Bayesian Optimization with Inexact Acquisition: Is Random Grid Search Sufficient?

Hwanwoo Kim, Chong Liu & Yuxin Chen Contact: hwanwoo.kim@duke.edu

Duke University · The University of Chicago

Overview

- We theoretically study the effect of the **inexact acquisition function solutions** in BO.
- We theoretically justify **random grid search as a valid acquisition solver**.
- We **empirically validate the efficiency of random grid search** over the existing acquisition function solvers.

Background

Goal Maximize a blackbox function $f : \mathcal{X} \to \mathbb{R}$, where

- f is not necessarily convex nor differentiable.
- Only source of information about f is through its noise-corrupted evaluations $y_t = f(x_t) + \epsilon_t$, which are typically expensive.

Gaussian Process Denote generic query locations by $X_t = \{x_1, \dots, x_t\}$ and the corresponding observations by $Y_t = [y_1, \dots, y_t]^\top$. Given a positive-definite kernel function k, Gaussian process interpolation with a prior $\mathrm{GP}(0, k)$ yields the following posterior predictive mean and variance:

$$\mu_t(x) = k_t(x)^\top (K_{tt} + \tau I)^{-1} Y_t,$$
$$\sigma_t^2(x) = k(x,x) - k_t(x)^\top (K_{tt} + \tau I)^{-1} k_t(x),$$

where $k_t(x) = [k(x, x_1), \dots, k(x, x_t)]^\top$ and K_{tt} is a $t \times t$ matrix with entries $(K_{tt})_{i,j} = k(x_i, x_j)$.

Popular Bayesian Optimization Strategies: GP-UCB / GP-TS

Input: Kernel k; Total number of iterations T; Initial query locations X_0; Initial noise-free observations Y_0

For $t = 1, \cdots, T$:

1. Compute posterior mean/variance using all query locations and function evaluations (X_{t-1}, Y_{t-1}).
2. For GP-UCB, obtain
   - $x_t = \arg\max \mu_{t-1}(x) + \beta_t \sigma_{t-1}(x)$

   For GP-TS, obtain
   - $x_t = \arg\max_{D_t} f_t(x)$, $f_t \sim \mathrm{GP}(\mu_{t-1}, \beta_t^2 \sigma_{t-1}^2)$
3. Set $X_t = X_{t-1} \cup \{x_t\}$, $Y_t = Y_{t-1} \cup \{f(x_t)\}$.

Performance Metric: The cumulative regret is defined as

$$R_T = \sum_{t=1}^{T} f(x^*) - f(x_t).$$

Inaccuracy Measure

- Measure the worst-case accumulated inaccuracy

$$M_T = \sum_{t=1}^{T} (1 - \tilde{\eta}_t), \quad 0 < \tilde{\eta}_t \le \frac{\alpha_t(x_t)}{\alpha_t^*}$$

- If M_T grows sublinearly, one can show that existing convergence results for GP-UCB and GP-TS still hold [KLC].

Regret Analysis with Random Grid Search

- Employ a sequence of random grids $\{\mathcal{X}_t\}_{t \in \mathbb{N}}$ whose size grows linearly with the iteration count, i.e., $|\mathcal{X}_t| = \Theta(t)$. Therefore, we allow in exactness in optimizing acquisition functions.
- For GP-UCB, obtain
  - $x_t = \arg\max_{\mathcal{X}_t} \mu_{t-1}(x) + \beta_t \sigma_{t-1}(x)$

  For GP-TS, obtain
  - $x_t = \arg\max_{\mathcal{X}_t} f_t(x)$, $f_t \sim \mathrm{GP}(\mu_{t-1}, \beta_t^2 \sigma_{t-1}^2)$
- Impose a function class assumption: reproducible kernel Hilbert space associated with a kernel k, denoted by $\mathcal{H}_k$.
- Cumulative regret in terms of maximum information gain [VKP21]
  - $\gamma_T = O\left(\log^{d+1} T\right)$ for exponential kernel
  - $\gamma_T = O\left(T^{\frac{d}{d+2\nu}} \log^{\frac{2\nu}{2\nu+d}} T\right)$ for Matérn kernel

Theorem 0.1 (Grid search GP-UCB [KLC]). *Under mild assumptions on kernel, suppose $f \in \mathcal{H}_k$ with $\|f\|_{\mathcal{H}_k} \le B$. With probability $1 - \delta$, the random grid GP-UCB with $\beta_t = B + R\sqrt{2(\gamma_{t-1} + 1 + \log(2/\delta))}$ yields*

$$R_T = \mathcal{O}\left(\gamma_T \sqrt{T}\right) + \tilde{\mathcal{O}}\left(T^{\frac{d-1}{d}}\right).$$

Theorem 0.2 (Grid search GP-TS [KLC]). *Under mild assumptions on kernel, suppose $f \in \mathcal{H}_k$ with $\|f\|_{\mathcal{H}_k} \le B$. With probability $1 - \delta$, the random grid GP-TS with $\beta_t = B + R\sqrt{2(\gamma_{t-1} + 1 + \log(3/\delta))}$ yields*

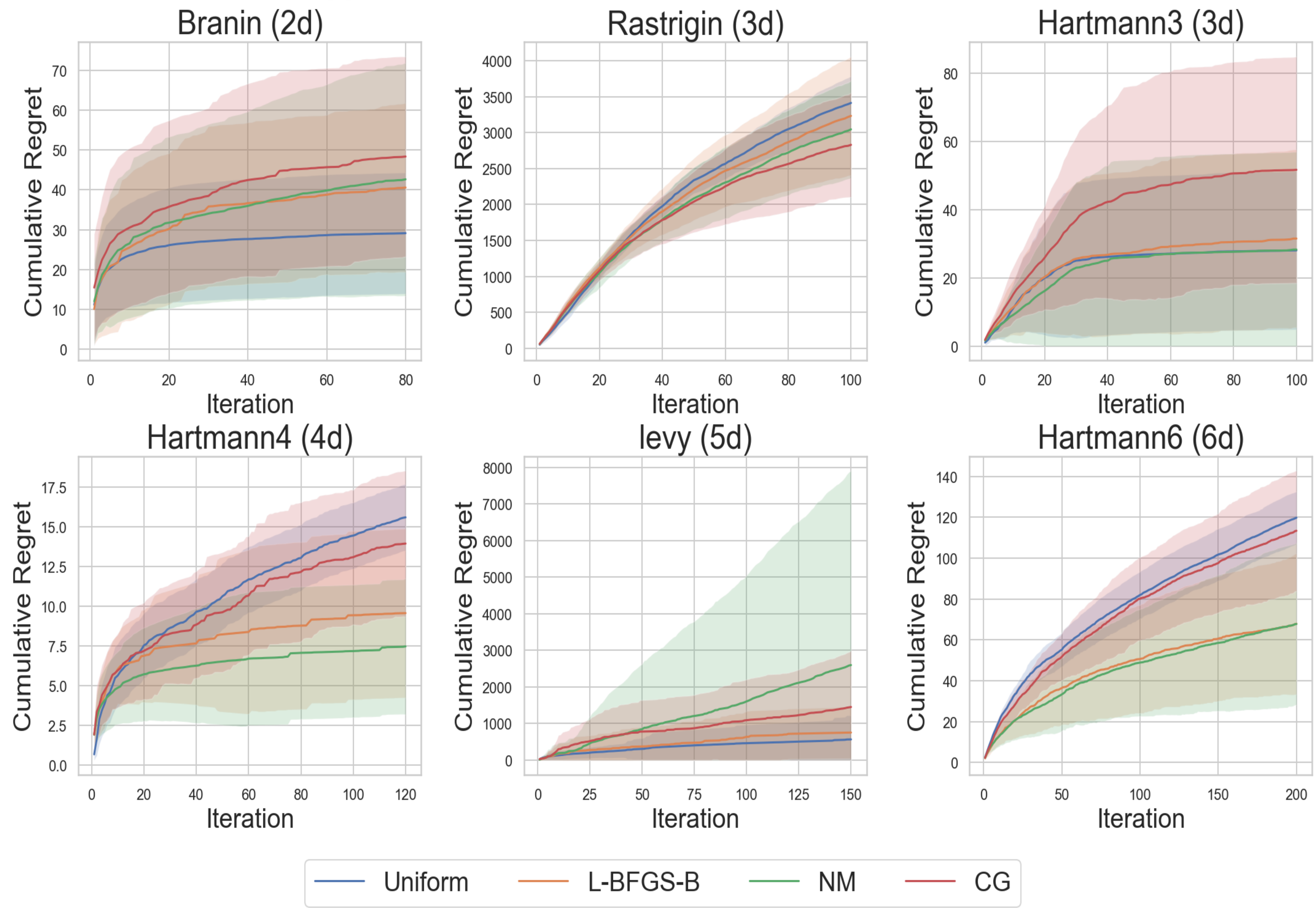
$$R_T = \mathcal{O}\left(\gamma_T \sqrt{T \log T}\right) + \tilde{\mathcal{O}}\left(T^{\frac{d-1}{d}}\right).$$

Remark. *Previous GP-TS convergence requires $\mathcal{X}_t = \Theta(t^{2d})$.*

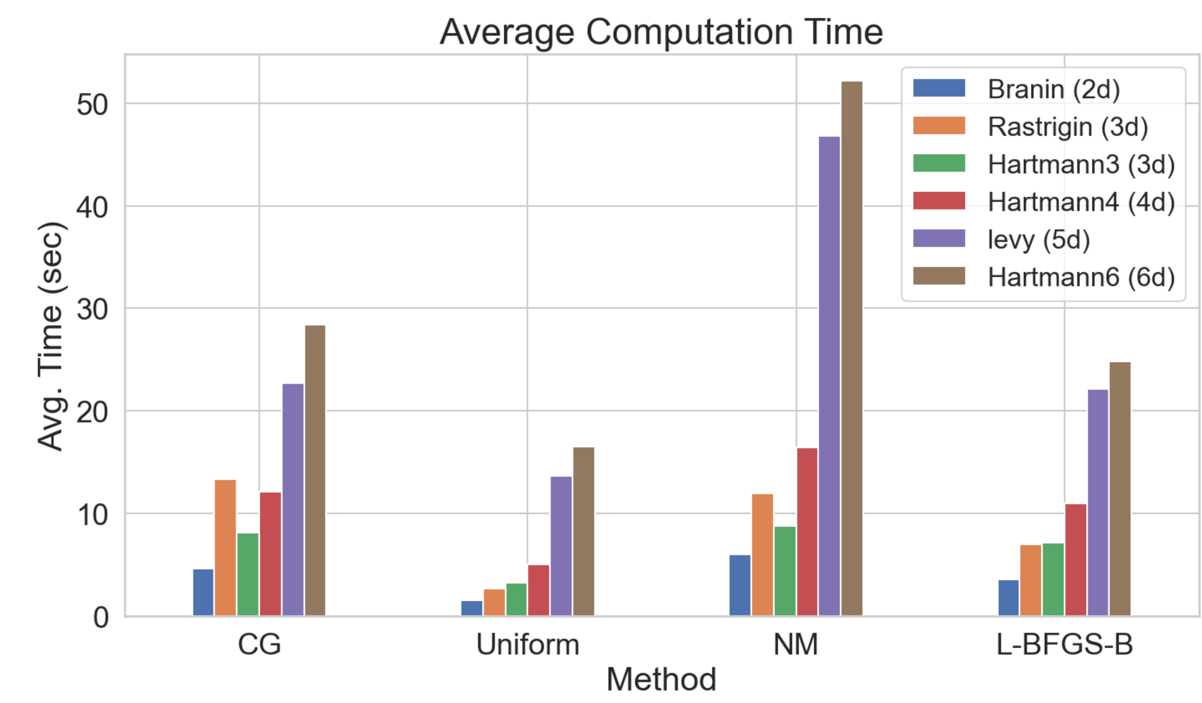
Numerical Experiments

- Performance comparison of BO with a range of acquisition function solvers: Uniform (uniform random grid search), L-BFGS-B (Limited memory Broyden-Fletcher-Goldfarb-Shanno), NM (Nelder Mead), CG (Conjugate Gradient)

Cumulative Regret Plot



Computation Time



References

[KLC] Hwanwoo Kim, Chong Liu, and Yuxin Chen. Bayesian optimization with inexact acquisition: Is random grid search sufficient? In *The 41st Conference on Uncertainty in Artificial Intelligence*.

[VKP21] Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 82–90. PMLR, 2021.