FROM PREDICTION TO RANKING IN OFFLINE MODEL-BASED OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Offline model-based optimization (MBO) aims to discover high-performing designs using only a fixed dataset of past evaluations. Most existing methods learn a surrogate by pointwise regression, implicitly assuming that accurate value prediction is sufficient for reliable optimization. However, once the optimizer reaches a high-quality design, successful optimization depends on whether the surrogate ranks it above the remaining candidates, rather than on predicting their objective values accurately. To formalize this idea, we separate optimizer reachability from surrogate-based selection and define an optimization-oriented ranking risk between ε-optimal and non-ε-optimal candidates. We show that this ranking risk directly controls final selection error, whereas pointwise regression reaches the same optimization target only indirectly through regional prediction errors and the objective-margin distribution. The resulting bound further identifies the mismatch between training pairs and optimizer-induced target pairs as a key source of error, and relates this mismatch to the geometric separation of high-quality candidates from the offline data support. Guided by this analysis, we propose distribution-aware ranking (DAR), which concentrates training on high-quality-versus-remaining pairs to better match the optimization-relevant comparison distribution. Experiments on five Design-Bench tasks validate the predicted extrapolation difficulty and show that DAR achieves the best average rank among the evaluated methods.

1 INTRODUCTION

Offline model-based optimization (MBO) seeks high-performing designs for applications such as protein engineering (Reddy et al., 2024; Chen et al., 2023b; Takizawa et al., 2025), materials discovery (Shin et al., 2025), and engineering design (Kumar et al., 2022; Qian et al., 2025). It aims to maximize an unknown objective using only a fixed dataset of evaluated designs, without additional objective evaluations during optimization (Trabucco et al., 2022; Kim et al., 2026; Qian et al., 2025). This setting is relevant when testing new designs requires costly physical experiments or simulations. A common approach learns a surrogate model from the offline dataset and then optimizes its predictions to propose new designs (Fu & Levine, 2021; Qi et al., 2022). Although the surrogate predicts objective values, its role in optimization is to guide the selection of designs with high true performance.

Many surrogate-based offline MBO methods learn objective values through pointwise regression, commonly by minimizing mean squared error (MSE) on the offline dataset (Fu & Levine, 2021; Qi et al., 2022; Chen et al., 2023b). During optimization, however, candidate designs can move beyond the offline data distribution, where prediction errors may drive the optimizer toward designs with overestimated objective values (Trabucco et al., 2021; Yu et al., 2021). Existing methods address this problem through conservative predictions on out-of-distribution designs (Trabucco et al., 2021) or robust model adaptation around candidate solutions (Yu et al., 2021). Co-teaching exchanges pseudo-labeled data among surrogates (Yuan et al., 2023), while parallel mentoring uses pairwise supervision to improve the reliability of a surrogate ensemble (Chen et al., 2023a). Alongside these efforts to improve surrogate reliability under distribution shift, an important question is how the training objective should reflect the goal of selecting high-quality designs.

The purpose of offline MBO is to identify designs with high true objective values, so surrogate quality should be assessed by the designs it helps select. Once search has reached a near-optimal candidate,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ranking it above all inferior candidates in the generated set ensures that selecting the highest-scoring candidate returns a near-optimal design. In contrast, MSE penalizes deviations in predicted values without distinguishing errors that alter candidate selection from those that leave it unchanged. In their experiments, Tan et al. (2025) find that ranking quality is more strongly correlated with final design quality than MSE, motivating their ranking-based surrogate models. The next question is which training comparisons best reflect the distinction between near-optimal candidates and the inferior alternatives encountered during search.

Choosing these comparisons is difficult because the pairs available for training need not represent those encountered during optimization. Pairs drawn from the offline dataset compare observed designs, whereas the comparisons relevant to final selection involve near-optimal candidates and inferior alternatives produced by the optimizer. These candidates may lie outside the observed data support, so accurate ranking on training pairs alone need not ensure reliable selection during search. This mismatch raises two practical questions: how to construct training pairs that emphasize the comparisons governing final selection, and how separation between near-optimal candidates and the training data affects ranking on unseen designs.

In this work, we separate candidate generation from final selection to study how training-pair construction affects ranking-based offline MBO. The analysis connects the distribution of training comparisons to those encountered during search and motivates a distribution-aware pairing strategy. Our main contributions are summarized as follows.

- We bound the probability of selecting an inferior design using ranking error, conditional on reaching a near-optimal candidate and under stated sampling assumptions. The bound exposes the roles of training-pair mismatch and geometric data coverage.
- We develop distribution-aware ranking (DAR), which emphasizes comparisons between high-quality observations and the remaining data. These observations serve as an empirical proxy for near-optimal candidates, guiding pair construction toward the comparisons relevant to final design selection.
- Experiments on Branin examine ranking behavior under geometric shift, while five Design-Bench tasks evaluate final design quality. DAR achieves the best average rank across these tasks among the evaluated methods.

2 Problem Setup and Optimization Target

We first specify the design quality that offline optimization should achieve and the information available to the optimizer. Let $\mathcal{X} \subseteq \mathbb{R}^d$ denote the design space and $f : \mathcal{X} \to \mathbb{R}$ the unknown objective. Offline MBO has access only to a fixed dataset $S = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, where $y_i = f(\boldsymbol{x}_i) + \xi_i$ and ξ_i denotes observation noise. No additional evaluations of f are available during optimization. Let $\boldsymbol{x}^\star \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$ be a true maximizer. The objective is to return a design $\widehat{\boldsymbol{x}}_{\mathrm{MBO}}$ whose simple regret $f(\boldsymbol{x}^\star) - f(\widehat{\boldsymbol{x}}_{\mathrm{MBO}})$ is as small as possible.

Let $\widehat{h}$ denote the surrogate learned from S. We describe the candidates produced by a finite search procedure without assuming access to a global maximizer of $\widehat{h}$ over $\mathcal{X}$. Given an initialization $\boldsymbol{x}_0$ and search budget T, an optimizer $\mathcal{A}$ generates a candidate set $\mathcal{C} := \mathcal{C}_{\mathcal{A}}(\widehat{h}, \boldsymbol{x}_0, T) \subseteq \mathcal{X}$, and returns the candidate with the largest surrogate score:

$$\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \arg\max_{\boldsymbol{x}\in\mathcal{C}} \widehat{h}(\boldsymbol{x}). \tag{1}$$

The set $\mathcal{C}$ records the optimizer, its initialization, and its finite search budget. Under (1), only scores assigned to candidates in $\mathcal{C}$ determine the final selection; surrogate behavior elsewhere matters through its effect on the search procedure.

We next specify which designs count as successful. For a tolerance $\varepsilon \geq 0$, define the ε-optimal region, its complement, and the event that the candidate set reaches the target region:

$$\begin{aligned} \mathcal{X}_\varepsilon &:= \{\boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}^\star) - f(\boldsymbol{x}) \leq \varepsilon\}, \\ \mathcal{X}_\varepsilon^c &:= \{\boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}^\star) - f(\boldsymbol{x}) > \varepsilon\}, \\ \mathcal{R}_\varepsilon &:= \{\mathcal{C} \cap \mathcal{X}_\varepsilon \neq \varnothing\}. \end{aligned} \tag{2}$$

We refer to $\mathcal{X}_\varepsilon^c$ as the non-ε-optimal region. The definitions in (2) separate two requirements: the search procedure must generate at least one candidate in $\mathcal{X}_\varepsilon$, and the surrogate must select such a candidate from $\mathcal{C}$. We isolate the first requirement through the following coverage condition.

Assumption 1 (ε-reachability)**.** For the specified initialization scheme and search budget,

$$\Pr(\mathcal{R}_\varepsilon) \geq 1 - \delta_{\mathrm{opt}}.$$

Assumption 1 does not require $\mathcal{A}$ to know $\mathcal{X}_\varepsilon$; it only bounds the chance that the candidate set misses this region. The condition must be justified for the chosen optimizer. Given this condition, our analysis asks whether the surrogate selects a high-quality design after search has reached the target region.

Conditional on $\mathcal{R}_\varepsilon$, surrogate-based selection succeeds whenever every reachable ε-optimal candidate is ranked above every reachable non-ε-optimal candidate. The following proposition states this requirement as a condition on the surrogate scores.

Proposition 2 (Reachability and cross-region ranking)**.** *On $\mathcal{R}_\varepsilon$, suppose $\widehat{h}(\boldsymbol{x}^+) > \widehat{h}(\boldsymbol{x}^-)$ for every $\boldsymbol{x}^+ \in \mathcal{C} \cap \mathcal{X}_\varepsilon$ and $\boldsymbol{x}^- \in \mathcal{C} \cap \mathcal{X}_\varepsilon^c$. Then $\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon$, and hence $f(\boldsymbol{x}^\star) - f(\widehat{\boldsymbol{x}}_{\mathrm{MBO}}) \leq \varepsilon$.*

Proposition 2 identifies cross-region ordering as sufficient for reliable selection; it does not require a correct ordering within either region. To distinguish this selection requirement from candidate generation, we combine Assumption 1 with a decomposition of the failure event:

$$\Pr(f(\boldsymbol{x}^\star) - f(\widehat{\boldsymbol{x}}_{\mathrm{MBO}}) > \varepsilon) \leq \delta_{\mathrm{opt}} + \Pr(\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon^c \mid \mathcal{R}_\varepsilon)\,. \tag{3}$$

The first term in (3) bounds failure to generate an ε-optimal candidate, while the conditional probability measures failure to select such a candidate after one has been reached. Thus, generating a good candidate and selecting it are assessed separately. The following analysis conditions on $\mathcal{R}_\varepsilon$ and studies the remaining selection error through cross-region ranking risk.

3 Ranking for Offline Optimization

Conditional on reaching an ε-optimal candidate, the remaining question is whether the surrogate selects it. We connect ranking to this decision, contrast it with regression, and identify how training pairs and data coverage affect the selection guarantee.

3.1 Ranking Error and Candidate Selection

To quantify selection errors, fix a sampling rule over generated candidates. Conditional on $\mathcal{R}_\varepsilon$, let X^+ and X^- be independent draws from this rule restricted to $\mathcal{X}_\varepsilon$ and $\mathcal{X}_\varepsilon^c$, with laws ρ_ε^+ and ρ_ε^-, respectively. Assume these laws exist whenever both regions are represented; without non-ε-optimal candidates, selection succeeds. For $h \in \mathcal{H}$, define the target ranking risk:

$$\begin{aligned} Q_\varepsilon &:= \rho_\varepsilon^+ \otimes \rho_\varepsilon^-, \\ \ell_{01}(h; \boldsymbol{x}, \boldsymbol{x}') &:= \mathbf{1}\{h(\boldsymbol{x}) \leq h(\boldsymbol{x}')\}, \\ \mathcal{E}_\varepsilon(h) &:= \mathbb{E}_{(X^+, X^-) \sim Q_\varepsilon}[\ell_{01}(h; X^+, X^-)]. \end{aligned} \tag{4}$$

The risk in (4) counts comparisons in which a non-ε-optimal candidate scores at least as high as an ε-optimal one. Suppose that, conditional on $\mathcal{R}_\varepsilon$, the candidate pool contains one reference $X^+ \sim \rho_\varepsilon^+$ and at most K non-ε-optimal candidates, each paired with X^+ according to Q_ε. Selection failure requires at least one such comparison error. A union bound and (3) therefore give the following selection bounds:

$$\begin{aligned} \Pr(\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon^c \mid \mathcal{R}_\varepsilon) &\leq K\mathcal{E}_\varepsilon(\widehat{h}), \\ \Pr(f(\boldsymbol{x}^\star) - f(\widehat{\boldsymbol{x}}_{\mathrm{MBO}}) > \varepsilon) &\leq \delta_{\mathrm{opt}} + K\mathcal{E}_\varepsilon(\widehat{h}). \end{aligned} \tag{5}$$

Here K depends on the terminal candidate pool. Multiple ε-optimal candidates are allowed if the chosen reference satisfies the same pair-distribution condition; independence among negative candidates is unnecessary. Lemma 7 in Appendix C gives the formal argument.

Training must learn these comparisons from offline data. Let Q_{tr} be the training-pair distribution, with the preferred design first, and draw $Z_{\mathrm{tr}} = \{(\boldsymbol{x}_k, \boldsymbol{x}_k')\}_{k=1}^N \sim Q_{\mathrm{tr}}^N$ independently conditional

on the fixed offline dataset. For a pairwise surrogate loss satisfying $\ell_{01} \le \widetilde{\ell} \le B$ and its class $\mathcal{G} := \{(\boldsymbol{x}, \boldsymbol{x}') \mapsto \widetilde{\ell}(h; \boldsymbol{x}, \boldsymbol{x}') : h \in \mathcal{H}\}$, define the following learning quantities:

$$\begin{aligned} \widehat{\mathcal{E}}_{\mathrm{tr}}(h) &:= \frac{1}{N} \sum_{k=1}^{N} \widetilde{\ell}(h; \boldsymbol{x}_k, \boldsymbol{x}'_k), \\ \Gamma_N(\mathcal{G}, \delta) &:= 2\widehat{\mathfrak{R}}_{Z_{\mathrm{tr}}}(\mathcal{G}) + 3B\sqrt{\frac{\log(2/\delta)}{2N}}, \\ \Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}}) &:= \sup_{h \in \mathcal{H}} \left| \mathbb{E}_{Q_\varepsilon} \widetilde{\ell}(h) - \mathbb{E}_{Q_{\mathrm{tr}}} \widetilde{\ell}(h) \right|. \end{aligned} \tag{6}$$

The terms in (6) measure training loss, finite-sample complexity, and training–target pair mismatch. Here $\widehat{\mathfrak{R}}$ is empirical Rademacher complexity, defined in Appendix B. The following result connects them to candidate selection.

Theorem 3 (Conditional ranking learnability)**.** *With probability at least $1 - \delta$ over $Z_{\mathrm{tr}} \sim Q_{\mathrm{tr}}^N$, the following bounds hold simultaneously for every $h \in \mathcal{H}$:*

$$\begin{aligned} \mathcal{E}_\varepsilon(h) &\le \widehat{\mathcal{E}}_{\mathrm{tr}}(h) + \Gamma_N(\mathcal{G}, \delta) + \Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}}), \\ \Pr(\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon^c) &\le \delta_{\mathrm{opt}} + K\Big[\widehat{\mathcal{E}}_{\mathrm{tr}}(\widehat{h}) + \Gamma_N(\mathcal{G}, \delta) + \Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}})\Big]. \end{aligned} \tag{7}$$

The second statement assumes the conditional candidate-pool model used in (5)*.*

Because Theorem 3 holds uniformly, it applies to the surrogate learned from the same training pairs. The key implication of (7) is that fitting more pairs does not by itself resolve a mismatch in which designs are compared. Even with a fixed model class and loss, the pair sampler changes both training emphasis and Δ_ℓ. The bound thus motivates learning comparisons representative of those encountered during search.

3.2 Ranking versus Regression

Regression controls the same selection risk through prediction errors and objective gaps. Let $e_h(\boldsymbol{x}) := h(\boldsymbol{x}) - f(\boldsymbol{x})$ and, for $1 \le p < \infty$, define $R_+^{(p)}(h) := \mathbb{E}_{\rho_\varepsilon^+}|e_h(X)|^p$ and $R_-^{(p)}(h) := \mathbb{E}_{\rho_\varepsilon^-}|e_h(X)|^p$. For an analysis margin $\gamma > 0$, let $M_\varepsilon(\gamma) := \Pr_{Q_\varepsilon}(0 < f(X^+) - f(X^-) < \gamma)$ measure the mass of difficult small-gap pairs.

Proposition 4 (Regression-to-ranking reduction)**.** *For every $h \in \mathcal{H}$, $1 \le p < \infty$, and $\gamma > 0$,*

$$\mathcal{E}_\varepsilon(h) \le M_\varepsilon(\gamma) + \frac{2^{p-1}}{\gamma^p}\left(R_+^{(p)}(h) + R_-^{(p)}(h)\right). \tag{8}$$

Equation (8) exposes two obstacles for regression: prediction errors in both candidate regions and pairs with small objective gaps. Decreasing γ reduces the unresolved small-gap mass but amplifies regional errors by γ^{-p}; no uniform gap is assumed. Ranking instead supervises the comparison event in (4), focusing on distinctions that affect selection. Appendix D provides the proof and further regression comparisons.

3.3 Training Pairs and Geometric Coverage

Which training comparisons are useful depends on their proximity to those encountered during search. Equip ordered pairs with $d_{\mathrm{pair}}((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')) := \|\boldsymbol{x} - \boldsymbol{z}\|_2 + \|\boldsymbol{x}' - \boldsymbol{z}'\|_2$. Assume $\widetilde{\ell}(h; \cdot, \cdot)$ is uniformly L-Lipschitz under this metric over $h \in \mathcal{H}$, and the relevant distributions have finite first moments. Lemma 9 in Appendix C gives $\Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}}) \le LW_1(Q_\varepsilon, Q_{\mathrm{tr}})$, yielding a geometric interpretation of pair mismatch.

Corollary 5 (Distribution-aware ranking bound)**.** *Under the conditions of Theorem 3 and the Lipschitz condition above, with probability at least $1 - \delta$, every $h \in \mathcal{H}$ satisfies*

$$\begin{aligned} &\text{(a)} \quad \mathcal{E}_\varepsilon(h) \le \widehat{\mathcal{E}}_{\mathrm{tr}}(h) + \Gamma_N(\mathcal{G}, \delta) + LW_1(Q_\varepsilon, Q_{\mathrm{tr}}); \\ &\text{(b)} \quad \mathcal{E}_\varepsilon(h) \le \widehat{\mathcal{E}}_{\mathrm{tr}}(h) + \Gamma_N(\mathcal{G}, \delta) + L\left[W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) + W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-)\right], \end{aligned} \tag{9}$$

where part (b) *additionally assumes $Q_{\mathrm{tr}} = \mu_{\mathrm{tr}}^+ \otimes \mu_{\mathrm{tr}}^-$.*

Part (a) of (9) applies to joint pair samplers, including DAR's mixture; part (b) separates the two coordinates under independent sampling. Global pairing can devote most comparisons to ordering within quality regions, whereas final selection depends on comparisons across them. This motivates emphasizing high-quality-versus-remaining pairs when observed high-quality designs approximate the reachable ε-optimal candidates. Their value lies in representing the comparisons relevant to final selection.

Useful comparisons also depend on where the data lie. To separate extrapolation from mismatch within observed support, let $\mathcal{M} := \operatorname{supp}(\mu_{\mathrm{tr}}^{+})$ be nonempty and closed, and let $\Pi_{\mathcal{M}}$ be a measurable nearest-point projection. The next result resolves the positive-coordinate term in part (b) of (9) into off-support displacement D_{out} and residual within-support mismatch D_{in}.

Corollary 6 (Geometric coverage bound)**.** *Under the conditions of Corollary 5*(b)*, the quantities defined in the first line below yield, with probability at least $1-\delta$, the following simultaneous bound for every $h \in \mathcal{H}$:*

$$\begin{aligned} D_{\mathrm{out}} &:= \mathbb{E}_{X\sim\rho_{\varepsilon}^{+}} \operatorname{dist}(X, \mathcal{M}), \qquad D_{\mathrm{in}} := W_1((\Pi_{\mathcal{M}})_{\#}\rho_{\varepsilon}^{+}, \mu_{\mathrm{tr}}^{+}), \\ \mathcal{E}_{\varepsilon}(h) &\le \widehat{\mathcal{E}}_{\mathrm{tr}}(h) + \Gamma_N(\mathcal{G}, \delta) + L\big[D_{\mathrm{out}} + D_{\mathrm{in}} + W_1(\rho_{\varepsilon}^{-}, \mu_{\mathrm{tr}}^{-})\big]. \end{aligned} \tag{10}$$

In (10), D_{out} measures how far reachable near-optimal candidates lie beyond positive training support, while D_{in} measures the mismatch remaining after projection onto that support. Together, they distinguish extrapolation distance from how training comparisons cover observed designs. This motivates DAR's use of high-quality observations as a proxy for target candidates, and the Branin diagnostic of ranking farther from the observed data.

4 Distribution-Aware Ranking (DAR)

DAR turns the comparison target in (9) into a training rule based on observed designs. Since the true ε-optimal region is unknown, high-quality observations serve as a proxy. The central choice is to devote most training comparisons to distinguishing these designs from the remaining data.

Empirical target candidates. Given $S = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$, let $\tau \in (0,1)$ specify the empirical top fraction and define the label threshold

$$q_{1-\tau} := \mathrm{Quantile}_{1-\tau}(\{y_i\}_{i=1}^{m}).$$

This threshold partitions the data into high-quality and remaining subsets:

$$S_{\tau}^{+} := \{\boldsymbol{x}_i : y_i \ge q_{1-\tau}\}, \qquad S_{\tau}^{-} := S \setminus S_{\tau}^{+}.$$

Unlike the objective-gap tolerance ε, τ specifies an empirical quantile. The two subsets approximate the regions in (2) using observed labels; they need not coincide with the true regions, and label noise can affect membership near the quantile threshold.

Training comparisons and loss. With probability $1-\lambda$, DAR samples $\boldsymbol{x}_1 \in S_{\tau}^{+}$ and $\boldsymbol{x}_2 \in S_{\tau}^{-}$. With probability $\lambda \in [0,1]$, it samples both designs from S_{τ}^{+} and orders them by their observed labels. The first component emphasizes candidate selection; the second encourages ordering within the high-quality subset.

For a pair with $\boldsymbol{x}_1$ preferred over $\boldsymbol{x}_2$, the surrogate h_θ is trained with margin β using the following ranking loss:

$$\mathcal{L}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \max\big(0,\, \beta - \big(h_\theta(\boldsymbol{x}_1) - h_\theta(\boldsymbol{x}_2)\big)\big). \tag{11}$$

For $\beta > 0$, the calibrated surrogate $\widetilde{\ell} := \mathcal{L}/\beta$ charges at least one for every inversion, as required by Theorem 3. This rescaling preserves the empirical minimizers of (11); bounded score ranges or loss clipping provide the upper bound needed for the concentration argument.

What the pair sampler changes. Global ranking (Tan et al., 2025) spends many comparisons within the two empirical regions. To quantify the difference, suppose all m labels are distinct and $|S_{\tau}^{+}| = \tau m$. Only $\tau m(1-\tau)m$ of the $\binom{m}{2}$ unordered pairs cross between these regions, giving the following pair proportion:

$$r_{\mathrm{cross}} = \frac{\tau m(1-\tau)m}{\binom{m}{2}} = \frac{2\tau(1-\tau)m}{m-1} \approx 2\tau(1-\tau). \tag{12}$$

For the illustrative choice $\tau = 0.05$, (12) gives about 9.5% cross-region pairs. DAR with $\lambda = 0.1$ allocates 90% to them, about 9.5 times the global-pairing proportion. Thus, the sampler redirects finite training effort toward the comparisons used as a proxy for candidate selection.

To relate this allocation to the theory, let $\widehat{\rho}_\tau^+$ and $\widehat{\rho}_\tau^-$ be the empirical distributions on S_τ^+ and S_τ^-, respectively. Their product defines the empirical cross-region distribution:

$$\widehat{Q}_\tau := \widehat{\rho}_\tau^+ \otimes \widehat{\rho}_\tau^-.$$

Let Q_{intra} denote label-ordered pairs sampled within S_τ^+. The two sampling components induce the following training distribution:

$$Q_{\mathrm{DAR}} := (1-\lambda)\widehat{Q}_\tau + \lambda Q_{\mathrm{intra}}. \tag{13}$$

In (13), Q_{DAR} instantiates the theory's Q_{tr}. This mixture need not factorize, so part (a) of (9) applies directly. If Q_ε, $\widehat{Q}_\tau$, and Q_{intra} have finite first moments under d_{pair}, Proposition 16 in Appendix E gives the following alignment bound:

$$W_1(Q_\varepsilon, Q_{\mathrm{DAR}}) \le W_1(Q_\varepsilon, \widehat{Q}_\tau) + \lambda W_1(\widehat{Q}_\tau, Q_{\mathrm{intra}}). \tag{14}$$

The first term in (14) measures how well the empirical cross-region pairs represent target comparisons. The second bounds the additional mismatch from within-region pairing and scales with λ. Setting $\lambda = 0$ gives $Q_{\mathrm{DAR}} = \widehat{Q}_\tau$; a small positive λ retains local ordering supervision while limiting its contribution to the bound. DAR's alignment therefore depends on the representativeness of its observed proxy.

Design optimization. Correct ordering does not fix the scale of gradient-based search. Replacing h_θ by $ah_\theta + b$ with $a > 0$ preserves ranking but changes gradient magnitudes and the fixed-margin loss in (11). Independently trained ranking surrogates (Tan et al., 2025) can therefore require different search-step scales.

Following Tan et al. (2025), we compute the prediction mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma} > 0$ on the training set and optimize $\tilde{h}_\theta(\boldsymbol{x}) = (h_\theta(\boldsymbol{x}) - \tilde{\mu})/\tilde{\sigma}$. This transformation preserves ordering and reduces arbitrary scale variation across models, facilitating comparable search hyperparameters; it does not guarantee gradients matching those of a regression surrogate. Starting from S_τ^+, DAR applies gradient ascent for the prescribed budget. Algorithm 1 in Appendix F gives the complete procedure.

5 Experiments

We examine whether learning to rank promising designs improves offline optimization, and which choices in DAR make that supervision useful. Branin exposes how the surrogate orders designs beyond the observed data; Design-Bench (Trabucco et al., 2022) tests the quality of the returned candidates across continuous and discrete tasks. We then use hyperparameter ablations to explain the balance between concentrating comparisons and retaining sufficient training support.

5.1 Ranking Behavior on Branin

Setup. Following the shifted-data protocol (Trabucco et al., 2022), we retain the worst 60% of collected Branin designs as the offline dataset. This leaves little training support around the three global optima and makes the location of high surrogate scores consequential for search. We compare MSE as a pointwise baseline, RaM (Tan et al., 2025) using RankCosine (Qin et al., 2008), and DAR using the margin loss in (11).

Where does the surrogate direct search? Figure 1 compares the learned landscapes with the true objective. MSE produces a smooth surface that suppresses distinct high-quality regions, especially near the upper-left optimum. RaM recovers a clearer multimodal ordering, while DAR places high-score regions around all three optima. Because the surrogates have different output scales, the informative feature is where each model assigns high scores. This comparison suggests that ranking supervision preserves structure useful for finding promising designs; the next diagnostic measures the reliability of that ordering.

How does ordering change farther from the data? We set $\tau = 5\%$, construct S_τ^+, and expand the comparison set to $S_{\tau,d}^- := \{\boldsymbol{x} \notin S_\tau^+ : \mathrm{dist}(\boldsymbol{x}, \mathcal{M}) \le d\}$. The inversion frequency between these sets

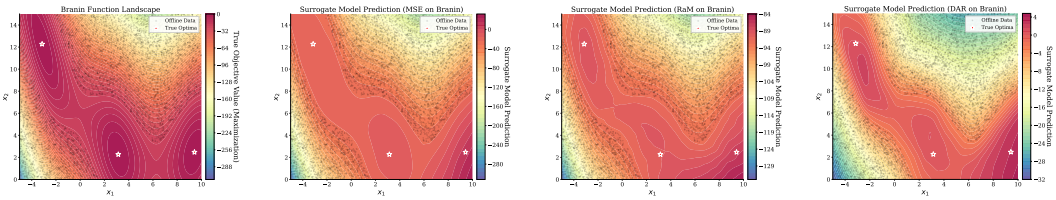


Figure 1: Branin landscapes under shifted offline data. **(a)** The true objective and observed designs (black dots), restricted to the worst 60%. **(b)–(d)** Surrogates trained by MSE, RaM, and DAR. Compare the locations of high-score regions within each panel; the surrogates have different output scales.

is an empirical proxy for the cross-region risk in (4). Increasing d includes more distant alternatives, while the high-quality reference set stays fixed. Here τ specifies an empirical fraction, distinct from the objective-gap tolerance ε in the theoretical selection target.

Figure 2 shows that DAR has the lowest inversion error at every evaluated radius. All three errors rise sharply outside the most local region, then fluctuate at elevated levels. This pattern is consistent with the geometric mismatch in (10): comparisons become harder when their locations depart from observed support. The diagnostic illustrates this difficulty without directly estimating the terms of the bound.

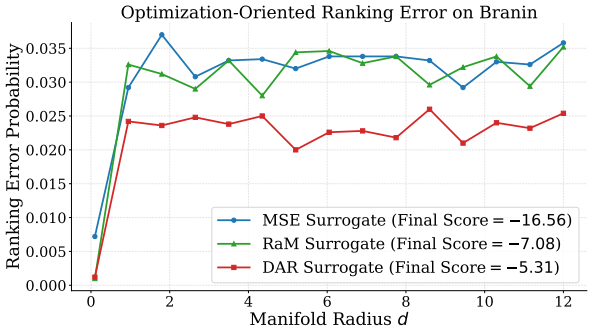


Figure 2: Branin ranking error as the comparison region expands with radius d. The legend reports final objective values.

The final Branin scores follow the same method ordering: DAR reaches -5.31, RaM -7.08, and MSE -16.56. Lower inversion error therefore accompanies a better final design in this study. The landscapes show where promising candidates may be found, and the radius diagnostic shows whether they remain well ordered against broader alternatives. Since RaM and DAR also differ in their loss, this comparison evaluates the complete methods; the sampling ablation below examines the role of pair allocation within DAR.

5.2 Final Design Quality on Design-Bench

Tasks and comparison methods. We next test whether the behavior observed on Branin translates to useful designs on five Design-Bench tasks (Trabucco et al., 2022). Ant Morphology (Brockman et al., 2016) and D'Kitty Morphology (Ahn et al., 2020) optimize robot structures in 60 and 56 dimensions for locomotion speed. Superconductor (Hamidieh, 2018) searches an 86-dimensional material space for high critical temperature. TF-Bind-8 and TF-Bind-10 (Trabucco et al., 2022; Barrera et al., 2016) optimize DNA sequences of lengths 8 and 10 for transcription-factor binding affinity. These tasks test the same ranking-based approach across morphology, materials, and biological sequence design.

The baselines cover two routes from offline observations to proposed designs. PGS (Chemingui et al., 2024), FGM (Grudzien et al., 2024), Match-OPT (Hoang et al., 2024), and RaM (Tan et al., 2025) learn a forward surrogate and optimize its predictions. GTG (Yun et al., 2024) and ROOT (Dao et al., 2025) instead model the design distribution and generate candidates. RaM provides the closest comparison to DAR because both learn rankings. The other methods establish whether this approach remains competitive with a broader set of offline optimizers.

Implementation and evaluation. Following Trabucco et al. (2021); Tan et al. (2025), the surrogate is an MLP with two layers of 2048 units and ReLU activations. We use the baseline settings from their original papers. DAR uses the same $\tau = 0.2$, $\lambda = 0.1$, and $\beta = 0.4$ across tasks, allocating

90% of its training pairs to comparisons across the empirical quality threshold. Appendix F gives the complete algorithm, and Table 2 reports the full hyperparameter ablations.

For each run, we evaluate the 128 proposed solutions and report their 100th percentile ground-truth score, following Design-Bench (Trabucco et al., 2022). Scores are normalized as $(y - y_{\min})/(y_{\max} - y_{\min})$, using the extremes of the full unseen dataset. Table 1 reports means and standard deviations over ten independent seeds; comparison results are inherited or reproduced as stated in its caption. The metric measures the best design in the returned batch, so a higher value means that search reaches at least one better candidate. It does not measure the average quality of all proposed designs.

Table 1: 100th percentile normalized scores on Design-Bench. The best and runner-up mean scores on each task are shown in **bold** and underlined, respectively; lower average rank is better. S(best) denotes the best score in the offline dataset. Results for comparison methods are inherited from Tan et al. (2025) or reproduced by us.

Method	Ant	D'Kitty	Superconductor	TF-Bind-8	TF-Bind-10	Avg. Rank
S(best)	0.565	0.884	0.400	0.439	0.467	/
PGS (Chemingui et al., 2024)	0.715 ± 0.046	0.954 ± 0.022	0.444 ± 0.020	0.889 ± 0.061	0.634 ± 0.040	5.4
FGM (Grudzien et al., 2024)	0.923 ± 0.023	0.944 ± 0.014	0.481 ± 0.024	0.811 ± 0.079	0.611 ± 0.008	5.8
Match-OPT (Hoang et al., 2024)	0.933 ± 0.016	0.952 ± 0.008	0.504 ± 0.021	0.824 ± 0.067	0.655 ± 0.050	4.0
GTG (Yun et al., 2024)	0.855 ± 0.044	0.942 ± 0.017	0.480 ± 0.055	0.910 ± 0.040	0.619 ± 0.029	5.6
RaM (Tan et al., 2025)	0.940 ± 0.028	0.951 ± 0.017	**0.514 ± 0.026**	<u>0.982 ± 0.012</u>	<u>0.675 ± 0.049</u>	<u>2.6</u>
ROOT (Dao et al., 2025)	**0.958 ± 0.012**	**0.967 ± 0.001**	0.451 ± 0.032	0.977 ± 0.015	0.652 ± 0.020	3.0
DAR (Ours)	<u>0.941 ± 0.013</u>	<u>0.959 ± 0.006</u>	<u>0.506 ± 0.008</u>	**0.988 ± 0.008**	**0.686 ± 0.056**	**1.6**

Consistent performance across task families. DAR achieves the best average rank, 1.6, compared with 2.6 for RaM and 3.0 for ROOT in Table 1. It is the only method ranked first or second on all five tasks. DAR has the highest mean on both sequence tasks and ranks second on the three continuous tasks, where ROOT leads on Ant and D'Kitty and RaM leads on Superconductor. Its advantage is therefore broad competitiveness across different design spaces. DAR also exceeds S(best) on every task, showing improvement beyond the best observed offline design.

Relative to RaM, DAR improves the mean on four of five tasks. Its standard deviations on Ant, D'Kitty, and Superconductor are 0.013, 0.006, and 0.008, compared with 0.028, 0.017, and 0.026 for RaM. This lower variability accompanies the stronger average rank, although individual gains can be small: the Ant means differ by only 0.001. On TF-Bind-10, DAR has a higher mean and a slightly larger standard deviation. These results support consistent performance across tasks without implying a large or less variable gain on every benchmark.

5.3 Which Training Comparisons Matter?

To examine DAR's design choices, we vary one hyperparameter at a time, fixing the others at $\lambda = 0.1$, $\tau = 0.2$, and $\beta = 0.4$. Table 2 reports the complete taskwise scores and average ranks for all tested settings. Ranks compare settings within each ablation, rather than methods in the main benchmark.

Retain some ordering within the top set. In Table 2a, adding a small intra-region component improves the average rank from 3.4 at $\lambda = 0$ to 1.6 at $\lambda = 0.1$, while further increases do not consistently help. At the default $\tau = 0.2$, (12) gives approximately 32% cross-region pairs under global pairing; DAR assigns 90%, about 2.8 times as many. Most comparisons therefore distinguish promising designs from the remaining data, while a small fraction discriminates among promising designs themselves.

This result also clarifies (14). Its additional mismatch term grows with λ, yet the best observed setting is not $\lambda = 0$. The bound limits one source of distribution mismatch, without capturing every effect of pair allocation on learning and search. The ablation favors a small positive mixture weight empirically. Reallocating pairs uses the same labels and supplies no observations near unseen optima, so its value still depends on the quality and coverage of the offline data.

Balance selectivity with the size of the top set. Table 2b gives the best average rank, 1.4, at $\tau = 0.2$ and the weakest, 4.6, at $\tau = 0.05$; nearby settings 0.15 and 0.25 remain competitive. The smallest top set particularly affects Superconductor, whose mean is 0.405 rather than 0.506 at $\tau = 0.2$ (Table 2b). A smaller top set is more selective but provides fewer positive examples. Broadening it increases

Table 2: DAR ablations on five Design-Bench tasks. Scores are reported as mean ± standard deviation; bold and underlining mark the best and runner-up means within each ablation. Lower average rank is better.

(a) Intra-region ratio λ; fixed $\tau = 0.2$ and $\beta = 0.4$.

λ	Ant	D'Kitty	Superconductor	TF-Bind-8	TF-Bind-10	Avg. Rank
0.0	0.900 ± 0.032	$\underline{0.950 \pm 0.012}$	0.496 ± 0.009	$\mathbf{0.989 \pm 0.003}$	0.670 ± 0.069	$\underline{3.4}$
0.1	$\mathbf{0.941 \pm 0.013}$	$\mathbf{0.959 \pm 0.006}$	$\underline{0.506 \pm 0.008}$	$\underline{0.988 \pm 0.008}$	$\underline{0.686 \pm 0.056}$	**1.6**
0.2	0.914 ± 0.032	0.949 ± 0.007	$\mathbf{0.507 \pm 0.009}$	0.979 ± 0.019	0.674 ± 0.058	3.4
0.3	0.923 ± 0.024	0.946 ± 0.006	0.496 ± 0.029	0.984 ± 0.008	0.647 ± 0.036	3.6
0.4	0.919 ± 0.013	0.943 ± 0.005	0.476 ± 0.047	0.978 ± 0.014	0.642 ± 0.035	5.4
0.5	$\underline{0.937 \pm 0.018}$	0.943 ± 0.005	0.473 ± 0.050	0.981 ± 0.008	$\mathbf{0.699 \pm 0.096}$	3.6

(b) Top-sample fraction τ; fixed $\lambda = 0.1$ and $\beta = 0.4$.

τ	Ant	D'Kitty	Superconductor	TF-Bind-8	TF-Bind-10	Avg. Rank
0.05	0.924 ± 0.015	0.944 ± 0.006	0.405 ± 0.037	0.941 ± 0.018	0.654 ± 0.029	4.6
0.1	0.930 ± 0.012	0.942 ± 0.008	0.498 ± 0.012	0.961 ± 0.033	0.674 ± 0.054	4.0
0.15	$\underline{0.931 \pm 0.021}$	0.946 ± 0.007	$\underline{0.505 \pm 0.011}$	0.973 ± 0.021	0.674 ± 0.055	2.6
0.2	$\mathbf{0.941 \pm 0.013}$	$\underline{0.959 \pm 0.006}$	$\mathbf{0.506 \pm 0.008}$	$\mathbf{0.988 \pm 0.008}$	$\underline{0.686 \pm 0.056}$	**1.4**
0.25	0.912 ± 0.018	$\mathbf{0.959 \pm 0.008}$	0.503 ± 0.009	$\underline{0.978 \pm 0.013}$	$\mathbf{0.689 \pm 0.155}$	$\underline{2.4}$

(c) Ranking margin β; fixed $\lambda = 0.1$ and $\tau = 0.2$.

β	Ant	D'Kitty	Superconductor	TF-Bind-8	TF-Bind-10	Avg. Rank
0.0	$\underline{0.936 \pm 0.025}$	$\mathbf{0.960 \pm 0.011}$	$\underline{0.505 \pm 0.008}$	0.982 ± 0.015	$\underline{0.672 \pm 0.060}$	$\underline{2.4}$
0.1	0.919 ± 0.036	0.951 ± 0.008	0.498 ± 0.011	0.975 ± 0.017	0.640 ± 0.045	5.6
0.2	0.920 ± 0.022	0.953 ± 0.010	0.504 ± 0.005	0.983 ± 0.010	0.641 ± 0.029	3.8
0.3	0.927 ± 0.022	0.952 ± 0.008	0.503 ± 0.007	$\underline{0.987 \pm 0.004}$	0.659 ± 0.025	3.6
0.4	$\mathbf{0.941 \pm 0.013}$	$\underline{0.959 \pm 0.006}$	$\mathbf{0.506 \pm 0.008}$	$\mathbf{0.988 \pm 0.008}$	$\mathbf{0.686 \pm 0.056}$	**1.2**
0.5	0.931 ± 0.012	0.948 ± 0.005	0.503 ± 0.006	0.983 ± 0.016	0.634 ± 0.028	4.4

training support while lowering the quality threshold. The observed pattern favors an intermediate fraction for these tasks.

Separate preferred designs with a positive margin. In Table 2c, $\beta = 0.4$ attains average rank 1.2, compared with 2.4 for $\beta = 0$. It has the highest mean on four tasks, while zero margin is slightly better on D'Kitty (Table 2c). Thus, score separation in (11) improves aggregate performance, although larger margins are not uniformly better. Together, the ablations favor concentrating supervision on a sufficiently large high-quality subset, retaining some internal ordering, and encouraging separation between preferred and lower-quality designs.

6 Conclusion

We frame offline MBO around learning the comparisons that govern final design selection. Conditional on reaching a near-optimal candidate, our analysis connects selection error to ranking error and identifies pair-distribution mismatch and data coverage as key challenges. DAR turns this perspective into a training rule that emphasizes high-quality-versus-remaining comparisons. Branin diagnostics and five Design-Bench tasks support its effectiveness, with DAR achieving the best average rank among the evaluated methods. These findings highlight training-pair construction as a practical choice for improving offline optimization. An open direction is to adapt training comparisons to the changing search distribution while accounting for uncertainty beyond the observed data support.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript editing, technical consistency checking, and experimental code development. All AI-assisted content and code were reviewed and verified by the authors. The authors take full responsibility for all claims, results, and materials presented in this work.

ETHICS STATEMENT

This work develops a general methodology for offline model-based optimization and evaluates it on established benchmark tasks. It does not involve human participants, personal data collection, or new physical or biological experiments. Some benchmark tasks concern biological sequence design, materials, and robotic morphology; our experiments are limited to benchmark evaluation and should not be interpreted as validating or recommending generated designs for real-world deployment. Applications in safety-critical scientific or engineering settings would require appropriate domain-specific validation and safeguards.

REPRODUCIBILITY STATEMENT

The problem setup and optimization target are defined in Section 2, the ranking learnability analysis and its comparison with regression are developed in Section 3, and the proposed distribution-aware ranking (DAR) method is described in Section 4. Section 5 specifies the benchmark tasks, surrogate architecture, evaluation protocol, shared DAR hyperparameters, repeated-seed reporting, and hyperparameter ablations. Appendix B states the notation and technical conditions, Appendices C–E provide proofs and supporting theoretical results, and Appendix F gives the complete algorithmic description of DAR.

REFERENCES

Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and J. Zico Kolter. Differentiable convex optimization layers. In *Advances in Neural Information Processing Systems 32*, pp. 9562–9574. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/9ce3c52fc54362e22053399d3181c638-Abstract.html.

Michael Ahn, Henry Zhu, Kristian Hartikainen, Hugo Ponte, Abhishek Gupta, Sergey Levine, and Vikash Kumar. ROBEL: Robotics benchmarks for learning with low-cost robots. In *Proceedings of the Conference on Robot Learning*, volume 100 of *Proceedings of Machine Learning Research*, pp. 1300–1313. PMLR, 2020. URL https://proceedings.mlr.press/v100/ahn20a.html.

Brandon Amos and J. Zico Kolter. OptNet: Differentiable optimization as a layer in neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 136–145. PMLR, 2017. URL https://proceedings.mlr.press/v70/amos17a.html.

Luis A. Barrera, Anastasia Vedenko, Jesse V. Kurland, Julia M. Rogers, Stephen S. Gisselbrecht, Elizabeth J. Rossin, Jaie Woodard, Luca Mariani, Kian Hong Kock, Sachi Inukai, Trevor Siggers, Leila Shokri, Raluca Gordân, Nidhi Sahni, Chris Cotsapas, Tong Hao, Song Yi, Manolis Kellis, Mark J. Daly, Marc Vidal, David E. Hill, and Martha L. Bulyk. Survey of variation in human transcription factors reveals prevalent DNA binding changes. *Science*, 351(6280):1450–1454, 2016. doi: 10.1126/science.aad2257. URL https://pubmed.ncbi.nlm.nih.gov/27013732/.

Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 950–959. PMLR, 2020. URL https://proceedings.mlr.press/v119/blondel20a.html.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. *arXiv preprint arXiv:1606.01540*, 2016. URL https://arxiv.org/abs/1606.01540.

Chris J. C. Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, pp. 89–96. ACM, 2005. doi: 10.1145/1102351.1102363. URL https://doi.org/10.1145/1102351.1102363.

Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, pp. 129–136. ACM, 2007. doi: 10.1145/1273496.1273513. URL https://doi.org/10.1145/1273496.1273513.

Yassine Chemingui, Aryan Deshwal, Trong Nghia Hoang, and Janardhan Rao Doppa. Offline model-based optimization via policy-guided gradient search. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(10):11230–11239, 2024. doi: 10.1609/aaai.v38i10.29001. URL https://ojs.aaai.org/index.php/AAAI/article/view/29001.

Can Chen, Christopher Beckham, Zixuan Liu, Xue Liu, and Christopher Pal. Parallel-mentoring for offline model-based optimization. In *Advances in Neural Information Processing Systems 36*, pp. 76619–76636, 2023a. doi: 10.52202/075280-3347. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/f189e7580acad0fc7fd45405817ddee3-Abstract-Conference.html.

Can Chen, Yingxue Zhang, Xue Liu, and Mark Coates. Bidirectional learning for offline model-based biological sequence design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 5351–5366. PMLR, 2023b. URL https://proceedings.mlr.press/v202/chen23ao.html.

Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems 30*, pp. 3730–3739. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/0070d23b06b1486a538c0eaa45dd167a-Abstract.html.

Manh Cuong Dao, Phi Le Nguyen, Thao Nguyen Truong, and Trong Nghia Hoang. Boosting offline optimizers with surrogate sensitivity. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 10072–10090. PMLR, 2024. URL https://proceedings.mlr.press/v235/dao24b.html.

Manh Cuong Dao, The Hung Tran, Phi Le Nguyen, Thao Nguyen Truong, and Trong Nghia Hoang. ROOT: Rethinking offline optimization as distributional translation via probabilistic bridge. In *Advances in Neural Information Processing Systems 38*, pp. 98783–98817, 2025. URL https://papers.nips.cc/paper_files/paper/2025/hash/8e357df853b7352fe91d92c83e2f90b2-Abstract-Conference.html.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems 30*, pp. 5484–5494. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3fc2c60b5782f641f76bcefc39fb2392-Abstract.html.

Othman El Balghiti, Adam N. Elmachtoub, Paul Grigas, and Ambuj Tewari. Generalization bounds in the predict-then-optimize framework. *Mathematics of Operations Research*, 48(4):2043–2065, 2023. doi: 10.1287/moor.2022.1330. URL https://pubsonline.informs.org/doi/10.1287/moor.2022.1330.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022. doi: 10.1287/mnsc.2020.3922. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3922.

Justin Fu and Sergey Levine. Offline model-based optimization via normalized maximum likelihood estimation. In *Proceedings of the 9th International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=FmMKSO4e8JK.

Javier González, Zhenwen Dai, Andreas Damianou, and Neil D. Lawrence. Preferential Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1282–1291. PMLR, 2017. URL https://proceedings.mlr.press/v70/gonzalez17a.html.

Kuba Grudzien, Masatoshi Uehara, Sergey Levine, and Pieter Abbeel. Functional graphical models: Structure enables offline data-driven optimization. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2449–2457. PMLR, 2024. URL https://proceedings.mlr.press/v238/grudzien24a.html.

Kam Hamidieh. A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science*, 154:346–354, 2018. doi: 10.1016/j.commatsci.2018.07.052. URL https://www.sciencedirect.com/science/article/pii/S0927025618304877.

Minh Hoang, Azza Fadhel, Aryan Deshwal, Jana Doppa, and Trong Nghia Hoang. Learning surrogates for offline black-box optimization via gradient matching. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 18374–18393. PMLR, 2024. URL https://proceedings.mlr.press/v235/hoang24a.html.

Minsu Kim, Jiayao Gu, Ye Yuan, Taeyoung Yun, Zixuan Liu, Yoshua Bengio, and Can Chen. Offline model-based optimization: Comprehensive review. *Transactions on Machine Learning Research*, 2026. URL https://openreview.net/forum?id=QcSZWo1TLl.

Siddarth Krishnamoorthy, Satvik Mehul Mashkaria, and Aditya Grover. Diffusion models for black-box optimization. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17842–17857. PMLR, 2023. URL https://proceedings.mlr.press/v202/krishnamoorthy23a.html.

Aviral Kumar, Amir Yazdanbakhsh, Milad Hashemi, Kevin Swersky, and Sergey Levine. Data-driven offline optimization for architecting hardware accelerators. In *Proceedings of the 10th International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=GsH-K1VIyy.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80:1623–1701, 2024. doi: 10.1613/jair.1.15320. URL https://jair.org/index.php/jair/article/view/15320.

Petrus Mikkola, Milica Todorović, Jari Järvi, Patrick Rinke, and Samuel Kaski. Projective preferential Bayesian optimization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 6884–6892. PMLR, 2020. URL https://proceedings.mlr.press/v119/mikkola20a.html.

Quoc Phong Nguyen, Sunil Gupta, Svetha Venkatesh, Bryan Kian Hsiang Low, and Patrick Jaillet. Active set ordering. In *Advances in Neural Information Processing Systems 37*, pp. 108518–108541, 2024. doi: 10.52202/079017-3445. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c4546b4f9e1a44ed15c253dd43307dd5-Abstract-Conference.html.

Yann Ollivier, Ludovic Arnold, Anne Auger, and Nikolaus Hansen. Information-geometric optimization algorithms: A unifying picture via invariance principles. *Journal of Machine Learning Research*, 18(18):1–65, 2017. URL https://www.jmlr.org/papers/v18/14-467.html.

Han Qi, Yi Su, Aviral Kumar, and Sergey Levine. Data-driven offline decision-making via invariant representation learning. In *Advances in Neural Information Processing Systems 35*, pp. 13226–13237, 2022. doi: 10.52202/068431-0961. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/559726fdfb19005e368be4ce3d40e3e5-Abstract-Conference.html.

Hong Qian, Yiyi Zhu, Xiang Shu, Shuo Liu, Yaolin Wen, Xin An, Huakang Lu, Aimin Zhou, Ke Tang, and Yang Yu. SOO-Bench: Benchmarks for evaluating the stability of offline black-box optimization. In *Proceedings of the 13th International Conference on Learning Representations*, volume 2025, pp. 81720–81784, 2025. URL https://openreview.net/forum?id=bqf0aCF3Dd.

Tao Qin, Xu-Dong Zhang, Ming-Feng Tsai, De-Sheng Wang, Tie-Yan Liu, and Hang Li. Query-level loss functions for information retrieval. *Information Processing & Management*, 44(2):838–855, 2008. doi: 10.1016/j.ipm.2007.07.016. URL https://www.sciencedirect.com/science/article/pii/S0306457307001276.

Aniketh Janardhan Reddy, Xinyang Geng, Michael H. Herschl, Sathvik Kolli, Aviral Kumar, Patrick D. Hsu, Sergey Levine, and Nilah M. Ioannidis. Designing cell-type-specific promoter sequences using conservative model-based optimization. In *Advances in Neural Information Processing Systems 37*, pp. 93033–93059, 2024. doi: 10.52202/079017-2954. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/a9619dd0f0d54a5cf7734add1dc38cd1-Abstract-Conference.html.

Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided representation learning for domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32 (1):4058–4065, 2018. doi: 10.1609/aaai.v32i1.11784. URL https://ojs.aaai.org/index.php/AAAI/article/view/11784.

Dong-Hee Shin, Young-Han Son, Hyun Jung Lee, Deok-Joong Lee, and Tae-Eui Kam. Offline model-based optimization for real-world molecular discovery. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55205–55254. PMLR, 2025. URL https://proceedings.mlr.press/v267/shin25g.html.

Shuuki Takizawa, Keita Mori, Naoto Tanishiki, Dai Yoshimura, Atsushi Ohta, and Reiji Teramoto. Safe model based optimization balancing exploration and reliability for protein sequence design. *Scientific Reports*, 15(1):27568, 2025. doi: 10.1038/s41598-025-12568-5. URL https://www.nature.com/articles/s41598-025-12568-5.

Rong-Xi Tan, Ke Xue, Shen-Huan Lyu, Haopu Shang, Yao Wang, Yaoyuan Wang, Sheng Fu, and Chao Qian. Offline model-based optimization by learning to rank. In *Proceedings of the 13th International Conference on Learning Representations*, volume 2025, pp. 47623–47650, 2025. URL https://openreview.net/forum?id=sb1HgVDLjN.

Brandon Trabucco, Aviral Kumar, Xinyang Geng, and Sergey Levine. Conservative objective models for effective offline model-based optimization. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 10358–10368. PMLR, 2021. URL https://proceedings.mlr.press/v139/trabucco21a.html.

Brandon Trabucco, Xinyang Geng, Aviral Kumar, and Sergey Levine. Design-Bench: Benchmarks for data-driven offline model-based optimization. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 21658–21676. PMLR, 2022. URL https://proceedings.mlr.press/v162/trabucco22a.html.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658. URL https://ojs.aaai.org/index.php/AAAI/article/view/3982.

Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. Listwise approach to learning to rank: Theory and algorithm. In *Proceedings of the 25th International Conference on Machine Learning*, pp. 1192–1199. ACM, 2008. doi: 10.1145/1390156.1390306. URL https://doi.org/10.1145/1390156.1390306.

Wenjie Xu, Wenbin Wang, Yuning Jiang, Bratislav Svetozarevic, and Colin Jones. Principled preferential Bayesian optimization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 55305–55336. PMLR, 2024. URL https://proceedings.mlr.press/v235/xu24y.html.

Sihyun Yu, Sungsoo Ahn, Le Song, and Jinwoo Shin. RoMA: Robust model adaptation for offline model-based optimization. In *Advances in Neural Information Processing Systems 34*, pp. 4619–4631, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/24b43fb034a10d78bec71274033b4096-Abstract.html.

Ye Yuan, Can Chen, Zixuan Liu, Willie Neiswanger, and Xue Liu. Importance-aware co-teaching for offline model-based optimization. In *Advances in Neural Information Processing Systems 36*, pp. 55718–55733, 2023. doi: 10.52202/075280-2432. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/ae8b0b5838ba510daff1198474e7b984-Abstract-Conference.html.

Taeyoung Yun, Sujin Yun, Jaewoo Lee, and Jinkyoo Park. Guided trajectory generation with diffusion models for offline model-based optimization. In *Advances in Neural Information Processing Systems 37*, pp. 83847–83876, 2024. doi: 10.52202/079017-2665. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/98904db124b2a2463e8c59ec33fc7150-Abstract-Conference.html.

Tailin Zhou, Zhilin Chen, Wenlong Lyu, Zhitang Chen, Danny H. K. Tsang, and Jun Zhang. Learning design-score manifold to guide diffusion models for offline optimization. *npj Artificial Intelligence*, 2(1):4, 2026. doi: 10.1038/s44387-025-00055-1. URL https://www.nature.com/articles/s44387-025-00055-1.

A Related Work

Offline MBO lies at the intersection of black-box optimization, decision-aligned learning, ranking, and distributional generalization. We organize the most relevant literature according to the role assigned to the surrogate and to the distribution on which it must generalize. This organization distinguishes our contribution from the use of a ranking loss alone: the central object in our analysis is the optimizer-induced distribution of comparisons that can change the terminal decision.

Offline Model-Based Optimization

Offline MBO commonly follows a two-stage pipeline in which a surrogate is fitted to a fixed dataset and subsequently optimized without further access to the objective. The central difficulty is that optimization actively seeks inputs on which prediction errors are favorable, often moving beyond the support of the offline data. Existing methods address this failure mode through several complementary mechanisms. Normalized maximum-likelihood estimation models uncertainty on unfamiliar inputs (Fu & Levine, 2021); conservative objective models penalize high surrogate values on adversarially generated designs (Trabucco et al., 2021); and robust model adaptation improves local reliability around model-generated candidates (Yu et al., 2021). Co-teaching and mentoring approaches instead use interacting surrogates or pseudo-labels to stabilize learning in optimization-relevant regions (Yuan et al., 2023; Chen et al., 2023a). Design-Bench and SOO-Bench provide standardized evaluations of performance and stability across continuous and discrete design tasks (Trabucco et al., 2022; Qian et al., 2025).

Recent work has broadened this algorithmic landscape. Sensitivity regularization constrains how rapidly a surrogate can change around the data (Dao et al., 2024), while gradient matching relates the learned gradient field to downstream optimization quality (Hoang et al., 2024). Structural assumptions can reduce the coverage required for high-dimensional optimization (Grudzien et al., 2024), and diffusion models can learn inverse conditional generators or improvement trajectories rather than optimize a pointwise regressor directly (Krishnamoorthy et al., 2023; Yun et al., 2024). Most directly related to our work, RaM formulates surrogate learning as a ranking problem and demonstrates that ranking quality is more predictive of final design quality than standard regression accuracy (Tan et al., 2025). Our analysis complements these algorithmic advances by asking when a surrogate risk controls terminal selection. It separates candidate reachability from selection, identifies cross-region inversions as the decision-relevant errors, and makes the construction of the training-pair distribution part of the learning problem.

Decision-Focused and Optimization-Aligned Learning

A related line of research studies the mismatch between predictive accuracy and the quality of a downstream decision (Mandi et al., 2024). Task-based learning trains probabilistic models through a stochastic optimization objective (Donti et al., 2017), and decision-focused learning differentiates through combinatorial optimization to favor predictions that induce better decisions (Wilder et al., 2019). Differentiable optimization layers provide a complementary mechanism for end-to-end learning through quadratic and disciplined convex programs (Amos & Kolter, 2017; Agrawal et al., 2019). The predict-then-optimize framework formalizes this principle through the smart predict-then-optimize loss, which evaluates the decision induced by predicted problem parameters rather than their pointwise estimation error (Elmachtoub & Grigas, 2022). Its statistical analysis further shows how decision loss can generalize out of sample under structural conditions on the feasible region (El Balghiti et al., 2023). Collectively, these works establish that the appropriate learning objective depends on how predictions are ultimately used.

Offline MBO shares this objective-alignment principle but differs in the information and decision structure. In conventional predict-then-optimize settings, the downstream optimization problem is known and predictions typically parameterize its objective or constraints. In offline MBO, the true black-box objective cannot be queried, and the learned surrogate both guides search and scores the candidates produced by that search. Consequently, the relevant target distribution is endogenous: changing the surrogate or optimizer changes the comparisons on which correct generalization is required. Our framework therefore aligns learning with a terminal black-box selection event rather than backpropagating through a fixed, known optimization problem.

Learning to Rank and Comparison-Based Optimization

Classical learning to rank provides pointwise, pairwise, and listwise formulations for learning scoring functions. RankNet learns from pairwise preferences using a probabilistic ranking loss (Burges et al., 2005), whereas ListNet treats a list, rather than an isolated pair, as the basic learning instance (Cao et al., 2007). Subsequent work studies the consistency of listwise losses and efficient differentiable ranking operators (Xia et al., 2008; Blondel et al., 2020). These formulations motivate the use of relative supervision when absolute scores are not the evaluation target. The ranking problem studied here, however, has a different target. In standard learning-to-rank applications, the query-conditioned collection of objects is generally supplied by the application. In offline MBO, the target pairs are generated by the learned surrogate and downstream optimizer. Moreover, a globally correct ordering is unnecessary: the risk isolates inversions between reachable ε-optimal and non-ε-optimal candidates because these are the comparisons that can change terminal selection. Pair sampling therefore determines not only computational emphasis but also the training distribution that must transfer to an optimizer-induced target.

Comparison-based optimization offers a second point of contact. Preferential Bayesian optimization learns a latent objective from actively acquired duels when scalar feedback is unavailable (González et al., 2017), with extensions to projective queries in high-dimensional spaces (Mikkola et al., 2020); recent theory provides regret and best-solution guarantees from preference feedback (Xu et al., 2024). Information-geometric optimization and related evolutionary methods use ranks or quantiles of newly evaluated populations to update a search distribution and obtain invariance to increasing transformations of the objective (Ollivier et al., 2017). Active Set Ordering similarly studies black-box maximization through the information contained in relative orderings rather than complete function recovery (Nguyen et al., 2024). These methods use comparisons as online feedback or as a search primitive. Our setting instead begins with fixed scalar observations, permits no new objective or preference queries, and must learn an ordering model that generalizes to unseen candidates produced by offline search.

Distribution Shift, Coverage, and Learnability

Distribution shift in offline MBO is coupled to optimization: the optimizer deliberately moves toward predicted high-value regions, which are often weakly represented in the data. Uncertainty modeling, conservative objectives, local adaptation, and sensitivity control can all be viewed as responses to this coverage gap (Fu & Levine, 2021; Trabucco et al., 2021; Yu et al., 2021; Dao et al., 2024). Other work studies conditions under which structure or gradient information makes optimization possible despite limited coverage (Grudzien et al., 2024; Hoang et al., 2024). These approaches primarily control the surrogate or the search trajectory. Our geometric analysis instead decomposes the transfer difficulty faced by a ranking surrogate into displacement beyond the support of positive training examples and residual mismatch within that support.

Optimal transport has also been used to quantify and reduce distribution mismatch in supervised transfer. For example, Wasserstein representation learning aligns source and target features (Shen et al., 2018), while joint-distribution optimal transport aligns feature–label distributions and relates the transport objective to target error (Courty et al., 2017). Classical domain-adaptation formulations treat the source and target domains as externally specified. In offline MBO, the target domain relevant to the decision is a distribution over ordered candidate pairs induced by the learned surrogate and search procedure. By making this endogenous pair distribution explicit, our analysis connects its inversion risk to terminal selection and separates finite-sample complexity, training–target pair mismatch, marginal transport, and off-support geometric displacement. DAR then acts on the component available at training time by reshaping comparisons toward an empirical proxy for the downstream decision distribution.

B Notation and Technical Conditions

Table 3 consolidates the notation used in the optimization, learning, and distribution-shift analyses. The dataset size m, sampled-pair size N, objective tolerance ε, empirical top fraction τ, analysis margin γ, and DAR margin β are distinct quantities.

Table 3: Notation used throughout the paper.

Symbol	Meaning
$\mathcal{X}, f, \boldsymbol{x}^\star$	Design space, unknown objective, and a true maximizer of f.
$S = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^m$	Fixed offline dataset of m evaluated designs, with $y_i = f(\boldsymbol{x}_i) + \xi_i$.
$h, \widehat{h}, \mathcal{H}$	Generic surrogate, learned surrogate, and surrogate class used in the theoretical analysis.
$\mathcal{A}, T, \mathcal{C}$	Search procedure, search budget, and abbreviated optimizer-generated candidate set.
$\widehat{\boldsymbol{x}}_{\mathrm{MBO}}$	Candidate selected by maximizing $\widehat{h}$ over $\mathcal{C}$.
$\varepsilon, \mathcal{X}_\varepsilon, \mathcal{X}_\varepsilon^c$	Objective-gap tolerance, ε-optimal region, and non-ε-optimal region.
$\mathcal{R}_\varepsilon, \delta_{\mathrm{opt}}$	Reachability event and search-failure probability.
$X^+, X^-, \rho_\varepsilon^+, \rho_\varepsilon^-$	Reachable ε-optimal and non-ε-optimal candidates and their conditional laws.
Q_ε	Optimization-relevant target pair distribution $\rho_\varepsilon^+ \otimes \rho_\varepsilon^-$.
$Q_{\mathrm{tr}}, Z_{\mathrm{tr}}, N$	Training-pair distribution, i.i.d. ordered pair sample $Z_{\mathrm{tr}} \sim Q_{\mathrm{tr}}^N$, and sampled-pair size.
$\mu_{\mathrm{tr}}^+, \mu_{\mathrm{tr}}^-$	First- and second-coordinate marginals when Q_{tr} factorizes.
$\ell_{01}, \widetilde{\ell}, \mathcal{G}$	The 0–1 inversion loss, a calibrated pairwise surrogate, and its induced loss class.
$\mathcal{E}_\varepsilon, \widehat{\mathcal{E}}_{\mathrm{tr}}$	Target inversion risk and empirical training surrogate risk.
Γ_N, Δ_ℓ	Statistical complexity term and surrogate-loss discrepancy between pair distributions.
$e_h, R_+^{(p)}, R_-^{(p)}, M_\varepsilon(\gamma)$	Pointwise residual, regional p-power risks, and small-objective-margin mass.
W_1, d_{pair}	1-Wasserstein distance and the additive Euclidean metric on ordered design pairs.
$\mathcal{M}, \Pi_{\mathcal{M}}$	Positive-coordinate training support and a measurable nearest-point projection.
$D_{\mathrm{out}}, D_{\mathrm{in}}$	Off-support displacement and residual within-support mismatch.
τ, S_τ^+, S_τ^-	Empirical top fraction, high-quality subset, and remaining subset used by DAR.
$\lambda, \beta, Q_{\mathrm{DAR}}$	Intra-region pair ratio, training margin, and DAR training-pair distribution.

The formal terms ε-*optimal* and *non-*ε*-optimal* always refer to $\mathcal{X}_\varepsilon$ and $\mathcal{X}_\varepsilon^c$. In formal method definitions, the empirical terms *high-quality subset* and *remaining subset* refer to S_τ^+ and S_τ^-; they do not presume oracle knowledge of f. When sampling designs or defining $S_\tau^\pm$, S denotes its design coordinates, with the corresponding labels retained for comparisons.

Pairwise losses and statistical complexity

For an ordered pair whose first coordinate is preferred, the inversion loss is $\ell_{01}(h; \boldsymbol{x}, \boldsymbol{x}') = \mathbf{1}\{h(\boldsymbol{x}) \le h(\boldsymbol{x}')\}$. The main results assume a measurable surrogate loss satisfying $\ell_{01}(h; \boldsymbol{x}, \boldsymbol{x}') \le \widetilde{\ell}(h; \boldsymbol{x}, \boldsymbol{x}') \le B$ for all $h \in \mathcal{H}$ and all relevant pairs. The lower inequality is the calibration condition needed to control inversions; bounded score ranges or clipping can enforce the upper inequality used for concentration.

For completeness, the empirical Rademacher complexity of a real-valued function class $\mathcal{F}$ on a sample $Z = \{z_i\}_{i=1}^n$ is

$$\widehat{\mathfrak{R}}_Z(\mathcal{F}) := \mathbb{E}_\sigma \left[\sup_{g \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sigma_i g(z_i) \right],$$

where the σ_i are independent Rademacher variables. Applied to $\mathcal{G}$, this definition gives the complexity component in $\Gamma_N(\mathcal{G}, \delta)$.

Transport regularity

The distribution-aware results assume finite first moments and uniform pairwise Lipschitz continuity:

$$\left| \widetilde{\ell}(h; \boldsymbol{x}, \boldsymbol{x}') - \widetilde{\ell}(h; \boldsymbol{z}, \boldsymbol{z}') \right| \le L(\|\boldsymbol{x} - \boldsymbol{z}\|_2 + \|\boldsymbol{x}' - \boldsymbol{z}'\|_2) \quad \text{for every } h \in \mathcal{H}.$$

These conditions ensure that the surrogate-loss expectation shift is controlled by the 1-Wasserstein distance on $(\mathcal{X} \times \mathcal{X}, d_{\mathrm{pair}})$.

C Proofs of the Main Theoretical Results

This section proves the selection and distribution-shift results in Sections 2 and 3. The regression reduction is proved in Appendix D. Auxiliary lemmas are stated where they enter the proof chain.

Proof of Proposition 2

Proof. Because $\mathcal{R}_\varepsilon$ occurs, choose any $\boldsymbol{x}^+ \in \mathcal{C} \cap \mathcal{X}_\varepsilon$. If a maximizer in (1) belonged to $\mathcal{X}_\varepsilon^c$, its surrogate score would be at least $\widehat{h}(\boldsymbol{x}^+)$, contradicting the assumed strict cross-region ordering. Thus $\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon$, and (2) gives the stated regret bound. □

Selection-to-ranking connection

Lemma 7 (Selection-to-ranking)**.** *Conditional on $\mathcal{R}_\varepsilon$, suppose the candidate pool contains $X^+ \sim \rho_\varepsilon^+$ and at most K candidates $X_1^-, \dots, X_K^- \in \mathcal{X}_\varepsilon^c$, with each $(X^+, X_j^-) \sim Q_\varepsilon$. Under exact maximization over the candidate set,*

$$\Pr(\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon^c \mid \mathcal{R}_\varepsilon) \le K\mathcal{E}_\varepsilon(\widehat{h}).$$

Proof. Selection of a non-ε-optimal candidate implies that at least one X_j^- satisfies $\widehat{h}(X^+) \le \widehat{h}(X_j^-)$. Therefore,

$$\Pr(\widehat{\boldsymbol{x}}_{\mathrm{MBO}} \in \mathcal{X}_\varepsilon^c \mid \mathcal{R}_\varepsilon) \le \sum_{j=1}^{K} \Pr\Big(\widehat{h}(X^+) \le \widehat{h}(X_j^-)\Big) = K\mathcal{E}_\varepsilon(\widehat{h}),$$

where no independence among the X_j^- is required. Combining this result with (3) proves (5). □

Uniform convergence on sampled pairs

Lemma 8 (Pair-sample uniform convergence)**.** *With probability at least $1-\delta$ over $Z_{\mathrm{tr}} \sim Q_{\mathrm{tr}}^N$,*

$$\sup_{h \in \mathcal{H}} \left| \mathbb{E}_{Q_{\mathrm{tr}}} \widetilde{\ell}(h) - \widehat{\mathcal{E}}_{\mathrm{tr}}(h) \right| \le \Gamma_N(\mathcal{G}, \delta).$$

Proof. Every function in $\mathcal{G}$ takes values in $[0, B]$. The standard two-sided Rademacher uniform-convergence inequality therefore gives

$$\sup_{h \in \mathcal{H}} \left| \mathbb{E}_{Q_{\mathrm{tr}}} \widetilde{\ell}(h) - \widehat{\mathcal{E}}_{\mathrm{tr}}(h) \right| \le 2\widehat{\mathfrak{R}}_{Z_{\mathrm{tr}}}(\mathcal{G}) + 3B\sqrt{\frac{\log(2/\delta)}{2N}}.$$

The right-hand side is $\Gamma_N(\mathcal{G}, \delta)$ by (6). The offline dataset is fixed; the independence assumption is imposed only on the sampled ordered pairs. □

Proof of Theorem 3

Proof. Fix $h \in \mathcal{H}$. Surrogate domination and the definition of Δ_ℓ in (6) imply

$$\begin{aligned} \mathcal{E}_\varepsilon(h) = \mathbb{E}_{Q_\varepsilon} \ell_{01}(h) &\le \mathbb{E}_{Q_\varepsilon} \widetilde{\ell}(h) \\ &\le \mathbb{E}_{Q_{\mathrm{tr}}} \widetilde{\ell}(h) + \Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}}). \end{aligned}$$

Lemma 8 holds simultaneously for all $h \in \mathcal{H}$. Substitution proves the first inequality in (7). Applying it to the data-dependent $\widehat{h}$ and then invoking Lemma 7 and Assumption 1 gives

$$\Pr(f(\boldsymbol{x}^\star) - f(\widehat{\boldsymbol{x}}_{\mathrm{MBO}}) > \varepsilon) \le \delta_{\mathrm{opt}} + K\left[\widehat{\mathcal{E}}_{\mathrm{tr}}(\widehat{h}) + \Gamma_N(\mathcal{G}, \delta) + \Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}})\right],$$

which is the second claim. □

LIPSCHITZ EXPECTATION SHIFT

Lemma 9 (Wasserstein expectation shift)**.** *Under the transport regularity conditions in Appendix B, every $h \in \mathcal{H}$ satisfies*

$$\left|\mathbb{E}_{Q_\varepsilon}\widetilde{\ell}(h) - \mathbb{E}_{Q_{\mathrm{tr}}}\widetilde{\ell}(h)\right| \le L W_1(Q_\varepsilon, Q_{\mathrm{tr}}).$$

Proof. Fix $h \in \mathcal{H}$ and write

$$g(\boldsymbol{x}, \boldsymbol{x}') := \widetilde{\ell}(h; \boldsymbol{x}, \boldsymbol{x}').$$

By the transport regularity condition, g is L-Lipschitz under d_{pair}:

$$\big|g(\boldsymbol{x}, \boldsymbol{x}') - g(\boldsymbol{z}, \boldsymbol{z}')\big| \le L\, d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) \quad \text{for all } (\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}') \in \mathcal{X} \times \mathcal{X}.$$

Let Q_ε and Q_{tr} be the two probability measures on $\mathcal{X} \times \mathcal{X}$ endowed with the metric d_{pair}. By the primal formulation of the 1-Wasserstein distance, we have

$$W_1(Q_\varepsilon, Q_{\mathrm{tr}}) = \inf_{\pi \in \Gamma(Q_\varepsilon, Q_{\mathrm{tr}})} \int d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big)\, \mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big),$$

where $\Gamma(Q_\varepsilon, Q_{\mathrm{tr}})$ denotes the set of all couplings π on $(\mathcal{X} \times \mathcal{X})^2$ whose first marginal is Q_ε and second marginal is Q_{tr}.

Fix any $\pi \in \Gamma(Q_\varepsilon, Q_{\mathrm{tr}})$. By the marginal constraints on π,

$$\begin{aligned}
\mathbb{E}_{Q_\varepsilon} g(\boldsymbol{x}, \boldsymbol{x}') - \mathbb{E}_{Q_{\mathrm{tr}}} g(\boldsymbol{x}, \boldsymbol{x}') &= \int g(\boldsymbol{x}, \boldsymbol{x}')\, \mathrm{d}Q_\varepsilon(\boldsymbol{x}, \boldsymbol{x}') - \int g(\boldsymbol{z}, \boldsymbol{z}')\, \mathrm{d}Q_{\mathrm{tr}}(\boldsymbol{z}, \boldsymbol{z}') \\
&= \int \Big(g(\boldsymbol{x}, \boldsymbol{x}') - g(\boldsymbol{z}, \boldsymbol{z}')\Big)\, \mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big).
\end{aligned}$$

Taking absolute values and applying the Lipschitz property of g yields

$$\begin{aligned}
|\mathbb{E}_{Q_\varepsilon} g(\boldsymbol{x}, \boldsymbol{x}') - \mathbb{E}_{Q_{\mathrm{tr}}} g(\boldsymbol{x}, \boldsymbol{x}')| &\le \int \big|g(\boldsymbol{x}, \boldsymbol{x}') - g(\boldsymbol{z}, \boldsymbol{z}')\big|\, \mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) \\
&\le \int L\, d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big)\, \mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) \\
&= L \int d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big)\, \mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big).
\end{aligned}$$

Because this inequality holds for every coupling π, taking the infimum over $\Gamma(Q_\varepsilon, Q_{\mathrm{tr}})$ gives

$$\big|\mathbb{E}_{Q_\varepsilon}\widetilde{\ell}(h) - \mathbb{E}_{Q_{\mathrm{tr}}}\widetilde{\ell}(h)\big| = \big|\mathbb{E}_{Q_\varepsilon} g - \mathbb{E}_{Q_{\mathrm{tr}}} g\big| \le L\, W_1(Q_\varepsilon, Q_{\mathrm{tr}}).$$

Since h was arbitrary, the result holds uniformly over $\mathcal{H}$. □

PRODUCT-PAIR WASSERSTEIN DECOMPOSITION

Lemma 10 (Product-pair decomposition)**.** *If $Q_\varepsilon = \rho_\varepsilon^+ \otimes \rho_\varepsilon^-$ and $Q_{\mathrm{tr}} = \mu_{\mathrm{tr}}^+ \otimes \mu_{\mathrm{tr}}^-$, then the pairwise Wasserstein distance satisfies*

$$W_1(Q_\varepsilon, Q_{\mathrm{tr}}) \;\le\; W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) + W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-).$$

Proof. We have

$$Q_\varepsilon = \rho_\varepsilon^+ \otimes \rho_\varepsilon^-, \qquad Q_{\mathrm{tr}} = \mu_{\mathrm{tr}}^+ \otimes \mu_{\mathrm{tr}}^-,$$

and the pairwise metric

$$d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) := \|\boldsymbol{x} - \boldsymbol{z}\|_2 + \|\boldsymbol{x}' - \boldsymbol{z}'\|_2.$$

The 1-Wasserstein distance between Q_ε and Q_{tr} is defined as

$$W_1(Q_\varepsilon, Q_{\mathrm{tr}}) := \inf_{\pi \in \Gamma(Q_\varepsilon, Q_{\mathrm{tr}})} \int d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big)\, \mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big),$$

where $\Gamma(Q_\varepsilon, Q_{\mathrm{tr}})$ is the set of all couplings with marginals Q_ε and Q_{tr}.

Let γ_1 be an optimal coupling between ρ_ε^+ and μ_{tr}^+, i.e.,

$$\gamma_1 \in \Gamma(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+), \qquad \int \|\boldsymbol{x} - \boldsymbol{z}\|_2 \,\mathrm{d}\gamma_1(\boldsymbol{x}, \boldsymbol{z}) = W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+),$$

and let γ_2 be an optimal coupling between ρ_ε^- and μ_{tr}^-, i.e.,

$$\gamma_2 \in \Gamma(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-), \qquad \int \|\boldsymbol{x}' - \boldsymbol{z}'\|_2 \,\mathrm{d}\gamma_2(\boldsymbol{x}', \boldsymbol{z}') = W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-).$$

Define the product coupling $\pi := \gamma_1 \otimes \gamma_2$ on $(\mathcal{X} \times \mathcal{X})^2$.

$$\pi\big(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{x}', \mathrm{d}\boldsymbol{z}, \mathrm{d}\boldsymbol{z}'\big) := \gamma_1(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{z})\, \gamma_2(\mathrm{d}\boldsymbol{x}', \mathrm{d}\boldsymbol{z}').$$

Because the first marginals of γ_1 and γ_2 are ρ_ε^+ and ρ_ε^-, the first pair marginal of π is Q_ε. Likewise, its second pair marginal is Q_{tr}. Hence $\pi \in \Gamma(Q_\varepsilon, Q_{\mathrm{tr}})$.

Next, we compute the transportation cost of π:

$$\begin{aligned}\int d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) \,\mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) &= \int \big(\|\boldsymbol{x} - \boldsymbol{z}\|_2 + \|\boldsymbol{x}' - \boldsymbol{z}'\|_2\big) \,\mathrm{d}\gamma_1(\boldsymbol{x}, \boldsymbol{z})\, \mathrm{d}\gamma_2(\boldsymbol{x}', \boldsymbol{z}') \\ &= \int \|\boldsymbol{x} - \boldsymbol{z}\|_2 \,\mathrm{d}\gamma_1(\boldsymbol{x}, \boldsymbol{z}) + \int \|\boldsymbol{x}' - \boldsymbol{z}'\|_2 \,\mathrm{d}\gamma_2(\boldsymbol{x}', \boldsymbol{z}') \\ &= W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) + W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-).\end{aligned}$$

Since $W_1(Q_\varepsilon, Q_{\mathrm{tr}})$ is the infimum of the transportation cost over all couplings, we conclude that

$$W_1(Q_\varepsilon, Q_{\mathrm{tr}}) \;\le\; \int d_{\mathrm{pair}}\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) \,\mathrm{d}\pi\big((\boldsymbol{x}, \boldsymbol{x}'), (\boldsymbol{z}, \boldsymbol{z}')\big) = W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) + W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-),$$

which proves the desired inequality. □

Proof of Corollary 5

Proof. Taking the supremum over $h \in \mathcal{H}$ in Lemma 9 yields

$$\Delta_\ell(Q_\varepsilon, Q_{\mathrm{tr}}) \le L W_1(Q_\varepsilon, Q_{\mathrm{tr}}).$$

Substitution into (7) proves part (a) of (9). If $Q_{\mathrm{tr}} = \mu_{\mathrm{tr}}^+ \otimes \mu_{\mathrm{tr}}^-$, Lemma 10 further gives

$$W_1(Q_\varepsilon, Q_{\mathrm{tr}}) \le W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) + W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-).$$

Substituting this inequality into part (a) of (9) proves part (b) of (9). □

Projection decomposition

We first separate displacement to the training support from residual mismatch within that support.

Lemma 11 (Projection decomposition)**.** *Under the notation of Corollary 6,*

$$W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) \le \mathbb{E}_{\boldsymbol{x} \sim \rho_\varepsilon^+} \operatorname{dist}(\boldsymbol{x}, \mathcal{M}) + D_{\mathrm{in}}.$$

Proof. Let $\bar{\rho}_\varepsilon^+ := (\Pi_{\mathcal{M}})_{\#}\rho_\varepsilon^+$. Coupling each $\boldsymbol{x} \sim \rho_\varepsilon^+$ with $\Pi_{\mathcal{M}}(\boldsymbol{x})$ gives

$$W_1(\rho_\varepsilon^+, \bar{\rho}_\varepsilon^+) \le \mathbb{E}_{\boldsymbol{x} \sim \rho_\varepsilon^+} \|\boldsymbol{x} - \Pi_{\mathcal{M}}(\boldsymbol{x})\|_2 = \mathbb{E}_{\boldsymbol{x} \sim \rho_\varepsilon^+} \operatorname{dist}(\boldsymbol{x}, \mathcal{M}).$$

The triangle inequality and the definition $D_{\mathrm{in}} = W_1(\bar{\rho}_\varepsilon^+, \mu_{\mathrm{tr}}^+)$ complete the proof. □

Proof of Corollary 6

Proof. Using part (b) of (9), we obtain

$$\mathcal{E}_\varepsilon(h) \le \widehat{\mathcal{E}}_{\mathrm{tr}}(h) + \Gamma_N(\mathcal{G}, \delta) + L\big[W_1(\rho_\varepsilon^+, \mu_{\mathrm{tr}}^+) + W_1(\rho_\varepsilon^-, \mu_{\mathrm{tr}}^-)\big].$$

Lemma 11 bounds the first Wasserstein term by $D_{\mathrm{out}} + D_{\mathrm{in}}$. Substitution gives (10). □

D Regression Guarantees and Counterexamples

We first prove Proposition 4, then give complementary regression guarantees and a counterexample to ordering optimization quality by global prediction error.

Proof of Proposition 4

Define the separated-pair error

$$\mathcal{E}_{\varepsilon,\gamma}(h) := \mathbb{E}_{Q_\varepsilon}\left[\mathbf{1}\{h(X^+) \le h(X^-)\}\mathbf{1}\{f(X^+) - f(X^-) \ge \gamma\}\right].$$

We first establish the reduction used in the proposition.

Lemma 12 (Separated-pair regression reduction)**.** *For every $h \in \mathcal{H}$, $1 \le p < \infty$, and $\gamma > 0$,*

$$\mathcal{E}_{\varepsilon,\gamma}(h) \le \frac{2^{p-1}}{\gamma^p}\left(R_+^{(p)}(h) + R_-^{(p)}(h)\right).$$

Proof. If a target pair is inverted and has objective gap at least γ, then

$$\gamma \le f(X^+) - f(X^-) \le |e_h(X^+)| + |e_h(X^-)|.$$

Consequently, for every $1 \le p < \infty$,

$$\begin{aligned}\mathbf{1}\{h(X^+) \le h(X^-)\}\mathbf{1}\{f(X^+) - f(X^-) \ge \gamma\} &\le \frac{\left(|e_h(X^+)| + |e_h(X^-)|\right)^p}{\gamma^p} \\ &\le \frac{2^{p-1}}{\gamma^p}\left(|e_h(X^+)|^p + |e_h(X^-)|^p\right).\end{aligned}$$

Taking expectations under $Q_\varepsilon = \rho_\varepsilon^+ \otimes \rho_\varepsilon^-$ proves the lemma. □

Proof. Every pair in the support of Q_ε has a positive objective gap. Splitting the misranking event according to whether this gap is smaller than γ gives

$$\mathcal{E}_\varepsilon(h) \le M_\varepsilon(\gamma) + \mathcal{E}_{\varepsilon,\gamma}(h).$$

Applying Lemma 12 to the second term proves (8). □

Uniform-error guarantee

Define $r_\infty^+(h) := \operatorname{ess\,sup}_{X\sim\rho_\varepsilon^+} |e_h(X)|$ and $r_\infty^-(h) := \operatorname{ess\,sup}_{X\sim\rho_\varepsilon^-} |e_h(X)|$.

Proposition 13 (L_∞-to-ranking guarantee)**.** *If $r_\infty^+(h) + r_\infty^-(h) < \gamma$, then*

$$\mathcal{E}_{\varepsilon,\gamma}(h) = 0, \qquad \mathcal{E}_\varepsilon(h) \le M_\varepsilon(\gamma).$$

Proof. An inverted pair with objective gap at least γ would satisfy

$$\gamma \le |e_h(X^+)| + |e_h(X^-)| \le r_\infty^+(h) + r_\infty^-(h),$$

contradicting the assumption. Hence the separated-pair error vanishes. Splitting the full inversion event into small- and large-margin pairs proves the second claim. □

Global error need not order optimization quality

Proposition 14 (Non-monotonicity of global L_p error)**.** *There exist a finite objective f and surrogates h_A, h_B with unique maximizers $\widehat{\boldsymbol{x}}_A, \widehat{\boldsymbol{x}}_B$ such that, simultaneously for every $p \in [1,\infty]$ under the uniform distribution on the design set,*

$$\|h_A - f\|_{L_p} < \|h_B - f\|_{L_p}, \qquad f(\widehat{\boldsymbol{x}}_A) < f(\widehat{\boldsymbol{x}}_B).$$

Proof. Consider five uniformly weighted designs with

$$f = (0,\ 0.25,\ 0.50,\ 0.75,\ 1.00), \quad h_A = (0,\ 0.25,\ 0.50,\ 0.95,\ 0.90), \quad h_B = f + 0.30.$$

For every finite $p \geq 1$,

$$\|h_A - f\|_{L_p} = \left(\frac{0.20^p + 0.10^p}{5}\right)^{1/p} \leq 0.20 < 0.30 = \|h_B - f\|_{L_p},$$

and the strict inequality also holds for $p = \infty$. Nevertheless, $\widehat{\boldsymbol{x}}_A = \boldsymbol{x}_4$ and $\widehat{\boldsymbol{x}}_B = \boldsymbol{x}_5$, so $f(\widehat{\boldsymbol{x}}_A) = 0.75 < 1 = f(\widehat{\boldsymbol{x}}_B)$. □

FINITE-SAMPLE MSE SPECIALIZATION

For this subsection, let ρ_{data} denote the population distribution of offline designs and define $\rho_{\mathrm{mix}} := \frac{1}{2}\rho_\varepsilon^+ + \frac{1}{2}\rho_\varepsilon^-$. For a class $\mathcal{H}_{\mathrm{mse}}$, write $R^{\mathrm{mse}}_{\mathrm{data}}(h) := \mathbb{E}_{\rho_{\mathrm{data}}} e_h(X)^2$, $R^{\mathrm{mse}}_{\mathrm{mix}}(h) := \mathbb{E}_{\rho_{\mathrm{mix}}} e_h(X)^2$, and

$$\Delta_{\mathrm{mse}} := \sup_{h \in \mathcal{H}_{\mathrm{mse}}} \left| R^{\mathrm{mse}}_{\mathrm{mix}}(h) - R^{\mathrm{mse}}_{\mathrm{data}}(h) \right|.$$

Theorem 15 (Finite-sample MSE bound)**.** *Let $S_X = \{\boldsymbol{x}_i\}_{i=1}^m \sim \rho_{\mathrm{data}}^m$ and $\widehat{h}_{\mathrm{mse}} \in \arg\min_{h \in \mathcal{H}_{\mathrm{mse}}} \widehat{R}^{\mathrm{mse}}_S(h)$, where $\widehat{R}^{\mathrm{mse}}_S(h) := m^{-1}\sum_{i=1}^m e_h(\boldsymbol{x}_i)^2$. If $\mathcal{F}_{\mathrm{mse}} := \{\boldsymbol{x} \mapsto e_h(\boldsymbol{x})^2 : h \in \mathcal{H}_{\mathrm{mse}}\}$ is bounded in $[0,1]$, then, with probability at least $1-\delta$,*

$$\mathcal{E}_{\varepsilon,\gamma}(\widehat{h}_{\mathrm{mse}}) \leq \frac{4}{\gamma^2}\left[\widehat{R}^{\mathrm{mse}}_S(\widehat{h}_{\mathrm{mse}}) + 2\widehat{\mathfrak{R}}_{S_X}(\mathcal{F}_{\mathrm{mse}}) + 3\sqrt{\frac{\log(2/\delta)}{2m}} + \Delta_{\mathrm{mse}}\right].$$

Proof. Lemma 12 with $p = 2$ gives

$$\mathcal{E}_{\varepsilon,\gamma}(\widehat{h}_{\mathrm{mse}}) \leq \frac{4}{\gamma^2} R^{\mathrm{mse}}_{\mathrm{mix}}(\widehat{h}_{\mathrm{mse}}).$$

By definition, $R^{\mathrm{mse}}_{\mathrm{mix}}(\widehat{h}_{\mathrm{mse}}) \leq R^{\mathrm{mse}}_{\mathrm{data}}(\widehat{h}_{\mathrm{mse}}) + \Delta_{\mathrm{mse}}$. The standard Rademacher uniform-convergence bound for $\mathcal{F}_{\mathrm{mse}}$ yields

$$R^{\mathrm{mse}}_{\mathrm{data}}(\widehat{h}_{\mathrm{mse}}) \leq \widehat{R}^{\mathrm{mse}}_S(\widehat{h}_{\mathrm{mse}}) + 2\widehat{\mathfrak{R}}_{S_X}(\mathcal{F}_{\mathrm{mse}}) + 3\sqrt{\frac{\log(2/\delta)}{2m}}.$$

Combining the inequalities proves the result. This statement uses the noise-free targets $f(\boldsymbol{x}_i)$ to isolate the regression comparison; observation noise introduces an additional estimation term. □

E PROOF OF THE DAR ALIGNMENT BOUND

Proposition 16 (Alignment of the DAR pair distribution)**.** *Under the finite-first-moment conditions in Section 4, the mixture in* (13) *satisfies the alignment bound in* (14)*.*

Proof. Convexity of W_1 in its second argument and (13) give

$$W_1(Q_\varepsilon, Q_{\mathrm{DAR}}) \leq (1-\lambda) W_1(Q_\varepsilon, \widehat{Q}_\tau) + \lambda W_1(Q_\varepsilon, Q_{\mathrm{intra}}).$$

The triangle inequality bounds the last term by $W_1(Q_\varepsilon, \widehat{Q}_\tau) + W_1(\widehat{Q}_\tau, Q_{\mathrm{intra}})$. Collecting coefficients proves (14). □

F Detailed Algorithmic Description

Algorithm 1 gives DAR training and one search trajectory, whose output is a proposed candidate. The 128-proposal evaluation in Section 5.2 uses the best ground-truth score in the batch; selection by surrogate score is defined in (1).

Algorithm 1 Distribution-Aware Ranking (DAR) for Offline Model-Based Optimization

Require: Offline dataset $S = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^m$; top-sample proportion τ; margin β; intra-region ratio λ; step size η; number of steps T
Ensure: One proposed candidate $\boldsymbol{x}^{(T)}$

Compute $q_{1-\tau} \leftarrow \mathrm{Quantile}_{1-\tau}(\{y_i\}_{i=1}^m)$
$S_\tau^+ \leftarrow \{\boldsymbol{x}_i \mid y_i \geq q_{1-\tau}\}$
$S_\tau^- \leftarrow S \setminus S_\tau^+$
Initialize ranking model parameters θ
for each training iteration **do**
  if with probability $1-\lambda$ **then**
    Sample $\boldsymbol{x}_1 \sim S_\tau^+$ and $\boldsymbol{x}_2 \sim S_\tau^-$
  else
    Sample $\boldsymbol{x}_i, \boldsymbol{x}_j \sim S_\tau^+$; set $(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}_i, \boldsymbol{x}_j)$ if $y_i > y_j$ and swap otherwise
  end if
  $\mathcal{L} \leftarrow \max\{0, \beta - (h_\theta(\boldsymbol{x}_1) - h_\theta(\boldsymbol{x}_2))\}$
  Update θ by gradient descent on $\mathcal{L}$
end for
Evaluate h_θ on S and compute prediction mean $\widetilde{\mu}$ and standard deviation $\widetilde{\sigma} > 0$
Define $\widetilde{h}_\theta(\boldsymbol{x}) \leftarrow (h_\theta(\boldsymbol{x}) - \widetilde{\mu})/\widetilde{\sigma}$
Initialize $\boldsymbol{x}^{(0)} \sim S_\tau^+$
for $t = 1$ to T **do**
  $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{x}^{(t-1)} + \eta \nabla_{\boldsymbol{x}} \widetilde{h}_\theta(\boldsymbol{x}^{(t-1)})$
end for
return $\boldsymbol{x}^{(T)}$

G Discussion and Broader Implications

Our results recast learnability in offline MBO as decision-aligned generalization over optimizer-generated candidates. The relevant question is not whether a surrogate reconstructs the objective globally, but whether it preserves the comparisons governing selection among those candidates. Search determines which candidates enter the decision, whereas pair construction determines which comparisons receive statistical capacity. The training-pair distribution is therefore an algorithmic object, not merely a property inherited from the offline dataset. This perspective is related to classical learning to rank, including pairwise methods such as RankNet (Burges et al., 2005), and extends recent rank-based offline MBO (Tan et al., 2025). Here, however, the target distribution is endogenous to the learned surrogate and optimizer, the risk isolates cross-region inversions that can alter ε-optimal selection, and pair sampling determines the training distribution appearing in the transfer discrepancy. The framework thus defines an optimization-aligned ranking problem over an optimizer-induced target distribution, rather than simply importing a ranking loss into offline MBO.

This view also distinguishes offline MBO from other comparison-based optimization settings. Preferential Bayesian optimization can actively query pairwise preferences (González et al., 2017), while online rank-based optimizers use newly evaluated population rankings to guide search (Ollivier et al., 2017). Offline MBO instead relies on fixed historical observations, which we reorganize into decision-relevant pairs for ranking unseen candidates. Together with the ordering view of black-box maximization (Nguyen et al., 2024), this suggests that, in historical-data-to-selection pipelines, preserving decision-relevant orderings may matter more than globally accurate value prediction.