No-Regret Algorithms for Time-Varying Bayesian Optimization

Xingyu Zhou
Wayne State University
Detroit, USA
Email: xingyu.zhou@wayne.edu

Ness Shroff
The Ohio State University
Columbus, USA
Email: shroff.11@osu.edu

Abstract—In this paper, we consider the time-varying Bayesian optimization problem. The unknown function at each time is assumed to lie in an RKHS (reproducing kernel Hilbert space) with a bounded norm. We adopt the general variation budget model to capture the time-varying environment, and the variation is characterized by the change of the RKHS norm. We adapt the restart and sliding window mechanism to introduce two GP-UCB type algorithms: R-GP-UCB and SW-GP-UCB, respectively. We derive the first (frequentist) regret guarantee on the dynamic regret for both algorithms. Our results not only recover previous linear bandit results when a linear kernel is used, but complement the previous regret analysis of time-varying Gaussian process bandit under a Bayesian-type regularity assumption, i.e., each function is a sample from a Gaussian process.

I. INTRODUCTION

We consider the online black-box optimization of an unknown function f with only bandit feedback. At each time t, one query point x_t is selected and the (noisy) corresponding evaluation (reward) y_t is then observed. The objective is to find the optimal point using a minimal number of trials. This model is ubiquitous in bandit learning problems under various assumptions. For instance, when the domain is a finite set with independent actions (points), this is the typical multi-arm bandit (MAB) problem [1]. If the unknown reward function is linear, this represents the standard linear bandit problem [2]. In Bayesian optimization, the unknown function could be arbitrary (e.g., non-linear and non-convex). Under a Bayesian-type regularity assumption, the unknown function is assumed to be a sample from a Gaussian process (and hence also often called Gaussian process bandit) [3]. Another popular regularity assumption is a frequentist-type view. That is, the unknown function is a fixed function in an RKHS with a bounded norm [4].

This versatile model of online decision making has found many successful real-life applications, such as recommendation system, dynamic pricing, and network resource allocation. The standard application of this model often assumes that the unknown function f is fixed, which is, however, typically not the case in real-life scenarios. For instance, in the dynamic pricing problem, f often represents the market environment, which definitely evolves with time. Another intriguing example is the dynamic resource allocation in wireless communications, where the channel conditions change with time due to fading.

To this end, non-stationary bandit learning has recently drawn significant interest. In MAB, a variety of works have studied either abrupt changes or slowly changing distributions [5], [6]. In the stochastic linear bandit setting, [7] takes the first attempt by considering a general *variation budget* model (i.e., $\sum_{t=1}^{T-1} \|\theta_t - \theta_{t+1}\| \le P_T$), which is able to capture both abruptly-changing and slowly-changing environments. Under this model, the authors propose a sliding window UCB-type algorithm. Later, under the same model, [8] proposes a weighted-UCB algorithm and [9] introduces a simple restarting UCB algorithm. Recently, [10] has pointed out a key common gap in the analysis of the three works above on linear bandits (i.e., [7]–[9]). After fixing the gap, the regret bound of all three algorithms is on the order of $\widetilde{O}\left(d^{7/8}T^{3/4}(1 + P_T)^{1/4}\right)$, where d is the dimension of the action space and P_T is the total variation budget. In the case of a general non-linear function, [11] considers the GP bandit setting where each function f_t is a sample from a GP (i.e., Bayesian-type regularity) and establishes both lower and upper bounds. However, Bayesian optimization under a frequentist-type regularity with time-varying functions remains untouched. In fact, it was stated as one interesting future direction to explore in [11].

Motivated by this, in this paper, we consider a time-varying Bayesian optimization under a frequentist-type regularity. That is, the smoothness of the unknown time-varying functions f_t is indicated by the corresponding RKHS norm. The time-varying environment is also captured by the RKHS norm by $\sum_{t=1}^{T-1} \|f_t - f_{t+1}\|_{\mathcal{H}} \le P_T$. Under this regularity assumption and time-varying model, we made the following contributions.

- We adapt the restarting-UCB and sliding window UCB algorithm in linear bandits to the GP ban-

dit setting. We call them R-GP-UCB[1] and SW-GP-UCB. We derive the first (frequentist) regret guarantee on the dynamic regret of both algorithms. For a known P_T, both enjoy regret bound $\widetilde{O}(\gamma_T^{7/8}(1+P_T)^{1/4}T^{3/4})$, and when P_T is unknown, the regret bound is $\widetilde{O}(\gamma_T^{7/8}(1+P_T)T^{3/4})$, where γ_T is the maximum information gain.

- Our regret bounds include previous linear bandit (e.g., [7]–[9]) results as special case by choosing a linear kernel for the RKHS, under which $\gamma_T = O(d \ln T)$. Moreover, our results also complement the regret bounds under the Bayesian-type regularity assumptions in [11].

II. PROBLEM STATEMENT AND PRELIMINARIES

We consider the time-varying Bayesian optimization using the sequential decision-making model. Under this model, in each time $t = 1, 2, \ldots$, the learning agent sequentially chooses a query point $x_t \in \mathcal{X} \subset \mathbb{R}^d$, and obtains a noisy evaluation

$$y_t = f_t(x_t) + \eta_t,$$

where $f_t : \mathcal{X} \to \mathbb{R}$ is the unknown function that could vary with time, and η_t is the zero-mean noise. In particular, the query point x_t is chosen based on query points and rewards up to time $t - 1$. The noise sequence $\{\eta_s\}_{s=1}^{\infty}$ is conditionally R-sub-Gaussian for a fixed $R \geq 0$, i.e.,

$$\forall t \geq 1, \ \forall \lambda \in \mathbb{R}, \ \mathbb{E}\left[e^{\lambda \eta_t} \mid \mathcal{F}_{t-1}\right] \leq \exp(\frac{\lambda^2 R^2}{2}),$$

where $\mathcal{F}_{t-1} = \sigma(\{x_\tau, y_\tau\}_{\tau=1}^{t-1}, x_t)$ is the σ-algebra generated by the events so far.

Regularity Assumptions. In contrast to [11], we consider a non-Bayesian regularity condition in this paper. That is, for each time t, f_t is allowed to be an arbitrary function in a RKHS with a bounded norm. Specifically, the RKHS is denoted by $\mathcal{H}_k(\mathcal{X})$, which is completely determined by the corresponding chosen kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. For any function $g \in \mathcal{H}_k(\mathcal{X})$, it satisfies the *reproducing property*: $g(x) = \langle g, k(\cdot, x)\rangle_{\mathcal{H}}$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product defined on $\mathcal{H}_k(\mathcal{X})$. We assume that $\|f_t\|_{\mathcal{H}} \leq B$ for all $t \geq 1$, and the domain $\mathcal{X}$ is compact. The kernel function k is a continuous kernel with respect to a finite Borel measure ν whose support is $\mathcal{X}$, and $k(x, x) \leq 1$. The assumptions hold for practically relevant kernels (cf. [12]).

Time-varying Assumptions. We assume that the total variation of f_t satisfies the following budget

$$\sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_{\mathcal{H}} \leq P_T.$$

One nice feature of this variation budget model is that

[1]Note that R-GP-UCB was first proposed in [11] and analyzed in the Bayesian regularity condition.

it includes both slowly-changing and abruptly-changing environments.

The objective is to minimize the following *dynamic regret* given the total variation budget

$$\mathcal{R}(T) := \sum_{t=1}^{T} \max_{x \in \mathcal{X}} f_t(x) - f_t(x_t),$$

which is the cumulative regret against the optimal strategy that has full information of unknown varying functions.

To design our learning strategy, we will adopt a surrogate model, which helps us to update the estimate and uncertainty of the underlying function after each new query. In particular, we will use a GP prior and Gaussian likelihood, which is specified below. Note that, this surrogate model is only used for the algorithm design. That is, it will not affect the fact that each f_t is an element in RKHS and the noise is allowed to be sub-Gaussian. This is often termed as the *agnostic* setting [3].

Surrogate Model. We use $\mathcal{GP}(0, k(\cdot, \cdot))$ as an initial prior on the unknown black-box function f_t, and a Gaussian likelihood with the noise variables η_t drawn independently across t from $\mathcal{N}(0, \lambda)$. Conditioned on a set of observations $\mathcal{H}_t = \{(x_s, y_s), s \in [t]\}$, by the properties of GPs [13], the posterior distribution over f is $\mathcal{GP}(\mu_t(\cdot), k_t(\cdot, \cdot))$, where

$$\mu_t(x) = k_t(x)^T (K_t + \lambda I)^{-1} Y_t \quad (1)$$
$$k_t(x, x') = k(x, x') - k_t(x)^T (K_t + \lambda I)^{-1} k_t(x')$$
$$\sigma_t^2(x) = k_t(x, x), \quad (2)$$

in which Y_t is the reward vector $[y_1, y_2, \ldots, y_t]^T$, $k_t(x) = [k(x_1, x), \ldots, k(x_t, x)]^T$, and $K_t = [k(u, v)]_{u, v \in \mathcal{H}_t}$. Therefore, for every $x \in \mathcal{X}$, the posterior distribution of $f(x)$, given $\mathcal{H}_t$ is $\mathcal{N}(\mu_t(x), \sigma_t^2(x))$. The following term (named maximum information gain) often plays a key role in the regret bounds of GP based algorithms.

$$\gamma_t := \gamma_t(k, \mathcal{X}) = \max_{A \subset \mathcal{X} : |A| = t} \frac{1}{2} \ln |I_t + \lambda^{-1} K_A|, \quad (3)$$

where $K_A = [k(x, x')]_{x, x' \in A}$. It is a function of the kernel k and domain $\mathcal{X}$. For instance, if $\mathcal{X}$ is compact and convex, then we have $\gamma_t = O((\ln t)^{d+1})$ for k_{SE}, $O(t^{\frac{d(d+1)}{2\nu+d(d+1)}} \ln t)$ for $k_{\text{Matérn}}$, and $O(d \ln t)$ for a linear kernel [3].

III. MAIN RESULTS

In this section, we first introduce two Bayesian optimization algorithms based on GP-UCB for time-varying environments, followed by the corresponding upper bounds on the dynamic regret.

A. Algorithms

The two algorithms basically implement the restart and sliding window mechanism in the GP setting. Specifi-

Algorithm 1: R-GP-UCB

Input: Prior $\mathcal{GP}(0, k)$, parameters B, R, λ, δ, reset interval H

1. Initialization: $\beta_0 = 0$, $\mu(x) = 0$ and $\sigma_0(x) = 0 \ \forall x \in \mathcal{X}$
2. **for** $t = 1, 2, 3, \ldots, T$ **do**
3.   **if** $t \bmod H = 1$ **then**
4.   Reset to the initialization state, i.e., $t_0 = t$
5.   set $\beta_t = B + R\sqrt{2(\gamma_{t-t_0} + 1 + \ln(1/\delta))}$
6. $x_t = \arg\max_{x \in \mathcal{X}} \mu_{t-1}(x) + \beta_t \sigma_{t-1}(x)$
7.   choose x_t, observe reward y_t
8.   Use only the samples since t_0 to update μ_t and σ_t via (1) and (2)

Algorithm 2: SW-GP-UCB

Input: Prior $\mathcal{GP}(0, k)$, parameters B, R, λ, δ, window size w.

1. Initialization: $\beta_0 = 0$, $\mu(x) = 0$ and $\sigma_0(x) = 0 \ \forall x \in \mathcal{X}$
2. **for** $t = 1, 2, 3, \ldots, T$ **do**
3.   set $\beta_t = B + R\sqrt{2(\gamma_{t \wedge w} + 1 + \ln(1/\delta))}$
4. $x_t = \arg\max_{x \in \mathcal{X}} \mu_{t-1}(x) + \beta_t \sigma_{t-1}(x)$
5.   choose x_t, observe reward y_t
6.   Use only the samples between $t_0 = 1 \vee (t - w)$ and t to update μ_t and σ_t via (1) and (2)

cally, the first algorithm (Algorithm 1) is simply restart a GP-UCB type algorithm ((i.e., IGP-UCB in [4])) every H time steps, named R-GP-UCB. The intuition is clear: since the environment is chaning, we might need to discard our old estimates and restart to build our new estimates. The second algorithm (Algorithm 2) is based on the sliding window technique. That is, it only uses the most recent w samples for estimations.

As we mentioned before, the design ideas behind both algorithms are not new as they have been considered in the linear bandit or the Gaussian process bandit setting. Our key contribution is to derive the first (frequentist) regret guarantee on the dynamic regret in the agnostic GP setting.

B. Regret Bounds

In this section, we show that with a proper choice of the algorithm parameters (reset interval H and sliding window size w), both algorithms can achieve sub-linear dynamic regrets.

Theorem 1. *R-GP-UCB with a reset interval H achieves a high probability regret bound for probability parameter $\delta \in (0, 1)$,*

$$\mathcal{R}(T) = O\left(\sqrt{\gamma_H} H^{3/2} P_T + \beta(\delta) T \sqrt{\frac{\gamma_H}{H}}\right),$$

where $\beta(\delta) = \left(B + \frac{1}{\sqrt{\lambda}} R\sqrt{2\gamma_H + 2\ln(1/\delta)}\right)$.

Theorem 2. *SW-GP-UCB with a window size w achieves a high probability regret bound for probability parameter $\delta \in (0, 1)$,*

$$\mathcal{R}(T) = O\left(\sqrt{\gamma_w} w^{3/2} P_T + \beta_T(\delta) T \sqrt{\frac{\gamma_w}{w}}\right),$$

$\beta_T(\delta) = \left(B + \frac{1}{\sqrt{\lambda}} R\sqrt{2\gamma_w + 2\ln(T/\delta)}\right)$.

Remark. (1) From Theorems 1 and 2, we can see R-GP-UCB and SW-GP-UCB share similar regret bounds, with the difference being the β term. The additional T factor in $\beta_T(\delta)$ comes from the information loss due to the

sliding window. (2) In the case of a known P_T, one can set H(or w) $= \widetilde{O}(\gamma_T^{1/4}(T/P_T)^{1/2})$ to achieve a dynamic regret $\widetilde{O}(\gamma_T^{7/8}(1 + P_T)^{1/4} T^{3/4})$. (3) For an unknown P_T, one can set H(or w) $= \widetilde{O}(\gamma_T^{1/4} T^{1/2})$ to achieve a dynamic regret $\widetilde{O}(\gamma_T^{7/8}(P_T + 1) T^{3/4})$. (4) These results directly recover the results for linear bandits by choosing a linear kernel, the γ_T of which is $O(d \ln T)$.

IV. Proofs of Theorems

In this section, we present the proofs for Theorems 1 and 2. Compared to previous works on linear bandits, the main challenge of our proofs is that the feature map associated with the kernel function could be infinite dimension. As a result, previous finite-dimension results cannot be directly applied since the regret bounds there grow to infinity with d. To this end, one possibility is resort to the general operator theory and establish results that hold for a general separable Hilbert space as in [14]. Here, instead we will directly focus on RKHS and derive our results with an explicit (infinite) feature map enabled by Mercer's theorem. Note that, even this explicit feature map is not necessary (e.g., one can directly use $k(\cdot, x)$ as an implicit feature map as in [4], [15]). We choose to directly focus on RKHS and use the explicit feature map because it directly helps to reveal the connections between linear bandits and Gaussian process bandits.

A. Mercer Representation

The following version of Mercer's theorem is adapted from Theorem 4.1 and 4.2 in [16], which roughly says that the kernel function can be expressed in terms of the eigenvalues and eigenfunctions under mild conditions.

Theorem 3. *Let $\mathcal{X}$ be a compact metric space, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a continuous kernel with respect to a finite Borel measure ν whose support is $\mathcal{X}$. Then, there is an at most countable sequence $(\lambda_i, \phi_i)_{i \in \mathbb{N}}$, where $\lambda_i \geq 0$ and $\lim_{i \to \infty} \lambda_i = 0$ and $\{\phi_i\}$ forms an an orthonormal*

basis of $L_{2,\nu}(\mathcal{X})$, such that

$$k(x, x') = \sum_{i \in \mathbb{N}} \lambda_i \phi_i(x) \phi_i(x'), \quad x, x' \in \mathcal{X},$$

where the convergence is absolute and uniform over $x, x' \in \mathcal{X}$. Further, the RKHS of k is given by

$$\mathcal{H} = \left\{ f = \sum_{i \in \mathbb{N}} \theta_i \sqrt{\lambda_i} \phi_i : \|f\|_{\mathcal{H}} := \sum_{i \in \mathbb{N}} \theta_i^2 < \infty \right\},$$

and the inner-product is given by $\langle f, g \rangle_{\mathcal{H}} = \sum_{i \in \mathcal{N}} \alpha_i \beta_i$, for $f = \sum_{i \in \mathcal{N}} \alpha_i \sqrt{\lambda_i} \phi_i$ and $f = \sum_{i \in \mathcal{N}} \beta_i \sqrt{\lambda_i} \phi_i$.

Based on this result, we can explicitly define a feature map as $\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots)$ where $\varphi_i = \sqrt{\lambda_i} \phi_i$. Given a $\theta = (\theta_1, \theta_2, \dots)$, we denote $\theta^T \varphi(x) = \langle \theta, \varphi(x) \rangle$ for $\sum_{i \in \mathcal{N}} \theta_i \varphi_i(x)$ similar to the finite case. Therefore, we have $k(x, y) = \varphi(x)^T \varphi(y)$, and more importantly, $f(x) = \theta^T \varphi(x)$ for $f = \sum_{i \in \mathcal{N}} \theta_i \varphi_i$ and $\|f\|_{\mathcal{H}}^2 = \|\theta\|^2 := \sum_{i \in \mathcal{N}} \theta_i^2$. From this, one can easily see that the linear bandit in the finite dimension is just a special case of this general formulation.

We will also introduce some basic notations by using the feature map. For any $f_s \in \mathcal{H}$, we let θ_s denotes its corresponding parameter such that $\|f\|_{\mathcal{H}} = \|\theta_s\|_2$. Given a set $\{x_1, x_2, \dots, x_t\} \in \mathcal{X}$, we define a $t \times \infty$ matrix[2] Φ_t such that $\Phi_t^T = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_t))$. We also introduce the (infinite) design matrix $V_t = \Phi_t^T \Phi_t + \lambda I$ and noise vector $N_t = (\eta_1, \dots, \eta_t)$. For a positive definite matrix V, We also define the inner product $\langle \cdot, \cdot \rangle_V := \langle \cdot, V \cdot \rangle$ with the corresponding norm as $\|\cdot\|_V$.

B. Proof of Theorem 1

In this section, we present the proof for Theorem 1. Comparing this proof with the linear bandit case, one can gain more insight on the connections between them.

We will first establish the following bound on the estimate μ_t in terms of σ_t under R-GP-UCB.

Lemma 1. *Let t_0 be the most recent restart time before a given time $t \geq t_0$, then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, the following holds for any $x \in \mathcal{X}$ and any $t \geq t_0$,*

$$|\mu_{t-1}(x) - f_t(x)| \leq \frac{1}{\lambda} \sqrt{H 2(1 + \lambda) \gamma_H} \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}}$$
$$+ \beta_t \sigma_{t-1}(x),$$

where $\beta_t = \left(B + \frac{1}{\sqrt{\lambda}} R \sqrt{2\gamma_{t-t_0} + 2\ln(1/\delta)} \right)$.

Now, let $x_t^* = \max_{x \in \mathcal{X}} f_t(x)$ and $\xi_H := \frac{1}{\lambda} \sqrt{H 2(1 + \lambda) \gamma_H}$. Then, based on Lemma 1, for any $\delta \in (0, 1)$, we have with probability at least $1 - \delta$,

$$r_t = f_t(x_t^*) - f_t(x_t)$$

[2]Note that, in most cases of our proof, we can safely use linear algebra for the infinite matrices that come up, since all of them are compact operators (and hence a well-established spectral theory). Some are even Hilbert-Schmidt or trace-class operators, which can be arbitrarily approximated by sufficient 'large' matrices, see [17].

$$\overset{(a)}{\leq} \mu_{t-1}(x_t^*) + \xi_H \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}} + \beta_t \sigma_{t-1}(x_t^*) - f_t(x_t)$$

$$\overset{(b)}{\leq} \mu_{t-1}(x_t) + \xi_H \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}} + \beta_t \sigma_{t-1}(x_t) - f_t(x_t)$$

$$\overset{(c)}{\leq} 2\xi_H \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}} + 2\beta_t \sigma_{t-1}(x_t)$$

$$\overset{(d)}{=} 2\xi_H \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}} + 2\sqrt{\lambda} \beta_t \|\varphi(x_t)\|_{V_{t-1}^{-1}}$$

$$\overset{(e)}{\leq} 2\xi_H \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}} + 2\sqrt{\lambda} \beta(\delta) \|\varphi(x_t)\|_{V_{t-1}^{-1}}$$

where (a) and (c) follow from Lemma 1; (b) follows from the UCB-type algorithm; (d) holds by Claim 1, where $V_t = \lambda I + \Phi_t^T \Phi_t$ and $\Phi_t^T = (\varphi(x_{t_0}), \dots, \varphi(x_t))$, t_0 is the most recent rest time slot before t; in (e), $\beta(\delta) = \left(B + \frac{1}{\sqrt{\lambda}} R \sqrt{2\gamma_H + 2\ln(1/\delta)} \right)$. Thus, we have the following regret bound,

$$\mathcal{R}(T) \leq 2H^{3/2} P_T \sqrt{\gamma_H} + \sqrt{\lambda} \beta(\delta) \sum_{t=1}^{T} \|\varphi(x_t)\|_{V_{t-1}^{-1}}.$$

The sum $\sum_{t=1}^{T} \|\varphi(x_t)\|_{V_t^{-1}}$ needs further analysis. Note that we can divide the time horizon into blocks of size H, and reset starts at the beginning of each block.

$$\sum_{t=1}^{T} \|\varphi(x_t)\|_{V_t^{-1}} \leq \sum_{k=0}^{T/H-1} \sum_{t=kH+1}^{(k+1)H} \|\varphi(x_t)\|_{V_{t-1}^{-1}},$$

where for all $t \in [kH + 1, (k + 1)H]$, $t_0 = kH + 1$. Thus, for each block, we have by Lemma 4 in [4],

$$\sqrt{\lambda} \sum_{t=kH+1}^{(k+1)H} \|\varphi(x_t)\|_{V_{t-1}^{-1}} \leq \sqrt{4(H + 2)\gamma_H}.$$

Putting everything together, yields $\mathcal{R}(T) = O\left(\sqrt{\gamma_H} H^{3/2} P_T + \beta(\delta) T \sqrt{\frac{\gamma_H}{H}} \right)$.

Proof of Lemma 1. By a slight abuse of notation, here $k_t(x) = [k(x_{t_0}, x), \dots, k(x_t, x)]^T$ and $K_t = [k(x_u, x_v)]_{u,v \in \{t_0, \dots, t\}}$. That is, we only use the samples starting from the most recent reset to calculate μ_t and σ_t. Let $\Phi_t^T = (\varphi(x_{t_0}), \dots, \varphi(x_t))$, then the mean estimate $\mu_t(x)$ can be rewritten as

$$\mu_t(x) = k_t(x)^T (K_t + \lambda I)^{-1} Y_t$$
$$= [\Phi_t \varphi(x)]^T (\Phi_t \Phi_t^T + \lambda)^{-1} Y_t$$
$$= \langle \varphi(x), (\Phi^T \Phi_t + \lambda I)^{-1} \Phi_t^T Y_t \rangle,$$

where the last equality we have used the fact $(A^T A + \lambda I)^{-1} A^T = A^T (AA^T + \lambda I)^{-1}$, which holds by Lemma 3 of [18]. Recall that $f_t(x) = \langle \theta_t, \varphi(x) \rangle$ where θ_t is the parameters of f_t. Combining this with the result above, yields $|f_t(x) - \mu_{t-1}(x)| = |\langle \varphi(x), \theta_t - \hat{\theta}_{t-1} \rangle|$, where $\hat{\theta}_{t-1} = V_{t-1}^{-1} \Phi_{t-1}^T Y_{t-1}$. Note that, the term $\hat{\theta}_{t-1} - \theta_t$

can be rewritten as

$$\hat{\theta}_{t-1} - \theta_t$$
$$= V_{t-1}^{-1}\left(\sum_{s=t_0}^{t-1}\varphi(x_s)\varphi(x_s)^T(\theta_s - \theta_t)\right) + V_{t-1}^{-1}\Phi_{t-1}^T N_{t-1}$$
$$- \lambda V_{t-1}^{-1}\theta_t.$$

Thus, by the triangle inequality, we have

$$|f_t(x) - \mu_{t-1}(x)|$$
$$\leq \left|\langle\varphi(x), \Phi_{t-1}^T N_{t-1}\rangle_{V_{t-1}^{-1}}\right| + \lambda\left|\langle\varphi(x), \theta_t\rangle_{V_{t-1}^{-1}}\right| \quad (4)$$
$$+ \left|\langle\varphi(x), V_{t-1}^{-1}\left(\sum_{s=t_0}^{t-1}\varphi(x_s)\varphi(x_s)^T(\theta_s - \theta_t)\right)\rangle\right|. \quad (5)$$

From the above, we can see (5) is the additional term due to time-varying environments. We now turn to bound each term, respectively. The second term in (4) can be easily bounded under the boundedness assumptions.

$$\lambda\left|\langle\varphi(x), \theta_t\rangle_{V_{t-1}^{-1}}\right| \leq \lambda\|\varphi(x)\|_{V_{t-1}^{-1}}\left\|V_{t-1}^{-1/2}\theta_t\right\|$$
$$\leq \sqrt{\lambda}B\|\varphi(x)\|_{V_{t-1}^{-1}}, \quad (6)$$

where the last inequality follows from $V_{t-1}^{-1} \preceq \lambda^{-1}I$, and $\|\theta_t\| = \|f_t\|_{\mathcal{H}} \leq B$. For the first term in (4), we can bound it by using the RKHS-valued self-normalized inequality (e.g., Lemma 7 in [18]). First, let $\widetilde{V}_{t-1} = V_{t-1}/\lambda$, and hence we have[3]

$$\left|\langle\varphi(x), \Phi_{t-1}^T N_{t-1}\rangle_{V_{t-1}^{-1}}\right|$$
$$= \frac{1}{\lambda}\|\varphi(x)\|_{\widetilde{V}_{t-1}^{-1}}\left\|\sum_{s=t_0}^{t-1}\eta_s\varphi(x_s)\right\|_{\widetilde{V}_{t-1}^{-1}}$$
$$\leq \frac{1}{\lambda}\|\varphi(x)\|_{\widetilde{V}_{t-1}^{-1}}\sqrt{2R^2\lambda\ln\left(\det(\widetilde{V}_{t-1})^{1/2}/\delta\right)}. \quad (7)$$

Then, by applying Claim 1 below to (6) and (7), we can upper bound (4) as

$$(4) \leq \sigma_{t-1}(x)\left(B + \frac{1}{\sqrt{\lambda}}R\sqrt{2\gamma_{t-t_0} + 2\ln(1/\delta)}\right). \quad (8)$$

Claim 1. *The following equations hold for all $t \geq t_0$.*
$$\lambda\|\varphi(x)\|_{V_t^{-1}}^2 = \sigma_t^2(x)$$
$$\|\varphi(x)\|_{\widetilde{V}_t^{-1}}^2 = \sigma_t^2(x)$$
$$\ln(\det(\widetilde{V}_t)) \leq 2\gamma_{t-t_0} \leq 2\gamma_H.$$

Now, to bound (5), we will follow similar arguments in [7]. In particular, we have

$$(5) \overset{(a)}{\leq} \left\|V_{t-1}^{-1}\left(\sum_{s=t_0}^{t-1}\varphi(x_s)\varphi(x_s)^T(\theta_s - \theta_t)\right)\right\|$$

[3]Note that the main purpose of defining $\widetilde{V}_{t-1}$ here is to make sure that $\det(\widetilde{V}_{t-1})$ is well-defined. This is because $\frac{1}{\lambda}\Phi_t^T\Phi_t$ is a trace-class operator, and hence $\det(I + \frac{1}{\lambda}\Phi_t^T\Phi_t)$ is well defined via Fredholm determinant.

$$= \left\|V_{t-1}^{-1}\left(\sum_{p=t_0}^{t-1}\sum_{s=t_0}^{p}\varphi(x_s)\varphi(x_s)^T(\theta_p - \theta_{p+1})\right)\right\|$$
$$\leq \sum_{p=t_0}^{t-1}\left\|V_{t-1}^{-1}\sum_{s=t_0}^{p}\varphi(x_s)\varphi(x_s)^T(\theta_p - \theta_{p+1})\right\|$$
$$\overset{(b)}{\leq} \frac{1}{\lambda}\sqrt{H2(1+\lambda)\gamma_H}\sum_{p=t_0}^{t-1}\|f_p - f_{p+1}\|_{\mathcal{H}}, \quad (9)$$

where (a) follows from $\|\varphi(x)\| = \sqrt{k(x,x)} \leq 1$; (b) follows from the following claim and the fact that $\|\theta_p - \theta_{p+1}\| = \|f_p - f_{p+1}\|_{\mathcal{H}}$.

Claim 2. *For any $t_0 \leq p \leq t - 1$, the operator norm satisfies*

$$\left\|V_{t-1}^{-1}\sum_{s=t_0}^{p}\varphi(x_s)\varphi(x_s)^T\right\| \leq \frac{1}{\lambda}\sqrt{2H(1+\lambda)\gamma_H}.$$

Combining the bounds in (8) and (9), yields the final result of Lemma 1. We are left to present proofs for the claims. The results in the first claim are standard (cf. Appendix in [18]). We give a proof for Claim 2.

Proof of Claim 2. Previous works (including [7]–[9]) all bound the LHS of Claim 2 by one, which is not true as shown in [10]. To address this gap, we follow the key idea in [10] with slight modifications in step (c) below to handle the possibly infinite dimension in our case.

Denote the unit ball $\mathcal{B}(1) = \{z|\|z\| = 1\}$. Then, we have

$$\left\|V_{t-1}^{-1}\sum_{s=t_0}^{p}\varphi(x_s)\varphi(x_s)^T\right\|$$
$$= \sup_{z\in\mathcal{B}(1)}\left|z^T V_{t-1}^{-1}\left(\sum_{s=t_0}^{p}\varphi(x_s)\varphi(x_s)^T\right)z\right|$$
$$\overset{(a)}{\leq} \|z^*\|_{V_{t-1}^{-1}}\left\|\left(\sum_{s=t_0}^{p}\varphi(x_s)\varphi(x_s)^T\right)z\right\|_{V_{t-1}^{-1}}$$
$$\leq \|z^*\|_{V_{t-1}^{-1}}\left\|\sum_{s=t_0}^{p}\varphi(x_s)\|\varphi(x_s)\|\|z\|\right\|_{V_{t-1}^{-1}}$$
$$\leq \frac{1}{\sqrt{\lambda}}\left\|\sum_{s=t_0}^{p}\varphi(x_s)\right\|_{V_{t-1}^{-1}}$$
$$\overset{(b)}{\leq} \frac{1}{\sqrt{\lambda}}\sum_{s=t_0}^{p}\|\varphi(x_s)\|_{V_{t-1}^{-1}}$$
$$\overset{(c)}{\leq} \frac{1}{\lambda}\sum_{s=t_0}^{p}\sigma_{s-1}(x_s)$$
$$\overset{(d)}{\leq} \frac{1}{\lambda}\sqrt{H\sum_{s=t_0}^{p}\sigma_{s-1}^2(x_s)}$$
$$\overset{(e)}{\leq} \frac{1}{\lambda}\sqrt{H2(1+\lambda)\gamma_H}$$

where in (a) z^* is the optimizer; (b) holds by $\|x\|_{V_{t-1}^{-1}} \leq \|x\|/\sqrt{\lambda}$ and $\|z\| = 1$ and $\|\varphi(x_s)\| \leq 1$; (c) follows from the fact that $V_{t-1}^{-1} \preceq V_{s-1}^{-1}$ and Claim 1; (d) follows from Cauchy–Schwarz inequality since $p - t_0$ is at most H; (e) holds by $\sum_{s=t_0}^{p} \sigma_{t-1}^2(x_s) \leq 2(1+\lambda)\gamma_H$ by Lemma 6 in [18]. Note that this holds because σ_{t-1} in our case is only updated via data points starting from t_0 and the fact that $p - t_0 \leq H$.

Proof of Claim 1. We begin with the first equation. Note that $(\Phi_t^T \Phi_t + \lambda I)\varphi(x) = \Phi_t^T k_t(x) + \lambda\varphi(x)$. Combining this with the fact that $(\Phi_t^T \Phi_t + \lambda I)^{-1}\Phi_t^T = \Phi_t^T(\Phi_t \Phi_t^T + \lambda I)^{-1}$, yields

$$\varphi(x)$$
$$= \Phi_t^T(\Phi_t \Phi_t^T + \lambda I)^{-1} k_t(x) + \lambda(\Phi_t^T \Phi_t + \lambda I)^{-1}\varphi(x).$$

This directly leads to

$$\lambda \|\varphi(x)\|_{V_t^{-1}}^2$$
$$= \lambda\varphi(x)^T(\Phi_t^T \Phi_t + \lambda I)^{-1}\varphi(x)$$
$$= k(x,x) - k_t(x)^T(K_t + \lambda I)^{-1}k_t(x)$$
$$= \sigma_t^2(x).$$

which proves the first equation. The second equation follows directly from the first one. To prove the third equation, first note that

$$\ln\det(\widetilde{V}_t) = \ln\det(I + \frac{1}{\lambda}\Phi_t^T \Phi_t)$$
$$= \ln\det(I + \frac{1}{\lambda}\Phi_t \Phi_t^T)$$
$$= \ln\det(I + \frac{1}{\lambda}K_t).$$

Recall the definition of γ_t in (3), we now have

$$\ln\det(\widetilde{V}_t) \leq 2\gamma_{t-t_0} \leq 2\gamma_H.$$

$\square$

C. Proof of Theorem 2

The proof shares great similarity with the proof of Theorem 1. However, as pointed out by [8], due to the sliding window feature, one cannot directly apply the standard self-normalized inequality. We adopt the same trick proposed in [8] to handle this. The challenge here is to handle the possibly infinite dimension. To this end, we will apply Fatou's lemma. We first have the following bound. Due to the information loss by sliding window, there is a T factor in the β term.

Lemma 2. *For any $\delta \in (0,1)$, with probability at least $1-\delta$, the following holds for any $x \in \mathcal{X}$ and any $t \geq 1$,*

$$|\mu_{t-1}(x) - f_t(x)| \leq \sum_{s=t_0}^{t-1} \|f_s - f_{s+1}\|_{\mathcal{H}} + \beta_t\sigma_{t-1}(x),$$

where $\beta_t = \left(B + \frac{1}{\sqrt{\lambda}}R\sqrt{2\gamma_{t\wedge w} + 2\ln(T/\delta)}\right)$.

Proof. The key difference compared to Lemma 1 is a new bound on $\left\|\sum_{s=t_0}^{t-1} \eta_s\varphi(x_s)\right\|_{\widetilde{V}_{t-1}^{-1}}$, where $t_0 =$

$\max(1, t - w)$. We denote $S_{t-1} = \sum_{s=t_0}^{t-1} \eta_s\varphi(x_s)$, following the trick in [8] to handle the information loss due to sliding window, we further define $\widehat{V}_{u-1} = \sum_{s=1\vee t-w}^{u-1}(1/\lambda)\varphi(x_s)\varphi(x_s)^T + I$, $\widehat{S}_{u-1} = \sum_{s=1\vee t-w}^{u-1} \eta_s\varphi(x_s)$ and $\widehat{M}_{u-1}(q) = \exp(\frac{1}{\sqrt{\lambda}}q^T\widehat{S}_{u-1} - \frac{R^2}{2\lambda}q^T\widehat{V}_{u-1}(0)q)$, where $\widehat{V}_{u-1}(0) = \widehat{V}_{u-1} - I$. Note that by these definitions, $\widehat{S}_{t-1} = S_{t-1}$ and $\widehat{V}_{t-1} = \widetilde{V}_{t-1}$. We then take an infinite Gaussian random sequence (independent of all other randomness) $Q \sim N(0, I)$, and define $M_{u-1} = \mathbb{E}\left[\widehat{M}_{t-1}(Q)\right]$. By standard martingale arguments and sub-Gaussian noise assumption, we have $\mathbb{E}[M_{t-1}] \leq 1$. We now take a finite approximation by using the first d dimension of the feature map. In particular, we denote by $Q_d, M_{u-1,d}, \widehat{S}_{u-1,d}$ and $\widehat{V}_{u-1,d}$ truncated versions. Clearly, $\mathbb{E}[M_{t-1,d}] \leq 1$ for all d. Moreover, we can take the finite-version result for $M_{t-1,d}$ in [8]. That is, $M_{t-1,d} = \frac{1}{\det(\widetilde{V}_{t-1,d})^{1/2}} \exp(\frac{1}{2R^2\lambda}\|S_{t-1,d}\|_{\widetilde{V}_{t-1,d}^{-1}}^2)$. Now, by using Fatou's lemma, we can obtain that

$$\mathbb{P}\left[\lim_{d\to\infty} \frac{\|S_{t-1,d}\|_{\widetilde{V}_{t-1,d}^{-1}}^2}{2R^2\lambda\ln(\det(\widetilde{V}_{t-1,d})^{1/2}/\delta)} \geq 1\right]$$
$$\leq \mathbb{E}\left[\lim_{d\to\infty} \frac{\delta\exp(\frac{1}{2R^2\lambda}\|S_{t-1,d}\|_{\widetilde{V}_{t-1,d}^{-1}}^2)}{\det(\widetilde{V}_{t-1,d})^{1/2}}\right]$$
$$\leq \delta \lim_{d\to\infty} \mathbb{E}[M_{t-1,d}] \leq \delta.$$

Due to the sliding window, we cannot apply the the standard 'stopping time' trick. Instead, we use union bound to obtain, for any $t \geq 1$, with probability at least $1 - \delta$, $\|S_{t-1}\|_{\widetilde{V}_{t-1}^{-1}}^2 \leq 2R^2\lambda\ln(\det(\widetilde{V}_{t-1,d})^{1/2}T/\delta)$. $\square$

Now, following the same steps as in Theorem 1, we obtain

$$\mathcal{R}(T) \leq 2HP_T + \sqrt{\lambda}\beta_T(\delta)\sum_{t=1}^{T} \|\varphi(x_t)\|_{V_{t-1}^{-1}},$$

where $V_{t-1} = \sum_{s=1\vee t-w}^{t-1} \varphi(x_s)\varphi(x_s)^T + \lambda I$. Similar to the procedure in the proof of Theorem 1, we have $\sum_{t=1}^{T} \|\varphi(x_t)\|_{V_t^{-1}} \leq \sum_{k=0}^{T/w-1} \sum_{t=kw+1}^{(k+1)w} \|\varphi(x_t)\|_{V_{t-1}^{-1}}$. Let $W_{t-1}^{(k)} = \sum_{s=kw+1}^{t-1} \varphi(x_s)\varphi(x_s)^T + \lambda I$ (which is the V_{t-1} in the proof of Theorem 1), and hence for $t \in [kw, (k+1)w]$, $V_{t-1}^{-1} \preceq W_{t-1}^{(k)}$. Hence, we can now follow the remaining steps in Theorem 1 to obtain

$$\mathcal{R}(T) = O\left(wP_T + \beta_T(\delta)T\sqrt{\frac{\gamma_w}{w}}\right).$$

V. Conclusion

We studied the black-box bandit optimization under a time-varying environment. We consider the variation budget model, which is able to capture both slowly-changing and abruptly-changing environments. We derived the dynamic regret bounds for R-GP-UCB and SW-GP-UCB with a non-Bayesian regularity assumption.

REFERENCES

[1] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[2] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Improved algorithms for linear stochastic bandits," in *Advances in Neural Information Processing Systems*, 2011, pp. 2312–2320.

[3] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," *arXiv preprint arXiv:0912.3995*, 2009.

[4] S. R. Chowdhury and A. Gopalan, "On kernelized multi-armed bandits," *arXiv preprint arXiv:1704.00445*, 2017.

[5] P. Auer, P. Gajane, and R. Ortner, "Adaptively tracking the best bandit arm with an unknown number of distribution changes," in *Conference on Learning Theory*. PMLR, 2019, pp. 138–158.

[6] O. Besbes, Y. Gur, and A. Zeevi, "Optimal exploration–exploitation in a multi-armed bandit problem with non-stationary rewards," *Stochastic Systems*, vol. 9, no. 4, pp. 319–337, 2019.

[7] W. C. Cheung, D. Simchi-Levi, and R. Zhu, "Learning to optimize under non-stationarity," in *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019, pp. 1079–1087.

[8] Y. Russac, C. Vernade, and O. Cappé, "Weighted linear bandits for non-stationary environments," *arXiv preprint arXiv:1909.09146*, 2019.

[9] P. Zhao, L. Zhang, Y. Jiang, and Z.-H. Zhou, "A simple approach for non-stationary linear bandits," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 746–755.

[10] P. Zhao and L. Zhang, "Non-stationary linear bandits revisited," *arXiv preprint arXiv:2103.05324*, 2021.

[11] I. Bogunovic, J. Scarlett, and V. Cevher, "Time-varying gaussian process bandit optimization," in *Artificial Intelligence and Statistics*. PMLR, 2016, pp. 314–323.

[12] G. Riutort-Mayol, P.-C. Bürkner, M. R. Andersen, A. Solin, and A. Vehtari, "Practical hilbert space approximate bayesian gaussian processes for probabilistic programming," *arXiv preprint arXiv:2004.11408*, 2020.

[13] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Springer, 2003, pp. 63–71.

[14] Y. Abbasi-Yadkori, "Online learning for linearly parametrized control problems," 2013.

[15] X. Zhou and J. Tan, "Local differential privacy for bayesian optimization," *arXiv preprint arXiv:2010.06709*, 2020.

[16] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur, "Gaussian processes and kernel methods: A review on connections and equivalences," *arXiv preprint arXiv:1807.02582*, 2018.

[17] D. A. Miller, "An introduction to functional analysis for science and engineering," *arXiv preprint arXiv:1904.02539*, 2019.

[18] S. R. Chowdhury and A. Gopalan, "Bayesian optimization under heavy-tailed payoffs," in *Advances in Neural Information Processing Systems*, 2019, pp. 13 790–13 801.