Learning Region of Interest for Bayesian Optimization with Adaptive Level-Set Estimation

Fengxue Zhang[1], Jialin Song[2], James Bowden[4], Alexander Ladd[3],
Yisong Yue[4], Thomas A. Desautels[3], Yuxin Chen[1]

[1]University of Chicago, [2]NVIDIA, Inc, [3]Lawrence Livermore National Laboratory, [4]California Institute of Technology

Scan for Original Paper

Motivating Applications

Are you having trouble optimizing your scientific experiments?

Is it because your objective is *too complex*?

Is the global model failing *to capture useful locality*?

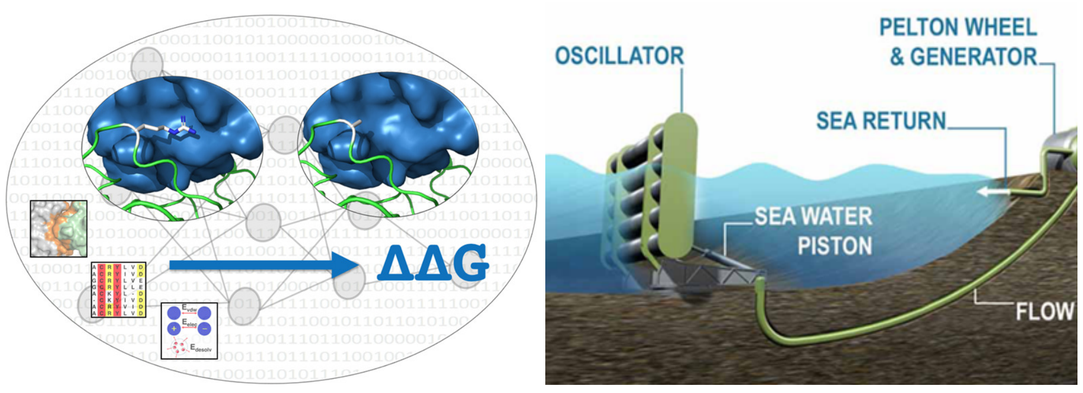


Figure 1: Left: Protein–protein Interactions; Right: Water Converter

Deep Kernel Learning

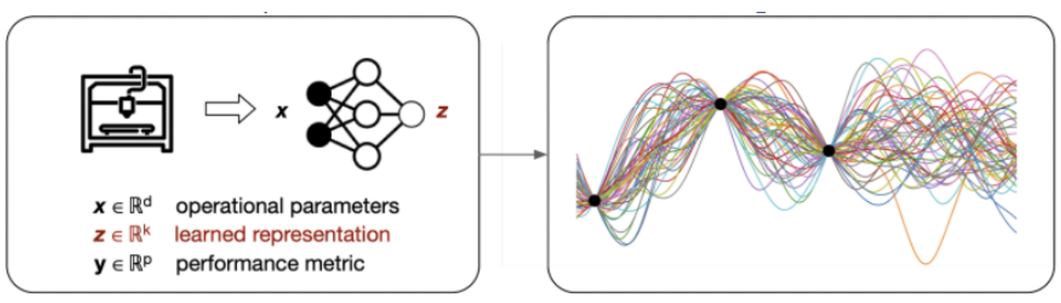


Figure 2: Deep Kernel Learning (DKL)

- A *latent space mapping* g to project input $\mathcal{X}$ to a latent space $\mathcal{Z}$
- An *objective mapping* $h : \mathcal{Z} \to \mathbb{R}$ such that $f(x) \approx h(g(x))$
- $h : \mathcal{Z} \to \mathbb{R}$ often modeled by *Gaussian Processes*
- [ZNC22] initialize the latent space mapping with the encoder of the pre-train auto encoder
- [TDHL20] periodically retrain the auto encoder to improve the latent space.

Global Estimation with DKL

- Learn the global partitioning on $\mathbf{X}$ with Deep Kernel Learning (DKL) [Wil+16] due to its scalability
- Underlying global function $f_g \coloneqq f$ is assumed to be drawn from a global Gaussian process.
- Maximizing the negative log-likelihood (NLL) $-\log\big(P(\mathbf{y}_t|X_t, \theta_{f_g,t})\big)$ [RW05].
- Sample unlabeled dataset from $\mathbf{X}$ to pre-trained an Auto-Encoder and use the parameters of its encoder to initialize the neural network q [FCT20].

Regions of Interest Filtering

With the confidence interval of the global $\mathcal{GP}_{f_g}$, we could define the upper confidence bound $UCB_{f_g}(\mathbf{x})$ and lower confidence bound $LCB_{f_g}(\mathbf{x})$ We attain the region of interest.

$$\hat{\mathbf{X}} = \{x \in \mathbf{X} |\ UCB_{f_g}(\mathbf{x}) > LCB_{f_g,max}\} \tag{1}$$

- The global $\mathcal{GP}_{f_g}$ enables a filtering on $\mathbf{X}$ to locate the region of interest $\hat{\mathbf{X}}$.
- It is desired for $\hat{\mathbf{X}}$ that with high probability, $\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x})$ is contained by $\hat{\mathbf{X}}$,
- Also $|\hat{\mathbf{X}}| \ll |\mathbf{X}|$, then $\hat{f}$, the objective function defined on $\hat{\mathbf{X}}$ is therefore of reduced complexity.

Bayesian Optimization on Partition(s)

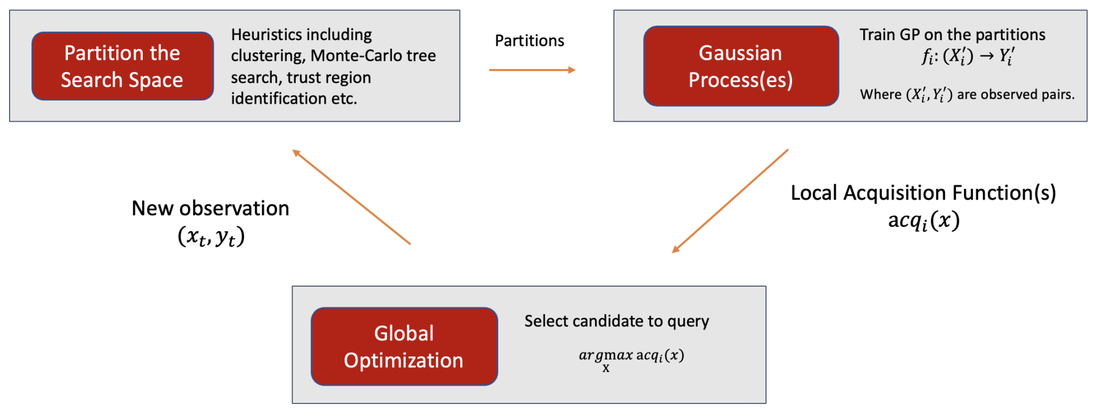


Figure 3: Bayesian Optimization on Partition(s)

Previous methods rely on heuristics that normally introduce *additional complexity* to the surrogate models, which could incur challenges in fine-tuning the *hyperparameters* of the heuristics.

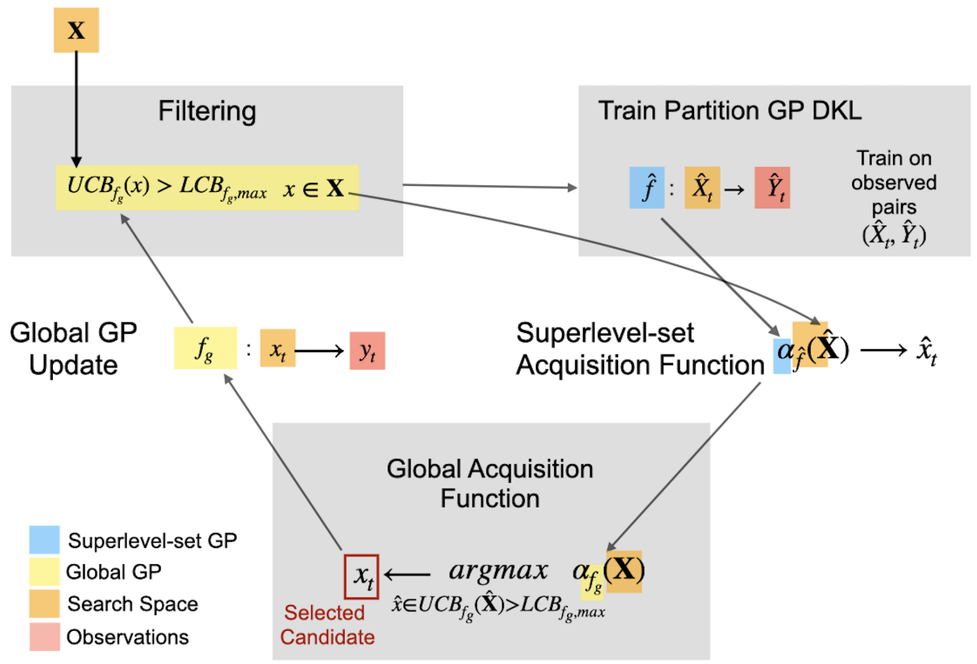
Algorithmic Details

Algorithm 1 **B**ayesian Optimization with **A**daptive **L**evel-**S**et **E**stimation (BALLET)

Input:Search space $\mathbf{X}$, initial observation $\mathbf{D}_0$, horizon T;
for $t = 1$ *to* T **do**
  Learn the global estimation $\mathcal{GP}_{f_g,t}$: $\theta_{f_g,t} \leftarrow \arg\max_{\theta_{f_g}} -\log(P(\mathbf{y}_t|X_{t-1}, \theta_{f_g}))$
  Partition by region of interest filtering: $\hat{\mathbf{X}}_t \leftarrow \{\mathbf{x} \in \mathbf{X} |\ UCB_{f_g,t}(\mathbf{x}) > LCB_{f_g,t,max}\}$
  Partition the historical observation: $\hat{\mathbf{D}} = \{(\mathbf{x}, y) \in \mathbf{D} | \mathbf{x} \in \hat{\mathbf{X}}_t\}$.
  Learn the superlevel-set GP: $\mathcal{GP}_{\hat{f},t}$: $\theta_{\hat{f},t} \leftarrow \arg\max_{\theta_{\hat{f}}} -\log(P(\mathbf{y}_t|\hat{\mathbf{X}}_t, \theta_{\hat{f}})$
  Optimize the superlevel-set acquisition function: $\mathbf{x}_{t+1} \leftarrow \arg\max_{\mathbf{x}\in\hat{\mathbf{X}}} \alpha_{\hat{f}}(\mathbf{x}))$
  Update $\mathbf{D}$: $\mathbf{D}_{t+1} \leftarrow \mathbf{D}_t \cup \{(\mathbf{x}_{t+1}, y_{t+1})\}$
Output: $\max_t y_t$

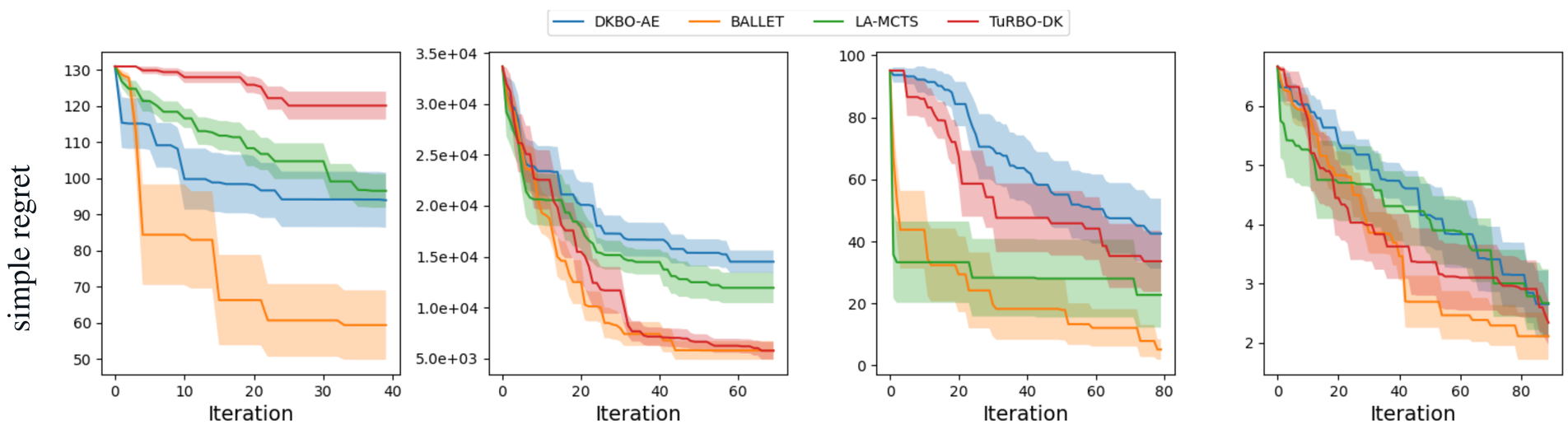
In contrast to existing work, we consider a *non-parametric model* for partitioning the search space, which has shown remarkable performance in real-world tasks while having *few hyperparameters* to maintain.

Bayesian Optimization with Adaptive Level-Set Estimation



The algorithm filters the search space $\mathbf{X}$ using the UCB and LCB estimated by a *global GP*. Then it feeds another GP the *regions of interest* of the search space $\hat{\mathbf{X}}$ and the filtered historical observations $(\hat{\mathbf{X}}_t, \hat{\mathbf{Y}}_t)$. Global optimization at each iteration is conducted on the *adaptive partition* using this GP and its acquisition function.

Experiment Results



(a) HDBO-200D (b) Water Converter (c) Nanophotonic (d) Rosetta