High-Dimensional Bayesian Optimization via Tree-Structured Additive Models

Eric Han,[1] Ishank Arora,[2] Jonathan Scarlett[1,3]

[1]School of Computing, National University of Singapore (NUS) [2]Indian Institute of Technology (BHU) Varanasi [3]Department of Mathematics & Institute of Data Science, NUS

eric_han@nus.edu.sg, ishank.arora.cse14@iitbhu.ac.in, scarlett@comp.nus.edu.sg

Highlights

Considering high-dimensional BO, we lower the computational resources required and facilitate faster model learning by *reducing the model complexity* while retaining the *sample-efficiency* of existing methods.

1. Trade-off expressiveness for scalability via **tree-structures**
2. Extend message passsing via zooming technique to **continuous** domains
3. **Hybrid** method, exploiting tree structures
   - 3.1 Grows tree via Gibbs sampling
   - 3.2 Edge mutation
4. Demonstrate the effectiveness of our approach in a variety of experiments

Introduction

Goal: Find $x_{\max} = \arg\max_{x \in \mathcal{X}} f(x)$ for black-box function $f : \mathcal{X} \to \mathbb{R}$

- Assume f decomposes as a sum of lower-dimensional functions [1]:

$$f(x) = \sum_{G \in \mathcal{G}} f^G\left(x^G\right), \tag{1}$$

where $G \subseteq \{1, \ldots, D\}$ denotes one subset of variables, and $\mathcal{G}$ represents the **tree-structured** additive structure (See example $h(x)$).

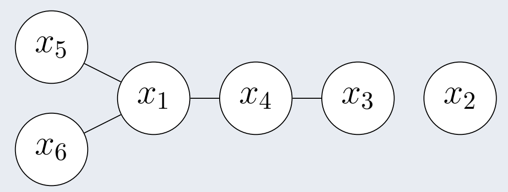


Figure: $h(x) = h^A(x_1, x_6) + h^B(x_1, x_5) + h^C(x_1, x_4) + h^D(x_3, x_4) + h^E(x_2)$.

- We model $f \sim \mathcal{GP}(\mu, \kappa)$, with each f^G being an independent sample from a Gaussian Process $\mathcal{GP}\left(\mu^G, \kappa^G\right)$, and

$$\mu(x) = \sum_{G \in \mathcal{G}} \mu^G\left(x^G\right),$$
$$\kappa(x, x') = \sum_{G \in \mathcal{G}} \kappa^G\left(x^G, x'^G\right). \tag{2}$$

- Focus on upper confidence bound (UCB) – global acquisition function $\phi_t(x)$ be the sum of the individual acquisition functions with respect to structure $\mathcal{G}$:

$$\phi_t(x) = \sum_{G \in \mathcal{G}} \phi_t^G\left(x^G\right),$$
$$\phi_t^G\left(x^G\right) = \mu_{t-1}^G\left(x^G\right) + \beta_t^{1/2} \sigma_{t-1}^G\left(x^G\right). \tag{4}$$

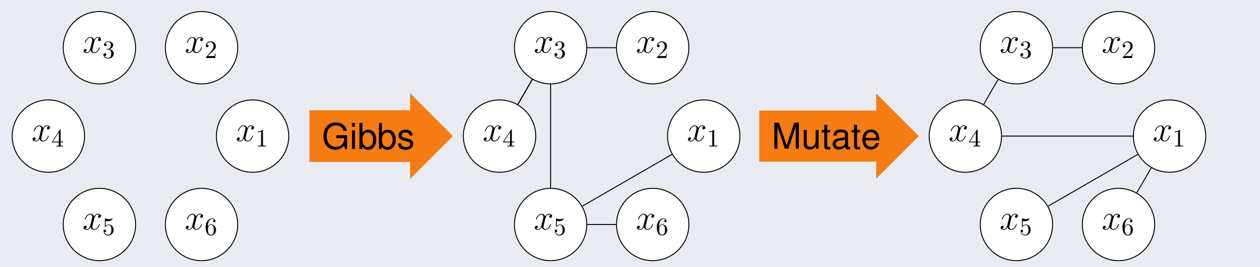
Additive GP-UCB on Tree Structures

Algorithm 1: TREE-GP-UCB

Initialize $\mathcal{D}_0 \leftarrow \{(x_t, y_t)\}_{x_t \in X_{\text{init}}}$
for $t = N_{\text{init}} + 1, \ldots, N_{\text{iter}}$ **do**
  if $t \mod C = 0$ **then**
    Learn $\mathcal{G} \leftarrow$ TREE-LEARNING (Alg. 3)
  Update $\mu_t^G, \sigma_t^G : \forall G \in \mathcal{G}$ (3)
  Optimize $x_t \leftarrow \arg\max_{x \in \mathcal{X}} \phi_t(x)$ (Alg. 2)
  Observe $y_t \leftarrow f(x_t) + \epsilon$
  Augment $\mathcal{D}_t \leftarrow \mathcal{D}_{t-1} \cup \{(x_t, y_t)\}$
return $\arg\max_{(x,y) \in \mathcal{D}} y$

Hybrid Dependency Tree Structure Learning

Exploiting tree-structures to enable efficient tree structure learning

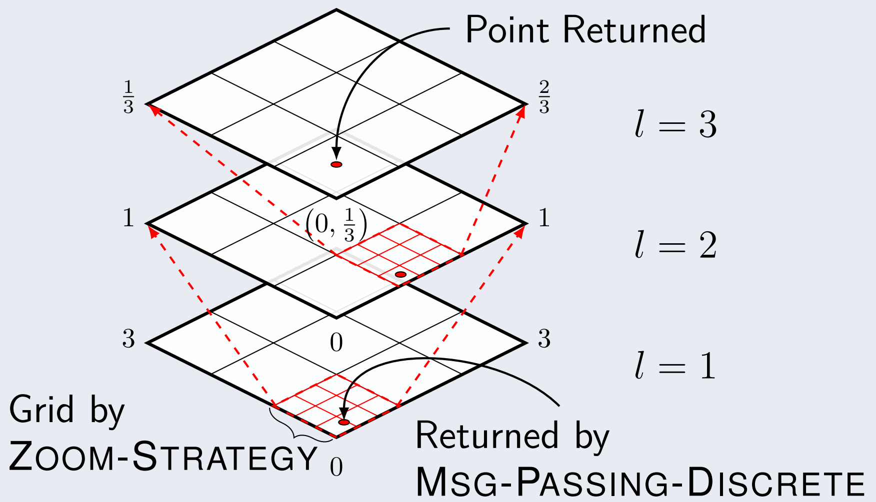


- **Gibbs**: Sample approximately, avoiding the difficult task of sampling directly from the high-dimensional distribution over tree structures.
- **Mutate**: Mutate edges while maintain tree structure diversity from one generation to another.

Optimize Acquisiton Functions via Zooming Technique

Extends [2], recursively zoom into the corresponding sub-domain

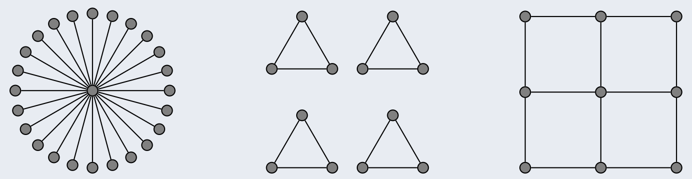
- Due to tree structure, message-passing computation is reduced to **quadratic of the domain**.
- Different zoom strategy can be employed
- Extend generalized additive models to continuous domains.



Experiments

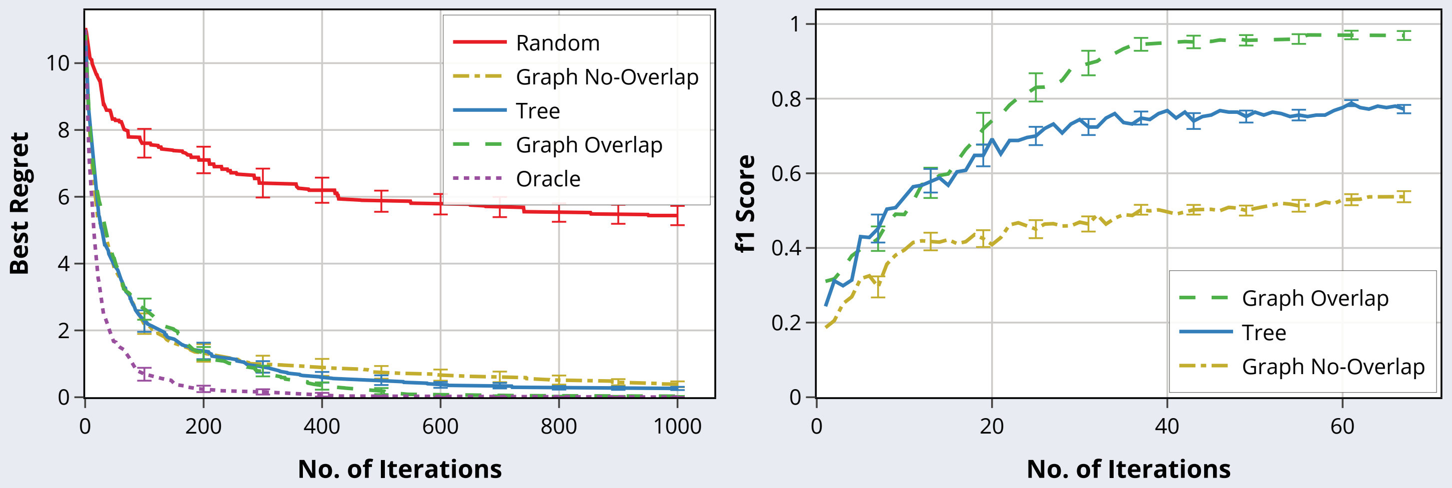
Experiments:

- Additive GP f (see right)
- Non-GP f



(a) Star-25 (b) Part-12 (c) Grid-3×3

Grid-3×3 (Continuous): Tree and Graph No-Overlap are not realizable; Tree remains competitive despite the poorer graph learning.



Scalability: Test Tree's scalability to higher dimensions up to 225D; Tree incurs the lowest cost, as the message passing cost is quadratic in grid size

Real: Tune lpsolve's 74 parameters, finding the best parameters with a time limit of 5 seconds.

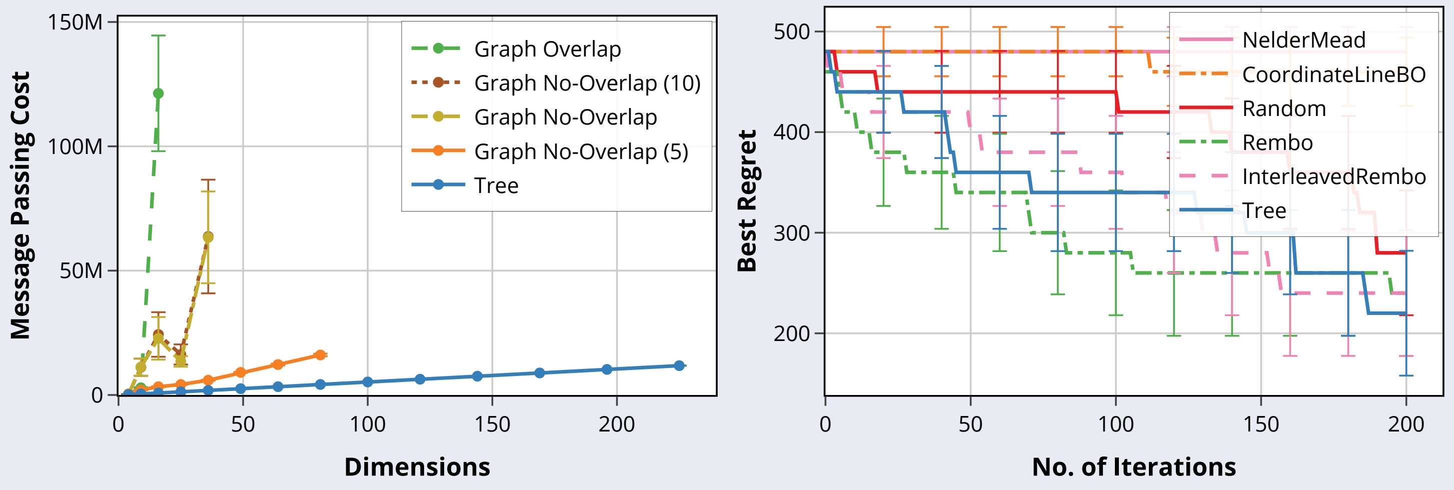


Figure: Scalability of Tree over dimensions Figure: Lpsolve-misc05inf Performance

Additional Experiments: HPOLib2, NAS, Black-box adversarial attack

References

[1] K. Kandasamy, J. Schneider, and B. Póczos, "High dimensional Bayesian optimisation and bandits via additive models," in *Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 295–304.

[2] P. Rolland, J. Scarlett, I. Bogunovic, and V. Cevher, "High-dimensional Bayesian optimization via additive models with overlapping groups," in *Int. Conf. Art. Intel. Stats. (AISTATS)*, 2018, pp. 298–307.