Overfitting in Adaptive Robust Optimization

Karl Zhu
Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA, USA
karlzhu@mit.edu

Dimitris Bertsimas
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA, USA
dbertsim@mit.edu

Abstract

Adaptive robust optimization (ARO) extends static robust optimization by allowing decisions to depend on the realized uncertainty — weakly dominating static solutions within the modeled uncertainty set. However, ARO makes previous constraints that were independent of uncertainty now dependent, making it vulnerable to additional infeasibilities when realizations fall outside the uncertainty set. This phenomenon of adaptive policies being brittle is analogous to overfitting in machine learning. To mitigate against this, we propose assigning constraint-specific uncertainty set sizes, with harder constraints given stronger probabilistic guarantees. Interpreted through the overfitting lens, this acts as regularization: tighter guarantees shrink adaptive coefficients to ensure stability, while looser ones preserve useful flexibility. This view motivates a principled approach to designing uncertainty sets that balances robustness and adaptivity.

1 Introduction

Robust optimization (RO) is one of the main frameworks (along with stochastic programming) for decision-making under uncertainty. RO produces a *robust* solution by requiring feasibility across all realizations within a specified uncertainty set $\mathcal{U}$. In this note, we consider a robust linear optimization problem with m robust constraints and uncertainty on the right-hand side (RHS):

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i + \boldsymbol{d}_i^\top \boldsymbol{u}, \quad \forall \boldsymbol{u} \in \mathcal{U}, \quad \forall i \in [m], \\ & \boldsymbol{x} \ge \boldsymbol{0}, \end{aligned}$$

where for certain parameters we have $\boldsymbol{c} \in \mathbb{R}^n$, and for each constraint i, we have $\boldsymbol{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, \boldsymbol{d}_i \in \mathbb{R}^p$. Finally, we have the perturbation parameter $\boldsymbol{u} \in \mathbb{R}^p$ which is uncertain but assumed to lie within $\mathcal{U}$. Here, $\boldsymbol{x}$ is fixed before $\boldsymbol{u}$ is realized, yielding a *static robust* solution.

Adaptive robust optimization (ARO) extends this framework by allowing some decisions to adjust after observing $\boldsymbol{u}$ [1]. A common tractable restriction is to use *affine decision rules*

$$\boldsymbol{x}(\boldsymbol{u}) = \boldsymbol{z} + \boldsymbol{V}\boldsymbol{u} \tag{1}$$

with decision variables $\boldsymbol{z} \in \mathbb{R}^n$ and $\boldsymbol{V} \in \mathbb{R}^{n\times p}$. Static solutions correspond to $\boldsymbol{V} = \boldsymbol{0}$, so adaptive policies always weakly dominate static ones within the set.

The size and geometry of $\mathcal{U}$ are chosen to balance feasibility and optimality. Common choices include the box (ℓ_∞ norm) [2], ellipsoid (ℓ_2 norm) [3], and budget sets (intersection of ℓ_∞ and ℓ_1 norms) [4]. A well-chosen uncertainty set generally should not contain all possible realizations of $\boldsymbol{u}$, as this would be overly conservative. Therefore, when evaluating solutions in simulation, one should not draw samples directly from the uncertainty set, but rather from an underlying distribution.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

A key, under-discussed drawback of ARO is that constraints that were originally independent of uncertainty become dependent. While static robust solutions often remain feasible outside the modeled set, adaptive policies may fail catastrophicallyfor example, violating variable non-negativity constraints. In this sense, adaptive policies are *brittle*: they achieve superior performance within $\mathcal{U}$ but generalize poorly outside it. This brittleness is directly analogous to *overfitting* in machine learning, where models with higher flexibility fit training data well but perform poorly out-of-sample.

This perspective motivates a rethinking of how uncertainty sets are specified in ARO. Not all constraints are equally critical: hard constraints demand stronger guarantees, while softer constraints may tolerate limited violations. Interpreted through the overfitting lens, this leads to a *regularization view*: tighter guarantees shrink adaptive coefficients to improve stability, while looser guarantees preserve flexibility. Although the equivalence of robustness and regularization has been studied extensively [5], [6], to our knowledge, there have been no similar discussions of this insight for ARO. In this note we illustrate brittleness using a renewable generation toy example, propose constraint-dependent uncertainty set sizes as a principled remedy for this brittleness, and show how robust counterparts (RC) impose regularization on adaptive coefficients.

2 Brittleness of Adaptive Policies

We illustrate brittleness with a toy model of production planning under renewables:

$$\begin{aligned}
\max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{s}} \quad & x_1 + x_2 + y_1 + y_2 - 100s_1 - 100s_2 \\
\text{s.t.} \quad & x_1 + y_1 - s_1 \leq 2 + u_1, \quad \forall (u_1, u_2) \in \mathcal{U}, \\
& x_2 + y_2 - s_2 \leq 2 + u_2, \quad \forall (u_1, u_2) \in \mathcal{U}, \\
& x_1, x_2, y_1, y_2, s_1, s_2 \geq 0.
\end{aligned}$$

Here a forecast of 2 units of renewable power is available in each period. Decisions x_1, y_1 allocate this supply in period 1, and x_2, y_2 in period 2. Shortfalls are covered by grid imports s_1, s_2, which are costly and non-renewable. Uncertainty in renewable availability is represented by a budget uncertainty set

$$\mathcal{U} = \{(u_1, u_2) : \|\boldsymbol{u}\|_\infty \leq \rho,\ \|\boldsymbol{u}\|_1 \leq \Gamma\}.$$

An optimal solution to the nominal problem (where $\mathcal{U} = \{\boldsymbol{0}\}$) is $x_i = y_i = 1,\ s_i = 0$ for $i = 1, 2$ with an objective value of 4, but any shortfall in renewables forces costly grid usage.

An optimal solution to the static robust problem with $(\rho, \Gamma) = (1, 1)$ is $x_i = 1, y_i = s_i = 0$, for $i = 1, 2$, with an objective value of 2. Notice that although the l_1 ball that defines $\mathcal{U}$ *couples* the uncertainties across constraints sets in $\mathcal{U}$, the static robust problem in general cannot exploit this coupling [7], giving an overly conservative solution that is equivalent to the uncertainty set being a box $\mathcal{U} = \{(u_1, u_2) : \|\boldsymbol{u}\|_\infty \leq \rho\}$.

In contrast, ARO is able to exploit this coupling [8]. Suppose we introduce a restricted affine decision rules that made the flows adapt to the current period of uncertainty:

$$x_i(u_i) = z_i^x + V_{i,i}^x u_i, \quad y_i(u_i) = z_i^y + V_{i,i}^y u_i, \quad i = 1, 2.$$

An optimal adaptive policy with $(\rho, \Gamma) = (1, 1)$ is $x_i(u_i) = 1,\ y_i(u_i) = 1 + u_i$, for $i = 1, 2$ with an objective value of 3 – significantly outperforming the static policy within $\mathcal{U}$. However, the adaptive policy is brittle: the non-negativity constraints, originally independent of u, now depend on u and may be violated. For instance, if $u_1 < -1$, then $y_1(u_1) < 0$, which is physically infeasible. Shortages can be met with imports, but negative flows cannot be implemented.

A natural remedy is to assign larger uncertainty radii to hard constraints (e.g., non-negativity) and smaller ones to softer constraints (e.g., grid imports). For example, with $(\rho, \Gamma) = (2, 2)$ on the non-negativity constraints only, an optimal adaptive solution is $x_i = y_i = 1 + \frac{1}{2}u_i$, for $i = 1, 2$ and maintains an objective value of 3 and remains feasible for all $u \geq -2$. In our example, only a non-negative RHS of renewable supply is physically sensible, so u_1, u_2 naturally both have a lower bound of -2. So the enlarged set recovers the guaranteed feasibility previously enjoyed by the static policy while retaining the benefits from ARO and uncertainty coupling.

3 Probabilistic and Deterministic Guarantees

From the toy example we saw that a uniform uncertainty set $\mathcal{U}$ is not appropriate in ARO, since many constraints that were previously independent of uncertainty become dependent once adaptivity is introduced. This motivates the use of *constraint-specific uncertainty sets* $\mathcal{U}_i$ for each constraint $i \in [m]$, sized according to its criticality. We distinguish two types:

- **Hard constraints** (e.g. flow non-negativity): must be satisfied in all realizations. These require deterministic guarantees, with uncertainty sets that fully cover the support of $\boldsymbol{u}$.
- **Softer constraints** (e.g., renewable allocations with non-renewable grid imports): can tolerate limited violations if backup resources exist. For these, probabilistic guarantees are appropriate, where ellipsoidal or budget sets provide explicit guarantees while preserving ARO's flexibility.

We now provide quantitative prescriptions. First, we recall probabilistic guarantees under Gaussian and distribution-free assumptions from the RO literature [9], suited for softer constraints. Then, we present deterministic RCs for bounded and semi-bounded supports, which are appropriate for hard constraints that must hold almost surely.

Probabilistic guarantees. For each constraint $i \in [m]$, we may require

$$\mathbb{P}\Big[\boldsymbol{a}_i^\top \boldsymbol{x}(\boldsymbol{u}) \ \le \ b_i + \boldsymbol{d}_i^\top \boldsymbol{u}\Big] \ \ge \ 1 - \varepsilon. \tag{2}$$

Different uncertainty sets $\mathcal{U}_i$ provide different sufficient conditions for (2).

- **Gaussian case.** If $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{x}(\boldsymbol{u})$ is affinely adaptive (1), then (2) holds iff

  $$\boldsymbol{a}_i^\top \boldsymbol{z} + (\boldsymbol{V}^\top \boldsymbol{a}_i - \boldsymbol{d}_i)^\top \boldsymbol{\mu} + \rho_{1-\varepsilon} \left\|\Sigma^{1/2}(\boldsymbol{V}^\top \boldsymbol{a}_i - \boldsymbol{d}_i)\right\|_2 \ \le \ b_i,$$

  where $\rho_{1-\varepsilon}$ is the $(1-\varepsilon)$-quantile of the standard normal. Equivalently, this is the RC with ellipsoidal uncertainty set

  $$\mathcal{U}_i = \left\{\boldsymbol{u} : \ \left\|\Sigma^{-1/2}(\boldsymbol{u} - \boldsymbol{\mu})\right\|_2 \le \rho_{1-\varepsilon}\right\}.$$

- **Distribution-free case.** If $u_1, \dots, u_p$ are independent, zero-mean, with support $[-1, 1]$, then:
  - *Ball-box uncertainty set.* For $\mathcal{U}_i = \{\boldsymbol{u} : \|\boldsymbol{u}\|_\infty \le 1, \ \|\boldsymbol{u}\|_2 \le \rho\}$, the RC is

    $$\boldsymbol{a}_i^\top \boldsymbol{z} + \rho \, \|\boldsymbol{V}^\top \boldsymbol{a}_i - \boldsymbol{d}_i - \boldsymbol{y}_i\|_2 + \|\boldsymbol{y}_i\|_1 \ \le \ b_i, \quad \boldsymbol{y}_i \in \mathbb{R}^p, \tag{3}$$

    which ensures $\mathbb{P}[\boldsymbol{a}_i^\top \boldsymbol{x}(\boldsymbol{u}) \le b_i + \boldsymbol{d}_i^\top \boldsymbol{u}] \ge 1 - \exp(-\rho^2/2)$. Choosing $\rho = \sqrt{2\ln(1/\varepsilon)}$ yields (2).
  - *Budget uncertainty set.* For $\mathcal{U}_i = \{\boldsymbol{u} : \|\boldsymbol{u}\|_\infty \le 1, \ \|\boldsymbol{u}\|_1 \le \Gamma\}$, the RC is

    $$\boldsymbol{a}_i^\top \boldsymbol{z} + \Gamma \, \|\boldsymbol{V}^\top \boldsymbol{a}_i - \boldsymbol{d}_i - \boldsymbol{y}_i\|_\infty + \|\boldsymbol{y}_i\|_1 \ \le \ b_i, \quad \boldsymbol{y}_i \in \mathbb{R}^p,$$

    which ensures $\mathbb{P}[\boldsymbol{a}_i^\top \boldsymbol{x}(\boldsymbol{u}) \ \le \ b_i + \boldsymbol{d}_i^\top \boldsymbol{u}] \ \ge \ 1 - \exp(-\Gamma^2/(2p))$. Choosing $\Gamma = \sqrt{2\ln(1/\varepsilon)}\sqrt{p}$ yields (2).

Deterministic guarantees. For hard constraints that must hold without exception, probabilistic guarantees are insufficient. In these cases, one must construct an uncertainty set that fully contains the support of $\boldsymbol{u}$ and enforce feasibility deterministically. This leads to robust counterparts over polyhedral box uncertainty sets, whose derivation follows standard duality arguments (see [10] for a tutorial). We state the results directly below.

Proposition 1 (Bounded support)**.** *Suppose u_k has support $[L_k, U_k]$ $\forall k \in [p]$, and $\boldsymbol{x}(\boldsymbol{u})$ is affinely adaptive (1). Then the following RC satisfies $\boldsymbol{a}_i^\top \boldsymbol{x}(\boldsymbol{u}) \le b_i + \boldsymbol{d}_i^\top \boldsymbol{u}$ w.p. 1:*

$$\boldsymbol{a}_i^\top \boldsymbol{z} + \boldsymbol{U}^\top \boldsymbol{\beta} - \boldsymbol{L}^\top \boldsymbol{\alpha} \ \le b_i, \quad \boldsymbol{\beta} - \boldsymbol{\alpha} = \boldsymbol{V}^\top \boldsymbol{a}_i - \boldsymbol{d}_i, \quad \boldsymbol{\alpha}, \boldsymbol{\beta} \ge \boldsymbol{0},$$

where $\boldsymbol{L} = (L_k)_{k=1}^p, \boldsymbol{U} = (U_k)_{k=1}^p$ and $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^p$.

Corollary 1 (Semi-bounded support)**.** *Under the same setup as Proposition 1, but instead if u_k is unbounded from below (resp. above), for each $k \in [p]$, the RC from Proposition 1 along with $\alpha_k = 0$ (resp. $\beta_k = 0$) satisfies $\boldsymbol{a}_i^\top \boldsymbol{x}(\boldsymbol{u}) \le b_i + \boldsymbol{d}_i^\top \boldsymbol{u}$ w.p. 1.*

4 A Regularization Perspective

We now interpret robust constraints through the lens of *regularization*, providing a mitigation for the brittleness of ARO in line with the overfitting analogy. Notice the RCs can be rearranged into explicit norm bounds on the adaptive coefficients. For example, with $\boldsymbol{y}_i = \mathbf{0}$, (3) can be rewritten as

$$\|\boldsymbol{V}^\top \boldsymbol{a}_i - \boldsymbol{d}_i\|_2 \;\leq\; \frac{\boldsymbol{b}_i - \boldsymbol{a}_i^\top \boldsymbol{z}}{\rho}, \qquad \rho > 0.$$

For illustration, we set $\boldsymbol{d}_i = \mathbf{0}$ and normalize $\boldsymbol{b}_i - \boldsymbol{a}_i^\top \boldsymbol{z} = 1$, and plot the resulting upper bound on $\|\boldsymbol{V}^\top \boldsymbol{a}_i\|_2$ against the probabilistic guarantee $1 - \varepsilon$. Figure 1 compares the bounds derived under a Gaussian assumption with those obtained from the distribution-free guarantee.

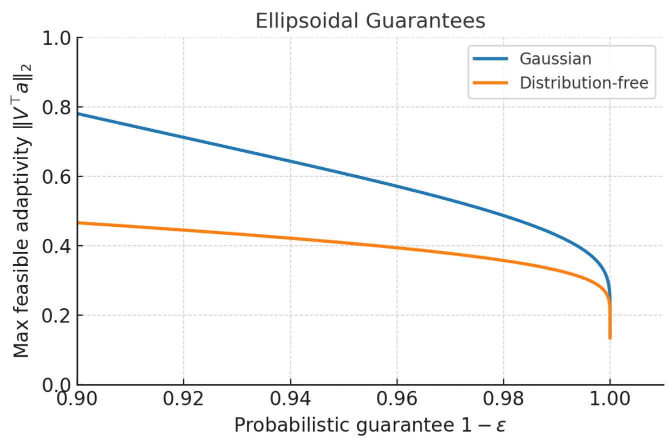

Figure 1: Ellipsoidal guarantees: maximum feasible adaptivity $\|\boldsymbol{V}^\top \boldsymbol{a}\|_2$ as a function of the probabilistic guarantee $1 - \varepsilon$. Under Gaussian assumptions, the underlying probabilistic bound is tighter, yielding less conservative feasibility regions. The distribution-free guarantee is looser, leading to more conservative regularization. In both cases, higher guarantees correspond to stronger regularization, shrinking adaptive flexibility and preventing brittle solutions.

In this view, the probabilistic guarantee $1 - \varepsilon$ serves as a *regularization parameter*: requiring stronger guarantees (smaller ε) tightens the constraint on $\boldsymbol{V}$, limiting adaptivity. In the limit as $\varepsilon \to 0$, we recover the static policy with $\boldsymbol{V} = \mathbf{0}$, eliminating brittleness. The resulting ℓ_2-norm bound directly parallels ridge regression in machine learning, where stronger regularization shrinks coefficients to improve stability. This mirrors the bias–variance tradeoff: greater adaptivity yields better in-set performance but poorer generalization, while stronger regularization reduces flexibility but stabilizes out-of-set behavior.

5 Discussion and Conclusion

Our analysis suggests several lessons for modeling with ARO.

- **Simulate beyond the set.** Evaluate performance on distributions whose support exceeds the modeled uncertainty set, just as machine learning models must be tested out-of-sample.
- **Differentiate constraints.** Treat some as *soft*, where limited violations are tolerable (e.g., supply–demand balance), and others as *hard*, where violations are catastrophic (e.g., flow non-negativity).
- **Use constraint-dependent uncertainty set sizes.** Assign larger sets to hard constraints, effectively regularizing their adaptive coefficients more strongly, while smaller sizes suffice for softer ones.

Overall, ARO provides flexibility but can overfit to the specified uncertainty sets, producing brittle out-of-set behavior. Interpreting robust counterparts as implicit regularization motivates constraint-dependent uncertainty sets as a principled way to balance adaptivity and stability.

References

[1] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, March 2004.

[2] A. L. Soyster. Technical NoteConvex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming. *Operations Research*, 21(5):1154–1157, October 1973.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, August 1999.

[4] Dimitris Bertsimas and Melvyn Sim. The Price of Robustness. *Operations Research*, 52(1):35–53, February 2004.

[5] Huan Xu, Constantine Caramanis, Shie Mannor, Shie Mannor, and Mcgill Ca. Robustness and Regularization of Support Vector Machines.

[6] Dimitris Bertsimas and Martin S. Copenhaver. Characterization of the equivalence of robustification and regularization in linear and matrix regression. *European Journal of Operational Research*, 270(3):931–942, November 2018.

[7] Ahmadreza Marandi and Dick den Hertog. When are static and adjustable robust optimization problems with constraint-wise uncertainty equivalent? *Mathematical Programming*, 170(2):555–568, 2018.

[8] Dimitris Bertsimas, Liangyuan Na, Bartolomeo Stellato, and Irina Wang. The Benefit of Uncertainty Coupling in Robust and Adaptive Robust Optimization. *INFORMS Journal on Optimization*, page ijoo.2023.0007, December 2024.

[9] Dimitris Bertsimas, Dick Den Hertog, and Jean Pauphilet. Probabilistic Guarantees in Robust Optimization. *SIAM Journal on Optimization*, 31(4):2893–2920, January 2021.

[10] Dimitris Bertsimas and Aurélie Thiele. Robust and Data-Driven Optimization: Modern Decision Making Under Uncertainty. In *Models, Methods, and Applications for Innovative Decision Making*, pages 95–122. INFORMS, September 2006.