COUNTERFACTUAL CREDIT GUIDED BAYESIAN OPTIMIZATION

A PREPRINT

Qiyu Wei
Department of Computer Science
The University of Manchester

Haowei Wang
Department of ISEM
National University of Singapore

Richard Allmendinger
Alliance Manchester Business School
The University of Manchester

Mauricio A. Álvarez
Department of Computer Science
The University of Manchester

ABSTRACT

Bayesian optimization has emerged as a prominent methodology for optimizing expensive black-box functions by leveraging Gaussian process surrogates, which focus on capturing the global characteristics of the objective function. However, in numerous practical scenarios, the primary objective is not to construct an exhaustive global surrogate, but rather to quickly pinpoint the global optimum. Due to the aleatoric nature of the sequential optimization problem and its dependence on the quality of the surrogate model and the initial design, it is restrictive to assume that all observed samples contribute equally to the discovery of the optimum in this context. In this paper, we introduce Counterfactual Credit Guided Bayesian Optimization (CCGBO), a novel framework that explicitly quantifies the contribution of individual historical observations through counterfactual credit. By incorporating counterfactual credit into the acquisition function, our approach can selectively allocate resources in areas where optimal solutions are most likely to occur. We prove that CCGBO retains sublinear regret. Empirical evaluations on various synthetic and real-world benchmarks demonstrate that CCGBO consistently reduces simple regret and accelerates convergence to the global optimum.

Keywords Bayesian optimization, Gaussian process, Credit Assignment, Counterfactual, Sublinear regret bound

1 Introduction

Bayesian Optimization (BO) is a powerful framework for optimizing expensive-to-evaluate black-box functions, demonstrating impressive efficiency across machine learning for hyperparameter tuning and experimental design tasks [Frazier, 2018, Wang et al., 2023]. Central to BO is the balance between exploration and exploitation [Garnett, 2023]. Traditionally, BO algorithms achieve this balance implicitly via acquisition functions [Archetti and Candelieri, 2019, Wilson et al., 2018] that guide sampling decisions based on the posterior uncertainty and predicted outcomes from Gaussian Process (GP) models [MacKay et al., 1998].

However, in BO the standard exploration–exploitation trade-off has limitations: it often wastes evaluations in suboptimal regions. BO is fundamentally a resource allocation problem. When budgets are extremely tight, finding the promising region early is more valuable than making marginal improvements later, relying solely on exploration and exploitation can be inefficient. Meanwhile, a limitation of conventional BO methods is the implicit assumption that all historical observations contribute equally to the optimization progress. In real-world landscapes, some specific samples provide more informative evidence of the global optimum than others [Koh and Liang, 2017], yet this heterogeneity is ignored and we may fail to identify key information. As a result, budget allocation is often inefficiently wasted in suboptimal areas, resulting in the budget not being reasonably allocated to areas that are more likely to have optimal values and slowing down convergence.

Some existing methods add regional constraints to avoid less informative or unsafe samples and focus on promising areas [Sui et al., 2015, Acerbi and Ji, 2017, Eriksson et al., 2019, Kim et al., 2020]. However, these methods often rely on manual thresholds and do not have adaptive capabilities, and using local optimization instead of global

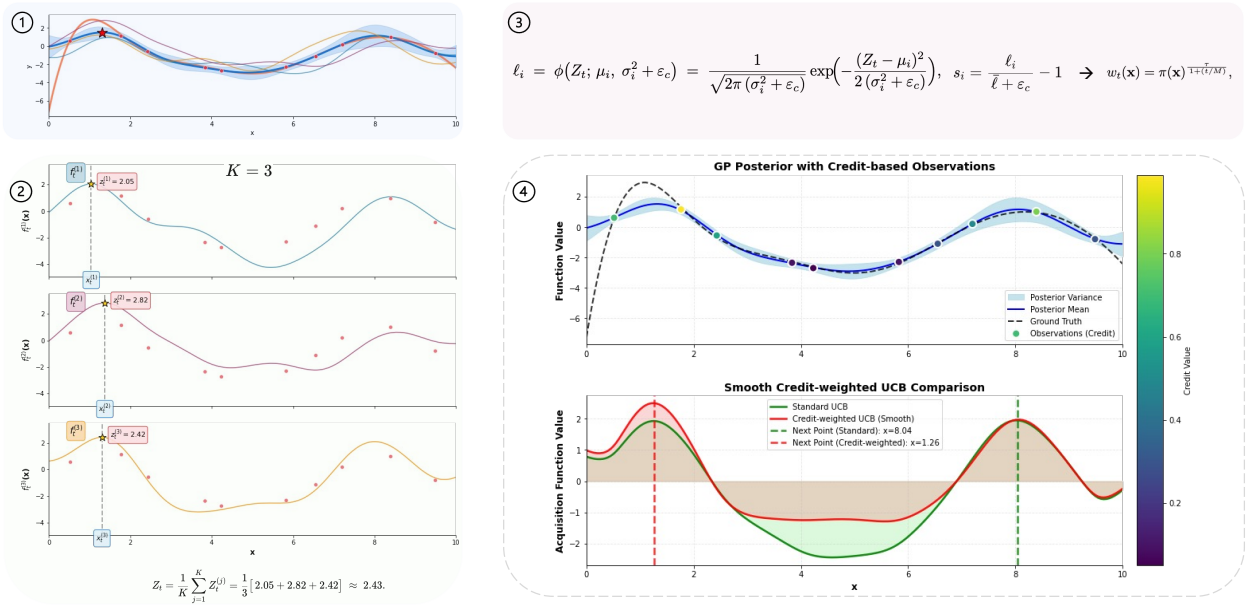

Figure 1: Illustration of Credit-Weighted UCB. Our goal is to maximize the objective function. (1) From the GP posterior at iteration t, we draw several paths; existing observations are shown as dots and the true optimum is marked by $\star$. (2) In this illustration, for $K = 3$ posterior samples, we compute each sample's maximizer $x_t^{(j)}$ and maximum $Z_t^{(j)}$, and form the Monte Carlo proxy of the global maximum Z_t. (3) For every observed point x_i, we assign a counterfactual credit ℓ_i, normalize and propagate it, then obtain the weight $w_t(x)$. (4) The right-hand color bar shows the counterfactual credit. By incorporating counterfactual credit into the UCB (red), our method concentrates exploitation on high-contribution regions, yielding a next query that targets the true optimum compared to the standard UCB (green).

optimization may lead to suboptimal results. In addition, some work uses expert knowledge to "Bring Human Back Into Loop" [Hvarfner et al., 2022b, 2024b], but this often requires good priors. Without accurate priors, performance can be worse than standard BO, and in real applications, we cannot have a good prior for all settings. However, these methods essentially rely on either external thresholds or expert priors. Therefore, they are unable to explicitly characterize which region is more promising from the observations itself.

To explicitly address these limitations, we propose Counterfactual Credit Guided Bayesian Optimization (CCGBO), which introduces an explicit mechanism to quantify the contribution of each historical observation using counterfactual reasoning [Harutyunyan et al., 2019, Mesnard et al., 2021, Meulemans et al., 2023]. Because credits are derived directly from the GP posterior, CCGBO does not require external priors. Specifically, counterfactual credit works as a proxy of contribution, assessing how significantly the prediction of the current optimum would degrade if a particular observation were absent, directly answering the question: "How significantly would our prediction of the current optimum deteriorate if a particular observation were absent?". By explicitly quantifying these contributions, CCGBO moves beyond traditional exploration-exploitation trade-off to a richer three-dimensional trade-off: exploration-exploitation-importance. Integrating counterfactual credit into the BO framework empowers BO to allocate sampling resources more efficiently and target areas that truly matter, enabling precise identification and prioritization of regions critically influencing rapid convergence to the optimum. Figure 1 visually illustrates how Credit-Weighted UCB effectively directs exploration toward areas of high contribution, demonstrating more targeted queries than standard UCB [Srinivas et al., 2010]. In summary, our contributions are threefold:

- **Counterfactual credit.** We introduce counterfactual credit as an efficient proxy for the per-sample contribution score of BO samples, which does not require manual specification. Quantifying and exploiting the varying counterfactual contribution credit of historical samples, enabling a trade-off on exploration-exploitation-importance;
- **Theoretical analysis.** We establish that the optimum proxy tracks the true optimum, which justifies CCGBO's early focusing on high-value regions. We further prove that counterfactual credit-weighted UCB retains sublinear regret, and show that it inherits the theoretical properties of the GP-UCB acquisition function;

- **Empirical validations.** We develop a counterfactual credit-weighted acquisition function, offering a modular toolkit that retains compatibility with any GP-BO backbone. We empirically validate the performance improvements provided by our counterfactual credit-guided approach across diverse optimization scenarios.

2 Related Works

Knowledge from Previous Experiments. In recent years, there has been growing interest in leveraging data from previous experiments to improve the efficiency of BO. *(1) OutlierBO.* Martinez-Cantin et al. [2018] addresses the issue that spurious observations can corrupt the GP surrogate and mislead acquisition. They propose a two-stage method using a Student-*t* likelihood in the GP to identify and remove outliers. *(2) Continuous optimization in non-stationary environments.* When the objective drifts over time, stale data may lose relevance. Bogunovic et al. [2016] extend GP-UCB to time-varying settings by periodically restarting the model to discard old data or gradually forgetting past observations via exponentially decaying weights. Building on this, Deng et al. [2022] introduces an explicit weight sequence ω_t into both the GP inference and acquisition computation, making recent points carry greater influence. *(3) Historical data utilization in multi-task and transfer learning.* In contexts where data from related tasks are available, Swersky et al. [2013] propose a multi-task GP that learns a task covariance kernel, thereby implicitly weighting historical data according to inter-task similarity. Feurer et al. [2018] fit a separate GP to each previous optimization run and combine their predictions with weights based on rank similarity between the source and target tasks. Hakhamaneshi et al. [2021] develop a multi-task BO, which trains a warm GP on large offline datasets and a cold GP on online samples, and dynamically trades off their suggestions based on task relevance. More recently, Mahboubi et al. [2025] introduced a point-by-point transfer strategy that evaluates the value of each candidate historical sample via a Gaussian mixture model and incorporates only the observations most aligned with the target task. Overall, these methods rely on pre-specified mechanisms. We instead compute per–sample weights using a counterfactual method in BO.

Embedding Priors over Function Optimum. Embedding informative priors over the location of a function's global optimum within BO has attracted growing interest. *(1) Expert-Driven Priors.* A number of works have sought to inject user beliefs directly into the acquisition process. Ramachandran et al. [2020] warp the input space according to a prior distribution over promising regions. Souza et al. [2021] and Hvarfner et al. [2022b] generalize this idea, defining a prior over the optimum that modulates any standard acquisition function. Huang et al. [2022] complement these by actively querying experts for pairwise preferences and encoding the resulting belief model into the GP prior. Wu et al. [2017] exploit the available gradient information as a form of expert knowledge, leading to the derivative-inclusive knowledge gradient algorithm that can drastically speed up convergence. *(2) Physics-Informed and Simulator-Based Priors.* In domains where there are partial physical models, integrating these laws into BO can produce substantial gains. Khatamsaz et al. [2023] embed known material-science relationships directly into GP kernels to optimize NiTi alloy processing, while Kobayashi et al. [2025] infuse epitaxial growth laws into the prior and demonstrate accurate extrapolation for III–V semiconductor deposition. Boltz et al. [2025] further show that using fast accelerator-simulation models as a GP mean function permits online tuning of particle accelerators with fewer expensive measurements. *(3) Functional-Structure Priors.* Beyond beliefs and data, exploiting the known structural properties of the objective can also guide BO. Li et al. [2017] inject monotonicity constraints via virtual derivative observations to enforce that the surrogate respects known increasing or decreasing trends. Kandasamy et al. [2015] address high-dimensional spaces by assuming an additive decomposition of the objective. However, these methods all depend heavily on having correctly specified priors, whether from experts, simulators, or structural assumptions, and can perform poorly or lose robustness when such prior information is unavailable, biased, or misspecified.

3 Preliminaries

In this section, we summarize BO setup and notation, and introduce a auxiliary prior augmented acquisition.

3.1 Bayesian Optimization (BO)

BO is a sequential decision-making framework that aims to find the global optimum of an expensive black-box function. BO employs a surrogate model, typically a GP, to approximate the unknown function, and an acquisition function to balance exploration and exploitation when selecting the next evaluation point.

Problem Setting. We consider the problem of maximizing an expensive black-box objective function $f : \mathcal{X} \to \mathbb{R}$, $\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$, where $\mathcal{X} \subset \mathbb{R}^d$ is a compact domain and each evaluation of $f(\mathbf{x})$ incurs high cost. At iteration t, we have collected observations $\mathcal{D}_t = \{(\mathbf{x}_i, y_i)\}_{i=1}^t$, $y_i = f(\mathbf{x}_i) + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$, where σ_n^2 denotes the variance of observation noise. BO maintains a GP surrogate that, conditioned on $\mathcal{D}_t$, yields a posterior

mean $\mu_t(\mathbf{x})$ and standard deviation $\sigma_t(\mathbf{x})$. An acquisition function $\alpha(\mathbf{x})$, built from μ_t and σ_t, then prescribes the next query point: $\mathbf{x}_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{X}\setminus\{\mathbf{x}_1,\dots,\mathbf{x}_t\}} \alpha(\mathbf{x})$, after which we observe $y_{t+1} = f(\mathbf{x}_{t+1}) + \varepsilon_{t+1}$, and update $\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(\mathbf{x}_{t+1}, y_{t+1})\}$. This process repeats until a stopping criterion is met, and we return $\mathbf{x}^*$ as our estimated optimizer.

Surrogate Model. We place a GP prior on f: $f(\mathbf{x}) \sim \mathcal{GP}\big(m(\mathbf{x}),\, k(\mathbf{x},\mathbf{x}')\big)$, where $m(\mathbf{x})$ is the prior mean and $k(\mathbf{x},\mathbf{x}')$ is a positive-definite covariance kernel. Conditioned on $\mathcal{D}_t$, the posterior at a test point $\mathbf{x}$ is $\mu_t(\mathbf{x}) = m(\mathbf{x}) + k\big(\mathbf{x},\mathbf{X}_{1:t}\big)\big[\mathbf{K}_{1:t} + \sigma_n^2\mathbf{I}\big]^{-1}\big(\mathbf{y}_{1:t} - m(\mathbf{X}_{1:t})\big)$, $\sigma_t^2(\mathbf{x}) = k(\mathbf{x},\mathbf{x}) - k\big(\mathbf{x},\mathbf{X}_{1:t}\big)\big[\mathbf{K}_{1:t} + \sigma_n^2\mathbf{I}\big]^{-1}k\big(\mathbf{X}_{1:t},\mathbf{x}\big)$, where $\mathbf{X}_{1:t} = [\mathbf{x}_1,\dots,\mathbf{x}_t]$, $\mathbf{y}_{1:t} = [y_1,\dots,y_t]^\top$, and $\mathbf{K}_{1:t}$ is the kernel matrix with entries $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$.

Acquisition Function. Given the posterior $\mathcal{N}\big(\mu_t(\mathbf{x}), \sigma_t^2(\mathbf{x})\big)$, BO selects the next point by maximizing an acquisition function. For the Upper Confidence Bound (UCB), we use $\mathrm{UCB}_t(\mathbf{x}) = \mu_t(\mathbf{x}) + \beta_t\,\sigma_t(\mathbf{x})$, $\beta_t > 0$, and choose $\mathbf{x}_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{X}\setminus\{\mathbf{x}_1,\dots,\mathbf{x}_t\}} \mathrm{UCB}_t(\mathbf{x})$. The new observation $(\mathbf{x}_{t+1}, y_{t+1})$ is added to $\mathcal{D}_{t+1}$, and the procedure repeats until termination. Upon completion, the best observed point $\mathbf{x}^*$ is returned.

3.2 Augmented Acquisition Functions

In many real-world tasks, one also has access to auxiliary information $a \in \mathcal{A}$ (e.g. domain clues, contextual features) that correlates with the optimizer's location. We encode this via a prior

$$\kappa(\mathbf{x} \mid a) = \Pr\big(\mathbf{x} = \arg\max_{\mathbf{x}'} f(\mathbf{x}') \;\big|\; a\big), \tag{1}$$

which places high mass on $\mathbf{x}$ likely to be optimal given a. We then augment any base acquisition $\alpha(\mathbf{x})$ by

$$\alpha_a\big(\mathbf{x}, a\big) = \alpha\big(\mathbf{x}\big)\;\kappa(\mathbf{x} \mid a),\;\; \mathbf{x}_{t+1} = \arg\max_{\mathbf{x}} \alpha_a(\mathbf{x}, a). \tag{2}$$

By forming $\alpha_a(\mathbf{x})$ as the product of the GP-based acquisition $\alpha(\mathbf{x})$ and the prior $\kappa(\mathbf{x} \mid a)$, we focus sampling on points where both the surrogate model predicts high utility and the auxiliary information indicates a high likelihood of optimality.

4 Counterfactual Credit Guided Bayesian Optimization (CCGBO)

In this section, we introduce CCGBO, we first motivate and define counterfactual credit, then derive a posterior-based estimator and its propagation to candidates, and finally integrate it into a credit-weighted acquisition.

4.1 Counterfactual Credit for BO

In BO, each observed sample contributes unequally to the discovery of the global optimum. At each iteration t, selecting a candidate point $\mathbf{x}_t$ not only yields an immediate function evaluation y_t, but also influences the future trajectory of the optimization and, consequently, the speed with which the true optimum is located. Our goal is to assign a credit $c(\mathbf{x}_t)$ to each observed sample that quantifies its contribution to the downstream search for the optimum. A natural "Monte Carlo" style approach is to simulate multiple complete BO trajectories $\{\tau_i\}$ starting from the current dataset and to estimate $\bar{R}(\mathbf{x}_t) = \frac{1}{K}\sum_{i=1}^{K} R\big(\tau_i \mid \mathbf{x}_t\big)$, where K is the number of simulated trajectories, and $R(\tau_i \mid \mathbf{x}_t)$ denotes the total discounted reward (e.g., reduction in best seen value) along trajectory τ_i given that sample $\mathbf{x}_t$ was selected at iteration t. However, because each simulated trajectory is subject to environmental noise and the stochasticity of the acquisition policy, the variance of this estimator is high. This forward and rollout-based sequential simulation leads to low sample efficiency in the optimization process, which is not suitable for real-time BO.

Ideally, one would like to perform counterfactual updates for all candidate samples in a single trajectory: Even if a particular point $\mathbf{x}'$ was not chosen at iteration t, we would still like to estimate "what would have happened" had we observed $f(\mathbf{x}')$. Existing BO methods do not provide this capability, since they only update the posterior at the actually sampled points. Therefore, we propose a counterfactual credit computation as a proxy of contribution that (i) leverages the GP posterior to estimate, for each candidate $\mathbf{x}$, the hypothetical contribution to finding the optimum, and (ii) can be evaluated in closed form and with low computational overhead at each iteration. Here we remodel the calculation method of contribution from *"how does choosing an observation* $\mathbf{x}$ *in the past affect the finding of optimal"* to *"given the optimal, how relevant was the choice of observation* $\mathbf{x}$ *in the past to achieve it?"*. If a sample contributes a lot about the optimal location, it usually means that it is located in or close to the area in space where high values are most likely to appear, and it will have a larger credit accordingly. Although observations with poor results will also contribute to finding the optimal value, their contribution is smaller than that of observations near the optimal value, which will be reflected in the credit. Incorporating counterfactual credit into BO enhances the traditional optimization

Algorithm 1 Counterfactual Credit Guided BO

Input: objective f, domain $\mathcal{X}$, initial data $\mathcal{D}_t$, total iterations N, credit weight λ, UCB factor $\{\beta_t\}$, decay τ, proxy samples K.
Output: estimated maximizer $\mathbf{x}^*$.
for $t = 1, \ldots, N$ **do**
  Train on $\mathcal{D}_t \to \mu_t,\ \sigma_t$.
  for $j = 1, \ldots, K$ **do**
    Sample $f_t^{(j)} \sim \mathcal{GP}(\mu_t, k_t)$.
    $\mathbf{x}_t^{(j)} \leftarrow \arg\max_{\mathbf{x}} f_t^{(j)}(\mathbf{x});\ Z_t^{(j)} \leftarrow f_t^{(j)}(\mathbf{x}_t^{(j)})$.
  end for
  Obtain optimistic estimate $Z_t \leftarrow \frac{1}{K}\sum_j Z_t^{(j)}$.
  Compute per-point credits $\{c_i\}$, credit field $\pi(\mathbf{x})$ and weighting factor $w_t(\mathbf{x})$.
  Obtain Counterfactual Credit-Weighted UCB $\alpha(\mathbf{x}) = \big[(1-\lambda) + \lambda\, w_t(\mathbf{x})\big]\,\big[\mu(\mathbf{x}) + \beta_t\,\sigma(\mathbf{x})\big]$.
  $\mathbf{x}_{\text{new}} \leftarrow \arg\max_{\mathbf{x}\in X_{\text{cand}}} \alpha(\mathbf{x}),\ y_{\text{new}} \leftarrow f(\mathbf{x}_{\text{new}})$.
  $\mathcal{D}_{t+1} \leftarrow \mathcal{D}_t \cup \{(x_{\text{new}},\ y_{\text{new}})\}$.
end for
return $\mathbf{x}^* = \arg\max_{(x,y)\in\mathcal{D}_t} y$.

strategy by assigning credit to previously evaluated points, effectively guiding the optimization toward regions with higher potential.

4.2 Computation of Counterfactual Credit

At iteration t, let the dataset be $\mathcal{D}_t = \{(\mathbf{x}_i, y_i)\}_{i=1}^t$, and let the GP posterior at any $\mathbf{x} \in \mathcal{X}$ have mean $\mu_t(\mathbf{x})$ and standard deviation $\sigma_t(\mathbf{x})$. Rather than directly using the observed maximum $\max_i y_i$, we construct a current global optimum proxy Z_t via a Monte Carlo approximation to the UCB criterion: Draw K independent sample paths from the GP posterior $f_t^{(j)}(\cdot) \sim \mathcal{GP}\big(\mu_t(\cdot),\, k_t(\cdot,\cdot)\big),\ j = 1, \ldots, K$. For each sample j, we approximate its maximizer $\mathbf{x}_t^{(j)} = \arg\max_{\mathbf{x}\in\mathcal{X}} f_t^{(j)}(\mathbf{x})$, and record the corresponding current global optimal proxy $Z_t^{(j)} = f_t^{(j)}\big(\mathbf{x}_t^{(j)}\big)$. We then aggregate the K sample maxima into a single proxy $Z_t = \frac{1}{K}\sum_{j=1}^K Z_t^{(j)}$. This Monte Carlo proxy captures both the posterior mean variance trade-off and uncertainty about the true global maximum. Denote for brevity at each observed location $\mathbf{x}_i$: $\mu_i = \mu_t(\mathbf{x}_i),\ \sigma_i = \sigma_t(\mathbf{x}_i)$. We compute a likelihood score quantifying how likely each $\mathbf{x}_i$ is to have produced Z_t:

$$\begin{aligned} \ell_i &= \phi\big(Z_t;\, \mu_i,\, \sigma_i^2 + \varepsilon_c\big) \\ &= \frac{1}{\sqrt{2\pi\,(\sigma_i^2 + \varepsilon_c)}} \exp\Big(-\frac{(Z_t - \mu_i)^2}{2\,(\sigma_i^2 + \varepsilon_c)}\Big), \end{aligned} \tag{3}$$

where $\phi(\cdot; m, v)$ is the Gaussian density with mean m, variance v, and $\varepsilon_c > 0$ is a small constant used to prevent division by zero and stabilize the denominator. Next, we introduce the unconditional baseline $\bar{\ell} = \frac{1}{t}\sum_{j=1}^t \ell_j$. And define the raw counterfactual score $s_i = \frac{\ell_i}{\bar{\ell}+\varepsilon_c} - 1$, so that $s_i > 0$ indicates $\mathbf{x}_i$ positively contributes to achieving the global proxy Z_t. Finally, we normalize and bound these scores into credits $c_i \in [r_{\min},\, r_{\max}]$. Let t be the current number of observations. We first calculate a normalized rank $r_i = \frac{\big|\{\,j:1\le j\le t,\ s_j \le s_i\}\big| - 1}{t-1} \in [0, 1]$. Then we linearly map this rank into the credit interval: $c_i = r_{\min} + (r_{\max} - r_{\min})\, r_i$, where $r_{\min} = 0.1$ and $r_{\max} = 1$. We set $r_{\min} = 0.1$ to avoid collapsing any weight to zero, which would prematurely exclude regions. The resulting $\{c_i\}$ form our counterfactual credits, which can be seamlessly incorporated into augmented acquisition functions, thus realizing a three-way exploration-exploitation-credit strategy in BO. Next, we will explain how to extend the discrete credits c_i to arbitrary candidate points.

4.3 Counterfactual Credit-Weighted Acquisition Function

In standard BO, acquisition functions such as UCB balance exploration and exploitation based on the posterior predictive mean $\mu(\mathbf{x})$ and standard deviation $\sigma(\mathbf{x})$: $\mathrm{UCB}(\mathbf{x}) = \mu(\mathbf{x}) + \beta_t\,\sigma(\mathbf{x})$, where $\beta_t > 0$ is a tunable parameter. However, this criterion treats all past observations as equally informative, ignoring their varying contributions. To address this, we

introduce a credit-weighted acquisition function that integrates counterfactual credits into an acquisition function. Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^t$ and $\{c_i\}_{i=1}^t$ denote the evaluated points and their credits. We proceed in two steps:

(1) *Propagating Credits to Continuous Candidates.* Given a set of candidate points $X_{\text{cand}} = \{\mathbf{x}\}$, we estimate a candidate credit $c(\mathbf{x})$ by averaging the credits of its H nearest neighbors in the evaluated set: $c(\mathbf{x}) = \frac{1}{H}\sum_{j\in\mathcal{N}_H(\mathbf{x})} c_j$, where $\mathcal{N}_H(\mathbf{x})$ are the indices of the H nearest $\mathbf{x}_j$ to $\mathbf{x}$ under Euclidean distance. This simple KNN-based propagation yields a smooth credit field $\pi(\mathbf{x}) = \frac{c(\mathbf{x})}{\max_j c_j}$ over $\mathcal{X}$.

(2) *Credit-Weighted UCB.* We then modify the UCB acquisition by incorporating the propagated credit weight:

$$\text{UCB}_{\text{credit}}(\mathbf{x}) = \big[(1-\lambda) + \lambda\, w_t(\mathbf{x})\big]\,\big[\mu(\mathbf{x}) + \beta_t\,\sigma(\mathbf{x})\big], \tag{4}$$

where $\lambda \in [0,1]$ modulates the strength of the credit influence, and the weighting factor $w_t(\mathbf{x})$ is computed based on the counterfactual evaluation of each observation's contribution:

$$w_t(\mathbf{x}) = \pi(\mathbf{x})^{\frac{\tau}{1+(t/M)}} \tag{5}$$

is a per-iteration weight. Here $\pi(\mathbf{x})$ is the normalized counterfactual credit for point $\mathbf{x}$, $\tau > 0$ controls the sensitivity to credit differences, t is the current iteration index, and the constant M sets the "half-life" of credit influence. At $t = 0$, $w_0(\mathbf{x}) = \pi(\mathbf{x})^\tau$ strongly biases early sampling toward high-credit regions; by $t = M$, the exponent halves to $\tau/2$, reducing extra weighting. As t grows further, the denominator dominates and $w_t(\mathbf{x}) \to 1$, smoothly reverting the acquisition back to the standard UCB form $\mu + \beta_t\,\sigma$. In practice, M and τ govern the decay of the credit effect, we therefore set them accordingly to match different evaluation budgets.

5 Theoretical Analysis

Here, we first establish that our Monte Carlo proxy of the current optimum tracks the true optimum. Let Z_t denote the average optimum of K posterior sample paths drawn at iteration t. Theoretical analysis shows that Z_t concentrates around the true optimal value $f(\mathbf{x}^\star)$ with high probability, with a decomposition into a model bias term and an MC error term. We now state the assumptions and theorem.

Assumptions 1 There exists $\beta_t > 0$ such that the GP-UCB confidence event $\mathcal{E}_t := \Big\{\, |f(\mathbf{x}) - \mu_t(\mathbf{x})| \le \beta_t\,\sigma_t(\mathbf{x}) \;\; \forall \mathbf{x} \in \mathcal{X} \,\Big\}$ holds with probability at least $1-\delta$ [Srinivas et al., 2010]. And there exist constants $c_0, c_1 > 0$ such that for the zero-mean posterior process $\varepsilon_t := f_t - \mu_t$ we have $\mathbb{E}\big[\sup_{\mathbf{x}\in\mathcal{X}} \varepsilon_t(\mathbf{x}) \;\big|\; \mathcal{D}_t\big] \le c_0 S_t$ [Adler and Taylor, 2007], and, conditionally on $\mathcal{D}_t$, the centered variables $Z_t^{(j)} - \mathbb{E}[Z_t^{(j)} \mid \mathcal{D}_t]$ are $(c_1 S_t)$-sub-Gaussian (ψ_2-norms are $\le c_1 S_t$) [Ledoux and Talagrand, 2013].

Theorem 5.1. *Let $\mathcal{X} \subset \mathbb{R}^d$ be compact, and at iteration t let the GP posterior be $f_t \sim \mathcal{GP}(\mu_t, k_t)$ with $\sigma_t^2(\mathbf{x}) = k_t(\mathbf{x},\mathbf{x})$ and $S_t = \sup_{\mathbf{x}\in\mathcal{X}} \sigma_t(\mathbf{x}) < \infty$. Draw K i.i.d. posterior sample paths $\{f_t^{(j)}\}_{j=1}^K$, and define $Z_t^{(j)} = \max_{\mathbf{x}\in\mathcal{X}} f_t^{(j)}(\mathbf{x})$ and $Z_t = \frac{1}{K}\sum_{j=1}^K Z_t^{(j)}$. Let $\mathbf{x}^\star \in \arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$. For any $\delta' \in (0,1)$, there exists a constant $C > 0$ (depending only on c_1) such that*

$$\Pr\left(\big|Z_t - f(\mathbf{x}^\star)\big| \le \beta_t S_t + C\, S_t \sqrt{\tfrac{\log(1/\delta')}{K}}\right) \ge 1 - \delta - \delta'. \tag{6}$$

Remark 1. By Theorem 5.1, with high probability we have $|Z_t - f(\mathbf{x}^\star)| \le (\beta_t + c_0)S_t + C S_t\sqrt{\log(1/\delta')/K}$, so the MC proxy Z_t is close to the true optimum $f(\mathbf{x}^\star)$. Consequently, for observed points $\mathbf{x}_i$ that lie in or near the optimal region, the quantity $(Z_t - \mu_t(\mathbf{x}_i))^2$ tends to be smaller, which makes the likelihood score $\ell_i = \phi\big(Z_t;\, \mu_t(\mathbf{x}_i),\, \sigma_t^2(\mathbf{x}_i) + \varepsilon_c\big)$ larger. Then resulting credit field $\pi(\mathbf{x})$ attains higher values around the optimum, making $w_t(\mathbf{x}) = \pi(\mathbf{x})^{\tau/(1+t/M)}$ also larger. Hence the credit-weighted acquisition $\alpha_t(\mathbf{x}) = \big[(1-\lambda) + \lambda\, w_t(\mathbf{x})\big]\big[\mu_t(\mathbf{x}) + \beta_t\,\sigma_t(\mathbf{x})\big]$ is amplified near high-credit areas. On the high-probability event, this mechanism makes the sampling tendency toward high-value regions, which promotes earlier hits near the optimum and thus faster early simple-regret decay.

We then analyze the cumulative regret of the Credit-Weighted UCB variant of CCGBO. Our strategy closely follows the classic GP-UCB analysis of Srinivas *et al.* [Srinivas et al., 2010] and [Hvarfner et al., 2022b], with the key modification that at each iteration t the standard UCB acquisition is rescaled by a factor $w_t(\mathbf{x}) = \big[c(\mathbf{x})\big]^{\frac{\tau}{1+(t/M)}}$, where $c(\mathbf{x})$ is the normalized counterfactual credit and $\tau > 0$ is the decay parameter. We show that this rescaling only incurs a constant multiplicative penalty in the cumulative regret bound.

Assumption 2 (Lipschitz on derivatives) There exist constants $a, b > 0$ such that for every $j = 1, \ldots, p$,

$$\Pr\Big(\sup_{\mathbf{x}\in\mathcal{X}}\Big|\tfrac{\partial f}{\partial x_j}(\mathbf{x})\Big| > L\Big) \;\le\; a\,\exp\big(-L^2/b^2\big). \tag{7}$$

Under this assumption and the usual choice of $\beta_t = O\big(\sqrt{\log(t^2\pi^2/6\delta)}\big)$, with probability at least $1-\delta$ we have simultaneously for all t and $\mathbf{x}$: $|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \;\le\; \beta_t\,\sigma_{t-1}(\mathbf{x})$.

Theorem 5.2. *Define for each t,* $A_t = 1 - \lambda + \lambda\, c_{\min}^{\tau/(1+t/M)}$, $B_t = 1 - \lambda + \lambda\, c_{\max}^{\tau/(1+t/M)}$. *Observe that* $0 < A_t \le B_t < \infty$ *and* $B_t/A_t \to 1$ *as* $t \to \infty$. *Under Assumption 5 and the above high-probability event, the cumulative regret of CCGBO satisfies*

$$R_N^{\mathrm{ccg}} \;\le\; \max_{1\le t\le N}\Big(\tfrac{B_t}{A_t}\Big)\, R_N, \tag{8}$$

where $R_N = O\big(\sqrt{N\,\gamma_N\,\beta_N}\big)$ *is the standard GP-UCB bound from [Srinivas et al., 2010] and* γ_N *is the maximum information gain. In particular, since* $\max_t B_t/A_t = B_1/A_1 < \infty$, *CCGBO matches the vanilla GP-UCB rate up to a constant factor that tends to 1 when* $\lambda \to 0$ *(so* $A_t = B_t = 1$*) or* $c_{\max}/c_{\min} \to 1$.

The cumulative regret of the standard GP–UCB algorithm is known to satisfy $R_N = O\big(\sqrt{N\,\gamma_N\,\beta_N}\big)$. In contrast, for CCGBO we obtain the bound $R_N^{\mathrm{ccg}} \;\le\; \max_{1\le t\le N}\Big(\frac{B_t}{A_t}\Big)\, R_N$. Thus, the regret of CCGBO exceeds that of GP–UCB by at most the constant factor $\max_t(B_t/A_t)$, which is controlled by the extremal credit weights $c_{\min}, c_{\max}$ and the decay parameters τ and M. The proof proceeds by comparing the two acquisition values under the high-probability confidence event, introducing A_t and B_t to extract the influence of the credit weights from the regret analysis.

The theoretical analysis shows that credits computed from Z_t are near the optimal region with high possibility, and also the counterfactual credits does not destroy the sublinear convergence rate of GP-UCB. CCGBO preserves the sublinear convergence rate of GP–UCB up to a controllable constant, demonstrating that adding counterfactual credit only incurs a slight overhead at the worst-case constant level.

6 Experiments

We validated CCGBO's performance against baseline methods on both synthetic functions and real-world tasks. The following are the experimental setup and results on the benchmarks.

6.1 Experimental Setup

We include the following baselines: First, Standard GP-UCB [Srinivas et al., 2010] and Random Search are the most natural baselines for BO. Then, we include Weighted Gaussian Process UCB (WGP) [Deng et al., 2022] and Time-Varying Gaussian Process UCB (RGP) [Bogunovic et al., 2016] to cover the non-stationary scenario. By comparing with WGP and RGP, we show that CCGBO's credit mechanism can also eliminate low-value observations during the optimization process. Next, OutlierBO [Martinez-Cantin et al., 2018] represents robust outlier handling. In contrast, CCGBO's credit assignment naturally downweights low-contribution points (including noise and outliers). Finally, PiBO [Hvarfner et al., 2022b] and ColaBO [Hvarfner et al., 2024b] embody user-prior, requiring expert-supplied beliefs about the optimum. By comparing with PiBO and COLABO, we highlight that our can achieve competitive results without any external prior. WGP, RGP, and OutlierBO are grouped as knowledge from previous experiments methods, whereas PiBO and ColaBO belong to Embedding Priors over the Function Optimum methods. We demonstrate the superiority of our algorithm by comparing it with standard, non-stationary, robust, and prior guidance methods. Note on PiBO and ColaBO, for a fair comparison, we do not elicit any external priors from users. Instead, we construct the user-provided prior by taking the location of the best point in the initial set as the prior mean for both PiBO and ColaBO.

In all experiments, each function evaluation is corrupted by independent Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ with $\sigma_n^2 = 0.01$, to simulate realistic measurements. The size of the initial design is set adaptively to $n_{\mathrm{init}} = \max\big(2\,d,\ 10\big)$, where d denotes the dimensionality of the problem. We employ a Matérn $5/2$ kernel for the GP, with hyperparameters learned via maximum likelihood estimation. The iteration budget is 100. The regret is averaged on 50 independent experiments. We set the "half-life" parameter $M = 20$, the Monte Carlo UCB proxy sample number $K = 25$, and the number of nearest neighbors $H = 5$. We employed eight benchmarks, five synthetic and three real-world problems. For clarity, all the tasks aim for function maximization. All experiments were conducted on a MacBook Pro with Apple M2 Pro (10-core CPU, 16 GB RAM).

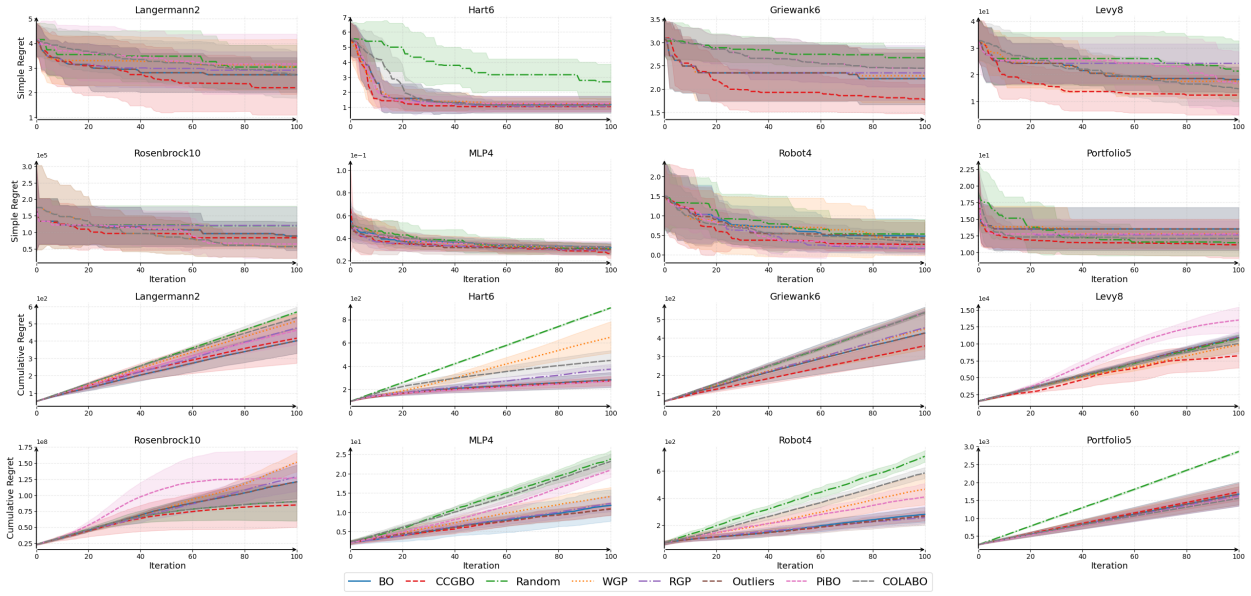

Figure 2: Cumulative regret and Simple regret versus iteration for eight benchmark functions.

Table 1: Area under simple regret metric across benchmarks (lower is better).

TASK	BO	CCGBO	RANDOM	WGP	RGP	OUTLIER	PIBO	COLABO
Langermann2	292.3 ± 69.0	$\mathbf{268.4 \pm 103.0}$	338.3 ± 84.5	324.0 ± 94.7	304.7 ± 73.0	292.3 ± 69.0	339.6 ± 103.1	329.9 ± 71.2
Hart6	160.3 ± 58.5	$\mathbf{135.1 \pm 27.8}$	376.5 ± 80.2	160.7 ± 42.2	152.5 ± 47.5	160.3 ± 58.5	157.3 ± 44.1	181.1 ± 37.3
Griewank6	235.0 ± 52.0	$\mathbf{203.3 \pm 35.5}$	280.6 ± 23.4	236.5 ± 54.1	238.3 ± 55.8	235.0 ± 52.0	266.0 ± 31.1	266.0 ± 31.1
Levy8	2128.3 ± 510.4	$\mathbf{1514.8 \pm 620.3}$	2530.9 ± 963.2	2116.1 ± 563.2	2446.2 ± 735.0	2128.3 ± 510.4	2368.8 ± 930.9	2132.4 ± 536.3
Rosenbrock10 ($1e4$)	1085.7 ± 436.6	966.1 ± 394.6	1221.9 ± 574.5	948.0 ± 406.6	1218.1 ± 559.9	1085.7 ± 436.6	969.7 ± 428.3	$\mathbf{947.9 \pm 406.6}$
MLP4	3.6 ± 0.3	$\mathbf{3.3 \pm 0.4}$	3.7 ± 0.6	3.6 ± 0.5	$\mathbf{3.3 \pm 0.3}$	3.4 ± 0.6	3.6 ± 0.5	3.7 ± 0.3
Robot4	71.1 ± 39.2	$\mathbf{48.5 \pm 26.0}$	80.6 ± 42.8	71.1 ± 59.8	50.6 ± 23.9	66.0 ± 36.4	52.2 ± 24.1	60.7 ± 26.4
Portfolio5	1360.3 ± 304.7	$\mathbf{1170.3 \pm 171.3}$	1277.0 ± 225.4	1320.2 ± 218.5	1290.1 ± 214.6	1360.3 ± 304.7	1287.6 ± 223.3	1244.4 ± 154.9

Synthetic Test Problems. We evaluate five synthetic test functions: *Langermann2* is a 2-dimensional multimodal function on $[0, 10]^2$; *Hartmann6* is a six-dimensional function on $[0, 1]^6$; *Griewank6* is a nonconvex six-dimensional landscape defined on $[-600, 600]^6$; *Levy8* is an 8-dimensional valley-shaped function on $[-10, 10]^8$; and *Rosenbrock10* is a 10-dimensional multimodal test function on $[-5, 10]^{10}$. We put high-dim synthetic test problems in Supplementary Materials.

Real-World Test Problems. We further evaluated CCGBO on three real-world problems. (1) The NAS experiment models a breast cancer prediction problem from the UCI repository [Dua and Graff, 2017] of searching for optimal hyperparameters as a BO problem. Specifically, each query (x_t, y_t) corresponds to a choice of $([32, 128], [1e-6, 1.0], [1e-6, 1.0], [1, 8]$ for batch size, Learning Rate, Learning Rate decay, hidden dim, respectively), where y_t is the test classification error. (2) The Robot4 [Kaelbling and Lozano-Pérez, 2017] problem simulates a 4-dimensional robot pushing task using the Box2D physics engine, where each query corresponds to robot initial position $\in [-5, 5]^2$, initial robot angle control $\in [0, 2\pi]$and simulation steps $\in [1, 30]$, with y_t being the Euclidean distance between the pushed object's final position and target. (3) The Portfolio problem [Bruni et al., 2016] represents a 5-dimensional financial portfolio optimization task, where each query corresponds to asset normalized weight allocations $\in [0, 1]^5$.

6.2 Comparison of CCGBO Against Baselines

Figures 2 summarize the results on eight benchmarks, show the mean and standard error of the simregret and cumregret on 8 benchmark problems. We also report the area under simple regret in Table 1 as a summary metric of the entire optimization trajectory, defined as $\text{AUSR} = \frac{1}{T-1} \sum_{t=2}^{T} \frac{r_{t-1}+r_t}{2}$, where r_t denotes the simple regret at iteration t. It measures the remaining gap between the best value found so far and the true optimum. We report the observations below:

(1) Faster regret convergence: Across all benchmarks, CCGBO attains the fastest drop of simple regret especially in the early stage and settling at the low regret plateau, outperforming Standard GP-UCB. This is also confirmed in the results reported in the Table 1. This shows that compared with the baseline method, counterfactual credit weighting can more effectively prioritize high-value areas and avoid redundant exploration in "less promising" areas. Especially when there are large flat areas or multiple local extremes in the objective function, counterfactual credit can quickly focus on the most informative sampling points, thereby significantly accelerating the convergence speed. Meanwhile, because CCGBO established advantages in the early stage, it has a better understanding of the global optimum in the early stage, which is helpful for the subsequent optimization process, and simple regret maintained an advantage at most tasks.

(2) Lower cumulative regret: CCGBO maintains an advantage over most baselines, while random search quickly accumulates regret and specialized schemes like WGP and RGP only partially close the gap.This confirming that CCGBO preserves the sublinear regret rate while delivering practical gains.

(3) Robustness without artificial priors: Unlike PiBO and ColaBO, which rely on user-supplied beliefs, CCGBO achieves equal or better performance without external priors. The optimization effects of PiBO and ColaBO vary on different benchmarks due to the fact that the prior settings are sometimes good and sometimes bad. Moreover, our method outperforms Outlier-Robust BO even in the noisy setting, since low-credit observations are naturally downweighted by the counterfactual credit mechanism.

Our algorithm is plug-and-play, it also achieves good results on other different aquisition functions and high-dimensional problems. We put the experiment on other acquisition functions (TS, EI, JES etc.), high-dim problems (25 to 1000 dimension) and ablation study on M, K, H in Supplementary Materials. Due to space limitations, details is provided in Supplementary Materials.

Together, these results confirm that CCGBO delivers a general improvement over existing BO techniques, accelerating convergence, reducing cumulative loss, and obviating the need for user-specified priors. It should be noted that the time spent on calculating credit is very small compared to the optimization process and can be ignored.

7 Conclusion

In this paper, we introduced CCGBO, an efficient framework that integrates counterfactual credit assignments into the acquisition process to explicitly quantify and exploit the varying contribution of past observations. By augmenting the traditional exploration–exploitation trade-off with a third importance dimension, our method adaptively focuses sampling effort on regions most likely to yield rapid improvements toward the global optimum. We derived an estimator for per-sample credit based on the GP posterior and a Monte Carlo proxy, incorporated this into a Credit-Weighted UCB acquisition function, and showed that the resulting algorithm retains faster simple regret convergence and sublinear cumulative regret. Empirical results on synthetic test function benchmarks and real-world hyperparameter-tuning task demonstrate that CCGBO outperforms standard BO and recent baseline alternatives, especially in faster convergence rate.

References

L. Acerbi and W. Ji. Practical bayesian optimization for model fitting with bayesian adaptive direct search. In *Advances in Neural Information Processing Systems 30*, pages 1836–1846, 2017.

R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer, 2007.

S. Ament, S. Daulton, D. Eriksson, M. Balandat, and E. Bakshy. Unexpected improvements to expected improvement for bayesian optimization. In *Advances in Neural Information Processing Systems 36*, pages 20577–20612, 2023.

F. Archetti and A. Candelieri. The acquisition function. In *Bayesian Optimization and Data Science*, pages 57–72. Springer, 2019.

I. Bogunovic, J. Scarlett, and V. Cevher. Time-varying gaussian process bandit optimization. In *International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 314–323. JMLR, 2016.

T. Boltz, J. L. Martinez, C. Xu, K. R. Baker, Z. Zhu, J. Morgan, R. Roussel, D. Ratner, B. Mustapha, and A. L. Edelen. Leveraging prior mean models for faster bayesian optimization of particle accelerators. *Scientific Reports*, 15(1): 12232, 2025.

R. Bruni, F. Cesarone, A. Scozzari, and F. Tardella. Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models. *Data in brief*, 8:858, 2016.

Y. Deng, X. Zhou, B. Kim, A. Tewari, A. Gupta, and N. B. Shroff. Weighted gaussian process bandits for non-stationary environments. In *International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 6909–6932. PMLR, 2022.

D. Dua and C. Graff. UCI Machine Learning Repository, 2017. URL http://archive.ics.uci.edu/ml.

D. Eriksson, M. Pearce, J. R. Gardner, R. Turner, and M. Poloczek. Scalable global optimization via local bayesian optimization. In *Advances in Neural Information Processing Systems 32*, pages 5497–5508, 2019.

M. Feurer, B. Letham, and E. Bakshy. Scalable meta-learning for bayesian optimization using ranking-weighted gaussian process ensembles. In *AutoML Workshop at ICML*, volume 7, page 5, 2018.

P. I. Frazier. A tutorial on bayesian optimization. *CoRR*, abs/1807.02811, 2018.

R. Garnett. *Bayesian optimization*. Cambridge University Press, 2023.

K. Hakhamaneshi, P. Abbeel, V. Stojanovic, and A. Grover. JUMBO: scalable multi-task bayesian optimization using offline data. *CoRR*, abs/2106.00942, 2021.

A. Harutyunyan, W. Dabney, T. Mesnard, M. G. Azar, B. Piot, N. Heess, H. van Hasselt, G. Wayne, S. Singh, D. Precup, and R. Munos. Hindsight credit assignment. In *Advances in Neural Information Processing Systems 32*, pages 12467–12476, 2019.

D. Huang, L. Filstroff, P. Mikkola, R. Zheng, and S. Kaski. Bayesian optimization augmented with actively elicited expert knowledge. *CoRR*, abs/2208.08742, 2022.

C. Hvarfner, F. Hutter, and L. Nardi. Joint entropy search for maximally-informed bayesian optimization. In *Advances in Neural Information Processing Systems 35*, pages 11494–11506, 2022a.

C. Hvarfner, D. Stoll, A. L. F. Souza, M. Lindauer, F. Hutter, and L. Nardi. pibo: Augmenting acquisition functions with user beliefs for bayesian optimization. In *The Tenth International Conference on Learning Representations*, 2022b.

C. Hvarfner, E. O. Hellsten, and L. Nardi. Vanilla bayesian optimization performs great in high dimensions. In *Proceedings of the 41nd International Conference on Machine Learning*. PMLR, 2024a.

C. Hvarfner, F. Hutter, and L. Nardi. A general framework for user-guided bayesian optimization. In *The Twelfth International Conference on Learning Representations*, 2024b.

L. P. Kaelbling and T. Lozano-Pérez. Learning composable models of parameterized skills. In *2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017*, pages 886–893. IEEE, 2017.

K. Kandasamy, J. G. Schneider, and B. Póczos. High dimensional bayesian optimisation and bandits via additive models. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 295–304. PMLR, 2015.

D. Khatamsaz, R. Neuberger, A. M. Roy, S. H. Zadeh, R. Otis, and R. Arróyave. A physics informed bayesian optimization approach for material design: application to niti shape memory alloys. *npj Computational Materials*, 9 (1):221, 2023.

Y. Kim, R. Allmendinger, and M. López-Ibáñez. Safe learning and optimization techniques: Towards a survey of the state of the art. In *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, pages 123–139. Springer, 2020.

W. Kobayashi, T. Otsuka, Y. K. Wakabayashi, and G. Tei. Physics-informed bayesian optimization suitable for extrapolation of materials growth. *npj Computational Materials*, 11(1):36, 2025.

P. W. Koh and P. Liang. Understanding black-box predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1885–1894. PMLR, 2017.

M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

C. Li, S. Rana, S. Gupta, V. Nguyen, and S. Venkatesh. Bayesian optimization with monotonicity information. In *Workshop on Bayesian optimization at neural information processing systems (NIPSW)*, 2017.

D. J. MacKay et al. Introduction to gaussian processes. *NATO ASI series F computer and systems sciences*, 168: 133–166, 1998.

N. Mahboubi, J. Xie, and B. Huang. Point-by-point transfer learning for bayesian optimization: An accelerated search strategy. *Comput. Chem. Eng.*, 194:108952, 2025.

R. Martinez-Cantin, K. Tee, and M. McCourt. Practical bayesian optimization in the presence of outliers. In *International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1722–1731. PMLR, 2018.

T. Mesnard, T. Weber, F. Viola, S. Thakoor, A. Saade, A. Harutyunyan, W. Dabney, T. S. Stepleton, N. Heess, A. Guez, E. Moulines, M. Hutter, L. Buesing, and R. Munos. Counterfactual credit assignment in model-free reinforcement learning. volume 139 of *Proceedings of Machine Learning Research*, pages 7654–7664. PMLR, 2021.

A. Meulemans, S. Schug, S. Kobayashi, N. Daw, and G. Wayne. Would I have gotten that reward? long-term credit assignment by counterfactual contribution analysis. In *Advances in Neural Information Processing Systems 36*, pages 68685–68735, 2023.

A. Ramachandran, S. Gupta, S. Rana, C. Li, and S. Venkatesh. Incorporating expert prior in bayesian optimisation via space warping. *Knowl. Based Syst.*, 195:105663, 2020.

A. L. F. Souza, L. Nardi, L. B. Oliveira, K. Olukotun, M. Lindauer, and F. Hutter. Bayesian optimization with a prior for the optimum. In *Machine Learning and Knowledge Discovery in Databases. Research Track - European Conference*, volume 12977 of *Lecture Notes in Computer Science*, pages 265–296. Springer, 2021.

N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, pages 1015–1022. PMLR, 2010.

Y. Sui, A. Gotovos, J. W. Burdick, and A. Krause. Safe exploration for optimization with gaussian processes. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 997–1005. PMLR, 2015.

K. Swersky, J. Snoek, and R. P. Adams. Multi-task bayesian optimization. In *Advances in Neural Information Processing Systems 26*, pages 2004–2012, 2013.

X. Wang, Y. Jin, S. Schmitt, and M. Olhofer. Recent advances in bayesian optimization. *ACM Comput. Surv.*, 55(13s): 287:1–287:36, 2023.

J. T. Wilson, F. Hutter, and M. P. Deisenroth. Maximizing acquisition functions for bayesian optimization. In *Advances in Neural Information Processing Systems 31*, pages 9906–9917, 2018.

J. Wu, M. Poloczek, A. G. Wilson, and P. I. Frazier. Bayesian optimization with gradients. In *Advances in Neural Information Processing Systems 30*, pages 5267–5278, 2017.

Appendix

A Iteration Illustration of CCGBO

Figure 3 shows that CCGBO locates the global optimum far more quickly than standard GP-UCB, reflected in a steeper decline of simple regret by iteration 4. Although CCGBO's posterior fit in non-critical regions can be less precise than that of standard UCB, our method deliberately allocates more evaluation budget to the high-value region. This focused sampling produces a more accurate local model around the true maximum, which in turn drives significantly better optimization performance.

B Proofs

Here, we provide the complete proofs for the Theorem introduced in Section 5.

B.1 Consistency of the MC proxy of the optimum

Assumptions 1 There exists $\beta_t > 0$ such that the GP-UCB confidence event $\mathcal{E}_t := \Big\{ |f(\mathbf{x}) - \mu_t(\mathbf{x})| \le \beta_t \, \sigma_t(\mathbf{x}) \;\; \forall \mathbf{x} \in \mathcal{X} \Big\}$ holds with probability at least $1 - \delta$ [Srinivas et al., 2010]. And there exist constants $c_0, c_1 > 0$ such that for the zero-mean posterior process $\varepsilon_t := f_t - \mu_t$ we have $\mathbb{E}\big[\sup_{\mathbf{x}\in\mathcal{X}} \varepsilon_t(\mathbf{x}) \;\big|\; \mathcal{D}_t \big] \le c_0 S_t$ [Adler and Taylor, 2007], and, conditionally on $\mathcal{D}_t$, the centered variables $Z_t^{(j)} - \mathbb{E}[Z_t^{(j)} \mid \mathcal{D}_t]$ are $(c_1 S_t)$-sub-Gaussian (equivalently, their ψ_2-norms are $\le c_1 S_t$) [Ledoux and Talagrand, 2013].

Theorem B.1. *Let $\mathcal{X} \subset \mathbb{R}^d$ be compact, and at iteration t let the GP posterior be $f_t \sim \mathcal{GP}(\mu_t, k_t)$ with $\sigma_t^2(\mathbf{x}) = k_t(\mathbf{x}, \mathbf{x})$ and $S_t = \sup_{\mathbf{x}\in\mathcal{X}} \sigma_t(\mathbf{x}) < \infty$. Draw K i.i.d. posterior sample paths $\{f_t^{(j)}\}_{j=1}^K$, and define $Z_t^{(j)} = \max_{\mathbf{x}\in\mathcal{X}} f_t^{(j)}(\mathbf{x})$ and $Z_t = \frac{1}{K}\sum_{j=1}^K Z_t^{(j)}$. Let $\mathbf{x}^\star \in \arg\max_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$. For any $\delta' \in (0,1)$, there exists a constant $C > 0$ (depending only on c_1) such that*

$$\Pr\left(\big| Z_t - f(\mathbf{x}^\star) \big| \le \underbrace{\beta_t S_t}_{\text{model bias}} + \underbrace{C\, S_t \sqrt{\tfrac{\log(1/\delta')}{K}}}_{\text{MC error}} \right) \;\ge\; 1 - \delta - \delta'. \tag{9}$$

Proof. By Assumptions 1. Since $\mathcal{X}$ is compact, maxima exist. Decompose $|Z_t - f(\mathbf{x}^\star)| \le T_1 + T_2$, where $T_1 = |Z_t - \mathbb{E}[Z_t^{(j)} \mid \mathcal{D}_t]|$ and $T_2 = |\mathbb{E}[Z_t^{(j)} \mid \mathcal{D}_t] - f(\mathbf{x}^\star)|$. By standard sub-Gaussian concentration for sample means (conditionally on $\mathcal{D}_t$), $T_1 \le C S_t \sqrt{\log(1/\delta')/K}$ with conditional probability $\ge 1 - \delta'$. Next, write $f_t^{(j)} = \mu_t + \varepsilon_t^{(j)}$ with zero-mean Gaussian $\varepsilon_t^{(j)}$; then

$$\mathbb{E}[Z_t^{(j)} \mid \mathcal{D}_t] = \mathbb{E}\Big[\sup_{\mathbf{x}\in\mathcal{X}} \big(\mu_t(\mathbf{x}) + \varepsilon_t^{(j)}(\mathbf{x}) \big) \;\Big|\; \mathcal{D}_t \Big]. \tag{10}$$

Using $\sup(a + b) \le \sup a + \sup b$, symmetry of $\varepsilon_t^{(j)}$ (since it is zero-mean Gaussian), we get $\big|\mathbb{E}[Z_t^{(j)} \mid \mathcal{D}_t] - \sup_{\mathbf{x}} \mu_t(\mathbf{x})\big| \le c_0 S_t$. On $\mathcal{E}_t$, $\big|\sup_{\mathbf{x}} \mu_t(\mathbf{x}) - f(\mathbf{x}^\star)\big| \le \beta_t S_t$, hence $T_2 \le (\beta_t + c_0) S_t$; absorbing c_0 into β_t yields $T_2 \le \beta_t S_t$. Finally, a union bound gives total probability at least $1 - \delta - \delta'$, which proves (6). □

B.2 Regret Bound

We denote the instantaneous regret at step t by $r_t = f(\mathbf{x}^*) - f(\mathbf{x}_t)$, where $\mathbf{x}^* = \arg\max_{\mathbf{x}} f(\mathbf{x})$, and the cumulative regret by $R_N = \sum_{t=1}^N r_t$. We will analyze the Credit-Weighted UCB (CCGBO) acquisition $\alpha_t(\mathbf{x}) = \big(1 - \lambda + \lambda\, w_t(\mathbf{x})\big)\big[\mu_{t-1}(\mathbf{x}) + \beta_t\, \sigma_{t-1}(\mathbf{x})\big]$, where $w_t(\mathbf{x}) = \big[c(\mathbf{x})\big]^{\frac{\tau}{1+t/M}}$, and we further assume $0 < c_{\min} \le c(\mathbf{x}) \le c_{\max} < \infty \; \forall \mathbf{x} \in \mathcal{X}$, $\lambda \in [0,1]$ is the credit-influence parameter, and $\{\beta_t\}$ is chosen as in [Srinivas et al., 2010]. We make the following standard assumption on the GP sample paths:

Assumption 2 (Lipschitz on derivatives) There exist constants $a, b > 0$ such that for every $j = 1, \ldots, p$,

$$\Pr\Big(\sup_{\mathbf{x}\in\mathcal{X}} \Big| \tfrac{\partial f}{\partial x_j}(\mathbf{x}) \Big| > L \Big) \;\le\; a \, \exp\big(-L^2/b^2\big). \tag{11}$$

Under this assumption and the usual choice of $\beta_t = O\big(\sqrt{\log(t^2\pi^2/6\delta)}\big)$, with probability at least $1-\delta$ we have simultaneously for all t and $\mathbf{x}$: $|f(\mathbf{x}) - \mu_{t-1}(\mathbf{x})| \le \beta_t\,\sigma_{t-1}(\mathbf{x})$.

Theorem B.2. *Define for each t, $A_t = 1-\lambda+\lambda\,c_{\min}^{\tau/(1+t/M)}$, $B_t = 1-\lambda+\lambda\,c_{\max}^{\tau/(1+t/M)}$. Observe that $0 < A_t \le B_t < \infty$ and $B_t/A_t \to 1$ as $t \to \infty$. Under Assumption 5 and the above high-probability event, the cumulative regret of CCGBO satisfies*

$$R_N^{\mathrm{ccg}} \;\le\; \max_{1\le t\le N}\Big(\tfrac{B_t}{A_t}\Big)\; R_N, \tag{12}$$

where $R_N = O\big(\sqrt{N\,\gamma_N\,\beta_N}\big)$ is the standard GP-UCB bound from [Srinivas et al., 2010] and γ_N is the maximum information gain. In particular, since $\max_t B_t/A_t = B_1/A_1 < \infty$, CCGBO matches the vanilla GP-UCB rate up to a constant factor that tends to 1 when $\lambda \to 0$ (so $A_t = B_t = 1$) or $c_{\max}/c_{\min} \to 1$.

Proof. Condition on the event $\mathcal{E}$ that for all $t, \mathbf{x}$, $f(\mathbf{x}) \le \mu_{t-1}(\mathbf{x}) + \beta_t\,\sigma_{t-1}(\mathbf{x})$, $f(\mathbf{x}) \ge \mu_{t-1}(\mathbf{x}) - \beta_t\,\sigma_{t-1}(\mathbf{x})$. By definition, $\mathbf{x}_t = \arg\max_{\mathbf{x}} a_t(\mathbf{x})$, so for the true maximizer $\mathbf{x}^*$, $a_t(\mathbf{x}^*) \le a_t(\mathbf{x}_t)$. Thus

$$\begin{aligned} &A_t\big[\mu_{t-1}(\mathbf{x}^*) + \beta_t\,\sigma_{t-1}(\mathbf{x}^*)\big] \\ &\le \big(1-\lambda+\lambda\,w_t(\mathbf{x}^*)\big)\big[\mu_{t-1}(\mathbf{x}^*) + \beta_t\,\sigma_{t-1}(\mathbf{x}^*)\big] \\ &\le a_t(\mathbf{x}_t) \le B_t\big[\mu_{t-1}(\mathbf{x}_t) + \beta_t\,\sigma_{t-1}(\mathbf{x}_t)\big]. \end{aligned} \tag{13}$$

Rearranging gives $\mu_{t-1}(\mathbf{x}^*)+\beta_t\,\sigma_{t-1}(\mathbf{x}^*) \le \Big(\frac{B_t}{A_t}\Big)\big[\mu_{t-1}(\mathbf{x}_t)+\beta_t\,\sigma_{t-1}(\mathbf{x}_t)\big]$. Since $f(\mathbf{x}_t) \ge \mu_{t-1}(\mathbf{x}_t) - \beta_t\,\sigma_{t-1}(\mathbf{x}_t)$, we obtain

$$\begin{aligned} r_t^{\mathrm{ccg}} &= f(\mathbf{x}^*) - f(\mathbf{x}_t) \\ &\le \Big(\tfrac{B_t}{A_t}\Big)\big[\mu_{t-1}(\mathbf{x}_t) + \beta_t\,\sigma_{t-1}(\mathbf{x}_t)\big] \\ &\quad - \big[\mu_{t-1}(\mathbf{x}_t) - \beta_t\,\sigma_{t-1}(\mathbf{x}_t)\big] \\ &\le \Big(\tfrac{B_t}{A_t}\Big)\,2\,\beta_t\,\sigma_{t-1}(\mathbf{x}_t) = \Big(\tfrac{B_t}{A_t}\Big)\,r_t^{\mathrm{UCB}}. \end{aligned} \tag{14}$$

where we define $r_t^{\mathrm{UCB}} = 2\,\beta_t\,\sigma_{t-1}(\mathbf{x}_t)$ as the standard UCB instantaneous regret bound. Summing over $t = 1, \dots, N$ gives

$$\begin{aligned} R_N^{\mathrm{ccg}} &= \sum\nolimits_{t=1}^{N} r_t^{\mathrm{ccg}} \\ &\le \max_{1\le t\le N}\Big(\tfrac{B_t}{A_t}\Big)\sum\nolimits_{t=1}^{N} r_t^{\mathrm{UCB}} \\ &= \max_{1\le t\le N}\Big(\tfrac{B_t}{A_t}\Big)\; R_N, \end{aligned} \tag{15}$$

and the result follows by the standard bound $R_N = O(\sqrt{N\,\gamma_N\,\beta_N})$. □

C CCGBO with other acquisition functions

CCGBO is a plug-and-play module and can be combined with acquisition functions beyond UCB. The counterfactual credit-based weight can be simply multiplied by other acquisition functions as a coefficient. As suggested, we have evaluated it with several others, including TS in Figure 4, logEI [Ament et al., 2023] in Figure 5, and JES [Hvarfner et al., 2022a] in Figure 6. We observe that CCGBO yields similar performance gains when paired with JES and TS, just as with UCB, whereas with logEI (an EI-family variant) the improvement is negligible. We believe this difference stems from the inherently greedy nature of EI-style methods: since they already concentrate heavily on nearby "promising" regions, multiplying by credit weights does not materially change their ranking or introduce substantial additional guidance. In contrast, acquisitions with explicit or implicit exploration (e.g., UCB, TS, JES) benefit more from the extra bias toward high-contribution regions while retaining their ability to explore.

These findings further confirm the generality of our credit-weighting approach across a variety of acquisition strategies, with performance improving or at least not degrading in most cases.

D High-dimensional Tasks

Tackling the intrinsic challenges of very high-dimensional BO is not the primary focus or contribution of our paper, but rather a distinct research topic in its own right. To nonetheless demonstrate CCGBO's behavior, we tested all methods on high-dimensional benchmarks.

However, directly comparing against algorithms specifically designed for high-D BO would be unfair, since those methods adapt their kernels or acquisition functions to that setting. Instead, we adopted the straightforward "Vanilla Bayesian Optimization Performs Great in High Dimensions" [Hvarfner et al., 2024a] strategy: in a standard GP + acquisition framework, one simply shifts the log-lengthscale prior upward by $\frac{1}{2}\log D$ (i.e. scales the lengthscale by $\sqrt{D}$) and fixes the signal variance to 1. We applied this modified GP backbone across all CCGBO and baseline algorithms to ensure a fair comparison.

In the experiments, we follow [Hvarfner et al., 2024a] to employ Embedded Test Functions to evaluate the performance of Bayesian optimization algorithms in high-dimensional spaces. These functions are constructed by embedding classical low-dimensional benchmark functions into higher-dimensional spaces. Given a d_{eff}-dimensional benchmark function $f : \mathbb{R}^{d_{eff}} \to \mathbb{R}$, we embed it into a D-dimensional space ($D \gg d_{eff}$) to construct a new objective function $f_{embed} : \mathbb{R}^D \to \mathbb{R}$: $f_{embed}(\mathbf{x}) = f(\mathbf{x}_{1:d_{eff}})$ where $\mathbf{x}_{1:d_{eff}}$ denotes the first d_{eff} components of the input vector $\mathbf{x} \in \mathbb{R}^D$. This means that the objective function value depends only on the first d_{eff} dimensions (effective dimensions), while the remaining $D - d_{eff}$ dimensions have no effect on the function value. In this work, two classical benchmark functions are adopted as bases: (1) Levy function: 4 effective dimensions, characterized by multiple local optima, used to evaluate global search capability. (2) Hartmann function: 6 effective dimensions, with a complex non-convex landscape, widely used in optimization benchmarking. These functions are embedded into 25, 100, 300, and 1000-dimensional spaces for experiments. For example, `Levy4_1000` denotes embedding the 4-dimensional Levy function into a 1000-dimensional space.

The resulting performance on high-dimensional tasks is reported in Figure 7. When these algorithms all get some performance improvements with the help of the algorithm proposed in "Vanilla Bayesian Optimization Performs Great in High Dimensions", they no longer struggle in high-dimensional tasks like standard GP. In such a comparison, our CCGBO still achieves the good results.

E Ablation Study on "half-life" parameter M

Figure 8 shows how varying the credit decay constant M affects simple regret. Recall that in our credit-weighted UCB, the per-iteration exponent on the normalized credit $\pi(\mathbf{x})$ is $\tau/(1 + n/M)$, so M controls how quickly the extra credit weighting fades: (1)Small M (e.g. $M = 10$). The exponent $\tau/(1 + n/M)$ drops rapidly as n increases, causing $w_n(\mathbf{x}) \to 1$ after only a few iterations. Consequently, the acquisition function reverts almost immediately to vanilla UCB, and the benefit of credit guidance is lost, slowing late-stage convergence. (2))Large M (e.g. $M \geq 30$). The credit exponent decays very slowly, so credit influence remains strong throughout most of the run. This over-emphasis on early high-credit regions leads to excessive exploitation and a premature plateau in the simple regret curve. (3))Moderate M (e.g. $15 \leq M \leq 25$). The exponent decays at a controlled rate: early iterations are properly biased by reliable credit signals, while later iterations smoothly return to a balanced exploration–exploitation regime. In our experiments, $M = 20$ yielded the lowest final simple regret and the most consistent downward trend.

F Ablation Study on MC surrogate

Figure 9 illustrates how the number of GP posterior samples K used in the Monte Carlo surrogate affects optimization performance: (1))Small K (e.g. $K \leq 10$). The proxy Z_t is estimated from very few samples, making the resulting credit scores noisy. Early iterations therefore receive unstable guidance, and the algorithm behaves similarly to standard UCB with erratic credit weighting, yielding modest gains over the baseline. (2)Moderate K (e.g. $15 \leq K \leq 30$). The proxy becomes sufficiently accurate to produce reliable credit signals, while still incurring moderate computation. In this regime, CCGBO attains the fastest and most consistent reduction in simple regret, with $K \approx 20 - 30$ achieving the lowest final regret. (3))Large K (e.g. $K \geq 60$). Further increasing K yields diminishing returns: the proxy is more precise, but the extra computational cost grows. These results demonstrate that a moderate MC surrogate size (around $K = 20 - 30$) strikes the best balance between proxy fidelity and computational efficiency, enabling CCGBO to leverage accurate counterfactual credits for rapid convergence without over-exploitation or excessive cost.

G Ablation Study on Nearest Neighbors H

Figure 10 evaluates how the number of neighbors H used in the KNN-based credit propagation affects optimization performance: (1)Small H (e.g. $H = 2 - 3$). Credit propagation is highly localized: each candidate inherits credit from only a few nearest samples, resulting in a very rugged credit field. Early iterations show erratic behavior and slower convergence due to noisy weight estimates. (2)Moderate H (e.g. $H = 4 - 7$)). The credit field balances locality and smoothness. Candidates in high-credit regions receive coherent weighting, while edge points still retain

exploratory potential. This setting yields the fastest decline in simple regret and the lowest final regret. (3)Large H (e.g. $H \geq 10$). Credit smoothing becomes excessive: distant, low-credit samples unduly influence each candidate. The credit contrast diminishes, causing the acquisition to revert toward uniform UCB behavior and slowing late-stage improvement. These results indicate that a moderate neighbor count ($H \approx 5 - 7$) provides the best trade-off between a stable yet discriminative credit field, enabling CCGBO to effectively leverage historical contributions without over-smoothing.

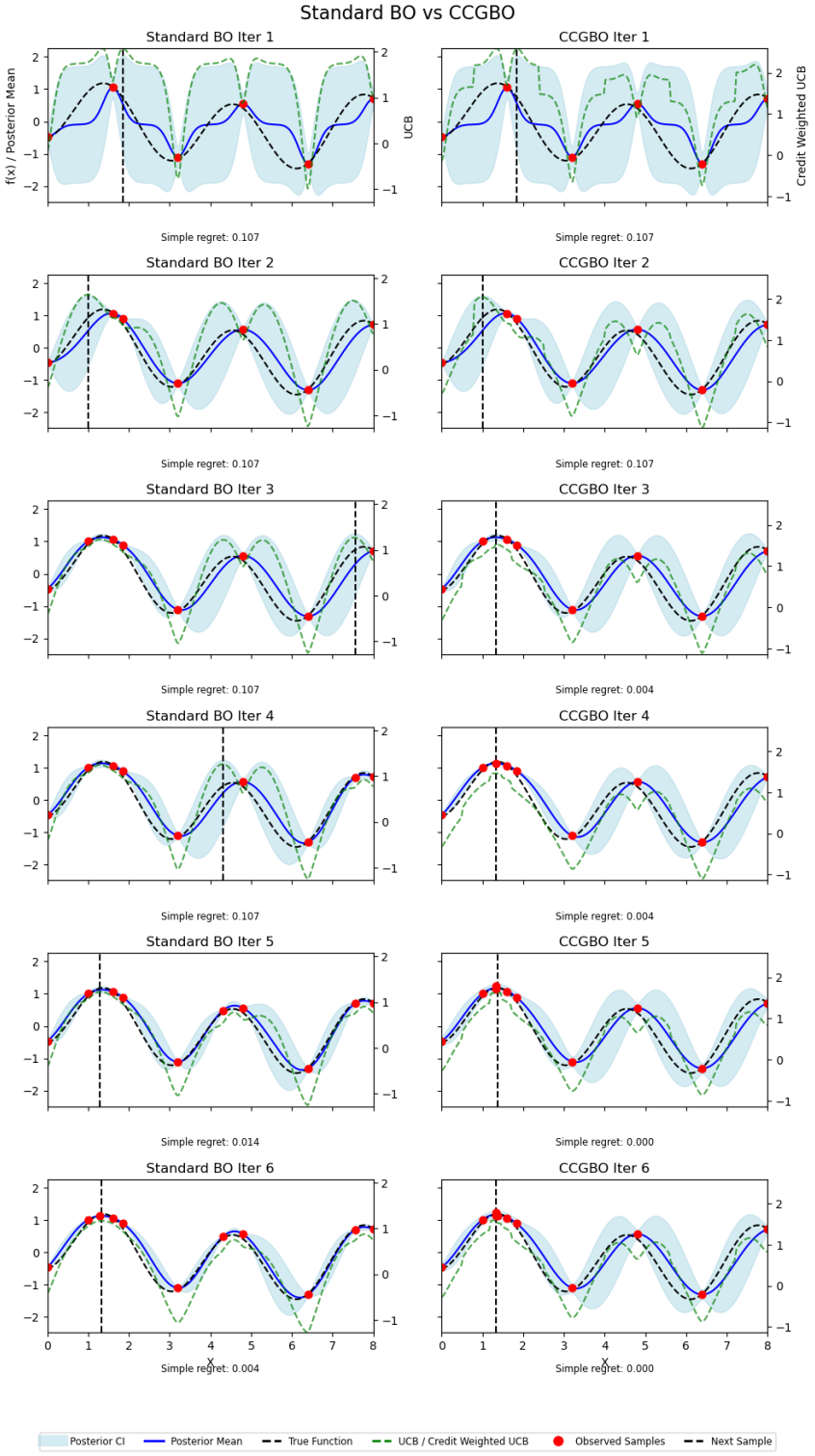

Figure 3: **Sequential optimization steps for Standard BO (left) vs. CCGBO (right).** Each row shows iterations 1 through 6 on a one-dimensional toy function. Blue curves and shaded regions: GP posterior mean and 95% credible interval. Black dashed curve: true objective. Green dashed line: standard UCB (*left*) or credit-weighted UCB (*right*). Red dots: observed samples. Vertical dashed line: next evaluation location. Simple regret is reported below each subplot.

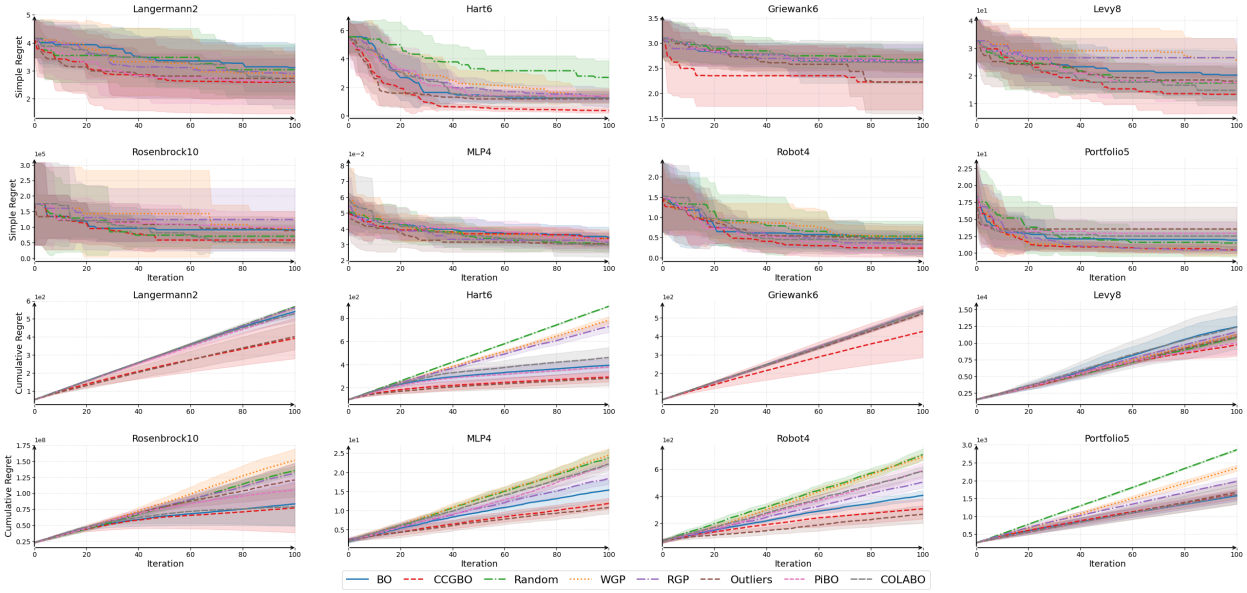

Figure 4: Cumulative regret and Simple regret versus iteration for 8 benchmark functions on TS. Each subplot plots cumulative regret over iterations t, comparing the eight baselines.

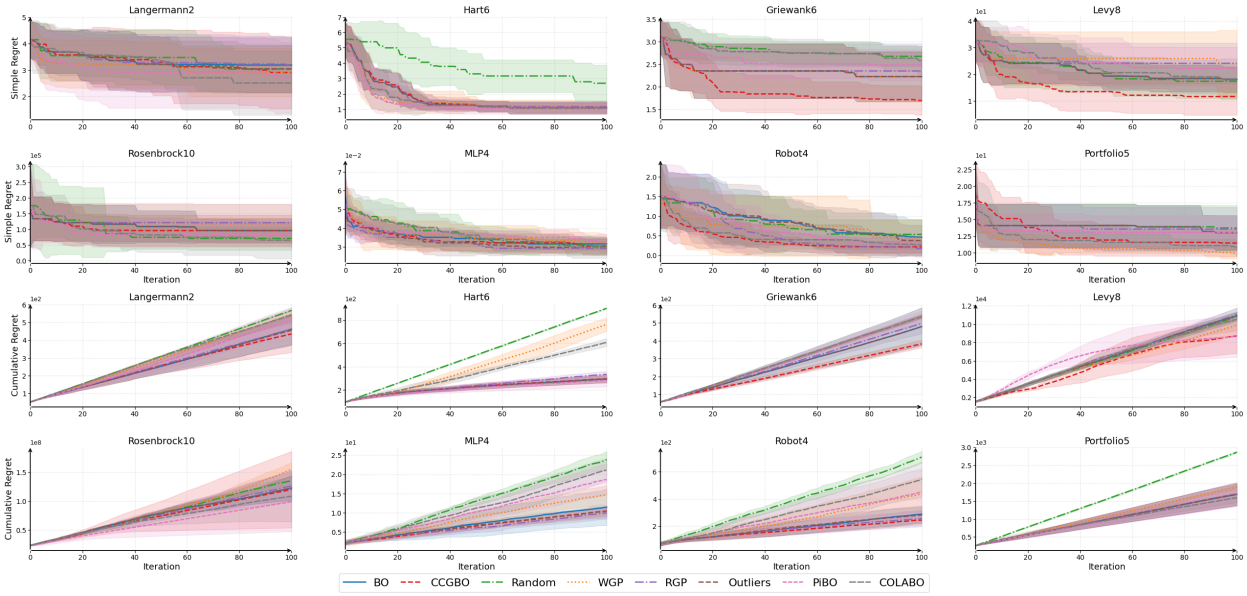

Figure 5: Cumulative regret and Simple regret versus iteration for 8 benchmark functions on LogEI. Each subplot plots cumulative regret over iterations t, comparing the eight baselines.

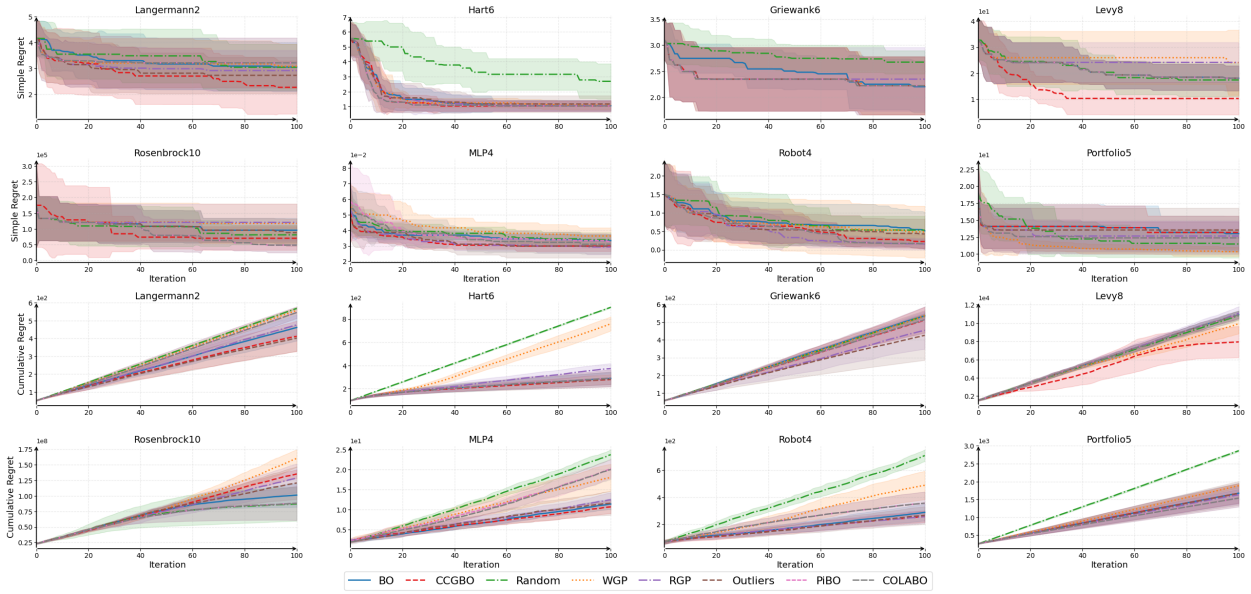

Figure 6: Cumulative regret and Simple regret versus iteration for 8 benchmark functions on JES. Each subplot plots cumulative regret over iterations t, comparing the eight baselines.

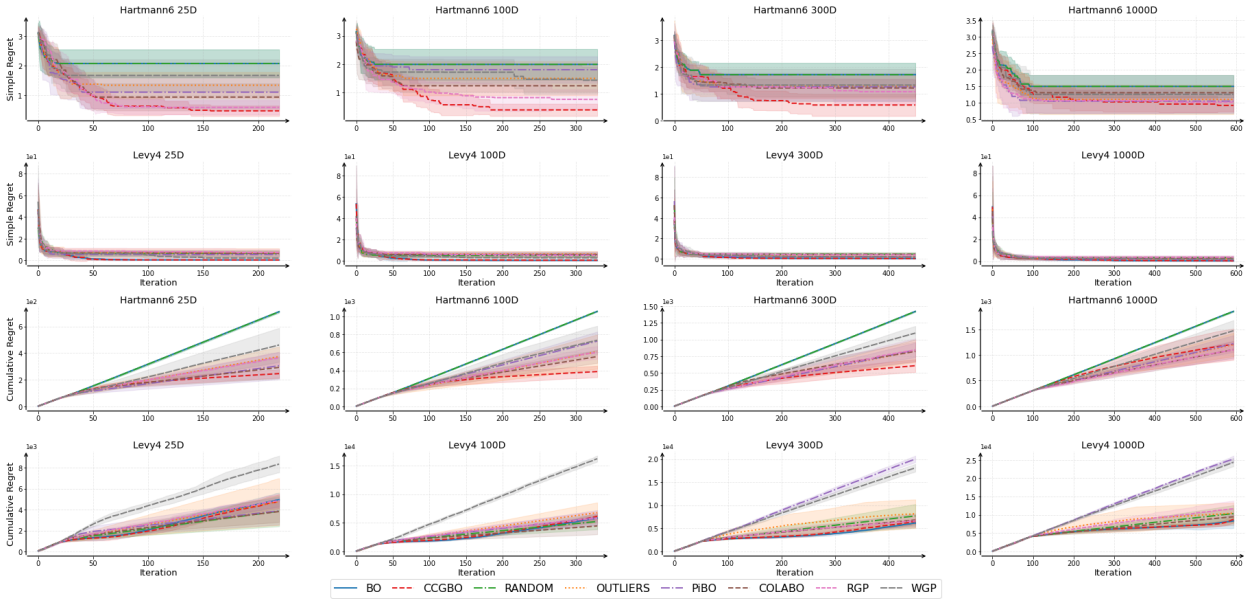

Figure 7: Simple regret versus iteration for eight high-dimensional test problems. Each subplot shows the mean simple regret over iterations t for CCGBO and baselines.

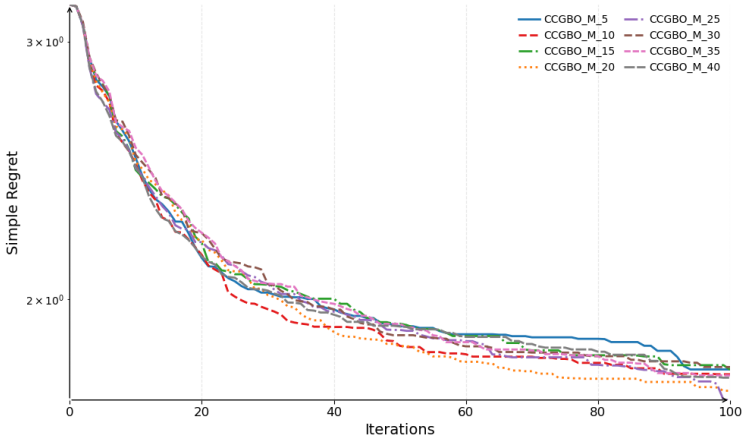

Figure 8: **Ablation of "half-life" parameter M on simple regret.** Curves show simple regret over iterations for CCGBO with $M \in \{5, 10, 15, 20, 25, 30, 35, 40\}$.

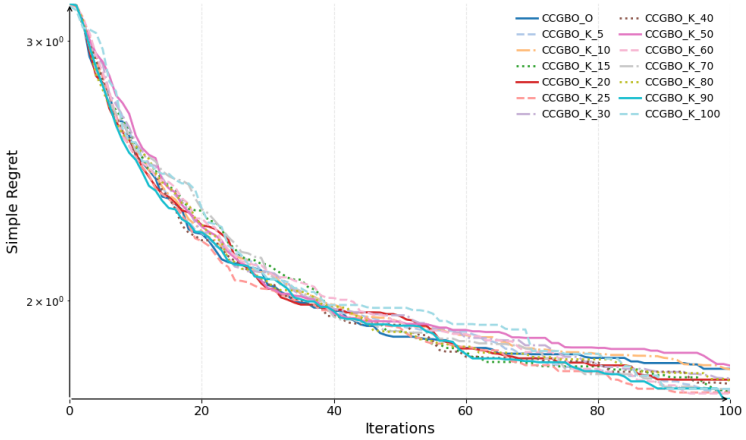

Figure 9: **Ablation of MC surrogate size K on simple regret.** Each curve plots the mean simple regret over iterations for CCGBO using different values of the credit decay parameter $M \in \{5, 10, 15, 20, 25, 30, 35, 40\}$.

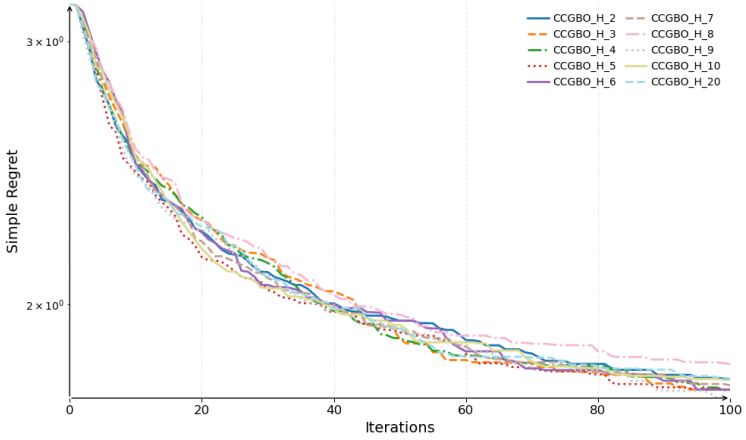

Figure 10: **Ablation of KNN neighbor parameter H on simple regret.** Each curve shows the mean simple regret over 50 runs for CCGBO using different numbers of nearest neighbors $H \in \{2, 3, \ldots, 10, 20\}$ to propagate discrete credits to continuous candidate points.