We Still Don't Understand High-Dimensional Bayesian Optimization

Colin Doumont[1,2], Donney Fan[1,6], Natalie Maus[3], Jacob R. Gardner[3], Henry Moss[4,5], Geoff Pleiss[1,6]

[1]Vector Institute [2]Tübingen AI Center [3]University of Pennsylvania [4]Lancaster University [5]University of Cambridge [6]University of British Columbia

Summary

We match or outperform existing high-dimensional Bayesian optimization (HDBO) methods with arguably the simplest method imaginable: **Bayesian linear regression** (after a spherical projection).

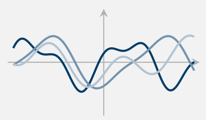
Background: High-dimensional BO

Search space: $\mathcal{X} = [-1, 1]^D \subseteq \mathbb{R}^D$

Well-known kernels in $\mathbb{R}^D$:

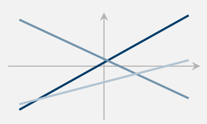
- RBF kernel [∞-dim. basis]:

$$k_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\tfrac{1}{2}\|\mathbf{z} - \mathbf{z}'\|^2\right), \quad \mathbf{z} = \left[\tfrac{x_1}{\ell_1}, \ \cdots, \ \tfrac{x_D}{\ell_D}\right].$$



- Linear kernel [$\mathcal{O}(D)$-dim. basis]:

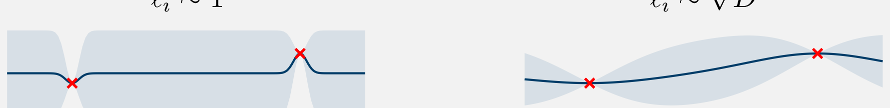
$$k_{\text{lin}}(\mathbf{x}, \mathbf{x}') = b_0 + b_1 \, \mathbf{x}^\top \mathbf{x}'.$$



Solutions to the curse of dimensionality:

1. Last ~10 years: structural assumptions
2. Last ~2 years [1, 2]:
   - non-linear surrogate, but simpler function
   - increase smoothness with $\ell_i \sim \sqrt{D}$
   - equivalent to rescaling $\mathcal{X} = [-1/\sqrt{D}, 1/\sqrt{D}]^D$

$\ell_i \sim 1$ $\qquad$ $\ell_i \sim \sqrt{D}$



What happens if we make the model even simpler?

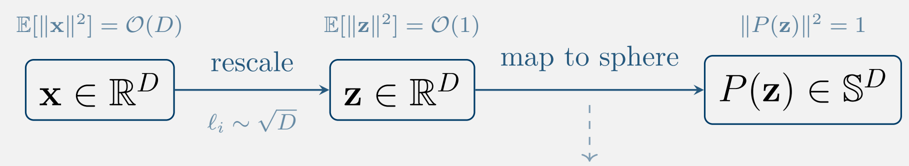
- Smoothness to its logical extreme → linear models
- Would this additional simplicity hurt or hinder high-dimensional BO?

Our Proposed Linear Model

Theorem

Bayesian linear models maximize acquisition on the search-space *boundary* (for EI, UCB, TS, etc.).

Fix: project inputs onto $\mathbb{S}^D$, then apply a linear kernel.



Inverse stereographic projection $P : \mathbb{R}^D \to \mathbb{S}^D$

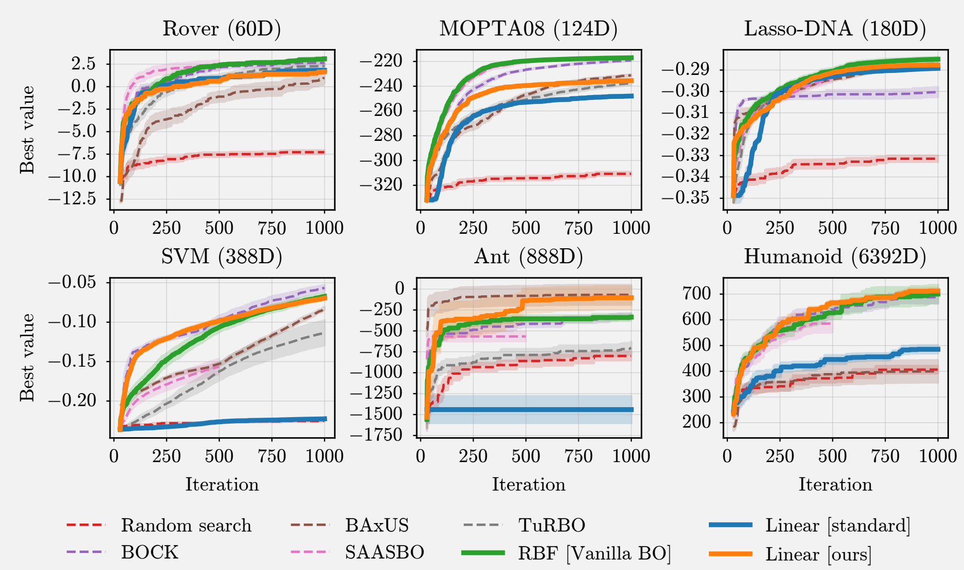
$$P(\mathbf{z}) = \tfrac{1}{\|\mathbf{z}\|^2+1}\left[2z_1, \ \cdots, \ 2z_D, \ \|\mathbf{z}\|^2 - 1\right]$$

$$k_{\text{lin}}(\mathbf{x}, \mathbf{x}') = b_0 + b_1 \, P(\mathbf{z})^\top P(\mathbf{z}')$$

Near identity: $P(\mathbf{z}) \approx [\mathbf{z}; \, 0]$ for typical $\mathbf{z}$ (i.e. $\|\mathbf{z}\|^2 \approx 1$).

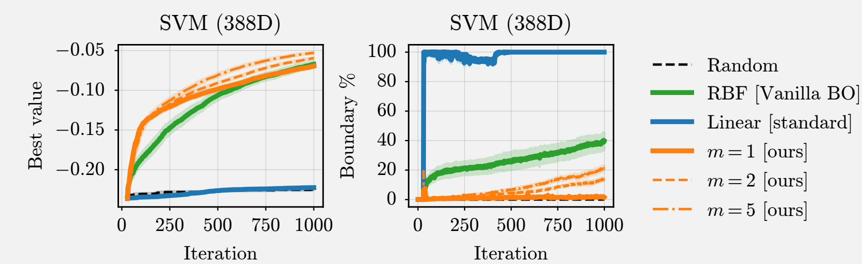
Main Results

Our linear model outperforms sophisticated baselines, and **matches** the performance of Vanilla BO (**SOTA**).

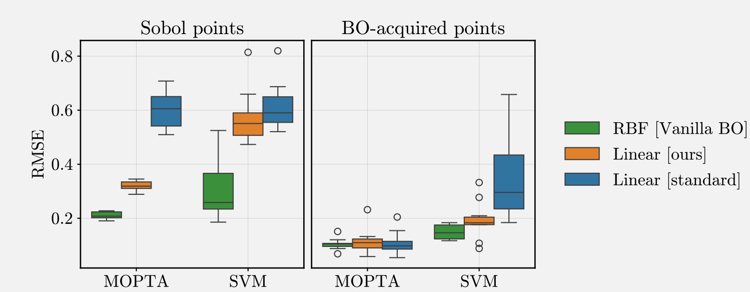


Ablations & Analyses

- **Representational power does not affect BO performance.** Higher-order polynomial extensions ($m > 1$) do not improve much upon our linear model ($m = 1$).
- **Our spherical projection prevents pathological boundary acquisition.** Standard linear models acquire points in corners of $\mathcal{X}$ (100% of dimensions at ±1), while our model acquires non-corner points (≈75% on some problems) or interior points (≈0% on `SVM`).



- **Spherical mappings affect BO performance, but not supervised regression performance.** On quasi-random data, linear models predict worse than RBF (left); however, on adaptive BO acquisitions, they match RBF accuracy (right), suggesting BO might operate in a more local region (where linear modeling is sufficient).



References

[1] C. Hvarfner, E. O. Hellsten, and L. Nardi. Vanilla Bayesian optimization performs great in high dimensions. In *ICML*, 2024.

[2] Z. Xu, H. Wang, J. M. Phillips, and S. Zhe. Standard Gaussian process is all you need for high-dimensional Bayesian optimization. In *ICLR*, 2025.