Time-varying Gaussian Process Bandit Optimization with Experts: no-regret in logarithmically-many side queries

Eliabelle Mauduit[1] ✉, Eloïse Berthier[2], and Andrea Simonetto[1]

[1] Unité des Mathématiques Appliquées, ENSTA, Institut Polytechnique de Paris, 91120 Palaiseau, France {eliabelle.mauduit,andrea.simonetto}@ensta.fr
[2] U2IS, ENSTA, Institut Polytechnique de Paris, 91120 Palaiseau, France
eloise.berthier@ensta.fr

Abstract. We study a time-varying Bayesian optimization problem with bandit feedback, where the reward function belongs to a Reproducing Kernel Hilbert Space (RKHS). We approach the problem via an upper-confidence bound Gaussian Process algorithm, which has been proven to yield no-regret in the stationary case.
The time-varying case is more challenging and no-regret results are out of reach in general in the standard setting. As such, we instead tackle the question of how many additional observations asked to an expert are required to regain a no-regret property. To do so, we formulate the presence of past observation via an uncertainty injection procedure, and we reframe the problem as a heteroscedastic Gaussian Process regression. In addition, to achieve a no-regret result, we discard long outdated observations and replace them with updated (possibly very noisy) ones obtained by asking queries to an external expert. By leveraging and extending sparse inference to the heteroscedastic case, we are able to secure a no-regret result in a challenging time-varying setting with only logarithmically-many side queries per time step. Our method demonstrates that minimal additional information suffices to counteract temporal drift, ensuring efficient optimization despite time variation.

Keywords: Gaussian Processes · Upper confidence bounds · Bandit feedback · Sparse inference · Time-varying optimization.

1 Introduction

We consider the problem of sequentially optimizing a reward function $f : \mathcal{D} \times \mathbb{R}_+ \to \mathbb{R}$ where $\mathcal{D} \subset \mathbb{R}^d$ is a compact convex set. In this configuration, the objective depends both on time and on a continuous decision space $\mathcal{D}$. At each discrete time step t, we obtain a noisy observation of the reward $y_t = f(x_t, t) + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. Our objective is to maximize the sum of rewards

$$\max_{(x_t)_t \in \mathcal{D}^T} \quad \sum_{t=1}^{T} \Big[f(x_t, t) =: f_t(x) \Big]. \tag{1}$$

At least in the static case, when f does not change in time, this type of problem has often been formulated via Bayesian optimization with bandit feedback [17], whereby an agent must take a sequence of actions while observing the corresponding sequence of rewards. Each action consists of picking a decision x to get an estimate of the reward at the corresponding point. The agent does not modify its environment through its actions and can thus *exploit* previous measurements to predict actions that offer the highest rewards, but should also *explore* new decisions where the value of the reward function is possibly high. For dynamic rewards, the setting is more challenging, as we will see.

In the time-varying case, the performance metric we are interested in is the dynamic cumulative regret, defined as

$$R_T = \sum_{t=1}^{T} \left(\max_{x \in \mathcal{D}} \; f_t(x) - f_t(x_t) \right) , \tag{2}$$

representing the cumulative loss in reward picking decision x_t at time t with respect to the best decision *at the same time step*. Algorithms that achieve an asymptotically vanishing average dynamic cumulative regret, as $\lim_{T\to\infty} R_T/T = 0$, are said to enjoy no-regret.

To derive our main theoretical results, we will work under two reasonable blanket assumptions. First, to model smoothness properties of the functions f_t, we assume that they all belong to a Reproducing Kernel Hilbert Space (RKHS) and have bounded RKHS norm. The RKHS associated with kernel k $(\mathcal{H}_k(\mathcal{D}), \langle .,. \rangle_k)$ is a subspace of $L_2(\mathcal{D})$ [14] and the associated inner product $\langle .,. \rangle_k$ is such that

$$\forall f \in \mathcal{H}_k(\mathcal{D}),\; f(x) = \langle f, k(x,.) \rangle_k \, .$$

The norm $\|f\|_k$ measures the smoothness of f with respect to the kernel function k, therefore assuming $\|f_t\|_k$ is bounded translates into regularity assumptions about the objective.

Assumption 1 *For all time steps, functions $x \mapsto f_t(x)$ belong to a Reproducing Kernel Hilbert Space with continuous bounded kernel k such that $\forall x \in \mathcal{D}$, $k(x,x) \leq M_k^2$ and they have bounded RKHS norm,*

$$\forall t,\; \|f_t\|_k \leq B. \tag{3}$$

Second, to model time variation, we assume boundedness of the variations as follows.

Assumption 2 *For the sequence of functions $(f_t)_{t=1}^T$, there exists a bounded constant Δ, such that,*

$$\forall t,\; \sup_{x \in \mathcal{D}} |f_{t+1}(x) - f_t(x)| \leq \Delta. \tag{4}$$

We further let $\Delta = 1$, without any loss of generality.

Assumption 2 ensures controlled temporal variations, limiting changes between consecutive iterations and provides a sound framework for uncertainty injection.

1.1 Related work

Bayesian optimization in the bandit feedback setting has been studied extensively in the static scenario: the landmark work of Srinivas and coauthors [17] proposes an upper-confidence bound algorithm based on a Gaussian Process model of the unknown function obtaining no-regret in several settings. In particular, first, they use the noisy observations of f to derive a possibly miss-specified estimation of its mean $\mu_t(x)$ and covariance $\sigma_t^2(x)$, via a Gaussian Process:

$$\mu_t(x) = k_t(x)^\top (K_t + \Sigma_t)^{-1} Y_t \tag{5}$$

$$\sigma_t^2(x) = k(x,x) - k_t(x)^\top (K_t + \Sigma_t)^{-1} k_t(x), \tag{6}$$

where $\Sigma_t := \sigma^2 I_t$, σ^2 being the noise variance of each observation y_i, $(K_t)_{i,j} = k(x_i, x_j)$, $k_t(x) = [k(x_1, x), \ldots, k(x_t, x)]$, k being the kernel or covariance function, and $Y_t = [y_1, \ldots, y_t]^\top$. Then, since the reward f is unknown, they propose choosing the next decision based on the upper-confidence bound proxy, as,

$$x_{t+1} = \arg\max_{x \in \mathcal{D}} \mu_t(x) + \beta_{t+1}\sigma_t(x), \tag{7}$$

where $(\beta_t)_{t\geq 1}$ is a sequence of positive parameters chosen to ensure a trade-off between exploration and exploitation and it is decisive in proving the convergence of the algorithm. Their algorithm, labeled GP-UCB, obtains no-regret in high probability when f is sampled from a GP, i.e., $f \sim \mathrm{GP}(0, k(x, x'))$ but also for arbitrary f with bounded RKHS norm. As a means of comparison for the square exponential kernel and f having a bounded RHKS norm, they obtain a $R_T = \tilde{\mathcal{O}}(\sqrt{T})$ result. Here, the notation $\tilde{\mathcal{O}}(\cdot)$ hides poly-logarithmic terms.

The cited work focused on a noise model whose distribution is identical across observations, also known as homoscedastic setting. Makarova and coauthors in [12] remove this assumption, define $\Sigma_t := \mathrm{diag}(\sigma_1^2, \ldots, \sigma_t^2)$ for noise model $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$, and deliver a regret bound that matches $R_T = \tilde{\mathcal{O}}(\sqrt{T})$ up to a multiplicative $\bar{\sigma} := \max\{\sigma_i\}$ factor, for the heteroscedastic setting.

The time-varying case has also received attention. The work of [2] extends the GP-UCB algorithm by considering a time-varying reward. They model the time variations by considering a spatio-temporal kernel with a forgetting factor ε, as

$$\forall\ t_i, t_j \leq t,\ \ k((x_{t_i}, t_i), (x_{t_j}, t_j)) = (1-\varepsilon)^{|t_i - t_j|/2} k(x_i, x_j), \tag{8}$$

where $k(\cdot,\cdot)$ is the static kernel. With this modeling, they propose two algorithms: R-GP-UCB runs GP-UCB on windows of size $w \in \mathbb{N}$ and resets at the start of each window. The second one, TV-GP-UCB, uses the spatiotemporal kernel (8). Under this setting, the authors showed that any GP bandit optimization incurs expected regret of at least $\mathbb{E}[R_T] = \Omega(T\varepsilon)$, meaning the algorithm does not enjoy no-regret for fixed ε. This lower bound is not surprising and it also appears in the multi-armed bandit literature [1]. Furthermore, TV-GP-UCB obtains a $R_T = \tilde{\mathcal{O}}(T)$ which implies an increasing average cumulative dynamic regret.

Building on the literature in dynamic (generalized) linear bandits [22, 23, 13, 21], in a series of papers [24, 5], new algorithms are proposed in the time-varying setting: a revised R-GP-UCB algorithm, a new sliding-window algorithm SW-GP-UCB, and a weighted algorithm W-GP-UCB. Under the RKHS setting, they either enjoy cumulative dynamic regrets of $\mathcal{O}(T)$ (matching the lower bound), or $\tilde{\mathcal{O}}(T)$ for the latter two (with our variation budget expressed in Assumption 2). The weighted algorithm is interesting, since it starts from a weighted kernel regression,

$$\hat{f} = \arg \min_{f \in \mathcal{H}_k(\mathcal{D})} \sum_{t=1}^{T} w_t (y_t - f(x_t))^2 + \lambda_t \|f\|_k^2, \tag{9}$$

where $\mathcal{H}_k(\mathcal{D})$ is the RKHS on set $\mathcal{D}$ and kernel k, w_t is a weight, and $\lambda_t \geq 0$ a parameter; they arrive then at the same iterations of Makarova and coauthors in [12] for the heteroscedastic setting, but with a growing-in-T noise variance.

Since no-regret is out of reach in the standard setting, the authors of [7] proposed an algorithm capable of dynamically capturing the changes of the objective function, and thereby acquiring more observations when needed. While this does not guarantee no-regret for a constant sampling time, they show an interesting trade-off between sampling and regret.

Dealing with a spatio-temporal kernel like (8) is theoretically challenging. The works of [20, 3] propose instead to inject uncertainty into old observations. Their starting point is to consider, at every time t, that the variance of old observations increases in time (either exponentially or linearly). This is easier to handle since it is now Σ_t that changes, but only on the diagonal. Regret results are not provided, but it is not difficult to see that this approach is equivalent to the weighted kernel regression in the RKHS settings and delivers the same $\tilde{\mathcal{O}}(T)$ regret.

In addition to regret analysis, another active field of research in GP regression involves optimization of algorithms complexity. Regression based on GP models becomes impracticable for large datasets as its time complexity scales as $\mathcal{O}\left(N^3\right)$, where N is the number of observations [16]. The idea of sparse Gaussian Process regression is to approximate the posterior by performing GP regression on a subset of $M \ll N$ inputs. In this way, the complexity becomes $\mathcal{O}\left(NM^2\right)$. The difficulty lies in the selection of the set of sparse inputs (also called pseudo or inducing inputs) and several techniques exist. For example, in [18], Titsias considers sparse inputs as variational parameters selected to minimize the Kullback-Leibler (KL) divergence between the exact and approximate posteriors. Leveraging this work, Burt et al. show in [4] that $M = \mathcal{O}(d \log^d(N))$ sparse inputs suffice to accurately approximate the posterior in terms of KL divergence. They make use of an approximation of a M-Determinantal Point Process (M-DPP) [11] to build the set of sparse inputs. M-DPPs define a probability distribution over input subsets of size M that favors the selection of dispersed and less correlated points.

Finally, the algorithms developed in the literature show affinity with online learning in the dynamic setting, e.g., [8].

1.2 Contributions

In this paper, we extend the literature in several ways.

- First, motivated by the fact that handling time-variations with spatio-temporal kernels is technically challenging, we embrace the uncertainty injection framework and we formulate the time-varying problem as a sequence of static regression problems, with growing-in-time uncertainty. This renders the GP problem a heteroscedastic one.
- Then, since a no-regret result is out of reach in this setting, we ask *how many additional queries one should pose to an expert* in order to regain the no-regret result that we enjoy in static settings. The answer to the queries are noisy evaluations (or predictions) of the function at a given time. To limit the number of queries, we leverage sparse inference and we estimate the error of updating past observations with the least number of observations as possible. We call our new GP-UCB algorithm SparQ-GP-UCB for sparse queries. The algorithm performs GP-UCB updates at every time step t by discarding past measurements taken at times τ farther away than $\mathcal{O}(\log(t))$ steps and asks new observations to an expert.
- We prove that SparQ-GP-UCB achieves a $R_T = \tilde{\mathcal{O}}(\sqrt{T})$ in $\tilde{\mathcal{O}}(1)$ additional queries per time step, and it exhibits a $\tilde{\mathcal{O}}(T^2)$ computational complexity. This makes SparQ-GP-UCB the first true no-regret time-varying Gaussian Process algorithm, at the expense of logarithmically-many side queries at each step.

2 Problem setting

2.1 Uncertainty injection

We recall our setting. We consider the problem of sequentially optimizing a reward function $f : \mathcal{D} \times \mathbb{R}_+ \to \mathbb{R}$ where $\mathcal{D} \subset \mathbb{R}^d$ is a compact convex set. In this configuration, the objective depends both on time and on a continuous decision space $\mathcal{D}$. At each step t, we obtain a noisy observation of the reward $y_t = f(x_t, t) + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We set $f_t(\cdot) := f(\cdot, t)$ for convenience.

Our approach of the problem is to inject uncertainty into old measurements and to consider each optimization problem depending on functions f_t as a sequence of separate snapshots.

Under Assumption 2 on the boundedness of function variations, at time T, we can consider that past observations are noisy observations of the current function f_T with zero-mean noise and variance that is increased depending on how old the observations are. In particular, we use independent noise random variables $\epsilon_{t,T}$ and model

$$y_t = f_T(x_t) + \epsilon_{t,T}, \qquad \epsilon_{t,T} \sim \mathcal{N}(0, \sigma^2([T-t]^2 + 1)), \quad t \leq T, \tag{10}$$

that is the noise standard deviation increases linearly in time. This is similar to the approach of [3] involving a Wiener process and it is well-motivated by the fact that the maximum variation of the function between t and T is $T-t$. In fact, if we model $f_{t+1}(x) = f_t(x) + v$, with $v \in \mathcal{U}(-1,1)$, i.e., the uniform distribution on $[-1,1]$, then the observation y_t of function f_t can be interpreted as an observation of function f_T with noise $\mathbb{E}[\epsilon_{t,T}] = \mathbb{E}[(T-t)v + \epsilon_t] = 0$ and variance $\mathbb{E}[\|\epsilon_{t,T}\|^2] = \sigma^2(\frac{1}{3\sigma^2}[T-t]^2 + 1)$. The latter justifies the expression of the noise, up to asymptotically-unimportant constants.

At time t, then, we would like to maximize the reward of $f_t(x)$ by choosing the next action based on past observation $Y_t = [y_1, \ldots, y_t]^\top$, each with its own zero-mean noise and time-dependent variance. We approach this as a heteroscedastic Gaussian Process and perform the update,

$$\mu_t(x) = k_t(x)^\top (K_t + \Sigma_t)^{-1} Y_t \tag{11}$$

$$\sigma_t^2(x) = k(x,x) - k_t(x)^\top (K_t + \Sigma_t)^{-1} k_t(x), \tag{12}$$

where $\Sigma_t := \mathrm{diag}(\mathrm{Var}(\epsilon_{1,t}), \ldots, \mathrm{Var}(\epsilon_{t,t}) = \sigma^2)$, the kernel matrix $(K_t)_{i,j} = k(x_i, x_j)$, and $k_t(x) = [k(x_1,x), \ldots, k(x_t,x)]$. We choose the next decision as,

$$x_{t+1} = \arg\max_{x \in \mathcal{D}} \mu_t(x) + \beta_{t+1}\sigma_t(x), \tag{13}$$

where $(\beta_t)_{t \geq 1}$ is a sequence of parameters chosen to ensure a trade-off between exploration and exploitation.

As said, a basic version of this update would lead an increasing average regret. To limit the regret, we consider only recent observations and summarize and update the remaining ones.

2.2 Sparse inference

To summarize and update past observations, we leverage and extend recent results from sparse inference provided in [4]. Consider Y_T observations performed at $X_T = [x_1, \ldots, x_T]$ points, as well as the mean and variance function coming from a GP regression on these points. Burt and coauthors in [4] offer an algorithm to select $\tilde{\mathcal{O}}(1)$ points in the domain $\mathcal{D}$ which would deliver the same mean and variance up to a tunable multiplicative error term. We summarize their main result in the following proposition.

Proposition 1 ([4]). *Consider the problem of estimating an unknown function $f : \mathcal{D} \to \mathbb{R}$ via T noisy observations, $y_t = f(x_t) + \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ acquired at i.i.d. training inputs X_T. Let f be a sample path of a Gaussian Process with zero mean and kernel k. Consider a squared exponential kernel function k for simplicity. Let $\mu_0(x)$ and $\sigma_0^2(x)$ represent the mean and variance of the Gaussian Process regression performed on the observations.*

Select a tolerance level $\eta \leq 1/5$. Then, there exists an algorithm that selects $\tilde{\mathcal{O}}(1) < T$ points in the domain $\mathcal{D}$ and their observations $y_t = f(x_t) + \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, such that if we let $\mu_1(x)$ and $\sigma_1^2(x)$ represent the mean and variance of the Gaussian Process regression performed on the new points and observations, *we obtain in high probability,*

$$
\begin{aligned}
|\mu_1 - \mu_0| \leq \sigma_0\sqrt{\eta} \leq \frac{\sigma_1\sqrt{\eta}}{\sqrt{1-\sqrt{3\eta}}}, \\
|1 - \sigma_1^2/\sigma_0^2| \leq \sqrt{3\eta}.
\end{aligned}
\tag{14}
$$

Proposition 1 is a condensed version of Proposition 1, Theorem 14, and Corollary 22 in [4].

A possible algorithm proposed in the paper to determine the sparse inputs is an approximate determinantal point process (DPP). Such algorithm selects $M < T$ points in order to minimize the difference between the KL divergence of the exact posterior and approximated one. Specifically, for $\epsilon > 0$, one can use an MCMC algorithm, as specified in Algorithm 1 of [4] to obtain an ϵ approximation of a M-DPP, with T inputs, with a computational complexity that is upper bounded by $\mathcal{O}\left(TM^3(\log\log T + \log M + \log 1/\epsilon^2)\right)$.

The most important feature of the DPP algorithm in [4] and Proposition 1 is that these results do not depend on the observations values Y_T, but only on the points X_T where these observations are taken. We will see next how this is key in devising our sparse algorithm and, along the way, how we can extend Proposition 1 to the heteroscedastic and deterministic setting.

3 SparQ-GP-UCB Algorithm

With all the previous preliminaries in place, we are now ready for the main algorithm: SparQ-GP-UCB.

The algorithm works in rounds. At each time step t, we consider the problem of maximizing the regret f_t, with observations $Y_t = [y_1, \ldots, y_t]^\top$ at points $X_t = [x_1, \ldots, x_t]$. The observations are properly injected with uncertainty, so that their variance grows in time as,

$$\epsilon_{i,t} \sim \mathcal{N}(0, \sigma^2([t-i]^2 + 1)), \quad i \leq t. \tag{15}$$

The **first** step of the algorithm is to discard observations that have variance greater than $g(t)$, where $g(t) = o\left(t^{1/4}\right)$. We take $g : t \mapsto \sigma^2 \log(t)$ as an illustrative example but any function $g : t \mapsto g(t) = o\left(t^{1/4}\right)$ would work with no change in the proof arguments (and we further discuss it in the proof).

Second, we act as if we had access to updated noisy observations for the discarded measurements, with noise being zero-mean and with $\bar{\sigma}^2$ variance. With this pretend observations and the most recent ones with noise less than $\sigma^2 \log(t)$, we perform sparse variational inference. We use the approximate DPP algorithm in [4] (Algorithm 1) to find the locations $X^{\mathrm{E}} = [x_1^{\mathrm{E}}, \ldots]$, with $|X^{\mathrm{E}}| = \tilde{\mathcal{O}}(1) \ll t$ at which to ask an expert for noisy updated observation with zero-mean and variance $\bar{\sigma}^2$. The new expert-delivered observations, together with the most recent ones are guaranteed to be a good approximation of the pretend setting.

Third, we let $Y_t^{\mathrm{s}}, X_t^{\mathrm{s}}$ being the set of expert-delivered observations together with the most recent ones with noise less than $\sigma^2 \log(t)$ and the points at which they are taken. With this, we can compute the mean and variance as,

$$\mu_t^{\mathrm{s}}(x) = k_t^{\mathrm{s}}(x)^\top (K_t^{\mathrm{s}} + \Sigma_t^{\mathrm{s}})^{-1} Y_t^{\mathrm{s}} \tag{16}$$

$$(\sigma^{\mathrm{s}})_t^2(x) = k(x,x) - k_t^{\mathrm{s}}(x)^\top (K_t^{\mathrm{s}} + \Sigma_t^{\mathrm{s}})^{-1} k_t^{\mathrm{s}}(x), \tag{17}$$

where Σ_t^{s} is a diagonal matrix containing all the observation variances up to $\max\{\bar{\sigma}^2, \sigma^2 \log(t)\}$, and the kernel elements K_t^{s}, k_t^{s}, are evaluated on X_t^{s}.

And finally, we compute the next decision, via the UCB proxy:

$$x_{t+1} = \arg\max_{x \in \mathcal{D}} \mu_t^{\mathrm{s}}(x) + \beta_{t+1} \sigma_t^{\mathrm{s}}(x). \tag{18}$$

The algorithm is summarized in Algorithm 1. We remark the need for performing sparse inference based on Algorithm 1 of [4], whose details are reported in the Appendix.

Algorithm 1 SparQ-GP-UCB

Input: Domain $\mathcal{D}$, kernel k
1: **for** $t = 1, 2, \dots$ **do**
2: Sample $y_t = f_t(x_t) + \epsilon_t$
3: Discard all the observations with a noise $> \sigma^2 \log(t)$
4: Perform sparse inference on X_t to obtain locations X^{E} of cardinality $\tilde{\mathcal{O}}(1)$
5: Query an expert to obtain updated observations on X^{E} for f_t
6: Perform Bayesian updates (16)-(17) to obtain μ_t^{s} and σ^{s} using $(X_t^{\mathrm{s}}, Y_t^{\mathrm{s}})$
7: Choose the next action x_{t+1} via (18)
8: **end for**

3.1 Main results

In this subsection, we report the main results for our algorithm. They are both given for a squared exponential kernel for simplicity, but they can easily be extended to other standard kernels (Matérn for example).

Theorem 1. *(Regret bound for SparQ-GP-UCB) Take any $0 < \delta \leq 1$ and consider a sequence of reward functions $(f_t)_t$ and the observations $y_t = f_t(x_t) + \epsilon_t$, for $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d.. Let Assumptions 1 and 2 hold and consider a squared exponential kernel k. Let T be a time horizon and $(x_t)_{t=1}^T$ the set of actions chosen by SparQ-GP-UCB (Algorithm 1) and set $(\beta_t)_{t=1}^T$ as*

$$\beta_t = \sqrt{2 \log \left(\frac{2|\Sigma_t^{\mathrm{s}} + K_{tt}^{\mathrm{s}}|^{1/2}}{\delta |\Sigma_t^{\mathrm{s}}|^{1/2}} \right)} + \|f_t\|_k.$$

Then, with probability at least $1-\delta$, by asking to an expert $\mathcal{O}\left(\log^d(t)\right)$ queries per time step, SparQ-GP-UCB attains a cumulative dynamic regret of

$$R_T = \mathcal{O}\left(\sqrt{Td\log^{d+5}(T)}\sqrt{\log\left(\frac{1}{\delta}\right) + d\log^{d+3}\left(d\log\left(T\right)\right)}\right) = \tilde{\mathcal{O}}(\sqrt{T}). \quad (19)$$

The asymptotic bound given in Eq.(19) implies no-regret with probability $1-\delta$ and matches the static case up to poly-logarithmic factors. The main steps of the proof are given in Section 5.

We can now discuss briefly the impact of Assumption 2 for the regret bound. A common metric in the literature to account for the time-varying nature of the objective is the variation budget V_T [8, 5, 1] defined as,

$$\forall T,\, V_T = \sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_k. \quad (20)$$

Let $x \in \mathcal{D}$. Then, by the reproducing property of RKHS,

$$|f_{t+1}(x) - f_t(x)| = |\langle f_{t+1} - f_t, k(x,.)\rangle| \leq \|f_{t+1} - f_t\|_k \|k(x,.)\|_k, \quad (21)$$

where we applied Cauchy Schwarz in the RKHS to obtain the inequality. From the reproducing property, $\|k(x,.)\|_k^2 = k(x,x)$. As we are working with bounded kernels (Assumption 1) and $\forall x \in \mathcal{D}$, $k(x,x) \leq M_k^2$, we take the infinite norm in the left side of Eq. (21) and sum over $t = 1$ to $T-1$ to obtain:

$$\sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_\infty \leq M_k V_T\,. \quad (22)$$

By Assumption 2, SparQ-GP-UCB does work even in the case of $\sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_\infty = (T-1)$, meaning that our algorithm can achieve no-regret even for a variation budget that grows linearly in time. This improves the result of [5] that requires $V_T = o(T)$ when V_T is known and $V_T = o(T^{1/4})$ otherwise to obtain sublinear regret.

Along with a no-regret result, we also provide a computational complexity estimate as follows.

Theorem 2. *Under the same setting of Theorem 1, the computational complexity of SparQ-GP-UCB is upper bounded by $\mathcal{O}\left(T^2\log(T)\log^{3d}\left(\frac{T}{\log(T)}\right)\right) = \tilde{\mathcal{O}}(T^2)$.*

The theorem shows how SparQ-GP-UCB is actually less computationally expensive than running a basic Bayesian update on the whole T measurement set, which can be bounded as $O(T^3)$.

3.2 Role of the expert

In SparQ-GP-UCB, the "expert" mechanism is not meant to be a human oracle, nor does it need to act as a perfectly accurate surrogate model. Instead, it serves as a means to partially refresh or correct stale information from previous observations in a principled and computationally bounded way.

More precisely, at each time step t, we are allowed to query the current value of the objective f_t at a small number $\mathcal{O}\left(\log^d(t)\right)$ of previously observed points, selected via a Q_t-DPP sampling over X_t.

This mechanism is abstracted as an "expert call", but it is not assumed to be human or even a separate model. Rather, it reflects limited access to the current function values at previously observed locations, which can be interpreted in several realistic ways:

- **Wireless sensor networks:** In Internet-of-Things applications [26], sensors might collect data continuously but transmit selectively due to bandwidth or power constraints. Revisiting previous locations or reactivating a subset of sensors is often feasible, though costly — thus motivating a trade-off.
- **Physics-based monitoring:** In tasks such as environmental monitoring [27] where the underlying phenomenon is governed by a partial differential equation that needs to be simulated, the "expert" corresponds to access to the simulation itself. While running the simulator to evaluate the objective at a new point can be computationally expensive, it is often possible — though still costly — to re-run the simulator at previous input points to obtain updated objective values, reflecting changes in the underlying system.
- **Continual learning in ML systems:** For adaptive hyperparameter tuning or online systems [25], logs or cached evaluations might allow querying recent values again (e.g., checking performance of previous configurations on a new data batch).

We emphasize that the expert is not required to provide perfectly accurate information, but rather noisy or approximate values, consistent with a sub-Gaussian noise model. This is crucial in practice and aligns with many systems where re-evaluation is possible but noisy (e.g., due to changing conditions).

4 Numerical results

In this section, we compare the performance of SparQ-GP-UCB with four existing algorithms (TV-GP-UCB [2], W-GP-UCB [5], R-GP-UCB and SW-GP-UCB [24]) in a time-varying environment, on both a synthetic and a real-life dataset. We also run standard GP-UCB to show how it performs in time-varying settings. For all baseline methods, hyperparameters suchs as window size (R-GP-UCB, SW-GP-UCB), temporal kernel hyperparameter (TV-GP-UCB) and observations weights (W-GP-UCB) were set according to the recommendations provided by their respective authors.

4.1 Synthetic data

Observations are generated by perturbing the function

$$\begin{aligned} f\colon \mathcal{D} \times \mathbb{R}_+ &\to \mathbb{R}_+ \\ (x,t) &\mapsto \exp(-0.05(x - 5\sin(0.1t))^2) + 0.5\cos(0.2x) + 1.5 \end{aligned}$$

with noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$, where the sampling noise variance σ^2 is set to 0.01. We take the domain $\mathcal{D} = [-50, 50]$. Let $t \geq 0$. Then, by the mean value theorem

$$\sup_{x\in\mathcal{D}} |f(x,t+1) - f(x,t)| \leq \sup_{x,t} \left| \frac{\partial f(x,t)}{\partial t} \right| \leq 1,$$

and Assumption 2 holds. We plot the average and standard deviation of the cumulative regret of each algorithm for $T = 500$ iterations and $\delta = 0.05$ over 40 realizations using the squared exponential kernel, whose parameters have been fine-tuned by maximizing the log marginal likelihood of the data. For all four methods, we plot the mean and standard deviation of the average regret at each iteration. By selecting the number of queries $Q_T = 6\log(T)$ in line with the result of Proposition 2, we expect the average regret of SparQ-GP-UCB to vanish asymptotically. Furthermore, since the variation budget is not $V_T = o(T^{1/4})$ in our setting, due to the periodicity of $f \mapsto f(x,t)$, R-GP-UCB and SW-GP-UCB are not expected to have sublinear cumulative regret bounds.

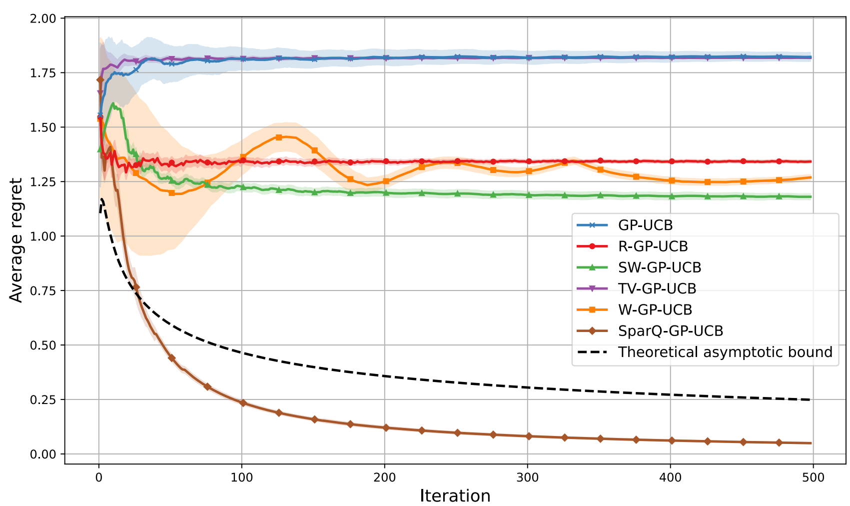

Fig. 1. Average regret of GP-UCB variants in the time-varying setting.

We can see in Figure 1 that SparQ-GP-UCB is the only method that converges to the dynamic optimum of the objective on average. Moreover, it falls below the theoretical bound (black curve) established in Eq. (19). Standard TV methods (TV-GP-UCB, W-GP-UCB, R-GP-UCB and SW-GP-UCB) struggle to track the optimum and have linear cumulative regret ($R_T \approx 1.8$ for TV-GP-UCB, $R_T \approx 1.3$ for W-GP-UCB, $R_T \approx 1.34T$ for R-GP-UCB and $R_T \approx 1.18T$ for SW-GP-UCB). As expected, the average regret of standard GP-UCB grows slowly, suggesting a slightly superlinear average regret. In summary, at the cost of $\tilde{\mathcal{O}}(1)$ additional observations per iteration, SparQ-GP-UCB is the only method capable of accurately optimizing a time-varying objective with weak assumptions on its temporal variations.

4.2 Real data

To evaluate the effectiveness of SparQ-GP-UCB, we conducted experiments on a real-world dataset consisting of daily ozone level measurements collected from 28 sensors distributed across the New York City area over the course of two years [28]. This dataset presents a naturally time-varying environment, making it an ideal testbed for adaptive Bayesian optimization methods.

Again, we benchmarked SparQ-GP-UCB against several state-of-the-art baselines: GP-UCB, R-GP-UCB, SW-GP-UCB, TV-GP-UCB, and W-GP-UCB. The evaluation metric used was cumulative regret, plotted as average regret over time to highlight long-term performance trends.

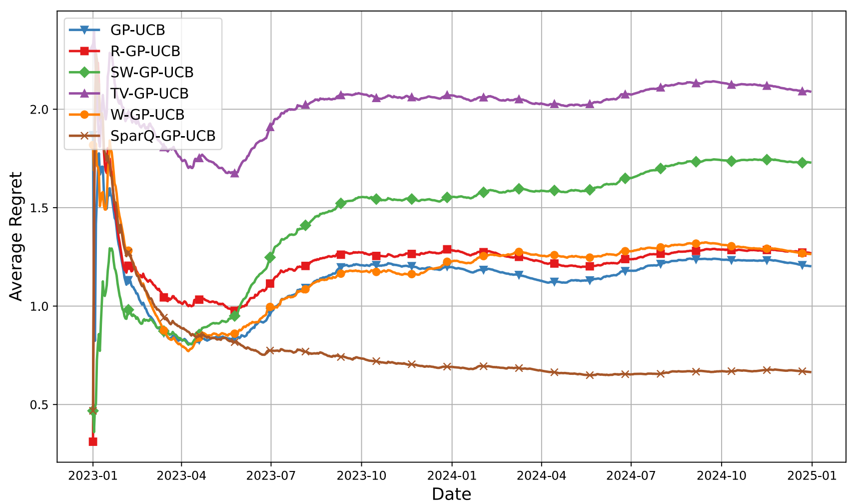

Fig. 2. Numerical performances of GP-UCB variants on real data.

Figure 2 shows the evolution of average regret across time steps (here one day). We set $\delta = 0.05$ and compute the kernel hyperparameters by maximizing the log marginal likelihood of the data. Although the ozone data is not generated by a model that respects our RKHS assumptions, we can see that SparQ-GP-UCB achieves significantly lower regret compared to other method. This indicates its superior ability to adapt to underlying non-stationary reward dynamics. Interestingly, GP-UCB achieves performance comparable to its time-varying counterparts. This may be attributed to characteristics of the ozone dataset — for example, the location of the maximum ozone level might be approximately stationary over time. Unlike the four time-varying baselines, SparQ-GP-UCB still consistently outperforms GP-UCB, suggesting that it effectively balances adaptation to temporal changes without overcompensating.

These results validate the robustness and adaptivity of our proposed method in capturing temporal variations and optimizing over dynamic environments.

5 Proofs of main theorems and additional results

5.1 Proof of Theorem 1

The regret proof is based on a few ingredients and extensions of previous work in [18, 12, 4]. We proceed as follows: first we extend Proposition 1 to the heteroscedastic and deterministic setting. Then we extend the regret results of [12] incorporating a sub-linearly growing maximum uncertainty, as well as the multiplicative error coming from the sparse inference. Then we combine the two results.

It is convenient to define the set of pretend observations together with the latest ones with noise $< \sigma^2 \log(T)$ as Y_T^{v} taken at points X_T^{v}. The cardinality of the two sets is T. These sets are not the same as the set of sparse plus latest one observations, indicated as Y_T^{s} and X_T^{s} whose cardinality is $\tilde{\mathcal{O}}(1)$.

The first proposition extends the sparse approximation to our setting, and in particular, it is an extension of Corollary 22 in [4] to heteroscedastic GPs and deterministic inputs X_T^{v} in a compact domain.

Proposition 2. *Let X_T^{v} be the set of actions chosen by SparQ-GP-UCB. Assume that pretend plus latest observations $Y_T^{\mathrm{v}}|X_T^{\mathrm{v}}$ are conditionally Gaussian distributed. Then, under the same assumptions as Theorem 1, for any $\eta > 0$ and any t, there exists an approximation level $\varepsilon_t = \mathcal{O}\left(\frac{\eta}{t}\right)$ and number of queries $Q_t = \mathcal{O}\left(\log^d\left(\frac{t}{\eta}\right)\right)$ such that running SparQ-GP-UCB with an ε_t-approximate Q_t-DPP provides a posterior distribution P_T^{s} satisfying*

$$\mathbb{E}[\mathrm{KL}[P_T^{\mathrm{s}} \| P_T^{\mathrm{v}}]] \leq \eta, \tag{23}$$

where P_T^{v} is the posterior distribution on $(X_T^{\mathrm{v}}, Y_T^{\mathrm{v}})$, and KL *is the KL divergence.*

Proof. We give the proof in the Appendix.

While the result of Equation (23) is given in expectation, we can also use Markov's inequality implying,

$$\mathrm{KL}[P_T^{\mathrm{s}} \| P_T^{\mathrm{v}}] \leq \frac{2\eta}{\delta}, \tag{24}$$

with probability $1 - \delta/2$.

Let us now recall Proposition 1 of [4].

Proposition 3. *[Proposition 1 of [4]] Let P and Q be the real and approximate posteriors with means μ_p, μ_q and variances σ_p^2 and σ_q^2. Suppose $2\mathrm{KL}[Q\|P] \leq \eta \leq \frac{1}{5}$ and let $x \in \mathbb{R}^d$. Then,*

$$|\mu_p(x) - \mu_q(x)| \leq \sigma_p(x)\sqrt{\eta} \leq \frac{\sigma_q(x)\sqrt{\eta}}{\sqrt{1-\sqrt{3\eta}}} \quad \text{and} \quad |1 - \sigma_q^2/\sigma_p^2| < \sqrt{3\eta}.$$

Since by Proposition 2 we have a way to bound the error coming from considering a sparse setting instead of the pretend setting, and by Propostion 3, we know how this translates into a multiplicative error of mean and variance, we are now ready for the regret result.

5.2 Regret proof

Proof. In SparQ-GP-UCB algorithm, the posterior mean $\mu_t^{\mathrm{s}}(.)$ and variance $(\sigma_t^{\mathrm{s}})^2(.)$ at step t are obtained by performing regression on the sparse observations $(X_{t-1}^{\mathrm{s}}, Y_{t-1}^{\mathrm{s}})$. The instantaneous regret of SparQ-GP-UBC at step t is:

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

where,

$$x_t^* = \underset{x \in \mathcal{D}}{\operatorname{argmax}}\ f_t(x) \quad \text{and} \quad x_t = \underset{x \in \mathcal{D}}{\operatorname{argmax}}\ \mu_{t-1}^{\mathrm{s}}(x) + \beta_t \sigma_{t-1}^{\mathrm{s}}(x).$$

By leveraging the definition of confidence bounds acquisition functions $x \mapsto \mathrm{ucb}_t(x) = \mu_{t-1}^{\mathrm{s}}(x) + \beta_t\sigma_{t-1}^{\mathrm{s}}(x)$ and $x \mapsto \mathrm{lcb}_t(x) = \mu_{t-1}^{\mathrm{s}}(x) - \beta_t\sigma_{t-1}^{\mathrm{s}}(x)$, it is possible to bound the cumulative regret with probability $1 - \delta/2$. To do that, we leverage the concentration bound provided in [9].

Proposition 4. *(Lemma 7, [9]) Take any $0 < \delta \leq 1$ and let $f_T \in \mathcal{H}_k(\mathcal{D})$ and $\mu_T(.)$ and $\sigma_T^2(.)$ be the posterior mean and covariance functions of $f_T(.)$ after observing (X_T, Y_T) points. Then, for any $x \in \mathcal{D}$, the following holds with probability at least $1 - \delta/2$:*

$$\forall t \in \{1, \dots, T\}, \qquad |\mu_{t-1}(x) - f_t(x)| \leq \beta_t \sigma_{t-1}(x) \tag{25}$$

where

$$\beta_t = \left(\sqrt{2\log\left(\frac{2\det\left(\Sigma_t + K_{tt}\right)^{1/2}}{\delta \det\left(\Sigma_t\right)^{1/2}}\right)} + \|f_t\|_k \right). \tag{26}$$

Proposition is valid for the heteroscedastic setting. As such, with this in place, and with probability $1 - \delta/2$:

$$r_t \leq \mathrm{ucb}_t(x_t^*) - \mathrm{lcb}_t(x_t) \leq \mathrm{ucb}_t(x_t) - \mathrm{lcb}_t(x_t) = 2\beta_t\sigma_{t-1}^{\mathrm{s}}(x_t).$$

Now we bound the cumulative regret at iteration T:

$$R_T = \sum_{t=1}^{T} r_t \leq 2\beta_T \sum_{t=1}^{T} \sigma_{t-1}^{\mathrm{s}}(x_t).$$

If we denote by $\sigma^{\mathrm{v}}_{t-1}(.)$ the posterior variance of the regression on the pretend plus latest observations $(X^{\mathrm{v}}_{t-1}, Y^{\mathrm{v}}_{t-1})$, Proposition 3 gives

$$\sigma^{\mathrm{s}}_{t-1}(x_t) \leq \sigma^{\mathrm{v}}_{t-1}(x_t)\sqrt{1+\sqrt{3}\eta},$$

with probability $1-\delta/2$, as long as we have a number of queries $\mathcal{O}(\log^d(t/\eta'))$ with $\eta' = \delta\eta/4$.

Therefore, the cumulative regret can be bounded with probability $1-\delta$ (for the union bound) as follows,

$$R_T \leq 2\beta_T\sqrt{1+\sqrt{3}\eta}\sum_{t=1}^{T}\sigma^{\mathrm{v}}_{t-1}(x_t).$$

The observations Y^{v}_{t-1} have been built such that their noise variance can be uniformly bounded by $\sigma^2\log(t-1)$. By following the exact same computation steps of Makarova et al. in [12] (Appendix A.1.1 Step 4) and replacing their fixed upper bound $\bar{\rho}$ by a logarithmically increasing upper bound $\sigma^2\log(T)$ we get

$$R_T \leq 2\beta_T\sqrt{1+\sqrt{3}\eta}\sqrt{2T(1+(\sigma^2\log(T))^2)\gamma_T}\,,$$

where γ_T is the maximum information gain at step T. Finally,

$$R_T = \mathcal{O}\left(\beta_T\sqrt{T\log^2(T)\gamma_T}\right). \tag{27}$$

Let us now bound β_T and γ_T.

In SparQ-GP-UCB, the ucb acquisition function is computed using the approximate posterior mean and variance. We thus have:

$$\beta_T = \sqrt{2\log\left(\frac{2|\Sigma^{\mathrm{s}}_T + K^{\mathrm{s}}_{TT}|^{1/2}}{\delta|\Sigma^{\mathrm{s}}_T|^{1/2}}\right)} + \|f_T\|_k.$$

By the definition of information gain with the sparse plus recent observations (see, e.g., [12]), we have

$$\gamma_{Q_T} \geq \log\left(\frac{|\Sigma^{\mathrm{s}}_T + K^{\mathrm{s}}_{TT}|}{|\Sigma^{\mathrm{s}}_T|}\right),$$

so that,

$$\beta_T = \mathcal{O}\left(\sqrt{\log\left(\frac{2}{\delta}\right) + \gamma_{Q_T}}\right). \tag{28}$$

If we combine bounds (28) and (27), we have a new expression for the regret bound:

$$R_T = \mathcal{O}\left(\sqrt{\left(\log\left(\frac{2}{\delta}\right) + \gamma_{Q_T}\right)\left(T\log^2(T)\gamma_T\right)}\right). \tag{29}$$

Again, by replacing $\bar{\rho}$ by $\sigma^2 \log(T)$ in Makarova et al. proof (Appendix A.1.3) and using $Q_T = \mathcal{O}\left(\log^d(T)\right)$, we can bound the information gains γ_T and γ_{Q_T} for a squared exponential kernel[3]:

$$\gamma_T = \mathcal{O}\left(d \log^{d+3}(T)\right), \tag{30}$$

$$\gamma_{Q_T} = \mathcal{O}\left(d \log^{d+3}\left(\log^d(T)\right)\right) = \mathcal{O}\left(d \log^{d+3}\left(d \log(T)\right)\right). \tag{31}$$

Finally, if we inject bounds (30) and (31) into (29):

$$\begin{aligned} R_T &= \mathcal{O}\left(\sqrt{\left(\log\left(\frac{1}{\delta}\right) + d \log^{d+3}\left(d \log(T)\right)\right)\left(T \log^2(T) d \log^{d+3}(T)\right)}\right) \\ &= \mathcal{O}\left(\sqrt{\left(\log\left(\frac{1}{\delta}\right) + d \log^{d+3}\left(d \log(T)\right)\right)\left(T d \log^{d+5}(T)\right)}\right) \end{aligned} \tag{32}$$

This proves Theorem 1. □

A closer look at the proof of Theorem 1 shows that one could choose to keep all the measurements with variance less than $g(T) = o(T^{1/4})$, as discussed in Section 3, instead limiting at the ones with variance less than $\sigma^2 \log(T)$. Since the maximum variance enters twice in the regret as a power of 2, then the final regret would read $R = \tilde{\mathcal{O}}(\sqrt{T}\sqrt{g^4(T)}) = o(T)$, leading to a sublinear cumulative regret and a no-regret result.

5.3 Proof of Theorem 2

The computational complexity of the algorithm proposed by Burt et al. [4] to obtain a ε approximation of a M-DPP from a set of N inputs is bounded as $\mathcal{O}\left(NM^3(\log\log N + \log M + \log 1/\varepsilon^2)\right)$, see their Section 4.2.2.

The cost of the GP regression with M training inputs is $\mathcal{O}(M^3)$ [16], and the complexity of SparQ-GP-UCB is dominated by the computation of the M-DPP. Thus, for T iterations in Algorithm 1 and with Q_T the number of sparse inputs at the end of the process, the computational complexity of SparQ-GP-UCB is T times the worst complexity of the DPP:

$$\mathcal{O}\left(T\left(Q_T\right)^3\left(\log\log T + \log Q_T + \log 1/\varepsilon_T^2\right)\right).$$

In Proposition 2, for fixed precision η, we show that $Q_T = \mathcal{O}\left(\log^d(T)\right)$, suffices to obtain a ε_T-approximation of a Q_T-DPP, with $\varepsilon_T = \mathcal{O}\left(\frac{1}{T}\right)$. By substituting these estimates into the complexity, we obtain a total computational complexity of SparQ-GP-UCB of $\mathcal{O}\left(T^2 \log(T) \log^{3d}(T)\right)$. □

[3] The information gain in a homoscedastic case for a SE kernel is $\mathcal{O}(d \log^{d+1}(T))$ to which we multiply a factor $\log^2(T)$ in our setting for the heteroscedastic case.

6 Conclusion

In this work, we provide a general framework to obtain sublinear regret bounds for GP optimization of a time-varying objective f in the bandit setting. The function f is assumed to belong to a RKHS with a bounded norm. We model time variations through uncertainty injection by linearly increasing the noise standard deviation of the data over time. We recover no-regret by asking $\tilde{\mathcal{O}}(1)$ additional side queries to an expert at each iteration. Future research will explore strategies to reduce the number of expert queries, such as retaining and reusing past responses to avoid querying the expert at every iteration.

Acknowledgments. This work was partly supported by the Agence Nationale de la Recherche (ANR) with the projects ANR AccelAILearning and ANR-23-CE48-0011-01.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Besbes, O., Gur, Y., Zeevi, A.: Stochastic multi-armed-bandit problem with non-stationary rewards. In: Advances in Neural Information Processing Systems. vol. 27 (2014)
2. Bogunovic, I., Scarlett, J., Cevher, V.: Time-varying Gaussian Process bandit optimization. In: International Conference on Artificial Intelligence and Statistics. vol. 51, pp. 314–323 (2016)
3. Brunzema, P., Von Rohr, A., Trimpe, S.: On controller tuning with time-varying Bayesian optimization. In: Conference on Decision and Control. pp. 4046–4052 (2022)
4. Burt, D.R., Rasmussen, C.E., van der Wilk, M.: Convergence of sparse variational inference in Gaussian Processes regression. Journal of Machine Learning Research **21**(131), 1–63 (2020)
5. Deng, Y., Zhou, X., Kim, B., Tewari, A., Gupta, A., Shroff, N.B.: Weighted Gaussian Process bandits for non-stationary environments. In: International Conference on Artificial Intelligence and Statistics. vol. 151, pp. 6909–6932 (2022)
6. Derezinski, M., Calandriello, D., Valko, M.: Exact sampling of determinantal point processes with sublinear time preprocessing. In: Advances in Neural Information Processing Systems. vol. 32 (2019)
7. Imamura, H., Charoenphakdee, N., Futami, F., Sato, I., Honda, J., Sugiyama, M.: Time-varying Gaussian Process bandit optimization with non-constant evaluation time. arXiv:2003.04691 (2020)
8. Jadbabaie, A., Rakhlin, A., Shahrampour, S., Sridharan, K.: Online Optimization : Competing with Dynamic Comparators. In: International Conference on Artificial Intelligence and Statistics. vol. 38, pp. 398–406 (2015)
9. Kirschner, J., Krause, A.: Information directed sampling and bandits with heteroscedastic noise. In: Conference On Learning Theory. pp. 358–384 (2018)
10. Koltchinskii, V., Giné, E.: Random matrix approximation of spectra of integral operators. Bernoulli **6**(1), 113–167 (2000)

11. Kulesza, A.: Determinantal point processes for machine learning. Foundations and Trends in Machine Learning **5**(2–3), 123–286 (2012)
12. Makarova, A., Usmanova, I., Bogunovic, I., Krause, A.: Risk-averse heteroscedastic Bayesian optimization. In: Advances in Neural Information Processing Systems. vol. 34, pp. 17235–17245 (2021)
13. Russac, Y., Faury, L., Cappé, O., Garivier, A.: Self-concordant analysis of generalized linear bandits with forgetting. In: International Conference on Artificial Intelligence and Statistics. vol. 130, pp. 658–666 (2021)
14. Scholkopf, B., Tsuda, K., Vert, J.P.: Kernel Methods in Computational Biology. The MIT Press (2004)
15. Shawe-Taylor, J., Williams, C., Cristianini, N., Kandola, J.: On the eigenspectrum of the gram matrix and the generalization error of kernel-PCA. IEEE Transactions on Information Theory **51**(7), 2510–2522 (2005)
16. Snelson, E., Ghahramani, Z.: Sparse Gaussian Processes using pseudo-inputs. In: Advances in Neural Information Processing Systems. vol. 18 (2005)
17. Srinivas, N., Krause, A., Kakade, S.M., Seeger, M.W.: Information-theoretic regret bounds for Gaussian Process optimization in the bandit setting. IEEE Transactions on Information Theory **58**(5), 3250—-3265 (2012)
18. Titsias, M.: Variational learning of inducing variables in sparse Gaussian Processes. In: International Conference on Artificial Intelligence and Statistics. vol. 5, pp. 567–574 (2009)
19. Titsias, M.K.: Variational model selection for sparse Gaussian Process regression. Tech. rep., University of Manchester, UK (2008)
20. Van Vaerenbergh, S., Lázaro-Gredilla, M., Santamaria, I.: Kernel recursive least-squares tracker for time-varying regression. IEEE Transactions on Neural Networks and Learning Systems **23**, 1313–1326 (2012)
21. Werge, N., Akgül, A., Kandemir, M.: BOF-UCB: A Bayesian-optimistic frequentist algorithm for non-stationary contextual bandits. arXiv:2307.03587 (2023)
22. Wu, Q., Iyer, N., Wang, H.: Learning contextual bandits in a non-stationary environment. In: International ACM SIGIR Conference on Research and Development in Information Retrieval (2018)
23. Zhao, P., Zhang, L., Jiang, Y., Zhou, Z.H.: A simple approach for non-stationary linear bandits. In: International Conference on Artificial Intelligence and Statistics. vol. 108, pp. 746–755 (2020)
24. Zhou, X., Shroff, N.: No-regret algorithms for time-varying Bayesian optimization. In: Annual Conference on Information Sciences and Systems. pp. 1–6 (2021)
25. Wang, L., Zhang, X., Su, H., Zhu, J.: A comprehensive survey of continual learning: Theory, method and application. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(8), 5362–5383 (2024)
26. Jamali-Rad, H., Campman, X.: Internet of things-based wireless networking for seismic applications. Geophysical Prospecting **66**(4), 833–853 (2018)
27. Roy, V., Simonetto, A., Leus, G.: Spatio-temporal sensor management for environmental field estimation. Signal Processing **128**, 369–381 (2016)
28. United States Environmental Protection Agency, *Download Daily Data*, https://www.epa.gov/outdoor-air-quality-data/download-daily-data.