Batched High-dimensional Bayesian Optimization via Structural Kernel Learning

Florian Klein

University of Cambridge, UK

11 November 2025

Introduction

Introduction
- Motivation

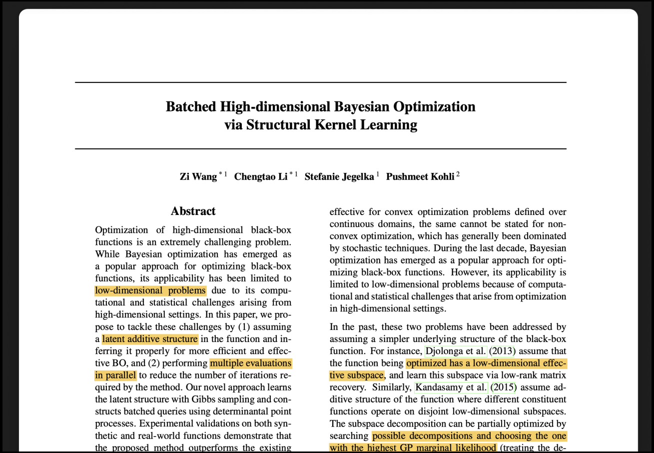
Approach
- Structural Kernel Learning
- Batched Bayesian Optimization

Results

Discussion
- Methodology
- Impact
- Future Work

Conclusion

References



Batched High-dimensional Bayesian Optimization via Structural Kernel Learning

Zi Wang [*1] **Chengtao Li** [*1] **Stefanie Jegelka** [1] **Pushmeet Kohli** [2]

Abstract

Optimization of high-dimensional black-box functions is an extremely challenging problem. While Bayesian optimization has emerged as a popular approach for optimizing black-box functions, its applicability has been limited to low-dimensional problems due to its computational and statistical challenges arising from high-dimensional settings. In this paper, we propose to tackle these challenges by (1) assuming a latent additive structure in the function and inferring it properly for more efficient and effective BO, and (2) performing multiple evaluations in parallel to reduce the number of iterations required by the method. Our novel approach learns the latent structure with Gibbs sampling and constructs batched queries using determinantal point processes. Experimental validations on both synthetic and real-world functions demonstrate that the proposed method outperforms the existing

effective for convex optimization problems defined over continuous domains, the same cannot be stated for non-convex optimization, which has generally been dominated by stochastic techniques. During the last decade, Bayesian optimization has emerged as a popular approach for optimizing black-box functions. However, its applicability is limited to low-dimensional problems because of computational and statistical challenges that arise from optimization in high-dimensional settings.

In the past, these two problems have been addressed by assuming a simpler underlying structure of the black-box function. For instance, Djolonga et al. (2013) assume that the function being optimized has a low-dimensional effective subspace, and learn this subspace via low-rank matrix recovery. Similarly, Kandasamy et al. (2015) assume additive structure of the function where different constituent functions operate on disjoint low-dimensional subspaces. The subspace decomposition can be partially optimized by searching possible decompositions and choosing the one with the highest GP marginal likelihood (treating the de-

Figure: [5]

Motivation: Bayesian Optimization

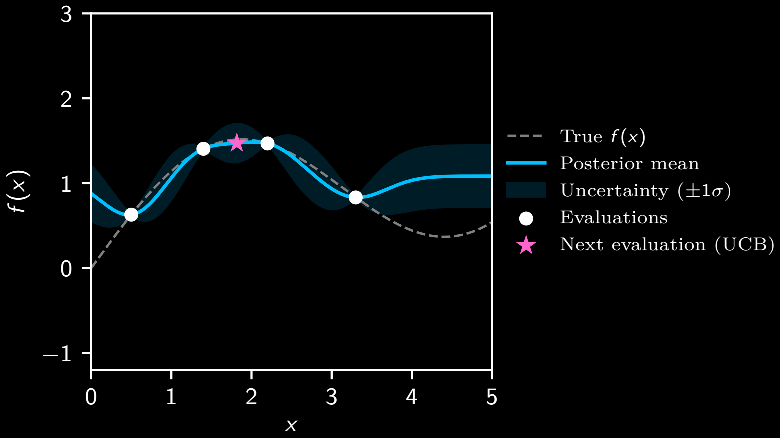
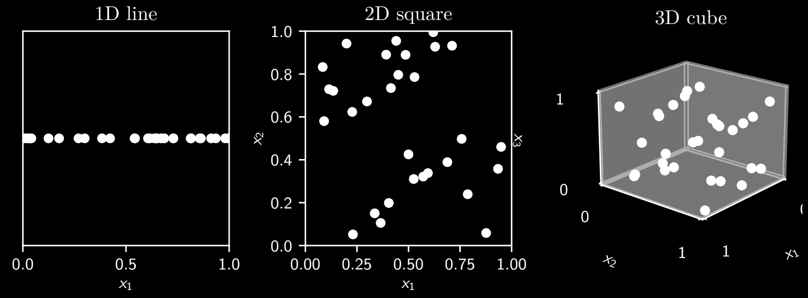


Figure: Illustration of Bayesian optimization in 1D

Motivation: Bayesian Optimization



BO struggles in high-dimensional settings due to:

- Computational challenges (scaling of GP)
- Statistical challenges (sample-inefficient exploration)

Motivation: High-Dimensional Bayesian Optimization

High-dimensional black-box optimization
$f : \mathbb{R}^D \to \mathbb{R}$ (expensive evaluations)

Computational challenge
Standard GP BO scales poorly in high D

Statistical challenge
Sample-inefficient exploration in high D

This paper

Latent additive structure
$f(x) = \sum_m f_m(x_{A_m})$
$\Rightarrow$ groups inferred via Gibbs sampling

Parallel batch BO
Batched queries constructed with determinantal point processes (DPPs)

Scalable high-dimensional Bayesian optimization
Fewer iterations, better performance
on synthetic and real-world tasks

Background: Bayesian Optimization in High Dimensions

Problem

We aim to optimize an expensive black-box function

$$f^* = \max_{x \in \mathcal{X}} f(x), \quad f : \mathcal{X} \subset \mathbb{R}^D \to \mathbb{R},$$

where each evaluation of $f(x)$ is expensive (e.g. experiment, simulation).

Bayesian Optimization

BO places a Gaussian Process prior

$$f \sim \mathcal{GP}(\mu(x), k(x, x')),$$

and selects query points iteratively by maximizing an *acquisition function*

$$x_{t+1} = \arg\max_{x \in \mathcal{X}} a_t(x), \qquad a_t(x) = \mu_t(x) + \beta_t^{1/2} \sigma_t(x)$$

Background: Bayesian Optimization in High Dimensions

Challenge: High Dimensionality.

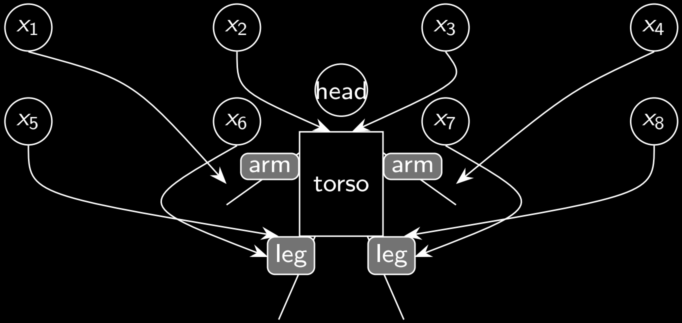
- GP inference is $\mathcal{O}(n^3)$ in number of observations n.
- Posterior variance grows exponentially with D (curse of dimensionality).

However:

Most high-dimensional objectives have *low-dimensional structure*:

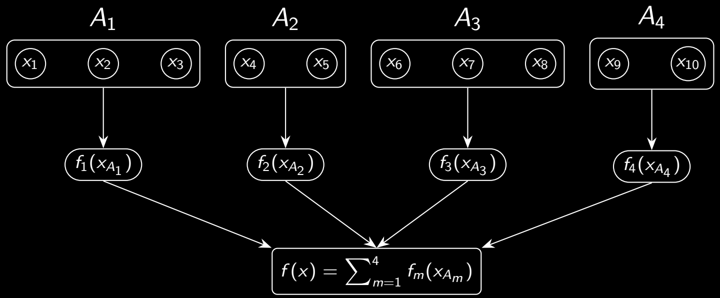
$$f(x) \approx \sum_{m=1}^{M} f_m(x_{A_m}), \qquad A_i \cap A_j = \emptyset.$$

Intuition: Robot Control with Low-dimensional Structure



$$\text{Walking speed} \approx f_{\text{torso}}(x_1, x_2) + f_{\text{arms}}(x_3, x_4) + f_{\text{legs}}(x_5, \ldots, x_8)$$

Background: Low-Dimensional Additive Structure



$$f(x) \approx f_1(x_1, x_2, x_3) + f_2(x_4, x_5) + f_3(x_6, x_7, x_8) + f_4(x_9, x_{10})$$

Background: Prior Approaches to High-Dimensional BO

1. **Additive structure assumption**
   - Function decomposed as

     $$f(x) = \sum_m f_m(x_{A_m}), \quad A_m \text{ disjoint subsets of features}$$

     (Kandasamy *et al.* [3])
   - But assumes the decomposition is known or uses heuristic search
2. **Low-dimensional embeddings**
   - Random embeddings (Wang *et al.* [6])
   - Subspace learning (Djolonga *et al.* [2])
3. **Batch-parallel Bayesian Optimization**
   - Multiple queries per iteration using diversity-promoting methods such as:
     - Determinantal Point Processes (Kathuria *et al.* [4])
   - Still computationally expensive in high dimensions

Gap Addressed and Proposed Solution

- **Main Gap:**
  No existing method *learns the additive structure automatically* **and** supports efficient *parallel (batched)* evaluations.

- **This Paper:**
  - **Structural Kernel Learning (SKL)**:
    Automatically derive additive groups $\{A_m\}$ within the GP kernel via **Gibbs sampling**, no need for prior knowledge of structure:

$$f(x) = \sum_{m=1}^{M} f_m(x_{A_m}), \qquad A_i \cap A_j = \emptyset$$

  - **Batched Bayesian Optimization:**
    Uses **group-wise Determinantal Point Processes (DPPs)** to select query points in parallel, reducing total iterations.

Structural Kernel Learning (SKL): Learning Additive Structure

- **Model assumption:**

$$f(x) = \sum_{m=1}^{M} f_m(x_{A_m}), \quad f_m \sim \mathcal{GP}(0, k^{(m)}), \quad A_i \cap A_j = \emptyset$$

The subsets A_m (feature groups) are **unknown**.

- **Latent decomposition as random variable:**
Each input dimension j is assigned to a group via

$$z_j \sim \mathrm{Categorical}(\theta), \qquad \theta \sim \mathrm{Dirichlet}(\alpha)$$

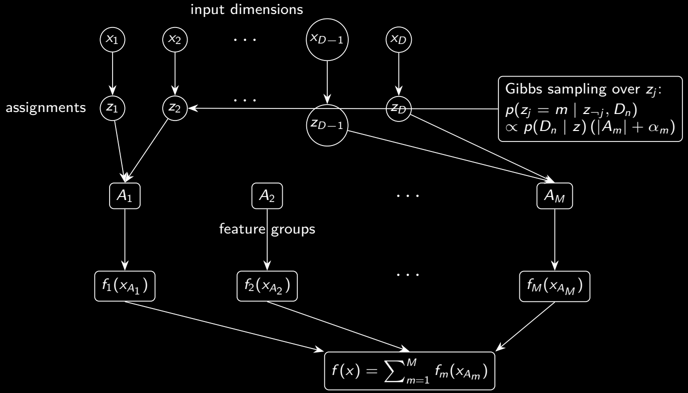
giving $A_m = \{j : z_j = m\}$.

- **Inference via Gibbs sampling** (MCMC method):
Sample each z_j conditional on all others:

$$p(z_j = m \mid z_{\neg j}, D_n) \propto p(D_n \mid z)\,(|A_m| + \alpha_m)$$

where $p(D_n \mid z)$ is the GP marginal likelihood for decomposition z.

Structural Kernel Learning: Learning Additive Structure



What does this mean?

- Posterior over z measures uncertainty in structure.
- The best decomposition is selected by sampling.

Batched Bayesian Optimization via Structured Kernels

- **Key idea.**
  Use the learned additive structure
  $$f(x) = \sum_{m=1}^{M} f_m(x_{A_m}), \quad f_m \sim \mathcal{GP}(0, k^{(m)}),$$
  each component f_m gives diverse candidates in its low-dimensional subspace.
- **Group-wise Diversity.**
  For each group m:
  $$X_t^{(m)} \sim \mathrm{DPP}\left(K_t^{(m)}\right), \quad K_t^{(m)}(x, x') = k_t^{(m)}(x, x'),$$
- **Combining subspaces.**
  Full-dimensional batch points are formed by combining samples across groups:
  $$x_{t,b} = (x_b^{(1)}, x_b^{(2)}, \dots, x_b^{(M)}), \qquad b = 1, \dots, B.$$
  Scales as $O(M)$ DPPs in subspaces instead of one exponential DPP in $\mathbb{R}^D$.

Overview: Empirical Evaluation

- **Goal:** Test if Structural Kernel Learning (SKL)
  $\Rightarrow$ (1) correctly infers additive structure
  $\Rightarrow$ (2) improves high-dimensional BO performance.
- **Benchmarks:**
  - Synthetic additive functions ($D = 2-100$)
  - Real-world robotics tasks (Box2D pushing, Biped Walker)
- **Comparisons:**
  - Known / No / Fully / Partially learned decompositions
  - Existing batch BO methods: PE, DPP, Random

Example: Empirical Evaluation

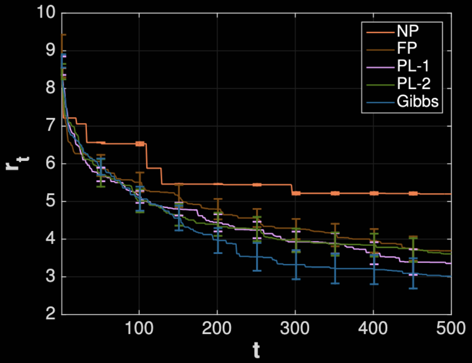


Figure: Simple regret of tuning the 14 parameters for a robot pushing task. Learning decompositions with Gibbs is more effective than partial learning (PL-1, PL-2), no partitions (NP), or fully partitioned (FP). Learning decompositions with Gibbs helps BO to find a better point for this tuning task. [5]

Discussion

1. **Strength:** Strong combination of two research directions — structure learning in Gaussian Processes and batched BO.
2. **But:** Evaluation remains mostly *empirical*.
   No formal convergence or regret guarantees for the batched variant.

Discussion: Methodological Gaps

1. **Limited theoretical analysis**
   - The Gibbs-sampling–based structure learning lacks convergence proofs.
   - No theoretical bounds for SKL or group-wise DPP batching.
2. **Scalability constraints**
   - Experiments reach $D = 100$; unclear performance in even higher dimensions or continuous large-scale domains.

Discussion

What did the paper do well and contribute?

1. **Structure discovery in high dimensions**
   - First Bayesian method to learn additive kernel structure automatically via Gibbs sampling.
2. **Scalable batched optimization**
   - Introduced group-wise DPP sampling — exponential speed-up compared to full DPPs.
3. **Empirical validation across synthetic and robotic tasks**
   - Consistently lowest regret vs. state-of-the-art baselines.

Impact

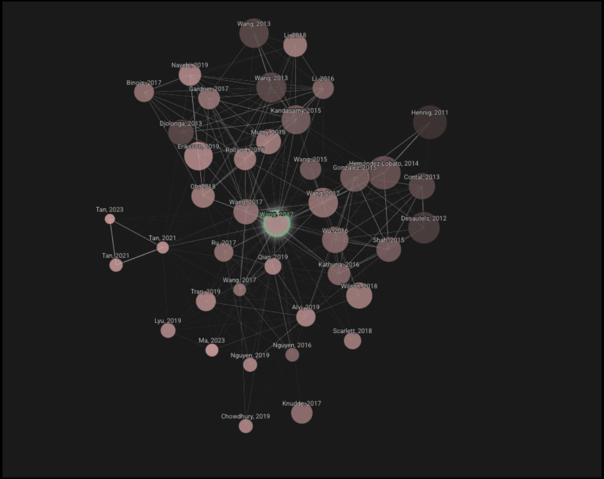


Figure: 164 citations; Connected Papers Graph [1]

Future Work

It might be interesting to explore:

1. **Theoretical regret guarantees** for learned additive structures and batch selection.
2. **Continuous and non-additive latent structures**.
3. **Applications to real-world AutoML tasks** (e.g., hyperparameter tuning, neural architecture search).

Conclusion

To summarize:

- **Structural Kernel Learning** to get additive structure in high-dimensional BO.
- **Group-wise DPP batching** for efficient parallel exploration.
- Empirically achieves state-of-the-art performance!

Thank you for your attention!

Any questions?

References I

Connected Papers | Find and explore academic papers, `https://www.connectedpapers.com/main/342449a88837d2ac86136870a5a9afd480c8fe8f/Batched-High%20dimensional-Bayesian-Optimization-via-Structural-Kernel-Learning/prior`

Djolonga, J., Krause, A., Cevher, V.: High-Dimensional Gaussian Process Bandits

Kandasamy, K., Schneider, J., Poczos, B.: High Dimensional Bayesian Optimisation and Bandits via Additive Models (May 2016). https://doi.org/10.48550/arXiv.1503.01673, `http://arxiv.org/abs/1503.01673`

Kathuria, T., Deshpande, A., Kohli, P.: Batched Gaussian Process Bandit Optimization via Determinantal Point Processes (Nov 2016). https://doi.org/10.48550/arXiv.1611.04088, `http://arxiv.org/abs/1611.04088`

Wang, Z., Li, C., Jegelka, S., Kohli, P.: Batched High-dimensional Bayesian Optimization via Structural Kernel Learning

References II

Wang, Z., Hutter, F., Zoghi, M., Matheson, D., de Freitas, N.: Bayesian Optimization in a Billion Dimensions via Random Embeddings (Jan 2016). https://doi.org/10.48550/arXiv.1301.1942, http://arxiv.org/abs/1301.1942