Is Vanilla Bayesian Optimization Sufficient in High-Dimensional Permutation Spaces?

Ke Xue [1 2] **Zhi-Xuan Gao** [1 2] **Ruo-Tong Chen** [1 2] **Rong-Xi Tan** [1 2] **Zhengbo Lu** [3 4] **Xudong Hou** [3 4] **Junxuan Fan** [3 4] **Chao Qian** [1 2]

Abstract

High-dimensional Bayesian Optimization (BO) has recently witnessed a paradigm shift. Accumulating evidence suggests that "Vanilla BO", i.e., standard Gaussian Processes with robust kernel initialization and length-scale scaling, can achieve state-of-the-art performance in high-dimensional tasks, casting doubt on the necessity of complex dimensionality reduction techniques. However, this narrative largely overlooks permutation spaces, a distinct and ubiquitous class of discrete combinatorial problems critical to various domains such as industrial applications and scientific discovery. In this work, we revisit the efficacy of vanilla BO in high-dimensional permutation spaces on these challenging tasks. We demonstrate that standard global kernels fail in this setting due to the concentration of measure and topological ruggedness inherent to factorial search spaces. Contrary to trends in continuous optimization, we find that dimensionality reduction techniques (i.e., variable selection) are not outdated but essential for restoring kernel informativeness, yielding substantial improvements in sample efficiency and optimization quality. Our findings urge a reevaluation of vanilla BO's generality, as well as highlight the enduring value of dimensionality reduction for intractable high-dimensional tasks.

[1]State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China [2]School of Artificial Intelligence, Nanjing University, Nanjing, China [3]State Key Laboratory of Critical Earth Material Cycling and Mineral Deposits, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China [4]Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing, China. Correspondence to: Chao Qian <qianc@lamda.nju.edu.cn>.

Submitted to the AI for Science workshop (ICML 2026).

1. Introduction

Bayesian optimization (BO) has emerged as a principled framework for optimizing expensive black-box functions, particularly in scenarios where gradients are unavailable and function evaluations are costly (Shahriari et al., 2015; Frazier, 2018; Garnett, 2023). Classical BO with Gaussian processes (GPs) (Rasmussen & Williams, 2006) as a surrogate model often struggles in high-dimensional space (Binois & Wycoff, 2022). To address this, the community has traditionally relied on dimensionality reduction techniques, such as additive models (Kandasamy et al., 2015; Rolland et al., 2018; Mutný & Krause, 2018), embeddings (Wang et al., 2016; Letham et al., 2020; Papenmeier et al., 2022), and axis-aligned variable selection (Li et al., 2017; Eriksson & Jankowiak, 2021; Song et al., 2022), to mitigate the curse of dimensionality.

However, high-dimensional BO has recently witnessed a paradigm shift. Accumulating evidence suggests that "Vanilla BO", specifically standard GPs equipped with robust kernel initialization and dimension-aware length-scale scaling, can achieve state-of-the-art performance in high-dimensional continuous tasks (Hvarfner et al., 2024; Xu et al., 2025; Papenmeier et al., 2025b). This trend has cast doubt on the necessity of dimensionality reduction techniques, suggesting that the perceived failure of standard GPs is often a result of poor configuration rather than inherent modeling constraints.

While this narrative is compelling for continuous optimization, it largely overlooks permutation spaces, which represent a distinct and ubiquitous class of discrete combinatorial problems critical to industrial applications and scientific discovery (Jiao & Vert, 2015; Deshwal et al., 2022; Oh et al., 2022a; Xie & Chen, 2025). Unlike continuous domains where the search volume typically grows polynomially, the size of permutation spaces scales factorially as $n!$, making exploration dramatically harder. In this work, we revisit the efficacy of vanilla BO in high-dimensional permutation spaces and demonstrate that its generality is fundamentally limited by the concentration of measure and topological ruggedness inherent to factorial search spaces. These pathologies cause global kernel matrices to become nearly

uninformative because most candidates appear equidistant to the observed data, a phenomenon we define as Geometric Collapse.

Contrary to the prevailing trends in continuous optimization, our findings urge a re-evaluation of the "Vanilla Hypothesis" for permutations. We find that dimensionality reduction techniques are not outdated but essential for restoring kernel informativeness. Specifically, we propose VaPBO, a principle-driven framework that performs **va**riable selection over **p**ermutation elements with **BO** to artificially restore variance into the distance distribution. By optimizing over active sub-permutations while employing a structure-aware fill-in strategy, our framework effectively restores kernel discriminative power and consistently surpasses vanilla methods across factorial search spaces. We validate these findings through an extensive empirical suite ranging from classical combinatorial problems, specifically the traveling salesman (Reinelt, 1991), quadratic assignment (Burkard et al., 1997), and traveling thief problems (Bonyadi et al., 2013), to critical but challenging real-world applications, including industrial chip placement from electronic design automation (EDA) (Mirhoseini et al., 2021) and biodiversity history reconstruction from AI-for-Science (Fan et al., 2020; Tang et al., 2024; Lu et al., 2026). These results underscore that dimensionality reduction remains a structural necessity for intractable high-dimensional tasks, yielding substantial improvements in both sample efficiency and optimization quality.

Our contributions are summarized as follows:

- **Diagnosis**: We identify two high-dimensional pathologies in $\mathbb{S}_n$, specifically distance concentration and landscape ruggedness, that make global kernel-based BO uninformative and cause the "Vanilla BO is sufficient" hypothesis to fail in permutation spaces.
- **Methodology**: We show that dimensionality reduction via variable selection is a structural necessity in this domain. We propose a principle-driven reduction framework that restores kernel discriminative power against factorial search-space growth.
- **Application**: We empirically validate the framework on important tasks derived from EDA and AI-for-Science, where it significantly outperforms strong vanilla baselines.

2. Background

This section formalizes Bayesian optimization in permutation space and reviews the Mallows and Merge kernels as the architectural foundations of permutation modeling. We then contextualize these methods within the recent shift toward "Vanilla BO" in the high-dimensional BO tasks.

2.1. Bayesian Optimization in Permutation Spaces

Let $[n] = \{1, \dots, n\}$ denote n items. A permutation can be represented as a ranking vector $\pi \in \mathbb{S}_n$, where $\pi(i)$ is the position (rank) of item i. The search space is the symmetric group $\mathbb{S}_n$ with cardinality $|\mathbb{S}_n| = n!$. We consider expensive black-box functions over permutations $f : \mathbb{S}_n \to \mathbb{R}$.

Bayesian Optimization on $\mathbb{S}_n$. BO over permutations runs as in the continuous case, but replaces Euclidean kernels with permutation ones. Given observations $\mathcal{D}_t = \{(\pi_i, y_i)\}_{i=1}^t$ with $y_i = f(\pi_i) + \epsilon_i$, a GP surrogate $\mathcal{GP}(\mu_0, K(\pi, \pi'))$ is fitted to obtain a posterior mean $\mu_t(\pi)$ and uncertainty $\sigma_t(\pi)$. An acquisition function $\alpha_t(\pi)$ such as expected improvement (EI) (Mockus, 1994; Jones et al., 1998) is then maximized over $\mathbb{S}_n$ to propose the next evaluation point. Designing the kernel function K is the main modeling choice of GP. Continuous relaxation (CR) offers a simple and direct approach for permutation optimization, utilizing general-purpose BO to search within a continuous unit hypercube and inducing discrete permutations via the *argsort* operator. While it is easy to implement, CR inherently sacrifices the essential algebraic and structural information of the symmetric group by treating permutations as unstructured Euclidean vectors. Recently, several permutation-specific kernels have been proposed.

- **Mallows kernel.** Mallows kernel adapts the RBF framework to the symmetric group via the Kendall-tau distance d_{KT} (Jiao & Vert, 2015; Deshwal et al., 2022). This distance counts the number of discordant pairs between two rankings. We can represent this kernel through an implicit feature mapping $\Phi_{Mal}(\pi) \in \{0,1\}^{n(n-1)/2}$, where each element records the outcome of a pairwise comparison between items i and j:

$$K_{\text{Mallows}}(\pi, \pi') = \exp\big(-\lambda d_{\text{KT}}(\pi, \pi')\big) = \exp\left(-\frac{\|\Phi_{\text{Mal}}(\pi) - \Phi_{\text{Mal}}(\pi')\|_2^2}{2\ell^2}\right).$$

  While mathematically robust, the Mallows kernel is plagued by an $O(n^2)$ feature dimensionality. In high-dimensional regimes ($n > 100$), this quadratic growth imposes a prohibitive computational overhead.

- **Merge kernel.** To circumvent the quadratic bottleneck, the Merge kernel (Xie & Chen, 2025) leverages the divide-and-conquer logic of merge sort to produce a structured scalable representation $\Phi_{\text{Mer}}(\pi) \in \{0,1\}^{\Theta(n \log n)}$. This construction achieves the information-theoretic lower bound for lossless permutation encoding, providing a scalable alternative that significantly improves convergence speed in intractable search spaces. Detailed formulations of the permutation distance and calculation of the Merge kernel are provided in Appendix A.1 and A.2, respectively.

However, these kernels still model the full $\mathbb{S}_n$ globally. Our main claim is that (as shown in Section 3.1), in high dimensions, global modeling can become unreliable due to distance concentration and rugged landscapes, motivating principled dimensionality reduction.

2.2. High-Dimensional Bayesian Optimization: Dimensionality Reduction and the Vanilla Paradigm Shift

To navigate the curse of dimensionality, high-dimensional BO has traditionally relied on imposing structural priors to restrict the search space. These methodologies typically fall into three categories: (i) Additive models (Kandasamy et al., 2015; Rolland et al., 2018; Mutný & Krause, 2018), which decompose the objective into a sum of lower-dimensional components; (ii) Subspace embeddings (Wang et al., 2016; Letham et al., 2020; Papenmeier et al., 2022), which project the high-dimensional input onto a latent manifold; and (iii) Axis-aligned variable selection (Li et al., 2017; Eriksson & Jankowiak, 2021; Song et al., 2022), which identifies and prioritizes a sparse subset of influential dimensions.

However, this reliance on structural complexity has been recently challenged by the "Vanilla Hypothesis". Recent advances demonstrate that the perceived failure of standard GP in high dimensions is often a consequence of poor hyperparameter configuration rather than an inherent modeling limitation (Hvarfner et al., 2024; Xu et al., 2025). By simply employing dimension-aware length-scale scaling ($l \propto \sqrt{D}$), a "Vanilla" GP can match or even exceed the performance of specialized architectures. MSR (Papenmeier et al., 2025b) further provides a nuanced diagnosis, suggesting that many structural high-dimensional BO methods offer marginal gains primarily because they implicitly improve hyperparameter optimization or local search dynamics.

This emerging consensus of "Vanilla Hypothesis" prompts a critical re-evaluation of the high-dimensional BO landscape. While structural reduction may indeed be redundant in continuous domains, we argue that this narrative overlooks the fundamentally different topological pathologies of permutation spaces. As we will demonstrate, the factorial growth and geometric collapse inherent to $\mathbb{S}_n$ render vanilla global modeling uninformative, making dimensionality reduction not a redundant heuristic, but a structural necessity.

3. Principle-Driven Approach for High-Dimensional Permutation Spaces

We first present an empirical diagnosis in Section 3.1, identifying two fundamental pathologies, i.e., geometric concentration of measure and topological ruggedness, which cause standard GP surrogates to fail as dimensionality increases. Motivated by these failures, Section 3.2 introduces VaPBO, a principle-driven framework that restores kernel informativeness. We detail how VaPBO leverages sub-permutation variable selection to restore variance and employs a structure-aware fill-in strategy to navigate the permutation spaces effectively.

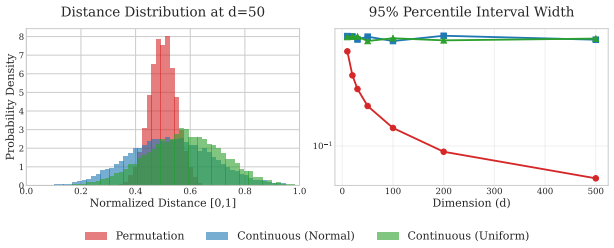

Figure 1. Empirical diagnosis of geometric collapse and concentration of measure in permutation spaces. (**Left**) At a representative dimension of $d = 50$, the normalized distance distribution between random pairs in $\mathbb{S}_n$ (red) is significantly more concentrated around the mean than in Gaussian (blue) or Uniform (green) continuous spaces. This concentration phenomenon renders global kernels uninformative, as most candidates appear nearly equidistant to the observed data. (**Right**) The 95% percentile interval width of normalized distances exhibits a sharp, factorial-rate decay as dimensionality n increases. This "Geometric Collapse" indicates that standard kernels fail to maintain discriminative power in high dimensions, providing the fundamental motivation for the structural dimensionality reduction.

3.1. Analyzing the Properties of High-Dimensional Permutation Spaces

Recent studies in continuous domains suggest that Vanilla BO can achieve state-of-the-art performance in high dimensions without subspace techniques. In this section, we investigate whether this "Vanilla Hypothesis" holds for the permutation space $\mathbb{S}_n$. Through the empirical analysis, we identify two fundamental pathologies, i.e., geometric concentration and topological ruggedness, that render global permutation kernels uninformative as n increases.

The Curse of Dimensionality: Geometric Collapse of the Permutation Space. Unlike continuous search spaces where the volume grows polynomially as c^d, the cardinality of the symmetric group $\mathbb{S}_n$ scales factorially ($n! \gg c^d$), creating a fundamentally more sparse and complex optimization landscape. In high-dimensional regimes, this complexity is further compounded by the Concentration of Measure phenomenon, which dictates that the distances between randomly sampled permutations converge tightly around a central mean with vanishing variance as n increases. This phenomenon is empirically diagnosed in Figure 1. At a relatively low dimension of $d = 50$ (Left), the normalized distance distribution for permutations is significantly more peaked than its continuous counterparts (Uniform or Gaussian), indicating that the vast majority of the space

Table 1. Comparison of Negative Log-Likelihood (NLL) results for different methods after 200 iterations. Lower values indicate better model calibration and uncertainty estimation. ID denotes the results for the in-distribution test set, while OOD represents the results for out-of-distribution points with better actual quality.

Metric	CR	Mallows	Merge
`Chip Placement` (ID)	0.72	1.52	1.51
`Biodiversity` (ID)	0.54	1.51	1.51
`Chip Placement` (OOD)	1015.7	221.78	840.08
`Biodiversity` (OOD)	1251.32	854.67	1035.9

resides at a nearly constant distance from any reference point. More critically, as shown in the right sub-figure, the 95% percentile interval width of these distances exhibits a sharp, factorial-rate decay, which is a pathology we define as geometric collapse. This collapse implies that in high dimensions, any unobserved candidate π^* appears nearly equidistant to all previously observed points in the dataset.

Impact on GP Modeling From a modeling perspective, geometric collapse poses a catastrophic challenge for global permutation kernels. When distances concentrate, the resulting Gram matrix tends toward an identity or a constant matrix, effectively stripping the GP of its ability to distinguish between near and far regions. Consequently, the surrogate model fails to propagate information from the training data to unobserved regions, causing the posterior mean to remain flat and the acquisition function to become uninformative.

Topological Ruggedness and Surrogate Fragility. The issues of geometric collapse induce significant topological ruggedness, thereby posing a formidable challenge for GP regression when applying BO to high-dimensional permutation tasks. To evaluate model performance, we utilize Negative Log-Likelihood (NLL) as the primary metric, as it accounts for both predictive accuracy and uncertainty quantification (Deshwal et al., 2022; Papenmeier et al., 2025b), which is a crucial balance for effective exploration and exploitation in BO. To further comprehensively assess the surrogate models' capacity to capture the underlying structure of high-dimensional permutation spaces, we evaluate their performance on two distinct test sets: an In-Distribution (ID) set, mirroring the training data distribution, and an Out-of-Distribution (OOD) set comprising points with superior objective quality. This experimental design, a standard practice in offline model-based optimization (Kim et al., 2025; Trabucco et al., 2022), is crucial for measuring a model's predictive ability on unseen, high-performing candidates.

The results in Table 1 reveal that while Continuous Relaxation (CR) achieves the lowest NLL on ID data, it has the worst NLL on OOD data. The low ID-NLL of CR is a deceptive indicator of performance, as it originates from overfitting to structured noise rather than capturing the underlying combinatorial topology. In contrast, permutation-aware Mallows and Merge kernels exhibit superior OOD robustness. They provide a compact and structure-aware representation, effectively avoiding the structural inconsistencies of continuous relaxation and achieving superior optimization performance. Despite these specialized kernels' improvements, the overall high NLL values across all methods underscore the persistent challenges of fitting accurate surrogates in high-dimensional permutation spaces.

Kernel Selection in High-Dimensional Permutation Spaces. In high-dimensional permutation spaces, the choice of kernel is governed by a critical trade-off between the ability to capture global structural invariance and the search efficiency driven by the compactness of the feature space. While the Mallows (Jiao & Vert, 2015; Deshwal et al., 2022) kernels offer robust distance metrics, their $O(n^2)$ feature dimensionality leads to significant computational and statistical redundancy as the number of objects n grows, severely limiting their sample efficiency in high-dimensional regimes. In this work, the recently proposed Merge kernel (Xie & Chen, 2025) is chosen because it leverages the divide-and-conquer logic of merge sort to produce a structured representation that matches the information-theoretic lower bound of $\Theta(n \log n)$ for lossless permutation encoding (Knuth, 1998). The construction of the Merge kernel provides a scalable approach and significantly outperforms both Mallows kernels and continuous relaxation methods in both final regret and convergence speed.

Discussions on the Loss of Right-Invariance of Merge Kernel. Although the Merge kernel offers compact feature dimensions and computational efficiency, it sacrifices a crucial feature of traditional permutation kernels, i.e., the right-invariance. A kernel K on $\mathbb{S}_n$ is *right-invariant* if $K(\pi, \pi') = K(\pi \circ \sigma, \pi' \circ \sigma)$ for any $\sigma \in \mathbb{S}_n$, meaning that the similarity depends only on the *relative ordering* (i.e., relabeling items does not change distances). This property is desirable for classical permutation optimization where the objective is defined purely over pairwise order relations, where traditional distance-based kernels such as Mallows satisfy it (Jiao & Vert, 2015). However, we argue that right invariance is not a necessary property for practical problems. In many physical and scientific domains, the absolute index of a permutation element carries intrinsic semantic weight that purely relative metrics fail to resolve. Take, for instance, the biodiversity optimization task from AI-for-Science. In this regime, a permutation $\pi \in \mathbb{S}_{2n}$ represents the true temporal sequence of global biological events, such as the first and last appearance datums of various species. In such contexts, right-invariance is not a virtue but a limitation, as it treats all positions as S_n-equivalent. The Merge kernel, by anchoring its feature map to a fixed merge-sort path, implicitly captures these absolute positional sensitivities.

3.2. Restoring Informative GP via Dimensionality Reduction

As illustrated in the right sub-figure of Figure 1, the changing rate of concentration in permutation spaces is significantly steeper than that in continuous spaces, suggesting that dimensionality reduction is potentially more effective for permutations. To mitigate the pathologies identified in Section 3.1, i.e., the concentration of distance measures and topological ruggedness, we propose VaPBO (Variable Selection of Permutation with Bayesian Optimization). We argue that applying recent vanilla techniques directly to the full permutation space yields limited improvements, as they cannot fundamentally counteract the intrinsic concentration of measure. Therefore, to restore the discriminative power of the GP, it is imperative to actively project the high-dimensional $\mathbb{S}_n$ onto a lower-dimensional subspace where the kernel's distance distribution regains sufficient variance.

However, performing such a reduction on permutation structures is non-trivial and necessitates specialized architectural designs. Unlike continuous domains where axes are independent, dimensionality reduction in permutation spaces is non-trivial due to the strict combinatorial constraints (i.e., the bijectivity of the permutation). VaPBO addresses this via a dropout-style variable selection approach (Li et al., 2017) tailored for $\mathbb{S}_n$, integrated with a Structure-Aware Fill-in strategy to handle inactive dimensions. Detailed algorithm flow of VaPBO is provided in Appendix A.3.

Permutation Subspace Definition. We define the variable selection mechanism on the ranking representation. Let $\pi \in \mathbb{S}_n$ be the permutation vector where $\pi(i)$ denotes the position of item i. In each BO iteration, we randomly select a subset of active items $\mathcal{I} \subset \{1, \dots, n\}$ with cardinality $|\mathcal{I}| = n_a < n$. The optimization is then performed on the subspace configuration $\mathbf{z} \in \{1, \dots, n\}^{n_a}$, which represents the positions of the active items:

$$\mathbf{z} = \{\pi(i) \mid i \in \mathcal{I}\}.$$

Crucially, while standard variable selection in Euclidean space fixes inactive dimensions to a seen value, fixing variables in $\mathbb{S}_n$ is constrained by the mutual exclusion of ranks. The positions assigned to $\mathcal{I}$ dictate the available slots for the inactive set $\mathcal{I}^c$, creating a dependency that must be managed by the fill-in strategy.

Subspace Modeling with Merge Kernel. To enable sample-efficient learning, we construct the surrogate model strictly over the active subspace. For any two permutations π, π', their similarity is computed based solely on the active items $\mathcal{I}$. We employ the Merge Kernel on the sub-permutation features:

$$K_{sub}(\pi, \pi'; \mathcal{I}) = K_{\text{Merge}}(\Phi(\pi)_{\mathcal{I}}, \Phi(\pi')_{\mathcal{I}}),$$

where $\Phi(\cdot)_{\mathcal{I}}$ extracts the merge-sort comparison features relevant only to the items in $\mathcal{I}$. By restricting the kernel to n_a dimensions, we artificially inject variance into the distance distribution, preventing the Gram matrix from collapsing to the identity, which is a necessary condition for effective posterior inference in high-dimensional settings.

Structure-Aware Fill-in Strategies. A core challenge in dimensionality reduction-based permutation BO is determining the configuration of the *inactive items* $\mathcal{I}^c$. Since the objective function $f(\pi)$ requires a full permutation for evaluation, we employ a fill-in strategy Ψ that maps a subspace configuration $\mathbf{z}$ to a valid full permutation $\pi = \Psi(\mathbf{z}; \mathcal{H}_t)$, potentially utilizing history $\mathcal{H}_t$. A naive approach is to assign the items in $\mathcal{I}^c$ to the remaining available positions uniformly at random. While this ensures unbiased exploration, it disregards the high degree of conditional dependence between variables. To combat landscape ruggedness and exploit known high-quality substructures, we introduce a Best-k Position fill-in strategy. We maintain a set of the top-k best solutions found so far, denoted as $\mathcal{P}_k^* = \{\pi_1^*, \dots, \pi_k^*\}$. For a candidate subspace configuration $\mathbf{z}$ (defining positions for items $\mathcal{I}$), we select a reference template $\pi_{ref} \in \mathcal{P}_k^*$ that minimizes the positional distance in the active subspace:

$$\pi_{ref} = \arg\min_{\pi \in \mathcal{P}_k^*} \sum_{i \in \mathcal{I}} (\pi(i) - \mathbf{z}_i)^2.$$

The full permutation is constructed by fixing the active items according to $\mathbf{z}$ and filling the remaining positions using the relative ordering of $\mathcal{I}^c$ found in π_{ref}. This acts as a data-driven soft trust region, stabilizing the optimization by ensuring that unoptimized dimensions adhere to structures known to yield low objective values.

Acquisition Optimization. Optimizing the acquisition function $\alpha(\pi)$ over the combinatorial subspace $z \in \{1, \dots, n\}^{n_a}$ remains a challenge. However, the reduced dimensionality n_a allows us to employ more sophisticated search strategies that would be computationally prohibitive in the full permutation space. As suggested in previous studies such as RAASP (Eriksson et al., 2019; Papenmeier et al., 2025b), we follow the enhanced local search strategy and adapt it for our dimensionality reduction framework. For our acquisition optimization, we employ a multi-restart strategy starting from the current best observed points and random initializations to avoid local optima. The detailed three specialized optimization strategies within the VaPBO framework are shown as follows:

- **Local Search** (as default strategy): Following the intuition of RAASP, we employ an enhanced local search strategy that explores the neighborhood of candidate solutions via swap operations. Crucially, the acquisition value for any candidate is evaluated by projecting it onto the

active subspace $\mathcal{I}$ and querying the low-dimensional GP model.

- **Evolutionary Algorithm (EA)**: To facilitate broader exploration and escape from narrow local basins, we implement a population-based variant (VaPBO-EA). This solver extends the local search foundation by maintaining a diverse population of sub-permutations and introducing structured crossover operators, which recombine high-quality substructures to effectively navigate disjoint regions of the acquisition landscape.
- **Gradient-Based Optimization (Grad)**: Leveraging the structured representation of the Merge kernel, we implement a gradient-assisted solver (VaPBO-Grad). This method utilizes gradients calculated in the feature space as a proxy to guide the search, substituting discrete local edits with direction-aware updates. Similar to our LS approach, this solver also employs a multi-start strategy to ensure robustness against the inherent non-convexity of the acquisition surface.

The detailed analysis of their behaviors is provided in Appendix B due to space limitations.

4. Experiments

In this section, we provide a comprehensive empirical validation of our methods. Our experiments are designed to systematically address the question: can structural dimensionality reduction restore the optimization power that global vanilla BO loses in high-dimensional permutation spaces?

4.1. Settings and Tasks.

Compared methods. We consider the following approaches in our experiments.

- Merge-BO (Xie & Chen, 2025). BO with Merge kernels, which is a recent state-of-the-art method in high-dimensional permutation spaces, demonstrates superior performance than TurBO (Eriksson et al., 2019) (handling permutations by continuous relaxation) and BOPS (Deshwal et al., 2022) (previous advanced BO method in permutation spaces). We use its official implementation[1].
- Vanilla Merge-BO. It is Merge-BO with recently proposed advanced techniques. We use the implementation of Vanilla BO in Xu et al. (2025)[2].
- VaPBO, VaPBO-EA, and VaPBO-Grad. VaPBO with different acquisition optimization strategies. For all variants of VaPBO, we default to setting the active dimension n_a to 50% of the total dimension n. We also conduct additional sensitivity experiments concerning n_a.

We implement all BO loops using the BoTorch[3] framework (Balandat et al., 2020). For the GP surrogates, we utilize Maximum Likelihood Estimation (MLE) to optimize hyperparameters, as recent evidence suggests it mitigates gradient vanishing in high-dimensional settings (Papenmeier et al., 2025b). While Vanilla-Merge employs LogEI, all other methods use EI as the acquisition function. The total evaluation budget is tailored to the complexity of the tasks: 50 evaluations for combinatorial benchmarks and 80 evaluations for the more complex EDA and AI-for-Science tasks. Data is initialized using an initial design of 5 random permutations for combinatorial benchmarks and 10 random permutations for complex applications, respectively. Detailed settings and configurations can be found in our supplementary files.

Tasks. We evaluate on a suite of permutation BO benchmarks spanning both controlled combinatorial problems and large-scale real applications (Deshwal et al., 2022; Xie & Chen, 2025). We primarily emphasize performance on the latter two categories (EDA and AI-for-science), since they reflect the real industrial and scientific regimes where scalable permutation BO is most needed.

Combinatorial Optimization Tasks. We evaluate VaPBO on three representative combinatorial tasks spanning various scales and complexities:

- **Traveling Salesman Problem (TSP).** TSP is a canonical benchmark for route planning and sequence optimization (Reinelt, 1991). We seek the shortest Hamiltonian cycle across n cities, utilizing TSPLIB instances derived from industrial electronic drilling applications.
- **Quadratic Assignment Problem (QAP).** QAP is a classic facility location problem (Burkard et al., 1997). This task involves mapping n facilities to locations to minimize the total cost defined by the product of flow and distance.
- **Traveling Thief Problem (TTP).** TTP is a hybrid NP-hard challenge coupling TSP with the Knapsack Problem (Bonyadi et al., 2013). The objective is non-linearly coupled via weight-dependent velocity, where item collection increases travel costs.

Industrial Applications: Chip placement in EDA. We introduce chip placement in our experiments, a challenging and representative industrial task from EDA (Mirhoseini et al., 2021).

[1] https://openreview.net/forum?id=7QtKdabBP9

[2] https://github.com/XZT008/Standard-GP-is-all-you-need-for-HDBO

[3] https://github.com/meta-pytorch/botorch/

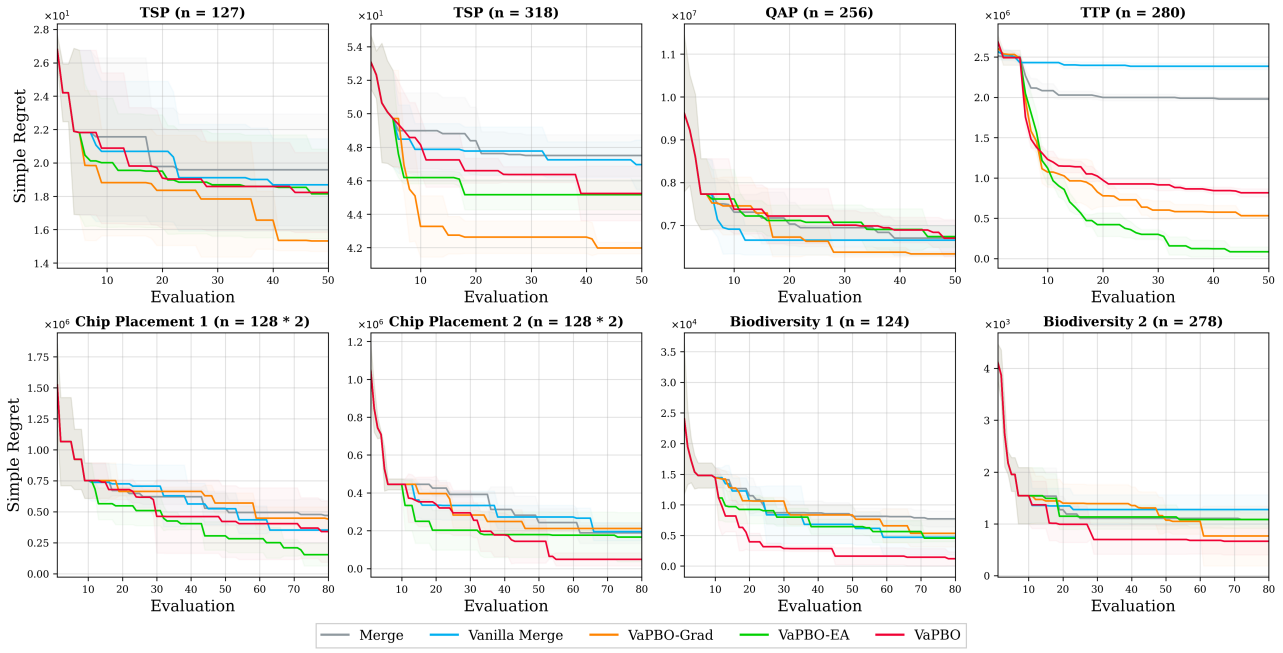

Figure 2. We compare VaPBO to baseline methods on eight high-dimensional combinatorial optimization tasks with five independent runs. For each method, we depict the mean value of the best minimum found at a given iteration. The top row presents four tasks of combinatorial optimization. The bottom row displays two tasks of chip placement and two tasks of biodiversity optimization.

- **Problem Significance**. Chip placement is a critical stage in the physical design of modern integrated circuits. It involves determining the physical locations of various electronic components on a silicon die. The quality of the placement directly impacts key design metrics, including power, performance, and area.

- **Permutation Optimization Formulation.** An important approach for modeling the chip placement task is the Sequence Pair representation (Murata et al., 1996; Oh et al., 2022b; Xue et al., 2025), which represents non-overlapping rectangular electronic components in a 2D chip canvas. Under this formulation, a placement solution is uniquely defined by two distinct permutations of n components: (π_+, π_-), transforming the chip placement problem into a discrete search over a joint permutation space $(\mathbb{S}_n \times \mathbb{S}_n)$. The relative geometric positions of any two blocks (e.g., left-of, right-of, above, or below) are determined by their relative order in these two permutations. This extremely high-dimensional and constrained space serves as a rigorous testbed for evaluating the efficiency of high-dimensional BO over permutations.

AI-for-Science: Biodiversity History Reconstruction. We introduce a novel and challenging benchmark from quantitative paleobiology focused on the high-resolution reconstruction of global biodiversity history across deep time (Fan et al., 2020; Tang et al., 2024; Lu et al., 2026).

- **Problem Significance.** Reconstructing the history of global biodiversity across deep time is fundamental to understanding how life responds to major environmental shifts, such as climate cooling or mass extinctions. However, the fossil record is inherently fragmented; local stratigraphic sections provide only incomplete and biased snapshots of biological history. To synthesize these disparate records into a unified global timeline, the problem is formulated as a large-scale combinatorial optimization task over permutation spaces.

- **Permutation Optimization Formulation**. We consider a set of n biological events, specifically the First Appearance Datums and Last Appearance Datums of various species. The objective is to find a globally optimal permutation $\pi \in \mathbb{S}_{2n}$ that represents the "true" temporal sequence of these events. For a dataset with n species, the optimizer must order $2n$ events, leading to a factorial search space of $(2n)!$. The goal is to minimize a loss function $f(\pi)$ that measures the contradictions between the proposed theoretical sequence and the observed physical evidence in local geological sections. Unlike standard permutations, this task involves critical biological and geological constraints that the optimizer must respect.

4.2. Main Results

Is Vanilla BO Sufficient in High-Dimensional Permutation Spaces? Our primary empirical results, summarized in Figures 2 and 3, provide a definitive answer: No. While

Table 2. Comparison of Negative Log-Likelihood (NLL) results for different methods after 200 iterations. Lower values indicate better model calibration and uncertainty estimation. ID denotes the results for the in-distribution test set, while OOD represents the results for out-of-distribution points with better actual quality.

Metric	Merge	Vanilla Merge	VaPBO
`Chip Placement` (ID)	1.51	52.45	**1.30**
`Biodiversity` (ID)	1.51	77.21	**1.30**
`Chip Placement` (OOD)	840.08	40.08	**5.06**
`Biodiversity` (OOD)	1035.9	52.87	**11.03**

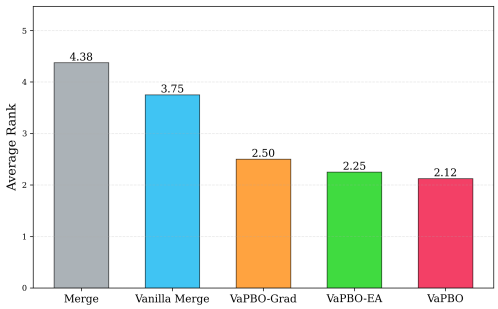

Figure 3. Aggregated average rank comparison.

the Vanilla Merge-BO shows marginal improvements over the baseline Merge-BO, it consistently fails to close the gap with our reduction-based framework. As predicted by our diagnosis in Section 3.1, simple hyperparameter scaling cannot compensate for the Geometric Collapse of the kernel. In the `TTP` ($n = 280$) and `Biodiversity 2` ($n = 278$) tasks, the vanilla surrogate becomes nearly uninformative, leading to stagnant optimization curves.

Superior Performance of VaPBO. Across all eight benchmarks, VaPBO and its variants (EA and Grad) achieve superior sample efficiency and final regret. VaPBO often identifies higher-quality solutions within 20 iterations than vanilla baselines find in 80. As shown in the Average Rank Comparison of Figure 3, VaPBO maintains a consistent lead, validating that dimensionality reduction is not merely an outdated strategy but a structural necessity for restoring GP discriminative power in $\mathbb{S}_n$. Furthermore, the NLL results in Table 2 demonstrate that by projecting onto an active subspace, VaPBO restores model calibration, achieving significantly lower OOD error than global kernels.

Sensitivity Analysis of Active Dimensions n_a. The choice of the active dimension n_a is a critical hyperparameter in VaPBO, as it dictates the degree of dimensionality reduction. Figure 4 illustrates the sensitivity of optimization performance to various ratios of n_a/n. Across all benchmarks, the $n_a/n = 75\%$ configuration performs significantly worse than its more compressed counterparts. This empirically confirms our diagnosis: as n_a remains high, the sub-permutation space still suffers from the Geometric

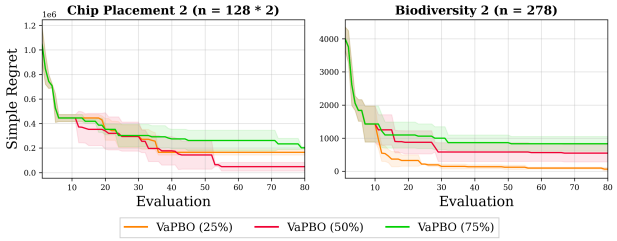

Figure 4. Sensitivity of the subspace dimension n_a. We evaluate the impact of the active dimension ratio (n_a/n) on two tasks. The results show that while moderate-to-aggressive reduction (25% and 50%) consistently outperforms high-dimensional modeling (75%).

Collapse of the distance distribution. The resulting kernel remains relatively uninformative, failing to provide the GP with the necessary contrast to guide the acquisition function effectively. While the optimal n_a/n may vary depending on the intrinsic dimensionality and landscape properties of the task, reduction itself is essential. The fact that even a 25% subspace can match or exceed the performance of larger subspaces validates VaPBO's core design principle: less is more when it comes to kernel-based modeling in $\mathbb{S}_n$.

Analysis of the Acquisition Function Optimization. Using Observation Traveling Salesperson Distance (OTSD) (Papenmeier et al., 2025a) as a metric for investigation, we find that VaPBO with local search behaves similarly to random search. This finding highlights a key distinction from continuous optimization problems. Details are provided in Appendix B due to space limitations.

5. Conclusion

VaPBO uses random active-item selection: it exposes distance collapse but misses objective-informed dependencies. Its gains also rely on a strong local-search shell; the GP improves guidance, yet improved OOD NLL indicates calibration rather than guaranteed online gains. The Merge kernel is not right-invariant, which suits tasks where identities and positions carry semantics but may not suit symmetry-dominated objectives. Five-run corrected tests still leave uncertainty under expensive budgets.

We revisited vanilla BO in high-dimensional permutation spaces and found that continuous-domain conclusions do not transfer automatically. VaPBO combines stochastic subspaces, Merge-kernel modeling, and structure-aware completion. The experiments show that variable selection and local search drive most gains, with surrogate-guided acquisition improving the best high-dimensional rank. Overall, permutation BO requires joint attention to geometry, calibration, and search operators.

Impact Statement

This paper presents work whose goal is to advance the field of Bayesian optimization in high-dimensional permutation spaces. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

Acknowledgments

The authors thank anonymous reviewers for their insightful and valuable comments. This work was supported by the National Science and Technology Major Project (2022ZD0116600), the National Science Foundation of China (624B2069), and the Fundamental Research Funds for the Central Universities (14380020).

References

Balandat, M., Karrer, B., Jiang, D. R., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E. BoTorch: A framework for efficient Monte-Carlo Bayesian optimization. In *Advances in Neural Information Processing Systems 33*, pp. 21524–21538, Virtual, 2020.

Binois, M. and Wycoff, N. A survey on high-dimensional Gaussian process modeling with application to Bayesian optimization. *ACM Transactions on Evolutionary Learning and Optimization*, 2(2):1–26, 2022.

Bonyadi, M. R., Michalewicz, Z., and Barone, L. The travelling thief problem: The first step in the transition from theoretical problems to realistic problems. In *2013 IEEE congress on evolutionary computation*, pp. 1037–1044, 2013.

Burkard, R. E., Karisch, S. E., and Rendl, F. QAPLIB - A quadratic assignment problem library. *J. Glob. Optim.*, 10(4):391–403, 1997.

Deshwal, A., Belakaria, S., Doppa, J. R., and Kim, D. H. Bayesian optimization over permutation spaces. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 6515–6523, 2022.

Eriksson, D. and Jankowiak, M. High-dimensional Bayesian optimization with sparse axis-aligned subspaces. In *Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence*, pp. 493–503, Virtual, 2021.

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

Fan, J.-x., Shen, S.-z., Erwin, D. H., Sadler, P. M., MacLeod, N., Cheng, Q.-m., Hou, X.-d., Yang, J., Wang, X.-d., Wang, Y., et al. A high-resolution summary of cambrian to early triassic marine invertebrate biodiversity. *Science*, 367(6475):272–277, 2020.

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

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

Hvarfner, C., Hellsten, E. O., and Nardi, L. Vanilla Bayesian optimization performs great in high dimensions. In *Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria, 2024.

Jiao, Y. and Vert, J.-P. The kendall and mallows kernels for permutations. In *International Conference on Machine Learning*, pp. 1935–1944, 2015.

Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

Kandasamy, K., Schneider, J. G., and Póczos, B. High dimensional Bayesian optimisation and bandits via additive models. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 295–304, Lille, France, 2015.

Kim, M., Gu, J., Yuan, Y., Yun, T., Liu, Z., Bengio, Y., and Chen, C. Offline model-based optimization: Comprehensive review. *arXiv:2503.17286*, 2025.

Knuth, D. E. *The Art of Computer Programming: Sorting and Searching, Volume 3*. Addison-Wesley Professional, 1998.

Letham, B., Calandra, R., Rai, A., and Bakshy, E. Re-examining linear embeddings for high-dimensional Bayesian optimization. In *Advances in Neural Information Processing Systems 33*, pp. 1546–1558, Vancouver, Canada, 2020.

Li, C., Gupta, S., Rana, S., Nguyen, V., Venkatesh, S., and Shilton, A. High dimensional Bayesian optimization using dropout. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 2096–2102, Melbourne, Australia, 2017.

Lu, Z., Xue, K., Deng, Y., et al. Complex marine ecological response during the eocene-oligocene revealed by global foraminiferal record. *Nature Communications*, 17:3954, 2026. doi: 10.1038/s41467-026-70541-w.

Mirhoseini, A., Goldie, A., Yazgan, M., Jiang, J. W., Songhori, E., Wang, S., Lee, Y.-J., Johnson, E., Pathak,

O., Nazi, A., et al. A graph placement methodology for fast chip design. *Nature*, 594(7862):207–212, 2021.

Mockus, J. Application of Bayesian approach to numerical methods of global and stochastic optimization. *Journal of Global Optimization*, 4(4):347–365, 1994.

Murata, H., Fujiyoshi, K., Nakatake, S., and Kajitani, Y. VLSI module placement based on rectangle-packing by the sequence-pair. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 15(12):1518–1524, 1996.

Mutný, M. and Krause, A. Efficient high dimensional Bayesian optimization with additivity and quadrature Fourier features. In *Advances in Neural Information Processing Systems 31*, pp. 9005–9016, Montreal, Canada, 2018.

Oh, C., Bondesan, R., Gavves, E., and Welling, M. Batch Bayesian optimization on permutations using the acquisition weighted kernel. In *Advances in Neural Information Processing Systems*, volume 35, pp. 6843–6858, 2022a.

Oh, C., Bondesan, R., Kianfar, D., Ahmed, R., Khurana, R., Agarwal, P., Lepert, R., Sriram, M., and Welling, M. Bayesian optimization for macro placement. *arXiv:2207.08398*, 2022b.

Papenmeier, L., Nardi, L., and Poloczek, M. Increasing the scope as you learn: Adaptive Bayesian optimization in nested subspaces. In *Advances in Neural Information Processing Systems 35*, New Orleans, LA, 2022.

Papenmeier, L., Cheng, N., Becker, S., and Nardi, L. Exploring exploration in Bayesian optimization. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pp. 3388–3415, Rio de Janeiro, Brazil, 2025a.

Papenmeier, L., Poloczek, M., and Nardi, L. Understanding high-dimensional Bayesian optimization. In *Proceedings of the 42nd International Conference on Machine Learning*, Vancouver, Canada, 2025b.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2006.

Reinelt, G. TSPLIB - A traveling salesman problem library. *INFORMS J. Comput.*, 3(4):376–384, 1991.

Rolland, P., Scarlett, J., Bogunovic, I., and Cevher, V. High-dimensional Bayesian optimization via additive models with overlapping groups. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, pp. 298–307, Playa Blanca, Spain, 2018.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

Song, L., Xue, K., Huang, X., and Qian, C. Monte Carlo tree search based variable selection for high dimensional Bayesian optimization. In *Advances in Neural Information Processing Systems 35*, New Orleans, LA, 2022.

Tang, Q., Zheng, W., Zhang, S., Fan, J., Riedman, L. A., Hou, X., Muscente, A., Bykova, N., Sadler, P. M., Wang, X., et al. Quantifying the global biodiversity of proterozoic eukaryotes. *Science*, 386(6728):eadm9137, 2024.

Trabucco, B., Geng, X., Kumar, A., and Levine, S. Design-Bench: Benchmarks for data-driven offline model-based optimization. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 21658–21676, Baltimore, MD, 2022.

Wang, Z., Hutter, F., Zoghi, M., Matheson, D., and de Feitas, N. Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55(1):361–387, 2016.

Xie, Z. and Chen, L. From sorting algorithms to scalable kernels: Bayesian optimization in high-dimensional permutation spaces. *arXiv:2507.13263*, 2025.

Xu, Z., Wang, H., Phillips, J. M., and Zhe, S. Standard Gaussian process is all you need for high-dimensional Bayesian optimization. In *Proceedings of the 13th International Conference on Learning Representations*, Singapore, 2025.

Xue, K., Chen, R.-T., Tan, R.-X., Lin, X., Shi, Y., Xu, S., Yuan, M., and Qian, C. Bboplace-bench: Benchmarking black-box optimization for chip placement. *arXiv:2510.23472*, 2025.

Zaefferer, M., Stork, J., and Bartz-Beielstein, T. Distance measures for permutations in combinatorial efficient global optimization. In *International Conference on Parallel Problem Solving from Nature*, pp. 373–383, 2014.

A. Details of methods.

A.1. Distances on $\mathbb{S}_n$.

A key ingredient for modeling permutations is a notion of similarity/distance. Following standard practice in permutation BO (Deshwal et al., 2022; Oh et al., 2022a; Xie & Chen, 2025), we highlight two widely used distances:

(i) Kendall distance counts the number of discordant pairs

$$d_{\mathrm{KT}}(\pi, \pi') \;=\; n_d(\pi, \pi') \;=\; \sum_{1 \le i < j \le n} \mathbb{I}\big[(\pi(i) - \pi(j))(\pi'(i) - \pi'(j)) < 0\big].$$

(ii) Position-based distances measure absolute position discrepancies, e.g., the Spearman footrule

$$d_{\mathrm{pos}}(\pi, \pi') \;=\; \sum_{i=1}^{n} \big|\pi(i) - \pi'(i)\big|.$$

These distances induce different invariances and inductive biases; see Jiao & Vert (2015); Zaefferer et al. (2014) for a systematic discussion.

A.2. Details of the Merge Kernel (Xie & Chen, 2025)

To mitigate the pathologies of geometric collapse and distance concentration in high-dimensional permutation spaces, VaPBO utilizes the Merge Kernel as its primary surrogate component. This section formalizes the construction of the merge-sort-inspired feature mapping $\Phi_{Mer}(\cdot)$ and discusses the structural modifications required to ensure its validity for GP inference, as shown in Algorithm 1.

Theoretical Motivation: Sorting as Feature Generation The fundamental insight behind the Merge Kernel is that any comparison-based sorting algorithm defines a unique mapping from a permutation $\pi \in \mathbb{S}_n$ to a sequence of binary outcomes. By recording these outcomes (i.e., whether a swap or a specific relative ordering was observed), we can generate a feature vector that losslessly encodes the structural information of the permutation.

However, a critical requirement for a kernel function is that the feature mapping must be deterministic and reside in a fixed-dimensional Hilbert space. While standard sorting algorithms are often adaptive (with data-dependent execution paths), the Merge Kernel framework introduces a fixed comparison path. This ensures that for any input, the algorithm traverses a predetermined sequence of comparisons, making the resulting feature vector $\Phi_{Mer}(\pi)$ compatible with standard RBF kernels.

Algorithm and Implementation The mapping $\Phi_{Mer}(\pi)$ is constructed via a recursive divide-and-conquer procedure. Unlike vanilla merge sort, which terminates merging once a sub-list is exhausted, our implementation employs **Fixed-Length Padding**. This structural constraint ensures that the feature dimensionality remains constant at $\Theta(n \log n)$, which corresponds to the information-theoretic lower bound for lossless permutation encoding.

Algorithm 1 Merge Feature Mapping with Fixed Comparison Path

Function MERGEFEATURE(π, n):
Input: Permutation $\pi \in \mathbb{S}_n$, dimension n
Output: Binary feature vector V
if $n = 1$ **then**
  Return: [] // Base case: single element
end if
if $n = 2$ **then**
  Return: $[\mathbb{I}(\pi[0] > \pi[1])]$ // Base case: single comparison
end if
$mid \leftarrow \lfloor n/2 \rfloor$
$V_L \leftarrow$ MergeFeature($\pi[0 : mid], mid$) // Recursive split
$V_R \leftarrow$ MergeFeature($\pi[mid : n], n - mid$)
// Track merge path using sorted sub-sequences
$\hat{\pi}_l \leftarrow$ Sort($\pi[0 : mid]$), $\hat{\pi}_r \leftarrow$ Sort($\pi[mid : n]$)
$V_{\text{Merge}} \leftarrow []$, $i \leftarrow 0$, $j \leftarrow 0$
while $i <$ length($\hat{\pi}_l$) **and** $j <$ length($\hat{\pi}_r$) **do**
  if $\hat{\pi}_l[i] > \hat{\pi}_r[j]$ **then**
    V_{Merge}.append(1); $j \leftarrow j + 1$
  else
    V_{Merge}.append(0); $i \leftarrow i + 1$
  end if
end while
Apply fixed padding to V_{Merge} to ensure length $n - 1$
Return: $V_L \oplus V_R \oplus V_{\text{Merge}}$ // Concatenate features

A.3. Detailed flow of VaPBO

This section provides a detailed procedural walkthrough of the Variable Selection for Permutation Bayesian Optimization (VaPBO) framework as formalized in Algorithm 2.

Stage 1: Dynamic Variable Selection. The algorithm begins by identifying a subset of active items $\mathcal{I} \subset \{1, \ldots, n\}$ with cardinality n_a, where $n_a < n$. This selection is updated periodically every N_{update} iterations to ensure that the optimizer does not become trapped in a stagnant subspace. By focusing only on n_a elements, we effectively project the factorial search space onto a lower-dimensional manifold where the distance distribution retains sufficient variance for effective modeling. Note that our selection of dropout stems from its simplicity, as it avoids introducing complex algorithmic components. It functions as the simplest variable selection algorithm, and in practice, any variable selection algorithm can replace it.

Stage 2: Subspace Projection and Modeling. Once the active set is defined, all previously observed permutations in the dataset $\mathcal{D}_{t-1}$ are projected into their subspace representations $Z_{1:t-1}$. The GP surrogate is then fitted strictly over this active subspace using the Merge Kernel. This construction ensures that the kernel similarity is computed based solely on the structural comparison features relevant to the items in $\mathcal{I}$, thereby preventing the Gram matrix from collapsing into an uninformative identity matrix.

Stage 3: Acquisition Optimization. The acquisition function α_t is maximized exclusively within the combinatorial subspace $z \in \{1, \ldots, n\}^{n_a}$. Although the underlying search space remains discrete, the reduced dimensionality allows for more intensive search strategies. While local search serves as the default optimizer, VaPBO can be extended to utilize evolutionary algorithms or gradient-assisted solvers in the feature space to escape local optima and navigate the rugged acquisition landscape.

Stage 4: Structure-Aware Reconstruction. To evaluate the proposed candidate z_t^* in the objective function f, it must be mapped back to a valid full permutation $\pi_t \in \mathbb{S}_n$. VaPBO employs the Best-k Position fill-in strategy, which identifies a reference template from the top-k historical solutions. By fixing the active items according to z_t^* and filling the remaining slots using the relative ordering found in the template, the algorithm creates a data-driven soft trust region that stabilizes the optimization against landscape ruggedness.

Algorithm 2 VaPBO: Variable Selection of Permutation with Bayesian Optimization

Input: Black-box function $f : \mathbb{S}_n \to \mathbb{R}$, dimension n, active dimension n_a, update frequency n_{update}, fill-in strategy Ψ
Initialize: Data $\mathcal{D}_0 \leftarrow \text{InitialDesign}()$, history $\mathcal{H}_0 \leftarrow \mathcal{D}_0$
for $t = 1, \ldots, T$ **do**
  // Variable Selection
  if $t \pmod{n_{\text{update}}} == 0$ **then**
    Sample active index set $\mathcal{I}_t \subset \{1, \ldots, n\}$ with $|\mathcal{I}_t| = n_a$ uniformly at random
  else
    Keep previous active set $\mathcal{I}_t \leftarrow \mathcal{I}_{t-1}$
  end if
  // Subspace Projection & Modeling
  Project observed permutations: $\mathbf{Z}_{1:t-1} \leftarrow \{\pi(\mathcal{I}_t) \mid (\pi, y) \in \mathcal{D}_{t-1}\}$
  Fit GP surrogate $\mathcal{M}_t$ using $K_{\text{sub}}(\mathbf{z}, \mathbf{z}') = K_{\text{Merge}}(\Phi(\mathbf{z}), \Phi(\mathbf{z}'))$
  // Acquisition Optimization (in Subspace)
  Find candidate: $\mathbf{z}_t^* \leftarrow \arg\max_{\mathbf{z}} \alpha_t(\mathbf{z} \mid \mathcal{M}_t)$ via local search
  // Structure-Aware Fill-in
  Construct full permutation $\pi_t \leftarrow \Psi(\mathbf{z}_t^*; \mathcal{H}_{t-1})$ // Note: Ψ fixes $\mathcal{I}_t$ items and fills $\mathcal{I}_t^c$ using historical bests
  // Evaluation & Update
  Evaluate $y_t \leftarrow f(\pi_t)$ and update $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(\pi_t, y_t)\}$
  Update history/strategy state (e.g., update Top-K list in Ψ)
end for
Return: Best solution found π_{best}

B. Additional Results

Analysis of the acquisition function optimization. To understand the search dynamics that lead to VaPBO's success, we analyze the Observation Traveling Salesperson Distance (OTSD) (Papenmeier et al., 2025a). This metric serves as a proxy for the algorithm's exploration efficiency by measuring the shortest path connecting all evaluated points. Our results in Figure 5 reveal a counter-intuitive phenomenon: VaPBO, while being the most performant, exhibits a sampling diversity that most closely mirrors random search. We believe that this is due to the highly rugged nature of the high-dimensional permutation space. This finding is not entirely consistent with conclusions from some high-dimensional continuous problems, thereby highlighting the unique characteristics and research value of high-dimensional permutation problems.

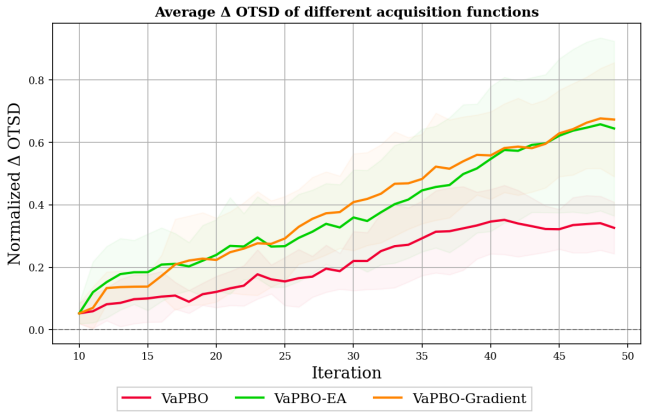

Figure 5. Exploration Dynamics via OTSD. We plot the Normalized Δ OTSD across iterations to characterize the exploration behavior of different acquisition optimizers. Lower values indicate a sampling pattern that more closely resembles the random search.