Improving sample efficiency of high dimensional Bayesian optimization with MCMC

Zeji Yi[1*]   ZEJIY@ANDREW.CMU.EDU

Yunyue Wei[1*]   WEIYY20@MAILS.TSINGHUA.EDU.CN

Chu Xin Cheng[2*]   CCHENG2@CALTECH.EDU

Kaibo He[1]   HKB21@MAILS.TSINGHUA.EDU.CN

Yanan Sui[1]   YSUI@TSINGHUA.EDU.CN

[1]Tsinghua University, [2]California Institute of Technology.

Abstract

Sequential optimization methods are often confronted with the curse of dimensionality in high-dimensional spaces. Current approaches under the Gaussian process framework are still burdened by the computational complexity of tracking Gaussian process posteriors and need to partition the optimization problem into small regions to ensure exploration or assume an underlying low-dimensional structure. With the idea of transiting the candidate points towards more promising positions, we propose a new method based on Markov Chain Monte Carlo to efficiently sample from an approximated posterior. We provide theoretical guarantees of its convergence in the Gaussian process Thompson sampling setting. We also show experimentally that both the Metropolis-Hastings and the Langevin Dynamics version of our algorithm outperform state-of-the-art methods in high-dimensional sequential optimization and reinforcement learning benchmarks.

Keywords: Bayesian Optimization, Markov Chain Monte Carlo, High Dimensional Optimization

1. Introduction

With broad applications in real-world engineering problems, black-box function optimization is an essential task in machine learning (Snoek et al., 2012; Hernández-Lobato et al., 2017). These non-convex sequential optimization problems often lack gradient information. Bayesian optimization (BO) is a popular sampling-based online optimization approach for solving expensive optimization problems. It has been successfully applied to problems such as online learning and sequential decision-making. BO builds a surrogate model for modeling the objective function and optimizes the acquisition function to propose new samples.

Similar to many other numerical problems, BO algorithms are also susceptible to the curse of dimensionality. The search space would grow exponentially large as the function dimension increases and become intractable under a limited computation budget. Common acquisition functions also tend to over-explore the uncertainty boundary region and lack exploitation in high dimensional input space (Oh et al., 2018). Recent developments in high dimensional BO include constructing trust regions and space partitions to improve the probability of sampling in promising regions, which effectively ameliorates the problem of over-exploration (Eriksson et al., 2019; Wang et al., 2020). To evaluate the acquisition function on a continuous domain, these methods often discretize

---

*These authors contributed equally to this work. Code and Appendices are available at https://drive.google.com/drive/folders/1fLUHIduB3-pR78Y1YOhhNtsDegaOqLNU?usp=sharing.

the search space using Sobolev sequence (Sobol', 1967). However, this kind of discretization could be inadequate in high dimensional space as the size of the discretization set is limited, hindering the exploitation of potential good regions.

In our setting, we investigate probabilistic acquisition such as Thompson sampling (TS) for optimizing unknown function modeled with Gaussian processes (GP). Yet, directly applying algorithms that perform well in low-dimensional problems to high-dimensional domains often suffers due to the high-computation demand and the phenomenon of over-exploration. Our proposed method aims to improve optimization performance in high-dimensional spaces by adapting to the search domain and specifically discretizing it in promising regions. To achieve this, we utilize Markov Chain Monte Carlo (MCMC) for BO, which is a widely-used technique known for its ability to effectively sample from high-dimensional posterior distributions. We introduce MCMC-BO, a method that transitions candidate points from their original positions towards the approximated stationary distribution of TS. By only tracking a batch size number of points during transitions, MCMC-

Figure 1: Illustration of MCMC-BO. The contours are 2d Rastrigin function. (**Left**): BO algorithms propose points to be sampled. The optimization performance is restricted by insufficient discretization. (**Right**): Points are adjusted by MCMC-BO , reaching regions with higher value.

BO significantly reduces the storage requirements in comparison to an entire mesh while still maintaining theoretical performance guarantees. It can be easily generalized to different scenarios with few additional hyper-parameters. Based on the basic MH version of MCMC-BO . We also propose the Langevin dynamics (LD) version which transits following the gradient of the log-likelihood. Overall, MCMC-BO can serve as a versatile algorithm that can be linked to most existing BO methods providing a posterior for any candidate points for continuous problems. We summarize our contributions as follows:

- We propose MCMC-BO, a Bayesian optimization algorithm which performs adaptive local optimization in high-dimensional problems that achieves both time and space efficiency.
- We provide theoretical guarantees on the convergence of MCMC-BO. To our knowledge, this is the first regret bound on high-dimensional Bayesian optimization problem which can deal with the scaling of dimensions with limited candidate points per round and avoid the overuse of memory.
- We experimentally show that MCMC-BO , combined with both MH and LD, outperforms other strong baselines on various high-dimensional tasks.

2. Related Work

The optimization of black-box functions has been broadly used in many scenarios, such as hyper-parameter tuning and experimental design (Snoek et al., 2012; Hernández-Lobato et al., 2017). Evolutionary algorithms are a class of methods with a long history and good performance to tackle the black-box optimization problem. One representative is CMA-ES, which adaptively adjusts the
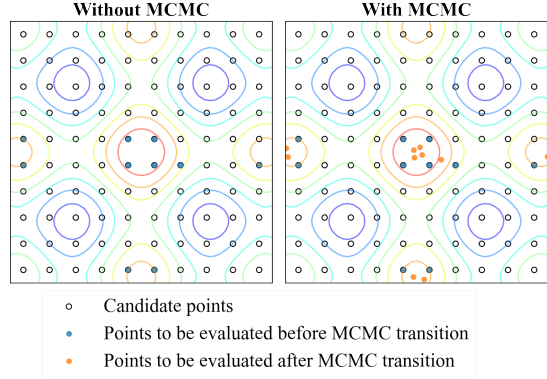
covariance matrix to generate new samples (Hansen et al., 2003). These kinds of problems can also be formalized under the Bayesian optimization framework (Shahriari et al., 2015; Frazier, 2018), which is also more efficient. However, conventional BO algorithms are limited to low-dimensional cases and a small number of observations. There are two major components in BO: modeling and performing acquisition. To improve scalability and efficiency, various methods have been proposed for these two aspects. Regarding modeling, there exist a large amount of works which scale BO to higher dimensions via approximated GP (Hensman et al., 2013), dimension reduction (Nayebi et al., 2019) or function decomposition (Wang et al., 2018). We leave the detailed discussion of this part in the appendix.

In the acquisition phase, balancing the exploration-exploitation trade-off is important. High-dimensional optimization problems often suffer from inadequate exploitation. To prevent over-exploring uncertain points near the domain boundary, methods like strong priors, dimension dropout, and cylindrical kernels have been proposed to allocate more sample points in the center region (Eriksson and Jankowiak, 2021; Li et al., 2018; Oh et al., 2018). Full exploration in high-dimensional space requires unrealistic computational resources and time. Therefore, adequate exploitation of optimal points determines final performance. Previous algorithms for optimizing high-dimensional spaces include TuRBO and LA-MCTS. TuRBO limits the candidate set to a small box around the best points and adjusts the box based on later evaluations (Eriksson et al., 2019).LA-MCTS partitions the region with a constructed tree, recommending promising regions to improve algorithm performance (Wang et al., 2020). Both methods focus on dividing promising regions, but maintaining a high-precision discretized grid for such regions is costly and can limit computation storage. The trade-off between regret and storage emphasizes the need for improvements in current methods. Therefore, we propose attention to the ability to congregate candidate points around the optimum in our algorithm.

Focusing on probabilistic acquisition, which involves using stochastic samples from the model posterior, sampling from the posterior distribution can be a bottleneck, especially for problems without a closed form expression for its posterior distribution. To address this, approximate sampling techniques are commonly used to generate samples from posterior approximations (Chapelle and Li, 2011; Lu and Van Roy, 2017). These approaches have shown effective practical performance (Riquelme et al., 2018; Urteaga and Wiggins, 2017). However, maintaining performance over arbitrary time horizons while using approximate sampling remains unclear. Recent works have explored the application of LD, a stochastic sampling approach, in Multi-armed Bandit (MAB) problems (Mazumdar et al., 2020; Xu et al., 2022). Instead of restricting to a parameterized utility function and optimizing the parameter space, we directly employ LD on the continuous and infinite action space. We believe this approach is more general and effective as it allows direct acquisition of candidate points without the need for optimization in the high-dimensional space.

3. Background

3.1. Modeling with Gaussian processes

Black-box optimization under uncertainty aims to find

$$\mathbf{x}^* \in D \text{ such that } f(\mathbf{x}^*) \geq f(\mathbf{x}), \quad \forall \mathbf{x} \in D \tag{1}$$

where $f : D \to \mathbb{R}$ is an unknown function. In BO, the unknown objective f is viewed as a probability distribution, and GP emerges as an effective way to estimate the unknown function, maintain

an uncertainty estimate, and can be sequentially updated with available information. A GP is characterized by its mean function $\mu(\cdot)$ and kernel $k(\cdot, \cdot)$. The prior distribution of $f(x)$ is assumed to be $\mathcal{N}(0, k(x, x))$. Under this prior, the sample points $A_T := [x_1, \cdots, x_T]$ and the observations $[y_1, \cdots y_T]$ follows the multivariate distribution $\mathcal{N}(0, K_T + \sigma^2 I)$, where $K_T = [k(x, x')]_{x, x' \in A_T}$. Denote $k_T(x) = [k(x_i, x)]_{i=1:T}^T \in \mathbb{R}^{T \times 1}$, the posterior distribution over f is thus Gaussian with mean $\mu_T(x)$ and covariance $k_T(x, x')$ that satisfy: $\mu_T(x) = k_T(x)^T (K_T + \sigma^2 I)^{-1} y_T, k_T(x, x') = k(x, x') - k_T(x)^T (K_T + \sigma^2 I)^{-1} k_T(x')$

3.2. Gaussian process and Thompson sampling

In online learning tasks where the actions x_t have to be sequentially chosen at each time step t, a common metric for measuring the performance is regret, which is defined using the optimal action $x^* := \arg \max_{x \in D} f(x)$ such that $r_t := f(x^*) - f(x_t)$ for the instantaneous regret at time t. The goal of optimization is to minimize the sum of r_t, which entails finding optimal points while effectively exploring the sample space. To achieve this, we build upon ideas inspired by TS, which is to select the best point using a stochastic sample from the most recent model. TS is a classical MAB algorithm that can balance the exploitation-exploration trade-off. In every round, the algorithm estimates the reward of each arm by sampling from their respective distributions and plays the arm with the highest estimate. Thus, exploitation of optimal arms is achieved by heuristically playing the optimal arm, while exploration is guaranteed by the intrinsic uncertainty of probabilistic sampling. With an evolving GP, the algorithm first samples $f_t \sim GP(\mu_t(x), k_t(x, x'))$ and selects $x_t := \mathrm{argmax}_{x \in \Omega} f_t(x)$ as the candidate points. Chowdhury and Gopalan gave a theoretical analysis for the algorithm GP-TS, which combines GP and TS to optimize stochastic MAB problems (Chowdhury and Gopalan, 2017).

3.3. Markov Chain Monte Carlo

Markov Chain Monte Carlo is used in GP modeling to approximate intractable probability distributions (Hensman et al., 2015). Similarly, we hope to utilize MCMC to sample from a distribution on the action set corresponding to our acquisition approach.

Metropolis-Hastings (MH) is an MCMC sampling algorithm that can sample from a target distribution $\pi(x)$, $x \in \mathcal{X}$ known up to a constant, i.e. when we have knowledge of $\pi_d(x) = c \cdot \pi(x)$ (Metropolis et al., 1953). With a given proposal distribution $q(\cdot, x)$, the algorithm samples a candidate point $y \in \mathcal{X}$ given the current value x according to $q(y, x)$. Then the Markov Chain transitions to y with acceptance probability $\alpha(x, y) = \min \left\{ 1, \frac{\pi_d(y)q(y, x)}{\pi_d(x)q(x, y)} \right\}$. This ratio yields a high acceptance probability when the proposed point is likely to be a better choice than the current point, so the points transit towards regions with higher function values. The MH algorithm is guaranteed to converge to a stationary distribution that is exactly $\pi(\cdot)$.

Langevin dynamics (LD), known to converge to the steady state distribution $\pi(x)$, provides another efficient iterative sampling procedure. From any arbitrary point x_0 sampled from a prior distribution, the update rule is given by $x_{i+1} = x_i + \epsilon \cdot \nabla \log \pi(x) + \sqrt{2\epsilon} z_i, \quad z_i \sim \mathcal{N}(0, I)$.

Directly applying these two MCMC algorithms is challenging, as we do not have access to the complicated target distribution $\pi(x)$ induced by our probabilistic acquisition approach. In Section 4, we propose estimates for the acceptance probability $\alpha(x, y)$ and $\nabla \log \pi(x)$ that allows us to sample from the posterior distribution.

4. Algorithm Design

We propose MCMC-BO, a novel algorithmic framework that combines MCMC methods with GP to evaluate more points effectively while maintaining a dense, time-varying discretization near promising regions for optimization.

The key idea of our algorithm is inspired by TS in which points with larger probability of being optimal are chosen next. However, probabilistic acquisition in this manner is computationally demanding due to the required size of the discretization set and the complexity of the induced distribution. Coarse partitioning of the input domain naturally leads to regret, as the best possible point in the mesh may still be far from the optimum. Consequently, a large number of points is necessary to ensure a fine partition. In high-dimensional spaces, this need for numerous points can result in large matrices for sampling in stochastic acquisition approaches and slow predictions.This presents a dilemma: a coarse mesh leads to significant regret, while a fine mesh makes sampling unrealistic. To efficiently sample from an intractable distribution and identify better candidate points, we employ MCMC.

MCMC-BO offers great flexibility in choosing different MCMC algorithms to sample from the GP posterior. We provide two MCMC subroutines using MH and LD. Our proposed method is compatible with any existing BO algorithms with probabilistic acquisition functions over a continuous domain (see Appendix). The MH algorithm introduced in Section 3.3 can serve as a overarching framework using various acceptance ratios, and we propose an acceptance ratio that approximates the transition probability induced by the acquisition approach.

Algorithm 1: MCMC-BO

Input: Initial dataset $\mathcal{D}_0$, batch size m, MCMC transition number N, MCMC transition parameter Θ

1   **for** $t = 0, 1, \cdots$ **do**
2   Update posterior distribution $\text{GP}\big(\mu_t(x), k_t(x, x')\big)$ on f using $\mathcal{D}_t$
3   Create discretized candidate points $\boldsymbol{x}_t^0$ from continuous search domain
     // MCMC transition start
4   **for** $i = 0$ **to** $N - 1$ **do**
5   **for** $k = 0$ **to** $m - 1$ **do**
6   Perform **[MCMC routine]** using $\text{GP}\big(\mu_t(x), k_t(x, x')\big)$ on x_{tk}^i with parameter Θ
7   **end**
8   **end**
     // MCMC transition end
9   Observe $\boldsymbol{y}_t = f(\boldsymbol{x}_t^N)$
10    $\mathcal{D}_{t+1} \leftarrow \mathcal{D}_t \cup (\boldsymbol{x}_t^N, \boldsymbol{y}_t)$
11 **end**

TS implicitly induces a distribution on each candidate point $x \in D_t$, with the probability of being selected as $\mathbb{P}(f_t(x) \geq f_t(x')), \forall x' \in D_t$. Let x_p, x_o be two candidate points after n samples denoted by history $\mathcal{H} = \{(x_t, y_t)\}_{t=1}^n$. We can sample according to the distribution of TS using MH with an acceptance rate $\alpha = \min\left\{1, \frac{\mathbb{P}(f_t(x_p) \geq f_t(x), \forall x \in D_t) \cdot q(x_p | x_o)}{\mathbb{P}(f_t(x_o) \geq f_t(x), \forall x \in D_t) \cdot q(x_o | x_p)}\right\}$, where, D_t represents the discretization set of the domain D at time t, and x_p is determined by the transition kernel $q(\cdot \mid \cdot)$. However, computing the probability involved in the ratio, $\mathbb{P}\big(f_t(x_p) \geq f_t(x), \forall x \in D_t\big)$, requires

marginalizing over the joint distribution of a high-dimensional multivariate normal distribution, which is intractable. Therefore, we propose an alternative approach that captures a similar idea. We define the acceptance probability as $\alpha = \min\left\{1, \frac{\mathbb{P}(f_t(x_p) \geq f_t(x_o))q(x_p|x_o)}{\mathbb{P}(f_t(x_o) \geq f_t(x_p))q(x_o|x_p)}\right\}$. From GP regression, we can view the surrogate function values $y_p = f_t(x_p)$, $y_o = f_t(x_o)$ associated to selected points $\boldsymbol{x} := (x_p, x_o)$ as a Gaussian random vector with joint distribution as specified in Section 3.1. Using linear transformations of a Gaussian random vector, let $c^T = \begin{bmatrix} 1 & -1 \end{bmatrix}$, $K_n = [k(x_{1:n}, x_{1:n})] \in \mathbb{R}^{n \times n}$, and $k_n(\boldsymbol{x}) = [k(x_{1:n}, x_p), k(x_{1:n}, x_o)] \in \mathbb{R}^{n \times 2}$, we have that $y_p - y_o \sim \mathcal{N}\left(c^T\mu, c^T\Sigma c\right)$, where $c^T\mu = \left(k(x_p, x_{1:n}) - k(x_o, x_{1:n})\right)^T (K_n + \sigma^2 I)^{-1} y_{1:n}$, $c^T\Sigma c = \left(k(x_p, x_p) + k(x_o, x_o) - 2k(x_p, x_o)\right) - \left(k_n(\boldsymbol{x})c\right)^T (K_n + \sigma^2 I)^{-1}\left(k_n(\boldsymbol{x})c\right)$. Since $c^T\mu, c^T\Sigma c \in \mathbb{R}$, we have that $P(y_p - y_o > 0|\mathcal{H}) = \Phi\left(\frac{c^T\mu}{c^T\Sigma c}\right)$ (here $\Phi(\cdot)$ denotes the CDF of the standard normal). Therefore, the acceptance probability reduces to

$$\alpha = \min\left\{1, \frac{\Phi\left(\frac{c^T\mu}{c^T\Sigma c}\right)q(x_p|x_o)}{(1 - \Phi\left(\frac{c^T\mu}{c^T\Sigma c}\right))q(x_o|x_p)}\right\}. \tag{2}$$

A demonstration of MCMC-BO with MH is shown in Algorithm 2. MCMC-BO prepares a batch of m candidate points x_{tk}^i each round either from direct discretization or points to be executed from other algorithms, where t stands for the round number, $k = 1, 2, 3, \cdots, m$, and i stands for transition times. Then with a proposed transition kernel, often defined using Brownian motion $x^{i+1} - x^i \sim \mathcal{N}(0, \sigma)$, we generate m pairs of points x_{tk}^i, x_{tk}^{i+1}. We accept or decline the transition with the ratio in Equation 2. The random walk of Markov chain enables dense discretization of continuous space on more optimal regions, as illustrated in Fig 1.

The generality of our framework also extends to other sampling methods such as LD, where the samples transit following Equation ??. Previously, MH aims to sample from the intractable distribution corresponding to our acquisition approach by simplifying the ratio between two discretization points. Denoting the time-varying density $p_t(x) \propto \mathbb{P}(f_t(x) \geq f_t(x'), \forall x' \in D_t)$, LD updates requires computing $\nabla \log p_t(x)$.

Algorithm 2: [MCMC routine] with Metropolis-Hastings

Input: GP posterior $\mathrm{GP}\left(\mu_t(x), k_t(x, x')\right)$, point x_o, proposal distribution $q(\cdot \mid x) := \mathcal{N}(0, \sigma)$ with parameter $\Theta = \{\sigma\}$.

1 Sample $u \sim \mathrm{Unif}[0, 1]$

2 Sample $x_p \sim x_o + \mathcal{N}(0, \Theta)$

3 **if** $u \geq \min\left\{1, \frac{\mathbb{P}\left(f_t(x_p) \geq f_t(x_o)\right)q(x_p|x_o)}{\mathbb{P}\left(f_t(x_o) \geq f_t(x_p)\right)q(x_o|x_p)}\right\}$ **then**

4 $\quad\mid\quad x_p \leftarrow x_o$ // Reject the transition

5 **end**

Algorithm 3: [MCMC routine] with Langevin dynamics

Input: GP posterior $\mathrm{GP}\left(\mu_t(x), k_t(x, x')\right)$, point x, $\Theta = \{\epsilon\}$ as Langevin transition step.

1 Sample $z \sim \mathcal{N}(0, 1)$

2 Estimate $\nabla \log p_t(x)$ using (3)

3 $x \leftarrow x + \epsilon \cdot \nabla \log p_t(x) + \sqrt{2\epsilon} \cdot z$

Similar to MH's acceptance ratio, we propose a simple estimate using the derivative of the GP posterior's mean and covariance.

In our proposed simplification for MCMC-BO, we utilize MH to calculate the ratio between the "winning probabilities" of two discretization points (the probability of being $\arg\max$). This calculation is done specifically with respect to the posterior distribution on those two points, rather

than the entire domain. The term $\nabla \log p_t(x)$ easily incorporates this simplified ratio as: $\frac{\partial \log p_t(x)}{\partial x_i} = \frac{\frac{\partial p_t(x)}{\partial x_i}}{p_t(x)} = \lim_{h \to 0} \frac{\frac{p_t(x + e_i \cdot h) - p_t(x)}{h}}{p_t(x)} = \lim_{h \to 0} \frac{1}{h}\left(\frac{p_t(x + e_i h)}{p_t(x)} - 1\right)$Thus, the approximation of the log-likelihood in LD has the following equation:

$$\frac{\partial \log p_t(x)}{\partial x_i} \approx \lim_{h \to 0} \frac{1}{h}\left(\frac{\mathbb{P}\big(f_t(x + e_i h) > f_t(x)\big)}{\mathbb{P}\big(f_t(x) > f_t(x + e_i h)\big)} - 1\right) \approx \frac{1}{h}\left(\frac{\mathbf{p}_i(x, h)}{1 - \mathbf{p}_i(x, h)} - 1\right), \qquad (3)$$

where $\mathbf{p}_i(x, h) = \mathbb{P}\big(f_t(x + e_i h) > f_t(x)\big) = \Phi(\frac{c^T \mu}{\sqrt{c^T \Sigma c}})$. As in MH, this involves calculating the cumulative distribution function (CDF) of a bivariate Gaussian, with the difference being that $x_p \leftarrow x + e_i h$ and $x_o \leftarrow x$. Using this numerical differentiation of the likelihood ratio, we obtain the gradient of $\log p_t(x)$ and the LD form of MCMC-BO.
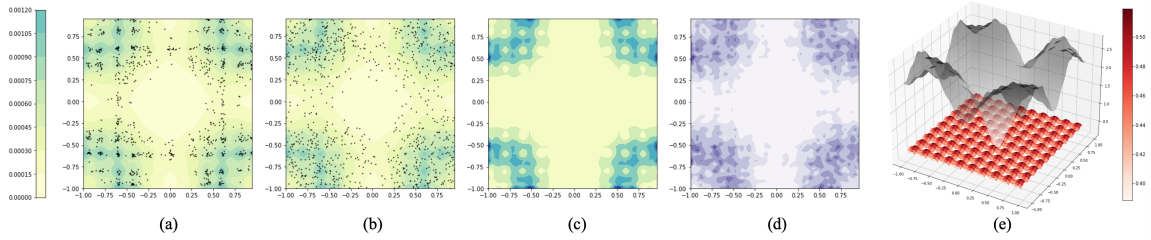

Figure 2: The figures are constructed from a 50×50 discretization of $D = [-1, 1]^2$. **(a)(b)** The stationary distribution achieved by the MH and Langevin version of MCMC-BO, respectively, and the congregated points obtained after convergence of the transition process from current GP information. **(c)** TS distribution simulated using Monte Carlo. **(d)** standard deviation of TS distribution over 10 trials of 10^6 samples. **(e)** GP posterior with surfaces being μ and Σ on which MCMC-BO transitions are performed.

5. Convergence Guarantee

We provide convergence guarantees to our proposed algorithmic framework and derive the subsequent regret bounds. Our proposed algorithm removes the dependence on the size of the discretization set; whereas the original GP-TS setting requires that to achieve the theoretical upper bound on regret. As we only track a batch of m points at a time, the algorithm complexity for MCMC-BO only depends on m and does not scale with the size of the discretization set $|D_t|$.

Notation and assumption. Let $\{x_t\}_{t=1}^{\infty}$ be an $\mathbb{R}^d$-valued discrete-time stochastic process adapted to filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Denote $x^* = \arg\max_{x \in D} f(x)$, then the cumulative regret over the horizon T is $R_T = \sum_{t=1}^{T} r_t$, where $r_t = f(x^*) - f(x_t)$. We also follow the convention and denote the maximum information gain at time t as γ_t. Further notation is attached in Appendix B.

Leveraging the properties of Reproducing Kernel Hilbert Space(RKHS), we make a common assumption as in Chowdhury and Gopalan (2017), that f is Lipschitz constant with constant $B \cdot L$, where L is a constant associated with Kernel $k(\cdot, \cdot)$.[1] In our analysis, we work with a discretization D_t where $|D_t|$ is finite. We denote $[x]_t := \arg\min_{x' \in D_t} \|x - x'\|_2$. For convex and compact domain $D \subset [0, r]^d$, the size constraint $|D_t| = \big(BLrdt^2\big)^d$ guarantees that for every $x \in D$, $|f(x) - f([x]_t)| \leq \|f\|_k L \|x - [x]_t\|_1 \leq 1/t^2$. Note that the size of the discretization is only

---

1. Further detail of assumption can be seen in Appendix B

needed in the theoretical analysis and influences the regret bound, but the algorithm complexity is independent of it. Additionally, the transition kernel of MCMC is defined as $P(x, y) = Q(x, y)\alpha(x, y), \forall x, y \in D_t$, where Q is the proposed distribution and α is the acceptance rate.

5.1. Overview of Proof

First, we show that the proposed approximated ratio of TS has a stationary distribution. Both MCMC algorithm (MH and LD) can converge to the stationary distribution with certain steps. Then the stationary distribution, though deviated from that of Thompson sampling, still inherits its good properties. Thus, by defining a benign set, we can prove that the probability that the sampling points located in the benign set grows with $O(1 - \frac{1}{T^2})$. An overview is presented here, and the detailed proof is deferred to Appendix C.

Lemma 1 *The proposed approximated posterior $P(x, \cdot)$, according to Alg.2 and Equation.2 does not yield a reversible Markov chain, but it still has a stationary distribution $\pi(x)$.*

In Fig 2 we demonstrate a 2D version of the stationary distribution of MCMC-BO . Note that the acceptance probability our algorithm proposed is inspired by TS but does not lead to the same exact stationary distribution of TS. Furthermore, we show that the approximated probability $P(x, \cdot)$ converges to the stationary distribution. In Sec 6 we also show that within a few hundred steps, the effect is significant enough. To obtain a bound on the overall regret, we begin by decomposing the instantaneous regret r_t into two parts. From the choice of discretization sets D_t, r_t depends on $[x^\star]_t$, which is the closest point to $x^\star$ in D_t. Therefore, we have that $r_t = \big(f(x^\star) - f([x^\star]_t)\big) + \big(f([x^\star]_t) - f(x_t)\big)$, where the difference $f(x^\star) - f([x^\star]_t)$ is bounded by $1/t^2$ by the regularity assumption. We proceed to bound $f([x^\star]_t) - f(x_t)$, which depends on the selected action x_t.

Regret increases when the approximation deviates from the underlying f. Thus, we hope that the function values of selected points x_t and the approximation at each time step t are not too far away. We define an event $E^f(t)$ and a benign set G_t in Appendix C for the convenience of Lemma 2. The benign points set bound the difference between function value and at the $t - 1$'s round GP's mean value.

Lemma 2 *For any filtration such that $E^f(t)$ is true, the transition probability from $[x^*]_t$ to any malignant point is bounded by*

$$P([x^*]_t, x') \le exp(-(\frac{c_t}{2} - v_t)^2)(1 - t^2), \quad \forall x' \in D_t \backslash G_t \tag{4}$$

Lemma 2 shows that transiting on the stationary distribution $\pi(x)$, the probability of choosing a malignant points is small. With Lemma 2, we can further show that the probability of playing any action from $D_t \backslash G_t$ is small; the regret from undesirable action is then bounded with Theorem 5 in Appendix B. By giving the bound of the transition probability between the best point and the malignant points, the total regret can then be bounded in terms of standard deviation of the chosen actions $\sigma_{t-1}(x_t)$. With the information theoretic lemma (Lemma 3 in Appendix) we can then upper bound $\sum_{t=1}^{T} \sigma_{t-1}(x_t)$ with $O\left(\sqrt{T\gamma_T}\right)$. As the difference $f([x^\star]_t) - f(x_t)$ in the decomposition of r_t is dependent on $\sigma_{t-1}(x_t)$, we finally arrive at the regret bound.

Theorem 3 *Let $\delta \in (0, 1), D \subset [0, r]^d$ be compact and convex, $\|f\|_k \le B, \{\varepsilon_t\}_t$ a conditionally R-sub-Gaussian sequence. For any T, we let $N(T)$ to be the transition number for each round.*

Running MCMC-BO *on a function f lying in the RKHS $H_k(D)$ for $N(T)$ transitions per round and with decision sets D_t chosen as above, with probability at least $1 - \delta$, the regret of* MCMC-BO *satisfies* $R_T = O\left(\sqrt{(\gamma_T + \ln(2/\delta))\, d\ln(BdT)}\left(\sqrt{T\gamma_T} + B\sqrt{T\ln(2/\delta)}\right)\right)$.

6. Experiments

In this section, we show the superiority of MCMC-BO on both high-dimensional synthetic functions and Mujoco tasks. We also provide the ablation study for the algorithm in Table **??** in Appendix. We further validate that our proposed method itself is an effective GP-based BO algorithm as other MAB algorithms. We compare MCMC-BO to state-of-the-art baselines of high-dimensional BO algorithms and EA algorithms.

For all BO algorithms, we utilize Thompson sampling to sample batches in each iteration and employ a scrambled Sobolev sequence to discretize the continuous search domain. The performance of MCMC-BO is evaluated against TuRBO and LA-MCTS, which serve as its BO components. The performance figures illustrate the mean performance of the algorithms with one standard error.
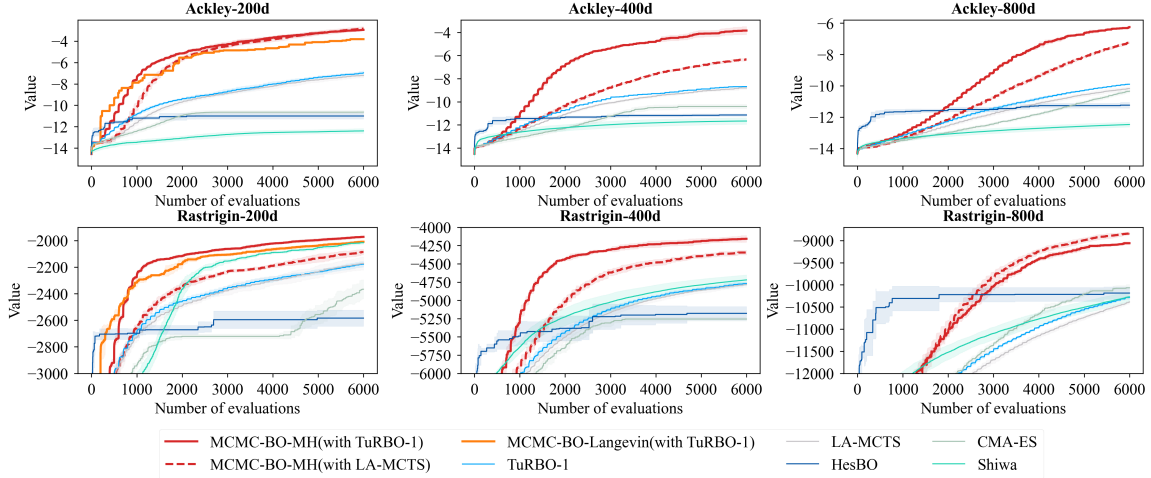

Figure 3: Optimization over high-dimensional synthetic functions.

6.1. High-Dimensional Synthetic Functions

We selected two popular synthetic problems, Ackley and Rastrigin, to optimize. For the function dimension, we chose $d = 200, 400$, and 800 to evaluate the performance in high-dimensional problems. We set the transition number to be the same as the function dimension. However, for the Langevin version of MCMC-BO, we ran only the $200d$ functions due to the lack of backward information for the posterior in the Gpytorch framework. All problems started with 200 initial points, and a batch of 100 points was sampled in each iteration.

Figure 4: Mujoco locomotion tasks.

Fig 3 suggests that MCMC-BO consistently outperforms other baselines on all functions. In higher dimensions $d = 800$, uniform discretization can not support good exploitation in such dimension, and TuRBO and LA-MCTS degenerate to the same level of performance as EA algorithms.
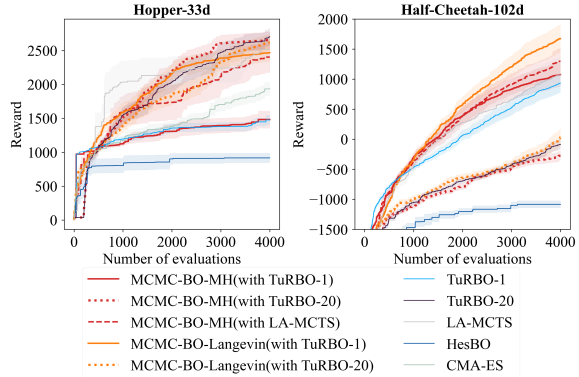
9

MCMC-BO achieves the best performance in all selected dimensions by allocating limited action points to more promising regions.

6.2. Mujoco Locomotion Task

Mujoco locomotion tasks are widely used benchmarks for reinforcement learning algorithms (Todorov et al., 2012). In our evaluation, we focus on Hopper and Half-Cheetah task, which have state spaces of dimensions 33 and 102, respectively. To assess the performance of the sampling-based algorithms, we optimize a linear policy: $\mathbf{a} = \mathbf{Ws}$ (Mania et al., 2018) , where the elements of the parameter matrix $\mathbf{W}$ are continuous and range from $[-1, 1]$. The reward is computed over 10 episodes for each policy proposal. Both tasks start with 200 initial points and sample a batch of 50 points in each iteration. We set the transition number to 200 on both tasks. Fig 6.1 shows the optimization performance of all algorithms. In the Hopper task, MCMC-BOwith TuRBO-20 algorithm converges faster than the original TuRBO-20. In higher dimensional Half-Cheetah task, MCMC-BO with TuRBO-1 still outperforms other baselines.

6.3. Performance on low-dimensional problems

We compare MCMC-BO with two popular BO algorithms, GP-TS and GP-UCB, on 1-2d synthetic functions. Figure 5 numerically demonstrates the convergence of MCMC-BO. We depict the distribution over the optimal value encoded as a violin plot, with horizontal bars 20% quantiles. GP-UCB only utilizes the diagonal information of the covariance matrix of $f(x_i)$, whereas TS uses the complete matrix information at once. As a compromise, MCMC-BO employs a divide-

Figure 5: We demonstrate the convergence of MCMC-BO compared with others on 1d and 2d functions.

and-conquer strategy by using a 2×2 sub-matrix at a time and still achieving an approximate stationary distribution of TS.

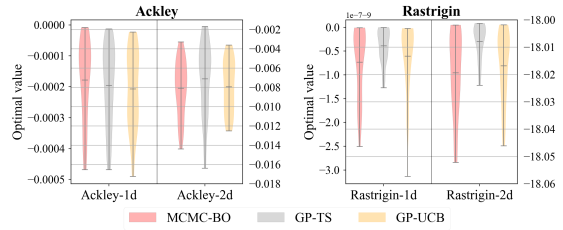## 7. Conclusion and Future Work

Sequential optimization in high-dimensional spaces has a profound impact on machine learning. In this paper, we propose MCMC-BO as a solution to improve the sample efficiency of high-dimensional BO. Our algorithm MCMC-BO offers versatile transitions to promising regions instead of maintaining a huge candidate set. MCMC-BO can be viewed as a GP-based bandit algorithm that yields an effective approximation for the induced TS distribution without the need to invert matrices of thousands of dimensions. We derive the regret bound of MCMC-BO under high-dimensional cases without memory overuse. We also conduct comprehensive evaluations to show that MCMC-BO can improve on existing popular high-dimensional BO baselines.

In future research, we aim to develop parallel computation mechanisms to further enhance computational speed. Additionally, implementing analytical backward computations holds the potential for significant acceleration. It is worth noting that MCMC-BO can transit in different irregular spaces. We also look forward to combining MCMC-BO with more complex space partition algorithms.

8. Acknowledgment

References

Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, pages 844–853. PMLR, 2017.

David Eriksson and Martin Jankowiak. High-dimensional bayesian optimization with sparse axis-aligned subspaces. In *Uncertainty in Artificial Intelligence*, pages 493–503. PMLR, 2021.

David Eriksson, Michael Pearce, Jacob Gardner, Ryan D Turner, and Matthias Poloczek. Scalable global optimization via local bayesian optimization. *Advances in neural information processing systems*, 32, 2019.

Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

Nikolaus Hansen, Sibylle D Müller, and Petros Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). *Evolutionary computation*, 11(1):1–18, 2003.

James Hensman, Nicolo Fusi, and Neil D Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

James Hensman, Alexander G Matthews, Maurizio Filippone, and Zoubin Ghahramani. Mcmc for variationally sparse gaussian processes. *Advances in Neural Information Processing Systems*, 28, 2015.

José Miguel Hernández-Lobato, James Requeima, Edward O Pyzer-Knapp, and Alán Aspuru-Guzik. Parallel and distributed thompson sampling for large-scale accelerated exploration of chemical space. In *International conference on machine learning*, pages 1470–1479. PMLR, 2017.

Cheng Li, Sunil Gupta, Santu Rana, Vu Nguyen, Svetha Venkatesh, and Alistair Shilton. High dimensional bayesian optimization using dropout. *arXiv preprint arXiv:1802.05400*, 2018.

Xiuyuan Lu and Benjamin Van Roy. Ensemble sampling. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Horia Mania, Aurelia Guy, and Benjamin Recht. Simple random search provides a competitive approach to reinforcement learning. *arXiv preprint arXiv:1803.07055*, 2018.

Eric Mazumdar, Aldo Pacchiano, Yian Ma, Michael Jordan, and Peter Bartlett. On approximate thompson sampling with Langevin algorithms. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 6797–6807. PMLR, 13–18 Jul 2020.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

Amin Nayebi, Alexander Munteanu, and Matthias Poloczek. A framework for bayesian optimization in embedded subspaces. In *International Conference on Machine Learning*, pages 4752–4761. PMLR, 2019.

ChangYong Oh, Efstratios Gavves, and Max Welling. Bock: Bayesian optimization with cylindrical kernels. In *International Conference on Machine Learning*, pages 3868–3877. PMLR, 2018.

Carlos Riquelme, G. Tucker, and Jasper Snoek. Deep bayesian bandits showdown: An empirical comparison of bayesian deep networks for thompson sampling. *ArXiv*, abs/1802.09127, 2018.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1): 148–175, 2015.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

Il'ya Meerovich Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pages 5026–5033. IEEE, 2012.

Iñigo Urteaga and Chris H Wiggins. Variational inference for the multi-armed contextual bandit. *ArXiv*, abs/1709.03163, 2017.

Linnan Wang, Rodrigo Fonseca, and Yuandong Tian. Learning search space partition for black-box optimization using monte carlo tree search. *Advances in Neural Information Processing Systems*, 33:19511–19522, 2020.

Zi Wang, Clement Gehring, Pushmeet Kohli, and Stefanie Jegelka. Batched large-scale bayesian optimization in high-dimensional spaces. In *International Conference on Artificial Intelligence and Statistics*, pages 745–754. PMLR, 2018.

Pan Xu, Hongkai Zheng, Eric V Mazumdar, Kamyar Azizzadenesheli, and Animashree Anandkumar. Langevin Monte Carlo for contextual bandits. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 24830–24850. PMLR, 17–23 Jul 2022.