Weighted Gaussian Process Bandits for Non-stationary Environments

Yuntian Deng
The Ohio State University

Xingyu Zhou
Wayne State University

Baekjin Kim
University of Michigan

Ambuj Tewari
University of Michigan

Abhishek Gupta
The Ohio State University

Ness Shroff
The Ohio State University

Abstract

In this paper, we consider the Gaussian process (GP) bandit optimization problem in a non-stationary environment. To capture external changes, the black-box function is allowed to be time-varying within a reproducing kernel Hilbert space (RKHS). To this end, we develop WGP-UCB, a novel UCB-type algorithm based on weighted Gaussian process regression. A key challenge is how to cope with infinite-dimensional feature maps. To that end, we leverage kernel approximation techniques to prove a sublinear regret bound, which is the first (frequentist) sublinear regret guarantee on weighted time-varying bandits with general nonlinear rewards. This result generalizes both non-stationary linear bandits and standard GP-UCB algorithms. Further, a novel concentration inequality is achieved for weighted Gaussian process regression with general weights. We also provide universal upper bounds and weight-dependent upper bounds for weighted maximum information gains. These results are of independent interest for applications such as news ranking and adaptive pricing, where weights can be adopted to capture the importance or quality of data. Finally, we conduct experiments to highlight the favorable gains of the proposed algorithm in many cases when compared to existing methods.

Proceedings of the 25[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

1 Introduction

There has been significant interest in developing the theoretical foundations and practical algorithms for solving bandit optimization problems. This interest has been driven by many practical applications, where one needs to sequentially select query points to maximize the cumulative reward (Bubeck and Cesa-Bianchi, 2012). Such sequential decision-making is usually based on noisy feedback from black-box functions defined over a possibly large domain space.

The most classical model is the multi-armed bandit (MAB) (Robbins, 1952) where query points are independent and finite. It is then extended to stochastic linear bandits (Auer, 2002; Abbasi-Yadkori et al., 2011), where the black-box function is linear, and query points become non-orthonormal. Gaussian process bandits (Srinivas et al., 2009; Chowdhury and Gopalan, 2017) further generalizes the previous two by allowing general black-box functions (e.g., non-linear and non-convex) by utilizing the representation power of the reproducing kernel Hilbert space (RKHS). These models of online decision-making have becoming ubiquitous in practical applications, such as change-points detection (Liu et al., 2018), personalized news recommendation (Li et al., 2010), and portfolio selection (Huo and Fu, 2017).

In real-world scenarios, the unknown function is often not fixed but varies over time. For example, the channel conditions in wireless networks are time-varying, and thus the Quality of Service is not static (Zhou et al., 2019). In recommender systems, users' preferences may change with growth, and the corresponding reward function for any recommending action is time-varying (Li et al., 2010). This has motivated recent studies in online-decision making under time-varying environments, such as in MAB (Besbes et al., 2014), linear bandits (Kim and Tewari, 2020), and Gaussian process bandits (Bogunovic et al., 2016). Roughly speaking, there are three commonly used techniques to handle non-stationarity – *restarting, sliding win-*

dow and *weighted penalty*. By restarting, the learning agent resets the learning process once in a while to directly discard all the old information at once (e.g., (Zhao et al., 2020; Besson and Kaufmann, 2019)) while under the sliding window, the learning agent gradually discard old information by only using the most recent data in the learning process (e.g., (Cheung et al., 2019)). Recently, Russac et al. (2019) proposes the weighted penalty approach, which puts more weight on the most recent data while penalizing outdated information via less weight. This approach can be viewed as a 'soft' way of discounting outdated information rather than completely dropping it as in restarting and sliding window methods. The weighted penalty approach has been shown to be beneficial when the black-box function is linear (Russac et al., 2019) or general continuously differentiable Lipschitz function (Russac et al., 2020). However, it remains an open problem whether one can achieve the advantages of the weighted penalty approach for general black-box functions, in particular the ones in an RKHS, which enjoys the uniform approximation of an arbitrary continuous function (Micchelli et al., 2006) (under a proper choice of the kernel).

Recently, Wei and Luo (2021) provides optimal results for (generalized) linear bandits, via maintaining different instances of base algorithms. It remains an open problem whether Gaussian Process bandit can achieve its regret bound. To be specific, we do not know whether GP bandits satisfies Assumption 1 in this paper and what is the form of $C(t)$ and $\Delta(t)$ for GP bandits. If $C(t)$ does not have the same polynomial form as its Theorem 2, its result cannot be extended to GP bandits.

In this paper, we take the first step to tackling this fundamental problem by proposing a novel weighted penalty algorithm with rigorous regret guarantees in the context of Gaussian process bandits. This is achieved by overcoming several key challenges. First, to fully utilize the representation power of an RKHS for general functions, our choice of kernel often has an infinite-dimensional feature space (e.g., Squared Exponential kernel). In this case, all existing regret analysis breaks down as regret bounds in these works have an explicit, growing dependence on the feature dimension (e.g., d). Moreover, the standard approach of resolving the dependence on d in the regret bounds for GP bandits does not apply in our case. This is because we are dealing with a weighted GP regression rather than a standard one, which directly raises three substantial challenges. First, we need to find a new self-normalized concentration inequality to show that the posterior mean under the weighted GP regression is still close to the true function in a certain sense, which helps to translate the cumulative regret into a sum of predictive variance. Then, we need to find a new technique to bounding this term by a properly defined so-called Maximum Information Gain (MIG). Finally, existing bounds on MIG also do not apply, and thus we have to derive the new ones in our weighted setting.

Contributions. In summary, our contributions can be summarized as follow.

First, we develop a general framework for the regret analysis under weighted GP regression by overcoming the aforementioned challenges. In particular, for general weighted GP regression, we establish the first self-normalized concentration inequality. Then, by novel applications of Quadrature Fourier features (QFF) approximation and Mercer's theorem, we present the first bounds for the sum of predictive variance and the corresponding MIGs in the weighted case. These results are not only the cornerstones in our setting, but also could be useful for other general GP regression settings.

Second, we propose a new algorithm - Weighted Gaussian Process Upper Confidence Bound (WGP-UCB) for non-stationary bandit optimization. It generalizes the standard GP-UCB algorithms (Srinivas et al., 2009; Chowdhury and Gopalan, 2017) in stationary environments to the time-varying case. This is also a significant generalization of discounted linear bandit (Russac et al., 2019) and discounted generalized linear UCB (Russac et al., 2020) by allowing the payoff function to be within a much broader class of functions (thanks to the use of RKHS).

Third, by a proper choice of the weighted scheme in WGP-UCB, we establish the first regret bound for the weighted penalty algorithm in the context of GP bandits by utilizing our novel results for the weighted GP regression. In particular, we have a regret bound $O(\dot{\gamma}_T^{7/8} B_T^{1/4} T^{3/4})$ if the variation budget B_T is known, and a regret bound $O(\dot{\gamma}_T^{7/8} B_T T^{3/4})$ if B_T is unknown, for both abruptly-changing and slowly-varying environments where $\dot{\gamma}_T$ is a properly defined maximum information gain. Note that this result directly recovers the existing weighted penalty results as special cases (e.g., a choice of the linear kernel).

Related Work. Online learning in changing environments has been well studied. In traditional MAB, (Auer et al., 2019) considers the situation where reward distributions may change abruptly several times, which is usually referred to switching bandits (Garivier and Moulines, 2011) or abruptly-changing environments. The regret bound usually depends on the number of changes, which relies on change-point detection (Cao et al., 2019; Liu et al., 2018). An alternative approach to quantify time-variations is variation budget (Besbes et al., 2014, 2015, 2019), which captures the cumulative

temporal variation of system parameters for the total time horizon.

In the stochastic linear bandits setting, we recall that there are mainly three strategies to deal with non-stationarity : *restarting* (Zhao et al., 2020), *sliding window* (Cheung et al., 2019), and *weighted penalty*. The last strategy leverages an increasing weight sequence to emphasize the impact of recent observations while gradually forgetting past observations. In Russac et al. (2019), exponentially increasing weights are used to develop the D-LinUCB algorithm based on the weighted least square estimator. Two variants of this algorithm are developed in Kim and Tewari (2020) based on perturbation techniques. Recently, a technical flaw in these three works (Cheung et al., 2019; Russac et al., 2019; Zhao et al., 2020) was identified in Zhao and Zhang (2021), which corrects the order of regret bounds in all three algorithms. For generalized linear models, sliding window and weighted penalty algorithms are developed in Russac et al. (2020), where the payoff functions are required to be continuously differentiable and Lipschitz.

For the Gaussian process bandits, there are two different assumptions on the black-box functions. The Bayesian setting assumes that the unknown function is a sample from a GP with a known kernel, while the frequentist setting (agnostic setting in Srinivas et al. (2009)) assumes that the unknown function is a fixed function in a reproducing kernel Hilbert space (RKHS) with bounded norm. Under the Bayesian setting, Bogunovic et al. (2016) proposes a discounted algorithm and a restarting algorithm, assuming that the evolution of Gaussian process obeys a simple Markov model. Under the frequentist setting, Zhou and Shroff (2021) introduces restarting and sliding window algorithms and obtains regret bounds based on maximum information gain. However, due to difficulties arising from the time variation and infinite-dimensional feature maps, the weighted penalty algorithm has not been studied yet under the frequentist setting, which is also the future direction as listed in Bogunovic et al. (2016).

2 Problem Statement and Preliminaries

In this section, we introduce the setting of our problem and necessary preliminaries.

We consider the non-stationary problem of sequentially maximizing reward function $f_t : D \to \mathbb{R}$ over a set of decisions $D \subset \mathbb{R}^d$. At each discrete time slot $t = 1, 2, \ldots$, the learning agent selects an action (query point) $x_t \in D$ and the reward $f_t(x_t)$ is observed through a noisy channel as $y_t = f_t(x_t) + \epsilon_t$ where ϵ_t is the zero mean noise. Denote the history as $\mathcal{H}_{t-1} = \{(x_s, y_s) : s \in \{1, 2, \ldots, t-1\}\}$. Conditioned on history $\mathcal{H}_{t-1}$, the noise sequence ϵ_t is R-sub-Gaussian for a fixed constant $R \geq 0$, i.e. $\forall t > 1, \forall \lambda \in \mathbb{R}, \mathbb{E}\left[e^{\lambda \epsilon_t} | \mathcal{F}_{t-1}\right] \leq \exp(\frac{\lambda^2 R^2}{2})$ where $\mathcal{F}_{t-1} = \sigma\left(\mathcal{H}_{t-1}, x_t\right)$ is the σ-algebra generated by actions and rewards observed so far.

The objective of the learning agent is to maximize the cumulative reward $\sum_{t=1}^{T} f_t(x_t)$. This is equivalent to minimize its *dynamic regret* R_T, which is defined as $R_T = \sum_{t=1}^{T} f_t(x_t^*) - f_t(x_t)$ where $x_t^* = \arg\max_{x \in D} f_t(x)$ is the attainable best action at time t for function $f_t(\cdot)$.

Regularity Assumptions: We assume that f_t is a fixed function in a Reproducing Kernel Hilbert Space (RKHS) with a bounded norm. Specifically, we assume that D is compact. The RKHS, denoted by $H_k(D)$, is completely specified by its kernel function $k(\cdot, \cdot)$, with an inner product $\langle \cdot, \cdot \rangle_H$ satisfying the reproducing property: $f(x) = \langle f, k(x, \cdot) \rangle_H$ for all $f \in H_k(D)$. The RKHS norm is given by $\|f\|_H := \sqrt{\langle f, f \rangle_H}$. We assume that f_t at each time t is bounded by $\|f_t\|_H \leq B$ for a fixed constant B. Moreover, we assume a bounded variance by restricting $k(x, x) \leq 1$. The assumptions hold for practically relevant kernels. One concrete example is Squared Exponential kernel, defined as $k_{SE}(x, x') = \exp(-s^2/2l^2)$ where scale parameter $l > 0$ and $s = \|x - x'\|_2$ specifies distance between two points.

Time-varying Budget: As the environment is time-varying, we assume that the total variation of f_t satisfies the following budget, $\sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_H \leq B_T$, including both abruptly-changing and slowly-changing environments.

Maximum Information Gain: We use $I(y_A; f_A)$ to denote the mutual information between $f_A = [f(x)]_{x \in A}$ and $y_A = f_A + \epsilon_A$, which quantifies the reduction in uncertainty about f after observing y_A at points $A \subset D$. Then the maximum information gain (Srinivas et al., 2009) is defined as, $\gamma_n := \max_{A \subset D : |A| = n} I(y_A; f_A) = \max_{A \subset D : |A| = n} \frac{1}{2} \log\det(I + \lambda^{-1} K_A)$, where $K_A = [k(x, x')]_{x, x' \in A}$.

Agnostic setting: We recall the agnostic setting in standard GP-UCB algorithm (Chowdhury and Gopalan, 2017) for the stationary environment. Gaussian process (GP) and Gaussian likelihood models are used to design this algorithm. $GP_D(0, k(\cdot, \cdot))$ is the prior for reward function f_t. The noise ϵ_t is drawn independently from $\mathcal{N}(0, \lambda)$. Conditioned on the history $\mathcal{H}_t$, it has the posterior distribution of f_t, $GP_D\left(\mu_t(\cdot), \sigma_t^2(\cdot)\right)$, where the posterior mean and vari-

ance are defined as

$$\mu_t(x) = k_t(x)^T(K_t + \lambda I)^{-1} y_{1:t} \quad (1)$$

$$\sigma_t^2(x) = k(x,x) - k_t(x)^T(K_t + \lambda I)^{-1} k_t(x) \quad (2)$$

where $y_{1:t} \in \mathbb{R}^t$ is the reward vector $[y_1, \ldots, y_t]^T$. For set of sampling points $A_t = \{x_1, \ldots, x_t\}$, the kernel matrix is $K_t = [k(x,x')]_{x,x' \in A_t} \in \mathbb{R}^{t \times t}$ and the vector $k_t(x) = [k(x_1,x), \ldots, k(x_t,x)]^T \in \mathbb{R}^t$. The GP prior and Gaussian likelihood are only used for algorithm design and do not affect the setting of reward function $f_t \in H_k(D)$ and noise ϵ_t (i.e., could be sub-Gaussian).

3 Weighted Gaussian Process Regression

In this section, we introduce a general weighted algorithm based on weighted GP regression. The key difference with standard GP regression is that we allow different weight for each data point. It is worth noting that this result is fairly generic in the sense that it can be applied in general situations where weights are used to associate with 'importance' in the data points. E.g., more weights are assigned to observations that are less noisy in weighted ridge regression (Zhou et al., 2021).

In particular, the weighted GP regression under a changing regularizer is defined by

$$\hat{f} = \arg\min_{f \in H_k(D)} \sum_{s=1}^{t-1} w_s (y_s - f(x_s))^2 + \lambda_t \|f\|_{\mathcal{H}}^2$$

where each data point is associated with a weight in computing the least square estimate. Due to this, standard posterior mean and variance in (1) and (2) fail to capture the statistics of $\hat{f}$. To this end, we have to carefully adjust the kernel vector and kernel matrix in (1) by incorporating proper weights. Specifically, let $W = \mathrm{diag}(\sqrt{w_1}, \sqrt{w_2}, \ldots, \sqrt{w_t}) \in \mathbb{R}^{t \times t}$. Then we define the weighted version of kernel matrix $\tilde{K}_t := WK_tW^T$ and weighted kernel vector $\tilde{k}_t(x) := Wk_t(x)$. We further define weighted observation $\tilde{y}_{1:t} := Wy_{1:t} = [\sqrt{w_1}y_1, \ldots, \sqrt{w_t}y_t]^T$. Finally, a weight-dependent regularizer is defined by $\lambda_t = \lambda w_t$. Then, $\hat{f}$ and its uncertainty are given by the following equations, respectively.

$$\tilde{\mu}_t(x) = \tilde{k}_t(x)^T(\tilde{K}_t + \lambda_t I_t)^{-1} \tilde{y}_{1:t} \quad (3)$$

$$\tilde{\sigma}_t^2(x) = k(x,x) - \tilde{k}_t(x)^T(\tilde{K}_t + \lambda_t I_t)^{-1} \tilde{k}_t(x) \quad (4)$$

One can see that (3) and (4) share the same structure as the standard ones in (1) and (2). This nice result directly enables us to design a UCB-type learning algorithm in the weighted case as follows.

Algorithm: The Weighted Gaussian Process-UCB algorithm (WGP-UCB) (Algorithm 1) uses a combination of the weighted posterior mean $\tilde{\mu}_{t-1}(x)$ and weighted standard deviation $\tilde{\sigma}_{t-1}(x)$ to construct an upper confidence bound (UCB) over the unknown function. It then chooses an action x_t at time t as follows:

$$x_t := \arg\max_{x \in D} \tilde{\mu}_{t-1}(x) + \beta_{t-1}\tilde{\sigma}_{t-1}(x) \quad (5)$$

where $\beta_t = B + \frac{1}{\sqrt{\lambda}} R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_t}$ and $0 < \delta < 1$. We note that this algorithm enjoys the same simplicity as the standard non-weighted one (Chowdhury and Gopalan, 2017; Srinivas et al., 2009). Meanwhile, there are substantial differences. In particular, besides the new posterior mean and variance, we also need to replace the MIG in the confidence width β_t by a weighted one, i.e., $\bar{\gamma}_t$. This term is defined as $\bar{\gamma}_t = \max_{A \subset D: |A| = t} \frac{1}{2} \log\det(I + \alpha_t^{-1} W^2 K_A W^{2T}) = \max_{A \subset D: |A| = t} \frac{1}{2} \log\det(I + \alpha_t^{-1} \bar{K}_t)$, where $W^2 = \mathrm{diag}(w_1, w_2, \ldots, w_t) \in \mathbb{R}^{t \times t}$, $\bar{K}_t = W^2 K_t W^{2T}$ is the double-weighted kernel matrix and $\alpha_t = \lambda w_t^2$.

In the following sections, we will develop a general framework for the regret analysis in the weighted GP regression, which recovers the standard analysis as special cases by choosing $w_s = 1$ for all $s \in [T]$ (Chowdhury and Gopalan, 2017; Srinivas et al., 2009). From a high-level perspective, the typical recipe of deriving regret bounds in GP bandits has three main steps. **(I)** One needs to first show that the true underlying function is close to the posterior mean within some distance given by the standard derivation. This concentration result is the cornerstone and is typically achieved by relying on the so-called self-normalized inequality. **(II)** Based on this concentration, one can bound the cumulative regret by a sum of predictive variance terms. This can be further upper bounded by the MIG. **(III)** The MIG will finally be upper bounded depending on the choice of kernels, which leads to the final regret bound. However, all the three key steps face new challenges in our weighted case, and hence we will conquer them one by one.

4 Confidence Bounds

In this section, we focus on deriving a new concentration inequality in the weighted case to show that the new posterior mean is still close to the true function in the non-stationary environment, which resolves the challenge **(I)** listed above. In particular, we present concentration results for both stationary and non-stationary environments.

To start with, we introduce a particular feature map via Mercer's Theorem, which will only be used in our analysis. The following version of Mercer's theorem (described by Theorem 1 next) is adapted from Theorem 4.1 and 4.2 in Kanagawa et al. (2018), which

Algorithm 1: Weighted Gaussian Process UCB (WGP-UCB)

Input : parameters $k(\cdot,\cdot), B, R, \lambda, \delta$, weights $\{\omega_t\}_{t=1}^T$.

for $t \geq 1$ **do**

  Set $\beta_{t-1} = B + \frac{1}{\sqrt{\lambda}} R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_{t-1}}$;

  Choose $x_t = \arg\max_{x \in D} \tilde{\mu}_{t-1}(x) + \beta_{t-1}\tilde{\sigma}_{t-1}(x)$;

  Observe reward $y_t = f_t(x_t) + \epsilon_t$;

  Update $\tilde{\mu}_t(x)$ and $\tilde{\sigma}_t(x)$ according to Equation (3) and (4).

end

roughly says that the kernel function can be expressed in terms of the eigenvalues and eigenfunctions under mild conditions.

Theorem 1. *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 a 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 orthonormal basis of $L_{2,\nu}(\mathcal{X})$, such that*

$$k(x, x') = \sum_{m=1}^{\infty} c_m \phi_m(x)\phi_m(x') \tag{6}$$

where $c_m \in \mathbb{R}^+$ and $\phi_m \in \mathcal{H}$ for $m \geq 1$. $\{c_m\}_{m=1}^{\infty}$ is the eigenvalue sequence in decreasing order. $\{\phi_m\}_{m=1}^{\infty}$ are the eigenfeatures (eigenfunctions) of k. The RKHS can also be represented in terms of $\{c_m, \phi_m\}_{m=1}^{\infty}$. i.e.,

$$\mathcal{H} = \Big\{ f(\cdot) = \sum_{m=1}^{\infty} \theta_m \sqrt{c_m}\phi_m(\cdot) : \|f\|_H := \|\theta\|_2 < \infty \Big\}.$$

Based on this theorem, we can explicitly define a feature map as $\varphi(x) = [\varphi_1(x), \varphi_2(x), \ldots]^T \in \mathbb{R}^M$ (M may be infinity) where $\varphi_m = \sqrt{c_m}\phi_m \in \mathcal{H}$ and $\varphi_m := D \to \mathbb{R}$. Given $\theta = [\theta_1, \theta_2, \ldots]^T \in \mathbb{R}^M$, we have reward function $f(x) = \theta^T\varphi(x)$ and kernel function $k(x, x') = \varphi^T(x)\varphi(x') \in \mathbb{R}$. Define $\Phi_t := [\varphi(x_1), \ldots, \varphi(x_t)]^T \in \mathbb{R}^{t\times M}$ and we get the $t \times t$ kernel matrix $K_t = \Phi_t\Phi_t^T$ and $k_t(x) = \Phi_t\varphi(x) \in \mathbb{R}^t$.

In our weighted case, we have the weighted feature matrix $\tilde{\Phi}_t := W\Phi_t$, weighted kernel vector $\tilde{k}_t(x) = \tilde{\Phi}_t\varphi(x)$ and weighted kernel matrix $\tilde{K}_t = \tilde{\Phi}_t\tilde{\Phi}_t^T$. Additionally, the double-weighted feature matrix $\bar{\Phi}_t := W^2\Phi_t$, and double-weighted kernel matrix $\bar{K}_t = \bar{\Phi}_t\bar{\Phi}_t^T$. Besides, we further define weighted Gram matrix $V_t = \sum_{s=1}^t w_s\varphi(x_s)\varphi(x_s)^T + \lambda_t I_{\mathcal{H}} \in \mathbb{R}^{M\times M}$ and double-weighted Gram matrix $\tilde{V}_t = \sum_{s=1}^t w_s^2\varphi(x_s)\varphi(x_s)^T + \alpha_t I_{\mathcal{H}} \in \mathbb{R}^{M\times M}$. The full list of notations is deferred to Appendix A.

This explicit feature map enables us to directly establish the following result, which states that our weighted GP bandit generalizes the weighted linear bandit. The details and proof are stated in Appendix B.1 Lemma 9, where $\hat{\theta}_t = V_t^{-1}\sum_{s=1}^t w_s\varphi(x_s)y_s$.

Remark 1. *The weighted linear bandits in (Russac et al., 2019, Equation 3) can be recovered by taking $\tilde{\mu}_t(x) = \varphi(x)^T\hat{\theta}_t$ and $\varphi(x) = x$.*

In the following, based on this explicit feature space, we will establish confidence bounds for our weighted GP bandit under both stationary and non-stationary environments.

Confidence bound under stationary environments. First we consider the stationary environment, where the reward function $f_t = f^*$ does not change with respect to time t. The following result shows how the posterior mean $\tilde{\mu}_t(x)$ is concentrated around the unknown reward function $f^*(x)$.

Theorem 2. *Let $f^* : D \to \mathbb{R}$ be a member of the RKHS of real-valued functions on D specified by kernel k, with RKHS norm bounded by $\|f^*\|_H \leq B$ and $\acute{\sigma}_t^2(x) = \lambda\|\varphi(x)\|^2_{V_t^{-1}\tilde{V}_tV_t^{-1}}$. Then, with probability at least $1-\delta$, the following concentration inequality holds:*

$$\begin{aligned}|f^*(x) - \tilde{\mu}_t(x)| &\leq \acute{\sigma}_t(x)B + \frac{\acute{\sigma}_t(x)}{\sqrt{\lambda}} R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_t} \\ &= \acute{\sigma}_t(x)\beta_t\end{aligned}$$

Proof Sketch for Theorem 2. Following similar steps in (Abbasi-Yadkori, 2013, Section 3.2), we first develop a self-normalized concentration bound on the weighted error sum $S_t = \sum_{s=1}^t w_s\varphi(x_s)\epsilon_s$. Then we bound $\|S_t\|_{\tilde{V}_t^{-1}}$ through double weighted information gain $\bar{\gamma}_t$. Finally we decompose $|f^*(x) - \tilde{\mu}_t(x)|$ into two terms $\|\varphi(x)\|_{V_t^{-1}\tilde{V}_tV_t^{-1}}$ and $(\|S_t\|_{\tilde{V}_t^{-1}} + \lambda_t\|\theta^*\|_{\tilde{V}_t^{-1}})$, and then bound them separately. The formal proofs and auxiliary lemmas are deferred to Appendix B.2.

We note that $\tilde{\mu}_t(x)$ here can be calculated by Equation (3). As $\varphi(x)$ is involved in V_t and $\tilde{V}_t$, we need to know the feature map $\varphi(x)$ before calculating $\acute{\sigma}_t(x)$, which is usually not practical. We resolve this issue in the following subsection by defining another predictive variance $\tilde{\sigma}_t(x)$.

With this confidence bound, we can claim that the standard kernelized bandit is only a special case of our weighted kernelized bandit. We defer the detailed explanation to Appendix B.2.3 via Lemma 12.

Remark 2. *The standard stationary case (IGP-UCB algorithm) (Chowdhury and Gopalan, 2017, Theorem 2) is recovered by taking $\lambda = 1$ and $w_t = 1$.*

Confidence bounds for non-stationary cases. In the non-stationary case, it is not guaranteed that

the actual reward function $f_t(x_t)$ always lies inside of confidence ellipsoid in Theorem 2 because of the time variations of environments. As did in weighted linear bandits (Russac et al., 2019), we introduce a surrogate parameter. $m_t(x) = \varphi(x)^T V_{t-1}^{-1}[\sum_{s=1}^{t-1} w_s \varphi(x_s) f_s(x_s) + \lambda w_{t-1}\theta_t^*]$, where $f_t^*(x) = \varphi(x)^T\theta_t^*$.

We note that this surrogate parameter $m_t(x)$ *is only used in the analysis of dynamic regret bound, and it is not involved in the implementation of our Algorithm 1.*

Leveraging this surrogate parameter $m_t(x)$, we can show that the new posterior mean is still close to the true function in the non-stationary environment. i.e., it satisfies $|m_t(x) - \tilde{\mu}_{t-1}(x)| \leq \tilde{\sigma}_{t-1}(x)\beta_{t-1}$ where $\tilde{\sigma}_t^2(x)$ is defined in Equation (4).

Theorem 3. *Let* $\mathcal{C}_t = \{f_t : |f_t(x) - \tilde{\mu}_{t-1}(x)| \leq \tilde{\sigma}_{t-1}(x)\beta_{t-1}, \forall x \in D\}$ *denote the confidence ellipsoid. Then,* $\forall \delta > 0$, $\mathcal{P}(m_t \in \mathcal{C}_t) \geq 1 - \delta$.

We remark that we cannot directly generalize weighted linear bandit (Russac et al., 2019) to nonlinear bandit by simply replacing A_s in Russac et al. (2019) with feature map $\varphi(x_s)$. This is because we can explicitly calculate the weighted gram $V_t = \sum_{s=1}^t \omega_s A_s A_s^T + \lambda_t I_d$ in linear case, while in the nonlinear case the weighted gram $V_t = \sum_{s=1}^t \omega_s \varphi(x_s)\varphi(x_s)^T + \lambda_t I_H$ cannot be explicitly calculated since the feature map $\varphi(x)$ is unknown. Therefore, calculating $\acute{\sigma}_t(x)^2 = \lambda ||\varphi(x)||^2_{V_t^{-1}\tilde{V}_t V_t^{-1}}$ is not practical. We overcome this by designing $\tilde{\sigma}_t(x)$ in Equation (4) (can be calculated without $\varphi(x)$) and $\tilde{\sigma}_t(x)$ plays the similar role as $\acute{\sigma}_t(x)$ in the confidence bound. The full proof is stated in Appendix B.3.

5 Dynamic Regret

In this section, we aim to resolve the challenge **(II)** listed at the end of Section 3 and obtain a sublinear regret bound for WGP-UCB (Algorithm 1). In particular, we resort to Quadrature Fourier Features (QFF) approximation to find an upper bound over the sum of predictive variance, which allows us to explicitly state the regret bound and analyze the order of regret bound. We further consider exponentially increasing weights of the form $w_t = \eta^{-t}$ to simplify the analysis, where $0 < \eta < 1$ is the discounting factor.

Quadrature Fourier Features (QFF) approximation. In some previous work (Abbasi-Yadkori et al., 2011; Russac et al., 2019), the feature dimension explicitly appears in the regret bound, which makes regret bound become trivial if the feature space is of infinite dimension. To overcome this, we find an approximate feature map $\breve{\varphi}$, such that the error of approximation is controlled in the infinite-dimensional feature space.

We consider a finite-dimension feature map $\breve{\varphi}(\cdot) : D \to \mathbb{R}^m$ such that it has a uniform approximation guarantee (Mutnỳ and Krause, 2019), i.e., for any $x, y \in D$, $\sup_{x,y} |k(x,y) - \breve{\varphi}(x)^T\breve{\varphi}(y)| \leq \varepsilon_m$. If $D = [0,1]^d$, for common kernels such as the Squared Exponential or the modified Matern kernel, we construct the feature map where $\bar{m} \in \mathbb{N}$ and $m = \bar{m}^d$,

$$\breve{\varphi}(x)_i = \begin{cases} \sqrt{v(\rho_i)}\cos\big(\frac{\sqrt{2}}{l}\rho_i^T x\big), \text{if } 1 \leq i \leq m \\ \sqrt{v(\rho_{i-m})}\sin\big(\frac{\sqrt{2}}{l}\rho_{i-m}^T x\big), \text{if } m+1 \leq i \leq 2m \end{cases}$$

where $v(\rho) = \prod_{j=1}^d \frac{2^{\bar{m}-1}\bar{m}!}{\bar{m}H_{\bar{m}-1}(\rho_j)^2}$ and H_i is the ith Hermite polynomial (Hildebrand, 1987). The set $\{\rho_1, \ldots, \rho_j\} = P_{\bar{m}} \times \ldots \times P_{\bar{m}}$ (d times) where $P_{\bar{m}}$ is the set of $\bar{m}$ roots of the ith Hermite polynomial H_i.

We then define $\breve{\Phi}_t = W[\breve{\varphi}(x_1), \ldots, \breve{\varphi}(x_t)]^T$, $\breve{k}_t(x) = \breve{\Phi}_t\breve{\varphi}(x)$, $\breve{K}_t = \breve{\Phi}_t\breve{\Phi}_t^T$, $\breve{V}_t = \breve{\Phi}_t^T\breve{\Phi}_t + \lambda_t I_{\mathcal{H}}$, $\breve{\sigma}_t^2(x) = k(x,x) - \breve{k}_t(x)^T(\breve{K}_t + \lambda_t I_t)^{-1}\breve{k}_t(x) = \lambda_t \|\breve{\varphi}(x)\|^2_{\breve{V}_t^{-1}}$, and $\breve{\gamma}_t = \frac{1}{2}\log\det(I + \lambda_t^{-1}\breve{\Phi}_t\breve{\Phi}_t^T)$. For SE kernel, QFF error is bounded by $\varepsilon_m = O(\frac{d2^{d-1}}{(\bar{m}l^2)^{\bar{m}}})$ (Chowdhury and Gopalan, 2019, Lemma 14).

Bounding the sum of predictive variance. Leveraging the QFF and the associated error bound, we can achieve a novel weight-dependent upper bound for the sum of predictive variance.

$$\sum_{t=1}^T \tilde{\sigma}_{t-1}(x_t) \leq \sqrt{4\lambda T\breve{\gamma}_T + 2\lambda m T^2\log(1/\eta)} + \frac{T\sqrt{\epsilon_m}}{1-\eta}.$$

i.e., we approximate it with some finite dimension results and we can show that the approximation error part is small, through $\frac{\beta_T T\sqrt{\varepsilon_m}}{1-\eta} = O(1)$ after properly tuning η and m. The detailed proof is in Appendix C.1 and C.2.

We have tried to simply extend standard results in Chowdhury and Gopalan (2017), however we found that we cannot bound $\sum_{t=1}^T \tilde{\sigma}_{t-1}(x_t)$ with weighted MIG $\tilde{\gamma}_t$, i.e., $\sum_{t=1}^T \tilde{\sigma}_{t-1}(x_t) \leq \sqrt{\lambda T\log\det(I + \lambda^{-1}\tilde{K}_T)}$, which cannot be bounded through $\tilde{\gamma}_t = \max_{A\subset D:|A|=t}\frac{1}{2}\log\det(I + \lambda_t^{-1}\tilde{K}_t)$ because $\lambda_t = \lambda\eta^{-t} > \lambda$. We overcome it by truncating the feature space via QFF and bound the finite part with a samll approximation error, as shown in Appendix C.2.

Regret bound of WGP-UCB with QFF approximation. With the novel upper bound above, we can state the dynamic regret bound of WGP-UCB with QFF approximation.

Theorem 4. *Let* $f_t \in H_k(D)$, $\|f_t\|_H \leq B$ *and* $k(x,x) \leq 1$. *Then, with probability at least* $1 - \delta$,

the dynamic regret R_T is bounded by

$$O\Big(\beta_T\sqrt{T\breve{\gamma}_T + mT^2\log(\frac{1}{\eta})} + c^{\frac{3}{2}}B_T\sqrt{\breve{\gamma}_t + mc\log(\frac{1}{\eta})} + \frac{B\eta^c}{1-\eta}T + B_T\frac{c^2\sqrt{\varepsilon_m}}{1-\eta} + \frac{\beta_T T\sqrt{\varepsilon_m}}{1-\eta}\Big)$$

where $c \geq 1$ is an integer, $0 < \eta < 1$, and $\beta_t = B + \frac{1}{\sqrt{\lambda}}R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_t}$.

Proof Sketch for Theorem 4. There are mainly three steps in this proof. First, we separate the stationary and non-stationary parts in the instantaneous regret r_t. They are bounded by $2\beta_{t-1}\tilde{\sigma}_{t-1}(x_t)$ and $2\sum_{p=t-c}^{t-1}\|f_p - f_{p+1}\|_H\frac{1}{\lambda}\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s) + \frac{4B\eta^c}{\lambda(1-\eta)}$, respectively. As pointed out by (Zhao and Zhang, 2021, p.4), the statement $\lambda_{\max}(V_{t-1}^{-1}\sum_{s=t-D}^{p}\eta^{-s}A_sA_s^T) \leq 1$ in (Russac et al., 2019, p.18) is not true. We fix this error in our proof as well, which introduces extra term $\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s)$. Secondly, we leverage the new bound for $\sum_{t=1}^{T}\tilde{\sigma}_{t-1}(x_t)$ developed above, which is $\sqrt{4\lambda T\breve{\gamma}_T + 2\lambda mT^2\log(1/\eta)} + \frac{T\sqrt{\varepsilon_m}}{1-\eta}$. Finally, we bound $\sum_{s=t-c}^{t}\acute{\sigma}_{t-1}(x_s)$ through $\sqrt{4\lambda c\breve{\gamma}_t + 2\lambda mc^2\log(1/\eta)} + \frac{c\sqrt{\varepsilon_m}}{1-\eta}$ with QFF, which is composed of finite approximation result and associated error. The full proof is in Appendix C.2-C.5.

Order analysis of regret bound. We start analysing the order of regret bound by define $\dot{\gamma}_T = \max\{\bar{\gamma}_T, \breve{\gamma}_T\}$. It is the maximum between double-weighted MIG and weighted MIG with QFF approximation, which is called *combined weighted MIG*. By optimally setting $c = \frac{\log T}{1-\eta}$ and $\bar{m} = \log_{4/e}(T^3\dot{\gamma}_T^{3/2})$, we have the order analysis as follows. The detail is deferred to Appendix C.6.

Corollary 5. *If B_T is known, the dynamic regret bound is $\tilde{O}(\dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4})$ by optimally choosing $\eta = 1 - \dot{\gamma}_T^{-1/4}B_T^{1/2}T^{-1/2}$. If B_T is unknown, the dynamic regret bound is $\tilde{O}(\dot{\gamma}_T^{7/8}B_TT^{3/4})$ by optimally choosing $\eta = 1 - \dot{\gamma}_T^{-1/4}T^{-1/2}$.*

Remark 3. *This regret bound achieves the same order as Zhou and Shroff (2021) where restarting and sliding window mechanisms are used. It is also a generalization of Zhao and Zhang (2021), which studied non-stationary linear bandit and fixed the error of largest eigenvalue in previous papers (Cheung et al., 2019; Russac et al., 2019; Zhao et al., 2020).*

6 Upper bounds on Maximum Information Gain

In this section, we aim to resolve the challenge **(III)** mentioned in Section 3, i.e., finding an explicit upper bound on MIG. In our case, we have multiple weighted MIGs and hence standard results fail. To resolve this issue, we generalize the idea in Vakili et al. (2021) to our weighted case by exploiting the tail properties in the feature maps given by Mercer's theorem.

In particular, our bounds on MIGs are based on a finite dimensional projection of the kernel, we start with outlining the details of this projection. For each element in K_t, we recall Equation (6) by Mercer's Theorem, where $c_m \in \mathbb{R}^+$ and $\phi_m \in \mathcal{H}_k$ for $m \geq 1$. $\{c_m\}_{m=1}^{\infty}$ is the eigenvalue sequence in decreasing order. $\{\phi_m\}_{m=1}^{\infty}$ are the eigenfeature of k. Similarly, for each element in double weighted kernel matrix $\bar{K}_t$, we have double weighted kernel function $\bar{k}(x_i, x_j) = w_iw_jk(x_i, x_j) = \sum_{m=1}^{\infty}w_iw_jc_m\phi_m(x_i)\phi_m(x_j)$.

Assumption 1. *(1) $\forall x, x' \in D, |k(x, x')| \leq \dot{k}$, for some $\dot{k} > 0$ (2) $\forall m \in \mathbb{N}, \forall x \in D, |\phi_m(x)| \leq \psi$, for some $\psi > 0$.*

In particular, we consider a N-dimensional projection (Vakili et al., 2021), where the N-dimensional feature space is $\Psi_N = [\phi_1(x), \phi_2(x), \ldots, \phi_N(x)]^T$. We keep the first N-dimension feature in kernel $\bar{k}_P(x_i, x_j) = w_iw_j\sum_{m=1}^{N}c_m\phi_m(x_i)\phi_m(x_j)$. The remaining part is $\bar{k}_O(x, x') = \bar{k}(x, x') - \bar{k}_P(x, x')$.

We define the following quantity based on the tail mass of the eigenvalues of m, $\delta_N = \sum_{m=N+1}^{\infty}c_m\psi^2$. Then for all $x, x' \in D$, we have $k_O(x, x') \leq \delta_N$. For some kernel k, if c_m diminishes at a sufficiently fast rate, then δ_N becomes arbitrarily small when N is large enough, which will be discussed in Corollary 8.

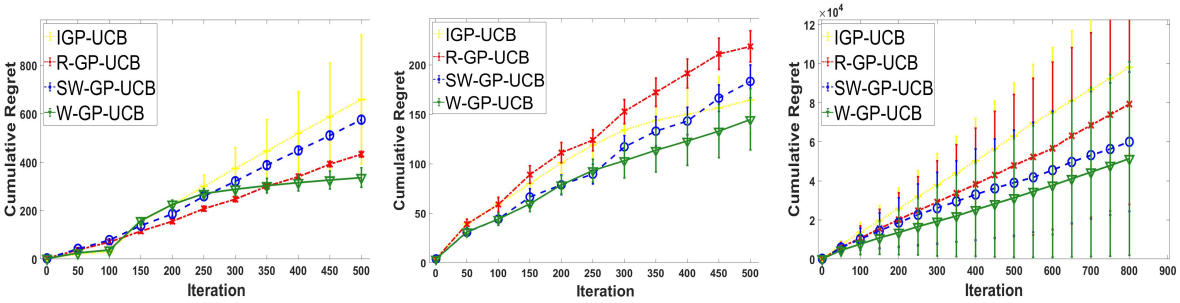
Universal Bound: Based on this eigendecay, we provide a universal upper bound for both $\bar{\gamma}_T$ and $\breve{\gamma}_T$, which states that the order $\tilde{O}(\log(T))$ holds for combined weighted MIG $\dot{\gamma}_T$ with any increasing weights $\{w_s\}_{s=1}^{t}$. The full proof is stated in Appendix D.1.

Theorem 6. *If Assumption 1 holds, $\dot{\gamma}_T = \max\{\bar{\gamma}_T, \breve{\gamma}_T\} \leq \frac{N}{2}\log\left(1 + \frac{\dot{k}T}{\lambda N}\right) + \frac{T}{2\lambda}\delta_N$*

The expression in Theorem 6 can be predigested as $\dot{\gamma}_T = O(N\log(T) + \delta_N T)$, which resolves challenge **(III)** mentioned in Section 3. To be more specific, the following remark provides an explicit form of the upper bound for SE kernel, which has a exponential eigendecay (Belkin, 2018; Vakili et al., 2021).

Remark 4. *For SE kernel, we have $c_m = O(exp(-m^{1/d}))$ and $\dot{\gamma}_T = O\big(\log^{d+1}(T)\big)$.*

Weight-dependent bound. Specifically, if the weights are exponentially increasing, we achieve a tighter upper bound for double weighted MIG $\bar{\gamma}_T$ and single weighted MIG with QFF $\breve{\gamma}_T$, respectively. This novel upper bound depends on the discount factor η and holds under any time horizon T.



(a) Abruptly-change, SE kernel (b) Slowly-change, SE kernel (c) Stock market data

Figure 1: Average cumulative regret of four algorithms in three different scenarios

Theorem 7. *If Assumption 1 holds and weight* $\omega_t = \eta^{-t}$*, the following upper bound on* $\bar{\gamma}_T$ *holds for all* $N \in \mathbb{N}$.

$$\bar{\gamma}_T \leq \frac{N}{2}\log\left(1+\frac{\dot{k}}{\lambda N(1-\eta^2)}\right)+\frac{1}{2\lambda(1-\eta^2)}\delta_N.$$

If the polynomial or exponential conditions on the the eigendecay of k are provided, a tighter bound is established.

Corollary 8. *1. Under the* (C_p, β_p) *polynomial eigendecay condition, i.e.,* $c_m \leq C_p m^{-\beta_p}$,

$$\bar{\gamma}_T \leq \left(\left(\frac{C_p\psi^2}{\lambda(1-\eta^2)}\right)^{\frac{1}{\beta_p}}\log^{-\frac{1}{\beta_p}}\left(1+\frac{\dot{k}}{\lambda(1-\eta^2)}\right)+1\right)\log\left(1+\frac{\dot{k}}{\lambda(1-\eta^2)}\right).$$

2. Under the $(C_{e,1}, C_{e,2}, \beta_e = 1)$ *exponential eigendecay condition, i.e.* $c_m \leq C_{e,1}\ exp(-C_{e,2}m^{\beta_e})$,

$$\bar{\gamma}_T \leq \left(\frac{1}{C_{e,2}}\left(\log\left(\frac{1}{1-\eta^2}\right)+C_{\beta_e}\right)+1\right)\log\left(1+\frac{\dot{k}}{\lambda(1-\eta^2)}\right)$$

where $C_{\beta_e} = \log\left(\frac{C_{e,1}\psi^2}{\lambda C_{e,2}}\right)$.

Similar results hold for $\check{\gamma}_T$ except that $\frac{1}{1-\eta}$ is replaced by $\frac{1}{1-\eta^2}$. For kernel with polynomial eigendecay condition, $\bar{\gamma}_T$ will play a role in the overall dynamic regret bound due to its leading term of $\left(\frac{1}{1-\eta^2}\right)^{1/\beta_p}$. However, for kernels with exponential eigendecay condition, it will not affect the overall dynamic regret bound since it only has the logarithmic dependency on $\frac{1}{1-\eta^2}$.

7 Experiments

We numerically compare the performance of IGP-UCB (Chowdhury and Gopalan, 2017), R-GP-UCB (Zhou and Shroff, 2021), SW-GP-UCB(Zhou and Shroff, 2021), WGP-UCB (Algorithm 1) on both synthetic and real-world data. The restarting period H, sliding window SW and exponential weight η are set order-wise by theory (Corollary 5, Remark 1 (Zhou and Shroff, 2021)).

Synthetic data. We develop experiments on both abruptly-changing environments and the slowly-varying environments. We generate the objective function $f \in H_k(D)$ where D is a discretization of $[0, 1]$ into 100 evenly spaced points. We use SE kernel with $l = 0.2$ as our kernel function $k(\cdot, x_i)$ where supporting points $x_i \in D$. The reward function is generated as $f(\cdot) = \sum_{i=1}^{M}\alpha_i k(\cdot, x_i)$ with $\alpha_i \in [-1, 1]$ uniformly sampled and $M = 100$. In the first experiment (Figure 1 (a)), we observe the empirical performance of all algorithms in an abruptly changing environment. The reward function changes at 2 points, i.e., before $t = 100$, $f_t = f_1^*$; for $t \in [100, 200]$, $f_t = f_2^*$; for $t \in [200, 500]$, $f_t = f_3^*$. The second experiment corresponds to a slowly-changing environment (Figure 1 (b)), where when $t <= T/2$, $f_t = f_4^* + (f_5^* - f_4^*)2t/T$, and when $t > T/2$, $f_t = f_5^* + (f_6^* - f_5^*)(2t - T)/T$. All f_i^*'s are randomly sampled within RKHS and the cumulative regret is averaged on 100 independent experiments with error bars in the figure.

Stock market data. We take the adjusted closing price of 29 stocks for 823 days.[1] We use the daily closing price as our time-varying reward function f_t and the empirical covariance of the stock price as our kernel function k. We assume that investors would like to buy one stock upon opening and sell it right before closing, i.e., they want to get much profit as possible after selling it on the same day. The regret is non-sublinear as the rewards in this dataset are heavy-tailed.

Observations. We find that WGP-UCB outperforms three algorithms over all experiments. Moreover, R-

[1] https://www.quandl.com/data/EOD-End-of-Day-US-Stock-Prices

GP-UCB and SW-GP-UCB completely drop outdated information and may not have enough information to make predictions. However, W-GP-UCB, retains outdated information through gradual discounts.

8 Conclusion

In this paper, we develop a framework for regret analysis under weighted Gaussian process regression by overcoming three critical challenges. We propose WGP-UCB algorithm for non-stationary bandit optimization and establish the first regret bound for weighted penalty algorithm in GP bandits. Our future direction is to improve the regret bound when time-varying budget B_T is unknown. It would be interesting to adopt adaptive weights based on non-stationarity detection, via maintaining different instances of algorithms with different starting times (Wei and Luo, 2021).

9 Acknowledgements

This work has been supported in part by NSF grants: 2112471 (also partly funded by DHS), IIS-2007055, CNS-2106933, and CNS-1901057 and a grant from the Army Research Office: W911NF-21-1-0244. We thank all reviewers for their comments and suggestions.

References

Abbasi-Yadkori, Y. (2013). Online learning for linearly parametrized control problems.

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In *NIPS*, volume 11, pages 2312–2320.

Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422.

Auer, P., Gajane, P., and Ortner, R. (2019). Adaptively tracking the best bandit arm with an unknown number of distribution changes. In *Conference on Learning Theory*, pages 138–158. PMLR.

Belkin, M. (2018). Approximation beats concentration? an approximation view on inference with smooth radial kernels. In *Conference On Learning Theory*, pages 1348–1361. PMLR.

Besbes, O., Gur, Y., and Zeevi, A. (2014). Stochastic multi-armed-bandit problem with non-stationary rewards. *Advances in neural information processing systems*, 27:199–207.

Besbes, O., Gur, Y., and Zeevi, A. (2015). Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244.

Besbes, O., Gur, Y., and Zeevi, A. (2019). Optimal exploration–exploitation in a multi-armed bandit problem with non-stationary rewards. *Stochastic Systems*, 9(4):319–337.

Besson, L. and Kaufmann, E. (2019). The generalized likelihood ratio test meets klucb: an improved algorithm for piece-wise non-stationary bandits. *arXiv preprint arXiv:1902.01575*.

Bogunovic, I., Scarlett, J., and Cevher, V. (2016). Time-varying gaussian process bandit optimization. In *Artificial Intelligence and Statistics*, pages 314–323. PMLR.

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*.

Cao, Y., Wen, Z., Kveton, B., and Xie, Y. (2019). Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 418–427. PMLR.

Cheung, W. C., Simchi-Levi, D., and Zhu, R. (2019). Learning to optimize under non-stationarity. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1079–1087. PMLR.

Chowdhury, S. R. and Gopalan, A. (2017). On kernelized multi-armed bandits. *arXiv preprint arXiv:1704.00445*.

Chowdhury, S. R. and Gopalan, A. (2019). Bayesian optimization under heavy-tailed payoffs. *arXiv preprint arXiv:1909.07040*.

Garivier, A. and Moulines, E. (2011). On upper-confidence bound policies for switching bandit problems. In *International Conference on Algorithmic Learning Theory*, pages 174–188. Springer.

Hildebrand, F. B. (1987). *Introduction to numerical analysis*. Courier Corporation.

Huo, X. and Fu, F. (2017). Risk-aware multi-armed bandit problem with application to portfolio selection. *Royal Society open science*, 4(11):171377.

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

Kim, B. and Tewari, A. (2020). Randomized exploration for non-stationary stochastic linear bandits. In *Conference on Uncertainty in Artificial Intelligence*, pages 71–80. PMLR.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670.

Liu, F., Lee, J., and Shroff, N. (2018). A change-detection based framework for piecewise-stationary multi-armed bandit problem. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Micchelli, C. A., Xu, Y., and Zhang, H. (2006). Universal kernels. *Journal of Machine Learning Research*, 7(12).

Mutnỳ, M. and Krause, A. (2019). Efficient high dimensional bayesian optimization with additivity and quadrature fourier features. *Advances in Neural Information Processing Systems 31*, pages 9005–9016.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535.

Russac, Y., Cappé, O., and Garivier, A. (2020). Algorithms for non-stationary generalized linear bandits. *arXiv preprint arXiv:2003.10113*.

Russac, Y., Vernade, C., and Cappé, O. (2019). Weighted linear bandits for non-stationary environments. In *Advances in Neural Information Processing Systems*, pages 12040–12049.

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

Vakili, S., Khezeli, K., and Picheny, V. (2021). On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 82–90. PMLR.

Wei, C.-Y. and Luo, H. (2021). Non-stationary reinforcement learning without prior knowledge: An optimal black-box approach. *arXiv preprint arXiv:2102.05406*.

Zhao, P. and Zhang, L. (2021). Non-stationary linear bandits revisited. *arXiv preprint arXiv:2103.05324*.

Zhao, P., Zhang, L., Jiang, Y., and Zhou, Z.-H. (2020). A simple approach for non-stationary linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 746–755. PMLR.

Zhou, D., Gu, Q., and Szepesvari, C. (2021). Nearly minimax optimal reinforcement learning for linear mixture markov decision processes. In *Conference on Learning Theory*, pages 4532–4576. PMLR.

Zhou, X. and Shroff, N. (2021). No-regret algorithms for time-varying bayesian optimization. In *2021 55th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE.

Zhou, Y., Shen, C., and van der Schaar, M. (2019). A non-stationary online learning approach to mobility management. *IEEE Transactions on Wireless Communications*, 18(2):1434–1446.

Weighted Gaussian Process Bandits for Non-stationary Environments: Supplementary Materials

Appendix

A List of notations

In this section we provide the full list of notations.

- Regularization and weight : $\lambda_t = \lambda w_t, \alpha_t = \lambda w_t^2, w_t = \eta^{-t}$
- Weighted observations : $\tilde{y}_{1:t} = W y_{1:t} = [\sqrt{w_1} y_1, \ldots, \sqrt{w_t} y_t]^T$
- Weight matrix : $W = \mathrm{diag}(\sqrt{w_1}, \sqrt{w_2}, \ldots, \sqrt{w_t})$
- Feature matrix : $\Phi_t = [\varphi(x_1), \ldots, \varphi(x_t)]^T$
- Weighted feature matrix : $\tilde{\Phi}_t = W\Phi_t, \bar{\Phi}_t = W^2\Phi_t, \breve{\Phi}_t = W[\breve{\varphi}(x_1), \ldots, \breve{\varphi}(x_t)]^T$
- Kernel vector : $k_t(x) = \Phi_t \varphi(x)$
- Weighted kernel vector : $\tilde{k}_t(x) = \tilde{\Phi}_t \varphi(x) = W\Phi_t\varphi(x), \breve{k}_t(x) = \breve{\Phi}_t \breve{\varphi}(x)$
- Kernel matrix : $K_t = \Phi_t \Phi_t^T$
- Weighted kernel matrix : $\tilde{K}_t = W\Phi_t\Phi_t^T W^T, \bar{K}_t = W^2\Phi_t\Phi_t^T W^{2T}, \breve{K}_t = \breve{\Phi}_t\breve{\Phi}_t^T$
- Weighted error sum : $S_t = \sum_{s=1}^t w_s \varphi(x_s)\epsilon_s$
- Weighted Gram matrix :
  $V_t = \sum_{s=1}^t w_s \varphi(x_s)\varphi(x_s)^T + \lambda_t I_{\mathcal{H}} = \tilde{\Phi}_t^T\tilde{\Phi}_t + \lambda_t I_{\mathcal{H}} = \Phi_t^T W^2 \Phi_t + \lambda_t I_{\mathcal{H}}$
- Weighted Gram matrix with QFF : $\breve{V}_t = \breve{\Phi}_t^T\breve{\Phi}_t + \lambda_t I_{\mathcal{H}}$
- Double weighted Gram matrix : $\tilde{V}_t = \sum_{s=1}^t w_s^2 \varphi(x_s)\varphi(x_s)^T + \alpha_t I_{\mathcal{H}} = \Phi_t^T W^4 \Phi_t + \alpha_t I_{\mathcal{H}}$
- Predictive variance : $\tilde{\sigma}_t^2(x) = k(x,x) - \tilde{k}_t(x)^T(\tilde{K}_t + \lambda_t I_t)^{-1}\tilde{k}_t(x) = \lambda_t \|\varphi(x)\|^2_{V_t^{-1}}$
- Predictive variance with QFF :
  $\breve{\sigma}_t^2(x) = \breve{k}(x,x) - \breve{k}_t(x)^T(\breve{K}_t + \lambda_t I_t)^{-1}\breve{k}_t(x) = \lambda_t \|\breve{\varphi}(x)\|^2_{\breve{V}_t^{-1}}$
- Loose predictive variance : $\acute{\sigma}_t^2(x) = \lambda \|\varphi(x)\|^2_{V_t^{-1}\tilde{V}_t V_t^{-1}}$
- Confidence bound, $\beta_t = B + \frac{1}{\sqrt{\lambda}} R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_t}$
- Weighted maximum information gain : $\tilde{\gamma}_t = \max_{A\subset D:|A|=t} \frac{1}{2}\log\det(I + \lambda_t^{-1} W K_A W^T) = \max_{A\subset D:|A|=t} \frac{1}{2}\log\det(I + \lambda_t^{-1} W\Phi_t\Phi_t^T W^T)$
- Weighted maximum information gain with QFF : $\breve{\gamma}_t = \frac{1}{2}\log\det(I + \lambda_t^{-1}\breve{\Phi}_t\breve{\Phi}_t^T)$
- Double weighted maximum information gain :
  $\bar{\gamma}_t = \max_{A\subset D:|A|=t} \frac{1}{2}\log\det(I + \alpha_t^{-1} W^2 K_A W^{2T}) =$
  $\max_{A\subset D:|A|=t} \frac{1}{2}\log\det(I + \alpha_t^{-1} W^2\Phi_t\Phi_t^T W^{2T})$
- Combined weighted maximum information gain : $\dot{\gamma}_t = \max\{\bar{\gamma}_t, \breve{\gamma}_t\}$

B Proof of Confidence Bounds

B.1 Connection with weighted linear bandits

The following lemma states that the linear case in Russac et al. (2019) can be recovered by taking $\tilde{\mu}_t(x) = \varphi(x)^T\hat{\theta}_t$ and $\varphi(x) = x$ in WGP-UCB algorithm (Algorithm 1).

Lemma 9. *Equation (3) is equivalent to* $\varphi(x)^T\hat{\theta}_t$ *where* $\hat{\theta}_t = V_t^{-1}\sum_{s=1}^t w_s\varphi(x_s)y_s$ *and* $V_t = \sum_{s=1}^t w_s\varphi(x_s)\varphi(x_s)^T + \lambda_t I_H$.

Proof. As $\hat{\theta}_t$ is the regularized weighted least-squares estimator of θ^* at time t in Russac et al. (2019), we have

$$\begin{aligned}\tilde{\mu}_t(x) &= \varphi(x)^T\hat{\theta}_t = \varphi(x)^T V_t^{-1}\sum_{s=1}^t w_s\varphi(x_s)y_s \\ &= \varphi(x)^T V_t^{-1}\tilde{\Phi}_t^T\tilde{y}_{1:t} = \varphi(x)^T(\tilde{\Phi}_t^T\tilde{\Phi}_t + \lambda_t I_H)^{-1}\tilde{\Phi}_t^T\tilde{y}_{1:t} \\ &= \varphi(x)^T\tilde{\Phi}_t^T(\tilde{\Phi}_t\tilde{\Phi}_t^T + \lambda_t I_t)^{-1}\tilde{y}_{1:t} = \tilde{k}_t(x)^T(\tilde{K}_t + \lambda_t I_t)^{-1}\tilde{y}_{1:t}.\end{aligned}$$

The second last equality holds by $V_t = \tilde{\Phi}_t^T\tilde{\Phi}_t + \lambda_t I_H$ and $(\tilde{\Phi}_t^T\tilde{\Phi}_t + \lambda_t I_H)^{-1}\tilde{\Phi}_t^T = \tilde{\Phi}_t^T(\tilde{\Phi}_t\tilde{\Phi}_t^T + \lambda_t I_t)^{-1}$. □

B.2 Confidence Bounds for stationary environments

In this section we present the detailed proof of confidence bounds for stationary environments.

B.2.1 Self-normalized Concentration

In the following lemma, we shows one concentration inequality about noise sequence ϵ_t. We define weighted error sum as $S_t = \sum_{s=1}^t w_s\varphi(x_s)\epsilon_s \in \mathbb{R}^M$.

Lemma 10. *With probability at least* $1-\delta$*, the following holds simultaneously over all* $t > 0$*:*

$$\|S_t\|_{\tilde{V}_t^{-1}} \leq R\sqrt{2\log(\frac{1}{\delta}) + \log(\det(I_t + \alpha_t^{-1}W^2\Phi_t\Phi_t^T W^{2T})} \tag{7}$$

Proof. This result is adopted from (Abbasi-Yadkori, 2013, Section 3.2).

Similar to (Abbasi-Yadkori, 2013, Equation 3.4), we have $S_t = \sum_{s=1}^t w_s\varphi(x_s)\epsilon_s$ where m_k is replaced by $w_s\varphi(x_s)$ and ϵ_s is R-sub-Gaussian noise. Following (Abbasi-Yadkori, 2013, Equation 3.5), this equation holds $\tilde{V}_t = \sum_{s=1}^t w_s^2\varphi(x_s)\varphi(x_s)^T + \alpha_t I_{\mathcal{H}} = \Phi_t^T W^4\Phi_t + \alpha_t I_{\mathcal{H}}$, where V is replaced by $\alpha_t I_{\mathcal{H}}$ and m_k is replaced by $w_s\varphi(x_s)$. Additionally, we can replace $M_{1:t}$ with $W^2\Phi_t$.

Following the analysis till (Abbasi-Yadkori, 2013, Corollary 3.6), we have the following inequality by replacing $M_{1:t}$ and V respectively,

$$\|S_t\|^2_{\tilde{V}_t^{-1}} \leq 2R^2\log\Big(\frac{\det(I_t + W^2\Phi_t(\alpha_t I_{\mathcal{H}})^{-1}(W^2\Phi_t)^T)^{1/2}}{\delta}\Big).$$

We can get the final result by taking the square root on both sides of the above inequality. □

As $\bar{\gamma}_t = \max_{A\subset D:|A|=t}\frac{1}{2}\log\det(I + \alpha_t^{-1}W^2\Phi_t\Phi_t^T W^{2T}) = \max_{A\subset D:|A|=t}\frac{1}{2}\log\det(I + \alpha_t^{-1}\bar{K}_t)$, we can bound the term $\|S_t\|_{\tilde{V}_t^{-1}}$ as follows,

Lemma 11. *With probability at least* $1-\delta$*, the following holds simultaneously over all* $t > 0$*:*

$$\|S_t\|_{\tilde{V}_t^{-1}} \leq R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_t}, \tag{8}$$

where $\bar{\gamma}_t = \max_{A\subset D:|A|=t}\frac{1}{2}\log\det(I_t + \alpha_t^{-1}\bar{K}_t)$.

We would like to highlight this bound is in terms of double-weighted kernel matrix $\bar{\gamma}_t$ instead of weighted kernel matrix $\tilde{\gamma}_t$.

B.2.2 Proof of Theorem 2

We would provide the detailed proof of Theorem 2 in this section.

Proof. As $\hat{\theta}_t$ is the regularized weighted least-squares estimator of θ^* at time t in Russac et al. (2019) and $V_t = \sum_{s=1}^{t} w_s \varphi(x_s)\varphi(x_s)^T + \lambda_t I_H$, we have

$$\begin{aligned}
\tilde{\mu}_t(x) &= \varphi(x)^T \hat{\theta}_t = \varphi(x)^T V_t^{-1} \sum_{s=1}^{t} w_s \varphi(x_s) y_s \\
&= \varphi(x)^T V_t^{-1} [\sum_{s=1}^{t} \left(w_s \varphi(x_s) f^*(x_s) + w_s \varphi(x_s) \epsilon_s \right)] \\
&= \varphi(x)^T V_t^{-1} [\sum_{s=1}^{t} w_s \varphi(x_s)\varphi(x_s)^T \theta^* + \lambda_t \theta^* - \lambda_t \theta^* + S_t] \\
&= \varphi(x)^T \theta^* - \lambda_t \varphi(x)^T V_t^{-1} \theta^* + \varphi(x)^T V_t^{-1} S_t.
\end{aligned}$$

We have $\tilde{\mu}_t(x) - f^*(x) = \tilde{\mu}_t(x) - \varphi(x)^T\theta^* = \varphi(x)^T V_t^{-1} S_t - \lambda_t \varphi(x)^T V_t^{-1} \theta^*$, therefore

$$\begin{aligned}
|\tilde{\mu}_t(x) - f^*(x)| &\leq \|\varphi(x)\|_{V_t^{-1}\tilde{V}_t V_t^{-1}} \left(\|V_t^{-1} S_t\|_{V_t \tilde{V}_t^{-1} V_t} + \|\lambda_t V_t^{-1} \theta^*\|_{V_t \tilde{V}_t^{-1} V_t} \right) \\
&\leq \|\varphi(x)\|_{V_t^{-1}\tilde{V}_t V_t^{-1}} \left(\|S_t\|_{\tilde{V}_t^{-1}} + \lambda_t \|\theta^*\|_{\tilde{V}_t^{-1}} \right)
\end{aligned}$$

Knowing that $\tilde{V}_t \succeq \alpha_t I_{\mathcal{H}}$ and $\tilde{V}_t$ is positive definite, we have $\|\theta^*\|_{\tilde{V}_t^{-1}} \leq \frac{1}{\sqrt{\alpha_t}} \|\theta^*\|_2$. With $\acute{\sigma}_t^2(x) = \lambda \|\varphi(x)\|^2_{V_t^{-1}\tilde{V}_t V_t^{-1}}$, we have

$$|\tilde{\mu}_t(x) - f^*(x)| \leq \frac{\acute{\sigma}_t(x)}{\sqrt{\lambda}} \left(\|S_t\|_{\tilde{V}_t^{-1}} + \frac{\lambda_t}{\sqrt{\alpha_t}} \|\theta\|_2 \right).$$

Given $\|f^*\|_H = \|\theta\|_2 \leq B$, we have

$$\begin{aligned}
|\tilde{\mu}_t(x) - f^*(x)| &\leq \frac{\acute{\sigma}_t(x)}{\sqrt{\lambda}} \left(\|S_t\|_{\tilde{V}_t^{-1}} + \frac{\lambda_t}{\sqrt{\alpha_t}} B \right) \\
&\leq \frac{\lambda_t}{\sqrt{\lambda \alpha_t}} \acute{\sigma}_t(x) B + \frac{\acute{\sigma}_t(x)}{\sqrt{\lambda}} R \sqrt{2 \log(\frac{1}{\delta}) + 2\bar{\gamma}_t} \\
&= \acute{\sigma}_t(x) B + \frac{\acute{\sigma}_t(x)}{\sqrt{\lambda}} R \sqrt{2 \log(\frac{1}{\delta}) + 2\bar{\gamma}_t} \\
&= \acute{\sigma}_t(x) \beta_t,
\end{aligned}$$

where $\beta_t = B + \frac{1}{\sqrt{\lambda}} R \sqrt{2 \log(\frac{1}{\delta}) + 2\bar{\gamma}_t}$. □

B.2.3 Proof of Remark 2

The following lemma shows that Equation (4) is equivalent to $\tilde{\sigma}_t^2(x) = \lambda_t ||\varphi(x)||^2_{V_t^{-1}}$.

Lemma 12. *Equation (4) is equivalent to* $\tilde{\sigma}_t^2(x) = \lambda_t ||\varphi(x)||^2_{V_t^{-1}}$.

Proof. As $I_{\mathcal{H}} - \tilde{\Phi}_t^T (\tilde{\Phi}_t \tilde{\Phi}_t^T + \lambda_t I_t)^{-1} \tilde{\Phi}_t = \lambda_t (\tilde{\Phi}_t^T \tilde{\Phi}_t + \lambda_t I_{\mathcal{H}})^{-1} = \lambda_t V_t^{-1}$, we have

$$\begin{aligned}
\tilde{\sigma}_t(x)^2 &= k(x,x) - \tilde{k}_t(x)^T (\tilde{K}_t + \lambda_t I_t)^{-1} \tilde{k}_t(x) \\
&= \varphi(x)^T \varphi(x) - \varphi(x)^T \tilde{\Phi}_t^T (\tilde{\Phi}_t \tilde{\Phi}_t^T + \lambda_t I_t)^{-1} \tilde{\Phi}_t \varphi(x) \\
&= \varphi(x)^T [I_{\mathcal{H}} - \tilde{\Phi}_t^T (\tilde{\Phi}_t \tilde{\Phi}_t^T + \lambda_t I_t)^{-1} \tilde{\Phi}_t] \varphi(x) \\
&= \varphi(x)^T \lambda_t V_t^{-1} \varphi(x) = \lambda_t ||\varphi(x)||_{V_t^{-1}}.
\end{aligned}$$

□

Assume $\lambda = 1$ and $w_t = 1$. Then the followings hold; $V_t = \tilde{V}_t$ and $\lambda_t = \lambda$, thus $\acute{\sigma}_t(x) = \tilde{\sigma}_t(x)$ and $\tilde{\mu}_t(x) = \mu_t(x)$. In the above Lemma 12, we have $\acute{\sigma}_t(x) = \sigma_t(x) = k(x,x) - k_t(x)^T(K_t + \lambda I)^{-1}k_t(x)$ and $\bar{\gamma}_t = \gamma_t$, which makes Theorem 2 equivalent to (Chowdhury and Gopalan, 2017, Theorem 2).

B.3 Confidence bounds for non-stationary cases

In this section we provide the relatively loose regret bound in terms of $\tilde{\sigma}_t(x)$ and then detailed proof of Theorem 3 that the surrogate parameter $m_t(x)$ lies in the confidence ellipsoid defined in Theorem 2 with high probability.

First, we further restrict that $\acute{\sigma}_t(x) \geq 0$ and $\tilde{\sigma}_t(x) \geq 0$, then we have the following lemma.

Lemma 13. *If $\{w_s\}_{s=1}^t$ is increasing, then $\acute{\sigma}_t(x) \leq \tilde{\sigma}_t(x)$.*

Proof. We recall that $\acute{\sigma}_t^2(x) = \lambda\|\varphi(x)\|^2_{V_t^{-1}\tilde{V}_t V_t^{-1}}$ by definition and $\tilde{\sigma}_t^2(x) = \lambda_t||\varphi(x)||^2_{V_t^{-1}}$ from Lemma 12. As $V_t = \sum_{s=1}^t w_s\varphi(x_s)\varphi(x_s)^T + \lambda w_t I_{\mathcal{H}}$, we have

$$\tilde{V}_t = \sum_{s=1}^t w_s^2\varphi(x_s)\varphi(x_s)^T + \lambda w_t^2 I_{\mathcal{H}} \leq w_t \sum_{s=1}^t w_s\varphi(x_s)\varphi(x_s)^T + \lambda w_t^2 I_{\mathcal{H}} \leq w_t V_t.$$

Therefore, $V_t^{-1}\tilde{V}_t V_t^{-1} \leq w_t V_t^{-1} V_t V_t^{-1} \leq w_t V_t^{-1}$ and $\acute{\sigma}_t^2(x) \leq \tilde{\sigma}_t^2(x)$ since $\lambda_t = \lambda w_t$. □

We would state the full proof of Theorem 3 as follows.

Proof of Theorem 3. We would obtain $\tilde{\mu}_t(x_t) = \varphi(x_t)^T V_t^{-1}\tilde{\Phi}_t^T\tilde{y}_{1:t}$ from the definition of posterior mean $\tilde{\mu}_t(x)$ and proof of Lemma 9. Then, we would get the followings,

$$\begin{aligned}
&m_t(x) - \tilde{\mu}_{t-1}(x) \\
&= \varphi(x)^T V_{t-1}^{-1}[\sum_{s=1}^{t-1}\eta^{-s}\varphi(x_s)f_s(x_s) + \lambda\eta^{-(t-1)}\theta_t^* - \sum_{s=1}^{t-1}\eta^{-s}\varphi(x_s)y_s] \\
&= \varphi(x)^T V_{t-1}^{-1}[\sum_{s=1}^{t-1}\eta^{-s}\varphi(x_s)f_s(x_s) + \lambda\eta^{-(t-1)}\theta_t^* - \sum_{s=1}^{t-1}\eta^{-s}\varphi(x_s)f_s(x_s) - \sum_{s=1}^{t-1}\eta^{-s}\varphi(x_s)\epsilon_s] \\
&= -\varphi(x)^T V_{t-1}^{-1}S_{t-1} + \lambda\eta^{-(t-1)}\varphi(x)^T V_{t-1}^{-1}\theta_t^*.
\end{aligned}$$

Then, the distance between surrogate parameter and posterior mean is bounded as,

$$\begin{aligned}
|m_t(x) - \tilde{\mu}_{t-1}(x)| &\leq |\varphi^T(x)V_{t-1}^{-1}S_{t-1}| + \lambda\eta^{-(t-1)}|\varphi(x)^T V_{t-1}^{-1}\theta_t^*| \\
&\leq \|\varphi(x)\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}}\|V_{t-1}^{-1}S_{t-1}\|_{V_{t-1}\tilde{V}_{t-1}^{-1}V_{t-1}} \\
&\quad + \lambda\eta^{-(t-1)}\|\varphi(x)\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}}\|V_{t-1}^{-1}\theta_t^*\|_{V_{t-1}\tilde{V}_{t-1}^{-1}V_{t-1}} \\
&\leq \|\varphi(x)\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_t^{-1}}\|S_{t-1}\|_{\tilde{V}_{t-1}^{-1}} + \lambda\eta^{-(t-1)}\|\varphi(x)\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}}\|\theta_t^*\|_{\tilde{V}_{t-1}^{-1}} \\
&\leq \frac{\acute{\sigma}_{t-1}(x)}{\sqrt{\lambda}}\|S_{t-1}\|_{\tilde{V}_{t-1}^{-1}} + \lambda_{t-1}\frac{\acute{\sigma}_{t-1}(x)}{\sqrt{\lambda}}\frac{1}{\sqrt{\alpha_{t-1}}}\|\theta_t^*\|_2 \\
&\leq \frac{\acute{\sigma}_{t-1}(x)}{\sqrt{\lambda}}R\sqrt{2\log(\frac{1}{\delta}) + 2\bar{\gamma}_{t-1}} + \acute{\sigma}_{t-1}(x)B \\
&\leq \acute{\sigma}_{t-1}(x)\beta_{t-1}.
\end{aligned}$$

The final two steps are because $\|\theta_t^*\|_{\tilde{V}_{t-1}^{-1}} \leq \frac{1}{\sqrt{\alpha_{t-1}}}\|\theta_t^*\|_2$ and $\|\theta_t^*\|_2 = \|f_t^*\|_H \leq B$. Due to the above Lemma 13, we obtain the following inequality, $|m_t(x) - \tilde{\mu}_{t-1}(x)| \leq \tilde{\sigma}_{t-1}(x)\beta_{t-1}$. □

C Proof of Regret Bound

In this section, we state the detailed analysis of dynamic regret of WGP-UCB (Algorithm 1). As $\lambda_t = \lambda w_t$, the weighted GP regression problem is equivalent to the following problem, where the time-dependent weight is $w'_{s,t} = w_s/w_t$ and regularization factor is time-independent.

$$\min_{f \in H_k(D)} \sum_{s=1}^{t-1} w'_{s,t}(y_s - f(x_s))^2 + \lambda \|f\|_{\mathcal{H}}^2$$

C.1 Approximation error

First, we would explicitly obtain the approximation error.

Lemma 14. *(QFF error (Chowdhury and Gopalan, 2019, Lemma 14)) If $D = [0,1]^d$, $k = k_{SE}$, then,*

$$\epsilon_m \leq d2^{d-1} \frac{1}{\sqrt{2}\bar{m}^{\bar{m}}} \left(\frac{e}{4l^2}\right)^{\bar{m}} = O\left(\frac{d2^{d-1}}{(\bar{m}l^2)^{\bar{m}}}\right).$$

Lemma 15. *Let $f \in H_k(D)$, $||f||_H \leq B$ and $k(x,x) \leq 1$ for all $x \in D$. Then we have $|\tilde{\sigma}_t(x) - \breve{\sigma}_t(x)| = O(\frac{\sqrt{\epsilon_m}}{1-\eta})$.*

Proof. First we define $a_t(x) = \tilde{k}_t(x) - \breve{k}_t(x)$, have $||a_t(x)||_2 \leq \epsilon_m \sqrt{\frac{\lambda}{\eta^t(1-\eta)}}$ as well as $||\tilde{k}_t(x)||_2 \leq \sqrt{\frac{\lambda}{\eta^t(1-\eta)}}$.

Similarly to the proof of Lemma 15 in Chowdhury and Gopalan (2019), we bound the approximation error between inverse kernel matrices as,

$$\begin{aligned}
&\|(\tilde{K}_t + \lambda_t I_t)^{-1} - (\breve{K}_t + \lambda_t I_t)^{-1}\|_2 \\
&= \|(\tilde{K}_t + \lambda_t I_t)^{-1}\Big((\tilde{K}_t + \lambda_t I_t) - (\breve{K}_t + \lambda_t I_t)\Big)(\breve{K}_t + \lambda_t I_t)^{-1}\|_2 \\
&= \|(\tilde{K}_t + \lambda_t I_t)^{-1}(\tilde{K}_t - \breve{K}_t)(\breve{K}_t + \lambda_t I_t)^{-1}\|_2 \\
&\leq \|(\tilde{K}_t + \lambda_t I_t)^{-1}\|_2 \|\tilde{K}_t - \breve{K}_t\|_2 \|(\breve{K}_t + \lambda_t I_t)^{-1}\|_2 \\
&\leq \frac{1}{\lambda_t} \frac{\epsilon_m \lambda}{\eta^t(1-\eta)} \frac{1}{\lambda_t} = \frac{\epsilon_m \lambda}{\eta^t(1-\eta)\lambda_t^2}.
\end{aligned}$$

The last inequality holds because $\|\tilde{K}_t - \breve{K}_t\|_2^2 \leq \sum_{1 \leq i,j \leq t} \Big((k(x_i,x_j) - \breve{k}(x_i,x_j))\lambda\sqrt{\eta^{-i-j}}\Big)^2 \leq \lambda^2 \epsilon_m^2 \sum_{1 \leq i \leq t} \eta^{-i} \sum_{1 \leq j \leq t} \eta^{-j} \leq \frac{\epsilon_m^2 \lambda^2}{\eta^{2t}(1-\eta)^2}$ and $\|(\tilde{K}_t + \lambda_t I_t)^{-1}\|_2 \leq \frac{1}{\lambda_t}$.

Therefore, we have

$$\begin{aligned}
&|\tilde{\sigma}_t^2(x) - \breve{\sigma}_t^2(x)| \\
&= |k(x,x) - \tilde{k}_t(x)^T(\tilde{K}_t + \lambda_t I_t)^{-1}\tilde{k}_t(x) - \breve{k}(x,x) - \breve{k}_t(x)^T(\breve{K}_t + \lambda_t I_t)^{-1}\breve{k}_t(x)| \\
&\leq |k(x,x) - \breve{k}(x,x)| + |\tilde{k}_t(x)^T(\tilde{K}_t + \lambda_t I_t)^{-1}\tilde{k}_t(x) - \breve{k}_t(x)^T(\breve{K}_t + \lambda_t I_t)^{-1}\breve{k}_t(x)| \\
&\leq \epsilon_m + |\tilde{k}_t(x)^T\Big((\tilde{K}_t + \lambda_t I_t)^{-1} - (\breve{K}_t + \lambda_t I_t)^{-1}\Big)\tilde{k}_t(x)| \\
&+ 2|a_t(x)^T(\breve{K}_t + \lambda_t I_t)^{-1}\tilde{k}_t(x)| + |a_t(x)^T(\breve{K}_t \lambda_t I_t)^{-1}a_t(x)| \\
&\leq \epsilon_m + \|(\tilde{K}_t + \lambda_t I_t)^{-1} - (\breve{K}_t + \lambda_t I_t)^{-1}\|_2 \|\tilde{k}_t(x)\|_2^2 \\
&+ 2\|a_t(x)\|_2 \|(\breve{K}_t + \lambda_t I_t)^{-1}\|_2 \|\tilde{k}_t(x)\|_2 + \|(\breve{K}_t \lambda_t I_t)^{-1}\|_2 \|a_t(x)\|_2^2 \\
&\leq \epsilon_m + \frac{\epsilon_m \lambda}{\eta^t(1-\eta)\lambda_t^2} \frac{\lambda}{\eta^t(1-\eta)} + 2\sqrt{\frac{\lambda}{\eta^t(1-\eta)}}\epsilon_m \frac{1}{\lambda_t}\sqrt{\frac{\lambda}{\eta^t(1-\eta)}} + \frac{1}{\lambda_t}\epsilon_m^2 \frac{\lambda}{\eta^t(1-\eta)} \\
&= O\left(\frac{\epsilon_m}{(1-\eta)^2}\right).
\end{aligned}$$

Then, the proof is completed from

$$|\tilde{\sigma}_t(x) - \breve{\sigma}_t(x)|^2 \leq |\tilde{\sigma}_t(x) + \breve{\sigma}_t(x)||\tilde{\sigma}_t(x) - \breve{\sigma}_t(x)| \leq |\tilde{\sigma}_t^2(x) - \breve{\sigma}_t^2(x)|.$$

□

C.2 Bound of $\sum_{t=1}^{T} \tilde{\sigma}_{t-1}(x_t)$

In this section we would describe the way to obtain the tight bound of $\sum_{t=1}^{T} \tilde{\sigma}_{t-1}(x_t)$

Lemma 16. $\sum_{t=1}^{T} \breve{\sigma}_{t-1}(x_t) \leq \sqrt{4\lambda T\breve{\gamma}_T + 2\lambda m T^2 \log(1/\eta)}$.

Proof. Assume the feature map has a finite dimension m, then

$$\sum_{t=1}^{T} \breve{\sigma}_{t-1}(x_t) \leq \sqrt{T\sum_{t=1}^{T} \breve{\sigma}_{t-1}^2(x_t)} \leq \sqrt{T\sum_{t=1}^{T} 2\lambda \log(1 + \frac{1}{\lambda}\breve{\sigma}_{t-1}^2(x_t))}$$
$$\leq \sqrt{T\sum_{t=1}^{T} 2\lambda \log(1 + \eta^{-(t-1)}||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}})} \leq \sqrt{2\lambda T\sum_{t=1}^{T} \log(1 + \eta^{-t}||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}})}.$$

Due to $\breve{V}_t \geq \breve{V}_{t-1} + \eta^{-t}\breve{\varphi}(x_t)\breve{\varphi}(x_t)^T \geq \breve{V}_{t-1}^{1/2}(I_{\mathcal{H}} + \eta^{-t}\breve{V}_{t-1}^{-1/2}\breve{\varphi}(x_t)\breve{\varphi}(x_t)^T\breve{V}_{t-1}^{-1/2})\breve{V}_{t-1}^{1/2}$, we have

$$\begin{aligned}\det(\breve{V}_t) &\geq \det(\breve{V}_{t-1})\det(I_{\mathcal{H}} + \eta^{-t/2}\breve{V}_{t-1}^{-1/2}\breve{\varphi}(x_t)\big(\eta^{-t/2}\breve{V}_{t-1}^{-1/2}\breve{\varphi}(x_t)\big)^T)\\ &\geq \det(\breve{V}_{t-1})(1 + \eta^{-t}||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}}),\end{aligned}$$

and then the following bound holds.

$$\sum_{t=1}^{T} \log(1 + \eta^{-t}||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}}) \leq \sum_{t=1}^{T} \log(\frac{\det(\breve{V}_t)}{\det(\breve{V}_{t-1})}) \leq \log(\Pi_{t=1}^{T}\frac{\det(\breve{V}_t)}{\det(\breve{V}_{t-1})}) \leq \log(\frac{\det(\breve{V}_T)}{\det(\breve{V}_0)}).$$

From matrix determinant lemma stating $\det(A + UV^T) = \det(I + V^TA^{-1}U)\det(A)$ and $V_0 = \lambda I_{\mathcal{H}} \in \mathbb{R}^{m\times m}$, $\det(\breve{V}_T)$ is decomposed as,

$$\begin{aligned}\det(\breve{V}_T) &= \det(\breve{\Phi}_T^T\breve{\Phi}_T + \lambda_T I_{\mathcal{H}}) = \det(I_t + \breve{\Phi}_T(\lambda_T I_{\mathcal{H}})^{-1}\breve{\Phi}_T^T)\det(\lambda_T I_{\mathcal{H}})\\ &= \det(I_t + \lambda_T^{-1}\breve{\Phi}_T\breve{\Phi}_T^T)\det(\eta^{-T}V_0) = \det(I_t + \lambda_T^{-1}\breve{\Phi}_T\breve{\Phi}_T^T)\eta^{-mT}\det(V_0)\end{aligned}$$

Thus, we get

$$\log(\frac{\det(\breve{V}_T)}{\det(\breve{V}_0)}) = \log\det(I_t + \lambda_T^{-1}\breve{\Phi}_T\breve{\Phi}_T^T) + mT\log(\eta^{-1}) \leq 2\breve{\gamma}_T + mT\log(1/\eta)$$

Therefore, we have

$$\sum_{t=1}^{T} \breve{\sigma}_{t-1}(x_t) \leq \sqrt{2\lambda T\sum_{t=1}^{T} \log(1 + \eta^{-t}||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}})} \leq \sqrt{4\lambda T\breve{\gamma}_T + 2\lambda m T^2\log(1/\eta)}.$$

□

By combining Lemma 15 and 16, we have the following lemma.

Lemma 17.

$$\beta_T\sum_{t=1}^{T} \tilde{\sigma}_{t-1}(x_t) \leq \beta_T\sqrt{4\lambda T\breve{\gamma}_T + 2\lambda m T^2\log(1/\eta)} + \frac{\beta_T T\sqrt{\epsilon_m}}{1-\eta}.$$

C.3 Bound of $\sum_{s=T-c}^{T} \acute{\sigma}_{t-1}(x_s)$

In this section we would describe the way to obtain the tight bound of partial sum $\sum_{s=t-c}^{t} \acute{\sigma}_{t-1}(x_s)$.

Lemma 18. $\sum_{s=t-c}^{t} \breve{\sigma}_{t-1}(x_s) \leq \sqrt{4\lambda c \breve{\gamma}_t + 2\lambda m c^2 \log(1/\eta)}$.

Proof. Similarly to the proof of Lemma 16, we get the following bound.

$$\sum_{s=t-c}^{t} \breve{\sigma}_{t-1}(x_s) \leq \sqrt{c \sum_{s=t-c}^{t} \breve{\sigma}^2_{t-1}(x_s)} \leq \sqrt{c \sum_{s=t-c}^{t} 2\lambda \log\left(1 + \lambda^{-1} \breve{\sigma}^2_{t-1}(x_s)\right)}$$
$$\leq \sqrt{4c\lambda \sum_{s=t-c}^{t} \frac{1}{2} \log\left(1 + \lambda^{-1} \breve{\sigma}^2_{t-1}(x_s)\right)} \leq \sqrt{4\lambda c \sum_{s=t-c}^{t} \frac{1}{2} \log(1 + \eta^{-t} ||\breve{\varphi}(x_s)||^2_{\breve{V}_{t-1}^{-1}})}.$$

Due to $\det(\breve{V}_t) \geq \det(\breve{V}_{t-1})(1 + \eta^{-t} ||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}})$, the upper bound can be derived as,

$$\sum_{t=T-c}^{T} \log(1 + \eta^{-t} ||\breve{\varphi}(x)||^2_{\breve{V}_{t-1}^{-1}}) \leq \sum_{t=T-c}^{T} \log(\frac{\det(\breve{V}_t)}{\det(\breve{V}_{t-1})}) \leq \log(\frac{\det(\breve{V}_T)}{\det(\breve{V}_{T-c})}).$$

From matrix determinant lemma, we have

$$\det(\breve{V}_{T-c}) = \det(I_t + \lambda_{T-c}^{-1} \breve{\Phi}_{T-c} \breve{\Phi}_{T-c}^T) \eta^{-m(T-c)} \det(\breve{V}_0) \geq \eta^{-m(T-c)} \det(\breve{V}_0).$$

Thus we get

$$\log(\frac{\det(\breve{V}_T)}{\det(\breve{V}_{T-c})}) \leq \log\det(I_t + \lambda_T^{-1} \breve{\Phi}_T \breve{\Phi}_T^T) + mc \log(\eta^{-1}) \leq 2\breve{\gamma}_T + mc \log(1/\eta).$$

Therefore, the partial sum $\sum_{s=t-c}^{t} \breve{\sigma}_{t-1}(x_s)$ is bounded as below.

$$\sum_{t=T-c}^{T} \breve{\sigma}_{t-1}(x_t) \leq \sqrt{4\lambda c \sum_{s=t-c}^{t} \frac{1}{2} \log(1 + \eta^{-t} ||\breve{\varphi}(x_s)||^2_{\breve{V}_{t-1}^{-1}})} \leq \sqrt{4\lambda c \breve{\gamma}_T + 2\lambda m c^2 \log(1/\eta)}.$$

□

By combining Lemma 13, 15 and 18, the following lemma can be derived.

Lemma 19. $\sum_{s=t-c}^{t} \acute{\sigma}_{t-1}(x_s) \leq \sqrt{4\lambda c \breve{\gamma}_t + 2\lambda m c^2 \log(1/\eta)} + \frac{c\sqrt{\epsilon_m}}{1-\eta}$

C.4 Preliminary results

The following lemma provides the upper bound the regret of WGP-UCB (Algorithm 1) in terms of $\tilde{\sigma}_{t-1}(x_t), \acute{\sigma}_{t-1}(x_t)$, c, and γ.

Lemma 20. *Let $f_t \in H_k(D)$, $\|f_t\|_H \leq B$ and $k(x,x) \leq 1$. Then, with probability at least $1-\delta$,*

$$R_T \leq 2\beta_T \sum_{t=1}^{T} \tilde{\sigma}_{t-1}(x_t) + \frac{2}{\lambda} c B_T \sum_{s=T-c}^{T} \acute{\sigma}_{s-1}(x_s) + \frac{4B\eta^c}{\lambda(1-\eta)} T,$$

where $c \geq 1$ is an integer and $0 < \eta < 1$.

Proof. The one time step regret r_t is decomposed into the stationary part (first two terms) and non-stationary part (remaining terms) as

$$r_t = f_t(x_t^*) - f_t(x_t) = m_t(x_t^*) - m_t(x_t) + f_t(x_t^*) - m_t(x_t^*) - (f_t(x_t) - m_t(x_t)).$$

We bound the stationary part as,

$$\begin{aligned} m_t(x_t^*) - m_t(x_t) &\leq \tilde{\mu}_{t-1}(x_t^*) + \beta_{t-1}\tilde{\sigma}_{t-1}(x_t^*) - (\tilde{\mu}_{t-1}(x_t) - \beta_{t-1}\tilde{\sigma}_{t-1}(x_t)) \\ &\leq \tilde{\mu}_{t-1}(x_t) + \beta_{t-1}\tilde{\sigma}_{t-1}(x_t) - (\tilde{\mu}_{t-1}(x_t) - \beta_{t-1}\tilde{\sigma}_{t-1}(x_t)) \\ &\leq 2\beta_{t-1}\tilde{\sigma}_{t-1}(x_t), \end{aligned}$$

where the first inequality holds by Theorem 3 stating $|m_t(x) - \tilde{\mu}_{t-1}(x)| \leq \beta_{t-1}\acute{\sigma}_{t-1}(x)$, and the second inequality works by the nature of UCB-type algorithm, i.e. x_t defined in Equation (5) is the arm chosen at time t and thus the following holds, $\tilde{\mu}_{t-1}(x_t^*) + \beta_{t-1}\tilde{\sigma}_{t-1}(x_t^*) \leq \tilde{\mu}_{t-1}(x_t) + \beta_{t-1}\tilde{\sigma}_{t-1}(x_t)$.

For the non-stationary part, we have $f_t(x) = \theta_t^{*T}\varphi(x)$ and $m(x) = \bar{\theta}_t^T\varphi(x)$ from Mercer theorem, and $\|f\|_H^2 = \|\theta\|_2^2 \leq B$. Then, we bound the non-stationary part in terms of distance between surrogate parameter $\bar{\theta}_t$ and true parameter θ^* as below,

$$\begin{aligned} f_t(x_t^*) - m_t(x_t^*) - (f_t(x_t) - m_t(x_t)) &= \left\langle \theta_t^* - \bar{\theta}_t, \varphi(x^*) - \varphi(x) \right\rangle \\ &\leq \|\varphi(x^*) - \varphi(x)\|_2 \|\theta_t^* - \bar{\theta}_t\|_2 \\ &\leq 2\|\varphi(x)\|_2 \|\theta_t^* - \bar{\theta}_t\|_2 \leq 2\|\theta_t^* - \bar{\theta}_t\|_2 \end{aligned}$$

where the last inequality holds due to $\|\varphi(x)\|_2^2 = \varphi(x)^T\varphi(x) = k(x,x) \leq 1$.

As pointed out by (Zhao and Zhang, 2021, p.4), the statement $\lambda_{max}(V_{t-1}^{-1}\sum_{s=t-D}^{p}\eta^{-s}A_sA_s^T) \leq 1$ in (Russac et al., 2019, p.18) is not true. We would fix this error in the following lemma 21 (proved at the end of this subsection). We recall some definitions of V_t, θ_t^*, and $\bar{\theta}_t$ as

$$\begin{aligned} V_t &= \sum_{s=1}^{t} w_s\varphi(x_s)\varphi(x_s)^T + \lambda_t I_H, \\ \theta_t^* &= V_{t-1}^{-1}\sum_{s=1}^{t-1} w_s\varphi(x_s)\varphi(x_s)^T\theta_t^* + V_{t-1}^{-1}\lambda_{t-1}\theta_t^*, \\ \bar{\theta}_t &= V_{t-1}^{-1}[\sum_{s=1}^{t-1}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\theta_s^* + \lambda\eta^{-(t-1)}\theta_t^*]. \end{aligned}$$

Lemma 21.

$$\left\|V_{t-1}^{-1}\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\right\|_2 \leq \frac{1}{\lambda}\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s)$$

Then we would bound the distance between surrogate parameter $\bar{\theta}_t$ and true parameter θ^* as below.

$$
\begin{aligned}
&\|\theta_t^* - \bar{\theta}_t\|_2 = \|V_{t-1}^{-1} \sum_{s=1}^{t} w_s \varphi(x_s)\varphi(x_s)^T(\theta_s^* - \theta_t^*)\|_2 \\
\leq &\|\sum_{s=t-c}^{t-1} V_{t-1}^{-1}\eta^{-s}\varphi(x_s)\varphi(x_s)^T(\theta_s^* - \theta_t^*)\|_2 + \|V_{t-1}^{-1}\sum_{s=1}^{t-c-1}\eta^{-s}\varphi(x_s)\varphi(x_s)^T(\theta_s^* - \theta_t^*)\|_2 \\
\leq &\|\sum_{s=t-c}^{t-1} V_{t-1}^{-1}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\sum_{p=s}^{t-1}(\theta_p^* - \theta_{p+1}^*)\|_2 + \|\sum_{s=1}^{t-c-1}\eta^{-s}\varphi(x_s)\varphi(x_s)^T(\theta_s^* - \theta_t^*)\|_{V_{t-1}^{-2}} \\
\leq &\|\sum_{p=t-c}^{t-1}\sum_{s=t-c}^{p} V_{t-1}^{-1}\eta^{-s}\varphi(x_s)\varphi(x_s)^T(\theta_p^* - \theta_{p+1}^*)\|_2 + \frac{1}{\lambda}\sum_{s=1}^{t-c-1}\eta^{t-1-s}\|\varphi(x_s)\varphi(x_s)^T(\theta_s^* - \theta_t^*)\|_2 \\
\leq &\sum_{p=t-c}^{t-1}\|V_{t-1}^{-1}\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\varphi(x_s)^T(\theta_p^* - \theta_{p+1}^*)\|_2 + \frac{2B}{\lambda}\sum_{s=1}^{t-c-1}\eta^{t-1-s} \\
\leq &\sum_{p=t-c}^{t-1}\|V_{t-1}^{-1}\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\|_2 \cdot \|(\theta_p^* - \theta_{p+1}^*)\|_2 + \frac{2B}{\lambda}\sum_{s=1}^{t-c-1}\eta^{t-1-s} \\
\leq &\sum_{p=t-c}^{t-1}\|\theta_p^* - \theta_{p+1}^*\|_2 \frac{1}{\lambda}\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s) + \frac{2B\eta^c}{\lambda(1-\eta)}.
\end{aligned}
$$

The third inequality holds by $V_t^{-2} \leq (\frac{\eta^{t-1}}{\lambda})^2 I_{\mathcal{H}}$, and the fourth inequality works due to $\|\theta^*\|_2 \leq B$ and $\|\varphi(x)\|_2^2 = k(x,x) \leq 1$. The last inequality holds from Lemma 21.

Accordingly, we would obtain the following upper bound for non-stationary part.

$$
f_t(x_t^*) - m_t(x_t^*) - (f_t(x_t) - m_t(x_t)) \leq 2\sum_{p=t-c}^{t-1}\|f_p - f_{p+1}\|_H \frac{1}{\lambda}\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s) + \frac{4B\eta^c}{\lambda(1-\eta)}.
$$

By combining bounds for both stationary and non-stationary part, the dynamic regret is bounded as,

$$
\begin{aligned}
R_T &= \sum_{t=1}^{T} r_t \\
&\leq 2\beta_T\sum_{t=1}^{T}\tilde{\sigma}_{t-1}(x_t) + 2\sum_{t=1}^{T}\sum_{p=t-c}^{t-1}\|f_p - f_{p+1}\|_H\frac{1}{\lambda}\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s) + \frac{4B\eta^c}{\lambda(1-\eta)}T \\
&\leq 2\beta_T\sum_{t=1}^{T}\tilde{\sigma}_{t-1}(x_t) + \frac{2}{\lambda}cB_T\sum_{s=T-c}^{T}\acute{\sigma}_{t-1}(x_s) + \frac{4B\eta^c}{\lambda(1-\eta)}T
\end{aligned}
$$

□

Proof of Lemma 21. We denote the unit ball as $\mathbb{B}(1) = \{z : \|z\|_2 = 1\}$ and the optimizer as $z^\star$.

$$
\begin{aligned}
\big\|V_{t-1}^{-1}\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\big\|_2 &= \sup_{z\in\mathbb{B}(1)}\big|z^T V_{t-1}^{-1}\big(\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\big)z\big| \\
&\le \|V_{t-1}^{-1}z^\star\|_{V_{t-1}\tilde{V}_{t-1}^{-1}V_{t-1}}\big\|\big(\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\varphi(x_s)^T\big)z^\star\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}} \\
&\le \|z^\star\|_{\tilde{V}_{t-1}^{-1}}\big\|\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\cdot\|\varphi(x_s)\|\cdot\|z^\star\|\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}} \\
&\le \frac{1}{\sqrt{\alpha_{t-1}}}\big\|\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}} \\
&\le \frac{\eta^{t-1}}{\sqrt{\lambda}}\big\|\sum_{s=t-c}^{p}\eta^{-s}\varphi(x_s)\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}}\Big(\text{Because } \tilde{V}_t \succeq \alpha_t I_{\mathcal{H}} \text{ and } \|z^*\|_{\tilde{V}_t^{-1}} \le \frac{1}{\sqrt{\alpha_t}}\|z^*\|_2\Big) \\
&\le \frac{1}{\sqrt{\lambda}}\big\|\sum_{s=t-c}^{p}\eta^{t-1-s}\varphi(x_s)\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}} \\
&\le \frac{1}{\sqrt{\lambda}}\big\|\sum_{s=t-c}^{p}\varphi(x_s)\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}}\Big(\text{Because } t-1-s>0 \text{ and } 0<\gamma<1\Big) \\
&\le \frac{1}{\sqrt{\lambda}}\sum_{s=t-c}^{p}\big\|\varphi(x_s)\big\|_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}} \le \frac{1}{\sqrt{\lambda}}\sum_{s=t-c}^{p}\frac{\acute{\sigma}_{t-1}(x_s)}{\sqrt{\lambda}} \le \frac{1}{\lambda}\sum_{s=t-c}^{p}\acute{\sigma}_{t-1}(x_s).
\end{aligned}
$$

□

C.5 Proof of Theorem 4

By combining Lemma 17 , 19 and 20, we would obtain the dynamic regret bound as below.

$$
\begin{aligned}
R_T \le\ & 2\beta_T\sqrt{4\lambda T\breve{\gamma}_T + 2\lambda m T^2\log(1/\eta)} + \frac{2}{\lambda}c^{3/2}B_T\sqrt{4\lambda\breve{\gamma}_t + 2\lambda mc\log(1/\eta)} \\
&+ \frac{4B\eta^c}{\lambda(1-\eta)}T + \frac{2}{\lambda}B_T\frac{c^2\sqrt{\epsilon_m}}{1-\eta} + \frac{2\beta_T T\sqrt{\epsilon_m}}{1-\eta},
\end{aligned}
$$

where $c \ge 1$ is an integer and $0 < \eta < 1$.

C.6 Proof of Corollary 5

In this section we provide the regret order analysis for WGP-UCB (Algorithm 1).

Lemma 22. *Let $c = \frac{\log T}{1-\eta}$ and $\bar{m} = \log_{4/e}(T^3\dot{\gamma}_T^{3/2})$. If B_T is known, by choosing $\eta = 1 - \dot{\gamma}_T^{-1/4}B_T^{1/2}T^{-1/2}$, the regret bound is $\tilde{O}(\dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4})$.*

Proof. Similarly to Russac et al. (2019), we define $\log(\frac{1}{\eta}) = -\log(\eta) \sim 1-\eta := X$ and $c := \frac{\log T}{1-\eta} = X^{-1}\log T$. By defining $X = \dot{\gamma}_T^{-1/4}B_T^{1/2}T^{-1/2}$ and neglecting the logarithmic factors, we analyse the terms in the following regret bound one by one.

For the first term $2\beta_T\sqrt{T}\sqrt{4\lambda\breve{\gamma}_T + 2\lambda mT\log(1/\eta)}$, we have the following

$$
\begin{aligned}
\beta_T &\sim \dot{\gamma}_T^{1/2} \\
\sqrt{4\lambda\breve{\gamma}_T + 2\lambda mT\log(1/\eta)} &\sim \dot{\gamma}_T^{1/2}T^{1/2}X^{1/2} \\
2\beta_T\sqrt{T}\sqrt{4\lambda\breve{\gamma}_T + 2\lambda mT\log(1/\eta)} &\sim \dot{\gamma}_T T X^{1/2} \sim \dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4}.
\end{aligned}
$$

For the second term $\frac{2}{\lambda}c^{3/2}B_T\sqrt{4\lambda\breve{\gamma}_t + 2\lambda mc\log(1/\eta)}$, we have the following

$$\begin{aligned} c^{3/2} &\sim X^{-3/2} \\ c\log(1/\eta) &\sim 1 \\ \frac{2}{\lambda}c^{3/2}B_T\sqrt{4\lambda\breve{\gamma}_t + 2\lambda mc\log(1/\eta)} &\sim \dot{\gamma}_T^{1/2}B_T X^{-3/2} \sim \dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4}. \end{aligned}$$

For the third term $\frac{4B\eta^c}{\lambda(1-\eta)}T$, we have

$$\begin{aligned} \eta^c = e^{c\log\gamma} = e^{\frac{\log\gamma}{1-\eta}\log T} = e^{-\log T} = T^{-1} \\ \frac{4B\eta^c}{1-\eta}T \sim X^{-1} \sim \dot{\gamma}_T^{1/4}B_T^{-1/2}T^{1/2}. \end{aligned}$$

For the fourth term $\frac{2}{\lambda}B_T\frac{c^2\sqrt{\epsilon_m}}{1-\eta}$, we have

$$\begin{aligned} c^2 &\sim X^{-2} \\ \epsilon_m^{1/2} &\sim T^{-3/2}\dot{\gamma}_T^{-3/4} \\ \frac{1}{1-\eta} &\sim X^{-1} \\ \frac{2}{\lambda}B_T\frac{c^2\sqrt{\epsilon_m}}{1-\eta} &\sim \dot{\gamma}_T^{-3/4}B_T T^{-3/2}X^{-3} \sim B_T^{-1/2} \end{aligned}$$

For the last term $\frac{2\beta_T T\sqrt{\epsilon_m}}{1-\eta}$, we have $\epsilon_m = O((e/4)^{\bar{m}}) = O(T^{-3}\dot{\gamma}_T^{-3/2})$. Then we have

$$\begin{aligned} \beta_T &\sim \dot{\gamma}_T^{1/2} \\ \epsilon_m^{1/2} &\sim T^{-3/2}\dot{\gamma}_T^{-3/4} \\ \frac{1}{1-\eta} &\sim X^{-1} \\ \frac{2\beta_T T\sqrt{\epsilon_m}}{1-\eta} &\sim \dot{\gamma}_T^{-1/4}T^{-1/2}X^{-1} \sim B_T^{-1/2}. \end{aligned}$$

By combining five terms, we complete the regret order analysis,

$$R_T \le 2B_T^{-1/2} + \dot{\gamma}_T^{1/4}B_T^{-1/2}T^{1/2} + \dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4} + \dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4} = \tilde{O}(\dot{\gamma}_T^{7/8}B_T^{1/4}T^{3/4}).$$

□

Lemma 23. *Let $c = \frac{\log T}{1-\eta}$ and $\bar{m} = \log_{4/e}(T^3\dot{\gamma}_T^{3/2})$. If B_T is unknown, by choosing $\eta = 1 - \dot{\gamma}_T^{-1/4}T^{-1/2}$, the regret bound is $\tilde{O}(\dot{\gamma}_T^{7/8}B_T T^{3/4})$.*

Proof. Similar to the proof above, we define $\log(\frac{1}{\eta}) = -\log(\eta) \sim 1-\eta := X$ and $c := \frac{\log T}{1-\eta} = X^{-1}\log T$. By defining $X = \dot{\gamma}_T^{-1/4}T^{-1/2}$ and neglecting the logarithmic factors, we analyse the terms in the following regret bound one by one.

For the first term, we have the following

$$2\beta_T\sqrt{T}\sqrt{4\lambda\breve{\gamma}_T + 2\lambda mT\log(1/\eta)} \sim \dot{\gamma}_T T X^{1/2} \sim \dot{\gamma}_T^{7/8}T^{3/4}$$

For the second term, we have the following

$$\frac{2}{\lambda}c^{3/2}B_T\sqrt{4\lambda\breve{\gamma}_T + 2\lambda mc\log(1/\eta)} \sim \dot{\gamma}_T^{1/2}B_T X^{-3/2} \sim \dot{\gamma}_T^{7/8}B_T T^{3/4}$$

For the third term, we have

$$\frac{4B\eta^c}{1-\eta}T \sim X^{-1} \sim \dot{\gamma}_T^{1/4}T^{1/2}$$

For the fourth term, we have

$$\frac{2}{\lambda}B_T\frac{c^2\sqrt{\epsilon_m}}{1-\eta} \sim \dot{\gamma}_T^{-3/4}B_T T^{-3/2}X^{-3} \sim B_T$$

For the last term, we have $\epsilon_m = O((e/4)^{\bar{m}}) = O(T^{-3}\dot{\gamma}_T^{-3/2})$. Then we have

$$\frac{2\beta_T T\sqrt{\epsilon_m}}{1-\eta} \sim \dot{\gamma}_T^{-1/4}T^{-1/2}X^{-1} \sim 1.$$

By combining five terms, we complete the regret order analysis,

$$R_T \leq 1 + B_T + \dot{\gamma}_T^{1/4}T^{1/2} + \dot{\gamma}_T^{7/8}B_T T^{3/4} + \dot{\gamma}_T^{7/8}T^{3/4} = \tilde{O}(\dot{\gamma}_T^{7/8}B_T T^{3/4}).$$

□

D Proof of Weighted Information Gain

In this section we provide two types of upper bounds of maximum information gain.

D.1 Universal Bound

In this section we present the proof of Theorem 6.

Proof of Theorem 6. The proof is composed of two following lemmas.

Lemma 24. $\bar{\gamma}_T \leq \frac{N}{2}\log\left(1 + \frac{\dot{k}T}{\lambda N}\right) + \frac{T}{2\lambda}\delta_N$

Proof. In the similar way as (Vakili et al., 2021, Theorem 3), we define T-by-T matrix $\bar{K}_P = [\bar{k}_P(x_i, x_j)]_{i,j=1}^T$ and $\bar{K}_O = [\bar{k}_O(x_i, x_j)]_{i,j=1}^T$. Then we have $\bar{K}_t = \bar{K}_P + \bar{K}_O$.

The mutual information is decomposed into two terms.

$$\begin{aligned}\bar{I}(y_t; f_t) &= \frac{1}{2}\log\det(I_t + \alpha_t^{-1}\bar{K}_t) \\ &= \frac{1}{2}\log\det(I_t + \alpha_t^{-1}\bar{K}_P) + \frac{1}{2}\log\det(I_t + \alpha_t^{-1}(I_t + \alpha_t^{-1}\bar{K}_p)^{-1}\bar{K}_O).\end{aligned}$$

To get the tighter bound, we specify $\alpha_t = \lambda w_t^2$. The first term is bounded as, where we define $\bar{K}_P = \bar{\Psi}_N\bar{C}_N\bar{\Psi}_N^T$ and $\bar{G}_t = \bar{C}_N^{1/2}\bar{\Psi}_N^T\bar{\Psi}_N\bar{C}_N^{1/2}$.

$$\begin{aligned}&\frac{1}{2}\log\det(I_t + \alpha_t^{-1}\bar{K}_P) \leq \frac{1}{2}N\log\left(\frac{1}{N}tr(I_N + \alpha_t^{-1}\bar{G}_t)\right) \\ &\leq \frac{1}{2}N\log\left(1 + \frac{1}{N}\alpha_t^{-1}\sum_{s=1}^t \bar{k}_p(x_s, x_s)\right) \leq \frac{1}{2}N\log\left(1 + \frac{1}{N}\frac{1}{\lambda w_t^2}\sum_{s=1}^t w_s^2 k_p(x, x)\right) \\ &\leq \frac{1}{2}N\log\left(1 + \frac{\dot{k}}{N}\frac{1}{\lambda}\sum_{s=1}^t \frac{w_s^2}{w_t^2}\right) \leq \frac{1}{2}N\log\left(1 + \frac{\dot{k}}{N}\frac{1}{\lambda}\sum_{s=1}^t 1\right) \leq \frac{N}{2}\log\left(1 + \frac{\dot{k}t}{\lambda N}\right).\end{aligned}$$

For the second term, as the largest eigenvalue of $(I_t + \alpha_t^{-1}\bar{K}_P)^{-1}$ is upper bounded by 1, we have $tr((I_t + \alpha_t^{-1}\bar{K}_P)^{-1}\bar{K}_O) \leq tr(\bar{K}_O)$. Then, we have

$$\frac{1}{2}\log\det(I_t+\alpha_t^{-1}(I_t+\alpha_t^{-1}\bar{K}_P)^{-1}\bar{K}_O) \le \frac{t}{2}\log\Big(\frac{1}{t}tr(I_t+\alpha_t^{-1}(I_t+\alpha_t^{-1}\bar{K}_P)^{-1}\bar{K}_O)\Big)$$
$$\le \frac{t}{2}\log\Big(\frac{1}{t}(t+\alpha_t^{-1}tr(\bar{K}_O))\Big) \le \frac{t}{2}\log\Big(\frac{1}{t}(t+\alpha_t^{-1}\sum_{s=1}^{t}\bar{k}_O(x_s,x_s))\Big)$$
$$\le \frac{t}{2}\log\Big(\frac{1}{t}(t+\frac{1}{\lambda w_t^2}\sum_{s=1}^{t}w_s^2 k_O(x_s,x_s))\Big) \le \frac{t}{2}\log\Big(\frac{1}{t}(t+\frac{1}{\lambda}\sum_{s=1}^{t}\frac{w_s^2}{w_t^2}\delta_N)\Big)$$
$$\le \frac{t}{2}\log\Big(\frac{1}{t}(t+\frac{1}{\lambda}\sum_{s=1}^{t}\delta_N)\Big) \le \frac{t}{2}\log\Big(\frac{1}{t}(t+\frac{t}{\lambda}\delta_N)\Big) \le \frac{t}{2}\log\Big(1+\frac{1}{\lambda}\delta_N\Big) \le \frac{t}{2\lambda}\delta_N.$$

Combining two terms, we provide the upper bound of maximal information gain for double weighted kernel matrix as,

$$\bar{\gamma}_T \le \frac{N}{2}\log\Big(1+\frac{\dot{k}T}{\lambda N}\Big)+\frac{T}{2\lambda}\delta_N.$$

□

Lemma 25. $\breve{\gamma}_T \le \frac{N}{2}\log\Big(1+\frac{\dot{k}T}{\lambda N}\Big)+\frac{T}{2\lambda}\delta_N$.

Proof. In the similar way of previous lemma, by replacing α_t with λ_t and $\bar{K}$ with $\tilde{K}$, we also provide the same upper bound of maximal information gain for weighted kernel matrix as,

$$\tilde{\gamma}_T \le \frac{N}{2}\log\Big(1+\frac{\dot{k}T}{\lambda N}\Big)+\frac{T}{2\lambda}\delta_N$$

where $\sum_{s=1}^{t}\frac{w_s}{w_t} \le \sum_{s=1}^{t}1 = t$ is used. The result follows as $\breve{\gamma}_t = \frac{1}{2}\log\det(I+\lambda_t^{-1}\breve{\Phi}_t\breve{\Phi}_t^T) \le \tilde{\gamma}_t = \max_{A\subset D:|A|=t}\frac{1}{2}\log\det(I+\lambda_t^{-1}WK_AW^T))$ since $\breve{\Phi}_t = W[\breve{\varphi}(x_1),\dots,\breve{\varphi}(x_t)]^T$ and $K_t=\Phi_t\Phi_t^T$. □

□

D.2 Weight dependent bound

We present the proof of Theorem 7. To get the tighter bound, we specify the weight $w_t=\eta^{-t}$ and thus $\alpha_t=\lambda\eta^{-2t}$.

Proof of Theorem 7. The proof is similar to proof of Lemma 24.

By replacing w_t by η^{-t} , we have

$$\frac{1}{2}\log\det(I_t+\alpha_t^{-1}\bar{K}_P) \le \frac{1}{2}N\log\Big(1+\frac{\dot{k}}{N}\frac{1}{\lambda}\sum_{s=1}^{t}\eta^{2t-2s}\Big)$$
$$\le \frac{N}{2}\log\Big(1+\frac{\dot{k}(1-\eta^{2t})}{\lambda N(1-\eta^2)}\Big) \le \frac{N}{2}\log\Big(1+\frac{\dot{k}}{\lambda N(1-\eta^2)}\Big).$$

For the second term,

$$\frac{1}{2}\log\det(I_t+\alpha_t^{-1}(I_t+\alpha_t^{-1}\bar{K}_P)^{-1}\bar{K}_O) \le \frac{t}{2}\log\Big(\frac{1}{t}(t+\frac{1}{\lambda}\sum_{s=1}^{t}\eta^{2t-2s}\delta_N)\Big)$$
$$\le \frac{t}{2}\log\Big(\frac{1}{t}(t+\frac{1-\eta^{2t}}{\lambda(1-\eta^2)}\delta_N)\Big) \le \frac{t}{2}\log\Big(1+\frac{1-\eta^{2t}}{t\lambda(1-\eta^2)}\delta_N\Big) \le \frac{1}{2\lambda(1-\eta^2)}\delta_N.$$

Combining two terms, we provide the upper bound of maximal information gain for double weighted kernel matrix as,

$$\bar{\gamma}_T \le \frac{N}{2}\log\Big(1+\frac{\dot{k}}{\lambda N(1-\eta^2)}\Big)+\frac{1}{2\lambda(1-\eta^2)}\delta_N.$$

In the similar way, we also provide the upper bound of maximal information gain for single weighted kernel matrix as,

$$\tilde{\gamma}_T \leq \frac{N}{2}\log\Big(1+\frac{\dot{k}}{\lambda N(1-\eta)}\Big)+\frac{1}{2\lambda(1-\eta)}\delta_N.$$

The result follows $\breve{\gamma}_T \leq \tilde{\gamma}_T$. □

We also present the proof of Corollary 8.

Proof of Corollary 8. Under the (C_p, β_p) polynomial eigendecay condition, we obtain the following bound on δ_N as

$$\delta_N = \sum_{m=N+1}^{\infty} \lambda_m \phi^2 \leq C_p N^{1-\beta_p}\phi^2.$$

By choosing $N = \lceil \big(\frac{C_p\phi^2}{\lambda(1-\eta^2)}\big)^{\frac{1}{\beta_p}} \log^{-\frac{1}{\beta_p}}(1+\frac{\dot{k}}{\lambda(1-\eta^2)})\rceil$,

$$\bar{\gamma}_T \leq \Big(\big(\frac{C_p\phi^2}{\lambda(1-\eta^2)}\big)^{\frac{1}{\beta_p}} \log^{-\frac{1}{\beta_p}}(1+\frac{\dot{k}}{\lambda(1-\eta^2)})+1\Big)\log(1+\frac{\dot{k}}{\lambda(1-\eta^2)}).$$

Under the $(C_{e,1}, C_{e,2}, \beta_e)$ exponential eigendecay condition, we obtain the following bound on δ_N as

$$\delta_N \leq \int_{z=N}^{\infty} C_{e,1}\exp(-C_{e,2}z^{\beta_e})\phi^2 dz.$$

Now, consider the case of $\beta_e = 1$ (skip the case of $\beta_e \neq 1$ for simplicity). Then,

$$\int_{z=N}^{\infty} \exp(-C_{e,2}z^{\beta_e})\phi^2 dz = \frac{1}{C_{e,2}}\exp(-C_{e,2}N).$$

With the similar logic, we choose $N = \lceil \frac{1}{C_{e,2}}\log\big(\frac{C_{e,1}\phi^2}{C_{e,2}\lambda(1-\eta^2)}\big)\rceil$, then we obtain the following bound,

$$\bar{\gamma}_T \leq \Big(\frac{1}{C_{e,2}}\big(\log(\frac{1}{1-\eta^2})+C_{\beta_e}\big)+1\Big)\log(1+\frac{\dot{k}}{\lambda(1-\eta^2)}).$$

□