Conformal Mixed-Integer Constraint Learning with Feasibility Guarantees

Presenter: Ben-Zheng Li

December 21, 2025

Reference: Daniel Ovalle,Lorenz T. Biegler, Ignacio E. Grossmann, Carl D. Laird, and Mateo Dulce Rubio, *Conformal Mixed-Integer Constraint Learning with Feasibility Guarantees*, *arXiv*, 2025.

Background I

- In many real-world optimization problems, **decision variables are mixed continuous and discrete**, and constraints are unknown or hard to model explicitly.
- True constraint depends on an unknown function:

$$h(x) \leq 0, \qquad h(\cdot) \text{ unknown}$$

- Instead, we observe data:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}, \qquad y_i \approx h(x_i)$$

- **Constraint Learning (CL)** [2, 1]:

$$h(x) \;\approx\; \widehat{h}(x)$$

  where $\widehat{h}(\cdot)$ is learned from data and embedded into optimization.

Background II

- Suppose we solve the optimization problem using the learned constraint:

$$\min_{x,z} f(x,z) \quad \text{s.t.} \quad \widehat{h}(x) \le 0,\ (x,z) \in \mathcal{X}$$

- The optimizer typically selects a solution near the **constraint boundary**:

$$\widehat{h}(x^\star) \approx 0$$

- However, the learned constraint is imperfect:

$$\widehat{h}(x) = h(x) + \varepsilon(x)$$

- Even small errors near the boundary can cause infeasibility:

$$\widehat{h}(x^\star) \le 0 \quad \not\Rightarrow \quad h(x^\star) \le 0$$

Literature Review

- **Mixed-Integer Constraint Learning (MICL)** [7, 5]
  - Learn surrogate constraint models from data
  - Embed learned constraints into mixed-integer optimization
  - **Limitation:**
    - High computational cost
    - No formal probabilistic feasibility guarantees

Literature Review

- **Mixed-Integer Constraint Learning (MICL)** [7, 5]
  - Learn surrogate constraint models from data
  - Embed learned constraints into mixed-integer optimization
  - **Limitation:**
    - High computational cost
    - No formal probabilistic feasibility guarantees
- **Trust-Region/Filter-based Optimization** [3]
  - Iteratively refine learned constraints via local sampling
  - Optimize only in regions where models are accurate
  - **Limitation:**
    - Strong assumptions on noise and smoothness
    - Requires access to true constraint evaluations

Literature Review

- **Mixed-Integer Constraint Learning (MICL)** [7, 5]
  - Learn surrogate constraint models from data
  - Embed learned constraints into mixed-integer optimization
  - **Limitation:**
    - High computational cost
    - No formal probabilistic feasibility guarantees
- **Trust-Region/Filter-based Optimization** [3]
  - Iteratively refine learned constraints via local sampling
  - Optimize only in regions where models are accurate
  - **Limitation:**
    - Strong assumptions on noise and smoothness
    - Requires access to true constraint evaluations
- **Conformal-Enhanced Optimization** [4, 6]
  - Use conformal prediction to quantify uncertainty
  - Embed uncertainty sets into optimization problems
  - **Limitation:**
    - Focus on parameters or stochastic constraints
    - Do not directly handle learned constraint functions

Desired Feasibility Guarantee

- Let $x^\star$ be the solution obtained using learned constraints.
- What truly matters is feasibility in the real system:

$$h(x^\star) \le 0$$

- This paper seeks a probabilistic guarantee:

$$\mathbb{P}\big(h(x^\star) \le 0\big) \ge 1 - \alpha$$

- The risk level $\alpha \in (0, 1)$ is user-specified.

Goal: Achieve finite-sample feasibility guarantees *without* specifying uncertainty sets or distributions.

C-MICL

- **C-MICL optimization problem:**

$$\begin{aligned} \min_{x,z} \quad & f(x,z) \\ \text{s.t.} \quad & g(x,z) \le 0, \quad (x,z) \in \mathcal{X}, \\ & \mathcal{C}(x) \subseteq \mathcal{Y} \end{aligned}$$

Key idea: Replace point constraints with conformal prediction sets, and enforce feasibility against the worst case within the set.

Roadmap: Conformal prediction $\rightarrow$ Mixed-integer formulation $\rightarrow$ Finite-sample feasibility guarantee

Conformal Prediction: Constructing $C(x)$

- Goal: construct a prediction set $C(x)$ such that

$$\mathbb{P}\big(h(x) \in C(x)\big) \geq 1 - \alpha$$

  without assuming a distribution for $h(x)$.
- Key idea: use a **calibration dataset** to quantify prediction uncertainty.
- Given a learned model $\widehat{h}(x)$, define a nonconformity score:

$$s_i = \big|y_i - \widehat{h}(x_i)\big|.$$

- Let $q_{1-\alpha}$ be the $(1-\alpha)$-quantile of $\{s_i\}$.

Interpretation: $q_{1-\alpha}$ quantifies how large the prediction error can be with probability at least $1 - \alpha$.

Conformal Prediction: Constructing $C(x)$

- By construction, the quantile $q_{1-\alpha}$ satisfies:

$$\mathbb{P}\Big(\big|h(x) - \widehat{h}(x)\big| \leq q_{1-\alpha}\Big) \geq 1 - \alpha.$$

- This implies:

$$-q_{1-\alpha} \leq h(x) - \widehat{h}(x) \leq q_{1-\alpha}.$$

- Rearranging terms yields:

$$h(x) \in \big[\, \widehat{h}(x) - q_{1-\alpha},\ \widehat{h}(x) + q_{1-\alpha} \,\big].$$

Conformal prediction set:

$$C(x) = \big[\, \widehat{h}(x) - q_{1-\alpha},\ \widehat{h}(x) + q_{1-\alpha} \,\big].$$

Remark. For regression constraints, $C(x)$ is an interval; for classification constraints, $C(x)$ is a label set $C(x) \subseteq \{0, 1\}$.

Embedding Prediction Sets into Optimization

- From conformal prediction, we obtain a prediction set $C(x)$ with coverage guarantee: $\mathbb{P}\big(h(x) \in C(x)\big) \geq 1 - \alpha$.
- To ensure safety, we enforce feasibility for *all* values in the prediction set:

$$\max_{y \in C(x)} y \leq 0.$$

- **Regression constraints.** For interval-valued sets

$$C(x) = [\,\widehat{h}(x) - q,\ \widehat{h}(x) + q\,],$$

  the robust constraint reduces to:

$$\widehat{h}(x) + q \leq 0.$$

Key observation: Conformal prediction transforms uncertain constraints into a tractable *shifted deterministic constraint*.

Remark. For classification constraints, feasibility is enforced by excluding infeasible labels from $C(x)$; the formulation is analogous.

Finite-Sample Feasibility Guarantee

Theorem. Finite-sample feasibility of C-MICL

Let $\mathcal{F}_N$ denote the feasible region of the C-MICL problem, defined as

$$\mathcal{F}_N = \big\{(x,z) \in \mathcal{X} : g(x,z) \leq 0,\ C(x) \subseteq \mathcal{Y}\big\}.$$

Under the conditions of Lemma 3.1 and Assumption 4.1, for any feasible solution $(x', z') \in \mathcal{F}_N$, we have

$$\mathbb{P}\big(h(x') \in \mathcal{Y}\big) \ \geq \ 1 - \alpha.$$

- The probability is taken over the randomness of the calibration data.
- The guarantee holds for **any feasible solution**, not only the optimum.
- The result is distribution-free and holds in finite samples.

Regression vs. Classification C-MICL

	Regression C-MICL	**Classification C-MICL**
Constraint type	Continuous constraint $h(x) \in \mathbb{R}$	Categorical constraint (feasible / infeasible)
Prediction set	Interval-valued: $C(x) = \hat{h}(x) \pm q\,\hat{u}(x)$	Label-valued: $C(x) \subseteq \mathcal{Y}$
Uncertainty modeling	Explicit uncertainty model $\hat{u}(x)$	No separate uncertainty model
Conformal score	Absolute or scaled residuals	Scores computed from logits
Feasibility enforcement	$\hat{h}(x) + q\hat{u}(x) \leq 0$	Exclude infeasible labels from $C(x)$

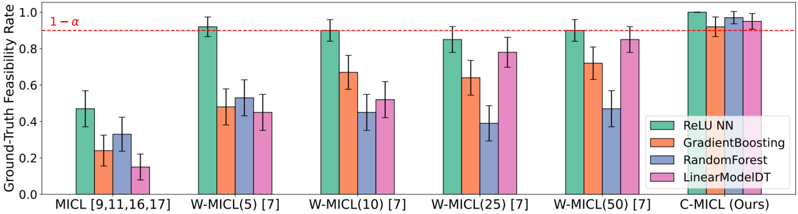
Key takeaway. Regression and classification are two instantiations of the same C-MICL framework and enjoy identical finite-sample feasibility guarantees.

Experimental Setup

- **Tasks**
  - Regression constrained optimization
  - Classification constrained optimization
  - Constraints are unknown and learned from data
- **Baselines**
  - MICL
  - W-MICL
  - C-MICL
- **Prediction models**
  - Neural networks for $\hat{h}(x)$
  - An explicit uncertainty model $\hat{u}(x)$ for regression
- **Evaluation metrics**
  - Feasibility violation rate
  - Objective value
  - Solution optimality gap

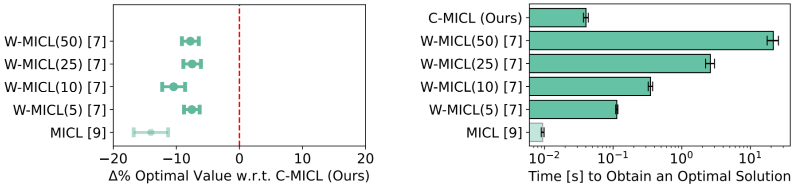
Goal. Evaluate whether C-MICL achieves finite-sample feasibility guarantees while maintaining competitive objective performance.

Main Results: Ground-Truth Feasibility



Key observation. C-MICL consistently achieves the target feasibility level $1-\alpha$ across all base prediction models, while prior MICL variants exhibit substantial violations and high variability.

Objective Quality vs. Computational Efficiency



Key takeaway: C-MICL achieves a favorable trade-off between *feasibility guarantees*, *solution quality*, and *computational efficiency*.

References

[1] Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *Management Science*, 2020.

[2] Priya Donti, Brandon Amos, and Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, 2017.

[3] Author et al. Trust region filter methods for optimization with learned constraints. *SIAM Journal on Optimization*, 2021.

[4] Lin et al. Conformalized optimization. *NeurIPS*, 2022.

[5] Tajmir-Riahi et al. Data-driven constraint learning for optimization. *Operations Research*, 2020.

[6] Zhao et al. Conformal chance-constrained optimization. *ICML*, 2021.

[7] Ruth Misener and Ben Bigler. Constraint learning in mixed-integer optimization. *Computers & Chemical Engineering*, 2021.