Conformal Contextual Robust Optimization

by Yash Patel, Sahana Rayan, Ambuj Tewari

May 2024

Problem Formulation

- Standard formulation is

$$w^*(x) := \min_{w \in \mathcal{W}} \quad \mathbb{E}[C^T w \mid x], \tag{1}$$

  w are decision variables, C an *unknown* cost parameter, x observed context, and $\mathcal{W}$ a compact feasible region
- Nominal: predict $\widehat{c} := f(x)$ and then optimize $\min_w \widehat{c}^T w$.

Problem Formulation

- Robust formulation

$$\begin{aligned} w^*(x) := \min_{w,\mathcal{U}} \max_{\widehat{c} \in \mathcal{U}(x)} \quad & f(w, \widehat{c}) \\ \text{s.t.} \quad & \mathcal{P}_{X,C}(C \in \mathcal{U}(X)) \geq 1 - \alpha, \end{aligned} \tag{2}$$

  where $\mathcal{U} : \mathcal{X} \to \Omega_{\mathcal{Y}}$ is a uncertainty region predictor
- Assume $c \in \mathcal{C}$, where $(\mathcal{C}, d)$ is a general metric space
- Let $f(w, c)$ be L-Lipschitz in c under the metric d for any fixed w

Problem Formulation

Denote $\Delta(x, c) := \min_w \max_{\widehat{c} \in \mathcal{U}(x)} f(w, \widehat{c}) - \min_w f(w, c)$

Coverage Bound

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

$$\mathcal{P}_{X,C}\left(0 \leq \Delta(X, C) \leq L \operatorname{diam}(\mathcal{U}(X))\right) \geq 1 - \alpha. \tag{3}$$

Conformal Prediction

- Interested in producing calibrated prediction regions:

$$\mathcal{P}_{X,Y}(Y \notin \mathcal{C}(X)) \leq \alpha$$

- $\mathcal{D} = \{(x^{(1)}, y^{(1)}), \ldots (x^{(N)}, y^{(N)})\}$ i.i.d. from $\mathcal{P}(X, Y)$

Conformal Prediction

1. Split conformal: $\mathcal{D} = \mathcal{D}_{\mathcal{T}} \cup \mathcal{D}_{\mathcal{C}}$
2. Train a predictor $\hat{f}(x)$ on $\mathcal{D}_{\mathcal{T}}$
3. Define a score function $s(x, y)$: should “act like” a residual(Ex: $s(x, y) = \|\hat{f}(x) - y\|$)

Conformal Prediction

4. Evaluate $\mathcal{S} = \{s(x^{(j)}, y^{(j)})\}_{j=1}^{N_{\mathcal{C}}}$ for $(x^{(j)}, y^{(j)}) \in \mathcal{D}_{\mathcal{C}}$
5. Define $\widehat{q}(\alpha)$ to be $\lceil (N_{\mathcal{C}} + 1)(1 - \alpha) \rceil / N_{\mathcal{C}}$ quantile of $\mathcal{S}$
6. For $\mathcal{C}(x) = \{y \mid s(x, y) \leq \widehat{q}(\alpha)\}$, $1 - \alpha \leq \mathcal{P}_{X,Y}(Y \in \mathcal{C}(X))$

Optimization

- Wish to solve $\min_w \max_{\widehat{c} \in \mathcal{U}(x)} f(w, \widehat{c})$
- Rewrite as $\min_{w \in \mathcal{W}} \phi(w)$ for $\phi(w) := \max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c})$
- $\phi(w)$ is convex (Danskin's Theorem)
- Solve via projected gradient descent:

$$\nabla_w \phi(w) = \nabla_w f(w, c^*)$$

where $c^* := \max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c})$

Optimization

- Want to have non-convex prediction regions $\mathcal{U}(x)$ to minimize resulting suboptimality
- Solving $c^* := \max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c})$ computationally intractable

Score Function

- Learn conditional generative model $q(C \mid X)$
- For a fixed K, define the score

$$s(x, c) = \min_k \{[d\left(\widehat{c}_k, c\right)]\}_{k=1}^K \tag{4}$$

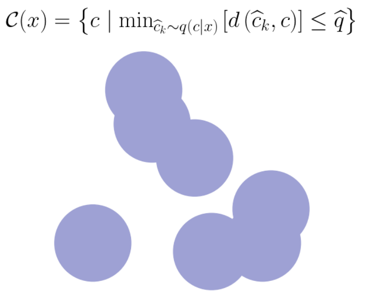
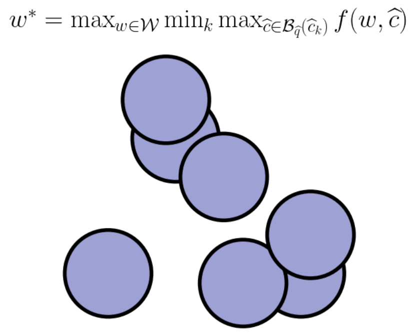


Figure 1: $\mathcal{C}(x) = \bigcup_{k=1}^K \mathcal{B}_{\widehat{q}}(\widehat{c}_k)$ where $\mathcal{B}_{\widehat{q}}$ is a ball of radius $\widehat{q}$

Optimization

- Maximum can efficiently computed as

$$\max_{\widehat{c} \in \mathcal{C}(x)} f(w, \widehat{c}) = \max_{k} \max_{\widehat{c} \in \mathcal{B}_{\widehat{q}}(\widehat{c}_k)} f(w, \widehat{c}) \tag{5}$$



Optimization

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
        $\left\{c_k^* \leftarrow \arg\max_{\widehat{c} \in \mathcal{B}_{\widehat{q}}(\widehat{c}_k)} f(w, \widehat{c})\right\}_{k=1}^{K}$
        $c^* \leftarrow \arg\max_{c_k^*} f(w, c_k^*)$
        $w \leftarrow \Pi_{\mathcal{W}}(w - \eta \nabla_w f(w, c^*))$
    end for
    Return $w$
end procedure
```

Results: Traffic Routing

- Robust traffic flow problem (RTFP) for network graph of Manhattan, where $|\mathcal{V}| = 4584$ and $|\mathcal{E}| = 9867$
- $\mathcal{P}(\widetilde{Y} \mid x)$ is probabilistic weather forecaster
- $\widetilde{Y}$ is post-processed to get edge cost C

Results: Traffic Routing

- RTFP formulation similar to knapsack:

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

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

where w_e is 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 $k \notin \{s, t\}$.

Results: Traffic Routing

	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)*

Results: Traffic Routing

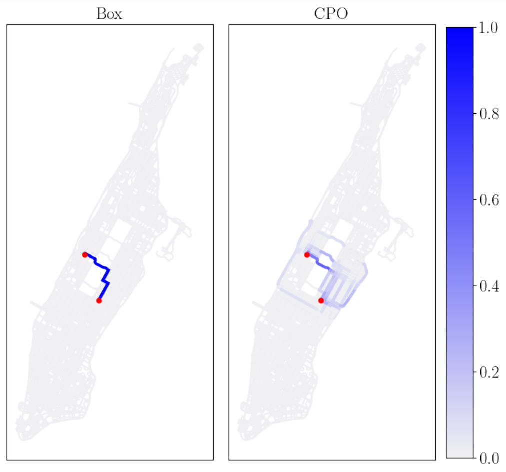


Figure 2: Solutions for RTFP under the Box (left) and CPO (right) regions.

Questions?

Visit our poster (#983) for more details!

Optimization

PGD Convergence
Let $\phi(w) := \max_{\widehat{c} \in \bigcup_{k=1}^{K} \mathcal{B}_{\widehat{q}}(\widehat{c}_k)} f(w, \widehat{c})$ for $\{\widehat{c}_k\}_{k=1}^{K} \subset \mathcal{C}$, $\widehat{q} \in \mathbb{R}^+$, and $f(w, c)$ convex-concave and L-Lipschitz in c for any fixed w. Let $w^* \in \mathcal{W}$ be a minimizer of ϕ. For any $\epsilon > 0$, define $T := \frac{L^2 ||w_0 - w^*||}{\epsilon^2}$ and $\eta := \frac{||w_0 - w^*||}{L\sqrt{T}}$. Then the iterates $\{w_t\}_{t=0}^{T}$ returned by Algorithm 1 satisfy

$$\phi\left(\frac{1}{T+1}\sum_{t=0}^{T} w_t\right) - \phi(w^*) \leq \epsilon. \tag{7}$$

- Consequence is “outer” iterations (i.e. T) has no dependence on K

K Selection

- Solely focus attention on choice of K on the "inner" optimization computational cost,
- $\max_k \phi_k(w)$ has linearly increasing cost with K
- Must be juxtaposed with the improved *statistical* efficiency of such prediction regions

K Selection

- Inflection point can be determined *prior* to optimization
- Only requires access to $q(C \mid X)$ and test samples to estimate the prediction region size
- Estimation of the volume of a union of hyperspheres is complicated due to overlapped regions

K Selection

- Inflection point can be determined *prior* to optimization
- Only requires access to $q(C \mid X)$ and test samples to estimate the prediction region size
- Estimation of the volume of a union of hyperspheres is complicated due to overlapped regions
- Volume estimated using Monte Carlo estimates from Voronoi cells:

$$\widehat{\ell}(\{\mathcal{B}_{\widehat{q}}(\widehat{c}_k)\}) := |\mathcal{B}_{\widehat{q}}| \sum_{k=1}^{K} \mathcal{P}_{C \sim U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))}(C \in V(\widehat{c}_k)), \qquad (8)$$

where $C \sim U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))$ denotes a uniform, $|\mathcal{B}_{\widehat{q}}|$ the volume of a hypersphere of radius $\widehat{q}$, and $V(\widehat{c}_k)$ the Voronoi cell of $\widehat{c}_k$, defined as $\{z \in \mathbb{R}^d \mid d(\widehat{c}_k, z) \leq d(\widehat{c}_{k'}, z), k' \neq k\}$

K Selection

- Choose K^* to be the inflection point
- $\arg\min_K |\widehat{\ell}_K - \widehat{\ell}_{K+1}| \leq \epsilon$ for some user-specified ϵ volume tolerance

K Selection

Algorithm 2 CPO

```
procedure VolumeEst
Inputs: Context $x$, CGM $q(C \mid X)$, Conformal quantile $\widehat{q}$
    $\{\widehat{c}_k\}_{k=1}^{K} \sim q(C_{1:K} \mid x)$
    $\left\{\{c_{k,m}\}_{m=1}^{M} \sim U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))\right\}_{k=1}^{K}$
    Return $|B_{\widehat{q}}| \sum_{k=1}^{K} \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}\left[c_{k,m} \in V(\widehat{c}_k)\right]$
end procedure
```

K Selection

Algorithm 3 CPO

```
procedure CPO
    Inputs: Context $x$, CGM $q(C \mid X)$, Optimization steps $T$, Desired coverage $1 - \alpha$, Max samples $K_{\max}$, Volume Tolerance $\epsilon$, Calibration sets $\mathcal{D}_{\mathcal{C}_1}, \mathcal{D}_{\mathcal{C}_2}$
    for $K \in \{1, \ldots K_{\max}\}$ do
        $s_K(x, c) \leftarrow \min_{\widehat{c}_k \in \{\widehat{c}_i\} \sim q(C_{1:K} | x)} \left[ d\left(\widehat{c}_k, c\right) \right]$
        $\mathcal{S}_K \leftarrow \left\{ s_K(x^{(i)}, c^{(i)}) \mid (x^{(i)}, c^{(i)}) \in \mathcal{D}_{\mathcal{C}_1} \right\}$
        $\widehat{q}_K \leftarrow \frac{\lceil (|\mathcal{D}_{\mathcal{C}_1}|+1)(1-\alpha) \rceil}{|\mathcal{D}_{\mathcal{C}_1}|}$ quantile of $\mathcal{S}_K$
        $\widehat{\ell}_K \leftarrow \frac{1}{|\mathcal{D}_{\mathcal{C}_2}|} \sum_{i=1}^{|\mathcal{D}_{\mathcal{C}_2}|} \textsc{VolumeEst}(x^{(i)}, q, \widehat{q}_K)$
    end for
    $K^* \leftarrow \arg\min_K \left| \widehat{\ell}_K - \widehat{\ell}_{K+1} \right| \leq \epsilon$
    Return $\textsc{CPO-Opt}(x, q, T, K^*, \widehat{q}_{K^*})$
end procedure
```

RPs

- Decisions are hard-to-interpret without intuition on $\mathcal{C}(x)$
- Summarize contents and variability with “representative points”:

$$\Xi(x) := \arg\min_{\widehat{\Xi} \in \zeta} \mathbb{E}_{C \sim U(\mathcal{C}(x))} \left[\min_{\widehat{\xi}^{(i)} \in \widehat{\Xi}} d(C, \xi^{(i)}) \right] \tag{9}$$

RPs

- Requires samples drawn from $U(\mathcal{C}(x))$
- M samples are initially drawn $\{c_{k,m}\}_{m=1}^{M} \sim U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))$ for each k
- Must be thinned in overlap regions
- Voronoi cells are overlap-free

$$V(\widehat{c}_k) := \{z \in \mathbb{R}^d \mid d(\widehat{c}_k, z) \leq d(\widehat{c}_{k'}, z), k' \neq k\}$$

- Discard $\{c_{k,m}\} \in V(\widehat{c}_{k'})$ for $k \neq k'$

RPs

- Size of region around RPs in $\mathcal{C}(x)$ gives intuition
- For each RP, look at size of projections $\{\pi_j\}_{j=1}^{J}$, where $J = \dim(\mathcal{C})$

$$\left|V_j^{(i)}\right| := \sum_{c \in V^{(i)}} d^2(\pi_j(c), \pi_j(\xi^{(i)})). \tag{10}$$

Optimization

Algorithm 4 CPO-RPs: QueryBall$(\mathcal{T}, x, r)$ is an assumed subroutine that returns all points in the kd tree $\mathcal{T}$ that are within a radius r of x.

```
procedure CPO-RPs
Inputs: Context $x$, CGM $q(C \mid X)$, RP count $N$, Conformal quantile $\widehat{q}$
    $\{\widehat{c}_k\}_{k=1}^{K} \sim q(C_{1:K} \mid x)$
    $\left\{\{c_{k,m}\}_{m=1}^{M} \sim U(\mathcal{B}_{\widehat{q}}(\widehat{c}_k))\right\}_{k=1}^{K}$
    $C \leftarrow \{c_{k,m} \mid c_{k,m} \in V(\widehat{c}_k)\}_{k=1,m=1}^{K,M}$
    $\mathcal{T} \leftarrow \textsc{KD-Tree}(C)$
    $\mathcal{E} \leftarrow \bigcup_i \{c_i \times \textsc{QueryBall}(\mathcal{T}, c_i, \widehat{q}) \mid c_i \in \mathcal{T}\}$
    $\{C_\ell\} \leftarrow \textsc{ConnectedComponents}(\mathcal{G}(C, \mathcal{E}))$
    $\Xi \leftarrow \bigcup_{\ell=1}^{L} \{\textsc{K-Means++}(C_\ell, N\left(\frac{|C_\ell|}{|C|}\right), d)\}$
    Return $\Xi$
end procedure
```

Results: SBI

- Fractional knapsack problem under complex $\mathcal{P}(C \mid X)$:

$$w^*(x) := \min_{w,\mathcal{U}} \max_{\widehat{c}\in\mathcal{U}(x)} \quad -\widehat{c}^T w \tag{11}$$

$$\text{s.t.} w \in [0,1]^n, p^T w \leq B, \mathcal{P}_{X,C}(C \in \mathcal{U}(X)) \geq 1-\alpha,$$

Results: SBI

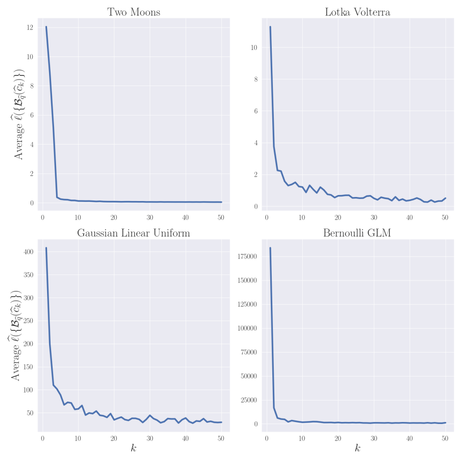


Figure 3: Average volume estimates $\widehat{\ell}(\{\mathcal{B}_{\widehat{q}}(\widehat{c}_k^{(i)})\})$ over $x^{(i)} \in \mathcal{D}_{\mathcal{C}_2}$ across SBI benchmarks.

Results: SBI

- RPs are not unique
- RP *objective minimum*, however, is unique
- Suboptimality can be assessed by measuring

$$\Delta(\Xi, \widehat{\Xi}) := \mathbb{E}_{C \sim U(\mathcal{C}(x))}\left[d(C, \widehat{\Xi}) - d(C, \Xi)\right]. \qquad (12)$$

Results: SBI

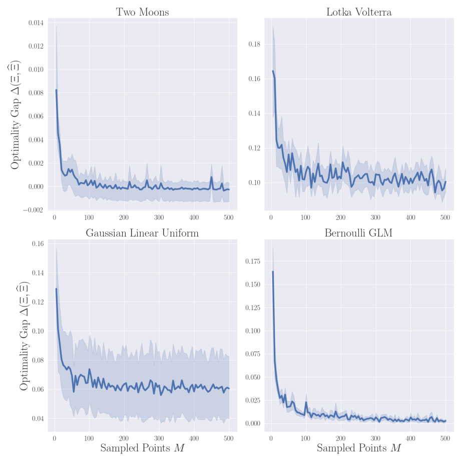


Figure 4: Suboptimality of the approximate representative points $\Delta(\Xi, \widehat{\Xi})$ decreases over increased sampling from the conformal prediction region.

Results: SBI

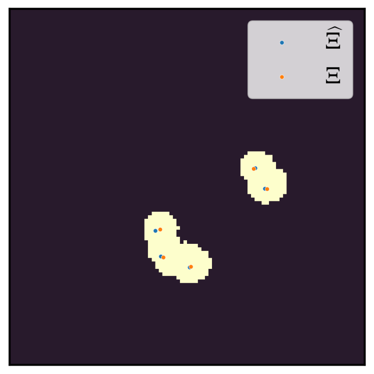


Figure 5: Recovery of exact RPs for two moons task

Results: SBI

	Box	PTC-B	Ellipsoid	PTC-E	CPO
Gaussian Uniform	0.94	0.96	0.95	0.95	0.95
Gaussian Mixture	0.95	0.93	0.94	0.93	0.94
Bernoulli GLM	0.96	0.95	0.95	0.94	0.94
Lotka Volterra	0.95	0.96	0.94	0.94	0.95
SIR	0.94	0.95	0.93	0.95	0.93
Two Moons	0.93	0.94	0.94	0.94	0.96

Results: SBI

	Box	PTC-B	Ellipsoid	PTC-E	CPO	Nominal
Gaussian Uniform	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	**-0.27 (0.35)**	**-0.43 (0.4)**	*-4.48 (0.56)*
Gaussian Mixture	0.0 (0.0)	**-6.6 (1.67)**	0.0 (0.0)	**-7.38 (1.78)**	**-7.77 (1.87)**	*-11.66 (1.23)*
Bernoulli GLM	0.0 (0.0)	**-0.18 (0.49)**	0.0 (0.0)	**-0.06 (0.25)**	**-0.18 (0.37)**	*-3.53 (0.27)*
Lotka Volterra	**-0.52 (0.02)**	-0.05 (0.24)	-0.02 (0.0)	-0.22 (0.18)	**-0.68 (0.26)**	*-1.88 (0.01)*
SIR	-0.16 (0.02)	-0.22 (0.09)	-0.08 (0.01)	-0.22 (0.06)	**-0.38 (0.05)**	*-0.52 (0.02)*
Two Moons	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	**-0.15 (0.11)**	*-0.38 (0.01)*

Results: Traffic Routing

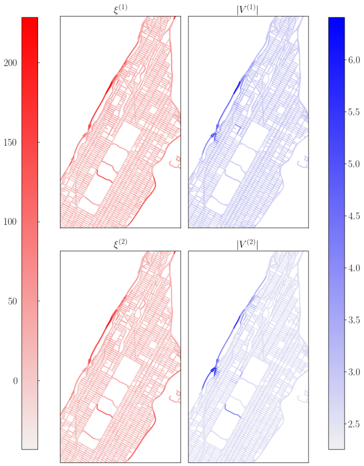


Figure 6: Two RPs for $\mathcal{C}(x)$ for travel time prediction (left) and the extents of their Voronoi cells (right).