In-Context Learning for Universal Bayesian Optimization

Xinyu Zhang, Nicolas Blumer, Daolang Huang, Conor Hassan, Samuel Kaski and **Julien Martinelli**

Learn the optimizer, not the surrogate. 🤯

Black-box optimization and Bayesian optimization

Black-box optimization (BBO): maximize an unknown costly function f **from noisy function calls** $y_t = f(x_t) + \epsilon_t$ only! **No gradient nor convexity information.**

Bayesian optimization (BO): the gold standard solution; fit a **surrogate** $p(f|\mathcal{D}_t)$, optimize an **acquisition** function $x_{t+1} = \arg\max_x \alpha(x|p(f|\mathcal{D}_t))$, then query y_{t+1}.

Caveats:

- surrogate/acquisition pair selection does not transfer across tasks
- acquisition optimization is a nontrivial problem in itself
- one-step acquisitions are mostly myopic; non-myopic BO is expensive.

In-context learning: amortized adaptation

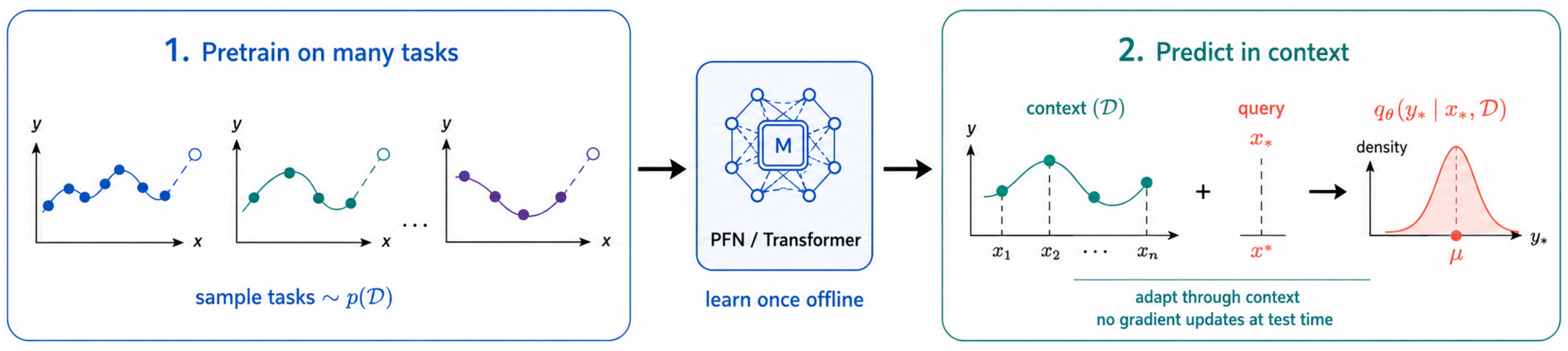
Classical supervised learning

$$\theta^\star = \arg\min_\theta \sum_j \ell(g_\theta(x_j), y_j), \qquad y_\star \approx g_{\theta^\star}(x_\star).$$

Learns a fixed input-output map; test examples are usually processed independently.

In-context learning

$$\mathcal{C}_n = \{(x_i, y_i)\}_{i=1}^n, \qquad g_\theta : (\mathcal{C}_n, x_\star) \mapsto \bar{y}_\star.$$



Prior fitted networks amortize Bayesian inference: prior samples are seen during pretraining; posterior reasoning is performed in context at deployment.

From In-context Learning to In-context Optimization

ICL **predicts** $\bar{y}_\star$ from context; ICOpt learns a **policy** over a query set that directly plays the acquisition role. No labels $x \mapsto \alpha(x)$, so the policy loss is trained by RL:

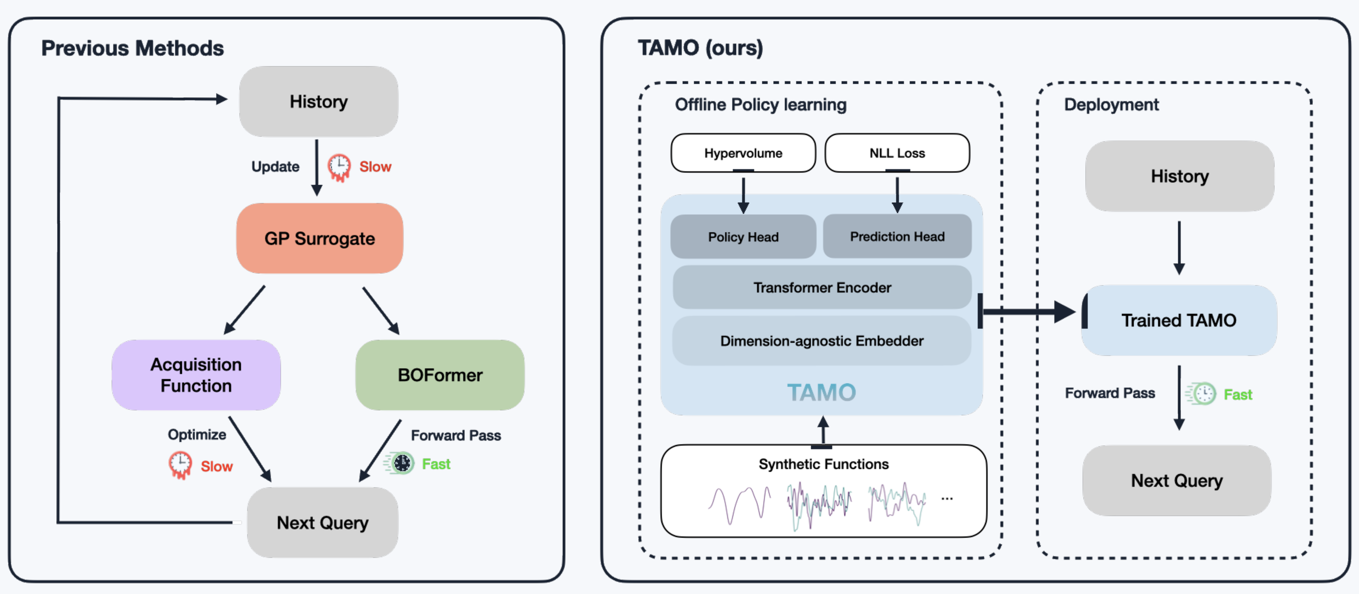
$$\mathcal{D}_t^{(c)} = \{(x_i, y_i)\}_{i=1}^t, \qquad \mathcal{D}_t^{(q)} = \{x_j^{(q)}\}_{j=1}^{N_q} \qquad x_{t+1} \sim \pi_\theta(\cdot \mid \mathcal{D}_t^{(c)}, \mathcal{D}_t^{(q)})$$

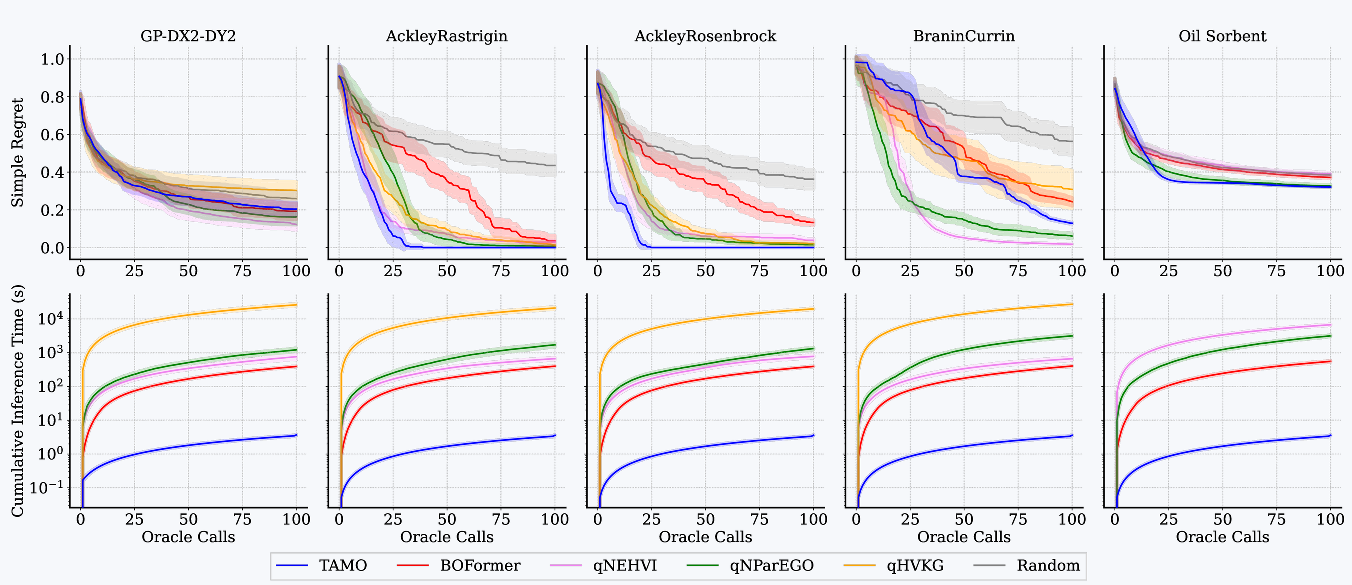
$$y_t^{\text{best}} = \max_{(x,y)\in\mathcal{D}_t^{(c)}} y, \qquad r_t = \frac{y_t^{\text{best}} - y_{\min}^\tau}{y_\star^\tau - y_{\min}^\tau} \qquad \mathcal{L}^{(\text{policy})}(\theta) = -\mathbb{E}_{\tau,\pi_\theta}\left[\sum_{t=1}^{T} \gamma^{t-1} r_t\right].$$

- **Offline:** simulate BO trajectories on many tasks and optimize cumulative reward.
- **Online:** adapt only through the context $\mathcal{D}_t^{(c)}$, without gradient updates.

TAMO: core In-context optimization module

A pretrained optimizer policy for scalar and multi-objective black-box design.

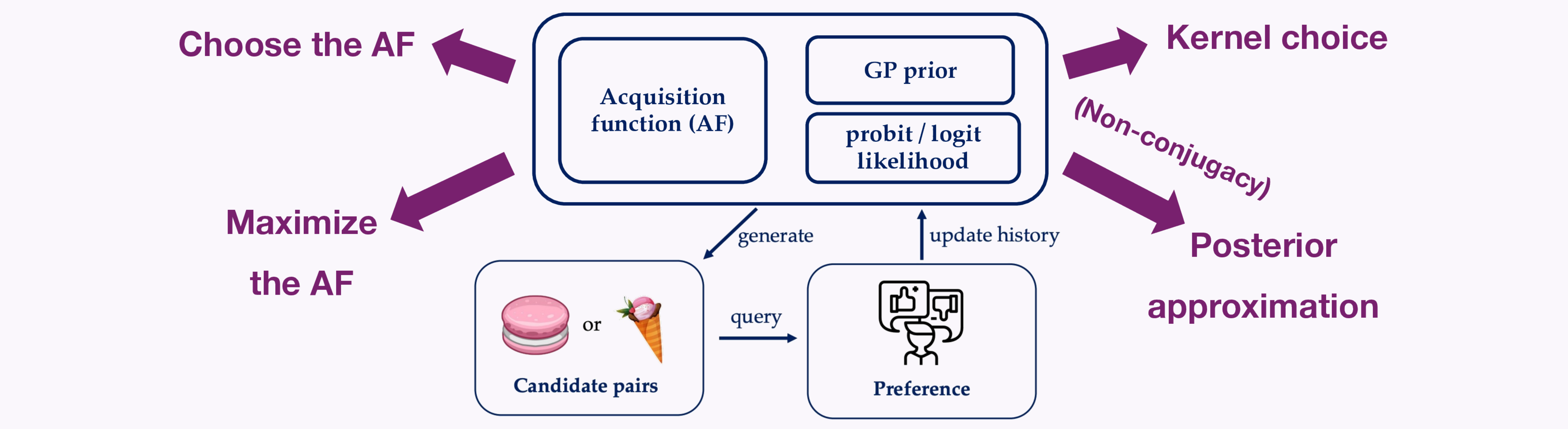




- **Performance:** competitive simple regret across synthetic and real black-box tasks.
- **Speed:** proposal time is amortized into one policy forward pass.

Extension 1 - PABBO: handling preferential queries

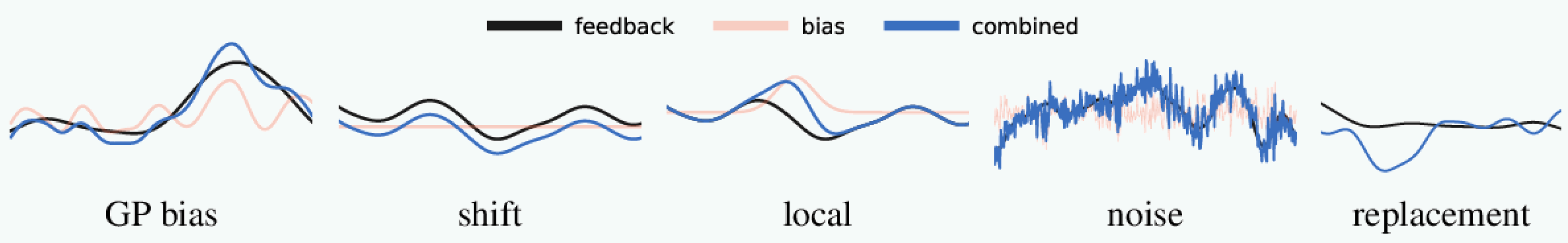
- Many BBO tasks are human-in-the-loop: e.g., visual design or controllers.
- For humans, providing scores $y(x)$ is difficult; comparisons $x \succ x'$ are more reliable.
- This problem is usually addressed by Preferential Bayesian Optimization (PBO).
- Yet, PBO uses a GP surrogate with likelihoods over pairwise preferences; these are non-Gaussian and require approximate posterior inference.



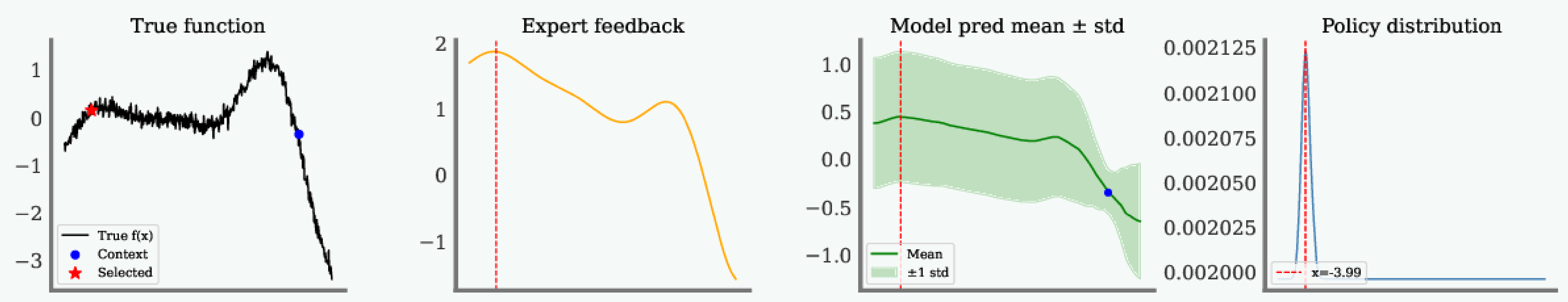
- As in TAMO, we meta-learn the preferential BO policy end-to-end.
- **Key lever: synthetic pretraining reveals both latent utilities $f(x)$ and comparisons $x \succ x'$ 🤯! $\rightarrow$ much richer supervision than deployment-time preferences alone.**

Extension 2 - FICBO: handling unreliable feedback sources

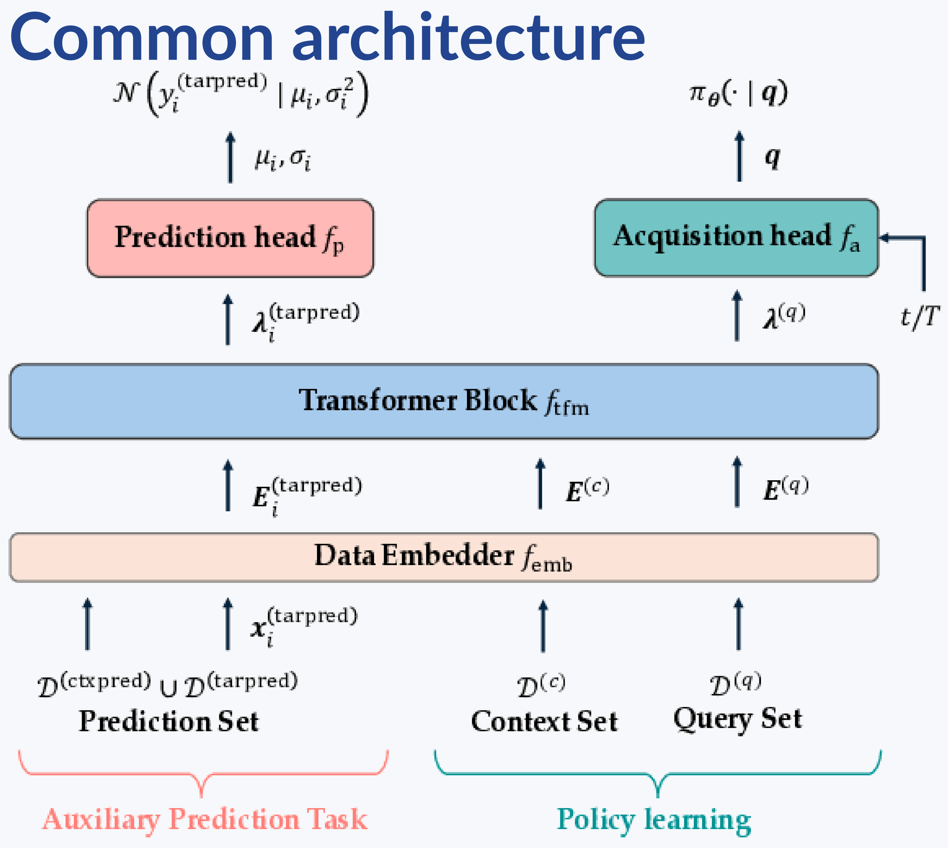
- Many BBO tasks come with auxiliary information, e.g. cheap experts or simulators.
- Classical BO can integrate these sources, but is brittle when their reliability varies.
- We can meta-learn a policy that infers at test time whether to trust a source.
- We pretrain on a novel synthetic feedback prior spanning useful, biased, noisy, and misleading sources.



- **Result: a feedback-conditioned policy that exploits helpful sources and discounts misleading ones 🤓.**



Common architecture



- The prediction head stabilizes policy training by learning the function landscape while the acquisition head learns to query.
- **Training objective** combines policy loss and auxiliary prediction:

$$\mathcal{L}^{(\text{policy})}(\theta) = -\mathbb{E}_{\tau,\pi_\theta}\left[\sum_{t=1}^{T} \gamma^{t-1} r_t\right]$$

$$\mathcal{L}^{(\text{pred})}(\theta) = -\mathbb{E}_\tau\left[\frac{1}{N_p}\sum_{i=1}^{N_p} \log p(y_i^p \mid x_i^p, \mathcal{D}^{(c)})\right]$$

$$\mathcal{L}(\theta) = \lambda_p \mathcal{L}^{(\text{pred})}(\theta) + \mathcal{L}^{(\text{policy})}(\theta)$$

Future work to achieve universal policies

- **Scaling to higher-dimensional problems.**
  - $\Longrightarrow$ Continuous policies beyond finite candidate pools.
  - $\Longrightarrow$ Trust region mechanisms learned from pretraining.
- **Interpretability**
  - $\Longrightarrow$ Probing the latent representations.
  - $\Longrightarrow$ Relate test-time tasks to pretraining tasks to diagnose prior transfer.

PABBO 😮
Preferential Amortized Black-Box Optimization
Xinyu Zhang, Daolang Huang, Samuel Kaski, **Julien Martinelli**
arXiv:2503.00924 – ICLR 2025 **(spotlight)**

TAMO 😶
In-Context Multi-Objective Optimization
Xinyu Zhang, Conor Hassan, **Julien Martinelli**, Daolang Huang, Samuel Kaski
arXiv:2512.11114 – ICLR 2026

FICBO 😯
In-Context Black-Box Optimization with Unreliable Feedback
Nicolas S. Blumer, **Julien Martinelli**, Samuel Kaski
arXiv:2605.06187 – submitted