Relation-Augmented Dueling Bayesian Optimization via Preference Propagation

Xiang Xia[1], **Xiang Shu**[1], **Shuo Liu**[1,2], **Yiyi Zhu**[1], **Yijie Zhou**[1], **Weiye Wang**[1], **Bingdong Li**[1] and **Hong Qian**[1*]

[1]Shanghai Institute of AI Education, and School of Computer Science and Technology, East China Normal University, Shanghai 200062, China
[2]Game AI Center, Tencent Inc, Shenzhen 518057, China
{10225102442, 51255901138}@stu.ecnu.edu.cn, seokliu@tencent.com, {51265901040, 10235102524, 10215102506}@stu.ecnu.edu.cn, {bdli, hqian}@cs.ecnu.edu.cn

Abstract

In black-box optimization, when directly evaluating the function values of solutions is very costly or infeasible, access to the objective function is often limited to comparing pairs of solutions, which yields dueling black-box optimization. Dueling optimization is solely based on pairwise preferences, and thus notably reduces cost compared with function value based methods. However, the optimization performance of dueling optimization is often limited due to that most existing dueling optimization methods do not make full use of the pairwise preferences collected. To better utilize these preferences, this paper proposes relation-augmented dueling Bayesian optimization (RADBO) via preference propagation. By considering solution similarity, RADBO aims to uncover the potential dueling relations between solutions within different preferences through the proposed preference propagation technique. Specifically, RADBO first clusters solutions using a Gaussian mixture model. After obtaining the solution set with the highest intra-cluster similarity, RADBO utilizes a directed hypergraph to model the potential dueling relations between solutions, thereby realizing relation augmentation. Extensive experiments are conducted on both synthetic functions and real-world tasks such as motion control, car cab design and spacecraft trajectory optimization. The experimental results disclose the satisfactory accuracy of augmented preferences in RADBO, and show the superiority of RADBO compared with existing dueling optimization methods. Notably, it is verified that, under the same evaluation cost budget, RADBO can be competitive with or even surpass the function value based Bayesian optimization methods with respect to optimization performance.

1 Introduction

Black-box optimization [Conn *et al.*, 2009; Yu *et al.*, 2025], also termed as derivative-free optimization, is a class of optimization methods designed for situations where the objective function is unknown, complex, or expensive to evaluate. It enables global search for the optimal solution, with Bayesian optimization (BO) [Garnett, 2023; Mei *et al.*, 2023; Shahriari *et al.*, 2016] as a representative. Due to the significant advantages and progress of black-box optimization, it has been widely applied in fields such as chemical synthesis [Shields *et al.*, 2021], reinforcement learning [Qian and Yu, 2021], machine learning [Freund and Schapire, 1997; Elsken *et al.*, 2019] and intelligent education [Li *et al.*, 2025].

In traditional black-box optimization, evaluating the numerical objective function values is typically necessary. However, in many real-world scenarios, acquiring the objective function values can be extremely costly or entirely infeasible [Brochu *et al.*, 2010] and it has been found that comparing two solutions by preferences is relatively cheaper than scoring solutions [Kahneman and Tversky, 1979], like A/B tests [Siroker and Koomen, 2013]. Thus, dueling or preferential optimization has been developed as an easier and cheaper alternative, e.g., dueling Bayesian optimization [González *et al.*, 2017]. Instead of relying on function values, dueling optimization leverages preferences (i.e., which solution is preferred between two solutions) to guide the optimization, extending black-box optimization to the scenarios where objective function values are unavailable. Dueling optimization has been successfully applied in a wide range of fields, such as visual design optimization [Koyama *et al.*, 2020] and robotic gait optimization [Li *et al.*, 2021], showcasing its adaptability and effectiveness across various domains.

However, most existing dueling optimization methods, such as dueling Bayesian optimization, typically make simple use of pairwise preferences, without delving deeper into the dueling relations among different preferences. It can be found that these insufficient utilization of pairwise preferences can significantly impact the performance of dueling optimization, while making a fuller use of the preferences can improve the dueling optimization process.

Problem. Although dueling optimization can optimize using only low-cost pairwise preferences, the insufficient exploitation of preferences of existing methods could significantly limit the optimization performance of dueling optimization, e.g., dueling Bayesian optimization.

Contribution. This paper focuses on alleviating the lim-

*Corresponding Author.

itations in dueling optimization performance caused by the insufficient utilization of the pairwise preferences. To this end, we propose the relation-augmented dueling Bayesian optimization (RADBO) method via preference propagation. RADBO aims to uncover the potential dueling relations between different preferences through the proposed preference propagation technique based on solution similarity. RADBO begins by clustering solutions using a Gaussian mixture model and selects the cluster with the highest intra-cluster similarity. It then models potential dueling relations between solutions with a directed hypergraph to achieve relation augmentation. The experimental results reveal that the preferences augmented by the preference propagation technique achieve satisfactory accuracy, indicating that RADBO further utilizes preferences, and comparative experiments verify its superiority over existing dueling optimization methods. Notably, it is verified that, within the same evaluation cost budget, the performance of RADBO can match or even surpass that of function value based Bayesian optimization methods, when function value evaluations are relatively more expensive compared with comparing solutions.

The following sections review related work and preliminaries, describe the proposed RADBO method, present experimental results, and conclude the paper.

2 Related Work

This section provides a brief overview of the related work, i.e., dueling Bayesian optimization.

Dueling Bayesian optimization (DBO) extends BO to scenarios where direct access to the objective function is unavailable, but information about user preferences can be obtained. González *et al.* [2017] propose a framework called preferential Bayesian optimization (PBO), which leverages pairwise preferences to fit a Gaussian process (GP) [Rasmussen and Williams, 2006] within preference function domain. The PBO employs the dueling-Thompson Sampling (DTS) to determine the potential optimal solution and the solution with high uncertainty as candidates for the next duel. Benavoli *et al.* [2021] prove that the true posterior distribution of the preference function is a skewed Gaussian process (SkewGP), and incorporate SkewGP to enhance the performance of dueling Bayesian optimization. Based on the work of Benavoli *et al.* [2021], Takeno *et al.* [2023] propose a practical method, which ensures high computational efficiency and low sample complexity. Due to the lack of theoretical guarantees for most acquisition functions in dueling Bayesian optimization, Astudillo *et al.* [2023] introduce qEUBO, a promising acquisition function with a grounded decision-theoretic justification. Guided by the optimism principle, POP-BO [Xu *et al.*, 2024] constructs a confidence set from preferences and employs an optimistic strategy that ensures a bound on cumulative regret, enabling it to effectively report an estimated best solution with guaranteed convergence. To address the dimensionality issue exacerbated by modeling the preference function, PE-DBO [Zhang *et al.*, 2023] extends the concept of intrinsic effective dimensionality to preference function. Despite these advancements, these methods still do not fully utilize the available pairwise preferences, which continues to impact the performance of dueling optimization.

Unlike PBO, kernel-self-sparring (KSS) [Sui *et al.*, 2017] and comp-GP-UCB (COMP-UCB) [Xu *et al.*, 2020] do not construct a surrogate model to fit the preference function. KSS uses a GP to model the function, where the function value represents the probability of one solution beating the optimal solution, rather than modeling a preference function. COMP-UCB employs the Borda function, inspired by the Borda score [Sui *et al.*, 2018], to replace the preference function and regards the average performance of all solutions as the basis for comparison. While these methods simplify the dueling optimization problems compared to the methods that model the preference function, they may still face challenges caused by the insufficient utilization of pairwise preferences, leading to a poor optimization performance.

DBO differs significantly from reinforcement learning from human feedback (RLHF) [Christiano *et al.*, 2017], and direct preference optimization (DPO) [Rafailov *et al.*, 2023; Liu *et al.*, 2025] in both their objectives and methodologies. While these methods all involve preference optimization, RLHF and DPO primarily focus on optimizing behavioral strategies based on human feedback to address sequential decision-making problems. RLHF emphasizes feedback-driven guidance in reinforcement learning, whereas DPO refines preference comparisons to enhance model performance. In contrast, DBO focuses on identifying the global optimum in black-box optimization problems through Bayesian methods, prioritizing efficient search in the function space. Therefore, although all three techniques involve preference optimization, they are applied in distinct problem domains with different methodological approaches.

3 Preliminaries

3.1 Dueling Optimization

Consider a black-box function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \subset \mathbb{R}^D$, which is costly to evaluate. The goal of global optimization is to find the optimal solution $\boldsymbol{x}^* = \text{argmax}_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$ in a D-dimensional continuous solution space. Instead of directly evaluating numerical function values, dueling optimization evaluates the objective function by comparing pairs of solutions $(\boldsymbol{x}, \boldsymbol{x}')$, i.e., duels. For each duel, we obtain a preference indicating which solution is better. This preference is represented as binary information (i.e., 0 indicates $\boldsymbol{x}'$ is better, while 1 indicates $\boldsymbol{x}$ is better). These preferences are then used to guide the dueling optimization process. Throughout this paper, each duel is treated as a column vector, represented by $[\boldsymbol{x}; \boldsymbol{x}'] \in \mathbb{R}^{2D}$, where the space with dimension $2D$ is called dueling solution space.

Preference Function. In dueling optimization, the preference of a duel $[\boldsymbol{x}; \boldsymbol{x}']$ is sampled from a Bernoulli distribution, where the probability reflects the likelihood that solution $\boldsymbol{x}$ is preferred over $\boldsymbol{x}'$. Under the assumption that the probability of solution $\boldsymbol{x}$ being preferred over $\boldsymbol{x}'$ is positively correlated with the difference in their objective function values, i.e., $P(\boldsymbol{x} \succ \boldsymbol{x}') \propto f(\boldsymbol{x}) - f(\boldsymbol{x}')$, the logistic function is commonly used to convert this difference into a probability. Therefore, the preference function in the dueling solution space can be formulated as $\pi_f([\boldsymbol{x}; \boldsymbol{x}']) = P(\boldsymbol{x} \succ$

$\boldsymbol{x}') = 1/(1 + e^{-[f(\boldsymbol{x})-f(\boldsymbol{x}')]})$, where $\pi_f([\boldsymbol{x}; \boldsymbol{x}'])$ represents the probability that solution $\boldsymbol{x}$ is preferred over solution $\boldsymbol{x}'$.

Copeland Score. To find the optimal solution $\boldsymbol{x}^*$, we introduce the concept of the *Condorcet winner* [González *et al.*, 2017], an extension from multi-armed bandit tasks, which is the solution that outperforms all others. However, in dueling optimization, a strict Condorcet winner cannot be obtained, so the solution with the highest Copeland score is selected as the best one. Due to that the objective function is continuous, the normalized Copeland score is defined as $S(\boldsymbol{x}) = \mathrm{Vol}(\mathcal{X})^{-1} \int_{\mathcal{X}} \mathbb{I}_{\{\pi_f([\boldsymbol{x};\boldsymbol{x}'])\geq 0.5\}} \, \mathrm{d}\boldsymbol{x}'$, where $\mathrm{Vol}(\mathcal{X})^{-1} = \int_{\mathcal{X}} 1 \mathrm{d}\boldsymbol{x}'$ is a normalizing constant that ensures $S(\boldsymbol{x})$ is in the $[0, 1]$ range and $\mathbb{I}_{\{\cdot\}}$ is the indicator function. For the optimal solution $\boldsymbol{x}^*$, $\pi_f([\boldsymbol{x}^*; \boldsymbol{x}']) \geq 0.5$ holds for all solutions, which implies that $S(\boldsymbol{x}^*) = \mathrm{Vol}(\mathcal{X})^{-1} \int_{\mathcal{X}} 1 \mathrm{d}\boldsymbol{x}' = 1$. The difficulty in calculating the normalized Copeland score limits its applicability in dueling optimization, thus the soft-Copeland score [González *et al.*, 2017] is adopted, which has the empirically same maximum as the normalized Copeland score. The soft-Copeland score is defined as $C(\boldsymbol{x}) = \mathrm{Vol}(\mathcal{X})^{-1} \int_{\mathcal{X}} \pi_f([\boldsymbol{x}; \boldsymbol{x}']) \mathrm{d}\boldsymbol{x}'$.

Dueling Bayesian Optimization. Dueling Bayesian Optimization [González *et al.*, 2017] is a representative method for dueling optimization. It leverages a Gaussian process (GP) to model the preference function in a dueling solution space. The process involves optimizing an acquisition function to determine the next duel, querying the preference function for the preference of the duel, and updating the dataset to refine the GP model. This cycle is repeated until a predefined number of iterations is reached. A detailed description can be found in the Appendix A and Appendix A-F can be found in https://github.com/X-Xia0828/RADBO.

3.2 Sampling Strategies in Dueling Bayesian Optimization

Rather than classifying dueling optimization methods based on the surrogate models (see Section 2), we categorize them according to whether the current best solution is used as one of the candidate solutions in the next duel. For methods where one solution in the duel is fixed, such as HB [Takeno *et al.*, 2023] and POP-BO [Xu *et al.*, 2024], the first solution is selected as the current best, while the second solution is resampled based on a given acquisition function. In this case, pairwise preferences are not entirely independent, as there is a common solution in the duels of consecutive comparisons, which allows a part of relations between different preferences to be inferred. However, this strategy limits the ability of methods to explore the solution space. In contrast, in the second type of methods, both solutions in a candidate duel are resampled through acquisition functions, with PBO [González *et al.*, 2017] being a typical example. PBO uses DTS to choose the potential optimal solution and the most uncertain one for the next duel, balancing exploration and exploitation. However, these approaches lead to pairwise preferences being more isolated, making it challenging to obtain the dueling relations between different preferences.

Algorithm 1 Relation-Augmented Dueling Bayesian Optimization (RADBO)

Input: Initial dataset $\mathcal{D}_M = \{[\boldsymbol{x}_i; \boldsymbol{x}'_i], p_i\}_{i=1}^M$, number of available duels N, boundary of subspace $\mathcal{X} \subset \mathbb{R}^D$ and preference propagation parameter k.
Procedure:
for $j = M$ **to** $M + N - 1$ **do**
  Fit a $\mathcal{GP}$ to $\mathcal{D}_j$ and perform preference propagation with parameter k to get the augmented dataset $\mathcal{D}_j^+$.
  Fit a $\mathcal{GP}^+$ to $\mathcal{D}_j^+$ and learn $\pi_{f_p,j}([\boldsymbol{x}; \boldsymbol{x}'])$.
  Sample a function $\pi_{\hat{f}_p}$ from $\mathcal{GP}^+$.
  $\boldsymbol{x}_{\text{next}} = \text{argmax}_{\boldsymbol{x}\in\mathcal{X}} \int_{\mathcal{X}} \pi_{\hat{f}_p}([\boldsymbol{x}; \boldsymbol{x}']; \mathcal{D}_j^+) \mathrm{d}\boldsymbol{x}'$.
  $\boldsymbol{x}'_{\text{next}} = \text{argmax}_{\boldsymbol{x}'\in\mathcal{X}} \sigma(\mathcal{GP}|\boldsymbol{x} = \boldsymbol{x}_{\text{next}}, \mathcal{D}_j)$.
  Run the duel $[\boldsymbol{x}_{\text{next}}; \boldsymbol{x}'_{\text{next}}]$ and obtain p_{j+1}.
  Augment $\mathcal{D}_{j+1} = \{\mathcal{D}_j \cup ([\boldsymbol{x}_{\text{next}}; \boldsymbol{x}'_{\text{next}}], p_{j+1})\}$.
end for
Fit a $\mathcal{GP}$ to $\mathcal{D}_{M+N}$ and find the solution $\boldsymbol{x}^*$ with the highest soft-Copeland score.
return $\boldsymbol{x}^*$.

In this paper, we focus on the second type of methods and aim to uncover the potential dueling relations through a preference propagation technique, thereby enhancing the performance of dueling optimization.

3.3 Directed Hypergraph

Directed hypergraphs [Bretto, 2013] are extensions of traditional graphs in which edges, called directed hyperedges, can connect multiple vertices from a source set to a target set, unlike traditional graphs that link pairs of vertex. This characteristic enables directed hypergraphs to naturally represent more complex, higher-order relationships, particularly excelling in modeling multi-party interactions. Consequently, directed hypergraphs are widely used in fields such as machine learning [Gao *et al.*, 2022], data mining [Ji *et al.*, 2020], and social network analysis [Lin *et al.*, 2009].

Formally, a directed hypergraph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of directed hyperedges. Each directed hyperedge $\varepsilon \in \mathcal{E}$ is an ordered pair of vertex subsets $(\mathcal{V}_s, \mathcal{V}_t)$, where $\mathcal{V}_s, \mathcal{V}_t \subseteq \mathcal{V}$ are disjoint source and target sets, with $\mathcal{V}_s \cap \mathcal{V}_t = \emptyset$. Directed hypergraphs provide a flexible way to model complex interactions between groups of vertices, avoiding the individual connections between each pair, as would be necessary in traditional graphs.

4 The Proposed Method

Although dueling optimization, e.g., preferential Bayesian optimization [González *et al.*, 2017], adapts well to scenarios where the objective function can only be evaluated through comparing a pair of solutions, the performance of the dueling optimization is also affected by the insufficient utilization of pairwise preferences (i.e., which solution is preferred). This section introduces the proposed method, relation-augmented dueling Bayesian optimization (RADBO), which aims to make fuller use of pairwise preferences and enhance the performance of dueling optimization through a preference prop-

agation technique. To clarify the explanation of the proposed method, we have included a notation section in Appendix B.

To make fuller use of the pairwise preferences and thus enhance the performance of dueling optimization, the RADBO method is proposed, with pseudo-code shown in Algorithm 1.

By utilizing a preference propagation technique (detailed in Section 4.1) to uncover potential dueling relations between different preferences, and employing PBO [González *et al.*, 2017] as the framework for this process, RADBO is proposed. The RADBO begins with an initial dataset $\mathcal{D}_M$, consisting of M evaluated pairwise preferences $\{[\boldsymbol{x}; \boldsymbol{x}'], p\}$, where p indicates whether one solution can beat the other (i.e., 0 means $\boldsymbol{x}'$ is better, and 1 means $\boldsymbol{x}$ is better.). In each iteration j, RADBO fits a Gaussian process $\mathcal{GP}$ as the surrogate model to the current dataset $\mathcal{D}_j$ and performs the preference propagation with parameter k to create an augmented dataset $\mathcal{D}_j^+$ (line 2). This augmented dataset includes both original preferences and additional dueling relations, allowing for fuller utilization of the existing pairwise preferences. A new GP model $\mathcal{GP}^+$ is then trained on $\mathcal{D}_j^+$ to learn the preference function $\pi_{f_p,j}([\boldsymbol{x}; \boldsymbol{x}'])$ (line 3). After training the surrogate models, the dueling-Thompson sampling [González *et al.*, 2017] acquisition function guides the sampling process. A sample function $\pi_{\hat{f}_p}$ is drawn from the new GP model $\mathcal{GP}^+$, which guides the selection of the first solution $\boldsymbol{x}_{\text{next}}$ (lines 4-5). Next, based on the $\mathcal{GP}$, the solution with the highest uncertainty is chosen as $\boldsymbol{x}'_{\text{next}}$ (line 6), resulting in a candidate duel $[\boldsymbol{x}_{\text{next}}; \boldsymbol{x}'_{\text{next}}]$. Then, the duel is evaluated, and the resulting preference p_{j+1} is used to update the dataset to $\mathcal{D}_{j+1}$ (lines 7-8). It is worth noting that the additional dueling relations generated by the preference propagation technique in each iteration do not carry over to the next iteration. After N iterations, the model $\mathcal{GP}$ is fit to the complete dataset $\mathcal{D}_{M+N}$, and the optimal solution $\boldsymbol{x}^*$ is determined based on the highest soft-Copeland score (line 10).

In the following sections, we will provide a detailed explanation of the preference propagation technique as well as the time and space complexity of the technique.

4.1 Preference Propagation Technique

Since insufficient utilization of pairwise preferences significantly impacts the performance of dueling optimization, a preference propagation technique is used to uncover potential relations between different preferences, enabling a fuller utilization of the preferences, with the pseudo-code in Appendix C. The preference propagation technique first clusters solutions using a clustering algorithm. Specifically, we employ the Gaussian mixture model [Reynolds *et al.*, 2009], which excels in capturing complex data distributions by modeling them as a combination of multiple Gaussian components. Other clustering algorithms such as k-means can also be used, and k-means yields similar results. After identifying the solution set with the highest intra-cluster similarity, the technique utilizes a hypergraph to model the dueling relations between solutions, achieving relation augmentation. This technique enables a fuller utilization of pairwise preferences, ultimately enhancing the optimization process.

Inspired by Sui *et al.* [2017], we construct an adaptive RBF kernel-based similarity model [Gardner *et al.*, 2018], denoted as $\mathcal{K}$. The kernel is initially set as RBF(1.0), and the model is used to fit a function where the value represents the probability that one solution beats the current optimal solution, ensuring that $\mathcal{K}$ operates in the D-dimensional solution space. Then, $\mathcal{K}$ can compute the covariance between any two solutions in the dataset, which can serve as a measure of similarity between solutions. Finally, these similarities are transformed into distances and clustering will be performed based on the distances, resulting in a set of solutions with the highest intra-cluster similarity (i.e., the smallest intra-cluster distance).

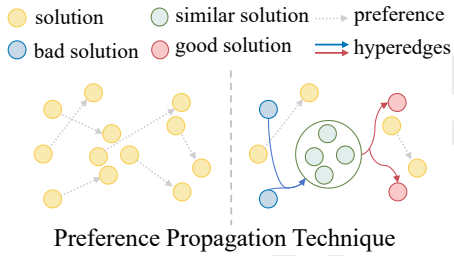

Figure 1: A diagram of the preference propagation technique. The pairwise preferences are modeled as a directed graph (left). During preference propagating, a set of similar solutions (green) is provided by a Gaussian mixture model and then a directed hypergraph is used to model the dueling relations between the different solutions (right).

As shown in Figure 1, pairwise preferences are modeled as a directed graph, where each vertex represents a solution, and each preference corresponds to a directed edge pointing from the worse solution to the better one. Next, preference propagation is conducted on the current dataset, with all solutions defined as set $\mathcal{V}$. The preference propagation technique first utilizes clustering based on the similarity model $\mathcal{K}$ to partition all solutions into k clusters and obtain a solution set with the highest intra-cluster similarity (the green circle), where the solutions in this set are termed *similar solutions* (the green vertices), and this set is defined as $\mathcal{V}_{\text{s}} \subseteq \mathcal{V}$. We assume that similar solutions exhibit analogous dueling relations, meaning that if solution A and solution B are similar and solution A is preferred over solution C, then solution B is also preferred over solution C. Subsequently, all solutions that can direct towards similar solutions via directed edges are termed *bad solutions* (the blue vertices), forming the set of the bad solutions $\mathcal{V}_{\text{bad}} \subseteq \mathcal{V}$, while all solutions that can be reached from similar solutions through directed edges are termed *good solutions* (the red vertices), forming the set of the good solutions $\mathcal{V}_{\text{good}} \subseteq \mathcal{V}$. The sets $\mathcal{V}_{\text{bad}}$, $\mathcal{V}_{\text{s}}$, and $\mathcal{V}_{\text{good}}$ have no intersection with each other. Finally, we construct a directed hypergraph $\mathcal{G}$ using two directed hyperedges. Specifically, ε_1 directs from the set of bad solutions to the set of similar solutions, i.e., ε_1 is an ordered pair of sets $(\mathcal{V}_{\text{bad}}, \mathcal{V}_{\text{s}})$, and ε_2 directs from the set of similar solutions to the set of good solutions, i.e., ε_2 is an ordered pair of sets $(\mathcal{V}_{\text{s}}, \mathcal{V}_{\text{good}})$.

Based on this directed hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} = \{\varepsilon_1, \varepsilon_2\}$, RADBO can uncover more potential dueling relations, i.e., all similar solutions are better than the bad solutions, and all good solutions are better than the similar solutions. Moreover, by leveraging the transitivity of preferences, we can also conclude that all good solutions are better than

the bad solutions. Thus, the preference propagation technique realizes relation augmentation based on the existing dataset, enabling a fuller utilization of the pairwise preferences.

4.2 Algorithmic Complexity Analysis

In this section, we analyze the improvements in time and space complexity achieved by using hypergraphs to model dueling relations between different solutions in the preference propagation technique.

The introduction of hypergraphs avoids the full connection that occurs when traditional graphs are used in the preference propagation technique. To establish the connections between the three solution sets, a traditional graph requires full connections from the bad solution set to the similar solution set, and from the similar solution set to the good solution set. We denote the quantities of bad solutions, similar solutions, and good solutions as n_1, n_2 and n_3, respectively. Specifically, in the case of using a traditional graph, the time complexity of modeling the relations between solutions is $O(n_1 \times n_2 + n_2 \times n_3)$, and the space complexity of the preference propagation technique is also $O(n_1 \times n_2 + n_2 \times n_3)$. However, when employing a hypergraph instead of a traditional graph, the two solution sets can be directly connected through a single hyperedge, resulting in the time complexity of modeling the relations reducing to $O(m)$, where m is the number of hyperedges and $m = 2$ in the preference propagation technique. Thus, the time complexity can also be expressed as $O(2)$. Additionally, the space complexity of the preference propagation technique also decreases to $O(m+n_1+n_2+n_3)$ with $m = 2$. The reduction in complexity brought about by the hypergraphs makes the preference propagation technique more efficient and practical.

5 Experiment

In this section, we compare RADBO with a series of dueling optimization algorithms through experiments on synthetic functions and real-world tasks. RADBO is implemented by BoTorch [Balandat *et al.*, 2020]. RADBO uses a Gaussian process with default parameters from the BoTorch library as the surrogate model, employs CMA-ES [Hansen *et al.*, 2003] as the optimizer of the acquisition function, and implements the Gaussian mixture model using the default parameters from the scikit-learn library. We compare RADBO with four dueling optimization methods, where both solutions in a candidate duel are resampled based on specific acquisition functions, rather than having one solution fixed as the current best, such as HB [Takeno *et al.*, 2023] and POP-BO [Xu *et al.*, 2024]. The methods include PBO [González *et al.*, 2017], KSS [Sui *et al.*, 2017], qEUBO [Astudillo *et al.*, 2023] and a simplified version of COMP-UCB [Xu *et al.*, 2020], which omits the second part of the optimization process that depends on function values. Specifically, PBO can be regarded as the version of RADBO after ablating the preference propagation technique. Further details of these methods are provided in the Appendix D. The experiments are designed to answer the following four significant questions.

Q1: Effectiveness and superiority: Can RADBO handle dueling optimization tasks and achieve better performance than other dueling optimization methods?

Q2: Utilization: Does RADBO uncover potential dueling relations based on the existing preferences and make fuller use of pairwise preferences?

Q3: The benefit of dueling optimization: Under a fixed budget, can RADBO match or even surpass the performance of function value based Bayesian optimization methods?

Q4: The impact of hyper-parameters: How sensitive is RADBO to changes in hyper-parameters?

The four questions are answered sequentially in this section. For all tasks, the best function value found so far is used as the evaluation criterion, with min-max scaling applied to the values to enable a more intuitive analysis of the results. The experimental code is publicly available at https://github.com/X-Xia0828/RADBO.

5.1 Experimental Settings

The Setting of Synthetic Functions. To evaluate the performance of RADBO, experiments are first conducted on synthetic functions. In this paper, we construct objective functions for evaluation in a standard setting based on different synthetic functions[1]. Specifically, let $f : \mathbb{R}^D \to \mathbb{R}$ be a base synthetic function, with its domain adjusted to $[-1, 1]^D$. The input is a D-dimensional vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_D]$, and the output is the function value $f(\boldsymbol{x})$. In the experiments, we evaluate RADBO on 18 synthetic functions. These synthetic functions collectively cover various optimization types, including multimodal landscapes, complex terrains, periodic variations, and convex optimization. All experiments on synthetic functions are maximization optimization.

The Setting of Real-world Tasks. To further explore the performance of RADBO and its applicability to real-world tasks, RADBO is evaluated on five real-world datasets. The first dataset is RobotPush problem [Eriksson *et al.*, 2019], which is a noisy 14-dimensional motion control problem involving optimizing the pre-image for pushing an object to a goal location. The second dataset is Sagas [Schlueter *et al.*, 2021], a 12-dimensional problem, which is designed for trajectory optimization problems, aiming to minimize the overall mission length to reach targets. The third dataset is a 10-dimensional problem, Cassini1-MINLP [Schlueter and Munetomo, 2019], which is designed to optimize a mixed-integer nonlinear programming problem (MINLP), allowing for flexible selection of any planet in the solar system. The remaining two tasks are a 5-dimensional animation optimization problem (Animation) and a 7-dimensional car cab design problem (Carcab). These real-world datasets are well-suited for dueling optimization. RobotPush is a noisy dataset where the noise affects the performance of function value based methods, while dueling optimization can mitigate the impact of noise to some extent. Cassini1-MINLP and Sagas are spacecraft trajectory optimization problems where evaluating the function value of a given solution may be very costly and time-consuming, while comparing a pair of solutions is much more manageable. Animation and Carcab are widely used in the prior work like qEUBO [Astudillo *et al.*, 2023]. All real-world tasks are maximization tasks.

[1] http://www.sfu.ca/~ssurjano/optimization.html

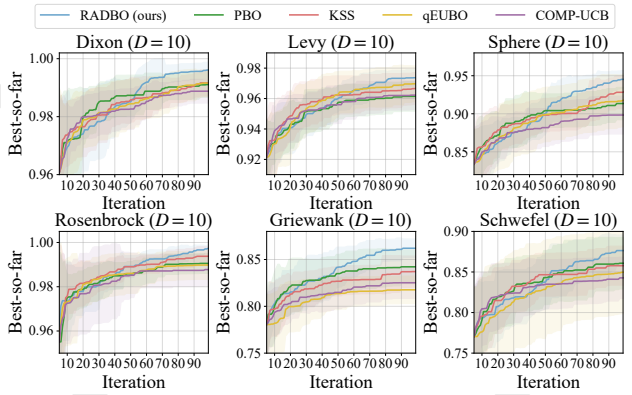

Figure 2: The best function value found by RADBO on synthetic functions are compared with different dueling optimization algorithms. All methods are evaluated with 5 initial duels, 95 iterations, and each experiment is repeated 20 times. The mean and standard deviation of the results are plotted. The horizontal axis of the plots represents the number of evaluations, and the vertical axis represents the best function value found by the algorithm.

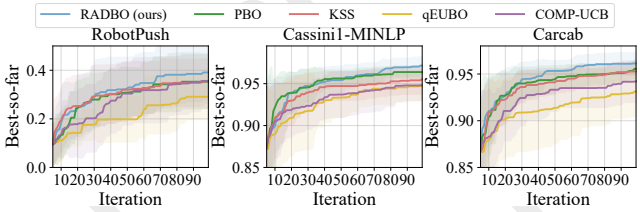

Figure 3: The best function value found by algorithms on three real-world datasets. Each experiment is repeated 20 times. The mean and standard deviation of the results are plotted. The horizontal axis of the plots represents the number of evaluations, and the vertical axis represents the best function value found by the algorithm. All methods are evaluated with 5 initial duels and 95 iterations.

5.2 The Performance of RADBO

To Q1: Effectiveness and Superiority. In the synthetic functions and real-world tasks experiments, $I = 500$ samples are employed to estimate the integral of the soft-Copeland score, and the GP model is initialized using $M = 5$ duels, followed by $N = 95$ duels for optimization. For RADBO, we use $k = 3$ to execute the preference propagation technique.

As shown in Figure 2, across most synthetic functions, RADBO consistently achieves better performance compared to the other optimization methods, showcasing its ability to handle dueling optimization tasks well. First, the RADBO curve converges relatively quickly and stays above other methods at around 50 iterations, indicating that it finds better solutions earlier in the optimization process. Moreover, RADBO shows a stable improvement in performance during optimization, particularly as other methods begin to converge around iterations 70 (a phenomenon we will explore further in the next section). Finally, the standard deviation of RADBO is relatively narrow in most cases, suggesting that its performance is more reliable compared to the other methods, particularly in challenging functions like Griewank.

As shown in Figure 3, across most real-world tasks, RADBO adapts well to the real-world tasks and achieves the best results. In RobotPush task, PBO, KSS, and COMP-UCB all achieve the similar final performance, as they are troubled by noise during optimization. However, due to the preference propagation technique, which uncovers many potential dueling relations from the existing preferences, RADBO can find the better solutions. In Cassini1-MINLP and Carcab tasks, RADBO exhibits a stable improvement throughout the optimization process, and ultimately achieves the best results.

The experiments on the remaining synthetic functions and real-world tasks yield similar conclusions, with results provided in the Appendix E. To verify that RADBO statistically outperforms other methods in most cases, the detailed results supported by t-tests are provided in the Appendix E.

In a nutshell, the experimental results verify that RADBO can handle dueling optimization tasks well and reflect the superiority of RADBO over other dueling optimization methods, which answers Q1.

To Q2: Utilization. To explore the utilization of pairwise preferences in RADBO and explain why RADBO shows a stable improvement in performance, we analyze the augmented preferences to better understand the factors driving the algorithm performance, as shown in Figure 4. The experiments are conducted on the Griewank function and three real-world tasks, with all settings consistent with those in the above section, and the experiments are repeated 20 times.

As shown in Figure 4, a significant number of augmented preferences maintaining satisfactory accuracy are newly added after preference propagation, which verifies that pairwise preferences are not fully utilized in previous work.

The Figure 4 (a) illustrates the results on the Griewank function, which we consider as an ideal environment. In the top plot, it is clear that as optimization progresses, the number of newly added augmented preferences significantly exceeds that of original preferences, with a faster growth rate as well. The bottom plot shows the mean accuracy of the augmented preferences, which increases steadily throughout the optimization process, consistently remaining above 0.5. Additionally, the lower accuracy of the augmented preferences during the early process of optimization may explain why RADBO performs worse than methods like PBO and KSS in certain situations, as shown in Figure 2, and as the accuracy of the augmented preferences increases, the performance of RADBO also improves rapidly. The Figure 4 (b) shows the results on the RobotPush task, and due to the presence of the noise, the accuracy of the augmented preferences is relatively low, but it remains consistently above 0.5. In this context, the preference propagation technique does not merely seek to propagate more relations, but instead uncovers a limited number of relations from the existing pairwise preferences, i.e., the scope of preference propagation is relatively narrow. This behavior ensures that the accuracy of the augmented preferences does not decline further, thereby preventing the newly generated dueling relations from affecting optimization performance. The Figure 4 (c) and (d) show the results on the Cassini1-MINLP and Carcab tasks, respectively. In both tasks, the augmented preferences all exhibit relatively high accuracy, which encourages the preference propagation technique to uncover more dueling relations from the existing

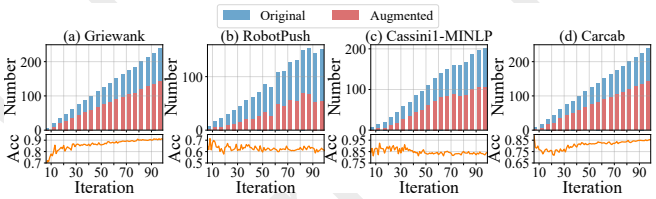

Figure 4: The utilization of RADBO on Griewank function and three real-world tasks. The figure shows the mean number of preferences in the original dataset (blue) and the augmented preferences **newly added** after preference propagation (red) in the top plot, with mean accuracy of the augmented preferences in the bottom plot. During the optimization process, RADBO uses a combination of original preferences and augmented preferences (blue + red). All settings are the same as Figure 2, 3. Each experiment is repeated 20 times.

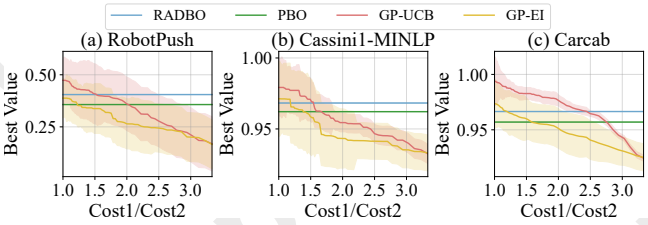

Figure 5: The best function value found by algorithms with a fixed budget of 100, where initialization uses 30 duels or solutions. Define the cost of observing the value of a function as Cost1, and the cost of comparing a pair of solutions as Cost2. The mean and standard deviation of the results are plotted. The vertical axis represents the best function value found by the algorithm and the horizontal axis of the plots represents is the Cost1/Cost2.

dataset, i.e., the scope of preference propagation is relatively broad. In the three real-world tasks, due to the complexity of the tasks, there is no gradual increase in accuracy of the augmented preferences as shown in Figure 4 (a).

In a nutshell, the results indicate that RADBO has effectively uncovered the potential dueling relations, thereby further utilizing the available preferences, which answers Q2.

5.3 Dueling Optimization vs. Function Value based Optimization

To Q3: The Benefit of Dueling Optimization. To verify that the performance of RADBO can match or even surpass that of function value based Bayesian optimization methods, RADBO and PBO (regarded as the ablated version of RADBO) are compared with the function value based methods, GP-UCB [Srinivas *et al.*, 2010] and GP-EI [Garnett, 2023]. With the results shown in Figure 5, all methods are tested on the real-world tasks and repeated 20 times. In Figure 5, the cost of evaluating the function value is defined as Cost1, while comparing a pair of solutions is defined as Cost2 = 1. Each method is allocated a budget of 100, with function value based methods initialized with 30 solutions and preference based methods with 30 duels. Other parameters match those in the previous section.

In Figure 5, we test different values of Cost1 and show the best function value found by each method under varying ratios of Cost1/Cost2. First, when the cost of observing a function value equals the cost of comparing a pair of solutions (Cost1 = Cost2), GP-UCB clearly outperforms preference based methods, which exhibits *a clear optimization performance gap between preference based and function value based methods*. However, as the cost of observing a function value increases, the advantage of preference based methods gradually becomes apparent, with RADBO surpassing GP-UCB and GP-EI when Cost1/Cost2 reaches approximately 1.5 to 2.5. This verifies that, under a fixed budget, RADBO can achieve optimization performance comparable to function value based Bayesian optimization methods. Finally, as Cost1 continues to increase, the optimization performance of preference based methods significantly exceeds that of GP-UCB, showcasing the advantages of preference based optimization in expensive black-box optimization problems.

In a nutshell, these results verify that, *under the same evaluation cost budget, when evaluating function values is relatively more expensive compared with comparing solutions, RADBO is competitive with or even surpass the function value based Bayesian optimization methods with respect to optimization performance*. It indicates for the first time that, if dueling relations between solutions within different preferences are fully and deeply exploited and utilized, dueling optimization could be more effective for expensive and costly optimization tasks, which answers Q3.

5.4 Hyper-parameter Analysis

To Q4: The Impact of Hyper-parameters. To explore the sensitivity of RADBO to different hyper-parameters, we conduct hyper-parameter experiments for k on 6 synthetic functions, with the results shown in Appendix F. It can be found that RADBO consistently outperforms PBO (the ablated version of RADBO) across different hyper-parameter k and is not significantly affected by changes in k, showcasing its insensitivity to hyper-parameter variations, which answers Q4. A more detailed analysis can be found in the Appendix F.

6 Conclusion

This paper aims to alleviate the limitations in dueling optimization performance caused by insufficient utilization of pairwise preferences. To address this limitation, we propose the method, relation-augmented dueling Bayesian optimization (RADBO), which enhances the performance of dueling optimization by capturing potential dueling relations between different solutions through the proposed preference propagation technique. Extensive experiments on synthetic functions and real-world tasks disclose the satisfactory accuracy of augmented preferences in RADBO, and exhibit the superiority of RADBO compared with existing dueling optimization methods. Notably, it is verified that, when function value evaluations are relatively more expensive than comparing solutions, the performance of RADBO can match or even surpass that of the function value based Bayesian optimization methods under the same cost budget.

In the future, we plan to conduct a theoretical justification for the proposed method. Specifically, the theoretical analysis includes convergence rate comparison between relation-augmented and traditional methods, and the accuracy of the augmented preferences. The challenge of these theoretical analyses stems mainly from the complexity of hypergraphs.

Ethical Statement

This work does not include any human subjects, personal data, or sensitive information. All testing datasets utilized are publicly accessible, and no proprietary or confidential information has been employed.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable and insightful comments. This work is supported by the National Natural Science Foundation of China (No. 62476091) and the National Undergraduate Training Program on Innovation and Entrepreneurship Grant (No. 202510269105G).

References

[Astudillo *et al.*, 2023] Raul Astudillo, Zhiyuan (Jerry) Lin, Eytan Bakshy, and Peter I. Frazier. qeubo: A decision-theoretic acquisition function for preferential Bayesian optimization. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pages 1093–1114, Valencia, Spain, 2023.

[Balandat *et al.*, 2020] Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. Botorch: A framework for efficient monte-carlo Bayesian optimization. In *Advances in Neural Information Processing Systems 33*, pages 21524–21538, Virtual Event, 2020.

[Benavoli *et al.*, 2021] Alessio Benavoli, Dario Azzimonti, and Dario Piga. Preferential Bayesian optimisation with skew Gaussian processes. In *Proceeedings of the 33th Genetic and Evolutionary Computation Conference*, pages 1842–1850, Lille, France, 2021.

[Bretto, 2013] Alain Bretto. Hypergraph theory. *An introduction. Mathematical Engineering. Cham: Springer*, 1, 2013.

[Brochu *et al.*, 2010] Eric Brochu, Vlad M Cora, and Nando De Freitas. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

[Christiano *et al.*, 2017] Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4299–4307, Long Beach, CA, 2017.

[Conn *et al.*, 2009] Andrew R. Conn, Katya Scheinberg, and Luís Nunes Vicente. *Introduction to Derivative-Free Optimization*, volume 8 of *MPS-SIAM series on optimization*. SIAM, 2009.

[Elsken *et al.*, 2019] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural architecture search: A survey. *Journal of Machine Learning Research*, 20:55:1–55:21, 2019.

[Eriksson *et al.*, 2019] David Eriksson, Michael Pearce, Jacob R. Gardner, Ryan Turner, and Matthias Poloczek. Scalable global optimization via local Bayesian optimization. In *Advances in Neural Information Processing Systems 33*, pages 5497–5508, Vancouver, Canada, 2019.

[Freund and Schapire, 1997] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[Gao *et al.*, 2022] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2022.

[Gardner *et al.*, 2018] Jacob R. Gardner, Geoff Pleiss, Kilian Q. Weinberger, David Bindel, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. In *Advances in Neural Information Processing Systems 31*, pages 7587–7597, Montréal, Canada, 2018.

[Garnett, 2023] Roman Garnett. *Bayesian Optimization*. Cambridge University Press, Cambridge, England, 2023.

[González *et al.*, 2017] Javier González, Zhenwen Dai, Andreas C. Damianou, and Neil D. Lawrence. Preferential Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 1282–1291, Sydney, Australia, 2017.

[Hansen *et al.*, 2003] Nikolaus Hansen, Sibylle D. Müller, and Petros Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[Ji *et al.*, 2020] Shuyi Ji, Yifan Feng, Rongrong Ji, Xibin Zhao, Wanwan Tang, and Yue Gao. Dual channel hypergraph collaborative filtering. In *Proceedings of the 26th International Conference on Knowledge Discovery and Data Mining*, pages 2020–2029, Virtual Event, 2020.

[Kahneman and Tversky, 1979] Daniel Kahneman and Amos Tversky. On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition*, 1979.

[Koyama *et al.*, 2020] Yuki Koyama, Issei Sato, and Masataka Goto. Sequential gallery for interactive visual design optimization. *ACM Transactions on Graphics*, 39(4):88, 2020.

[Li *et al.*, 2021] Kejun Li, Maegan Tucker, Erdem Biyik, Ellen R. Novoseller, Joel W. Burdick, Yanan Sui, Dorsa Sadigh, Yisong Yue, and Aaron D. Ames. ROIAL: region of interest active learning for characterizing exoskeleton gait preference landscapes. In *Proceedings of the 38th International Conference on Robotics and Automation*, pages 3212–3218, Xi'an, China, 2021.

[Li *et al.*, 2025] Mingjia Li, Hong Qian, Jinglan Lv, Mengliang He, Wei Zhang, and Aimin Zhou. Foundation model enhanced derivative-free cognitive diagnosis. *Frontiers of Computer Science*, 19(1):191318, 2025.

[Lin *et al.*, 2009] Yu-Ru Lin, Jimeng Sun, Paul C. Castro, Ravi B. Konuru, Hari Sundaram, and Aisling Kelliher. Metafac: community discovery via relational hypergraph factorization. In *Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining*, pages 527–536, Paris, France, 2009.

[Liu *et al.*, 2025] Shuo Liu, An Zhang, Guoqing Hu, Hong Qian, and Tat-Seng Chua. Preference diffusion for recommendation. In *Proceedings of the 13th International Conference on Learning Representations*, Singapore, Singapore, 2025.

[Mei *et al.*, 2023] Yongsheng Mei, Tian Lan, Mahdi Imani, and Suresh Subramaniam. A Bayesian optimization framework for finding local optima in expensive multimodal functions. In *Proceedings of the 26th European Conference on Artificial Intelligence*, volume 372, pages 1704–1711, Kraków, Poland, 2023.

[Qian and Yu, 2021] Hong Qian and Yang Yu. Derivative-free reinforcement learning: A review. *Frontiers of Computer Science*, 15(6):156336, 2021.

[Rafailov *et al.*, 2023] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36*, New Orleans, LA, 2023.

[Rasmussen and Williams, 2006] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.

[Reynolds *et al.*, 2009] Reynolds, Douglas A, et al. Gaussian mixture models. *Encyclopedia of Biometrics*, 741(659-663), 2009.

[Schlueter and Munetomo, 2019] Martin Schlueter and Masaharu Munetomo. A mixed-integer extension for esa's cassini1 space mission benchmark. In *Proceedings of the 18th Congress on Evolutionary Computation*, pages 912–919, Wellington, New Zealand, 2019.

[Schlueter *et al.*, 2021] Martin Schlueter, Mehdi Neshat, Mohamed Wahib, Masaharu Munetomo, and Markus Wagner. GTOPX space mission benchmarks. *SoftwareX*, 14:100666, 2021.

[Shahriari *et al.*, 2016] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.

[Shields *et al.*, 2021] Benjamin J. Shields, Jason M. Stevens, Jun Li, Marvin Parasram, Farhan N. Damani, Jesus I. Martinez Alvarado, Jacob M. Janey, Ryan P. Adams, and Abigail G. Doyle. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590(7844):89–96, 2021.

[Siroker and Koomen, 2013] Dan Siroker and Pete Koomen. *A/B Testing: The Most Powerful Way to Turn Clicks Into Customers*. Wiley Publishing, Hoboken, NJ, 2013.

[Srinivas *et al.*, 2010] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias 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, Haifa, Israel, 2010.

[Sui *et al.*, 2017] Yanan Sui, Vincent Zhuang, Joel W. Burdick, and Yisong Yue. Multi-dueling bandits with dependent arms. In *Proceedings of the 33th Conference on Uncertainty in Artificial Intelligence*, Sydney, Australia, 2017.

[Sui *et al.*, 2018] Yanan Sui, Masrour Zoghi, Katja Hofmann, and Yisong Yue. Advancements in dueling bandits. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 5502–5510, Stockholm, Sweden, 2018.

[Takeno *et al.*, 2023] Shion Takeno, Masahiro Nomura, and Masayuki Karasuyama. Towards practical preferential Bayesian optimization with skew Gaussian processes. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 33516–33533, Honolulu, HI, 2023.

[Xu *et al.*, 2020] Yichong Xu, Aparna Joshi, Aarti Singh, and Artur Dubrawski. Zeroth order non-convex optimization with dueling-choice bandits. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, volume 124 of *Proceedings of Machine Learning Research*, pages 899–908, Virtual Event, 2020.

[Xu *et al.*, 2024] Wenjie Xu, Wenbin Wang, Yuning Jiang, Bratislav Svetozarevic, and Colin N. Jones. Principled preferential Bayesian optimization. In *Proceedings of the 41th International Conference on Machine Learning*, volume 235, pages 55305–55336, Vienna, Austria, 2024.

[Yu *et al.*, 2025] Yang Yu, Hong Qian, and Yi-Qi Hu. *Derivative-Free Optimization: Theoretical Foundations, Algorithms, and Applications*. Springer, 2025.

[Zhang *et al.*, 2023] Yangwenhui Zhang, Hong Qian, Xiang Shu, and Aimin Zhou. High-dimensional dueling optimization with preference embedding. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, pages 11280–11288, Washington, DC, 2023.