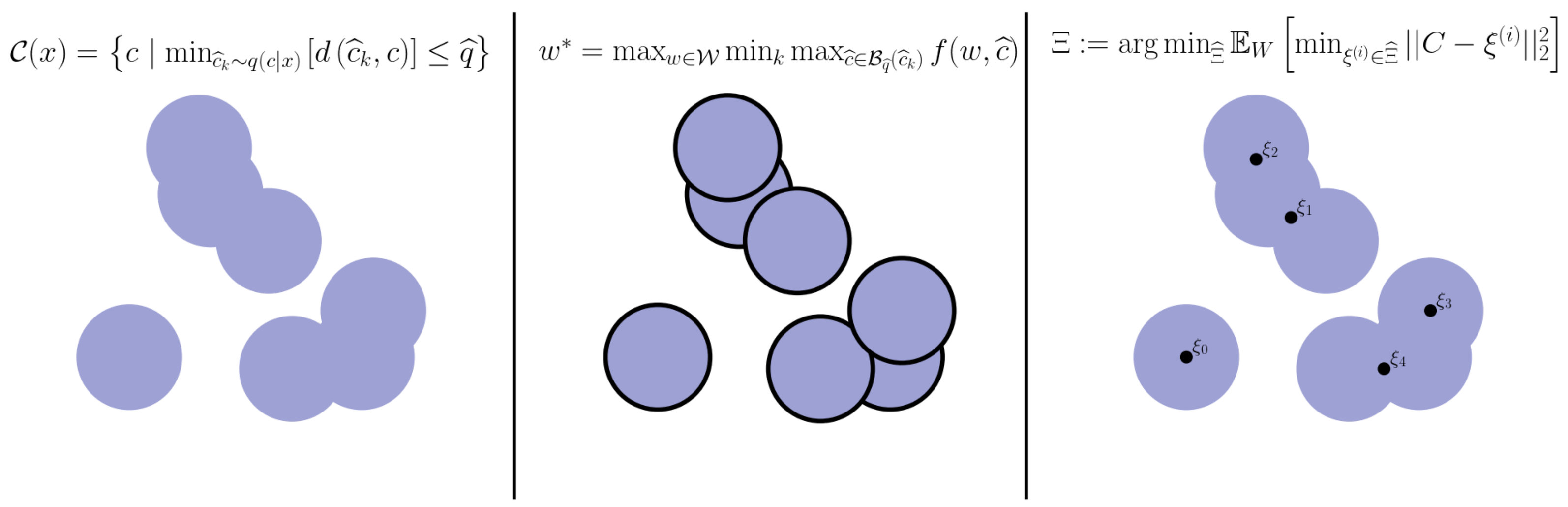
Conformal Contextual Robust Optimization

Yash Patel, Sahana Rayan, Ambuj Tewari

Department of Statistics, University of Michigan

Overview

Data-driven approaches to predict-then-optimize decision-making problems seek to mitigate the risk of uncertainty region misspecification in safety-critical settings. Current approaches, however, suffer from considering overly conservative uncertainty regions, often resulting in suboptimal decision-making. To this end, we propose Conformal-Predict-Then-Optimize (CPO), a framework for leveraging highly informative, nonconvex conformal prediction regions over high-dimensional spaces based on conditional generative models, which have the desired distribution-free coverage guarantees.



Setup

Let $c \in \mathcal{C}$, where $(\mathcal{C}, d)$ is a general metric space. Consider a convex-concave objective function $f(w, c)$ that are L-Lipschitz in c under the metric d for any fixed w, as follows:

$$w^*(x) := \min_{w, \mathcal{U}} \max_{\widehat{c} \in \mathcal{U}(x)} \quad f(w, \widehat{c}) \qquad \text{s.t.} \quad \mathcal{P}_{X,C}(C \in \mathcal{U}(X)) \ge 1 - \alpha, \tag{1}$$

For any $\mathcal{U}$, this robust counterpart to the nominal problem produces a valid upper bound if $c \in \mathcal{U}(x)$. Denote the suboptimality as $\Delta(x, c) := \min_w \max_{\widehat{c} \in \mathcal{U}(x)} f(w, \widehat{c}) - \min_w f(w, c)$.

Lemma

Consider any $f(w, c)$ that is L-Lipschitz in c under the metric d for any fixed w. Assume further that $\mathcal{P}_{X,C}(C \in \mathcal{U}(X)) \ge 1 - \alpha$. Then,

$$\mathcal{P}_{X,C}\left(0 \le \Delta(X, C) \le L \operatorname{diam}(\mathcal{U}(X))\right) \ge 1 - \alpha. \tag{2}$$

Method

We desire non-convex prediction regions $\mathcal{U}(x)$ to make the resulting suboptimality as small as possible. However, generically solving Equation 1 then becomes computationally intractable. Crucially, however, we can consider a particular conformal score that lends itself to a decomposition under which such optimization becomes tractable. For a fixed K and $\{\widehat{c}_k\}_{k=1}^K \sim q(C \mid x)$, let

$$s(x, c) = \min_k [d(\widehat{c}_k, c)]. \tag{3}$$

Let $\mathcal{C}(x)$ denote the resulting prediction regions under this score. With this score, exact solution of the min-max problem follows with gradient-based optimization on $\phi(w) := \max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c})$.

By Danskin's Theorem, $\nabla_w \phi(w) = \nabla_w f(w, c^*)$, where $c^* := \max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c})$. Efficient solution of this RO problem, therefore, reduces to being able to efficiently solve the maximization problem over $\mathcal{C}(x)$, which we can do with:

$$\max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c}) = \max_k \max_{\widehat{c} \in \mathcal{B}_{\widehat{q}}(\widehat{c}_k)} f(w, \widehat{c}), \tag{4}$$

where the maximum over a ball can be efficiently computed with traditional convex optimization techniques.

Algorithm 1 CPO-Opt

procedure CPO-Opt
  Inputs: Context x, CGM $q(C \mid X)$, Optimization steps T, Score samples K, Conformal quantile $\widehat{q}$
  $w \sim U(\mathcal{W}), \{\widehat{c}_k\}_{k=1}^K \sim q(C \mid x)$
  for $t \in \{1, \dots T\}$ **do**
    $\left\{c_k^* \leftarrow_{\widehat{c} \in \mathcal{B}_{\widehat{q}}(\widehat{c}_k)} f(w, \widehat{c})\right\}_{k=1}^K$
    $c^* \leftarrow_{c_k^*} f(w, c_k^*)$
    $w \leftarrow \Pi_{\mathcal{W}}(w - \eta \nabla_w f(w, c^*))$
  end for
  Return w
end procedure

Volume Estimation

The cost of performing Equation 4 linearly increases with K. However, it must also be juxtaposed with the improved statistical efficiency of such prediction regions. This inflection point can be identified prior to the optimization by estimating the volume of a union of hyperspheres for various values of K. Monte Carlo estimates of the prediction region volume are computed using Voronoi cells of the hypersphere centers which is given by:

$$\widehat{\ell}(\{\mathcal{B}_{\widehat{q}}(\widehat{c}_k)\}) := |\mathcal{B}_{\widehat{q}}| \sum_{k=1}^{K} \mathcal{P}_{C \sim U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))}(C \in V(\widehat{c}_k)) \tag{5}$$

where $U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))$ denotes an uniform distribution over the ball centered at $\widehat{c}_k$ and $V(\widehat{c}_k)$ the Voronoi cell of $\widehat{c}_k$, defined as $\{z \in \mathbb{R}^d \mid d(\widehat{c}_k, z) \le d(\widehat{c}_{k'}, z), k' \ne k\}$. K^* will be the inflection point if it is the smallest K such that $|\widehat{\ell}_K - \widehat{\ell}_{K+1}| \le \epsilon$ for some ϵ volume tolerance

Experiment

We similarly consider the robust traffic flow problem (RTFP) for a source-target pair (s, t) over the network graph of Manhattan, where $|\mathcal{V}| = 4584$ and $|\mathcal{E}| = 9867$. The precipitation $\widetilde{Y}$ was combined with the nominal speed limits to obtain the final travel costs c along edges:

$$w^*(x) := \min_w \max_{\widehat{c} \in \mathcal{U}(x)} \widehat{c}^T w \tag{6}$$

$$\text{s.t.} w \in [0, 1]^{\mathcal{E}}, Aw = b, \mathcal{P}_{X,C}(C \in \mathcal{U}(X)) \ge 1 - \alpha$$

where w_e represents the proportion of traffic routed along edge e, $C \in \mathbb{R}^{|\mathcal{E}|}$ is the edge weight vector, $A \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ is the node-arc incidence matrix, and $b \in \mathbb{R}^{|\mathcal{V}|}$ has entries $b_s = 1, b_t = -1$, and $b_k = 0$ for $k \notin \{s, t\}$.

	Box	PTC-B	Ellipsoid	PTC-E	CPO	*Nominal*
Coverage	0.94	0.93	0.94	0.92	0.94	–
Objective	7863.45 (0.0)	34559.03 (171.3)	7038.77 (0.0)	8807.68 (4.22)	**4171.22 (321.34)**	*299.50 (0.0)*

Discussion

This work suggests many directions for future work. We are actively pursuing the extension of CPO to robust LQR control and subsequently to the broader category of robust control. Also, leveraging CPO over function spaces would enable its use to distributionally robust optimization.

Github: https://github.com/yashpatel5400/csi

yppatel@umich.edu