi$  unbounded, the constraint in the satisficing formulation (20) is not feasible in stark contrast to a regret formulation.

Hence, satisficing formulations in Equation (17) may become infeasible and guarantees in the form (19) can be prohibitively restrictive. Indeed, for the satisficing formulation (17) to be feasible it is necessary that the oracle cost is finite and satisfies  $v_{wc}^*(\gamma) = \mathcal{O}(\gamma)$ . Furthermore, the satisficing guarantee in Equation (19) is relative to an arbitrary satisfaction level  $f_0$ . In fact, we show in Proposition 2 in the Appendix that a satisficing solution can sometimes be found as a robust solution for a robustness radius  $\gamma'_0$  which is determined by  $f_0$ .

We remark that the assumption in Proposition 2 that the cost function  $f(x, p)$  is strictly concave in  $p$  can be relaxed, resulting in a slightly weaker conclusion that there always exists a satisficing solution admitting a robust interpretation; see Appendix B for details including the proof of Proposition 2 and supporting examples.

**Theorem 1.** *Let  $P_\infty$  be a compact convex set admitting a continuous strictly convex function  $\psi : P_\infty \rightarrow \mathbb{R}_+$ . Let  $f(x, p)$  be jointly continuous, convex in  $x$ , and concave in  $p$  for every  $x \in X$ , with  $X$  compact convex. Let  $p \mapsto d(p_0, p)$  be lower semicontinuous and convex. Assume (17) is feasible and satisfies condition (18). Then there exists an optimal solution  $x_{sat}$  of (17) such that*

$$x_{sat} \in \cup_{\gamma \in [0, \gamma_{\max}]} X_{rob}(\gamma), \quad \gamma_{\max} := \max_{p \in P_\infty} d(p_0, p).$$

The assumption that  $P_\infty$  admits a continuous strictly convex function is mild: in finite dimensions it is always satisfied, e.g., by  $\psi(p) = \|p\|^2$ . In infinite-dimensional settings, such as when  $P_\infty$  is a set of probability distributions, the condition requires separate verification.

In summary, each of the three standard formulations exhibits a fundamental weakness when the prediction error level  $\gamma$  is unknown. Robust formulations yield vacuous guarantees whenever the uncertainty set must be taken large. Regret formulations are inherently optimistic and can produce decisions offering less protection than the nominal solution. Satisficing formulations, while offering prediction-relative guarantees, are under mild convexity assumptions nothing but robust formulations in disguise.

### 3. Robust Decisions with Wild Predictions

When we have access to wild predictions, standard performance guarantees can generally not be achieved. Indeed, as wild predictions can be arbitrarily bad or even have malicious intent, a robust formulation ignoring the predictions altogether will provide only vacuous worst-case guarantees. In what follows we introduce a decision framework that addresses the shortcomings of the classical robust, regret, and satisficing formulations identified in Section 2. The main idea is to control adversarial regret uniformly over all prediction error levels.

To make this rigorous, we introduce the univariate family of non-decreasing sets  $P_\gamma$ , e.g.,  $P_\gamma = \{p \in P : d(p_0, p) \leq \gamma\}$ , for  $\gamma \in \Gamma$ . We do assume  $X$  and  $P_\gamma$  to be compact for all  $\gamma \in \Gamma$ . We note that  $p$  may represent either a finite-dimensional parameter vector, recovering standard robust optimization (Ben-Tal and Nemirovski 2002), or a probability distribution with a suitable distance  $d$  such as the Wasserstein metric, recovering distributionally robust optimization (Delage and Ye 2010, Mohajerin Esfahani and Kuhn 2015); the running facility location example falls in the latter setting. Throughout, we write min and max when optimizing over  $X$  or  $P_\gamma$ , where attainment is guaranteed by compactness of these sets and continuity of  $f$ . We write inf and sup when optimizing over infinite-dimensional spaces such as sets of probability distributions, where attainment requires a separate argument. Inspired by the discussion in Section 2 we introduce the adversarial regret

$$A(x, \gamma) := \max_{p \in P_\gamma} f(x, p) - \min_{x' \in X} \max_{p' \in P_\gamma} f(x', p').$$

That is, the additional cost we may incur if the parameter  $p$  is chosen adversarially in  $P_\gamma$  compared to an oracle robust solution with access to  $\gamma$ . We remark that classical regret is larger than the adversarial regret, i.e.,  $A(x, \gamma) \leq \max_{p \in P_\gamma} R(x, p)$  as the former also takes into account how much performance is left on the table when  $p$  would be chosen beneficially. It was precisely this aspect which rendered regret optimization potentially optimistic as discussed in Section 2.

As mentioned earlier, the defining characteristic of decision-making with wild predictions is that parameter  $\gamma$  is fundamentally unknown and hence at a minimum we would like to have that our decision is weakly minimal in

$$\min_{x \in X} (A(x, \gamma))_{\gamma \in \Gamma}. \tag{21}$$

That is, a decision  $x^*$  for which there is no other decision  $y \in X$  so that  $A(y, \gamma) < A(x^*, \gamma)$  for all  $\gamma \in \Gamma$ . As its name suggests, though, this is a rather weak requirement. We observe indeed that the nominal and robust solution as discussed in Section 2 are weakly minimal as they minimize  $x \mapsto A(x, 0)$  and  $x \mapsto A(x, \gamma_0)$ , respectively. We point out that as the facility location example in Figures 1 and 2 illustrates, regret optimal solutions nevertheless can fail to be weakly optimal in problem (21). Finally, under Proposition 2's strict concavity assumption satisficing solutions are weakly minimal as

they minimize  $x \mapsto A(x, \gamma'_0)$  for some  $\gamma'_0 \geq 0$ ; more generally, a satisficing solution is weakly minimal whenever  $\Gamma$  is compact and  $\gamma \mapsto v_{wc}(x, \gamma)$  is continuous. See Proposition 3 in Appendix B for the formal statement and proof.

Rather than seeking a decision that merely minimizes  $A(x, \gamma)$  at a single fixed  $\gamma$  — as robust and satisficing formulations effectively do — we propose to bound the adversarial regret uniformly over all  $\gamma \in \Gamma$ , leading to the following class of decision formulations:

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & A(x, \gamma) = \max_{p \in P_\gamma} f(x, p) - \min_{x' \in X} \max_{p' \in P_\gamma} f(x', p') \leq \alpha \phi(\gamma) \quad \forall \gamma \in \Gamma. \end{aligned} \tag{GARO}$$

Here a nondecreasing rate function  $\phi : \Gamma \rightarrow \mathbb{R}_+$  controls the amount of adversarial regret our decision suffers as a function of the prediction error  $d(p_0, p^*)$ . We now list the main benefits of our formulation explicitly.

**Weak Optimality** Unlike a regret formulation which can yield decisions which are even more optimistic than a nominal formulation, the proposed formulation is guaranteed by construction to produce weakly minimal solutions in (21).

**Performance Guarantees** Unlike nominal or robust formulations, GARO simultaneously guarantees performance across all error levels  $\gamma \in \Gamma$ . That is, any minimizer  $(x_{garo}, \alpha_{garo})$  of GARO satisfies

$$v_{wc}(x_{garo}, \gamma) \leq v_{wc}^*(\gamma) + \alpha_{garo} \phi(\gamma) \quad \forall \gamma \in \Gamma. \tag{22}$$

With the constant rate  $\phi(\gamma) = 1$  this becomes the *absolute* guarantee

$$v_{wc}(x_{garo}, \gamma) - v_{wc}^*(\gamma) \leq \alpha_{garo} \quad \forall \gamma \in \Gamma, \tag{23}$$

bounding the gap to the oracle in absolute terms. With  $\phi(\gamma) = v_{wc}^*(\gamma)$  it becomes the *relative* guarantee

$$\frac{v_{wc}(x_{garo}, \gamma) - v_{wc}^*(\gamma)}{v_{wc}^*(\gamma)} \leq \alpha_{garo} \quad \forall \gamma \in \Gamma, \tag{24}$$

bounding the gap as a fraction of the oracle cost. In both cases  $x_{garo}$  tracks the oracle simultaneously at the typical error level  $\gamma_{\min}$  and at the worst-case level  $\gamma_{\max}$ , without committing to any particular level upfront.

**Choosing the Rate Function** The constant rate  $\phi(\gamma) = 1$  is the natural choice when the oracle cost may be zero or lacks a natural scale; the oracle-cost rate  $\phi(\gamma) = v_{wc}^*(\gamma)$  is preferable when the oracle

cost grows with  $\gamma$  and an absolute bound would become vacuously large. These are the two principal choices a practitioner faces: in either case a single instance of GARO is solved and  $\alpha_{garo}$  directly quantifies how closely the decision tracks the oracle benchmark.

We emphasize that GARO makes no a priori claim that  $\alpha_{garo}$  will be small. Its guarantee is one of “mere” optimality; if there is a decision with good absolute or relative performance guarantees then GARO will find it. A large  $\alpha_{garo}$  signals that the problem simply does not admit a decision that uniformly tracks the oracle across  $\Gamma$ . A concrete example in which  $\alpha_{garo}$  is provably small is the adaptive prediction setting of Section 4.2 where the relative rate function yields  $\alpha_{garo} \leq 1$ , recovering Lepskii’s classical factor-of-two guarantee. However, when a standard choice fails to yield a small  $\alpha_{garo}$ , a problem-dependent rate can salvage the situation as we illustrate in the running example below.

**Example 4.** With  $\phi(\gamma) = 1$  the GARO solution  $\mu_{garo}$  is depicted as a green star in Figure 1. It coincides neither with any robust location on the path  $\{\mu_{rob}(\gamma) : \gamma \in \Gamma\}$  (gray) nor with any regret-optimal location on the path  $\{\mu_{reg}(\gamma) : \gamma \in \Gamma\}$  (blue), confirming that GARO produces a genuinely distinct decision. Figure 2 shows that  $\mu_{garo}$  achieves a strictly smaller adversarial regret  $A(\mu_{garo}, \gamma)$  than every other proposed hub location across the entire range of  $\gamma$ , bounded uniformly by the guarantee  $\alpha_{garo} \approx 4.58 \times 10^{-2}$ .

For the bounded instance above in Figure 1, the absolute rate  $\phi(\gamma) = 1$  yields a reasonably small  $\alpha_{garo}$ , but as the support  $\Xi$  expands the absolute guarantee deteriorates. From Lemma 3 the oracle cost satisfies  $v_{wc}^*(\gamma) \geq \min(\gamma/2, \|\Xi\|) \|\Xi\|$ , so the oracle itself grows with the circumradius  $\|\Xi\|$ ; maintaining a uniformly small absolute gap  $A(\mu, \gamma) \leq \alpha_{garo}$  over all  $\gamma \in \Gamma$  therefore forces  $\alpha_{garo}$  to grow with  $\|\Xi\|$ .

In the limit  $\Xi = \mathbb{R}_+$  with  $\mathbb{P}_0 = \delta_0$  the absolute rate becomes outright infeasible: direct computation gives  $A(\mu, \gamma) = (\mu - \gamma)^2$ , which diverges as  $\gamma \rightarrow \infty$  for any fixed  $\mu$ . Similarly, as we pointed out in Lemma 3, we have  $v_{wc}^*(\gamma) = \infty$  for any  $\gamma > 0$  and hence considering a relative rate function is not a sensible option either. Note that the two failure modes are of opposite character: absolute infeasibility ( $\alpha_{garo} \rightarrow \infty$ ) signals that no decision uniformly tracks the oracle, while relative triviality ( $\alpha_{garo} \rightarrow 0$  for every  $\mu$ ) means all decisions do so vacuously. We can salvage the situation by considering instead  $\phi(\gamma) = (1 + \gamma)^q$  with  $q \geq 2$ : the resulting guarantee  $A(\mu, \gamma) \leq \alpha_{garo}(1 + \gamma)^q$  is meaningful for practically relevant prediction errors (small  $\gamma$ , where  $(1 + \gamma)^q \approx 1$ ) and degrades gracefully as predictions become increasingly wild, a reasonable price to pay when  $\Xi$  is unbounded. For  $q = 2$  the solution is  $\alpha_{garo} = 1$  and  $\mu_{garo} = [0, 1]$ , while for  $q > 2$  the unique minimizer satisfies  $(\mu_{garo})^2(1 + \mu_{garo})^{q-2} = 4(q - 2)^{q-2}/q^q$ .

The choice of rate function  $\phi$  encodes a design judgment about how much adversarial regret is

acceptable as the prediction error grows. Together with the distance function  $d$  which it shares with the uncertain optimization formulations in Section 2, the rate function shapes the nature of the GARO guarantee (22) and its appropriate specification should reflect the structure of the application at hand.

## 4. Performance Guarantees Across Prediction Regimes

The three prediction regimes introduced in Section 1.1 (classical, adaptive, and wild) call for different analytical tools, but GARO delivers meaningful guarantees in each. The following subsections establish these guarantees in turn, completing the picture begun in Section 3.

### 4.1. Classical Predictions

Classical predictions satisfy the guarantee (3). In case the event  $d(p_0, p^*) \leq \gamma_N(\delta)$  occurs it is straightforward to conclude that  $f(x_{rob}(\gamma_N(\delta)), p^*) \leq \max_{d(p_0, p) \leq \gamma_N(\delta)} f(x_{rob}(\gamma_N(\delta)), p) = v_{wc}^*(\gamma_N(\delta))$ . It follows that

$$\text{Prob}[f(x_{rob}(\gamma_N(\delta)), p^*) \leq v_{wc}^*(\gamma_N(\delta))] \geq 1 - \delta. \quad (25)$$

Such out-of-sample guarantees have been established throughout the robust optimization literature (Kuhn et al. 2019, Van Parys et al. 2021) since the probability of observing a disappointment event in which the actual cost exceeds the anticipated worst-case cost can be controlled explicitly by a judiciously designed robustness parameter  $\gamma_0$ . More recently, several authors (Van Parys et al. 2021, Lam 2019) have pointed out that the guarantee (25) does not necessitate the predictor to satisfy Equation (3) but rather can be established more directly as well. In any event, a key limitation of the out-of-sample guarantee (25) is that it reduces the disappointment risk of its decision to a single probability level. However, reporting multiple probability levels is either encouraged or legally required in many safety-critical applications. For example, the US Nuclear Regulatory Commission (U.S. Nuclear Regulatory Commission n.d.) mandates three tiers of probability assessments whereas NASA (National Aeronautics and Space Administration 2011, Chapter 13) promotes a continuous approach through reporting an exceedance curve which in our context means reporting

$$\text{Prob}[f(x_{rob}(\gamma_N(\delta)), p^*) \leq v_{wc}(x_{rob}(\gamma_N(\delta)), \gamma)] \quad (26)$$

for several radii  $\gamma \geq 0$  simultaneously. However, beyond the fact that  $v_{wc}(x_{rob}(\gamma_N(\delta)), \gamma) \leq v_{wc}^*(\gamma_N(\delta))$  for  $\gamma \leq \gamma_N(\delta)$ , the worst-case cost  $v_{wc}(x_{rob}(\gamma_N(\delta)), \gamma)$  for  $\gamma > \gamma_N(\delta)$  is uncontrolled. In stark contrast, the GARO solution  $x_{garo}$  satisfies  $A(x_{garo}, \gamma) \leq \alpha_{garo}\phi(\gamma)$  for *every*  $\gamma \in \Gamma$  by construction, independent of any chosen confidence level  $\delta$ . This deterministic property immediately controls the full exceedance

curve: for every  $\gamma \in \Gamma$ ,

$$\text{Prob} [f(x_{garo}, p^*) \leq v_{wc}^*(\gamma) + \alpha_{garo}\phi(\gamma)] \geq \text{Prob} [d(p_0, p^*) \leq \gamma]. \quad (27)$$

For any fixed  $\delta$ , substituting  $\gamma = \gamma_N(\delta)$  and applying the classical guarantee (3) yields a single-level bound

$$\text{Prob} [f(x_{garo}, p^*) \leq v_{wc}^*(\gamma_N(\delta)) + \alpha_{garo}\phi(\gamma_N(\delta))] \geq 1 - \delta,$$

which at that particular  $\delta$  is admittedly weaker than (25) by the additional term  $\alpha_{garo}\phi(\gamma_N(\delta))$ . The fundamental advantage of  $x_{garo}$  is therefore not at any single confidence level but across all levels simultaneously: the robust solution  $x_{rob}(\gamma_N(\delta_0))$ , designed for a specific  $\delta_0$ , provides no control over  $v_{wc}(x_{rob}(\gamma_N(\delta_0)), \gamma)$  for  $\gamma > \gamma_N(\delta_0)$  and hence no guarantee at higher confidence  $\delta < \delta_0$ . The single decision  $x_{garo}$ , by contrast, satisfies (27) simultaneously for every  $\delta \in (0, 1)$ , exactly supplying the exceedance curve that safety-critical reporting standards require.

## 4.2. Adaptive Predictions

In Section 1.1 we pointed out that adaptive predictions can be derived from classical predictions (Jain et al. 2022), and discussed two example settings in which  $\kappa$  represented an unknown moment in a heavy-tailed data setting and an unknown corruption level in a corrupt data setting. Consider the setting in which we have access to a family of classical estimators  $p_0(\kappa)$  parametrised in  $\kappa$ , all of which satisfy the pointwise guarantee in Equation (5). If we knew  $\kappa$  then we could use  $p_0(\kappa)$  as a classical prediction and guarantee

$$\text{Prob} \left[ f(x_{rob, \kappa}(\gamma_N(\delta, \kappa)), p^*) \leq v_{wc, \kappa}^*(\gamma_N(\delta, \kappa)) \right] \geq 1 - \delta \quad (28)$$

where  $x_{rob, \kappa}(\gamma)$  and  $v_{wc, \kappa}^*(\gamma)$  are the minimizers and minimum of  $\min_{x \in X} \max_{p \in P_{\gamma, \kappa}} f(x, p)$ , with  $P_{\gamma, \kappa} = \{p \in P : d(p_0(\kappa), p) \leq \gamma\}$ .

However,  $\kappa$  is unknown in practice. Neither robust nor regret formulations can directly exploit adaptive predictions satisfying guarantee (7): any decision calibrated to a single  $\kappa$  forces the practitioner to commit upfront, either choosing the pessimistic  $\kappa_{\max}$  (yielding an overcautious decision) or guessing  $\kappa$  (risking an underprotected one). The GARO formulation sidesteps this commitment entirely, generalizing and refining the adaptation strategy of Lepskii (1993), and in fact ensures

$$\text{Prob}_{\kappa} [f(x_{garo}, p^*) \leq v_{wc}^*(\gamma_N(\delta, \kappa)) + \alpha_{garo}\phi(\gamma_N(\delta, \kappa))] \geq 1 - \delta.$$

Rather than guaranteeing absolute performance, the previous inequality establishes that the performance of our decision is adaptive to the hardness of the prediction problem as captured by the

auxiliary parameter  $\kappa \in K \subseteq [\kappa_{\min}, \kappa_{\max}]$ . Crucially,  $x_{garo}$  does not depend on  $\kappa$ : a single decision simultaneously adapts to every level of problem hardness, just as in the classical prediction setting it supplies the full exceedance curve without committing to any particular confidence level  $\delta$ .

Let  $\beta > 1$  be an approximation constant and consider a discretized subset  $K_J = \{\kappa_0, \dots, \kappa_J\}$  of  $K$ . We have for any estimator  $p_0(\kappa_j)$  with  $\kappa_j \geq \kappa$  the pointwise guarantee

$$\text{Prob}_\kappa [d(p_0(\kappa_j), p^*) \leq \gamma_N(\delta, \kappa_j)] \geq 1 - \delta.$$

Given the above, we introduce the associated parameter set  $\Gamma_J = \{\gamma_j := \gamma_N(\delta/(J+1), \kappa_j)\}_{j=0}^J$  and define a nondecreasing family of sets

$$P_\gamma = \{p \in P : d(p_0(\kappa_j), p) \leq \gamma_j \quad \forall j \in [0, \dots, J] \text{ s.t. } \gamma_j \geq \gamma\} \quad (29)$$

for any  $\gamma \geq 0$ . Consider the event  $E_\kappa = \{p^* \in P_{\gamma_N(\delta/(J+1), \kappa^*)}\}$  where  $\kappa^* = \min\{\kappa' \in K_J : \kappa' \geq \kappa\}$ . Informally, if  $E_\kappa$  occurs then all of the applicable estimates  $p_0(\kappa_j)$  for  $\kappa_j \geq \kappa$  realize within their error margin bound  $d(p_0(\kappa_j), p^*) \leq \gamma_j$ . A simple union bound argument from the pointwise guarantee (5) implies now that we have

$$\text{Prob}_\kappa[E_\kappa] \geq 1 - |\{\kappa' \in K_J : \kappa' \geq \kappa\}|\delta/(J+1) \geq 1 - \delta. \quad (30)$$

**Theorem 2.** *The globalized adversarial regret formulation (GARO) with  $\phi(\gamma) = \min_{x \in X} \max_{p \in P_\gamma} f(x, p)$  and  $P_\gamma$  defined as in Equation (29) reduces to*

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \alpha \\ \text{s.t. } \max_{p \in P_{\gamma_j}} f(x, p) \leq (1 + \alpha) \min_{x' \in X} \max_{p' \in P_{\gamma_j}} f(x', p') \quad \forall j \in [0, \dots, J]. \end{aligned} \quad (31)$$

*Its solution  $(x_{garo}, \alpha_{garo})$  does not depend on  $\kappa \in K$  and satisfies on  $E_\kappa$  the guarantee*

$$f(x_{garo}, p^*) \leq (1 + \alpha_{garo}) v_{w_c, \kappa^*}^*(\gamma_N(\delta/(J+1), \kappa^*))$$

*where  $\kappa^* = \min\{\kappa' \in K_J : \kappa' \geq \kappa\}$ .*

*Proof.* The reduction from GARO to (31) follows from the observation that  $\phi$  is an increasing function and the definition of  $P_\gamma$  in Equation (29) implies that  $P_{\gamma_j} = P_\gamma$  for any  $\gamma \in [\gamma_j, \gamma_{j+1})$ .

By the choice of rate function  $\phi$  the solution  $(x_{garo}, \alpha_{garo})$  satisfies here a relative guarantee (24) and hence

$$\max_{p \in P_\gamma} f(x_{garo}, p) \leq (1 + \alpha_{garo}) \min_{x' \in X} \max_{p' \in P_\gamma} f(x', p') \quad \forall \gamma \geq 0.$$

We consider  $\gamma^* = \gamma_N(\delta/(J+1), \kappa^*)$  and observe that  $E_\kappa = \{p^* \in \mathcal{P}_{\gamma_N(\delta/(J+1), \kappa^*)}\}$  and so

$$f(x_{garo}, p^*) \leq (1 + \alpha_{garo}) \min_{x' \in X} \max_{p' \in P_{\gamma^*}} f(x', p').$$

Finally, observe that  $P_{\gamma^*} \subseteq P_{\gamma_N(\delta/(J+1), \kappa^*), \kappa^*}$  and hence  $f(x_{garo}, p^*) \leq (1 + \alpha_{garo}) v_{wc, \kappa^*}^*(\gamma_N(\delta/(J+1), \kappa^*))$ .

□

From Equation (30) and Theorem 2 it follows that

$$\text{Prob}_\kappa \left[ f(x_{garo}, p^*) \leq (1 + \alpha_{garo}) v_{wc, \kappa^*}^*(\gamma_N(\delta/(J+1), \kappa^*)) \right] \geq 1 - \delta$$

with  $\kappa^* = \min \{\kappa' \in K_J : \kappa' \geq \kappa\}$ , which should be compared to Equation (28). Note that (28) is an oracle bound, achievable only if  $\kappa$  is known. Relative to this oracle, GARO's bound is admittedly weaker in two respects: it carries the multiplicative overhead  $(1 + \alpha_{garo})$  and uses the reduced confidence  $\delta/(J+1)$  in place of  $\delta$ . Since  $\gamma_N(\delta, \kappa)$  is typically insensitive to  $\delta$ , the latter difference is minor and can be further controlled by an appropriate discretization of  $K$ . The fundamental advantage of  $x_{garo}$  is again not at any single  $\kappa$  but across all  $\kappa \in K$  simultaneously: the single  $x_{garo}$  satisfies the above bound for every  $\kappa$  without any knowledge of it.

We emphasize that GARO is not merely feasible but optimal among adaptive decisions: any decision  $\hat{x}$  satisfying  $f(\hat{x}, p^*) \leq (1 + \hat{\alpha}) v_{wc, \kappa^*}^*(\gamma_N(\delta/(J+1), \kappa^*))$  on  $E_\kappa$  must have  $\hat{\alpha} \geq \alpha_{garo}$ . That is,  $x_{garo}$  achieves the smallest possible multiplicative excess over the oracle cost, for each  $\kappa$ .

**Example 5.** To recover the adaptation strategy of Lepskii (1993) we consider a continuous cost function  $f(x, p) = d(x, p)$  satisfying a triangular inequality and again consider access to a family of classical predictors  $p_0(\kappa)$  for  $\kappa \in K$  satisfying the pointwise guarantee

$$\text{Prob}_\kappa [d(p_0(\kappa), p^*) \leq \gamma_N(\delta; \kappa)] \geq 1 - \delta.$$

The cost function reflects here that in classical estimation settings in the absence of any particular downstream decision tasks the norm of the error is the most classical measure of performance. Consider here the increasing family of sets  $P_{\gamma, \kappa} = \{p : d(p_0(\kappa), p) \leq \gamma\}$  with associated oracle cost  $v_{wc, \kappa}^*(\gamma) = \min_x \max_{p \in P_\gamma} f(x, p) = \gamma$  and define the increasing family of sets  $P_\gamma$  as in Equation (29). The associated globalized adversarial regret formulation (GARO) returns  $(x_{garo}, \alpha_{garo})$  which following Theorem 2 satisfies on  $E_\kappa$  the guarantee

$$d(x_{garo}, p^*) \leq (1 + \alpha_{garo}) \gamma_N(\delta/(J+1), \kappa^*).$$

**Lemma 1.** *We have  $E_\kappa \implies \alpha_{garo} \leq 1$ .*

*Proof.* Define here  $p_\gamma \in \arg \min_x \max_{p \in P_\gamma} d(x, p)$  and  $\|P_\gamma\| = \min_x \max_{p \in P_\gamma} d(x, p)$ . Observe that from the triangle inequality it follows that  $v_{wc}(x, \gamma) = \max_{p \in P_\gamma} d(x, p) \leq \max_{p \in P_\gamma} d(x, p_\gamma) + d(p_\gamma, p) = d(x, p_\gamma) + \|P_\gamma\|$  and  $v_{wc}^*(\gamma) = \|P_\gamma\|$ . Hence, we have that

$$\alpha_{garo} \leq \min_{x, \alpha \geq 0} \{ \alpha : d(x, p_\gamma) + \|P_\gamma\| \leq (1 + \alpha) \|P_\gamma\| \quad \forall \gamma \in \Gamma_J \}.$$

However, consider now  $\bar{x} \in P_{\gamma^*}$  and  $\bar{\alpha} = 1$  with  $\gamma^*$  the smallest  $\gamma \in \Gamma_J$  so that  $P_\gamma \neq \emptyset$ . We have

$$d(\bar{x}, p_\gamma) \leq \|P_\gamma\|$$

as  $\bar{x} \in P_{\gamma^*} \subseteq P_\gamma$  for any  $\gamma \geq \gamma^*$  from which the claim follows immediately.  $\square$

We have by Lemma 1 that in this adaptive prediction setting  $\alpha_{garo}$  is bounded above by 1 on  $E_\kappa$ . From Equation (30) it hence follows that

$$\text{Prob}_\kappa [d(x_{garo}, p^*) \leq 2\gamma_N(\delta/(J+1), \kappa^*)] \geq 1 - \delta.$$

For the particular geometric discretization  $K = \{\kappa_j = \kappa_{\min} \beta^j\}_{j=0}^J$  with  $J = \lceil \log_\beta(\kappa_{\max}/\kappa_{\min}) \rceil$  and  $\beta > 1$  we have  $\kappa^* \leq \beta \kappa$  and recover Equation (6).

### 4.3. Wild Predictions

In the wild prediction setting of Section 1.1, the prediction error  $\gamma^* = d(p_0, p^*)$  satisfies  $\gamma^* \leq \gamma_{wc}$  in the worst case yet is typically much smaller ( $\gamma^* \leq \gamma_{tp}$ ), with no probabilistic relationship between the two levels. Setting  $\Gamma = [\gamma_{tp}, \gamma_{wc}]$ , the performance guarantees (23) and (24) of Section 3 apply in their simplest, fully deterministic form: whatever the unknown  $\gamma^* \in [\gamma_{tp}, \gamma_{wc}]$  turns out to be,

$$f(x_{garo}, p^*) \leq v_{wc}^*(\gamma^*) + \alpha_{garo} \phi(\gamma^*),$$

without invoking any probability model on the prediction. This is the primary regime for which GARO was designed. Unlike the classical and adaptive settings, where statistical concentration inequalities are needed to give the deterministic guarantee probabilistic content, here the bound holds with certainty for any realisation of  $\gamma^* \in \Gamma$ . The classical and adaptive guarantees of the preceding subsections should therefore be seen as a bonus: GARO rewards the practitioner with probabilistic exceedance curves or adaptive oracle-tracking when a statistical model is available, but requires neither. Furthermore, by construction  $x_{garo}$  is weakly minimal in  $(A(x, \gamma))_{\gamma \in \Gamma}$ : no other single decision simultaneously reduces adversarial regret across the entire interval  $\Gamma = [\gamma_{tp}, \gamma_{wc}]$ , making it the natural choice precisely when

predictions cannot be trusted to carry statistical structure.

## 5. Solution Methods

The adversarial regret formulation (GARO) is computationally more appealing than a classical regret formulation which is already NP hard even for linear optimization problems (Averbakh and Lebedev 2005). Nevertheless, the adversarial regret formulation is more challenging to solve than its robust counterpart formulation stated in Equation (8). In what follows we will take the standard robust problem in Equation (8) as a computational primitive which we assume can be solved efficiently; see Ben-Tal and Nemirovski (2002). We point out that an exact tractable formulation of GARO is possible when  $v_{wc}$  is affine in  $\gamma$ : for concave rate functions  $\phi$  this reduces to just two constraints (Section 5.1), while for general  $\phi$  it yields a finite number of convex constraints via a conjugate representation (Section 5.2). Furthermore, in case  $\Gamma$  has finite cardinality (as was the case when working with Lepskii’s method for adaptive predictions in Section 4.2) an exact tractable formulation is guaranteed. Finally, we give two approximation algorithms to handle more general settings.

### 5.1. Nominal-Robust Formulations

In this setting we assume that the worst-case cost function is affine in the robustness parameter.

**Assumption 1.** *Let  $\gamma \mapsto v_{wc}(x, \gamma) = \max_{p \in P_\gamma} f(x, p)$  be affine for all  $x \in X$ .*

The following proposition shows that under the simplifying Assumption 1, only the interval extremes  $\{\min \Gamma, \max \Gamma\}$  of the set  $\Gamma$  are of interest when the rate function  $\phi$  is concave. We remark that this includes  $\phi(\gamma) = v_{wc}^*(\gamma) = \min_{x \in X} v_{wc}(x, \gamma)$  due to Assumption 1.

**Proposition 1.** *Let Assumption 1 hold and  $\phi$  be concave. Then, Problem GARO is equivalent to*

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & \max_{p \in P_{\min \Gamma}} f(x, p) - v_{wc}^*(\min \Gamma) \leq \alpha \phi(\min \Gamma), \\ & \max_{p \in P_{\max \Gamma}} f(x, p) - v_{wc}^*(\max \Gamma) \leq \alpha \phi(\max \Gamma). \end{aligned} \tag{32}$$

*Proof.* We can reformulate Problem GARO as

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & A(x, \gamma) = \max_{p \in P_\gamma} f(x, p) - v_{wc}^*(\gamma) \leq \alpha \phi(\gamma) \quad \forall \gamma \in \Gamma. \end{aligned} \tag{33}$$

The constraint function  $\max_{p \in P_\gamma} f(x, p) - v_{wc}^*(\gamma) - \alpha \phi(\gamma)$  is here convex in  $\gamma$ . It follows that the function attains its maximum on the interval extremities  $\min \Gamma$  and  $\max \Gamma$ .  $\square$

Formulation (32) is an uncertain optimization formulation with objective balancing nominal and robust performance. This balance has been extensively studied in the robust optimization literature; see Chassein and Goerigk (2016). Furthermore, the desire to balance performance when the prediction is exact ( $p^* = p_0$  and  $\min \Gamma = 0$ ) while retaining control over the performance when the prediction fails ( $p^* \in P_{\max \Gamma}$ ) has been the topic of the surging literature on algorithms with predictions (Mitzenmacher and Vassilvitskii 2022).

**Example 6** (Uncertain Linear Regression). Consider an uncertain linear regression problem  $\min_{x \in X} \|Ax - b\|$  where the matrix  $A$  is uncertain. We assume that we have a prediction  $A_0$  for the matrix  $A$  where we expect the operator norm  $\|A_0 - A\| = \max_{u \neq 0} \|(A - A_0)u\|/\|u\|$  to remain small. A well known result from the robust optimization literature is that Assumption 1 holds as we have  $v_{wc}(x, \gamma) := \max_{\|A - A_0\| \leq \gamma} \|Ax - b\| = \|A_0x - b\| + \gamma \|x\|$  revealing the classical connection between robustness and regularization. The GARO formulation of this problem with  $\Gamma = [0, \gamma_{\max}]$  for a concave rate function  $\phi$  simplifies to

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & \|A_0x - b\| - v_{wc}^*(0) \leq \alpha \phi(0), \\ & \|A_0x - b\| + \gamma_{\max} \|x\| - v_{wc}^*(\gamma_{\max}) \leq \alpha \phi(\gamma_{\max}). \end{aligned}$$

## 5.2. Linear Formulations with Polyhedral Norm Uncertainty

Adversarial robust optimization formulation (GARO) has an exact reformulation also in the context of uncertain linear optimization problems.

**Assumption 2.** Let  $X$  be a polyhedral set,  $P$  unconstrained, the distance function  $d(p, p') = \|p - p'\|$  defined by a polyhedral norm and  $f(x, p) = x^\top p$  a linear function.

A standard result from the robust optimization literature is that

$$v_{wc}(x, \gamma) := \max_{\|p - p_0\| \leq \gamma} x^\top p = x^\top p_0 + \gamma \|x\|_*$$

where  $\|\cdot\|_*$  is the dual norm. Hence, the worst-case function is affine in  $\gamma$  for any decision  $x$  and satisfies Assumption 1. Consequently, for concave rate functions  $\phi$  Proposition 1 applies and our associated (GARO) formulation reduces to simply balancing nominal and worst-case costs.

For general rate functions  $\phi$ , the situation is more complicated. As however the dual norm  $\|\cdot\|_*$  is also polyhedral and  $X$  a polytope, we have that the robust oracle cost

$$v_{wc}^*(\gamma) := \min_{x \in X} x^\top p_0 + \gamma \|x\|_* \tag{34}$$

is characterized as a linear representable minimization problem. A standard result from parametric linear optimization (Bertsimas and Tsitsiklis 1997) is that we can partition the interval  $\Gamma = \bigcup_{t=1}^T \Gamma_t$  such that on each interval  $\Gamma_t = [\gamma_t, \gamma_{t+1})$  an extreme point  $x_t$  is optimal in (34) so that  $v_{wc}^*(\gamma) = x_t^\top p_0 + \gamma \|x_t\|_\star$  for all  $\gamma \in \Gamma_t$ . The semi-infinite constraint in the adversarial robust formulation (GARO) hence reduces to  $T$  semi-infinite constraints

$$x^\top p_0 + \gamma \|x\|_\star - x_t^\top p_0 - \gamma \|x_t\|_\star \leq \alpha \phi(\gamma) \quad \forall \gamma \in \Gamma_t \quad \forall t \in [1, \dots, T]. \quad (35)$$

We now indicate that the latter constraints can be reformulated as  $T$  convenient convex constraints. Associate indeed with every interval  $\Gamma_t$  an increasing convex conjugate function given as

$$\phi_t^*(\Delta) := \max_{\gamma \in \Gamma_t} \Delta \gamma - \phi(\gamma) \quad (36)$$

and hence we can reformulate the semi-infinite constraints (35) as the convex constraints

$$(x - x_t)^\top p_0 + \alpha \phi_t^*((\|x\|_\star - \|x_t\|_\star)/\alpha) \leq 0 \quad \forall t \in [1, \dots, T].$$

Here the  $(\Delta, \alpha) \mapsto \alpha \phi_t^*(\Delta/\alpha)$  term is the perspective of  $\Delta \mapsto \phi_t^*(\Delta)$ , which is well-defined and convex for  $\alpha \geq 0$  (Combettes and Müller 2018).

### 5.3. Discretization Formulations

As  $\gamma \in \Gamma$  is a one-dimensional parameter, Problem GARO can be approximated by appropriately discretizing the interval  $\Gamma$ . For a given discretization  $\min \Gamma = \gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_{T+1} = \max \Gamma$ , the discretized version of GARO is defined here as the problem

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & \max_{p \in P_{\gamma_t}} f(x, p) - v_{wc}^*(\gamma_t) \leq \alpha \phi(\gamma_t) \quad \forall t \in [1, \dots, T+1]. \end{aligned} \quad (\text{GARO}_d)$$

We recall that the oracle costs  $v_{wc}^*(\gamma_t) := \min_{x' \in X} \max_{p' \in P_{\gamma_t}} f(x', p')$  require merely the solution of a classical robust optimization formulation. Hence, Problem  $(\text{GARO}_d)$  is an optimization problem with  $T+1$  robust constraints. Using standard dual representations, the maximization terms in  $(\text{GARO}_d)$  can in a wide range of circumstances be reformulated (see Ben-Tal et al. (2015)) resulting in an equivalent optimization problem amenable to off-the-shelf solvers. Alternatively, classical constraint generation methods for robust optimization problems could be applied to each of the  $T+1$  constraints to solve the problem.

**Assumption 3.** *Assume a selection  $x_{rob}(\gamma) \in X_{rob}(\gamma) = \arg \min_{x \in X} \max_{p \in P_\gamma} f(x, p)$  exists for all*

$\gamma \in \Gamma$ . We assume that  $\gamma' \mapsto \max_{p \in P_{\gamma'}} f(x_{rob}(\gamma), p)$  is Lipschitz continuous with constant  $L$  for all  $\gamma \in \Gamma$ .

In the following we study the effect of discretizing the interval  $\Gamma$ . The following theorem states that the discretized formulation (GAROD) returns a solution which violates the constraints in the adversarial robust formulation GARO by a value which is well controlled by the grid size under Assumption 3.

**Theorem 3.** *Let Assumption 3 hold and let  $\Delta := \max_{t=1, \dots, T} \gamma_{t+1} - \gamma_t$  be the grid size. Denote with  $L'$  the Lipschitz constant of  $\phi$  over  $\Gamma$ . An optimal solution  $(x_{garo}^{\Delta}, \alpha_{garo}^{\Delta})$  in (GAROD) satisfies for all  $\gamma \in \Gamma$  the inequality*

$$\max_{p \in P_{\gamma}} \left[ f(x_{garo}^{\Delta}, p) - \min_{x' \in X} \max_{p' \in P_{\gamma}} f(x', p') \right] - \alpha_{garo}^{\Delta} \phi(\gamma) \leq \Delta(L + \alpha_{garo}^{\Delta} L').$$

In other words,  $(x_{garo}^{\Delta}, \alpha_{garo}^{\Delta})$  is  $\Delta(L + \alpha_{garo}^{\Delta} L')$ -feasible in GARO. Since  $\alpha_{garo}^{\Delta}$  is available after solving the discretized problem, the bound  $\Delta(L + \alpha_{garo}^{\Delta} L') \leq \varepsilon$  can be verified a posteriori and the grid refined if necessary. Theorem 3 is a feasibility statement, but it implies an optimality guarantee on  $\alpha_{garo}^{\Delta}$  directly. Denote  $\phi_{\min} := \phi(\min \Gamma)$ . If  $\phi_{\min} = 0$ , the GARO constraint at  $\gamma = \min \Gamma$  forces  $A(x, \min \Gamma) = 0$ , so  $x_{garo} \in X_{rob}(\min \Gamma)$  and the problem essentially reduces to a standard robust formulation. In a nontrivial case  $\phi_{\min} > 0$ , the discretized problem GAROD has fewer constraints than GARO, giving  $\alpha_{garo}^{\Delta} \leq \alpha_{garo}$ . Theorem 3 further implies that  $(x_{garo}^{\Delta}, \alpha_{garo}^{\Delta} + \Delta(L + \alpha_{garo}^{\Delta} L')/\phi_{\min})$  is feasible in GARO, since  $\phi(\gamma) \geq \phi_{\min}$  for all  $\gamma \in \Gamma$ , and hence  $\alpha_{garo} \leq \alpha_{garo}^{\Delta} + \Delta(L + \alpha_{garo}^{\Delta} L')/\phi_{\min}$ . Combining,

$$0 \leq \alpha_{garo} - \alpha_{garo}^{\Delta} \leq \Delta(L + \alpha_{garo}^{\Delta} L')/\phi_{\min} \leq \Delta(L + \alpha_{garo} L')/\phi_{\min},$$

so  $\alpha_{garo}^{\Delta} \rightarrow \alpha_{garo}$  at rate  $\mathcal{O}(\Delta)$  as the grid is refined.

*Proof of Theorem 3.* We recall that we defined  $v_{wc}(x, \gamma) := \max_{p \in P_{\gamma}} f(x, p)$  and  $v_{wc}^*(\gamma) := \min_{x \in X} v_{wc}(x, \gamma)$ . Assume now that  $\gamma \in \Gamma$  is arbitrary. There exists a  $j \in \{1, \dots, T\}$  such that  $\gamma \in [\gamma_j, \gamma_{j+1})$ . Let  $x_j = x_{rob}(\gamma_j) \in X_{rob}(\gamma_j)$  be an optimal solution of the min-max problem with uncertainty set  $P_{\gamma_j}$ . We can bound

$$\begin{aligned} v_{wc}(x_{garo}^{\Delta}, \gamma) - v_{wc}^*(\gamma) &\leq v_{wc}(x_{garo}^{\Delta}, \gamma_{j+1}) - v_{wc}^*(\gamma_j) \\ &\leq v_{wc}^*(\gamma_{j+1}) - v_{wc}^*(\gamma_j) + \alpha_{garo}^{\Delta} \phi(\gamma_{j+1}) \\ &\leq v_{wc}(x_j, \gamma_{j+1}) - v_{wc}(x_j, \gamma_j) + \alpha_{garo}^{\Delta} \phi(\gamma_{j+1}) \end{aligned}$$

where the first inequality follows since  $v_{wc}(x, \gamma)$  is non-decreasing in  $\gamma$  for every  $x$ , the second inequality follows since  $(x_{garo}^{\Delta}, \alpha_{garo}^{\Delta})$  is a feasible solution in Problem (GAROD), and the third inequality follows

since  $x_j \in X$  is a minimizer of  $v_{wc}(x, \gamma_j)$ . We can continue bounding the last term as

$$\begin{aligned}
& v_{wc}(x_j, \gamma_{j+1}) - v_{wc}(x_j, \gamma_j) + \alpha_{garo}^\Delta \phi(\gamma_{j+1}) \\
& \leq (\gamma_{j+1} - \gamma_j)L + \alpha_{garo}^\Delta \phi(\gamma_{j+1}) \\
& = (\gamma_{j+1} - \gamma_j)L + \alpha_{garo}^\Delta \phi(\gamma) + \alpha_{garo}^\Delta (\phi(\gamma_{j+1}) - \phi(\gamma)) \\
& \leq (\gamma_{j+1} - \gamma_j)L + \alpha_{garo}^\Delta \phi(\gamma) + \alpha_{garo}^\Delta L'(\gamma_{j+1} - \gamma) \\
& \leq (\gamma_{j+1} - \gamma_j)L + \alpha_{garo}^\Delta \phi(\gamma) + \alpha_{garo}^\Delta L'(\gamma_{j+1} - \gamma_j)
\end{aligned}$$

where the first inequality follows from Assumption 3, the second inequality follows from Lipschitz continuity of  $\phi$  with constant  $L'$ , and the third inequality follows since  $\gamma_j \leq \gamma$ .  $\square$

In a wide array of problems, including those described in Sections 5.1 and 5.2, the mapping  $\gamma \mapsto v_{wc}(x, \gamma) := \max_{p \in P_\gamma} f(x, p)$  is Lipschitz continuous and hence Assumption 3 indeed holds. For instance, when the worst-case cost function  $\gamma \mapsto v_{wc}(x, \gamma)$  is concave in  $\gamma$ , then  $v_{wc}^*(\gamma) = \min_{x \in X} v_{wc}(x, \gamma)$  is a concave function as well and Assumption 3 holds with Lipschitz constant  $L = \partial_\gamma v_{wc}^*(0) = \partial_\gamma v_{wc}(x_{nom}, 0)$  by the envelope theorem.

**Example 7.** Reconsider the uncertain linear regression problem  $\min_{x \in X} \|Ax - b\|$  where the matrix  $A$  is uncertain. We have  $v_{wc}(x, \gamma) := \max_{\|A - A_0\| \leq \gamma} \|Ax - b\| = \|A_0x - b\| + \gamma \|x\|$  is affine and hence concave. It follows that Assumption 3 holds for  $L = \|x_{nom}\| = \|A_0^\dagger b\|$ . Similarly, the uncertain linear optimization problem in Section 5.2 where  $v_{wc}(x, \gamma) := \max_{\|p - p_0\| \leq \gamma} x^\top p = x^\top p_0 + \gamma \|x\|_*$  which is affine and hence concave. It follows that Assumption 3 holds for  $L = \|x_{nom}\|_*$ .

**Example 8.** Consider an uncertain stochastic optimization problem  $\min_{x \in X} \mathbb{E}_\mathbb{P}[\ell(x, \xi)]$  where the distribution  $\mathbb{P}$  is uncertain. We assume here that  $\mathbb{P}$  is supported on  $\Xi$  and that the  $\ell_{\min} \leq \ell(x, \xi) \leq \ell_{\max}$  for all  $x \in X$  and  $\xi \in \Xi$ . Suppose we have access to a prediction  $\mathbb{P}_0$  for which we expect the Kullback-Leibler distance to  $P$  to remain small. Given that the Kullback-Leibler distance behaves quadratically, i.e.,  $\text{KL}(\lambda \mathbb{P}_0 + (1 - \lambda)\mathbb{P}, \mathbb{P}) = \frac{\lambda^2}{2} \chi^2(\mathbb{P}_0, \mathbb{P}) + o(\lambda^2)$  with  $\chi^2(\mathbb{P}_0, \mathbb{P}) := \int (d\mathbb{P}_0/d\mathbb{P}(\xi) - 1)^2 d\mathbb{P}(\xi)$ , we will impose  $\text{KL}(\mathbb{P}_0, \mathbb{P}) \leq \gamma^2$ . A well known result from the optimization literature (Van Parys et al. 2021) is that the worst-case cost function satisfies the dual characterization

$$v_{wc}(x, \gamma) = \begin{cases} \max_{\mathbb{P}} \mathbb{E}_\mathbb{P}[\ell(x, \xi)] \\ \text{s.t. } \text{KL}(\mathbb{P}_0, \mathbb{P}) \leq \gamma^2 \end{cases}$$

$$= \begin{cases} \min_{\alpha} & \alpha - \exp(-\gamma^2) \exp(\mathbb{E}_{\mathbb{P}_0} [\log(\alpha - \ell(x, \xi))]) \\ \text{s.t.} & \max_{\xi \in \Xi} \ell(x, \xi) \leq \alpha \end{cases}$$

and can be evaluated efficiently using simple bisection search. Furthermore, as Lemma 6 points out the worst-case function is uniformly Lipschitz in the decision  $x \in X$  and hence Assumption 3 is satisfied.

#### 5.4. Constraint Generation

The number of discretization points required to achieve a small constraint violation via the discretization approach in the previous section can be large when working with large Lipschitz constants or a large interval  $\Gamma$ . In this case it could be beneficial to iteratively discretize  $\Gamma$  by calling a separation oracle which finds the  $\gamma$  value that violates the constraint in GARO; see Algorithm 1.

**Assumption 4.** *Let  $f(x, p)$  be Lipschitz continuous on  $x \in X$  uniformly for all  $p \in P_{\max \Gamma}$ ,  $X$ ,  $\Gamma$  and  $P_{\gamma}$  compact sets for every  $\gamma \in \Gamma$ , and  $\phi$  Lipschitz over  $\Gamma$ .*

**Assumption 5.** *We assume we can solve the univariate separation problem*

$$\max_{\gamma \in \Gamma} \max_{p \in P_{\gamma}} f(x, p) - \min_{x' \in X} \max_{p' \in P_{\gamma}} f(x', p') - \alpha \phi(\gamma)$$

for any  $x \in X$  and  $\alpha \geq 0$ .

We show that under Assumptions 4 and 5 constraint generation Algorithm 1 converges to an approximate solution in our adversarial regret optimization formulation.

**Theorem 4.** *Let Assumptions 4 and 5 hold and let GARO be feasible. Then, Algorithm 1 converges to an  $\varepsilon$ -feasible solution of our GARO.*

*Proof.* Since GARO is feasible there exists a solution  $(\bar{x}, A)$  satisfying

$$\max_{p \in P_{\gamma}} \left[ f(\bar{x}, p) - \min_{x' \in X} \max_{p' \in P_{\gamma}} f(x', p') \right] \leq A \phi(\gamma) \quad \forall \gamma \in \Gamma.$$

We can hence restrict  $(x, \alpha)$  in GARO to the compact set  $X \times [0, A]$  without loss of optimality.

Following the convergence analysis in (Mutapcic and Boyd 2009, Section 5.2), our cutting plane algorithm converges if the constraint function

$$g(x, \alpha; \gamma) = \max_{p \in P_{\gamma}} f(x, p) - \min_{x' \in X} \max_{p' \in P_{\gamma}} f(x', p') - \alpha \phi(\gamma)$$

---

**Algorithm 1** Constraint Generation Algorithm for GARO

---

**Require:**  $X \subset \mathbb{R}^n$  compact,  $f(x, p)$  Lipschitz continuous in  $x$  for every  $p \in P$ , interval  $\Gamma$ ,  $\varepsilon > 0$

**Ensure:**  $\varepsilon$ -optimal solution of Problem GARO

- 1: Set  $C \leftarrow \emptyset$
- 2: Set  $v_{wc}^* \leftarrow \infty$
- 3: **while**  $v_{wc}^* \geq \varepsilon$  **do**
- 4:     Calculate an optimal solution  $(x^*, \alpha^*)$  in the main problem

$$\begin{aligned} & \min_{x \in X, \alpha \geq 0} \alpha \\ & \text{s.t. } \max_{p \in P_\gamma} \left[ f(x, p) - \min_{x' \in X} \max_{p' \in P_\gamma} f(x', p') \right] \leq \alpha \phi(\gamma) \quad \forall \gamma \in C. \end{aligned}$$

- 5:     Calculate an optimal solution  $\gamma^*$  of the separation problem

$$v_{wc}^* \leftarrow \max_{\gamma \in \Gamma} \max_{p \in P_\gamma} \left[ f(x^*, p) - \min_{x' \in X} \max_{p' \in P_\gamma} f(x', p') \right] - \alpha^* \phi(\gamma).$$

- 6:     Set  $C \leftarrow C \cup \{\gamma^*\}$ .
  - 7: **end while**
  - 8: **return**  $(x^*, \alpha^*)$
- 

are Lipschitz continuous in the decision variables, i.e.,  $x$  and  $\alpha$ , uniformly over the quantifiers  $\gamma \in \Gamma$ . Since  $f$  is Lipschitz continuous in  $x$  uniformly over  $p \in P_{\max \Gamma}$  the worst-case cost  $(x, \alpha) \mapsto v_{wc}(x, \gamma) := \max_{p \in P_\gamma} f(x, p)$  must be Lipschitz continuous for every  $\gamma$  since

$$v_{wc}(x_1, \gamma) - v_{wc}(x_2, \gamma) \leq f(x_1, p^*) - f(x_2, p^*) \leq L \|x_1 - x_2\|$$

where  $p^*$  is a maximizer of  $\max_{p \in P_\gamma} f(x_1, p)$  and  $L$  the uniform Lipschitz constant of  $f(x, p)$  in  $x \in X$  over  $p \in P_{\max \Gamma}$ . Clearly  $(x, \alpha) \mapsto \alpha \phi(\gamma)$  is Lipschitz continuous (and linear) in  $\alpha$  with Lipschitz constant  $\max_{\gamma \in \Gamma} \phi(\gamma) < \infty$ , and the claim follows.  $\square$

By the same reasoning as in Section 5.3,  $\varepsilon$ -feasibility also implies an optimality guarantee in the nontrivial case  $\phi_{\min} > 0$ . At termination,  $(x^*, \alpha^*)$  satisfies  $A(x^*, \gamma) \leq \alpha^* \phi(\gamma) + \varepsilon$  for all  $\gamma \in \Gamma$ , so  $(x^*, \alpha^* + \varepsilon / \phi_{\min})$  is feasible in GARO and hence  $\alpha_{garo} \leq \alpha^* + \varepsilon / \phi_{\min}$ . Combined with  $\alpha^* \leq \alpha_{garo}$ , we obtain  $0 \leq \alpha_{garo} - \alpha^* \leq \varepsilon / \phi_{\min}$ .

The separation problem over  $\gamma$  which has to be solved in Algorithm 1 is a univariate maximization problem. When the function  $\gamma \rightarrow \max_{p \in P_\gamma} f(x, p)$  is Lipschitz this can be done to any desired accuracy using the Shubert–Piyavskii algorithm. If  $\gamma \rightarrow \max_{p \in P_\gamma} f(x, p)$  is concave, this could also be done by exploiting DC-programming algorithms. However, in certain situations when the objective function is linear, the separation problem can be calculated exactly; see Section 5.2.

**Example 9.** We consider a generalized version of the uncertain linear optimization problem discussed in Section 5.2 in which  $P_\gamma = \{p \in P : \|p - p_0\| \leq \gamma\}$  and where now  $P$  can be an arbitrary convex polytope. Consider support function  $h_P(u) = \max_{p \in P} u^\top p$ . In the linear case  $f(x, p) = x^\top p$  we have that  $v_{wc}(x, \gamma) := \max_{p \in P_\gamma} f(x, p) = \min_y y^\top p_0 + \|y\|_\star \gamma + h_P(x - y)$ .

Observe that as here  $\|\cdot\|_\star$  is a polyhedral norm,  $X$  a polytope and  $h_P$  a linear representable function both

$$\begin{aligned} v_{wc}(x, \gamma) &:= \min_y y^\top p_0 + \|y\|_\star \gamma + h_P(x - y) \quad \forall \gamma \in \Gamma \\ v_{wc}^\star(\gamma) &:= \min_{x \in X} \max_{p \in P_\gamma} f(x, p) = \min_{x' \in X, y'} y'^\top p_0 + \|y'\|_\star \gamma + h_P(x' - y') \quad \forall \gamma \in \Gamma \end{aligned} \quad (37)$$

are linear optimization problems. Hence, for any fixed decision  $x$  we can partition  $\Gamma = \cup_{t=1}^T \Gamma_t$  such that on each  $\Gamma_t = [\gamma_t, \gamma_{t+1})$  we can find extreme points  $y_t$  and  $(x'_t, y'_t)$  in (37) so that

$$\begin{aligned} v_{wc}(x, \gamma) &:= y_t^\top p_0 + \|y_t\|_\star \gamma + h_P(x - y_t) \quad \forall \gamma \in \Gamma_t \\ v_{wc}^\star(\gamma) &:= \min_{x \in X} \max_{p \in P_\gamma} f(x, p) = y_t'^\top p_0 + \|y_t'\|_\star \gamma + h_P(x'_t - y'_t) \quad \forall \gamma \in \Gamma_t \end{aligned}$$

Hence, using the same argument as in Section 5.2 we can write the separation problem as a maximum of at most  $T$  terms, i.e.,

$$\begin{aligned} &\max_{\gamma \in \Gamma} \max_{p \in P_\gamma} \left[ f(x, p) - \min_{x' \in X} \max_{p' \in P_\gamma} f(x', p') \right] - \alpha \phi(\gamma) \\ &= \max_{t=1}^T (y_t - y'_t)^\top p_0 + h_P(x - y_t) - h_P(x'_t - y'_t) + \alpha \phi_t^\star ((\|y'_t\|_\star - \|y_t\|_\star) / \alpha) \end{aligned}$$

where the convex conjugate  $\phi_t^\star$  associated with each  $\Gamma_t$  is stated in Equation (36).

## 6. A Minimum Knapsack Experiment

We apply the algorithmic framework of Section 5 to minimum knapsack problems with uncertainty sets constructed from randomly generated data to assess the practical performance of GARO. The code for the experiments is publicly available at <https://github.com/Jannisku/GARO.git>.

## 6.1. Setup and Baseline Methods

We consider as deterministic problem the minimum knapsack problem of the form

$$\begin{aligned} \min_x x^\top p \\ \text{s.t. } a^\top x \geq b \\ x \in [0, 100]^n. \end{aligned}$$

The bounds on the variables prevent unbounded problems when a random scenario  $p$  has negative entries. Furthermore, we refrain from applying an upper bound of one to each variable since the effect of the robust models in this case is only marginal. Intuitively, for the chosen variant a nominal solution is sparse, while robustness has the effect of reducing this sparsity.

We assume a training set  $\mathcal{D}$  and a test set  $\mathcal{T}$  of historical data points in 80/20 proportion. We then define the uncertainty set

$$P_\gamma := \{p \in \mathbb{R}^n : (p - p_0)^\top \tilde{\Sigma}^{-1} (p - p_0) \leq \gamma\}$$

where  $p_0$  is the empirical mean prediction and  $\tilde{\Sigma}$  the empirical covariance matrix of  $\mathcal{D}$ . If the smallest eigenvalue of  $\tilde{\Sigma}$  is smaller than  $10^{-4}$ , we add the matrix  $10^{-4}\mathbb{1}_n$  to  $\tilde{\Sigma}$ . For a given confidence level  $\rho \in [0, 1)$ , we define  $\gamma_\rho$  as the smallest value such that  $P_{\gamma_\rho}$  contains a fraction  $\rho$  of all training data points. We define the interval for the  $\gamma$ -parameter as  $\Gamma = [0, \gamma_{0.99}]$ . Choosing  $\gamma_{0.99}$  instead of  $\gamma_{1.0}$  avoids calibrating the interval to the single most extreme sample, yielding a more stable estimate of the effective data range.

We compare our method (GARO) against several baseline methods:

- (RO) robust optimization with ellipsoidal uncertainty set,
- (RO<sub>d</sub>) robust optimization with discrete uncertainty set,
- (SAT) robust satisficing with ellipsoidal uncertainty set, and
- (REG) classical robust regret with discrete uncertainty set.

We chose the discrete version of robust regret since the same problem with convex uncertainty sets is computationally intractable. We briefly describe below how each method is solved. We denote by  $X := \{x \in \mathbb{R}^n : a^\top x \geq b, x \in [0, 100]^n\}$  the feasible set.

**Classical Robust Optimization with Convex Uncertainty (RO)** For a given  $\gamma \in \Gamma$ , we solve the problem  $\min_{x \in X} \max_{p \in P_\gamma} x^\top p$  which can be reformulated as

$$\min_{x \in X} x^\top p_0 + \sqrt{\gamma} \sqrt{x^\top \tilde{\Sigma} x}. \quad (38)$$

**Classical Robust Optimization with Discrete Uncertainty (RO<sub>d</sub>)** For a given  $\gamma \in \Gamma$ , we collect all training scenarios which are contained in  $P_\gamma$ , i.e., we define  $\mathcal{D}(\gamma) := \mathcal{D} \cap P_\gamma$  and solve the problem  $\min_{x \in X} \max_{p \in \mathcal{D}(\gamma)} x^\top p$  which can be reformulated as

$$\begin{aligned} \min_{x \in X, \alpha} \quad & \alpha \\ \text{s.t.} \quad & x^\top p \leq \alpha \quad \forall p \in \mathcal{D}(\gamma). \end{aligned}$$

**Robust Satisficing (SAT)** We discretize the interval  $[0, \gamma_{0.99}]$  into  $T + 1 = 100$  equidistant values  $0 = \gamma_1 \leq \dots \leq \gamma_{T+1} = \gamma_{0.99}$ . We then calculate the oracle nominal cost  $v_{wc}^*(0) = \min_{x \in X} x^\top p_0$ . For a given factor  $\beta \in [1, \infty)$  we define the target value as  $f_0 = \beta v_{wc}^*(0)$  and then solve the robust satisficing problem

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & \max_{p \in P_{\gamma_t}} x^\top p \leq f_0 + \alpha \gamma_t \quad \forall t \in [1, \dots, T + 1]. \end{aligned}$$

**Robust Regret (REG)** The regret formulation for a given  $\gamma \in \Gamma$  solves  $\min_{x \in X} \max_{p \in P_\gamma} [x^\top p - \min_{x' \in X} x'^\top p]$ , which is computationally intractable due to the nonconvexity induced by the inner minimization. To allow computation, we consider an approximation in which the outer maximization ranges only over the training scenarios contained in  $P_\gamma$ . We define  $\mathcal{D}(\gamma) := \mathcal{D} \cap P_\gamma$  and solve the approximate problem  $\min_{x \in X} \max_{p \in \mathcal{D}(\gamma)} [x^\top p - \min_{x' \in X} x'^\top p]$  which can be reformulated as

$$\begin{aligned} \min_{x \in X, \alpha} \quad & \alpha \\ \text{s.t.} \quad & x^\top p \leq \alpha + \min_{x' \in X} x'^\top p \quad \forall p \in \mathcal{D}(\gamma). \end{aligned}$$

**Globalized Adversarial Regret Optimization (GARO)** We discretize the interval  $[0, \gamma_{0.99}]$  into  $T + 1 = 100$  equidistant values  $0 = \gamma_1 \leq \dots \leq \gamma_{T+1} = \gamma_{0.99}$ . For each  $\gamma_t$  we calculate the optimal value  $v_{wc}^*(\gamma_t)$  of the classical robust problem, i.e.,

$$v_{wc}^*(\gamma_t) := \min_{x \in X} \max_{p \in P_{\gamma_t}} x^\top p.$$

We use the power rate function  $\phi(\gamma) = (1 + \gamma)^q$  with parameter  $q \geq 0$  introduced in Example 4, which interpolates between the absolute guarantee ( $q = 0$ ) and increasingly permissive adversarial regret growth for larger  $q$ . We then solve the discretized global robust regret problem

$$\begin{aligned} \min_{x \in X, \alpha} \quad & \alpha \\ \text{s.t.} \quad & \max_{p \in P_{\gamma_t}} x^\top p - v_{wc}^*(\gamma_t) \leq \alpha(1 + \gamma_t)^q \quad \forall t \in [1, \dots, T + 1]. \end{aligned}$$

The maximization term can be reformulated as shown for the classical robust optimization baseline.

**Instance Generation** We generate 5 random problem instances and 5 random data sets, each containing  $m = 5000$  random observations on the profit vector  $p$ , which we split into training data  $\mathcal{D}$  and test data  $\mathcal{T}$  in 80/20 proportion. The problem instances are generated as follows:

- (i) We set  $n = 50$  and draw five different weight vectors  $a$  uniformly from the box  $[25, 100]^n$ .
- (ii) We define the knapsack capacity as  $b = \frac{2}{5} \sum_{i=1}^n a_i$ .

For the data sets we consider different types of randomly generated data:

- (a) **Gaussian:** We draw  $\mu$  uniformly from the box  $[0, 50]^n$  and the standard deviation  $\sigma_i$  uniformly from  $[0, \frac{1}{2}\mu_i]$  for each  $i \in [n]$ . We additionally draw a random orthonormal basis  $u^1, \dots, u^n \in \mathbb{R}^n$  and define the covariance matrix as  $\Sigma = \sum_{i=1}^n \sigma_i^2 u^i (u^i)^\top$ . We then sample  $m$  random points from  $\mathcal{N}(\mu, \Sigma)$ .
- (b) **Gaussian with Inverse Mean-Variance Relationship:** The points are generated as in the Gaussian case, except that the standard deviation  $\sigma_i$  is chosen uniformly from  $[0, 50 - \mu_i]$  for each  $i \in [n]$ . This ensures that large mean values are paired with small standard deviations and vice versa. This choice is motivated by the observation that in the robust case the size of the uncertainty has a larger impact on the ordering of entries in the worst-case scenario, which intuitively leads to more diverse solutions for different  $\gamma$  values.
- (c) **Heavy-Tail:** We draw  $m$  random points from the Pareto distribution with tail index 1.5, and add the all-ones vector  $\mathbf{1}_n$  to each drawn point.

## 6.2. Preliminary Analysis of Classical Robust Optimization

We start with an analysis of RO with Gaussian samples to introduce our evaluation metrics and show the drawbacks of the guarantees provided by the model. A critical question in RO is how to choose  $\gamma$  for the uncertainty set  $P_\gamma$  to obtain a solution which performs well out-of-sample under a given set of performance metrics. To determine a good value for  $\gamma$ , we solve RO with radius  $\gamma = \theta\gamma_{0.99}$  where  $\theta \in \{0.7^i : i = 0, 1, \dots, 25\}$ , which we denote by  $\text{RO}(\theta)$ . Note that for  $\theta = 0$  we obtain the nominal solution associated with the prediction  $p_0$ , while for  $\theta = 1$  we use an ellipsoid containing 99% of the training data. For each  $\gamma$  we solve each combination of the 5 problem instances and the 5 training data sets, resulting in 25 instances.

Each solution is evaluated on all test scenarios  $p \in \mathcal{T}$ , i.e., we calculate the objective value  $x^\top p$  for every solution  $x$  and test scenario  $p$ , and evaluate the average, worst-case and 90%-quantile performance. While robust optimization typically aims for solutions with good performance in the upper quantiles, these solutions should also not perform too poorly on average. Hence, decision makers are

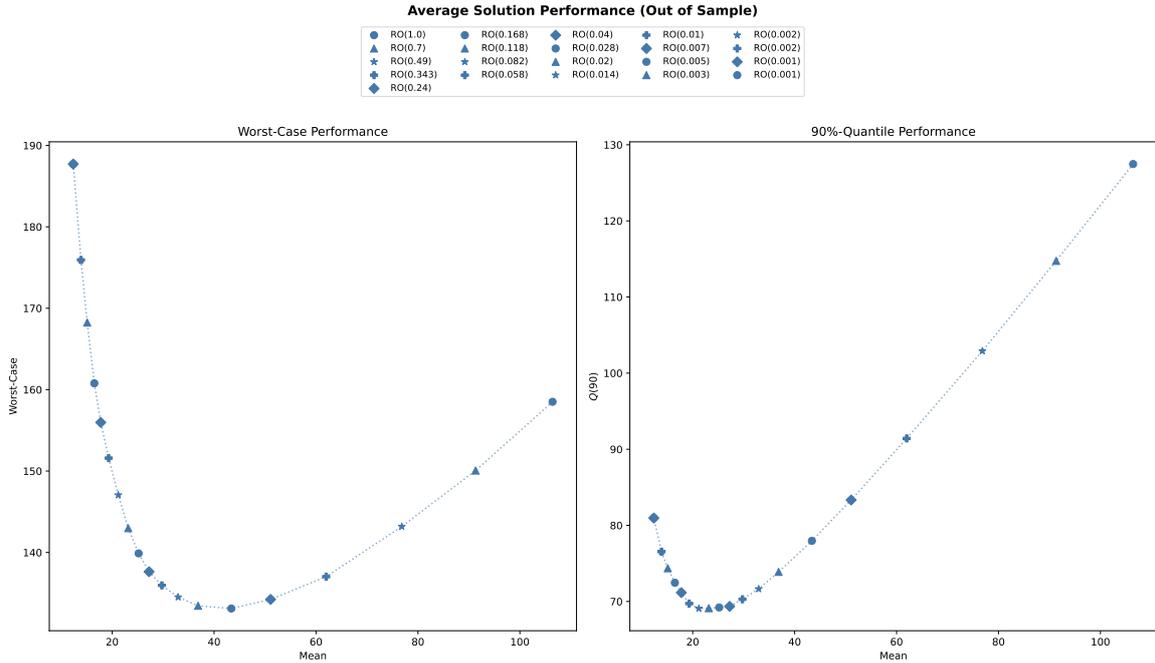


Figure 3: Out-of-sample performance of RO for the minimum knapsack problem with  $n = 50$  for Gaussian data. Mean vs. worst-case (left) and mean vs. 90%-quantile (right).

often interested in the trade-off between average and worst-case performance. In Figure 3 we show the average out-of-sample performance over all 25 instances, plotting average vs. worst-case (left) and average vs. 90%-quantile (right). The results show that good trade-offs between average and worst-case (or 90%-quantile) are achieved for very small  $\gamma$  radii. Beyond  $\theta = 0.168$ , the performance in both average and worst-case (or 90%-quantile) already deteriorates as  $\theta$  increases. The performance is also highly sensitive to  $\gamma$  for small values thereof.

In practice, finding a good value for  $\gamma$  as in the previous analysis can be done via an empirical out-of-sample performance test on a validation dataset. However, due to limited or biased data, such empirical tests can be highly imprecise and hence misleading regarding the future performance of the calculated solution. To obtain a more reliable performance guarantee, one can instead consider the guarantee offered by the RO model itself. More precisely, for an optimal solution  $x_{RO(\theta)}$  of  $RO(\theta)$  with optimal value  $\text{opt}(RO(\theta))$ , we have

$$x_{RO(\theta)}^\top p^\star \leq \begin{cases} \text{opt}(RO(\theta)) & \text{if } \gamma^\star = d(p_0, p^\star) \leq \theta\gamma_{0.99} \\ \infty & \text{otherwise.} \end{cases}$$

These performance guarantees, plotted as a function of  $d(p_0, p)$ , are shown in Figure 4. All values

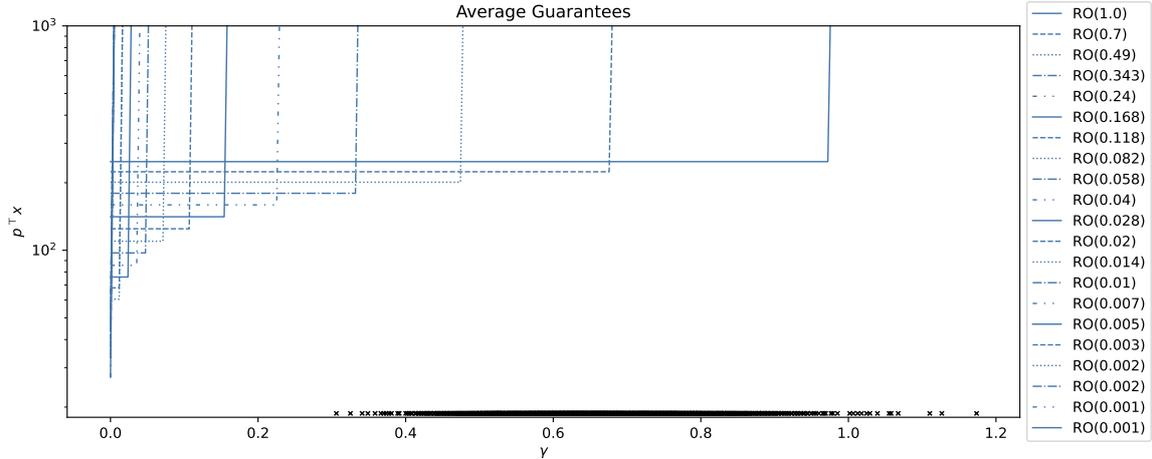


Figure 4: Performance guarantees of  $\text{RO}(\theta)$  in the minimum knapsack problem with  $n = 50$  for Gaussian data. Its guarantee is all-or-nothing: it is constant for  $d(p_0, p) \leq \theta\gamma_{0.99}$  and is vacuous beyond this threshold.

are averages over the 25 instances. For each test sample  $p \in \mathcal{T}$ , the value  $d(p_0, p)$  is indicated on the horizontal axis with a cross. The horizontal axis is normalized such that 1 corresponds to the largest  $\gamma_{0.99}$  over all problem instances. Since  $\gamma_{0.99}$  is computed from the training data only, test samples may exceed this value significantly.

The results show that the guarantees for the potential values of  $\theta$  which lead to good trade-offs ( $0 \leq \theta \leq 0.12$ ) are practically useless since they only hold for scenarios  $p$  close to the mean. However, this region does not contain a single test scenario and hence the robust optimization guarantee does not apply to any test scenario. In fact, the uncertainty sets  $P_\gamma$  which lead to a good out-of-sample performance do not contain a single training point. This stark observation highlights the impracticality of the guarantees classical robust formulations provide and the need for methods offering broader performance guarantees. Furthermore, the size of the uncertainty set does not provide any intuitive insight into the final performance of the corresponding solution. Indeed, the box-plots in Figure 5 suggest that the only notable benefit of increasing the uncertainty set is a reduction in the variance of the out-of-sample performance. This is unsurprising: from the reformulation (38), larger  $\gamma$  values assign greater weight to the standard deviation term in the objective.

### 6.3. Gaussian Data

We now turn to the main experiment comparing the different methods mentioned above. For the baseline methods and GARO we consider the following setups, chosen based on the best worst-case performance of each method in preliminary experiments. For RO we consider  $\gamma = \theta\gamma_{0.99}$  where  $\theta \in$

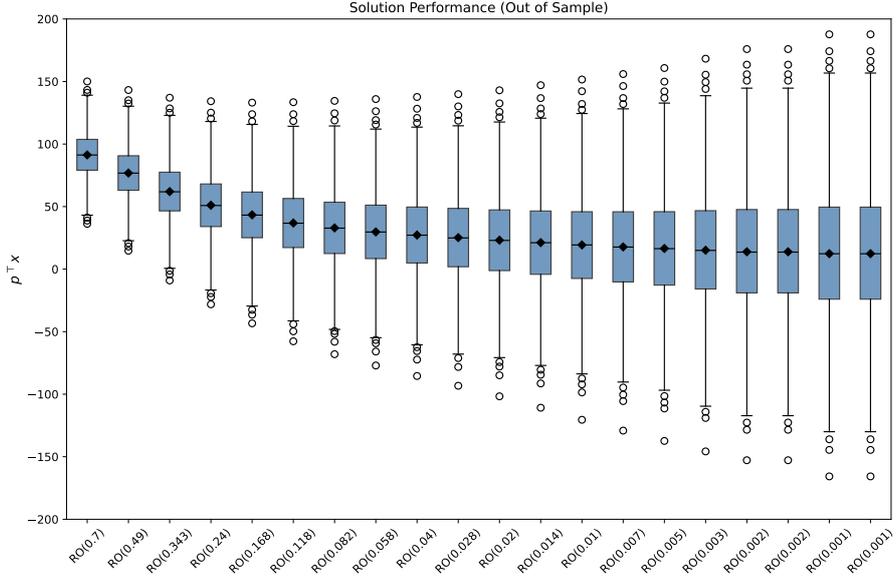


Figure 5: Boxplots of the out-of-sample objective values for the minimum knapsack problem with  $n = 50$  for Gaussian data. The diamonds denote the mean value.

$\{0, 0.02, \dots, 0.08\}$ , and for  $RO_d$  and REG we consider values  $\theta \in \{0.1, \dots, 0.5\}$ . For SAT we consider target values of  $\beta \in \{1.2, 1.4, \dots, 2.0\}$ , and for GARO we consider exponents  $q \in \{0, 0.5, \dots, 2.0\}$  in the rate function  $\phi(\gamma) = (1 + \gamma)^q$ . We denote the previously described methods as  $RO(\theta)$ ,  $RO_d(\theta)$ ,  $SAT(\beta)$ ,  $REG(\theta)$  and  $GARO(q)$ . For each type of data generation process we solve each combination of the 5 problem instances and the 5 training data sets, resulting in 25 instances.

In Figure 6 we show the average out-of-sample cost over all 25 instances, plotting average vs. worst-case (left) and average vs. 90%-quantile (right) for all methods. Each value is an average over all 25 instances. The points for  $REG(0.3)$ ,  $REG(0.4)$  and  $REG(0.5)$  are not shown in the plot since they lead to very large values on average and in worst-case and 90%-quantile costs.

The results show that the points of our method for  $q \geq 0.5$  provide a good trade-off lying close to the lower left corner of the plot and dominate the solutions of the other methods. The solutions of SAT provide similar values but are sensitive to the choice of the target level. In fact, for different target levels SAT will return one of the RO solutions for a certain  $\gamma$ -value, and indeed any point of the RO trade-off curve could be generated by setting a corresponding target level (see the discussion in Section 2.3).

The solutions of RO and  $RO_d$  provide good values for the worst-case costs with large increases in mean cost, where  $RO_d$  suffers most.  $RO_d$  deteriorates also for the 90%-quantile compared to RO. Only the nominal solution provides the best mean performance, at the cost of large worst-case and 90%-quantile costs (upper left corner). REG returns the nominal solution for  $\theta \in \{0, 0.1, 0.2\}$  but

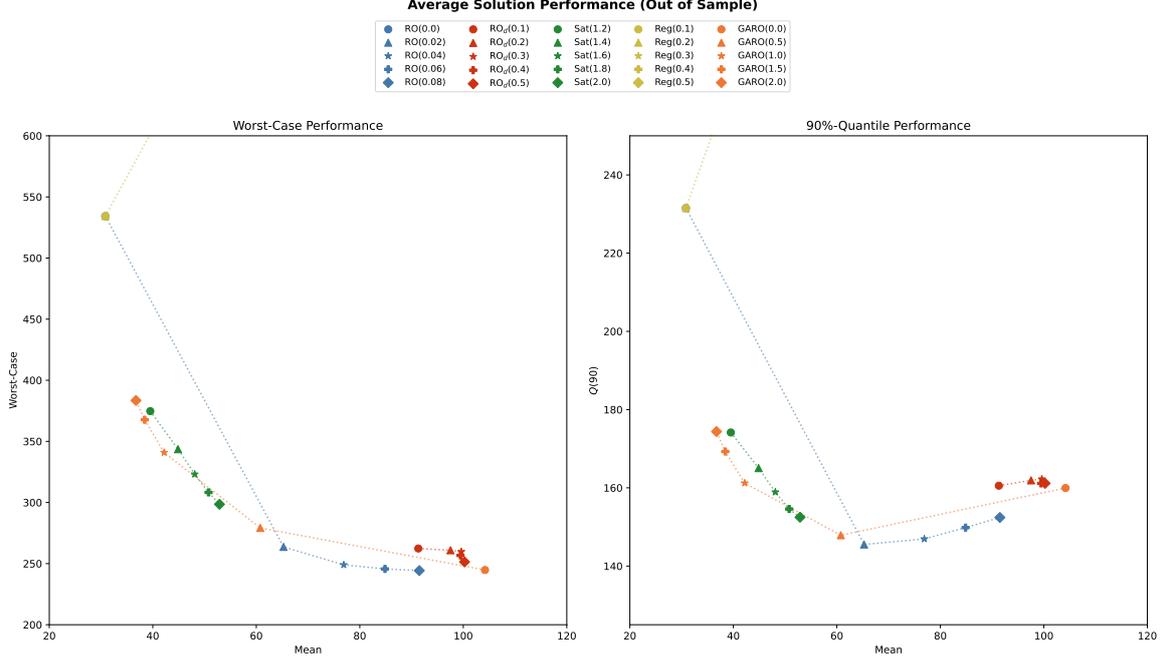


Figure 6: Out-of-sample performance for the minimum knapsack problem with  $n = 50$  for Gaussian data. Average vs. worst-case (left) and average vs. 90%-quantile (right).

leads to very poor performance in mean, worst-case and 90%-quantile costs for larger  $\theta$  values, which are not visible on the plot.

Some of the studied methods provide performance guarantees on future scenarios after the solution is calculated. We note that  $RO_d(\theta)$  only provides guarantees against scenarios in the training set  $\mathcal{D}(\gamma)$  and hence offers no formal out-of-sample guarantee. For an optimal solution  $x_{GARO(q)}$  of  $GARO(q)$  with corresponding optimal value  $\alpha_{GARO(q)}$ , the resulting performance guarantee is

$$x_{GARO(q)}^\top p^* \leq v_{wc}^*(\gamma^*) + \alpha_{GARO(q)}(1 + \gamma^*)^q$$

where  $\gamma^* = d(p_0, p^*)$ . Moreover, for an optimal solution  $x_{SAT(\beta)}$  of  $SAT(\beta)$  with corresponding optimal value  $\alpha_{SAT(\beta)}$ , the resulting performance guarantee is

$$x_{SAT(\beta)}^\top p^* \leq \beta v_{wc}^*(0) + \alpha_{SAT(\beta)} \gamma^*.$$

Finally, note that for REG we do not naturally obtain a guarantee depending only on  $\gamma^*$ .

These performance guarantees for Gaussian data, plotted as a function of  $d(p_0, p)$ , are shown in Figure 7. For each scenario  $p$ , the plot shows the guaranteed upper bound on the objective value provided by each method. All values are averages over the different problem and data instances.

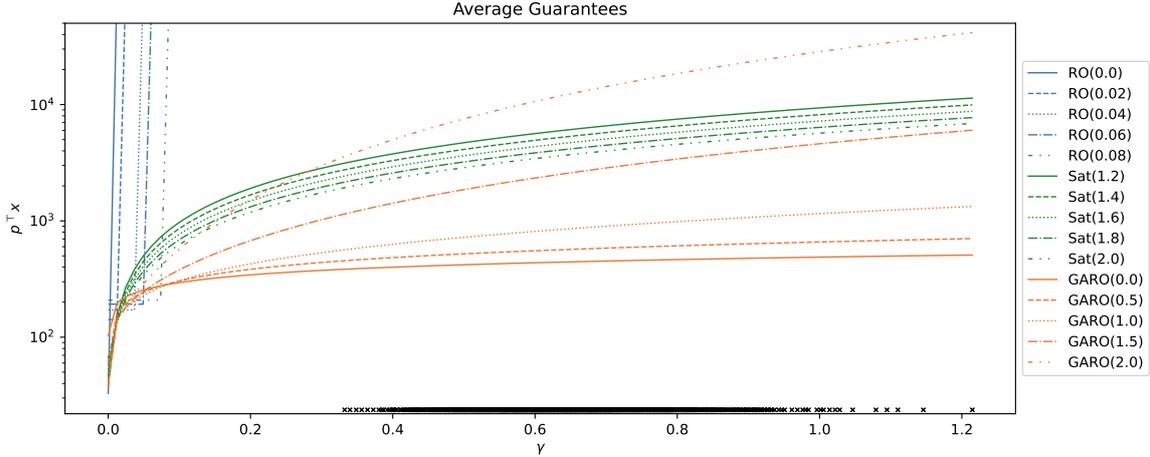


Figure 7: Performance guarantees for the minimum knapsack problem with  $n = 50$  for Gaussian data.

For each test sample  $p \in \mathcal{T}$ , the value  $d(p_0, p)$  is indicated on the horizontal axis with a cross. The horizontal axis is normalized such that 1.0 corresponds to the largest  $\gamma_{0.99}$  over all instances. Since  $\gamma_{0.99}$  is computed from the training set, test scenarios can have a normalized distance exceeding 1.0.

As shown in Figure 7, GARO provides significantly tighter guarantees throughout the test data region for nearly all values of  $q$ , compared to SAT and RO/ $RO_d$ . Only for a small region of scenarios very close to the nominal prediction  $p_0$  (which did not appear in the test set) are the guarantees of RO/ $RO_d$  slightly better than those of GARO. Since the uncertainty sets chosen for RO have a very small radius, the provided guarantees are practically vacuous: under Gaussian sampling the squared Mahalanobis distance of a test point to the training mean is approximately  $\chi_n^2$ -distributed with mean  $n$ , so the typical test scenario lies far outside the small uncertainty sets used by RO.

In Figure 8 we visualize the distribution of the out-of-sample performance. More precisely, for each method the returned optimal solution  $x$  is evaluated on each of the test samples, i.e., we calculate the objective value  $x^\top p$  for each  $p \in \mathcal{T}$  and the distribution of these values is shown in the boxplots. Each value is an average over all 25 instances.

The results indicate that nearly all methods (except REG and the nominal solution) perform well for the upper quantiles, while for the lower quantiles only REG achieves good values. At the same time, REG exhibits a larger variance compared to the other methods.

The average runtime is below or around one second for all methods except GARO which has a runtime of around 6 seconds on average. We note, however, that the REG runtimes reflect the approximate formulation over finitely many training scenarios; solving the exact regret problem would be substantially more expensive due to its inherent nonconvexity.

For Gaussian samples with inverse mean-variance relationship we show the tradeoff plots in Figure 9.

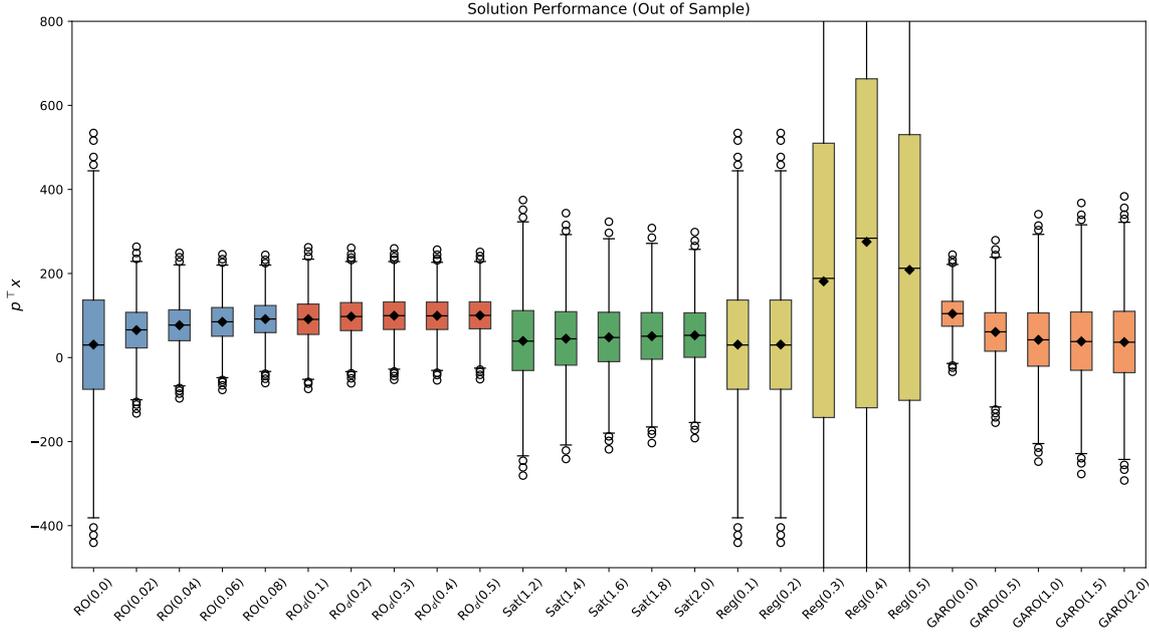


Figure 8: Boxplots of the out-of-sample objective values for the minimum knapsack problem with  $n = 50$  for Gaussian data. The diamonds denote the mean value.

The results show that the worst-case and 90%-quantile costs that GARO achieves are consistently close to the best performance across all methods. At the same time, the mean performance is even better than that of the nominal solution for  $q \geq 1$ . Meanwhile, SAT also achieves better mean performance than the nominal solution but is worse in terms of worst-case and 90%-quantile costs compared to GARO. RO and  $RO_d$  perform similarly to GARO in terms of worst-case and 90%-quantile costs but underperform in terms of mean performance. The reason GARO performs well here is that items in the knapsack problem with large mean costs have small variance and vice versa. For small uncertainty sets, items with small mean values are therefore beneficial; for larger uncertainty sets, however, these items become less attractive, since their variance eventually exceeds that of items with large means, which are not greatly affected by larger uncertainty sets owing to their small variance. We show the performance guarantees in Figure 15, and the out-of-sample boxplots in Figure 14 in the Appendix which show a similar picture as for classical Gaussian samples.

The runtimes show approximately the same picture as for classical Gaussian samples, but are slightly lower.

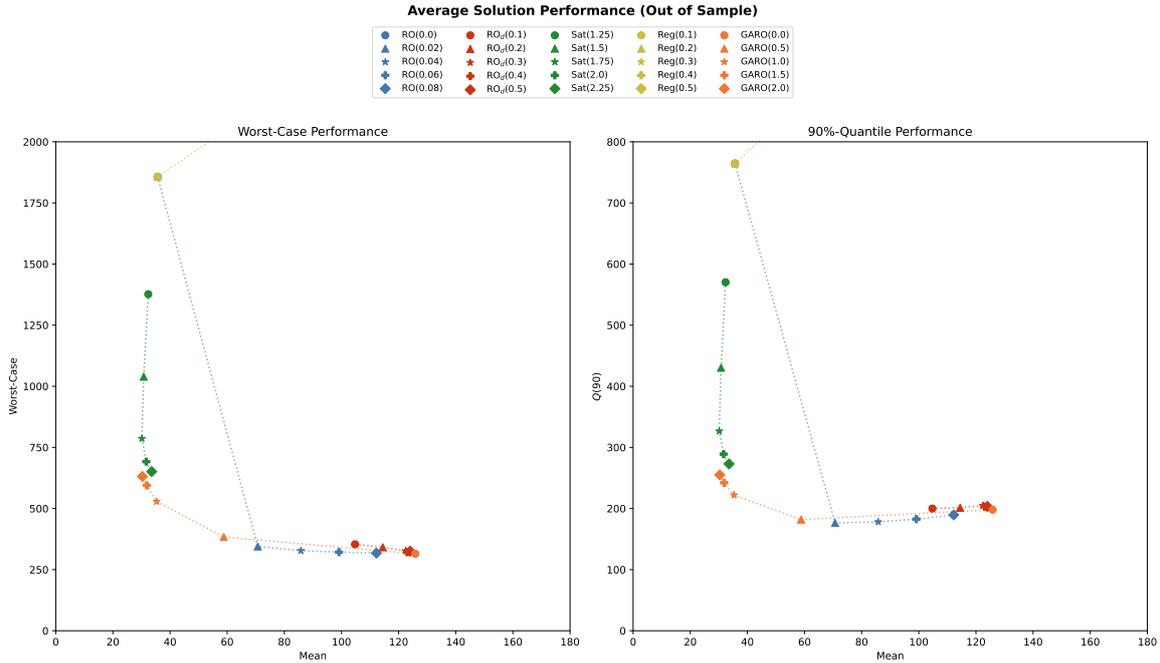


Figure 9: Out-of-sample performance for the minimum knapsack problem with  $n = 50$  for Gaussian data with inverse mean-variance relationship. Average vs. worst-case (left) and average vs. 90%-quantile (right).

#### 6.4. Heavy-Tail Data

For the heavy-tail samples we present the trade-off plots in Figure 10. The results indicate that RO, SAT and REG consistently achieve approximately the best worst-case performances, however with significant degradation in mean performance. The GARO method again yields solutions which dominate in terms of worst-case vs. mean trade-off. However, its solutions are more sensitive in terms of worst-case behavior, leading to larger worst-case values for  $q \geq 1$  compared to most of the benchmark methods, while achieving nearly the best worst-case value for  $q = 0$ . At the same time, GARO with  $q \geq 1$  achieves the best mean values together with the nominal solution. REG suffers most in terms of mean performance.

Interestingly, when considering the 90%-quantile cost, the trade-off curve changes. The 90%-quantile costs are significantly smaller than worst-case costs, owing to the properties of the Pareto distribution, which generates many samples with small values and a few with very large values. GARO for  $q \geq 0.5$  performs very well in terms of 90%-quantile cost while simultaneously achieving nearly the best mean behavior (only the nominal solution is better). The methods SAT, RO and REG do not achieve good 90%-quantile costs, performing even worse than the nominal solution and having significantly larger mean values than GARO. REG achieves the worst 90%-quantile costs followed by RO, SAT and  $RO_d$ .

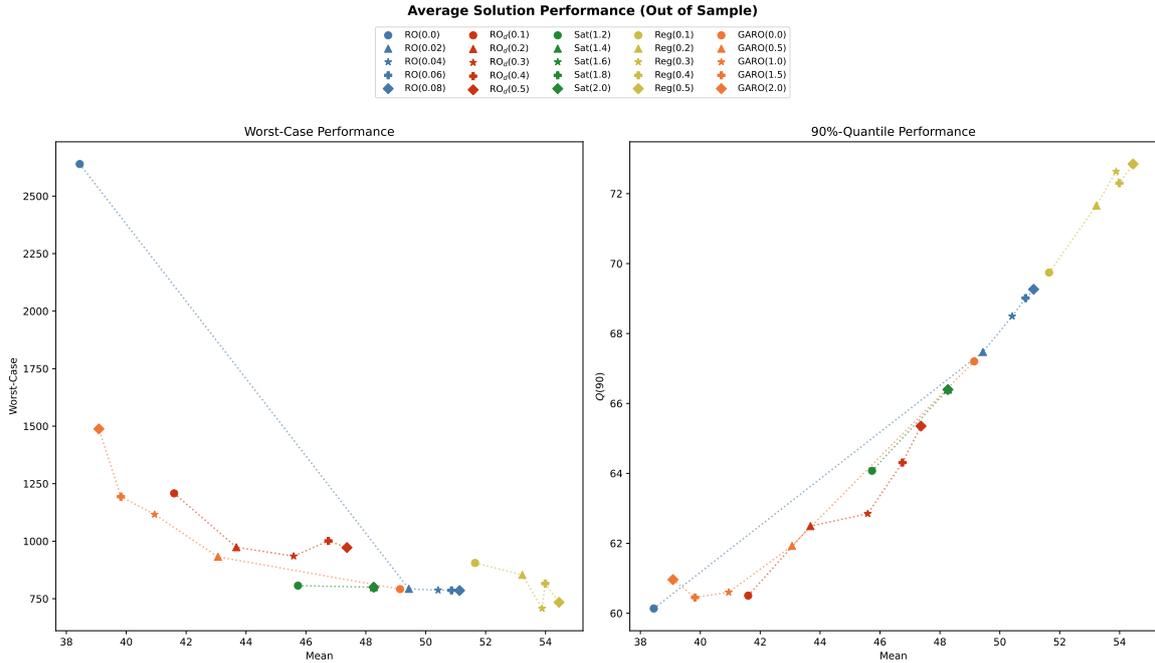


Figure 10: Out-of-sample performance for the minimum knapsack problem with  $n = 50$  for heavy-tail data. Average vs. worst-case (left) and average vs. 90%-quantile (right).

In Figure 11 we show the guarantees for each method as a function of  $\gamma$  for all test scenarios (right) and a zoomed-in view restricted to test scenarios with  $\gamma$  close to zero (left). Due to the heavy-tail distribution, most data points lie very close to  $p_0$  while a few lie far away. Note that what appears to be piecewise linear functions on the left is an artifact of the plot generation process, since we compute the function value on a grid of 100 evenly spaced  $\gamma$ -values connected by straight lines. The results show that GARO(0) and GARO(0.5) provide very good guarantees for heavy-tail data, both for scenarios close to  $p_0$  and those far from it. The previous trade-off results confirmed that GARO(0.5) also provided one of the best out-of-sample performances in terms of 90%-quantile cost. We show the out-of-sample boxplots in Figure 12.

The runtimes of RO,  $RO_d$ , and SAT are again below one second on average, while REG and GARO can have runtimes of up to 10 seconds where larger uncertainty sets lead to larger runtimes for REG. The runtime for GARO is stable over different  $q$ .

## 7. Conclusion

We introduced Globalized Adversarial Regret Optimization (GARO), a decision framework that controls adversarial regret uniformly across all prediction error levels. Unlike classical robust and regret

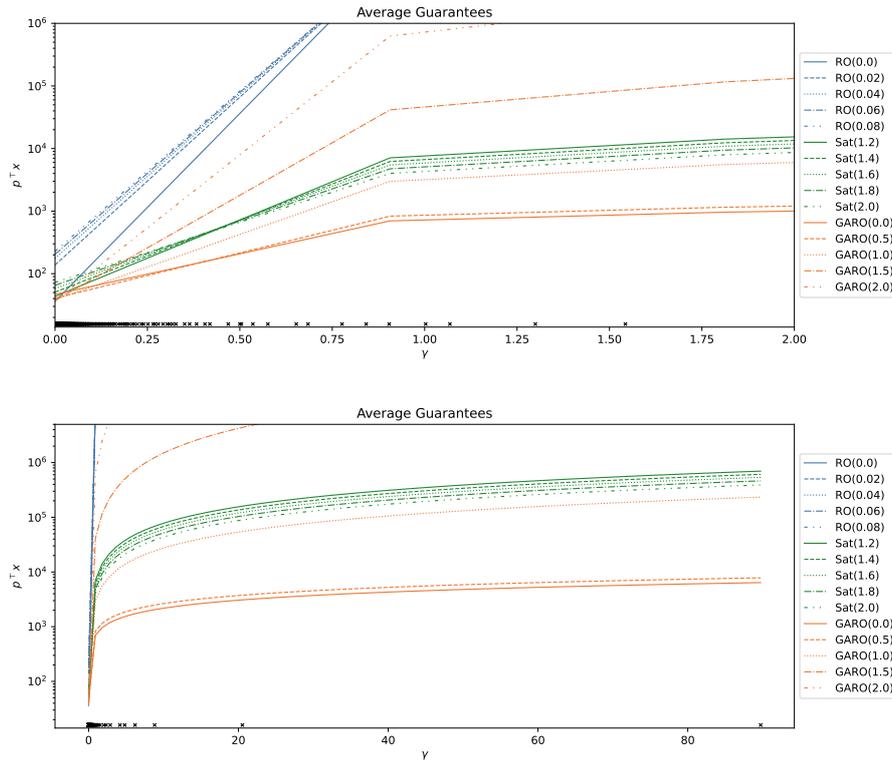


Figure 11: Performance guarantees for the minimum knapsack problem with  $n = 50$  for heavy-tail data. Zoomed view for small  $\gamma$  (top) and full range (bottom).

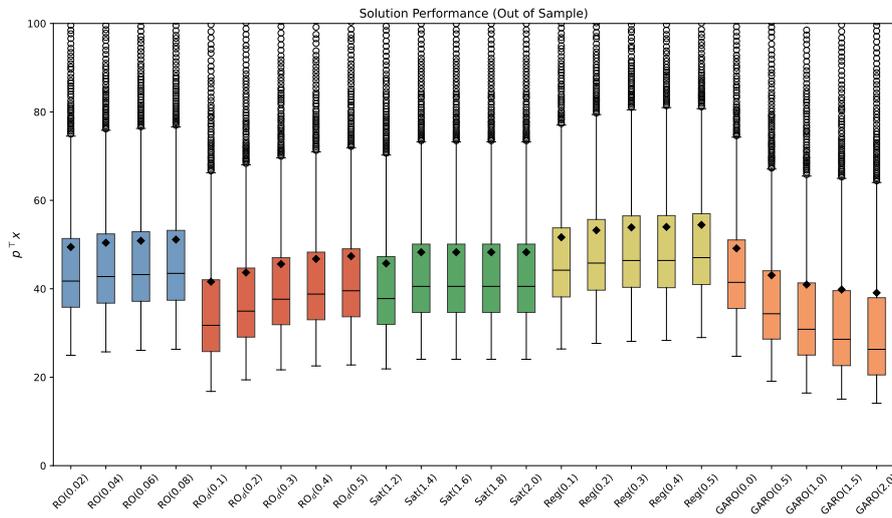


Figure 12: Boxplots for performance over all test scenarios for the knapsack problem with  $n = 50$  and heavy-tail data.

formulations, GARO does not require probabilistic calibration of the uncertainty set and provides meaningful performance guarantees even when predictions lack rigorous error bounds. We showed that satisficing formulations reduce to robust formulations under typical convexity assumptions, and that GARO with a relative rate function generalizes Lepski’s classical adaptation method to downstream decision problems. On the algorithmic side, we derived exact tractable reformulations for affine and linear-polyhedral settings and proposed a discretization and constraint-generation algorithm with convergence guarantees for the general case. Experiments on minimum knapsack problems show that GARO yields a favorable trade-off between worst-case and mean out-of-sample performance, and provides stronger global performance guarantees than existing methods.

In this work we focus on single-level problems where the uncertain parameters affect only the objective function. Extending the globalized adversarial regret perspective to multi-stage problems or to settings where uncertain parameters also affect the constraints is an interesting direction for future research. Furthermore, not much is known yet about the theoretical complexity of the problem, either in general or for specific problem structures.

**Acknowledgements** Bart P.G. van Parys gratefully acknowledges funding from NWO Vidi grant VI.Vidi.243.021.

## References

- Oleg V. Lepskii. Asymptotically minimax adaptive estimation. II. schemes without optimal adaptation: Adaptive estimators. *Theory of Probability & Its Applications*, 37(3):433–448, 1993.
- Paolo Toth and Daniele Vigo. *The Vehicle Routing Problem*. SIAM, 2002.
- Jayakrishnan Nair, Adam Wierman, and Bert Zwart. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*, volume 53. Cambridge University Press, 2022.
- Sujay Bhatt, Guanhua Fang, Ping Li, and Gennady Samorodnitsky. Nearly optimal Catoni’s M-estimator for infinite variance. In *International Conference on Machine Learning*, pages 1925–1944. PMLR, 2022.
- Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, pages 73–101, 1964.
- Peter J. Huber. *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1981.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539.
- Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016. doi: 10.1145/2939672.2939785.
- Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, et al. On the opportunities and risks of foundation models. Technical report, Stanford University, 2021. arXiv preprint arXiv:2108.07258.
- Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*, 2015. arXiv:1412.6572.
- Peter L. Bartlett, Dylan J. Foster, and Matus J. Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- Peter L. Bartlett and Shahar Mendelson. Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002.
- Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, 2009. doi: 10.1007/b13794.
- Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
- Matteo Fischetti and Michele Monaci. Light robustness. *Lecture Notes in Computer Science*, 5868:61–84, 2009.
- Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
- Leonard J. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46(253): 55–67, 1951.
- Panos Kouvelis and Gang Yu. *Robust discrete optimization and its applications*, volume 14. Springer Science & Business Media, 2013.

- Hassene Aissi, Cristina Bazgan, and Daniel Vanderpooten. Min–max and min–max regret versions of combinatorial optimization problems: A survey. *European journal of operational research*, 197(2):427–438, 2009.
- Daniel Zhuoyu Long, Melvyn Sim, and Minglong Zhou. Robust satisficing. *Operations Research*, 71(1):61–82, 2023.
- Aharon Ben-Tal, Ruud Brekelmans, Dick Den Hertog, and Jean-Philippe Vial. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing*, 29(2):350–366, 2017.
- Michael Mitzenmacher and Sergei Vassilvitskii. Algorithms with predictions. *Communications of the ACM*, 65(7):33–35, 2022.
- Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. In *Journal of the ACM*, volume 68, pages 1–25, 2021.
- Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- André Chassein and Marc Goerigk. A bicriteria approach to robust optimization. *Computers & Operations Research*, 66:181–189, 2016.
- Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.
- Aharon Ben-Tal and Arkadi Nemirovski. Robust optimization—methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.
- Ernst Roos and Dick den Hertog. Reducing conservatism in robust optimization. *INFORMS Journal on Computing*, 32(4):1109–1127, 2020.
- Abraham Wald. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, 46(2):265–280, 1945.
- Alekh Agarwal and Tong Zhang. Minimax regret optimization for robust machine learning under distribution shift. In *Conference on Learning Theory*, pages 2704–2729. PMLR, 2022.
- Georgia Perakis and Guillaume Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.
- Igor Averbakh and Vasilij Lebedev. On the complexity of minmax regret linear programming. *European Journal of Operational Research*, 160(1):227–231, 2005.
- Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
- Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *arXiv preprint arXiv:1505.05116*, 2015.
- Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. *Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning*, chapter 6, pages 130–166. INFORMS, 2019. doi: 10.1287/educ.2019.0198.

- Bart P. G. Van Parys, Peyman Mohajerin Esfahani, and Daniel Kuhn. From data to decisions: Distributionally robust optimization is optimal. *Management Science*, 67(6):3387–3402, 2021.
- Henry Lam. Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research*, 67(4):1090–1105, 2019.
- U.S. Nuclear Regulatory Commission. Probabilistic risk assessment (PRA). <https://www.nrc.gov/about-nrc/regulatory/risk-informed/pra>, n.d. Accessed: 2026-02-05.
- National Aeronautics and Space Administration. Probabilistic risk assessment procedures guide for NASA managers and practitioners. Technical Report NASA/SP-2011-3421, NASA, 2011.
- Ayush Jain, Alon Orlitsky, and Vaishakh Ravindrakumar. Robust estimation algorithms don't need to know the corruption level, 2022. arXiv preprint arXiv:2202.05453.
- Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization*, volume 6. Athena Scientific, 1997.
- Patrick L. Combettes and Christian L. Müller. Perspective functions: Properties, constructions, and examples. *Set-Valued and Variational Analysis*, 26(2):247–264, 2018.
- Aharon Ben-Tal, Dick Den Hertog, and Jean-Philippe Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1):265–299, 2015.
- Almir Mutapcic and Stephen Boyd. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3):381–406, 2009.
- Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.
- David G. Luenberger. *Optimization by Vector Space Methods*. Wiley, 1969.
- Claude Berge. *Topological Spaces*. Oliver and Boyd, Edinburgh, 1963.

## A. Supporting Results on the Facility Location Example

The Wasserstein distance is taken here as

$$W(\mathbb{P}, \mathbb{P}_0) := \inf \left\{ \int \|\xi - \xi'\| d\mathbb{T}(\xi, \xi') : \mathbb{T} \in \mathcal{P}(\mathbb{P}, \mathbb{P}_0) \right\}$$

where  $\mathcal{P}(\mathbb{P}, \mathbb{P}_0)$  denotes the set of all probability measures on  $\Xi \times \Xi$  with marginals  $\mathbb{P}$  and  $\mathbb{P}_0$ . The classical Kantorovich-Rubinstein duality theorem states that the Wasserstein distance can be characterized alternatively as

$$W(\mathbb{P}, \mathbb{P}_0) = \sup \left\{ \int \psi(\xi) d\mathbb{P}(\xi) - \int \psi(\xi) d\mathbb{P}_0(\xi) : \psi \text{ a } 1\text{-Lipschitz function} \right\}.$$

The following result states the classical dual characterization of Wasserstein distributional optimization.

**Lemma 2** (Blanchet and Murthy (2019)). *Let  $\int \|\xi - \mu\|^2 d\mathbb{P}_0(\xi) < \infty$ . Then, we have*

$$\max_{\mathbb{P} \in \mathcal{P}_\gamma} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) = \inf_{\lambda \geq 0} \lambda \gamma + \int \max_{\xi' \in \Xi} \|\xi' - \mu\|^2 - \lambda \|\xi' - \xi\| d\mathbb{P}_0(\xi).$$

**Lemma 3.** *We have*

$$v_{wc}^*(\gamma_0) \geq \max\left(0, 1 - \frac{\gamma_0}{2\|\Xi\|}\right) v(\mathbb{P}_0) + \min\left(\frac{\gamma_0}{2}, \|\Xi\|\right) \|\Xi\|. \quad (39)$$

*Proof of Lemma 3.* Let  $\gamma_0 > 2\|\Xi\|$ . Then as  $W(\mathbb{P}, \mathbb{P}_0) \leq \gamma_0$  for any  $\mathbb{P}$  it hence holds that

$$\max_{\mathbb{P} \in \mathcal{P}_{\gamma_0}} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) = \max_{\mathbb{P}} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) = \max_{\xi \in \Xi} \|\xi - \mu\|^2 \geq \|\Xi\|^2.$$

Let  $\gamma_0 \leq 2\|\Xi\|$  and take  $\Xi^* = \arg \max_{\xi \in \Xi} \|\xi - \mu\|$  which is a nonempty compact set by the extreme value theorem. Consider the parameter  $\alpha = \gamma_0 / (2 \max_{\xi \in \Xi} \|\xi - \mu\|) \leq \gamma_0 / (2\|\Xi\|) \leq 1$  and consider the following transport plan

$$\mathbb{T}_\alpha(\xi', \xi) = \mathbb{1}\{\xi = \xi'\} (1 - \alpha) + \alpha \mathbb{1}\left\{\xi' \in \arg \min_{\xi'' \in \Xi^*} \|\xi'' - \xi\|\right\}.$$

Informally, we transport from each location a fraction  $\alpha \in [0, 1]$  of the probability mass to a closest point in the set  $\Xi^*$ . Let us denote with  $\mathbb{P}_\alpha$  the associated pushforward probability measure. From the definition of the Wasserstein distance is trivial to observe that  $W(\mathbb{P}_\alpha, \mathbb{P}_0) \leq \max_{\xi \in \Xi} 2\|\xi - \mu\| \alpha \leq \gamma_0$ . Furthermore, observe that

$$\int \|\xi - \mu\|^2 d\mathbb{P}_\alpha(\xi) = (1 - \alpha) \int \|\xi - \mu\|^2 d\mathbb{P}_0(\xi) + \alpha \max_{\xi \in \Xi} \|\xi - \mu\|^2.$$

Hence,

$$\begin{aligned} \max_{\mathbb{P} \in \mathcal{P}_{\gamma_0}} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi) &\geq \int \|\xi - \mu\|^2 d\mathbb{P}_\alpha(\xi) \\ &\geq \left(1 - \frac{\gamma_0}{\max_{\xi \in \Xi} \|\xi - \mu\|}\right) \int \|\xi - \mu\|^2 d\mathbb{P}_0(\xi) + \frac{\gamma_0}{2} \max_{\xi \in \Xi} \|\xi - \mu\| \end{aligned}$$

from which the claim follows immediately.  $\square$

**Lemma 4.** *Let  $M(\gamma_0) := \{\mu_{nom}(\mathbb{P}) : \mathbb{P} \in \mathcal{P}_{\gamma_0}\}$  where  $\mu_{nom}(\mathbb{P}) := \arg \min_{\mu} \int \|\xi - \mu\|^2 d\mathbb{P}(\xi)$  is the unique minimizer. Then  $M(\gamma_0) \subseteq B[\mu_{nom}(\mathbb{P}_0), \gamma_0]$ .*

*Proof of Lemma 4.* Recall that the support function of the set  $M(\gamma_0)$  is defined as

$$h_{M(\gamma_0)}(c) := \max_{\mu} \left\{ c^\top \mu : \mu \in M(\gamma_0) \right\} = \|c\| \sup \left\{ \int \frac{c^\top \xi}{\|c\|} d\mathbb{P}(\xi) : \mathbb{P} \in \mathcal{P}_{\gamma_0} \right\}. \quad (40)$$

As the function  $c^\top \xi / \|c\|$  is 1-Lipschitz for any  $c$ , it follows from the Kantorovich-Rubinstein duality that

$$h_{M(\gamma_0)}(c) \leq c^\top \int \xi d\mathbb{P}_0(\xi) + \|c\| \gamma_0 = h_B(c) \quad (41)$$

where  $B$  denotes here a closed norm ball with center  $\mu_{nom}(\mathbb{P}_0) = \int \xi d\mathbb{P}_0(\xi)$  of radius  $\gamma_0$ . As the inequality holds for any  $c$ , it follows that  $M(\gamma_0) \subseteq \text{cl}(\text{conv}(M(\gamma_0))) \subseteq B$  from which the claim follows.  $\square$

## B. Supporting Results on Satisficing

**Proposition 2.** *Let  $f(x, p)$  be convex and lower semicontinuous in  $x$  and strictly concave and upper semicontinuous in  $p$  with  $X$  compact convex and  $P_\infty$  compact convex. Assume that  $p \mapsto d(p_0, p)$  is a convex lower semicontinuous distance function and (17) is feasible and satisfies condition (18). Then, we have*

$$X_{sat} \subseteq \cup_{\gamma \in [0, \gamma_{\max}]} X_{rob}(\gamma), \quad \gamma_{\max} := \max_{p \in P_\infty} d(p_0, p).$$

*Proof.* First, it is straightforward to observe that problem (17) is here equivalent to

$$\begin{aligned} \min_{x \in X, \alpha \geq 0} \quad & \alpha \\ \text{s.t.} \quad & (v_{wc}(x, \gamma) - f_0) / \gamma \leq \alpha \quad \forall \gamma \geq 0. \end{aligned}$$

As (17) admits a feasible point  $(x_s, \alpha_s)$ , the previous optimization problem satisfies the Slater constraint qualification condition at  $(x_s, \alpha_s + 1)$ . Furthermore, its minimum is attained as  $v_{wc}(x, \gamma)$  is lower semicontinuous in  $x$  for all  $\gamma \geq 0$  and  $X$  is a compact set and  $\alpha$  can be constrained without loss of generality to the compact set  $[0, \alpha_s]$ . Consider now an arbitrary minimizer  $(x_{sat}, \alpha_{sat})$ .

**Case  $\alpha_{sat} = 0$ .** We must have here that  $f_0 \geq \min_{x \in X} \max_{p \in P_\infty} f(x, p)$  for (17) to be feasible and hence from (18) we get  $f_0 = \min_{x \in X} \max_{p \in P_\infty} f(x, p)$  and hence  $x_{sat} \in X_{rob}(\gamma_{\max})$ .

**Case  $\alpha_{sat} > 0$ .** The constraint  $\alpha \geq 0$  is nonbinding and can be relaxed and hence (17) is here equivalent to

$$\begin{aligned} \min_{x \in X, \alpha \in \mathbb{R}} \quad & \alpha \\ \text{s.t.} \quad & (v_{wc}(x, \gamma) - f_0) / \gamma \leq \alpha \quad \forall \gamma \geq 0. \end{aligned}$$

We remark that as  $f(x, p)$  is convex in  $x$  this optimization problem is convex by construction. Introduce

the Lagrangian

$$L(x, \alpha, \mu) = \alpha + \int_{\mathbb{R}_+} \frac{(v_{wc}(x, \gamma) - f_0)}{\gamma} - \alpha d\mu(\gamma)$$

where  $\mu \in \mathcal{M}_+(\mathbb{R}_+)$  is a positive measure supported on  $\mathbb{R}_+$ . By Luenberger (1969, Theorem 8.3.1) Slater's constraint qualification implies that strong duality holds and the dual maximum is attained at some nonnegative measure  $\mu_{sat}$ . Combined with primal attainment of  $(x_{sat}, \alpha_{sat})$  Luenberger (1969, Theorem 8.4.1) gives that the triple  $(x_{sat}, \alpha_{sat}, \mu_{sat})$  is a saddle point of the Lagrangian  $L$ , i.e.,

$$\max_{\mu \in \mathcal{M}_+(\mathbb{R}_+)} L(x_{sat}, \alpha_{sat}, \mu) \leq L(x_{sat}, \alpha_{sat}, \mu_{sat}) = \alpha_{sat} \leq \min_{x \in X, \alpha \in \mathbb{R}} L(x, \alpha, \mu_{sat}).$$

We now indicate that  $\mu_{sat}$  must be a degenerate probability distribution. From the saddle point condition, we first observe that  $\alpha_{sat}$  minimizes  $L(x_{sat}, \alpha, \mu_{sat})$  over  $\alpha \in \mathbb{R}$ . Given that

$$L(x_{sat}, \alpha, \mu_{sat}) = \alpha \left( 1 - \int_{\mathbb{R}_+} d\mu_{sat}(\gamma) \right) + \int_{\mathbb{R}_+} \frac{(v_{wc}(x_{sat}, \gamma) - f_0)}{\gamma} d\mu_{sat}(\gamma)$$

we need that  $\int_{\mathbb{R}_+} d\mu_{sat}(\gamma) = 1$  and hence it follows that  $\mu_{sat}$  is a probability measure on  $\mathbb{R}_+$ . Second, we have that  $\alpha_{sat} = L(x_{sat}, \alpha_{sat}, \mu_{sat})$  which implies the complementarity condition

$$\int_{\mathbb{R}_+} \frac{(v_{wc}(x_{sat}, \gamma) - f_0)}{\gamma} - \alpha_{sat} d\mu_{sat}(\gamma) = 0.$$

Observe that as  $f(x, p)$  is concave in  $p$  and  $d(p_0, p)$  is convex in  $p$  it follows that  $v_{wc}(x, \gamma)$  is an increasing concave function for any  $x \in X$ . From primal feasibility it follows that  $(v_{wc}(x_{sat}, \gamma) - f_0)/\gamma - \alpha_{sat} \leq 0$  and hence  $\mu_{sat}$  must be supported on a convex set  $C = \{\gamma \geq 0 : v_{wc}(x_{sat}, \gamma) = f_0 + \alpha_{sat}\gamma\} = \{\gamma \geq 0 : v_{wc}(x_{sat}, \gamma) \geq f_0 + \alpha_{sat}\gamma\}$  where the worst-case function is affine and strictly increasing as we have here  $\alpha_{sat} > 0$ . Hence, we must have that  $C \subseteq [0, \max_{p \in P_\infty} d(p_0, p)]$  as for  $\gamma \geq \max_{p \in P_\infty} d(p_0, p)$  the worst-case function takes on a constant value  $v_{wc}(x_{sat}, \gamma) = \max_{p \in P_\infty} f(x_{sat}, p)$ .

However, the worst-case cost function  $\gamma \mapsto v_{wc}(x_{sat}, \gamma)$  is strictly concave following Lemma 5 and hence  $C = \{\gamma'_0\}$  must be a singleton. For the sake of contradiction, let the convex set  $C$  contain two distinct points  $\gamma_a$  and  $\gamma_b$ . Then for any  $t \in (0, 1)$  we have  $v_{wc}(x_{sat}, t\gamma_a + (1-t)\gamma_b) = f_0 + \alpha_{sat}(t\gamma_a + (1-t)\gamma_b) = t(f_0 + \alpha_{sat}\gamma_a) + (1-t)(f_0 + \alpha_{sat}\gamma_b) = tv_{wc}(x_{sat}, \gamma_a) + (1-t)v_{wc}(x_{sat}, \gamma_b)$  which contradicts the fact that  $v_{wc}(x_{sat}, \gamma)$  is strictly concave.

From the saddle point condition,  $x_{sat}$  finally minimizes  $L(x, \alpha_{sat}, \mu_{sat}) = \int_{\mathbb{R}_+} (v_{wc}(x, \gamma) - f_0)/\gamma d\mu_{sat}(\gamma) = (v_{wc}(x, \gamma'_0) - f_0)/\gamma'_0$  over  $x \in X$  and hence  $x_{sat} \in X_{rob}(\gamma'_0)$ .  $\square$

**Lemma 5.** *Let  $P$  be a compact convex set and  $p_0 \in P$ , let  $f(x, p)$  be strictly concave and upper*

semicontinuous in  $p$ , and let  $p \mapsto d(p_0, p)$  be a convex lower semicontinuous function. The function

$$\gamma \mapsto v_{wc}(x, \gamma) = \max \{f(x, p) : p \in P, d(p_0, p) \leq \gamma\}$$

is strictly concave on any set  $C$  where it is strictly increasing.

*Proof.* The feasible set  $F(\gamma) = \{p \in P : d(p_0, p) \leq \gamma\}$  is a nonempty, compact, and convex set for any  $\gamma \in C$ . Since  $f(x, p)$  is upper semicontinuous in  $p$ , the maximum defining  $v_{wc}(x, \gamma)$  is attained by the Weierstrass extreme value theorem, and since  $p \mapsto f(x, p)$  is strictly concave this maximizer  $p^*(\gamma)$  is also unique.

**Concavity.** Consider  $\gamma_1$  and  $\gamma_2$  and  $\lambda \in (0, 1)$ , and set  $\gamma_\lambda = \lambda\gamma_1 + (1 - \lambda)\gamma_2$ . Let  $p_i^* = p^*(\gamma_i)$  and  $p_\lambda = \lambda p_1^* + (1 - \lambda)p_2^*$ . Since  $P$  is convex,  $p_\lambda \in P$ . Since  $p \mapsto d(p_0, p)$  is convex,

$$d(p_0, p_\lambda) \leq \lambda d(p_0, p_1^*) + (1 - \lambda) d(p_0, p_2^*) \leq \lambda\gamma_1 + (1 - \lambda)\gamma_2 = \gamma_\lambda,$$

so  $p_\lambda \in F(\gamma_\lambda)$ . By feasibility and concavity of  $f$ ,

$$v_{wc}(x, \gamma_\lambda) \geq f(x, p_\lambda) \geq \lambda f(x, p_1^*) + (1 - \lambda) f(x, p_2^*) = \lambda v_{wc}(x, \gamma_1) + (1 - \lambda) v_{wc}(x, \gamma_2),$$

establishing concavity of  $v_{wc}$ .

**Strict concavity on  $C$ .** It suffices to show that for  $\gamma_1 < \gamma_2$  in  $C$ , the optimizers satisfy  $p_1^* \neq p_2^*$ , since then the strict concavity of  $f$  gives  $f(x, p_\lambda) > \lambda f(x, p_1^*) + (1 - \lambda) f(x, p_2^*)$ , and the argument above yields strict inequality  $v_{wc}(x, \gamma_\lambda) > \lambda v_{wc}(x, \gamma_1) + (1 - \lambda) v_{wc}(x, \gamma_2)$ . Suppose for contradiction that  $p_1^* = p_2^* = p^*$ . Then  $d(p_0, p^*) \leq \gamma_1$ , so  $p^*$  is feasible for every  $\gamma' \in [\gamma_1, \gamma_2]$ , giving  $v_{wc}(x, \gamma') \geq f(x, p^*) = v_{wc}(x, \gamma_1)$ . But  $p^*$  is also optimal at  $\gamma_2$ , so for any  $\gamma' \leq \gamma_2$  we have  $v_{wc}(x, \gamma') \leq v_{wc}(x, \gamma_2) = f(x, p^*) = v_{wc}(x, \gamma_1)$ . Hence  $\gamma \mapsto v_{wc}(x, \gamma)$  is constant on  $[\gamma_1, \gamma_2]$ , contradicting the assumption that  $\gamma \mapsto v_{wc}(x, \gamma)$  is strictly increasing on  $C$ .  $\square$

We remark that the assumption that the cost function  $f(x, p)$  is strictly concave in  $p$  can be relaxed, resulting in a slightly weaker result that there is always a satisficing solution which admits a robust interpretation. The following is a proof of Theorem 1, which proceeds via a regularization argument.

*Proof of Theorem 1.* For each  $\varepsilon > 0$  define

$$f^\varepsilon(x, p) := f(x, p) - \varepsilon \psi(p).$$

Since  $f(x, \cdot)$  is concave and  $-\varepsilon\psi$  is strictly concave,  $f^\varepsilon(x, \cdot)$  is strictly concave for every  $\varepsilon > 0$ . Define

$$v_{wc}^\varepsilon(x, \gamma) := \max_{p \in P_\gamma} f^\varepsilon(x, p), \quad v_{wc}(x, \gamma) := \max_{p \in P_\gamma} f(x, p).$$

**Step 1 (Uniform approximation and compactness).** Let  $M := \max_{p \in P_\infty} \psi(p) < \infty$ . Let  $p^\varepsilon \in \arg \max_{p \in P_\gamma} f^\varepsilon(x, p)$  and  $p^* \in \arg \max_{p \in P_\gamma} f(x, p)$ . Using optimality of  $p^\varepsilon$  for  $f^\varepsilon$  and  $p^*$  for  $f$  respectively:

$$\begin{aligned} v_{wc}^\varepsilon(x, \gamma) &= f(x, p^\varepsilon) - \varepsilon\psi(p^\varepsilon) \leq f(x, p^*) - \varepsilon\psi(p^\varepsilon) \leq v_{wc}(x, \gamma), \\ v_{wc}^\varepsilon(x, \gamma) &= f(x, p^\varepsilon) - \varepsilon\psi(p^\varepsilon) \geq f(x, p^*) - \varepsilon\psi(p^*) \geq v_{wc}(x, \gamma) - \varepsilon M, \end{aligned}$$

where the first inequality in the first line uses optimality of  $p^*$ , the second inequality in the first line uses  $\psi \geq 0$  and the first inequality in the second line uses optimality of  $p^\varepsilon$ . Hence

$$\max_{x \in X, \gamma \in [0, \gamma_{\max}]} |v_{wc}^\varepsilon(x, \gamma) - v_{wc}(x, \gamma)| \leq \varepsilon M \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Since  $v_{wc}^\varepsilon \leq v_{wc}$ , any  $(x, \alpha)$  feasible for SAT is also feasible for  $\text{SAT}^\varepsilon$ , so  $\alpha^\varepsilon \leq \alpha_{sat}$  for all  $\varepsilon > 0$ . By compactness of  $X \times [0, \alpha_{sat}] \times [0, \gamma_{\max}]$ , there exists a sequence  $\varepsilon_k \downarrow 0$  along which

$$(x^{\varepsilon_k}, \alpha^{\varepsilon_k}, \gamma^{\varepsilon_k}) \longrightarrow (\hat{x}, \hat{\alpha}, \hat{\gamma}) \in X \times [0, \alpha_{sat}] \times [0, \gamma_{\max}].$$

**Step 2 (Continuity of  $v_{wc}$ ).** We verify the conditions of Berge's maximum theorem (Berge 1963, p. 116) applied to  $v_{wc}(x, \gamma) = \max_{p \in P_\gamma} f(x, p)$  with parameter  $(x, \gamma) \in X \times [0, \gamma_{\max}]$ . First, each  $P_\gamma = \{p \in P_\infty : d(p_0, p) \leq \gamma\}$  is compact, as a closed sublevel set of the lower semicontinuous function  $d(p_0, \cdot)$  intersected with the compact set  $P_\infty$ , and nonempty since  $p_0 \in P_\infty$  satisfies  $d(p_0, p_0) = 0 \leq \gamma$ . The correspondence  $\gamma \mapsto P_\gamma$  is upper hemicontinuous: if  $\gamma_n \rightarrow \gamma$  and  $p_n \in P_{\gamma_n}$  with  $p_n \rightarrow p$  for  $n \rightarrow \infty$ , then lower semicontinuity of  $d$  gives  $d(p_0, p) \leq \liminf_{n \rightarrow \infty} d(p_0, p_n) \leq \liminf_{n \rightarrow \infty} \gamma_n = \gamma$ , so  $p \in P_\gamma$ . It is lower hemicontinuous: given  $p \in P_\gamma$  and  $\gamma_n \rightarrow \gamma$  for  $n \rightarrow \infty$ , set  $t_n = \min(1, \gamma_n/\gamma)$  (interpreting  $t_n = 1$  when  $\gamma = 0$ ) and  $p_n = t_n p + (1 - t_n) p_0 \in P_\infty$  by convexity of  $P_\infty$ . Convexity of  $d(p_0, \cdot)$  gives  $d(p_0, p_n) \leq t_n d(p_0, p) \leq t_n \gamma \leq \gamma_n$ , so  $p_n \in P_{\gamma_n}$ , and  $p_n \rightarrow p$  since  $t_n \rightarrow 1$ . Since  $f$  is jointly continuous, Berge's maximum theorem implies that  $v_{wc}(x, \gamma)$  is jointly continuous in  $(x, \gamma)$  on  $X \times [0, \gamma_{\max}]$ .

**Step 3 (Perturbed satisficing problem).** Since  $f^\varepsilon(x, \cdot)$  is strictly concave and upper semicontinuous, Proposition 2 applies to  $\text{SAT}^\varepsilon$  and yields an optimal solution  $(x^\varepsilon, \alpha^\varepsilon)$  with  $x^\varepsilon \in X_{rob}^\varepsilon(\gamma^\varepsilon)$  for some  $\gamma^\varepsilon \in [0, \gamma_{\max}]$ , i.e.,

$$v_{wc}^\varepsilon(x^\varepsilon, \gamma^\varepsilon) \leq v_{wc}^\varepsilon(x, \gamma^\varepsilon) \quad \forall x \in X.$$

**Step 4 ( $\hat{x}$  is a robust solution at  $\hat{\gamma}$ ).** Fix  $x \in X$ . For the right-hand side of the inequality in Step 3,

$$|v_{wc}^{\varepsilon_k}(x, \gamma^{\varepsilon_k}) - v_{wc}(x, \hat{\gamma})| \leq \underbrace{|v_{wc}^{\varepsilon_k}(x, \gamma^{\varepsilon_k}) - v_{wc}(x, \gamma^{\varepsilon_k})|}_{\leq \varepsilon_k M \rightarrow 0} + \underbrace{|v_{wc}(x, \gamma^{\varepsilon_k}) - v_{wc}(x, \hat{\gamma})|}_{\rightarrow 0 \text{ by Step 2}},$$

and for the left-hand side,

$$|v_{wc}^{\varepsilon_k}(x^{\varepsilon_k}, \gamma^{\varepsilon_k}) - v_{wc}(\hat{x}, \hat{\gamma})| \leq \underbrace{|v_{wc}^{\varepsilon_k}(x^{\varepsilon_k}, \gamma^{\varepsilon_k}) - v_{wc}(x^{\varepsilon_k}, \gamma^{\varepsilon_k})|}_{\leq \varepsilon_k M \rightarrow 0} + \underbrace{|v_{wc}(x^{\varepsilon_k}, \gamma^{\varepsilon_k}) - v_{wc}(\hat{x}, \hat{\gamma})|}_{\rightarrow 0 \text{ by Step 2}}.$$

Passing to the limit gives  $v_{wc}(\hat{x}, \hat{\gamma}) \leq v_{wc}(x, \hat{\gamma})$  for all  $x \in X$ , so  $\hat{x} \in X_{rob}(\hat{\gamma})$ .

**Step 5 ( $\hat{x}$  is feasible and optimal for SAT).** Since  $(x^{\varepsilon_k}, \alpha^{\varepsilon_k})$  is feasible for  $\text{SAT}^{\varepsilon_k}$ ,

$$v_{wc}^{\varepsilon_k}(x^{\varepsilon_k}, \gamma) \leq f_0 + \alpha^{\varepsilon_k} \gamma \quad \forall \gamma \in [0, \gamma_{\max}].$$

Fix  $\gamma \in [0, \gamma_{\max}]$ . Adding and subtracting  $v_{wc}(x^{\varepsilon_k}, \gamma)$ ,

$$\begin{aligned} v_{wc}(\hat{x}, \gamma) &\leq \underbrace{|v_{wc}(\hat{x}, \gamma) - v_{wc}(x^{\varepsilon_k}, \gamma)|}_{\rightarrow 0 \text{ by Step 2}} + \underbrace{|v_{wc}(x^{\varepsilon_k}, \gamma) - v_{wc}^{\varepsilon_k}(x^{\varepsilon_k}, \gamma)|}_{\leq \varepsilon_k M \rightarrow 0} + v_{wc}^{\varepsilon_k}(x^{\varepsilon_k}, \gamma) \\ &\leq |v_{wc}(\hat{x}, \gamma) - v_{wc}(x^{\varepsilon_k}, \gamma)| + \varepsilon_k M + f_0 + \alpha^{\varepsilon_k} \gamma. \end{aligned}$$

Passing to the limit as  $k \rightarrow \infty$  gives  $v_{wc}(\hat{x}, \gamma) \leq f_0 + \hat{\alpha} \gamma$  for all  $\gamma \in [0, \gamma_{\max}]$ , so  $(\hat{x}, \hat{\alpha})$  is feasible for SAT. Since  $\hat{\alpha} \leq \alpha_{sat}$  by Step 1 and  $(\hat{x}, \hat{\alpha})$  is feasible, optimality of  $\alpha_{sat}$  gives  $\hat{\alpha} = \alpha_{sat}$ , so  $\hat{x}$  is an optimal solution of SAT.  $\square$

The following examples illustrate that the assumptions in Proposition 2 and Theorem 1 are critical.

**Example 10.** Let here  $X = \{x^1 = (0, 2), x^2 = (1, 0), x^3 = (0.5, 1.7)\}$  (nonconvex),  $f(x, p) = p^\top x$ ,  $d(p_0, p) = \|p - p_0\|_2$ ,  $P = \mathbb{R}^2$  and  $p_0 = (1, 0)$ . The associated worst-case cost is

$$v_{wc}(x, \gamma) = \max_{d(p_0, p) \leq \gamma} p^\top x = p_0^\top x + \gamma \|x\|_2.$$

The worst-case costs are  $v(x^1, \gamma) = 2\gamma$ ,  $v(x^2, \gamma) = 1 + \gamma$  and  $v(x^3, \gamma) \approx 0.5 + 1.77\gamma$ . For  $0 \leq \gamma \leq 1$  solution  $x^1$  is robust optimal, while for  $1 \leq \gamma$  solution  $x^2$  is robust optimal; see Figure 13. In particular, note that  $x^3$  is always dominated in terms of worst-case cost by  $x^1$  or  $x^2$ .

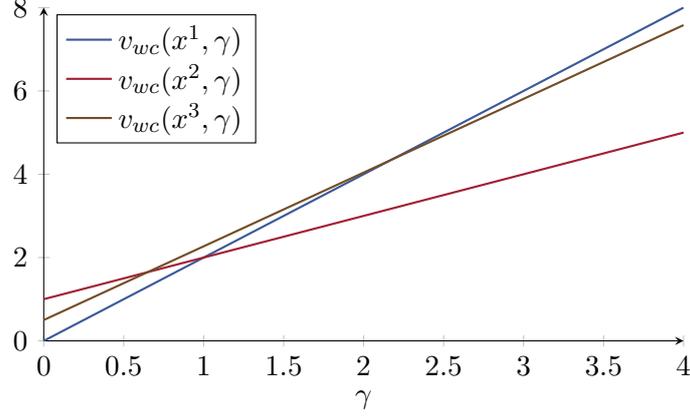


Figure 13: Worst-case costs for the three decisions in the nonconvex counterexample.

We consider a satisficing formulation

$$\begin{aligned} \min \quad & \alpha \\ \text{s.t.} \quad & \max_{d(p_0, p) \leq \gamma} p^\top x \leq f_0 + \alpha\gamma \quad \forall \gamma \geq 0 \\ & x \in X \end{aligned}$$

with satisfaction level  $f_0 = 1/2$ . Note that solution  $x^2$  is not feasible in the satisficing problem since  $p_0^\top x^2 = 1 > f_0$ . The optimal  $\alpha$  for  $x^3$  is  $\alpha^3 = \|x^3\|_2 \approx 1.77$  while for  $x^1$  it must hold  $\alpha^1 = \|x^1\|_2 = 2$ . Hence  $(x_{sat}, \alpha_{sat}) = (x^3, \|x^3\|)$  is the unique satisficing solution and yet  $x^3$  is not robustly optimal for any  $\gamma \geq 0$ .

**Example 11.** Let here  $X = [0, 2]$ ,  $f(x, p) = (x - 1)^2 + p$  (not strictly concave in  $p$ ),  $d(p_0, p) = |p - p_0|$ ,  $P = \mathbb{R}$  and  $p_0 = 0$ . The associated worst-case cost is

$$v_{wc}(x, \gamma) = (x - 1)^2 + \gamma$$

where clearly we have  $x_{rob}(\gamma) = 1$  for all  $\gamma \geq 0$  achieving oracle worst-case cost  $v_{wc}^*(\gamma) = \gamma$ . We consider now the satisficing formulation

$$\begin{aligned} \min \quad & \alpha \\ \text{s.t.} \quad & v_{wc}(x, \gamma) \leq f_0 + \alpha\gamma \quad \forall \gamma \geq 0 \\ & x \in X \end{aligned}$$

with satisfaction level  $f_0 = 1/4 > 0$  with solution set  $(X_{sat}, \alpha_{sat}) = (\{x : (x-1)^2 \leq f_0\}, 1)$  and hence there are satisficing solutions which are not robustly optimal.

**Proposition 3.** *Let  $\Gamma$  be compact and  $\gamma \mapsto v_{wc}(x, \gamma)$  be continuous for all  $x \in X$ . Suppose  $(x_{sat}, \alpha_{sat})$  is optimal in (17). Then  $x_{sat}$  is weakly minimal in (21).*

*Proof.* Suppose for contradiction that  $x_{sat}$  is not weakly minimal. Then there exists  $y \in X$  such that  $A(y, \gamma) < A(x_{sat}, \gamma)$  for all  $\gamma \in \Gamma$ . Adding  $v_{wc}^*(\gamma)$  to both sides gives

$$v_{wc}(y, \gamma) < v_{wc}(x_{sat}, \gamma) \leq f_0 + \alpha_{sat}\gamma \quad \forall \gamma \in \Gamma,$$

where the second inequality follows from the feasibility of  $(x_{sat}, \alpha_{sat})$  in (17). Since  $\Gamma$  is compact and  $\gamma \mapsto f_0 + \alpha_{sat}\gamma - v_{wc}(y, \gamma)$  is continuous and strictly positive on  $\Gamma$ , its minimum

$$\delta := \min_{\gamma \in \Gamma} (f_0 + \alpha_{sat}\gamma - v_{wc}(y, \gamma)) > 0$$

is attained by Weierstrass' extreme value theorem. Rearranging gives

$$v_{wc}(y, \gamma) \leq f_0 + \alpha_{sat}\gamma - \delta \leq f_0 + (\alpha_{sat} - \delta/\max \Gamma)\gamma \quad \forall \gamma \in \Gamma,$$

so  $(y, \alpha_{sat} - \delta/\max \Gamma)$  is feasible in (17), contradicting the optimality of  $(x_{sat}, \alpha_{sat})$ .

The assumption that  $\Gamma$  is compact is necessary: in Example 11 with  $\Gamma = \mathbb{R}_+$ , which is closed but not bounded (hence not compact), the adversarial regret simplifies to  $A(x, \gamma) = (x-1)^2$ , so the entire set of satisficing solutions  $\{x : (x-1)^2 \leq f_0\}$  fails to be weakly minimal, with the sole exception  $x_{sat} = 1$ .  $\square$

## C. Supporting Results on Algorithmic Aspects

**Lemma 6.** *We have that  $\gamma \mapsto v_{wc}(x, \gamma)$  is Lipschitz continuous with constant*

$$L = \sqrt{2}(\ell_{\max} - \ell_{\min}).$$

*Proof.* Define  $\psi(r) := v_{wc}(x, \sqrt{r})$  for  $r \geq 0$  and observe that it is concave and non-decreasing with  $\psi(0) = \mathbb{E}_{\mathbb{P}_0}[\ell(x, \xi)]$  (Van Parys et al. 2021). Pinsker's inequality gives  $\text{KL}(\mathbb{P}_0, \mathbb{P}) \leq r \implies \|\mathbb{P} - \mathbb{P}_0\|_{TV} \leq \sqrt{r/2}$ , where  $\|\cdot\|_{TV}$  denotes the total variation norm, so

$$\psi(r) - \psi(0) \leq (\ell_{\max} - \ell_{\min})\sqrt{r/2}. \tag{42}$$

For  $0 \leq \gamma_1 < \gamma_2$ , set  $r_i = \gamma_i^2$ . By concavity of  $\psi$ , the secant slope from  $r_1$  to  $r_2$  is at most the secant

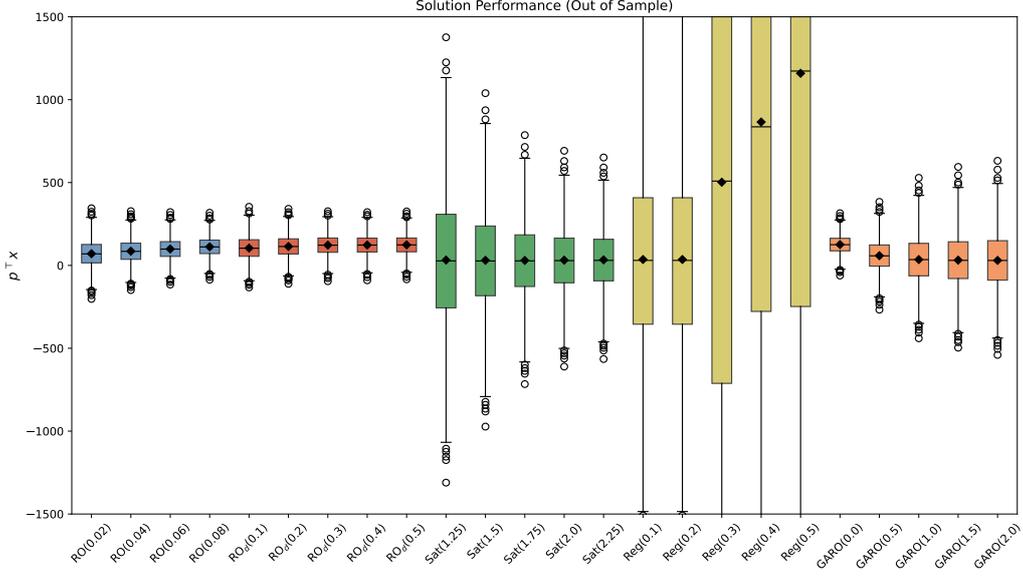


Figure 14: Boxplots for performance over all test scenarios for the knapsack problem with  $n = 50$  and Gaussian data with inverse mean-variance relationship. The diamonds denote the mean value.

slope from 0 to  $r_2$ :

$$\psi(r_2) - \psi(r_1) \leq \frac{r_2 - r_1}{r_2} (\psi(r_2) - \psi(0)) \leq \frac{(\gamma_2^2 - \gamma_1^2)(\ell_{\max} - \ell_{\min})}{\sqrt{2} \gamma_2}.$$

Using  $\gamma_2^2 - \gamma_1^2 = (\gamma_2 + \gamma_1)(\gamma_2 - \gamma_1) \leq 2\gamma_2(\gamma_2 - \gamma_1)$  yields

$$v_{wc}(x, \gamma_2) - v_{wc}(x, \gamma_1) \leq \sqrt{2} (\ell_{\max} - \ell_{\min})(\gamma_2 - \gamma_1).$$

□

## D. Supporting Results on Experiments

### D.1. Additional Results Experiment 1

In the following we present additional plots for the knapsack experiments.

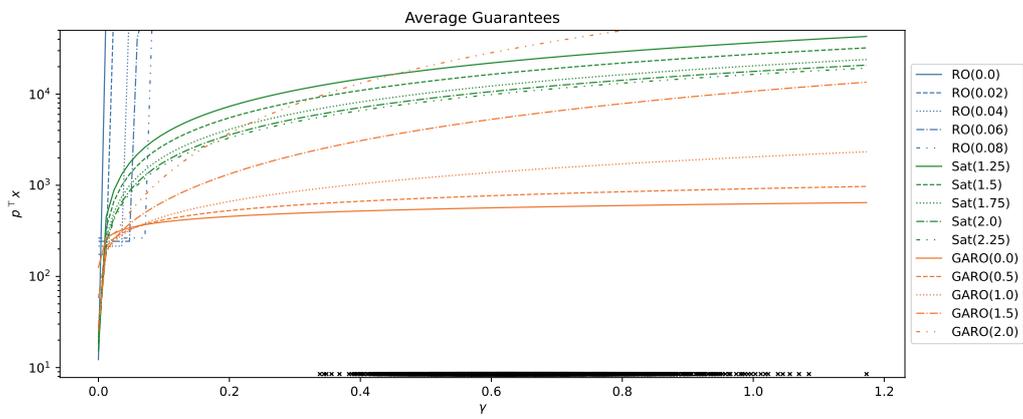


Figure 15: Performance guarantees for the minimum knapsack problem with  $n = 50$  for Gaussian data with inverse mean-variance relationship.