Relaxing the Additivity Constraints in Decentralized No-Regret High-Dimensional Bayesian Optimization

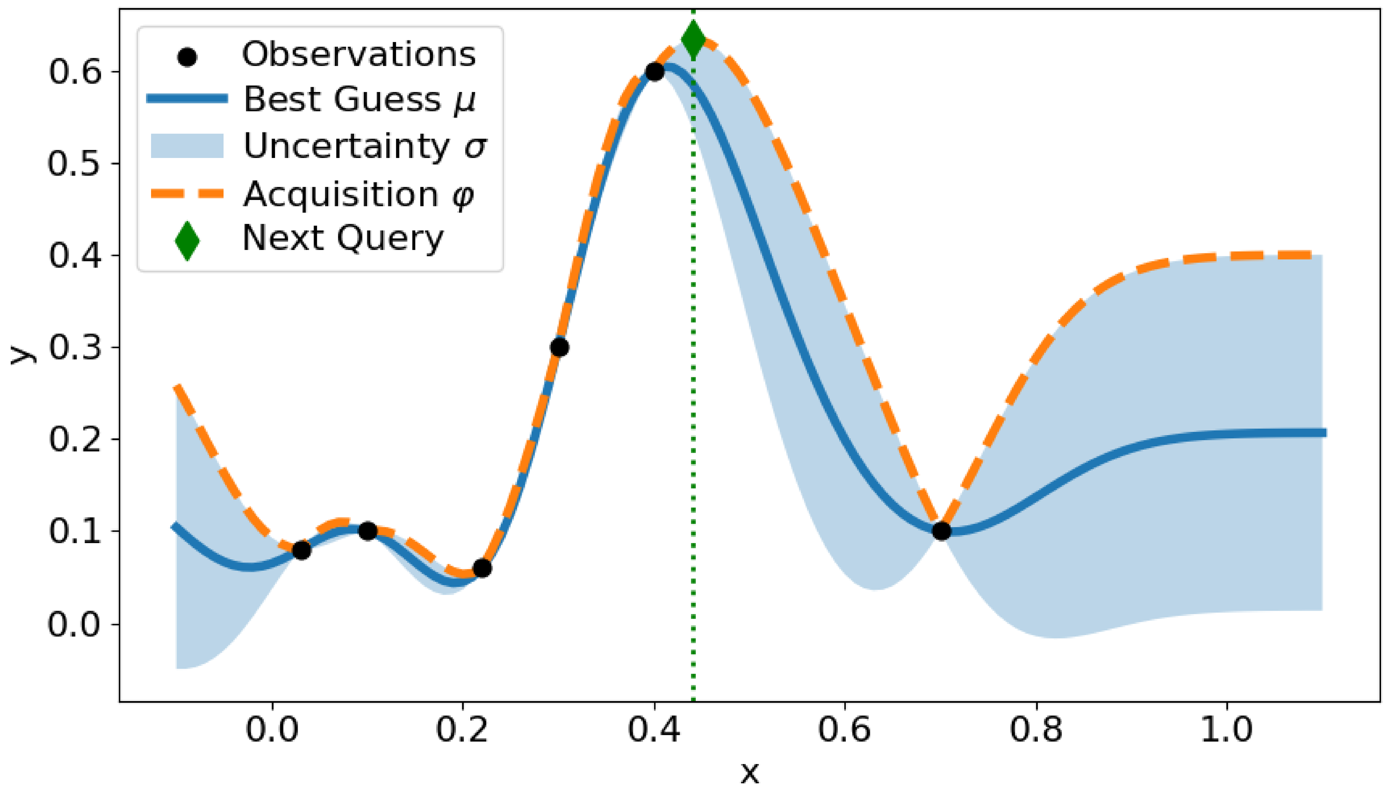
Anthony Bardou[1,2], Patrick Thiran[1], and Thomas Begin[2]

[1]IC, EPFL, Lausanne, Switzerland — [2]ENS Lyon, Université Lyon 1, CNRS, Lyon, France

EPFL · ENS DE LYON · cnrs · Lyon 1 · ICLR

PAPER · CODE

Bayesian Optimization

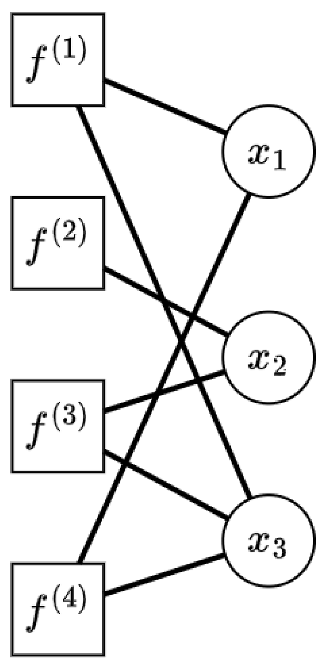


High-Dimensional BO with Additive Structure

The usual algorithms optimizing the acquisition function are **prohibitive in high-dimensional spaces**. To reduce the complexity, one can **assume an additive decomposition** $f = \sum_{i=1}^{n} f^{(i)}$, with each factor $f^{(i)}$ being a $\mathcal{GP}$ with kernel function $k^{(i)}$.

GP-UCB is optimized on each $f^{(i)}$ in parallel, leading to the optimization of $\varphi_t(\boldsymbol{x}) = \mu_t(\boldsymbol{x}) + \beta_t^{1/2} \sum_{i=1}^{n} \sigma_t^{(i)}(\boldsymbol{x})$. This also ensures no-regret performance. However:

- $\sum_{i=1}^{n} \sigma_t^{(i)}(\boldsymbol{x}) \geq \sqrt{\sum_{i=1}^{n} \left(\sigma_t^{(i)}(\boldsymbol{x})\right)^2} = \sigma_t(\boldsymbol{x})$, leading to **over-exploration**
- This strategy is less prohibitive if the additive decomposition is assumed to have **a low Maximum Factor Size (MFS)**



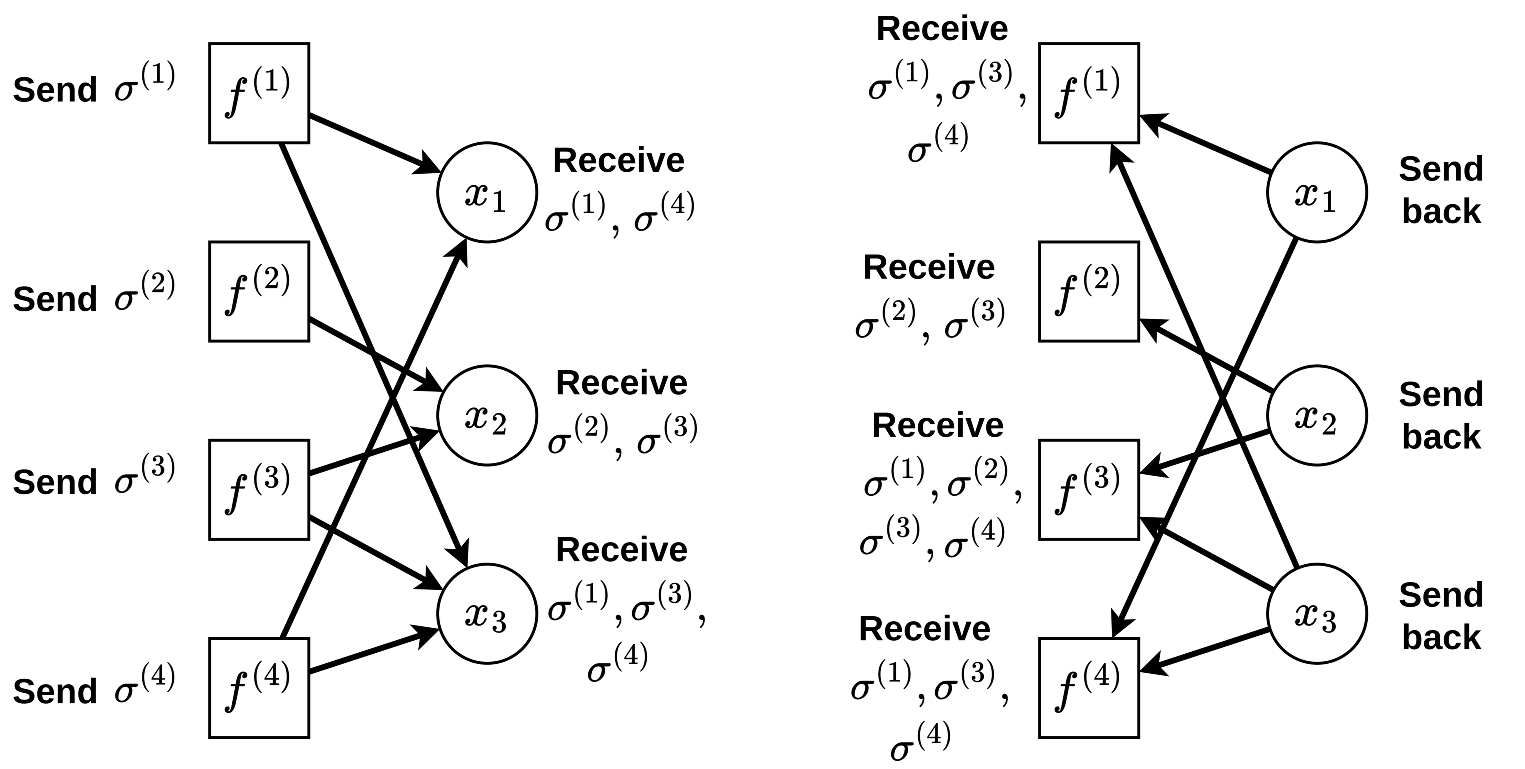
The Big Questions

- Can we address the decentralized GP-UCB over-exploration issue?
- Can we relax the low-MFS assumption in no-regret high-dimensional BO?

A New Acquisition Function

Through **message-passing** in the factor graph, a factor $f^{(i)}$ can access information from the set $\mathcal{N}_i$ of factors that share at least one input dimension with it. We propose

$$\varphi_t(\boldsymbol{x}) = \sum_{i=1}^{n} \varphi_t^{(i)}(\boldsymbol{x}) = \sum_{i=1}^{n} \mu_t^{(i)}(\boldsymbol{x}) + \beta_t^{1/2} \sqrt{\sum_{k \in \mathcal{N}_i} \left(\sigma_t^{(k)}(\boldsymbol{x}) / \mathcal{N}_k\right)^2}.$$



Theorem $\sigma_t(\boldsymbol{x}) \leq \sum_{i=1}^{n} \sqrt{\sum_{k \in \mathcal{N}_i} \left(\sigma_t^{(k)}(\boldsymbol{x}) / \mathcal{N}_k\right)^2} \leq \sum_{i=1}^{n} \sigma_t^{(i)}(\boldsymbol{x})$ for **any factor graph**.

DuMBO

We propose **DuMBO**, an algorithm that optimizes the augmented Lagrangian of φ_t, that is $\mathcal{L}_\eta(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \bar{\boldsymbol{x}}, \boldsymbol{\lambda})$ where

$$\mathcal{L}_\eta = \sum_{i=1}^{n} \varphi_t^{(i)}(\boldsymbol{x}^{(i)}) - \overbrace{\boldsymbol{\lambda}^{(i)\top}}^{\text{dual var. for } \varphi_t^{(i)}} (\boldsymbol{x}^{(i)} - \overbrace{\bar{\boldsymbol{x}}_{\mathrm{dom}\,\varphi_t^{(i)}}}^{\text{consensus var. projected onto dom } \varphi_t^{(i)}}) - \frac{\eta}{2} ||\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}_{\mathrm{dom}\,\varphi_t^{(i)}}||_2^2.$$

DuMBO uses ADMM to optimize $\mathcal{L}_\eta$ in a **decentralized fashion** and **scalably**, even when the additive decomposition **has a large MFS**.

Theorem Under a mild assumption, φ_t is a restricted prox-regular function and $\mathcal{L}_\eta$ satisfies the Kurdyka-Łojasiewicz condition.

Corollary ADMM can **always** find $x_{t+1} = \arg\max_{\boldsymbol{x} \in \mathcal{D}} \varphi_t(\boldsymbol{x})$ [1].

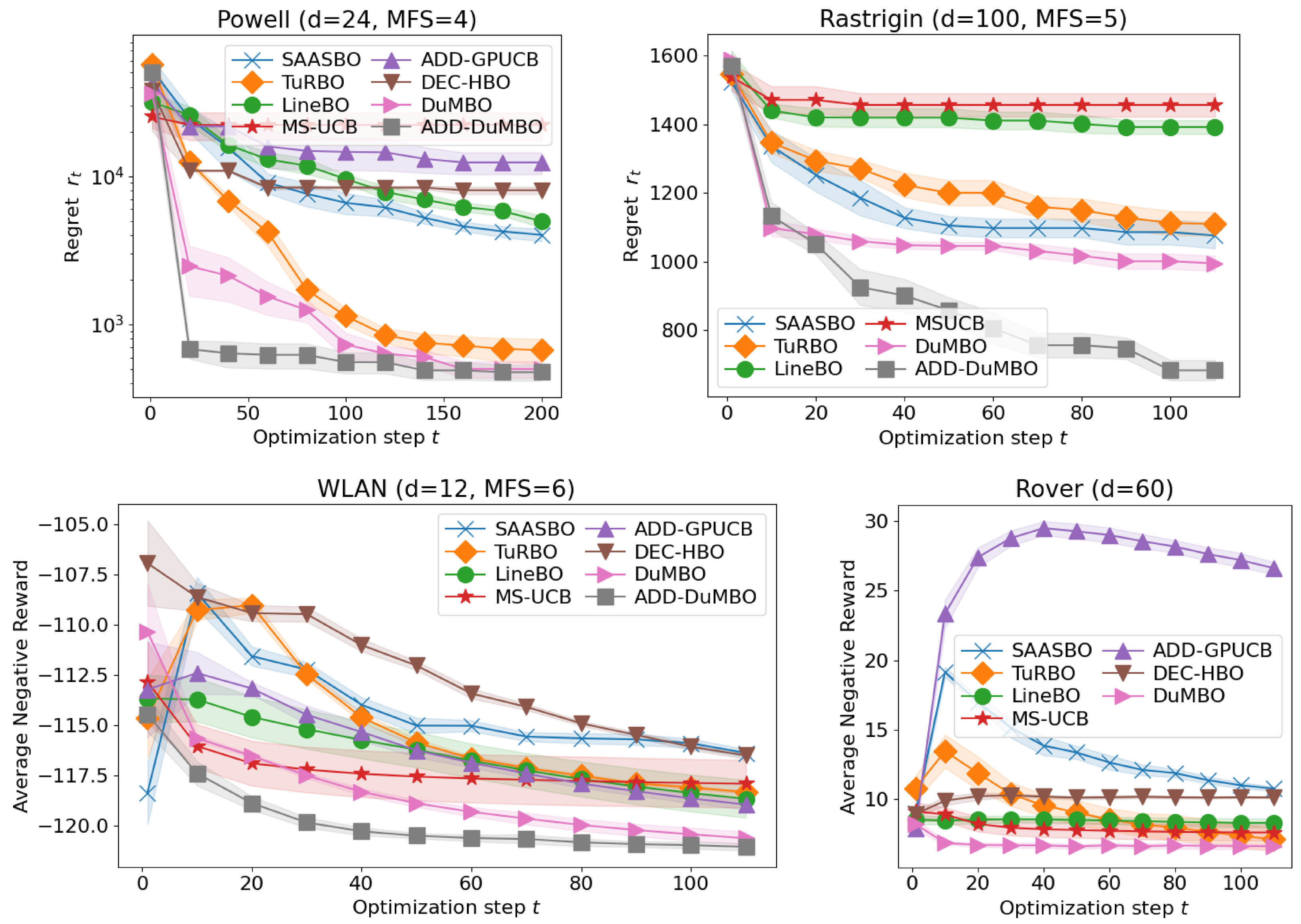
Asymptotic Guarantees

Theorem DuMBO has a **lower regret bound** than the regret bounds of some decentralized no-regret high-dimensional BO algorithms (e.g. [2]).

Corollary DuMBO ensures **no-regret performance**.

Numerical Results

Excellent empirical performance on **synthetic and real-world** problems. When the decomposition is provided instead of being inferred, DuMBO achieves **even better performance** (see ADD-DuMBO).



Summary

- Exploiting the factor graph of the decomposition can prevent over-exploration
- Some acquisition functions can be optimized by ADMM **without weakening BO asymptotic guarantees**
- DuMBO is a high-dimensional BO algorithm that offers simultaneously **a no-regret guarantee** and **excellent empirical performance**.

References

[1] Yu Wang, Wotao Yin, and Jinshan Zeng. "Global Convergence of ADMM in Nonconvex Nonsmooth Optimization". In: *J. Sci. Comput.* 78.1 (Jan. 2019), pp. 29–63. DOI: 10.1007/s10915-018-0757-z.

[2] Trong Nghia Hoang et al. "Decentralized High-Dimensional Bayesian Optimization with Factor Graphs". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. New Orleans, Louisiana, USA, 2018.

Contact: {anthony.bardou, patrick.thiran}@epfl.ch, thomas.begin@ens-lyon.fr