Data-Driven Conditional Robust Optimization

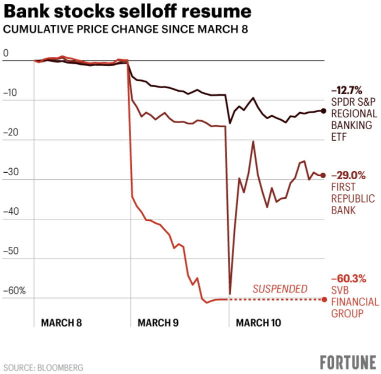
Abhilash Chenreddy[1] Nymisha Bandi[2] Erick Delage[1]

[1]GERAD, MILA, and HEC Montréal & [2]McGill University
Montréal, Canada

Optimization Seminar, Princeton University
March 27th, 2025

Motivating example

- Returns of different assets are unknown but may depend on historical returns, economic factors, investor sentiments via social media.
- Portfolio manager can formulate an allocation problem to minimize the value-at-risk (VaR) of the portfolio while preserving an expected return above a given target.



What is contextual stochastic optimization?

- Optimization problems arising in practice almost always involve **unknown parameters** $\xi \in \mathbb{R}^{m_\xi}$
- Oftentimes, there is a relationship between unknown parameters and some observable **contextual data** $\psi \in \mathbb{R}^{m_\psi}$

What is contextual stochastic optimization?

- Optimization problems arising in practice almost always involve **unknown parameters** $\xi \in \mathbb{R}^{m_\xi}$
- Oftentimes, there is a relationship between unknown parameters and some observable **contextual data** $\psi \in \mathbb{R}^{m_\psi}$

- **Contextual Optimization**:
  - Optimizes a policy, $\boldsymbol{x} : \mathbb{R}^{m_\psi} \to \mathcal{X}$
    - I.e., action $x \in \mathcal{X}$ is adapted to the observed context ψ
  - **Contextual Stochastic Optimization** problem minimizes the expected cost of running the policy over the joint distribution of (ψ, ξ):

$$\min_{\boldsymbol{x}(\cdot)} \mathbb{E}[c(\boldsymbol{x}(\psi), \xi)] \Leftrightarrow \quad \boldsymbol{x}^*(\psi) \in \operatorname*{arg\,min}_{x \in \mathcal{X}} \mathbb{E}[c(x, \xi) | \psi] \text{ a.s.}$$

(see survey Sadana et al. [2024])

What is conditional robust optimization?

- We introduce a novel **Contextual Robust Optimization** paradigm for solving contextual optimization problems in a risk-averse setting:

$$\text{(Robust-CO)} \qquad \min_{\boldsymbol{x}(\cdot)} \max_{\psi \in \mathcal{V}, \xi \in \mathcal{U}(\psi)} c(\boldsymbol{x}(\psi), \xi)$$

  where $\mathcal{U}(\psi)$ is a **conditional uncertainty set** designed to contain with high probability the realization of ξ conditionally on observing ψ.

What is conditional robust optimization?

- We introduce a novel **Contextual Robust Optimization** paradigm for solving contextual optimization problems in a risk-averse setting:

$$\text{(Robust-CO)} \qquad \min_{\boldsymbol{x}(\cdot)} \max_{\psi\in\mathcal{V},\xi\in\mathcal{U}(\psi)} c(\boldsymbol{x}(\psi),\xi)$$

  where $\mathcal{U}(\psi)$ is a **conditional uncertainty set** designed to contain with high probability the realization of ξ conditionally on observing ψ.
- A weak interchangeability property states:

$$\boldsymbol{x}^*(\cdot) \in \arg\min_{\boldsymbol{x}(\cdot)} \max_{\psi\in\mathcal{V},\xi\in\mathcal{U}(\psi)} c(\boldsymbol{x}(\psi),\xi)$$

$$\Leftarrow \quad \boldsymbol{x}^*(\psi) \in \underbrace{\arg\min_{x\in\mathcal{X}} \max_{\xi\in\mathcal{U}(\psi)} c(x,\xi)}_{\textbf{Conditional Robust Optimization}\text{ (CRO)}} \quad , \forall\, \psi \in \mathcal{V}$$

Desirable coverage properties for $\mathcal{U}(\psi)$

The field of conformal prediction identifies two important properties for conditional uncertainty sets

- Marginal coverage property: $\mathbb{P}(\xi \in \mathcal{U}(\psi)) \geq 1 - \epsilon$
- Conditional coverage property: $\mathbb{P}(\xi \in \mathcal{U}(\psi)|\psi) \geq 1 - \epsilon$ a.s.
- Conditional coverage $\Rightarrow$ Marginal coverage

E.g., target coverage $1 - \epsilon = 90\%$:

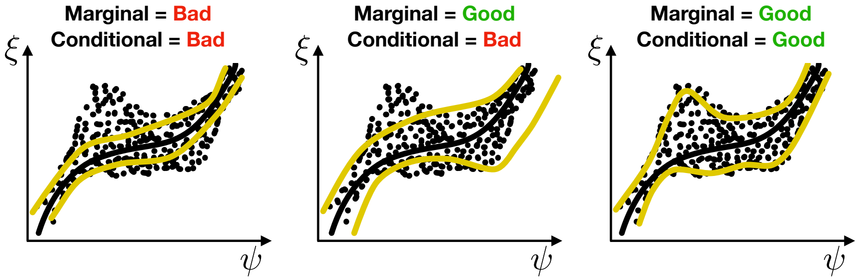


Image from Angelopoulos and Bates, A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification, CoRR, 2021.

Related work in operations research literature

- Contextual Stochastic Optimization:
  - Hannah et al. [2010], Bertsimas and Kallus [2020], ...: Conditional distribution estimation used to formulate and solve the CSO problem.
  - Donti et al. [2017], Elmachtoub and Grigas [2022], ...: End-to-end paradigm applied to solve the data driven CSO problem.
- Distributionally Robust CSO:
  - Bertsimas et al. [2022], McCord [2019], Wang and Jacquillat [2020], Kannan et al. [2020]: DRO approaches with ambiguity sets centered at the estimated conditional distribution
- Data-driven Robust Optimization:
  - Johnstone and Cox [2021], Goerigk and Kurtz [2023]: learns a "non-contextual" uncertainty set using conformal prediction and NN.
  - Ohmori [2021], Sun et al. [2023], Blanquero et al. [2023]: calibrates a set to cover the realizations of a calibrated conditional distribution
  - Sun et al. [2023], Patel et al. [2024] derive contextual sets using conformal prediction
  - Wang et al. [2023] learns a non-contextual set using a task-based obj. with marginal feasibility guarantees

Presentation overview

Outline

Deep Data-Driven Robust Optimization (DDDRO)

- Classic **non-contextual** RO model is written as

$$\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}} c(x, \xi),$$

Deep Data-Driven Robust Optimization (DDDRO)

- Classic **non-contextual** RO model is written as

$$\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}} c(x, \xi),$$

- Goerigk and Kurtz [2023] describe the uncertainty set $\mathcal{U}$ in the form,

$$\mathcal{U}(W, R) = \left\{ \xi \in \mathbb{R}^{m_\xi} : \|f_W(\xi) - \bar{f}_0\| \leq R \right\},$$

where $f_W : \mathbb{R}^{m_\xi} \to \mathbb{R}^d$ is a deep neural network (DNN).

Deep Data-Driven Robust Optimization (DDDRO)

- Classic **non-contextual** RO model is written as

$$\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}} c(x, \xi),$$

- Goerigk and Kurtz [2023] describe the uncertainty set $\mathcal{U}$ in the form,

$$\mathcal{U}(W, R) = \left\{ \xi \in \mathbb{R}^{m_\xi} : \|f_W(\xi) - \bar{f}_0\| \leq R \right\},$$

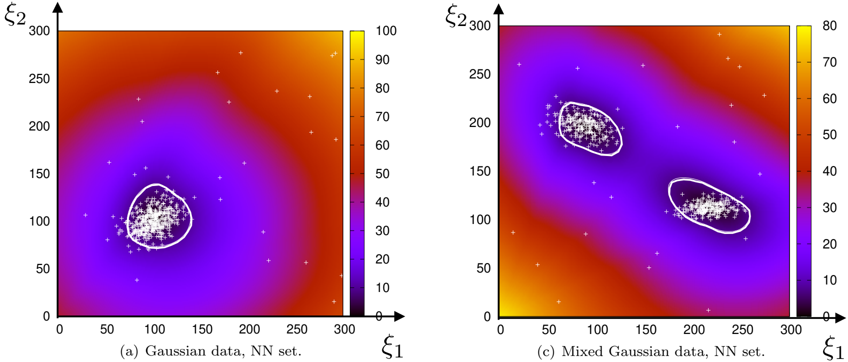
where $f_W : \mathbb{R}^{m_\xi} \to \mathbb{R}^d$ is a deep neural network (DNN).

- Given a dataset $\mathcal{D}_\xi = \{\xi_1, \xi_2 \ldots \xi_N\}$, $\mathcal{U}$ is designed by training a NN to minimize the one-class classification loss

$$\min_W \frac{1}{N} \sum_{i=1}^{N} \|f_W(\xi_i) - \bar{f}_0\|^2,$$

where $\bar{f}_0 := (1/N) \sum_{i \in [N]} f_{W_0}(\xi_i)$ and the radius R is calibrated for $1 - \epsilon$ coverage on the data set.

Illustrative examples



(a) Gaussian data, NN set.

(c) Mixed Gaussian data, NN set.

Images from Goerigk and Kurtz. Data-driven robust optimization using deep neural networks. Computers and Operational Research, 151(C), 2023

Solving robust optimization with deep uncertainty sets

- When using piecewise affine activation functions, $\mathcal{U}(W,R)$ can be represented as:

$$\mathcal{U}(W,R) := \left\{ \xi \;\middle|\; \begin{array}{c} \exists u \in \{0,1\}^{d\times K\times L},\ \zeta \in \mathbb{R}^{d\times L},\ \phi \in \mathbb{R}^{d\times L} \\ \sum_{k=1}^{K} u_j^{k,\ell} = 1,\ \forall j,\ell \\ \phi^1 = W^1\xi \\ \zeta_j^\ell = \sum_{k=1}^{K} u_j^{k,\ell} a_k^\ell \phi_j^\ell + \sum_{k=1}^{K} u_j^{k,\ell} b_k^\ell,\ \forall j,\ell \\ \phi^\ell = W^\ell \zeta^{\ell-1},\ \forall \ell \geq 2 \\ \sum_{k=1}^{K} u_j^{k,\ell} \underline{\alpha}_k^\ell \leq \phi_j^\ell \leq \sum_{k=1}^{K} u_j^{k,\ell} \overline{\alpha}_k^\ell,\ \forall j,\ell \\ \|\zeta^L - \bar{f}_0\| \leq R \end{array} \right\}$$

- The problem $\max_{\xi\in\mathcal{U}(W,R)} c(x,\xi)$ can therefore be formulated as a mixed-integer second order cone program when $c(x,\xi)$ is linear.

Solving robust optimization with deep uncertainty sets

- When using piecewise affine activation functions, $\mathcal{U}(W,R)$ can be represented as:

$$\mathcal{U}(W,R) := \left\{ \xi \;\middle|\; \begin{array}{c} \exists u \in \{0,1\}^{d\times K\times L},\ \zeta \in \mathbb{R}^{d\times L},\ \phi \in \mathbb{R}^{d\times L} \\ \sum_{k=1}^K u_j^{k,\ell} = 1,\ \forall j,\ell \\ \phi^1 = W^1\xi \\ \zeta_j^\ell = \sum_{k=1}^K u_j^{k,\ell} a_k^\ell \phi_j^\ell + \sum_{k=1}^K u_j^{k,\ell} b_k^\ell,\ \forall j,\ell \\ \phi^\ell = W^\ell \zeta^{\ell-1},\ \forall \ell \geq 2 \\ \sum_{k=1}^K u_j^{k,\ell}\underline{\alpha}_k^\ell \leq \phi_j^\ell \leq \sum_{k=1}^K u_j^{k,\ell}\overline{\alpha}_k^\ell,\ \forall j,\ell \\ \|\zeta^L - \bar{f}_0\| \leq R \end{array} \right\}$$

- The problem $\max_{\xi\in\mathcal{U}(W,R)} c(x,\xi)$ can therefore be formulated as a mixed-integer second order cone program when $c(x,\xi)$ is linear.
- This can be integrated in a cutting plane method for solving the RO:

$$\begin{aligned} \min_{x\in\mathcal{X},t} \quad & t \\ \text{subject to} \quad & c(x,\xi) \leq t,\ \forall \xi \in \mathcal{U}' \subset \mathcal{U}(W,R) \end{aligned}$$

Outline

Deep Cluster then Classify (DCC)

- We use $\mathcal{D} := \{(\psi_1, \xi_1), \ldots, (\psi_N, \xi_N)\}$ to design data-driven conditional uncertainty sets $\mathcal{U}(\psi)$.
- This approach reduces the side-information ψ to a set of K different clusters and designs customized sets, i.e., $\mathcal{U}(\psi) := \mathcal{U}_{a(\psi)}$
  - $a : \mathbb{R}^{m_\psi} \to [K]$ is a trained K-class cluster assignment function
  - Each $\mathcal{U}_k$, for $k = 1, \ldots, K$, is an uncertainty sets for ξ calibrated on the dataset $\mathcal{D}_\xi^k := \cup_{(\psi,\xi)\in\mathcal{D}:a(\psi)=k}\{\xi\}$ as in Goerigk and Kurtz [2023].

Deep clustering using auto-encoder/decoder networks

We use an auto-encoder and decoder network to identify $a(\cdot)$,

$$\mathcal{L}^1(V,\theta) := \frac{1-\alpha_K}{N}\sum_{i=1}^{N}\|g_{V_D}(g_{V_E}(\psi_i))-\psi_i\|^2 + \frac{\alpha_K}{N}\sum_{i=1}^{N}\|g_{V_E}(\psi_i)-\theta^{a(\psi_i)}\|^2\,,$$

where

$$a(\psi) := \operatorname*{argmin}_{k\in[K]}\|g_{V_E}(\psi)-\theta^k\|$$

and V_E and V_D are the network parameters.

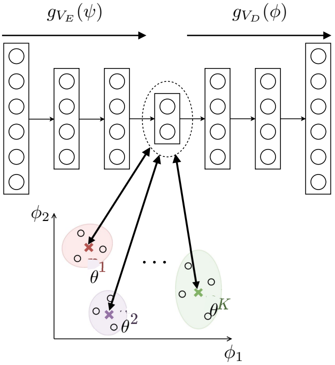


Image adapted from Fard et al. Deep k-means: Jointly clustering with k-means and learning representations. Pattern Recognition Letters, 138:185–192, 2020

Integrated DCC addresses shortcoming of DCC

1. DCC fails to tackle the conditional uncertainty set learning problem as a whole
   - Solution: IDCC optimizes V_E, V_D, θ, and $\{W^k\}_{k=1}^K$ jointly using a loss function that trades-off between the objectives used for clustering and each of the K versions of one-class classifiers

Integrated DCC addresses shortcoming of DCC

1. DCC fails to tackle the conditional uncertainty set learning problem as a whole
   - Solution: IDCC optimizes V_E, V_D, θ, and $\{W^k\}_{k=1}^K$ jointly using a loss function that trades-off between the objectives used for clustering and each of the K versions of one-class classifiers

2. DCC struggles for cases where clear separation of clusters isn't possible.
   - Solution: IDCC trains a parameterized random assignment policy $\tilde{a}(\psi) \sim \pi(\psi)$:

$$\mathbb{P}(\tilde{a}(\psi) = k) = \pi_k(\psi) := \frac{\exp\{-\beta\|g_V(\psi) - \theta^k\|^2\}}{\sum_{k'=1}^K \exp\{-\beta\|g_V(\psi) - \theta^{k'}\|^2\}}$$

   The **random** uncertainty set is $\tilde{\mathcal{U}}(\psi) := \mathcal{U}(W^{\tilde{a}(\psi)}, R^{\tilde{a}(\psi)})$

Experiments

Robust portfolio optimization with market data

- Decision model:

$$x^*(\psi) := \arg \min_{x:\sum_{i=1}^n x_i=1,\ x\geq 0} \max_{\xi\in\mathcal{U}(\psi)} -\xi^\intercal x$$

  which captures the need to invest one unit of wealth among the available assets while minimizing risk exposure.

Experiments

Robust portfolio optimization with market data

- Decision model:

$$x^*(\psi) := \arg \min_{x:\sum_{i=1}^n x_i=1,\, x\geq 0} \max_{\xi\in\mathcal{U}(\psi)} -\xi^\intercal x$$

  which captures the need to invest one unit of wealth among the available assets while minimizing risk exposure.

- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).

Experiments

Robust portfolio optimization with market data

- Decision model:

$$x^*(\psi) := \arg \min_{x:\sum_{i=1}^n x_i=1,\ x\geq 0} \max_{\xi\in\mathcal{U}(\psi)} -\xi^\intercal x$$

  which captures the need to invest one unit of wealth among the available assets while minimizing risk exposure.

- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).
- Market data from Yahoo! Finance: 70 different stocks during period from 01/01/2012 to 31/12/2019 (2017-2019 reserved for test).

Experiments

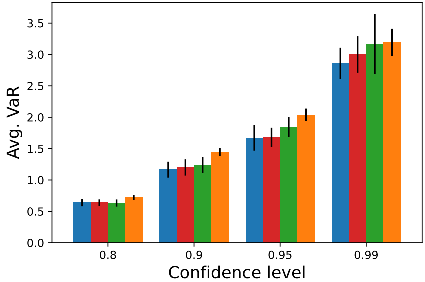
Robust portfolio optimization with market data

- Decision model:

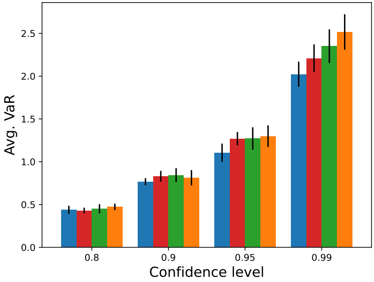
$$x^*(\psi) := \arg \min_{x:\sum_{i=1}^n x_i=1,\, x\geq 0} \max_{\xi\in\mathcal{U}(\psi)} -\xi^\mathsf{T} x$$

  which captures the need to invest one unit of wealth among the available assets while minimizing risk exposure.
- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).
- Market data from Yahoo! Finance: 70 different stocks during period from 01/01/2012 to 31/12/2019 (2017-2019 reserved for test).
- Performance metric: out-of-sample VaR of $-\xi^T x(\psi)$

Portfolio optimization: Comparison of avg. VaR across portfolio simulations



(a) 2018



(b) 2019

Ellipsoid DDDRO DCC IDCC

Outline

Task-based CRO

- The IDCC approach suffers from two issues:
  1. Training is done solely based on total variation measurements, disregarding entirely the out-of-sample performance of the solution obtained from robust optimization.

Task-based CRO

- The IDCC approach suffers from two issues:
  1. Training is done solely based on total variation measurements, disregarding entirely the out-of-sample performance of the solution obtained from robust optimization.
  2. While the calibration process encourages **marginal coverage** by making the coverage accurate for each cluster:

$$\mathbb{P}(\xi \in \mathcal{U}(\psi) | \tilde{a}(\psi) = k) \geq 1 - \epsilon \, \forall k \;\checkmark \;\Rightarrow\; \mathbb{P}(\xi \in \mathcal{U}(\psi)) \geq 1 - \epsilon \;\checkmark$$

it does not promote **conditional coverage** over all ψ:

$$\mathbb{P}(\xi \in \mathcal{U}(\psi) | \psi) \geq 1 - \epsilon \text{ a.s. } \times$$

Task-based CRO

- The IDCC approach suffers from two issues:
  1. Training is done solely based on total variation measurements, disregarding entirely the out-of-sample performance of the solution obtained from robust optimization.
  2. While the calibration process encourages **marginal coverage** by making the coverage accurate for each cluster:

$$\mathbb{P}(\xi \in \mathcal{U}(\psi) | \tilde{a}(\psi) = k) \geq 1 - \epsilon \, \forall k \; ✔ \; \Rightarrow \; \mathbb{P}(\xi \in \mathcal{U}(\psi)) \geq 1 - \epsilon \; ✔$$

  it does not promote **conditional coverage** over all ψ:

$$\mathbb{P}(\xi \in \mathcal{U}(\psi) | \psi) \geq 1 - \epsilon \text{ a.s. } ✖$$

- In this next part, we propose **Task-based Conditional Robust Optimization** that promotes **performance** and **conditional coverage**.

Estimate-then-Optimize with continuous adaptation

- We consider a continuously adapted conditional ellipsoidal set:

$$\mathcal{U}_\theta(\psi) := \{\ \xi \in \mathbb{R}^{m_\xi} : (\xi - \mu_\theta(\psi))^T \Sigma_\theta^{-1}(\psi)(\xi - \mu_\theta(\psi)) \leq R_\theta \},$$

Estimate-then-Optimize with continuous adaptation

- We consider a continuously adapted conditional ellipsoidal set:

$$\mathcal{U}_\theta(\psi) := \{\, \xi \in \mathbb{R}^{m_\xi} : (\xi - \mu_\theta(\psi))^T \Sigma_\theta^{-1}(\psi)(\xi - \mu_\theta(\psi)) \leq R_\theta \}\,,$$

- Given a data set $\mathcal{D} = \{(\psi_1, \xi_1), (\psi_2, \xi_2) \ldots (\psi_N, \xi_N)\}$, an estimate-then-optimize (ETO) approach takes the form:

$$\xrightarrow{\mathcal{D}} \boxed{\begin{array}{c}\text{Estimation:}\\ \min_\theta \mathcal{L}^{\xi|\psi}_{\text{NLL}}(f_\theta(\cdot), \mathcal{D})\end{array}} \xrightarrow{\mathcal{U}_{\theta^*}(\cdot)} \boxed{\begin{array}{c}\text{Robust optimization:}\\ x^*(\psi) := \arg\min_{x \in \mathcal{X}}\\ \max_{\xi \in \mathcal{U}_{\theta^*}(\psi)} c(x, \xi)\end{array}} \xrightarrow{x^*(\cdot)}$$

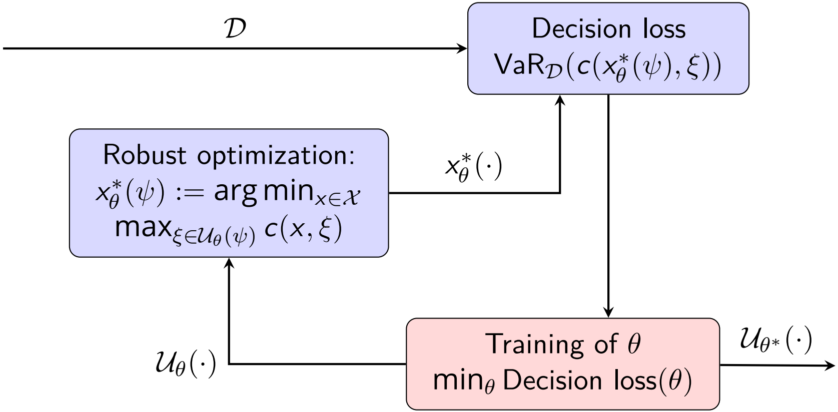
where $\mathcal{L}^{\xi|\psi}_{\text{NLL}}$ is the negative log likelihood for a conditional Gaussian density estimator (see Barratt and Boyd [2021]):

$$\xi \sim f_\theta(\psi) := \mathcal{N}(\mu_\theta(\psi), \Sigma_\theta(\psi))$$

and R_θ s.t. $\mathbb{P}_{\mathcal{D}}(\xi \in \mathcal{U}_\theta(\psi)) = 1 - \epsilon$

(Single) Task-based Set (TbS) training

A task-based approach learns the estimator by trying to minimize the decision loss, e.g. the portfolio risk based on VaR



Decision loss relaxation and derivatives

- Decision loss $\text{VaR}_{\mathcal{D}}(c(x^*_\theta(\psi), \xi))$ suffers from multiple local optima.

Figure 3: Simulation-based trade risk profile

Image from Mausser and Rosen, Beyond VaR: from measuring risk to managing risk, CIFEr, 1999.

Decision loss relaxation and derivatives

- Decision loss $\text{VaR}_{\mathcal{D}}(c(x_\theta^*(\psi), \xi))$ suffers from multiple local optima.

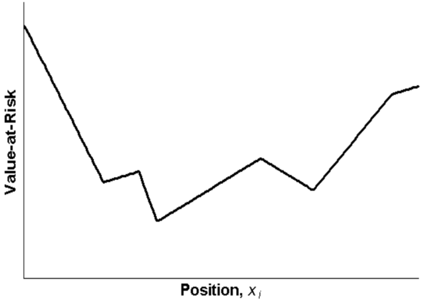


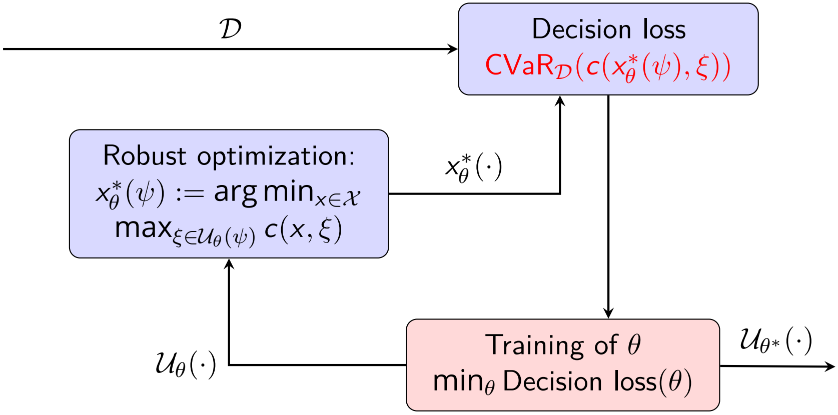
Figure 3: Simulation-based trade risk profile

Image from Mausser and Rosen, Beyond VaR: from measuring risk to managing risk, CIFEr, 1999.

- We therefore replace it with upper bound $\text{CVaR}_{\mathcal{D}}(c(x_\theta^*(\psi), \xi))$.

$$\frac{\partial \text{CVaR}_{\text{i}\sim N}(y_{\text{i}})}{\partial y_i} = \upsilon_i(y) \text{ with } \boldsymbol{\upsilon}(\boldsymbol{y}) \in \underset{\boldsymbol{\upsilon}\in\mathbb{R}_+^M:\mathbb{1}^T\boldsymbol{\upsilon}=1,\boldsymbol{\upsilon}\leq((1-\alpha)N)^{-1}}{\text{argmax}} \boldsymbol{\upsilon}^T\boldsymbol{y}$$

Decision loss relaxation and derivatives



Robust optimization reformulation and derivatives

- We assume that $c(x, \xi)$ is convex in x and concave in ξ, while $\mathcal{X}$ is a convex set.

Robust optimization reformulation and derivatives

- We assume that $c(x, \xi)$ is convex in x and concave in ξ, while $\mathcal{X}$ is a convex set.
- Using Fenchel duality, one can follow Ben-Tal et al. [2015] to reformulate the robust optimization problem as:

$$x_\theta^*(\psi) := \arg\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}_\theta(\psi)} c(x, \xi) = \arg\min_{v, x \in \mathcal{X}} \underbrace{\delta^*(v|\mathcal{U}_\theta(\psi)) - c_*(x, v)}_{f(x, v, \mathcal{U}_\theta(\psi))}$$

where the support function

$$\delta^*(v|\mathcal{U}_\theta(\psi)) := \sup_{\xi \in \mathcal{U}_\theta(\psi)} \xi^T v = \mu^T v + \sqrt{v^T \Sigma v}$$

while the partial concave conjugate function is defined as

$$c_*(x, v) := \inf_{\xi} v^T \xi - c(x, \xi)$$

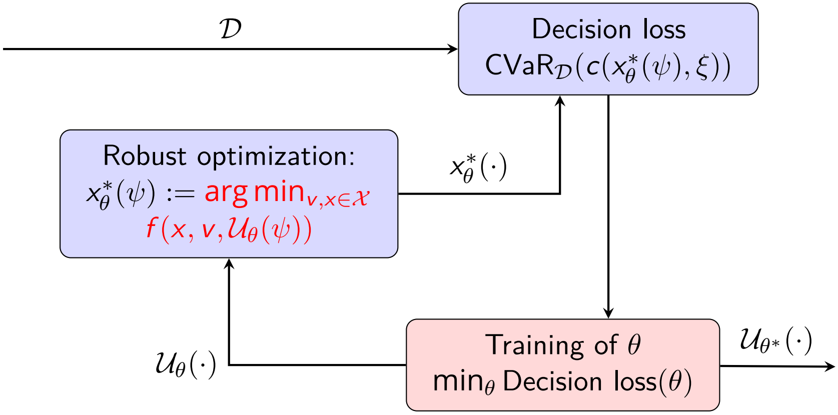
Robust optimization reformulation and derivatives

- We assume that $c(x,\xi)$ is convex in x and concave in ξ, while $\mathcal{X}$ is a convex set.
- Using Fenchel duality, one can follow Ben-Tal et al. [2015] to reformulate the robust optimization problem as:

$$x^*_\theta(\psi) := \arg\min_{x\in\mathcal{X}} \max_{\xi\in\mathcal{U}_\theta(\psi)} c(x,\xi) = \arg\min_{v,x\in\mathcal{X}} \underbrace{\delta^*(v|\mathcal{U}_\theta(\psi)) - c_*(x,v)}_{f(x,v,\mathcal{U}_\theta(\psi))}$$

- The derivatives of $x^*_\theta(\psi) := \arg\min_{v,x\in\mathcal{X}} f(x,v,\mathcal{U}_\theta(\psi))$ w.r.t. θ can be obtained using implicit differentiation (see Blondel et al. [2022])

Robust optimization reformulation and derivatives



Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

	ETO-ACPS	ETO-DbS	TbS	
Avg. CVaR	1.69 ± 0.05	1.64 ± 0.07	**1.03** ± 0.10	
Avg. VaR	1.12 ± 0.04	1.07 ± 0.02	**0.72** ± 0.07	
Avg. marginal cov.	**91%** $\pm 1.4\%$	$85\% \pm 7.8\%$	$23\% \pm 6.1\%$	

Second-task: Conditional coverage

Lemma

An uncertainty set $\mathcal{U}_\theta(\psi)$ has an a.s. conditional coverage of $1-\epsilon$ if and only if

$$\mathcal{L}_{CC}(\theta) := \mathbb{E}[\,(\mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi) - (1-\epsilon))^2\,] = 0$$

Second-task: Conditional coverage

Lemma

An uncertainty set $\mathcal{U}_\theta(\psi)$ has an a.s. conditional coverage of $1-\epsilon$ if and only if

$$\mathcal{L}_{CC}(\theta) := \mathbb{E}[\,(\mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi) - (1-\epsilon))^2\,] = 0$$

$\mathcal{L}_{\mathsf{CC}}(\theta)$ can be approximated using:

$$\widehat{\mathcal{L}}_{\mathsf{CC}}(\theta) := \mathbb{E}_{\mathcal{D}}[(g_{\phi^*(\theta)}(\psi) - (1-\epsilon))^2]$$

where $g_{\phi^*(\theta)}(\psi) \approx \mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi)$ is obtained using logistic regression of membership variable $y(\psi, \xi; \theta) := \mathbb{1}\{\xi \in \mathcal{U}_\theta(\psi)\}$ on ψ.

Second-task: Conditional coverage

Lemma

An uncertainty set $\mathcal{U}_\theta(\psi)$ has an a.s. conditional coverage of $1-\epsilon$ if and only if

$$\mathcal{L}_{CC}(\theta) := \mathbb{E}[\,(\mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi) - (1-\epsilon))^2\,] = 0$$

$\mathcal{L}_{\mathrm{CC}}(\theta)$ can be approximated using:

$$\widehat{\mathcal{L}}_{\mathrm{CC}}(\theta) := \mathbb{E}_{\mathcal{D}}[(g_{\phi^*(\theta)}(\psi) - (1-\epsilon))^2]$$

where $g_{\phi^*(\theta)}(\psi) \approx \mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi)$ is obtained using logistic regression of membership variable $y(\psi,\xi;\theta) := \mathbb{1}\{\xi \in \mathcal{U}_\theta(\psi)\}$ on ψ.

- I.e., letting the augmented data set

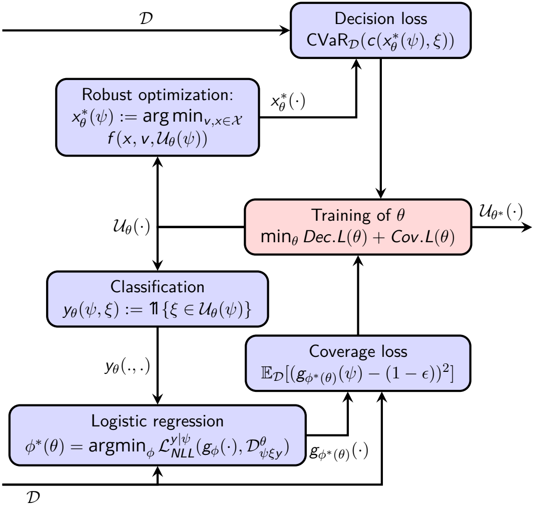
$$\mathcal{D}^\theta_{\psi\xi y} := \{(\psi_1, \xi_1, y(\psi_1, \xi_1; \theta)), \ldots, (\psi_N, \xi_N, y(\psi_N, \xi_N; \theta))\},$$

one solves $\phi^*(\theta) \in \mathrm{argmin}_\phi\, \mathcal{L}^{y|\psi}_{NLL}(g_\phi(\cdot), \mathcal{D}^\theta_{\psi\xi y})$ with

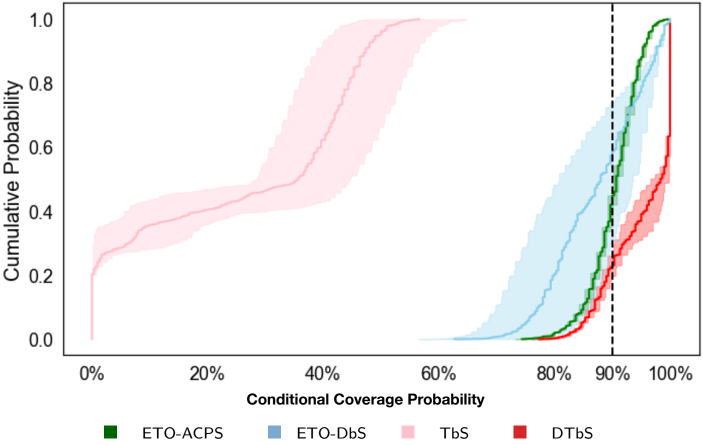
$$g_\phi(\psi) := \frac{1}{1 + \exp^{\phi^T\psi + \phi_0}}$$

Double Task-based Set (DTbS) training

We train $\mathcal{U}_\theta(\psi)$ using the two tasks: produce good decision + produce good conditional coverage:



Comparative study with GMM environment



(See uncertainty set animation (url))

Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

	ETO-ACPS	ETO-DbS	TbS	
Avg. CVaR	1.69 ± 0.05	1.64 ± 0.07	**1.03** ± 0.10	
Avg. VaR	1.12 ± 0.04	1.07 ± 0.02	**0.72** ± 0.07	
Avg. marginal cov.	**91%** $\pm 1.4\%$	$85\% \pm 7.8\%$	$23\% \pm 6.1\%$	

Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

	ETO-ACPS	ETO-DbS	TbS	!! DTbS !!
Avg. CVaR	1.69 ± 0.05	1.64 ± 0.07	**1.03** ± 0.10	**1.08** $\pm$ 0.13
Avg. VaR	1.12 ± 0.04	1.07 ± 0.02	**0.72** ± 0.07	**0.75** $\pm$ 0.10
Avg. marginal cov.	**91%** $\pm 1.4\%$	$85\% \pm 7.8\%$	$23\% \pm 6.1\%$	$92\% \pm 1.5\%$

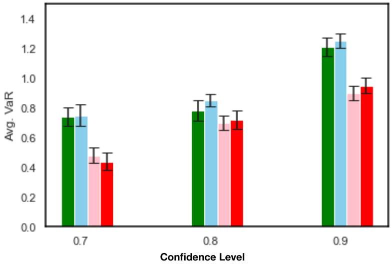
Portfolio optimization with market data

- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).
- Market data from Yahoo! Finance: 70 different stocks during period from 01/01/2012 to 31/12/2019 (2017-2019 reserved for test).
- Target confidence level of 70%, 80%, or 90%

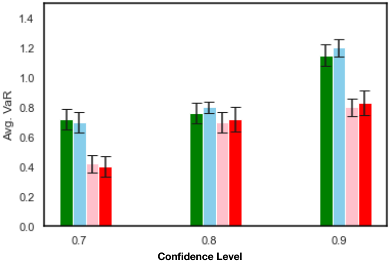
Model	Marginal coverage					
	2018			2019		
	70%	80%	90%	70%	80%	90%
ETO-ACPS	68%	78%	87%	71%	78%	89%
ETO-DbS	59%	75%	87%	61%	76%	86%
TbS	23%	24%	29%	26%	30%	32%
DTbS	71%	80%	93%	69%	78%	92%

Portfolio optimization with market data

- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).
- Market data from Yahoo! Finance: 70 different stocks during period from 01/01/2012 to 31/12/2019 (2017-2019 reserved for test).



(a) 2018



(b) 2019

ETO-ACPS ETO-DbS TbS DTbS

Outline

Concluding remarks

- We introduced a new contextual robust optimization approach for solving risk averse contextual optimization problems.
- In CRO, deep neural networks can be used to:
  - Represent richly structured uncertainty sets, e.g. DDDRO, IDCC
  - Adapt uncertainty set continuously to covariates, e.g. ETO-ACPS,..., DTbS.
- Two types of training procedures: "Estimate-then-optimize" vs. "Task-based"
- Two types of training objectives:
  - Statistical performance: achieving the right marginal/conditional coverage
  - Decision performance: producing decisions that achieve low VaR/CVaR

Thank you

Bibliography I

Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2022.

Shane Barratt and Stephen Boyd. Covariance prediction via convex optimization, 2021.

Aharon Ben-Tal, Dick den Hertog, and Jean-Philippe Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Program.*, 149(1–2):265–299, feb 2015.

Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, 2020.

Dimitris Bertsimas, Christopher McCord, and Bradley Sturt. Dynamic optimization with side information. *European Journal of Operational Research*, 2022.

Rafael Blanquero, Emilio Carrizosa, and Nuria Gómez-Vargas. Contextual uncertainty sets in robust linear optimization. 2023.

Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-López, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and modular implicit differentiation. *Advances in neural information processing systems*, 35:5230–5242, 2022.

Priya Donti, Brandon Amos, and J Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Adam N Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68 (1):9–26, 2022.

Marc Goerigk and Jannis Kurtz. Data-driven robust optimization using deep neural networks. *Computers and Operational Research*, 151(C), 2023.

Bibliography II

Lauren Hannah, Warren Powell, and David Blei. Nonparametric density estimation for stochastic optimization with an observable state variable. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

Chancellor Johnstone and Bruce Cox. Conformal uncertainty sets for robust optimization, 2021.

Rohit Kannan, Güzin Bayraksan, and James R Luedtke. Residuals-based distributionally robust optimization with covariate information. *arXiv preprint arXiv:2012.01088*, 2020.

H. Mausser and D. Rosen. Beyond var: from measuring risk to managing risk. In *Proceedings of the IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering*, 1999.

Christopher George McCord. *Data-driven dynamic optimization with auxiliary covariates*. PhD thesis, Massachusetts Institute of Technology, 2019.

Shunichi Ohmori. A predictive prescription using minimum volume k-nearest neighbor enclosing ellipsoid and robust optimization. *Mathematics*, 9(2):119, 2021.

Yash P Patel, Sahana Rayan, and Ambuj Tewari. Conformal contextual robust optimization. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

Utsav Sadana, Abhilash Chenreddy, Erick Delage, Alexandre Forel, Emma Frejinger, and Thibaut Vidal. A survey of contextual optimization methods for decision-making under uncertainty. *European Journal of Operational Research*, 2024.

Bibliography III

Chunlin Sun, Linyu Liu, and Xiaocheng Li. Predict-then-calibrate: A new perspective of robust contextual lp, 2023.

Irina Wang, Cole Becker, Bart Van Parys, and Bartolomeo Stellato. Learning for robust optimization. *arXiv preprint arXiv:2305.19225*, 2023.

Kai Wang and Alex Jacquillat. From classification to optimization: A scenario-based robust optimization approach. *Available at SSRN 3734002*, 2020.