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, **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)** [3, 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:

$$\min_x f(x) \quad \text{s.t.} \quad \widehat{h}(x) \leq 0$$

- 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:

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

Key insight: Optimization does not average out errors — it *amplifies* them.

Related Approaches and Their Limitations

- **Naive Constraint Learning** [3]
  - Core idea: learn a single surrogate constraint from data $h(x) \approx \widehat{h}(x)$
  - Optimization is performed using: $\widehat{h}(x) \leq 0$
  - Limitation: $\hat{h}(x) \leq 0$ does not guarantee true feasibility
- **Ensemble-based MICL** [4]
  - Core idea: learn multiple constraint models $\{\widehat{h}_j(x)\}_{j=1}^{M}$
  - Enforce feasibility for a subset of models: $\hat{h}_j(x) \leq 0$ for $j \in \mathcal{J}$
  - Limitation:
    - High computational cost
    - No formal probabilistic feasibility guarantee
- **Robust/Chance-Constrained Optimization** [2]
  - Core idea: protect against uncertainty in constraints

$$h(x,\xi) \leq 0 \quad \forall \xi \in \mathcal{U} \quad \text{or} \quad \mathbb{P}\big(h(x,\xi) \leq 0\big) \geq 1-\alpha$$

  - Limitation:
    - Require explicit uncertainty sets or distributions
    - Often overly conservative or hard to calibrate

Desired Feasibility Guarantee

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

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

- This paper seeks a probabilistic guarantee:

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

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

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

C-MICL: Core Idea

- **Problem with naive constraint learning:** optimization enforces a *point prediction* $\widehat{h}(x) \leq 0$, which is sensitive to model error.
- **Key idea:** replace point predictions with a *prediction set* $C(x)$ such that

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

- **Safe optimization:** enforce feasibility for all values in the set

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

Core idea: Use **conformal prediction** to construct prediction sets with finite-sample coverage guarantees.

Roadmap: Conformal prediction $\longrightarrow$ Mixed-integer formulation $\longrightarrow$ 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:

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

- This implies:

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

- 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

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

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.

Proof Sketch

- **Step 1: Marginal coverage via conformal prediction.** For any fixed decision x, conformal prediction guarantees

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

- **Step 2: Feasible solutions are restricted by prediction sets.** The C-MICL formulation enforces that for any feasible solution $(x, z) \in \mathcal{F}_N$, the entire prediction set satisfies

$$C(x) \subseteq \mathcal{Y}.$$

- **Step 3: Optimization cannot exploit model uncertainty.** Since feasibility is enforced at the *set level*, the optimizer cannot choose a decision x whose prediction set overlaps the infeasible region.
- **Conclusion.** Combining coverage and set-level feasibility, any feasible solution $(x', z') \in \mathcal{F}_N$ satisfies $\mathbb{P}\big(h(x') \in \mathcal{Y}\big) \geq 1 - \alpha$.

Key intuition: enforcing feasibility on the entire prediction set decouples the optimization decision from estimation uncertainty.

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 $\hat{u}(x)$ (heteroskedastic)	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 $C(x)$

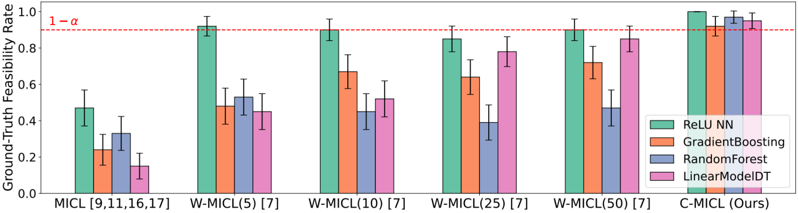
Key takeaway. Regression and classification are two instantiations of the same C-MICL framework and enjoy identical finite-sample feasibility guarantees.

Experimental Setup

- **Tasks**
  - Regression constrained optimization
  - Classification constrained optimization
  - Constraints are unknown and learned from data
- **Baselines**
  - MICL: point-estimate constraint learning
  - W-MICL: ensemble-based heuristic variants
  - C-MICL: conformal prediction with feasibility guarantees
- **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

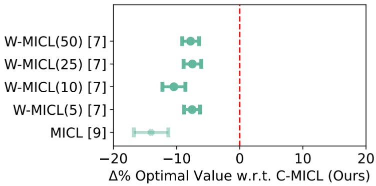
Goal. Evaluate whether C-MICL achieves finite-sample feasibility guarantees while maintaining competitive objective performance.

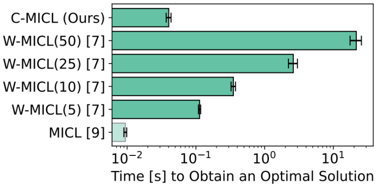
Main Results: Ground-Truth Feasibility



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] Dimitris Bertsimas and Melvyn Sim. The theory of robust optimization. *Operations Research*, 2011.

[3] 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.

[4] Radhika Mistry, Sidhant Misra, and Juan Pablo Vielma. Mixed-integer constraint learning. *Operations Research*, 2023.