Project Report: Optimization Under Uncertainty (0960335)

Student : JORNET Jeremy - 931215248

Joint Estimation and Robustness Optimization

Taozeng Zhu, Jingui Xie, Melvyn Sim

I. INTRODUCTION

In practice, optimization problems often rely on input parameters such as costs, demands, rates or capacities that are not known exactly but could be estimated from data. Those input parameters are estimated from imprecise data, which inevitably induces statistical errors in parameters that are nevertheless crucial for optimization problems and the resulting decision making. Classical Robust Optimization frameworks mitigate such risks through specified uncertainty sets. However, the size and shape of these sets is difficult to determine in practical terms, since they are neither interpretable nor directly linked to the estimation procedure, yet it critically determines the robustness of the solution. This raises the fundamental problem of how to systematically account for uncertainty in optimization models without relying on arbitrary uncertainty set sizes, and how to integrate the estimation procedure itself into the robustness analysis. Zhu et al. (2020) propose the *Joint Estimation and Robustness Optimization* framework which directly addresses this issue by integrating parameter estimation metric in the uncertainty set formulation and by automatically deriving its radius r.

The paper's problem is connected to earlier contributions within the literature. Bertsimas et al. (2018) extend the robust optimization paradigm by designing data-driven uncertainty sets, in particular through the use of statistical confidence intervals. Instead of choosing an arbitrary radius r, the size of the uncertainty set is calibrated by the variability of the data

Link to the article : https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2020.3898?journalCode=mnsc and the statistical properties of the estimators. This approach eliminates the need for subjective choices and provides a probabilistic interpretation of robustness: the confidence level (often 95%) quantifies the likelihood that the true parameters lie within the uncertainty set. Moreover, the estimation is integrated into the optimization model, as the uncertainty set reflects the precision of the chosen estimator. This way, the framework ties robustness directly to estimation accuracy and offers an interpretable protection against estimation errors than traditional formulations with arbitrary set radius.

In the DRO framework, several methodologies exist to determine the radius r automatically. In this paradigm, the ambiguity set is defined as a neighborhood of probability distributions around the empirical distribution. Ben-Tal et al. (2013) construct such ambiguity sets using ϕ-divergences, where the radius r quantifies the maximum divergence allowed between the empirical and the true distribution. The quantity r is not chosen arbitrarily but can be calibrated through concentration inequalities, thus tying it to the sample size and confidence level. Similarly, Esfahani and Kuhn (2018) propose Wasserstein-metric based DRO, where the ambiguity set is a Wasserstein ball centered at the empirical distribution. Here again, the radius r is determined from measure concentration results, ensuring that the true distribution lies within the Wasserstein ball with high probability. In both cases, the size of the uncertainty set is no longer an arbitrary choice but is derived from statistical principles.

II. Contribution and Main Result

A. *Contribution*

The main contribution of the paper is the development of the *Joint Estimation and Robustness Optimization (JERO)* framework, which integrates the parameter estimation procedure directly into the optimization problem and derives robustness as part of the model. Specifically, the novelty of the JERO framework lies in the model design itself. It constructs the uncertainty set using the estimation metric of the chosen procedure (OLS, LASSO, MLE) and jointly optimizes the decision variable and the uncertainty radius r. Moreover, in contrast to traditional approaches, JERO shifts the input choice to a goal cost, a task that is considerably more intuitive and natural.

The method searches for the largest radius r that remains consistent with this goal cost, thereby maximizing robustness.

In addition, the paper provides explicit JERO formulations for several standard estimation procedures. It further demonstrates the applicability of the framework through a real-world case study on health insurance reimbursement. Finally, the paper addresses the computational complexity of JERO applications by developing tractable reformulations, notably via second-order conic approximations, which enable efficient solution with state-of-the-art solvers.

B. Main results

In this section, we first focus on the theoretical main result of the paper, namely the JERO formulation and the proposed methodology to solve it. We then illustrate the framework flow with two explicit formulations corresponding to standard estimation procedures: Linear Regression under normal errors and Poisson Regression, which will also serve as the basis for the application in the next section. Secondly, we present the numerical experiment conducted in the paper and discuss its main results.

*1) **Theoretical results***

: **Starting Problem.** We will begin with the following classic optimization problem :

$$\begin{aligned} Z = \min_{x \in \mathcal{X}} \; & a_0(x, \hat{\beta}) \\ \text{s.t.} \;\; & a_i(x, \hat{\beta}) \leq \tau_i, \quad \forall i \in [I], \end{aligned} \tag{1}$$

$x \in \mathcal{X} \subseteq R^N$ is the decision variable, $\hat{\beta} \in \mathcal{W} \subseteq R^M$ is the input parameter. The model assume that the $a_i : \bar{\mathcal{X}} \times \bar{\mathcal{W}} \mapsto R$ are *saddle functions* on the domain $\bar{\mathcal{X}} \times \bar{\mathcal{W}} \subseteq R^N \times R^M$ (note $\mathcal{X} \subseteq \bar{\mathcal{X}},\ \mathcal{W} \subseteq \bar{\mathcal{W}}$).

The classical robust optimization approach redefines (1) into the following problem:

$$\begin{aligned} Z_R(r) = \min_{\tau,\, x \in \mathcal{X}} \; & \tau \\ \text{s.t.} \;\; & a_0(x, \beta) \leq \tau, \quad \forall \beta \in \mathcal{U}(r), \\ & a_i(x, \beta) \leq \tau_i, \quad \forall \beta \in \mathcal{U}(r),\ \forall i \in [I]. \end{aligned} \tag{2}$$

where $\mathcal{U}(r)$ is an *uncertainty set*.

Building the Uncertainty Set. The *uncertainty set* is usually a normed ball of radius r centered at the estimate $\hat{\beta}$, defined by $\mathcal{U}(r) = \left\{\beta \in \bar{W} \mid \|\beta - \hat{\beta}\| \leq r\right\}$. Given a data set D, we compute the parameter $\hat{\beta}$ by using an estimation procedure that solves $\hat{\beta} = \arg\min_{\beta \in \mathcal{W}} \ \rho(\beta; \mathcal{D})$. where $\rho(\beta; \mathcal{D})$ is referred to as the *estimation metric*. For example, in ordinary least-squares (OLS) estimation, ρ is the sum of squared prediction errors $\rho(\beta; \mathcal{D}) = \|z - Y\beta\|_2^2$.

The *Estimate Uncertainty Set* is defined as follows :

$$\mathcal{E}(r; \mathcal{D}) \left\{\beta \in \mathcal{W} \mid \rho(\beta; \mathcal{D}) \leq \hat{\rho} + r\right\}. \tag{3}$$

where $\hat{\rho} = \rho(\hat{\beta}; \mathcal{D})$ denotes the minimal value of the estimation metric, and $r \geq 0$ represents the allowable deviation from this optimum.

JERO Formulation. With the estimate uncertainty set now defined, it serves as the basis for the JERO optimization problem, which thereby integrates the estimation procedure into the optimization model. Then, the *joint estimation and robustness optimization* (JERO) model is :

$$\begin{aligned} Z_E(\tau_0) = \max_{x \in \mathcal{X},\, r \geq 0} \ & r \\ \text{s.t.} \ & a_0(x, \beta) \leq \tau_0, \quad \forall \beta \in \mathcal{E}(r; \mathcal{D}), \\ & a_i(x, \beta) \leq \tau_i, \quad \forall \beta \in \mathcal{E}(r; \mathcal{D}),\ \forall i \in [I]. \end{aligned} \tag{4}$$

Under JERO, the objective function is now the size of the uncertainty set and the goal is to find a solution x that remains feasible for the largest possible uncertainty set. Observe that we must specify the goal cost τ_0 in the problem.

Methodology to Solve a JERO Problem. We will now explain the proposed methodology to solve problem (4). The problem is not necessarily jointly convex in r and x. However, given $r \geq 0$, the constraints in (4) are convex in x. Then, let's look at the following problem :

$$
\begin{aligned}
Z_E^r(\tau_0) = \min_{x\in\mathcal{X},\, t}\ & t - \tau_0 \\
\text{s.t.}\ \ & a_0(x,\beta) \le t, \quad \forall \beta \in \mathcal{E}(r;\mathcal{D}), \\
& a_i(x,\beta) \le \tau_i, \quad \forall \beta \in \mathcal{E}(r;\mathcal{D}),\ \forall i \in [I].
\end{aligned} \tag{5}
$$

If the optimal objective value of (5) is non-positive, this implies that there exists a decision $x \in \mathcal{X}$ that satisfies the goal cost τ_0 as well as all the robust constraints for the given $r \ge 0$. Conversely, if the optimal objective value is positive, then no such decision x exists, meaning that the goal cost and robust constraints cannot be simultaneously satisfied for that value of r.

If problem (5) can be solved efficiently, then a sequence of such problems can be solved in order to obtain the solution to problem (4). This observation is precisely the motivation behind the proposed methodology in the paper, which employs a binary search scheme on r based on repeated solutions of (5). Algorithm 1 in Zhu et al. (2020) takes as input an oracle for solving problem (5), a precision level $\Delta > 0$, and an upper bound $\bar{r}$ on the radius that must be big enough; it outputs a solution x that is feasible for the largest possible uncertainty radius r up to error Δ. At each iteration, the algorithm sets $r = (r_1 + r_2)/2$ as the midpoint of the current search interval $[r_1, r_2]$ and tests this value by solving problem (5). If the test confirms feasibility ($Z_E^T(\tau_0) \le 0$), then r_1 is updated to r, thereby increasing the radius; otherwise, r_2 is updated to r, thereby decreasing it. By repeatedly halving the search interval, the algorithm converges to the maximum radius r consistent with the goal cost τ_0 and solves the original JERO formulation up to a controllable approximation error, requiring only a logarithmic number $O(\log(\bar{r}/\Delta))$ oracle calls to problem (5), see Proposition 1 in Zhu et al. (2020).

However, in order to solve problem (5), it is necessary to derive a tractable representation of its robust counterpart. For any $i \in [I] \cup \{0\}$, let's focus on $a_i(x,\beta) \le \tau_i, \forall \beta \in \mathcal{E}(r;\mathcal{D})$. To obtain a reformulation, we require additional assumptions. In particular, Proposition 2 holds under the condition that $\mathrm{ri}(\mathcal{E}(r;\mathcal{D})) \ne \emptyset$ and that $a_i(x,\cdot)$ is closed concave for all $x \in \bar{\mathcal{X}}$, while Proposition 3 relies on the convexity of $\rho(\beta;\mathcal{D})$ in β.

Proposition 2. Assume that $\mathrm{ri}(\mathcal{E}(r;\mathcal{D})) \ne \emptyset$ and that $a_i(x,\cdot)$ is closed concave for all $x \in \bar{\mathcal{X}}$. Then x satisfies (9) if and only if x and ν $(x \in \bar{\mathcal{X}}, \nu \in \bar{\mathcal{W}})$ satisfy $\delta_r^*(\nu) - a_i^*(x,\nu) \le \tau_i$.

Here $a_i^*(x,\nu)$ is the concave conjugate of $a_i(x,\nu)$, defined as $a_i^*(x,\nu) = \inf_{\beta\in\mathcal{W}}\{\beta'\nu - a_i(x,\beta)\}$; and $\delta_r^*(\nu)$ is the support function of the set $\mathcal{E}(r;\mathcal{D})$, defined as $\delta_r^*(\nu) = \sup_{\beta\in\mathcal{W}}\{\beta'\nu \mid \rho(\beta;\mathcal{D}) \leq \hat{\rho} + r\}$.

Proof Sketch of Proposition 2. We first reformulate the original constraint in equivalent max form. Introducing the uncertainty set indicator $\delta_r(\beta)$ turns it into an unconstrained problem. Applying Fenchel's duality theorem then yields $\max_{\beta\in\mathcal{E}(r;D)} a_i(x,\beta) = \min_\nu\{\delta_r^*(\nu) - a_i^*(x,\nu)\}$. Substituting this back into the original constraint gives the desired result. □

Proposition 3. Assume that $r > 0$ and that $\rho(\beta;\mathcal{D})$ is convex in β. Then the support function of $\mathcal{E}(r;\mathcal{D})$ can be represented as $\delta_r^*(\nu) = \inf_{\mu>0}\left\{(\hat{\rho}+r)\mu + \mu\rho^*\left(\frac{\nu}{\mu};\mathcal{D}\right)\right\}$, where $\rho^*(\nu;\mathcal{D})$ is the convex conjugate of the estimation metric $\rho(\beta;\mathcal{D})$ and is defined as $\rho^*(\nu;\mathcal{D}) = \sup_{\beta\in\bar{\mathcal{W}}}\left\{\beta'\nu - \rho(\beta;\mathcal{D})\right\}$.

Proof Sketch of Proposition 3. The desired result is obtained by applying strong duality to the problem defined by the support function. □

Concretely, Proposition 2 shows how to replace the original robust constraints by an equivalent finite condition using duality. Proposition 3 then provides an explicit representation of this support function in terms of the convex conjugate of the estimation metric. Together, these results yield a reformulation of the robust constraints, which is the key step for solving problem (5).

Given an estimation procedure for the parameters, solving problem (4) reduces to finding an explicit and tractable formulation of problem (5). This is achieved by applying Propositions 2 and 3 and then deriving the corresponding support function of the estimate uncertainty set. The paper provides explicit support functions for a variety of estimation procedures, including:

- **Least Squares Estimation:** Ordinary Least Squares (OLS), LASSO;
- **Maximum Likelihood Estimation (MLE):** Multivariate Normal, Finite Support, Linear Regression, Poisson Regression, Logistic Regression, Independent Marginal.

Also, the paper illustrates these derivations with concrete examples of explicit JERO formulations. In what follows, we focus exclusively on the support function derivations for MLE Linear Regression and Poisson Regression, since these are the models used in the numerical study.

JERO Linear Regression Normal. Consider data $\mathcal{D} = \{z, Y\}$, where $z \in R^P$ is the response vector and $Y \in R^{P \times M}$ is a full column rank design matrix. The linear regression model is $z = Y\beta + \epsilon$ where $\beta \in R^M$ is the vector of regression coefficients and ϵ represents i.i.d. normal errors. Then, the log-likelihood function is $\ell(\beta, \sigma^2; \mathcal{D}) = -\frac{1}{2}\left(\ln(2\pi\sigma^2) + \frac{1}{P\sigma^2}\|z - Y\beta\|_2^2\right)$ and the estimation metric would be :

$$\rho(\beta; \mathcal{D}) = \min_{\sigma^2 \in R_+} \frac{1}{2}\left(\ln(\sigma^2) + \frac{1}{P\sigma^2}\|z - Y\beta\|_2^2\right) = \frac{1}{2}\left(\ln\left(\frac{\|z-Y\beta\|_2^2}{P}\right) + 1\right). \tag{6}$$

Plugging the well known maximum likelihood estimator $\hat{\beta} = (Y'Y)^{-1}Y'z$, the corresponding optimal value of the estimation metric is $\hat{\rho} = \frac{1}{2}\left(\ln\left(\frac{\|z-Y(Y'Y)^{-1}Y'z\|_2^2}{P}\right) + 1\right)$.

Proposition 8. Let $r > 0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r; \mathcal{D})$, given the estimation metric (6), is

$$\begin{aligned} \delta_r^*(\nu) = \min_{w \in R^P} \; & \gamma(r)\|w\|_2 + z'w \\ \text{s.t. } & Y'w = \nu, \end{aligned} \tag{7}$$

where $\gamma(r) = \sqrt{P\exp(2\hat{\rho} + 2r - 1)}$.

Proof Sketch of Proposition 8. The constraint defines an ℓ_2 ball. Thus the support function reduces to $\max_\beta\{\nu^\top\beta : \|z - Y\beta\|_2 \leq \gamma(r)\}$, whose dual form gives the desired reformulation. $\square$

JERO Poisson Regression. Given data $\mathcal{D} = \{z, Y\}$ with $z \in N^P$ the response vector and $Y \in R^{P \times M}$ the matrix of explanatory variables, consider the same linear model as before but with respect to a Poisson distribution. The corresponding log-likelihood function is $\ell(\beta; \mathcal{D}) = \frac{1}{P}\sum_{i=1}^P (z_i \ln(\beta'Y_i) - \beta'Y_i - \ln(z_i!))$. Then, the estimation metric would be

$$\rho(\beta; \mathcal{D}) = \frac{1}{P}\sum_{i=1}^P (\beta'Y_i - z_i \ln(\beta'Y_i)). \tag{8}$$

Proposition 9. Let $r > 0$. Then the corresponding support function of the estimate uncertainty set $\mathcal{E}(r;\mathcal{D})$, given the estimation metric (8), is

$$\begin{aligned} \delta_r^*(\nu) = \min_{u\in R_+,\, v\in R_+^P,\, w\in R^P} \quad & (\hat{\rho}+r)Pu + e'w \\ \text{s.t.} \quad & \nu + \sum_{i=1}^{P} v_i Y_i - u\sum_{i=1}^{P} Y_i = 0, \\ & z_i u \ln(z_i u/v_i) - z_i u - w_i \le 0, \quad \forall i \in [P]. \end{aligned} \tag{9}$$

where $\hat{\rho}$ is the optimal value of the corresponding estimation metric.

Proof Sketch of Proposition 9. We first reformulate the support function problem by introducing auxiliary variables f_i to rewrite the log terms and make the problem easier to dualize. Applying strong duality then yields the desired form. □

Note that the last row constraints in (9) are difficult. To address this, the paper constructs a SOC approximation of the exponential cone, which enables the use of efficient solvers.

*2) **Numerical results***

: **Introduction** The numerical experiments are based on a real-world *patient flow problem* in the Chinese healthcare system. The data come from the New Rural Cooperative Medical Insurance program in Anhui province, China, and cover the year 2015. It includes 26 hospitals that admitted more than 25,500 cerebral infarction patients. The dataset contains detailed patient information (ID, age, gender, cost, reimbursement, admission and discharge dates, length of stay, location) as well as hospital information (ID, hospital level, number of beds, and location).

The study models hospital demand and evaluates reimbursement policies under parameter uncertainty. The proposed JERO framework is compared with deterministic models in terms of the resulting optimal total budget, with training and test sets defined by odd and even days. Robustness performance is assessed by the number of violated constraints on the test set.

In this case study, hospital demand is measured in *bed-days*, defined as the total number of days that hospital beds are occupied by patients. We assume the following explanatory model: Bed-days $= \beta_0 + \beta_1\,\text{LOS} + \beta_2\,\text{Cost} + \beta_3\,\text{Capacity} + \beta_4\,\text{Reimbursement}$, where LOS denotes the average length of stay, Cost the average medical expenditure, Capacity the number of beds in the hospital, and Reimbursement the average insurance coverage level.

The objective is to determine the *optimal average reimbursement per bed-day* for each hospital in order to improve total budget for deterministic models and respect goal budget with robustness for JERO models. This decision is subject to the following constraints that define a realistic problem:

- **Goal Cost for JERO models:** The total reimbursement across all hospitals must not exceed 120 million ¥. This *goal cost* corresponds to the current level of expenditure and is considered reasonable.
- **Policy constraint:** The reimbursement ratio for each hospital cannot fall below 40%, ensuring that patients receive a decent level of coverage.
- **Capacity constraint:** The system must ensure treatment for at least 25,500 patients, matching the observed flow in the real data.

In this case study, the estimation of the demand model parameters is carried out using maximum likelihood estimation (MLE). Specifically, the paper considers both linear regression under the assumption of normally distributed demand and Poisson regression. The corresponding support functions for both the JERO-Normal and JERO-Poisson models were presented earlier in this report. Here, however, we do not reproduce the detailed problem formulations of the deterministic and JERO optimization models. Instead, we focus directly on the numerical results and their implications.

Results

TABLE I

OPTIMAL REIMBURSEMENT RATES UNDER DIFFERENT OPTIMIZATION MODELS

	Current	Normal	Poisson	JERO-Normal	JERO-Poisson
Budget	¥102,196,990	¥91,724,427	¥91,206,700	¥102,151,504	¥102,151,504

Table I reports the optimal reimbursement budgets obtained under the different optimization models. The JERO-Normal and JERO-Poisson approaches both deliver solutions that respect the goal cost of 120 million ¥ and even remain slightly below it, demonstrating their ability to balance robustness with reasonable budget feasibility. The deterministic models (Normal and

Poisson) yield substantially lower reimbursement budgets, around 91 million ¥, which may appear attractive in terms of cost savings but do not incorporate robustness considerations.

Figure 1 illustrates the robustness of the JERO solutions when varying the goal cost.

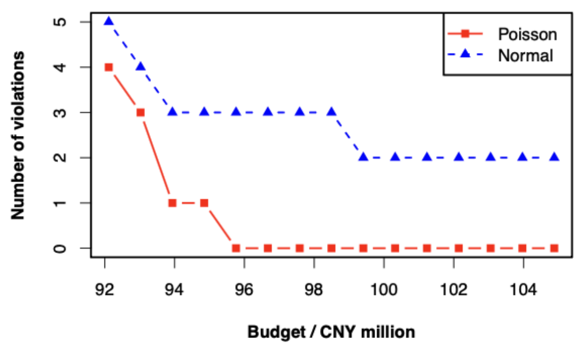

Fig. 1. Robustness of solutions

The paper considers 15 different budget levels, ranging from 92.1 to 104.9 million ¥, and evaluates the resulting solutions of JERO models by the number of violated constraints on the test set. We can observe that already from around 96 million ¥, the JERO-Poisson model achieves *zero* violations, perfectly robust solutions, while still operating at a budget below the actual expenditure of about 102 million ¥. Another observation is that JERO-Poisson consistently outperforms JERO-Normal in terms of robustness. As discussed in the paper, this superiority is likely due to the more plausible assumption of an underlying Poisson distribution for hospital demand. Unfortunately, the paper does not report the number of violated constraints for the deterministic models, which would have provided a clearer baseline for robustness comparison.

III. Summary and Assessment

The paper proposes a generic and well-motivated framework that integrates parameter estimation with robust optimization. Instead of relying on arbitrary uncertainty radius, JERO constructs data-driven uncertainty sets directly from estimation error and maximizes the feasible radius under a prescribed goal cost. The framework remains tractable and relatively simple, relying on binary search over subproblems. Explicit formulations are provided for several estimation procedures. The healthcare reimbursement case study demonstrates that JERO produces solutions that are fully robust on the test data while attaining a lower cost than the actual expenditure.

The paper shows several notable strengths. First, it integrates the parameter estimation procedure directly into the construction of the uncertainty set, thereby eliminating the need for an arbitrarily specified radius and deriving maximal robustness. This results in a framework that is both conceptually rigorous and practically relevant. Second, the approach remains computationally tractable, even in large-scale convex settings, while maintaining a relatively simple formulation. Third, the substitution of a target cost in place of a pre-specified uncertainty size both enhances interpretability and can often be identified in practice, for example by solving the deterministic model to obtain insights into suitable goal levels. Finally, the paper provides several concrete and diverse examples, which demonstrate the broad applicability of the framework and highlight its potential to address a wide range of optimization problems in real life.

The paper also presents several weaknesses. First, it relies on restrictive assumptions to ensure tractability, which may limit its applicability in more general settings. For example, maximum likelihood estimation in Gaussian mixture models yields a non-convex objective in the parameters, which violates the convexity condition required in Proposition 3. Second, the empirical validation is limited to a single case and the experimental design presents serious limitations: the paper addresses robustness only by comparing the two JERO variants rather than against the deterministic models, drifting from the initial objective of the experiment. Moreover, the comparison of budgets is not meaningful, since JERO is evaluated with a fixed goal budget while the deterministic models minimize total budget; without a robustness benchmark, such a comparison lacks interpretive value. Also, the absence of direct comparisons with standard robust optimization or DRO baselines makes it difficult to fully evaluate JERO's performance. Finally, Binary search algorithm requires $\bar{r}$, but the paper offers no guidance on how to set it.

IV. Extension

A. Regularization Uncertainty Sets

Motivation. The *Estimate Uncertainty Set* is defined as $\mathcal{E}(r; D) = \{\beta \in W \mid \rho(\beta; D) \leq \hat{\rho} + r\}$, as we saw at (3). This construction ties robustness directly to the value of the estimation metric.

Yet the resulting uncertainty set $\mathcal{E}(r; D)$ does not necessarily privilege statistically meaningful estimators: as r increases, it simply admits all β with larger estimation error, including potentially poor fits, without any mechanism to orient the set toward genuinely good estimators.

In practice, estimation procedures often include a regularization term designed to improve the generalization of parameters to unseen data. Under this perspective, the paradigm of uncertainty may shift a bit: uncertainty is understood as stemming from the risk of poor generalization.

Model Definition. Given a convex loss $\ell(\beta; D)$ and a convex regularizer $R(\beta)$, define the penalized estimation problem $\hat{\beta}(\lambda) \in \arg\min_{\beta \in W} \ \ell(\beta; D) + \lambda R(\beta), \lambda \geq 0$. For a fixed parameter $\Lambda \geq 0$, the *Regularization Uncertainty Set* is

$$\mathcal{U}_\Lambda = \bigcup_{\lambda \in [0,\Lambda]} \hat{\beta}(\lambda).$$

Thus $\mathcal{U}_\Lambda$ contains all estimators that would be obtained by solving the penalized problem with any regularization weight between 0 (no penalty) and Λ (specified maximal regularization).

Advantages. The Regularization Uncertainty Set approach exhibits several conceptual and practical advantages compared to the classical RO and Estimate Uncertainty Set from JERO :

1) **Estimator-awareness beyond fit.** Whereas $\mathcal{E}(r; D)$ is defined by proximity in estimation error, $\mathcal{U}_\Lambda$ encodes the bias–variance trade-off induced by regularization. It can be viewed as a natural extension of the Estimate Uncertainty Set proposed in JERO.
2) **Interpretability.** The specified Λ has an interpretable meaning: it reflects the maximum degree of regularization deemed plausible. Decision makers can relate Λ to model complexity or sparsity level. In this way, the role of the radius r is replaced by a specified Λ, avoiding the need for an arbitrary choice of r and ensuring that the uncertainty set remains intrinsically tied to the estimation procedure, in contrast to classical robust optimization.
3) **Convexity.** Classical regularization terms such as the ℓ_1 (Lasso) and ℓ_2 (Ridge) penalties are convex, and by convex additivity the overall objective remains convex. This preservation of convexity is a desirable property in robust optimization.
4) **Structural modeling.** By selecting different regularizers R, we can flexibly induce uncertainty sets that respect problem-specific structure (sparsity with Lasso, stability with Ridge,

...). In this way, we actively guide the parameters toward desired properties through the choice of regularization, while ensuring robustness to variability in parameters that share those properties.

B. Data-Space Uncertainty Sets

Motivation. The original JERO framework constructs uncertainty sets on the *estimated parameters* by enlarging the estimation metric around its optimal value. This construction can be seen as indirect: it places robustness on the parameters, whereas the fundamental postulate in JERO is that the data itself is inherently imprecise. In many applications, data are subject to measurement noise, reporting errors, or systematic biases. A more faithful way to capture this uncertainty is therefore to construct first the uncertainty set in the *data space*, and subsequently propagate this data-level perturbation to the parameters of interest. We seek to suggest a generic framework that addresses this motivation. Our intent is conceptual: we do not delve into the tractability of the robust counterparts.

Model Definition. Let $D \in R^{n\times m}$ denote the data matrix (rows as observations, columns as features), with associated parameter estimator $\hat{\beta}(D)$. Instead of directly defining an estimate uncertainty set as in JERO, we define a data uncertainty set $\mathcal{U}_D(r)$ as a "ball" around D according to one of the following generic approaches:

- *Norm-based bounds.* We constrain perturbations via norms applied at different levels. For example, feature-wise robustness can be enforced by $\|\Delta D_{:,j}\| \leq s_j$ (reflecting measurement accuracy of feature j), observation-wise robustness by $\|\Delta D_{i,:}\| \leq b_i$ (accounting for localized errors), and global robustness by $\|\Delta D\|_F \leq r$ (limiting total perturbation). These conditions can also be combined. Formally, the combined data-space uncertainty set is

$$\mathcal{U}_D(r) \;=\; \Big\{ D' \in R^{n\times m} \;\Big|\; \|D' - D\|_F \leq r,\;\; \|D'_{:,j} - D_{:,j}\| \leq s_j,\;\; \|D'_{i,:} - D_{i,:}\| \leq b_i \;\; \forall i, j \Big\}$$

- *Probabilistic divergence bounds:* Interpret the dataset D as sampled from an empirical distribution $\hat{P}_n$. Then construct an ambiguity set

$$\mathcal{U}_D(r) \;=\; \Big\{ D' \;\Big|\; \mathsf{dist}\big(\hat{P}'_n, \hat{P}_n\big) \leq r \Big\},$$

where dist is a statistical distance (like Wasserstein) and $\hat{P}_n'$ is the empirical distribution of D'. The data perturbation radius r can be interpreted as the maximum statistical deviation from the data empirical distribution that we are willing to tolerate.

Given a data-space uncertainty set, the induced parameter uncertainty set is defined by

$$\mathcal{U}_\beta(r) = \left\{ \beta \;\middle|\; \exists D' \in \mathcal{U}_D(r) \text{ such that } \beta = \hat{\beta}(D') \right\}.$$

Automatic Radius Selection via Goal-Cost Maximization. If we assume that the robust counterparts of the proposed data-space uncertainty sets are tractable, then, in full analogy with the JERO framework, the radius r *need not be specified*. Instead, r can be determined for both *norm-based* constructions and *probabilistic* constructions by solving a goal-cost problem that maximizes the admissible uncertainty size subject to feasibility of the optimization constraints at the prescribed target cost.

Calibration of r. Other calibration methods remain plausible:

- *Probabilistic calibration:* choose r as the largest radius for which the empirical and true distributions are statistically indistinguishable at level $1 - \alpha$ (e.g., concentration-based or test-based), therefore ensuring that the true law lies in $\mathcal{U}_D(r)$ with high probability.
- *Norm-based calibration:* set s_j from known instrument reliabilities of measurements or domain tolerances on feature-level perturbations; calibrate b_i using observation-level confidence or reliability scores; and interpret the global bound r as an aggregate tolerance level (for example, set as 5% of the overall data magnitude).

Advantages.

1) **Closer to JERO's postulate:** uncertainty is rooted in the intrinsic imprecision of data.
2) **Compatible with maximal radius derivation via goal cost:** norm bounds tie to measurement tolerances; Wasserstein radius admit probabilistic meaning and statistical control; both allow *automatic* selection of the largest admissible radius via goal-cost maximization under the assumption that robust counterparts are tractable.
3) **Flexibility:** multi-granular and abstract design that adapts to data quality and domain knowledge.

C. Automatic Bound Selection for Binary Search

Motivation. Algorithm 1 in Zhu et al. (2020) requires as input an upper bound $\bar{r}$ that is "large enough." However, $\bar{r}$ is not directly interpretable, and if it is chosen too small, the procedure will fail to capture the true maximizing radius, which is critical for obtaining the correct JERO solution. To avoid specifying an arbitrary bound, we propose a simple doubling scheme that automatically identifies a valid interval for proceeding with the binary search.

Procedure.

Algorithm 1 Doubling Scheme for Initial Bound

Input: Oracle that solves Problem (5), goal cost τ_0.
Initialize: $r \leftarrow 1$.
if $Z_E^r(\tau_0) > 0 \quad$ (infeasible) **then**
  Output: interval $[0, 1]$ and terminate.
end if
while true **do**
  $r_{\text{next}} \leftarrow 2r$.
  if $Z_E^{r_{\text{next}}}(\tau_0) > 0 \quad$ (infeasible) **then**
    Output: interval $[r, r_{\text{next}}]$ and terminate.
  else
    $r \leftarrow r_{\text{next}}$.
  end if
end while

Advantages.

1) Eliminates the need to specify an arbitrary $\bar{r}$ with a simple implementation.
2) Requires only logarithmic effort in the magnitude of the true bound.
3) Guarantees that the returned interval contains a valid "large enough" r and the returned interval is already refined for binary search.

References

Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

Bertsimas, D., Gupta, V., and Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, 167(2):235–292.

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166.

Zhu, T., Xie, J., and Sim, M. (2020). Joint estimation and robustness optimization. *Management Science*. Available at: https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2020.3898?journalCode=mnsc.