Proceedings of Machine Learning Research vol XX:1–27, 2019

Gaussian Process Optimization with Adaptive Sketching: Scalable and No Regret

Daniele Calandriello DANIELE.CALANDRIELLO@IIT.IT
LCSL - Istituto Italiano di Tecnologia, Genova, Italy & MIT, Cambridge, USA

Luigi Carratino LUIGI.CARRATINO@DIBRIS.UNIGE.IT
UNIGE - Università degli Studi di Genova, Genova, Italy

Alessandro Lazaric LAZARIC@FB.COM
FAIR - Facebook AI Research, Paris, France

Michal Valko MICHAL.VALKO@INRIA.FR
INRIA Lille - Nord Europe, SequeL team, Lille, France

Lorenzo Rosasco LROSASCO@MIT.EDU
LCSL - Istituto Italiano di Tecnologia, Genova, Italy & MIT, Cambridge, USA
UNIGE - Università degli Studi di Genova, Genova, Italy

Abstract

Gaussian processes (GP) are a popular Bayesian approach for the optimization of black-box functions. Despite their effectiveness in simple problems, GP-based algorithms hardly scale to complex high-dimensional functions, as their per-iteration time and space cost is at least *quadratic* in the number of dimensions d and iterations t. Given a set of A alternative to choose from, the overall runtime $\mathcal{O}(t^3 A)$ quickly becomes prohibitive. In this paper, we introduce BKB (*budgeted kernelized bandit*), a novel approximate GP algorithm for optimization under bandit feedback that achieves near-optimal regret (and hence near-optimal convergence rate) with near-constant per-iteration complexity and no assumption on the input space or covariance of the GP.

Combining a kernelized linear bandit algorithm (GP-UCB) with randomized matrix sketching technique (i.e., leverage score sampling), we prove that selecting inducing points based on their posterior variance gives an accurate low-rank approximation of the GP, preserving variance estimates and confidence intervals. As a consequence, BKB does not suffer from *variance starvation*, an important problem faced by many previous sparse GP approximations. Moreover, we show that our procedure selects at most $\widetilde{\mathcal{O}}(d_{\text{eff}})$ points, where d_{eff} is the *effective* dimension of the explored space, which is typically much smaller than both d and t. This greatly reduces the dimensionality of the problem, thus leading to a $\mathcal{O}(TAd_{\text{eff}}^2)$ runtime and $\mathcal{O}(Ad_{\text{eff}})$ space complexity.

Keywords: Sparse Gaussian Process Optimization; Kernelized Linear Bandit; Regret; Sketching; Bayesian Optimization; Black Box Optimization; Variance Starvation

1. Introduction

Efficiently selecting the best alternative out of a set of alternatives is an important problem in sequential decision making, with practical applications ranging from recommender systems (Li et al., 2010), to experimental design (Robbins, 1952). It is also the main focus of the research in bandits (Lattimore and Szepesvári) and Bayesian optimization (Mockus, 1989; Pelikan, 2005; Snoek et al., 2012), that study optimization under bandit feedback. In this setting, a learning algorithm sequentially interacts with a reward/utility function f. Over T interactions, the algorithm chooses a point $\mathbf{x}_t$ and it has only access to a noisy black-box evaluation of f at $\mathbf{x}_t$. The goal of the algorithm is to minimize the cumulative regret, which compares the reward accumulated at the points selected over time (i.e., $\sum_t f(\mathbf{x}_t)$) to the reward obtained by repeatedly selecting the optimum of the function (i.e., $T \max_x f(x)$). In this paper we focus on the Gaussian process optimization approach to this problem. In particular, we study the GP-UCB algorithm first introduced by Srinivas et al. (2010).

Starting from a Gaussian process prior over f, GP-UCB alternates between evaluating the function, and using the evaluations to build a posterior of f. This posterior is composed by a mean function μ that estimates the value of f, and a variance function σ that captures the uncertainty of μ. These two quantities are combined in a single upper confidence bound (UCB) that drives the selection of the evaluation points, and trades off between evaluating high-reward points (i.e., exploitation) and testing possibly sub-optimal points to reduce the uncertainty on the function (i.e., exploration). While other approaches to select promising points exist, such as expected improvement (EI) and maximum probability of improvement, it is not known if they can achieve low regret. The performance of GP-UCB has been extensively studied by (Srinivas et al., 2010; Chowdhury and Gopalan, 2017) and it provably achieves low regret both in a Bayesian and non-Bayesian framework. However, the main limiting factor to its applicability is its computational cost. When choosing between A alternatives, GP-UCB requires $\Omega(At^2)$ time/space to select each new point, and therefore does not scale to complex function that require many iterations to be optimized. Several approximations of GP-UCB have been suggested before (Quinonero-Candela et al., 2007; Liu et al., 2018):

Inducing points: The GP can be restricted to lie in the range of a small subset of inducing points. The subset should cover well the whole space for accuracy, but also be as small as possible for efficiency. Several methods, referred to as *sparse GPs*, have been proposed to select the inducing points and an approximation based on the subset. Popular instances of this approach are the subset of regressors (SoR, Wahba, 1990) and the deterministic training conditional (DTC, Seeger et al., 2003). While these methods are simple to interpret and efficient, they do not come with regret guarantees. Moreover, when the subset does not cover the space well, they suffer from *variance starvation* Wang et al. (2018), as they underestimate the variance of points far away from the inducing points.

Random Fourier features: Another approach is to use explicit feature expansions to approximate the GP covariance function, and embed the points in a low-dimensional space, usually exploiting some variation of Fourier expansions (Rahimi and Recht, 2007). Among these methods, Mutny and Krause (2018) recently showed that discretizing the posterior on a fine grid of quadrature Fourier features (QFF) incurs a negligible approximation error. This is sufficient, among other things, to prove that the maximum of the approximate posterior can be efficiently found, and that it is accurate enough that Thompson sampling with quadratures provably achieves low regret. However this approach does not extend to non-stationary kernels, and although its dependency on t is small, the approximation and posterior maximization procedure scales exponentially with the input dimension.

Variational inference: another approach is to replace the true GP likelihood with a variational approximation that can be optimized efficiently. Although recent methods provide guarantees on the approximate posterior mean and variance (Huggins et al., 2019), these guarantees only apply to GP regression and not to the harder optimization setting.

Linear case: in the linear bandit literature, several methods to reduce the complexity of algorithms such as LinUCB have been proposed. Kuzborskij et al. (2019) uses the Frequent Directions (FD, Ghashami et al., 2016) to project the design matrix data to a smaller subspace. Unfortunately, the size of the subspace has to be specified in advance and when the size is not sufficiently large, the method can suffer linear regret. Prior to the FD approach, CBRAP (Yu et al., 2017) used random projections instead, but faced similar issues. This turns out to be a fundamental weakness of all approach that do not adapt the *actual* size of the space defined by the sequence of points selected by the learning algorithm. Indeed, Ghosh et al. (2017) showed a lower bound for the linear case that shows that as soon as one single arm is misspecified we can suffer linear regret.

1.1. Contributions

In this paper, we show a way to adapt the size of the projected space online and devise an algorithm, BKB (budgeted kernel bandit), that achieves near-optimal regret guarantees with a computational complexity that can be drastically smaller than GP-UCB. This is achieved without assumptions on the dimensionality of the input or on the kernel function. BKB leverages several well-known tools from the literature: a DTC approximation of the posterior variance, based on inducing points, and a confidence interval construction based on state-of-the-art self-normalized concentration inequalities (Abbasi-Yadkori et al., 2011). It also introduces two novel tools: a selection strategy to select inducing points based on ridge leverage score (RLS) sampling Alaoui and Mahoney (2015) that is provably accurate, and a novel approximate confidence interval for the function that is both nearly as accurate as the one of GP-UCB, but also efficient to compute.

Ridge leverage score sampling was originally introduced for randomized kernel matrix approximation (Alaoui and Mahoney, 2015). In the context of GPs, RLS correspond to the posterior variance of a point, which allows adapting algorithms and their guarantees from the RLS sampling literature to the GP setting. This solves two critical problems in sparse GP and linear bandit approximations. First, BKB constructs estimates of the variance that are provably accurate, i.e. it does not suffer from variance starvation, which results in provably accurate confidence bounds as well. The only method with comparable guarantees, Thompson sampling with quadrature FF (Mutny and Krause, 2018), only applies to stationary kernels, and only in extremely low dimension. Moreover our approximation guarantees are qualitatively different, as they do not require a corresponding *uniform* approximation bound on the GP.

Second, BKB adaptively chooses the size of the inducing point set based on the *effective dimension* d_{eff} of the problem, also known as degrees of freedom of the model. This is crucial to achieve low regret, since fixed approximation schemes may suffer linear regret. Moreover, in a problem with A arms, using a set of $\mathcal{O}(d_{\text{eff}})$ inducing points results in an algorithm with $\mathcal{O}(Ad_{\text{eff}}^2)$ per-step runtime and $\mathcal{O}(Ad_{\text{eff}})$ space, a significant improvement over the $\mathcal{O}(At^2)$ time and $\mathcal{O}(At)$ space cost of GP-UCB.

Finally, while in our work we only address kernelized (GP) bandits, our work could be extended to more complex online learning problems, such as to recent advances in kernelized reinforcement learning (Chowdhury and Gopalan, 2019). Moreover inducing point methods have clear model interpretability, and our analysis provides insight both from a bandit and Bayesian optimization perspective, making it applicable to a large amount of downstream tasks.

2. Background

Notation. We use lower-case letters a for scalars, lower-case bold letters $\mathbf{a}$ for vectors, and upper-case bold letters $\mathbf{A}$ for matrices and operators, where $[\mathbf{A}]_{ij}$ denotes its element (i,j). We denote by $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}$ the norm with metric $\mathbf{A}$, and $\|\mathbf{x}\| = \|\mathbf{x}\|_{\mathbf{I}}$ with $\mathbf{I}$ the identity. Finally, we denote the first T integers as $[T] := \{1, \dots, T\}$.

Online optimization under bandit feedback. Let $f : \mathcal{A} \rightarrow \mathbb{R}$ be a reward function that we wish to optimize over a set of decisions $\mathcal{A}$, also called actions or arms. For simplicity, we assume that $\mathcal{A} = \{\mathbf{x}_i\}_{i=1}^{A}$ is a fixed finite set of A vectors in $\mathbb{R}^d$. We discuss how to relax these assumptions in Section 5. In optimization under bandit feedback, a learner aims to optimize f through a sequence of interactions. At each step $t \in [T]$ the learner **1)** chooses an arm $\mathbf{x}_t \in \mathcal{A}$, **2)** receives reward $y_t = f(\mathbf{x}_t) + \eta_t$, where η_t is a zero-mean noise, **3)** updates its model of the problem.

The goal of the learner is to minimize its cumulative regret $R_T = \sum_{t=1}^{T} f(\mathbf{x}_*) - f(\mathbf{x}_t)$ w.r.t. the best[1] arm $\mathbf{x}_*$, where $\mathbf{x}_* = \arg\max_{\mathbf{x}_i \in \mathcal{A}} f(\mathbf{x}_i)$. In particular, the objective of a *no-regret* algorithm is to have R_T/T go to zero as T grows as fast as possible. We recall that the regret is strictly related to the convergence rate and the optimization performance. In fact, let $\overline{x}_T$ be an arm chosen at random from the sequence of arms $(x_1, \ldots, x_T)$ selected by the learner, then $f(\mathbf{x}_*) - \mathbb{E}[f(x_T)] \leq R_T/T$.

Gaussian process optimization and GP-UCB

A popular no-regret algorithm for optimization under bandit feedback is GP-UCB, introduced by Srinivas et al. (2010) in the context of Gaussian process optimization. We first introduce the formal definition of a Gaussian process, and then briefly present the GP-UCB algorithm.

A *Gaussian process* $GP(\mu, k)$ (Rasmussen and Williams, 2006) is a generalization of the Gaussian distribution to a space of functions and it is defined by a mean function $\mu(\cdot)$ and covariance function $k(\cdot, \cdot)$. W.l.o.g. we consider zero-mean $GP(0, k)$ priors and bounded covariance $k(\mathbf{x}_i, \mathbf{x}_i) \leq \kappa^2$ for all $\mathbf{x}_i \in \mathcal{A}$. An important property of Gaussian processes is that if we combine a prior $f \sim GP(0, k)$, and assume the observation noise is zero-mean Gaussian (i.e., $\eta_t \sim \mathcal{N}(0, \xi^2)$), then the posterior distribution of f conditioned on a set of observations $\{(\mathbf{x}_s, y_s)\}_{s=1}^{t}$ is also a GP. More precisely, let $\mathbf{X}_t = [\mathbf{x}_1, \ldots, \mathbf{x}_t]^\mathsf{T} \in \mathbb{R}^{t \times d}$ be the matrix with all arms selected so far and $\mathbf{y}_t = [y_1, \ldots, y_t]^\mathsf{T}$ be the corresponding observations, then the posterior is still a GP and the mean and variance of the function at a test point $\mathbf{x}$ are defined as

$$\mu_t(\mathbf{x} \mid \mathbf{X}_t, \mathbf{y}_t) = \mathbf{k}_t(\mathbf{x})^\mathsf{T}(\mathbf{K}_t + \lambda\mathbf{I})^{-1}\mathbf{y}_t, \tag{1}$$

$$\sigma_t^2(\mathbf{x} \mid \mathbf{X}_t) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_t(\mathbf{x})^\mathsf{T}(\mathbf{K}_t + \lambda\mathbf{I})^{-1}\mathbf{k}_t(\mathbf{x}), \tag{2}$$

where $\lambda = \xi^2$, $\mathbf{K}_t \in \mathbb{R}^{t \times t}$ is the matrix $[\mathbf{K}_t]_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ constructed from all pairs $\mathbf{x}_i, \mathbf{x}_j$ in $\mathbf{X}_t$, and $\mathbf{k}_t(\mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}), \ldots, k(\mathbf{x}_t, \mathbf{x})]^\mathsf{T}$. Notice that $\mathbf{k}_t(\mathbf{x})$ can be interpreted as an *embedding* of an arm $\mathbf{x}$ “supported” over the arms $\mathbf{x}_1, \ldots, \mathbf{x}_T$ observed so far.

The GP-UCB *algorithm* is a Bayesian optimization algorithm that uses a Gaussian process $GP(0, k)$ as a prior for f. Inspired by the optimism-in-face-of-uncertainty principle, at each time step t, GP-UCB uses the posterior GP to compute the mean and variance of an arm $\mathbf{x}_i$ and obtain the score

$$u_t(\mathbf{x}_i) = \mu_t(\mathbf{x}_i) + \beta_t\sigma_t(\mathbf{x}_i), \tag{3}$$

where we use the short-hand notation $\mu_t(\cdot) = \mu(\cdot \mid \mathbf{X}_t, \mathbf{y}_t)$ and $\sigma_t(\cdot) = \sigma(\cdot \mid \mathbf{X}_t)$. Finally, GP-UCB chooses the maximizer $\mathbf{x}_{t+1} = \arg\max_{\mathbf{x}_i \in \mathcal{A}} u_t(\mathbf{x}_i)$ as the next arm to evaluate. According to the score u_t, an arm $\mathbf{x}$ is likely to be selected if it has high mean reward μ_t *and* high variance σ_t, i.e., its estimated reward $\mu_t(\mathbf{x})$ is very uncertain. As a result, selecting the arm $\mathbf{x}_{t+1}$ with the largest score trades off between collecting (estimated) large reward (i.e., *exploitation*) and improving the accuracy of the posterior (i.e., *exploration*). The parameter β_t balances between these two objectives and it

[1] We assume a consistent and arbitrary tie-breaking strategy.

must be properly tuned to guarantee low regret. Srinivas et al. (2010) proposes different approaches to tune β_t for different assumptions on f and $\mathcal{A}$.

While GP-UCB is interpretable, simple to implement and provably achieves low regret, it is computationally expensive. In particular, computing $\sigma_t(\mathbf{x})$ has a complexity at least $\Omega(t^2)$ for the matrix-vector product $(\mathbf{K}_{t-1} + \xi^2\mathbf{I})^{-1}\mathbf{k}_{t-1}(\mathbf{x})$. Multiplying this for T iterations and A arms results in an overall $\mathcal{O}(AT^3)$ computational cost, which does not scale to large number of iterations T.

3. Budgeted Kernel Bandits

In this section, we introduce the BKB (*budgeted kernel bandit*) algorithm, a novel efficient approximation of GP-UCB, and we provide guarantees for its computational complexity. The analysis in Section 4 shows that BKB can be tuned to significantly reduce the complexity of GP-UCB with a negligible impact on the regret. We begin by introducing the two major contributions of this section: an approximation of the GP-UCB scores "supported" only by a small subset $\mathcal{S}_t$ of *inducing points*, and a method to *incrementally and adaptively* construct an accurate subset $\mathcal{S}_t$.

3.1. The algorithm

The main complexity bottleneck to compute the scores in Equation 3 is due to the fact that after t steps, the posterior GP is "supported" on *all* t previously seen arms. As a consequence, evaluating Equations 1 and 2 requires computing a t dimensional vector $\mathbf{k}_t(\mathbf{x})$ and $t \times t$ matrix $\mathbf{K}_t$ respectively. To avoid this dependency we restrict both $\mathbf{k}_t$ and $\mathbf{K}_t$ to be supported on a subset $\mathcal{S}_t$ of m arms. This approach is part of the sparse Gaussian process approximation framework (Quinonero-Candela et al., 2007), or equivalently a linear bandit constrained on a subspace (Kuzborskij et al., 2019).

Approximated GP-UCB scores. Consider a subset of arm $\mathcal{S}_t = \{\mathbf{x}_i\}_{i=1}^m$ and denote by $\mathbf{X}_{\mathcal{S}_t} \in \mathbb{R}^{m\times d}$ the matrix with all arms in $\mathcal{S}_t$ as rows. Let $\mathbf{K}_{\mathcal{S}_t} \in \mathbb{R}^{m\times m}$ be the matrix constructed by evaluating the covariance k between any two pair of arms in $\mathcal{S}_t$ and $\mathbf{k}_{\mathcal{S}_t}(\mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}), \dots, k(\mathbf{x}_m, \mathbf{x})]^\mathsf{T}$. The Nyström embedding $\mathbf{z}_t(\cdot)$ associated with subset $\mathcal{S}_t$ is defined as the mapping[2]

$$\mathbf{z}_t(\cdot) = \left(\mathbf{K}_{\mathcal{S}_t}^{1/2}\right)^+ \mathbf{k}_{\mathcal{S}_t}(\cdot) : \mathbb{R}^d \to \mathbb{R}^m,$$

where $(\cdot)^+$ indicates the pseudo-inverse. We denote with $\mathbf{Z}_t(\mathbf{X}_t) = [\mathbf{z}_t(\mathbf{x}_1), \dots, \mathbf{z}_t(\mathbf{x}_t)]^\mathsf{T} \in \mathbb{R}^{t\times m}$ the associated matrix of points and we define $\mathbf{V}_t = \mathbf{Z}_t(\mathbf{X}_t)^\mathsf{T}\mathbf{Z}_t(\mathbf{X}_t) + \lambda\mathbf{I}$. Then, we approximate the posterior mean and variance of the function on an arm $\mathbf{x}_i$ as

$$\begin{aligned}\widetilde{\mu}_t(\mathbf{x}_i) =& \mathbf{z}_t(\mathbf{x}_i)^\mathsf{T}\mathbf{V}_t^{-1}\mathbf{Z}_t(\mathbf{x}_i)^\mathsf{T}\mathbf{y}_t\\ \widetilde{\sigma}_t^2(\mathbf{x}_i) =& \frac{1}{\lambda}\Big(k(\mathbf{x}_i, \mathbf{x}_i) - \mathbf{z}_t(\mathbf{x}_i)^\mathsf{T}\mathbf{Z}_t(\mathbf{X}_t)^\mathsf{T}\mathbf{Z}_t(\mathbf{X}_t)\mathbf{V}_t^{-1}\mathbf{z}_t(\mathbf{x}_i)\Big).\end{aligned}$$

As a result, the approximate scores are obtained as

$$\widetilde{u}_t(\mathbf{x}_i) = \widetilde{\mu}_t(\mathbf{x}_i) + \widetilde{\beta}_t\widetilde{\sigma}_t(\mathbf{x}_i), \tag{4}$$

where $\widetilde{\beta}_t$ is appropriately tuned to achieve small regret in the theoretical analysis of Section 4. Finally, at each time step t, BKB selects arm $\widetilde{\mathbf{x}}_{t+1} = \arg\max_{\mathbf{x}_i\in\mathcal{A}} \widetilde{u}_t(\mathbf{x}_i)$.

[2] Recall that in the exact version, $\mathbf{k}_t(\mathbf{x})$ can be seen as an embedding of any arm $\mathbf{x}$ into the space induced by all the t arms selected so far, i.e. using all selected points as inducing points.

Notice that in general, $\widetilde{\mu}_t$ and $\widetilde{\sigma}_t$ do not correspond to any GP posterior. In fact, if we were simply replacing the $k(\mathbf{x}_i, \mathbf{x}_i)$ in the expression of $\widetilde{\sigma}_t^2(\mathbf{x}_i)$ by its value in the Nyström embedding, i.e., $\mathbf{z}_t(\mathbf{x}_i)^\mathsf{T}\mathbf{z}_t(\mathbf{x}_i)$, then we would recover a classical sparse GP approximations, the *subset of regressors*. Using $\mathbf{z}_t(\mathbf{x}_i)^\mathsf{T}\mathbf{z}_t(\mathbf{x}_i)$ is known to cause variance starvation, i.e., it can severely underestimate the variance of a test point $\mathbf{x}_i$ when it is far from the points in $\mathcal{S}_t$. Our formulation of $\widetilde{\sigma}_t$ is known in Bayesian literature as the *deterministic training conditional* (DTC), where it is used as a heuristic to prevent variance starvation. However, DTC does *not* correspond to a GP since it violates consistency (Quinonero-Candela et al., 2007). In this work, we justify this approach rigorously, showing that it is crucial to prove approximation guarantees necessary both for the optimization process and for the construction of the set of inducing points.

Choosing the inducing points. A critical aspect to effectively keep the complexity of BKB low while preserving regret guarantees is to carefully choose the inducing points to include in the subset $\mathcal{S}_t$. As the complexity of computing $\widetilde{u}_t$ scales with the size m of $\mathcal{S}_t$, a smaller set gives a faster algorithm. Conversely, the difference between $\widetilde{\mu}_t$ and $\widetilde{\sigma}_t$ and their exact counterparts depends on the accuracy of the embedding $\mathbf{z}_t$, which increases with the size of the set $\mathcal{S}_t$. Moreover, even for a fixed m, the quality of the embedding greatly depends on *which* inducing points are included. For instance, selecting the same arm as inducing point twice, or two co-linear arms, does not improve accuracy as the embedding space does not change. Finally, we need to take into account two important aspects of sequential optimization when choosing $\mathcal{S}_t$. First, we need to focus our approximation more on regions of $\mathcal{A}$ that are relevant to the optimization problem (i.e., high-reward arms). Second, as these regions change over time, we need to keep adapting the composition and size of $\mathcal{S}_t$ accordingly.

Algorithm 1: BKB

```
Data: Arm set $\mathcal{A}$, $q$, $\{\beta_t\}_{t=1}^T$
Result: Arm choices $\mathcal{D}_T = \{(\widetilde{\mathbf{x}}_t, y_t)\}$
Select uniformly at random $\mathbf{x}_1$ and observe $y_1$;
Initialize $\mathcal{S}_1 = \{\mathbf{x}_1\}$;
for $t = \{1, \dots, T-1\}$ do
    Compute $\widetilde{\mu}_t(\mathbf{x}_i)$ and $\widetilde{\sigma}_t^2(\mathbf{x}_i)$ for all $\mathbf{x}_i \in \mathcal{A}$;
    Select $\widetilde{\mathbf{x}}_{t+1} = \arg\max_{\mathbf{x}_i \in \mathcal{A}} \widetilde{u}_t(\mathbf{x}_i)$ (Eq. 4);
    for $i = \{1, \dots, t+1\}$ do
        Set $\widetilde{p}_{t+1,i} = \overline{q} \cdot \widetilde{\sigma}_t^2(\widetilde{\mathbf{x}}_i)$;
        Draw $q_{t+1,i} \sim Bernoulli(\widetilde{p}_{t+1,i})$;
        If $q_{t+1} = 1$ include $\widetilde{\mathbf{x}}_i$ in $\mathcal{S}_{t+1}$;
    end
end
```

Taking into consideration these objectives, we choose to construct $\mathcal{S}_t$ by randomly subsampling only out of the set of arms $\widetilde{\mathbf{X}}_t$ evaluated so far. In particular, arms are selected for inclusion in $\mathcal{S}_t$ with a probability proportional to their posterior variance σ_t at step t. We report the selection procedure in Algorithm 1, with the complete BKB algorithm.

We initialize $\mathcal{S}_1 = \{\widetilde{\mathbf{x}}_1\}$ by selecting an arm uniformly at random. At each step t, after selecting $\widetilde{\mathbf{x}}_{t+1}$, we must re-generate $\mathcal{S}_t$ to reflect the changes in $\widetilde{\mathbf{X}}_{t+1}$. Ideally, we would sample each arm in $\widetilde{\mathbf{X}}_{t+1}$ proportionally to σ_{t+1}^2, but this would be too computationally expensive. For efficiency we first approximate σ_{t+1}^2 with σ_t^2. This is equivalent to ignoring the last arm and does not significantly impact the accuracy. We can then replace σ_t^2 with $\widetilde{\sigma}_t^2$ that was already efficiently computed when constructing Equation 4. Finally, given a parameter $\overline{q} \geq 1$, we set our approximate inclusion probability as $\widetilde{p}_{t+1,i} = \overline{q}\widetilde{\sigma}_t^2(\widetilde{\mathbf{x}}_s)$. The $\overline{q}$ parameter is used to increase the inclusion probability in order to boost the overall success probability of the approximation procedure at the expense of a small increase in the size of $\mathcal{S}_{t+1}$. Given $\widetilde{p}_{t+1,i}$, we start from an empty $\mathcal{S}_{t+1}$ and iterate over all $\widetilde{\mathbf{x}}_i$ for $i \in [t+1]$ drawing $q_{t+1,i}$ from a Bernoulli distribution with probability $\widetilde{p}_{t+1,i}$. If $q_{t+1,i} = 1$, $\widetilde{\mathbf{x}}_i$ is included in $\mathcal{S}_{t+1}$.

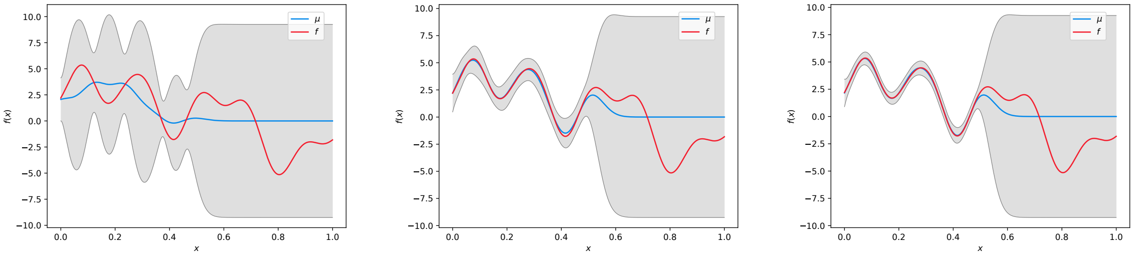

Fig. 1: We simulate a GP on $[0, 1] \in \mathbb{R}$ using Gaussian kernel with bandwidth $\sigma^2 = 100$. We draw a f from the GP, and give to BKB $t = \{6, 63, 215\}$ evaluations sampled uniformly in $[0, 0.5]$. We plot f, and $\widetilde{\mu}_t \pm 3\widetilde{\sigma}_t$.

Notice that while constructing $\mathcal{S}_t$ based on σ_t^2 is a common heuristic in the sparse GP literature, it has not been yet rigorously justified. In the next section, we show that this approach is equivalent to λ-ridge leverage score (RLS) sampling (Alaoui and Mahoney, 2015), a well studied tool in randomized linear algebra. We leverage the known results from this field to prove both accuracy and efficiency guarantees for our selection procedure.

3.2. Complexity analysis

Denote with $m_t = |\mathcal{S}_t|$ the size of the set $\mathcal{S}_t$ at step t. At each step, we first compute the embedding $\mathbf{z}_t(\mathbf{x}_i)$ of all arms in $\mathcal{O}(Am_t^2 + m_t^3)$ time, which corresponds to one inversion of $\mathbf{K}_{\mathcal{S}_t}^{1/2}$ and the matrix-vector product specific to each arm. We then rebuild the matrix $\mathbf{V}_t$ from scratch using all the arms observed so far. In general, it is sufficient to record the counters of the arms pulled so far, rather than the full list of arms, so that $\mathbf{V}_t$ can be constructed in $\mathcal{O}(\min\{t, A\}m_t^2)$ time. Then the inverse $\mathbf{V}_t^{-1}$ is computed in $\mathcal{O}(m_t^3)$ time. We can now efficiently compute $\widetilde{\mu}_t$, $\widetilde{\sigma}_t$, and $\widetilde{u}_t$ for all arms in $\mathcal{O}(Am_t^2)$ reusing the embeddings and $\mathbf{V}_t^{-1}$. Finally, computing all $q_{t+1,i}$ and $\mathcal{S}_{t+1}$ takes $\mathcal{O}(\min\{t+1, A\})$ using the estimated variances $\widetilde{\sigma}_t^2$. As a result, the per-step complexity is of order $\mathcal{O}\big((A + \min\{t, A\})m_T^2\big)$.[3] Space-wise, we only need to store the embedded arms and $\mathbf{V}_t$ matrix, which takes at most $\mathcal{O}(Am_T)$ space.

Bounding the size of $\mathcal{S}_T$. The size m_t of $\mathcal{S}_t$ can be expressed using the $q_{t,i}$ r.v. as the sum $m_t = \sum_{i=1}^t q_{t,i}$. In order to provide a bound on the total number of inducing points, which directly determines the computational complexity of BKB, we go through three major steps.

The first is to show that w.h.p., m_t is close to the sum $\sum_{i=1}^t \widetilde{p}_{t,i} = \sum_{i=1}^t \overline{q}\widetilde{\sigma}_t^2(\widetilde{\mathbf{x}}_i)$, i.e., close to the sum of the probabilities we used to sample each $q_{t,i}$. However the various $q_{t,i}$ are not independent, and each $\widetilde{p}_{t,i}$ is itself a r.v. Nonetheless all $q_{t,i}$ are conditionally independent given the previous $t-1$ steps, and this is sufficient to obtain the result.

The second and the most complex step is to guarantee that the random sum $\sum_{i=1}^t \widetilde{\sigma}_t^2(\widetilde{\mathbf{x}}_i)$ is close to $\sum_{i=1}^t \sigma_t^2(\widetilde{\mathbf{x}}_i)$, and at a lower level that each individual estimate $\widetilde{\sigma}_t^2(\cdot)$ is close to $\sigma_t^2(\cdot)$. To achieve this we exploit the connection between ridge leverage scores and posterior variance σ_t^2. In particular, we show that the variance estimator $\widetilde{\sigma}_t^2(\cdot)$ used by BKB is a variation of the RLS estimator recently introduced by Calandriello et al. (2017b) for RLS sampling. As a consequence, we can transfer the strong accuracy and size guarantees of RLS sampling to our optimization setting (see Appendix C). Note that anchoring the probabilities to the RLS (i.e. the sum of the posterior variances) means that

[3] Notice that $m_t \leq \min\{t, a\}$ and thus the complexity term $\mathcal{O}(m_t^3)$ is absorbed in the other terms.

the size of $\mathcal{S}_t$ naturally follows the effective dimension of the arms pulled so far. This strikes an adaptive balance between decreasing each individual probability to avoid $\mathcal{S}_t$ growing too large, while at the same time automatically increasing the effective degrees of freedom of the GP approximation when necessary.

The first two steps lead to $m_t \approx \sum_{i=1}^t \sigma_i^2(\widetilde{\mathbf{x}}_i)$, for which we need to derive a more explicit bound. In the GP literature, this quantity is bounded using the maximal information gain γ_T after T rounds. Denote with $\mathbf{X}_\mathcal{A} \in \mathbb{R}^{A \times d}$ the matrix with all arms as rows, with $\mathcal{D}$ a subset of these rows, potentially with duplicates, and with $\mathbf{K}_\mathcal{D}$ the associated kernel matrix. Then, Srinivas et al. (2010) define

$$\gamma_T = \max_{\mathcal{D} \subset \mathcal{A} : |\mathcal{D}| = T} \tfrac{1}{2} \log\det(\mathbf{K}_\mathcal{D}/\lambda + \mathbf{I}),$$

and show that $\sum_{i=1}^t \sigma_i^2(\widetilde{\mathbf{x}}_i) \leq \gamma_t$, and that γ_T itself can be bounded for specific $\mathcal{A}$ and kernel functions, e.g., $\gamma_T \leq \mathcal{O}(\log(T)^{d+1})$ for Gaussian kernels. Using the equivalence between RLS and posterior variance σ_t^2, we can also relate the posterior variance $\sigma_t^2(\widetilde{\mathbf{x}}_i)$ of the evaluated arms to the so-called GP's *effective dimension* d_{eff} or degrees of freedom

$$d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T) = \sum\nolimits_{i=1}^t \sigma_t^2(\widetilde{\mathbf{x}}_i) = \operatorname{Tr}(\mathbf{K}_T(\mathbf{K}_T + \lambda \mathbf{I})^{-1}), \tag{5}$$

using the following inequality by Calandriello et al. (2017a)

$$\log\det(\mathbf{K}_T/\lambda + \mathbf{I}) \leq \operatorname{Tr}(\mathbf{K}_T(\mathbf{K}_T + \lambda \mathbf{I})^{-1})\Big(1 + \log\Big(\tfrac{\|\mathbf{K}_T\|}{\lambda} + 1\Big)\Big). \tag{6}$$

We will use both RLS and d_{eff} to describe the behaviour of BKB's selection procedure.

Finally, we obtain the main result of this section.

Theorem 1 *For a desired* $0 < \varepsilon < 1$, $0 < \delta < 1$, *let* $\alpha = (1+\varepsilon)/(1-\varepsilon)$. *If we run* BKB *with* $\overline{q} \geq 6\alpha \log(4T/\delta)/\varepsilon^2$ *then with probability* $1-\delta$, *for all* $t \in [T]$ *and for all* $\mathbf{x} \in \mathcal{A}$ *we have*

$$\sigma_t^2(\mathbf{x})/\alpha \leq \widetilde{\sigma}_t^2(\mathbf{x}) \leq \alpha\sigma_t^2(\mathbf{x}) \qquad \textit{and} \qquad |\mathcal{S}_t| \leq 3(1+\kappa^2/\lambda)\alpha\overline{q}d_{\textit{eff}}(\lambda, \widetilde{\mathbf{X}}_t).$$

Computational complexity. We already showed that BKB's implementation with Nyström embedding required $\mathcal{O}(T(A + \min\{t, A\})m_T^3)$ time and $\mathcal{O}(Am_T)$ space. Combining this with Theorem 1 and the bound $m_T \leq \widetilde{\mathcal{O}}(d_{\text{eff}})$, we obtain a $\widetilde{\mathcal{O}}(TAd_{\text{eff}}^2 + \min\{t, A\})d_{\text{eff}}^3)$ time complexity. Whenever $d_{\text{eff}} \ll T$ and $T \ll A$ this is essentially a quadratic $\mathcal{O}(T^2)$ runtime, a large improvement over the quartic $\mathcal{O}(T^4) \leq \mathcal{O}(T^3A)$ runtime of GP-UCB.

Tuning $\overline{q}$. Note that although $\overline{q}$ must satisfy the condition reported in Theorem 1 for the result to hold, it is quite robust to uncertainty on the desired horizon T. In particular, the bound holds for *any* $\varepsilon > 0$, and even if we continue updating $\mathcal{S}_T$ after the T-th step, the bound still holds by implicitly increasing the parameter ε. Alternatively, after the T-th iteration the user can suspend the algorithm, increase $\overline{q}$ to suit the new desired horizon, and re-run only the subset selection subroutine on the arms selected so far.

Avoiding variance starvation. Another important consequence of Theorem 1 is that BKB's variance estimate are always close to the exact one up to a small constant factor. To the best of our knowledge, it makes BKB the first efficient and general GP algorithm that provably avoids variance starvation, which can be caused by two sources of error. The first source is the degeneracy, i.e. low-rankness, of the GP approximation, which causes the estimate to grow over-confident when the number of observed points grows and exceeds the degrees of freedom of the GP. BKB can adaptively choose its degrees of freedom as the size of $\mathcal{S}_t$ scales with the effective dimension. The second source of error arises when a point is far away from $\mathcal{S}_t$. Our use of a DTC variance estimator avoids under-estimation before we update the subset $\mathcal{S}_t$. Afterward, we can use guarantees on the quality of $\mathcal{S}_t$ to guarantee that we do not over-estimate the variance too much, exploiting a similar approach used to guarantee accuracy in RLS estimation. Both problems, and BKB's accuracy, are highlighted in Figure 1 using a benchmark experiment proposed by Wang et al. (2018).

Incremental dictionary update. At each step t, BKB recomputes the dictionary $\mathcal{S}_{t+1}$ from scratch by sampling each of the arms pulled so far with a suitable probability $\widetilde{p}_{t+1,i}$. A more efficient variant would be to build $\mathcal{S}_{t+1}$ by adding the new point $\mathbf{x}_{t+1}$ with probability $\widetilde{p}_{t+1,t+1}$ and including the points in $\mathcal{S}_t$ with probability $\widetilde{p}_{t+1,i}/\widetilde{p}_{t,i}$. This strategy is often used in the streaming setting to avoid storing all points observed so far and incrementally update the dictionary (see e.g., Calandriello et al., 2017b). Nonetheless, the stream of points, although arbitrary, is assumed to be generated *independently* from the dictionary itself. On the other hand, in our bandit setting, the points $\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \ldots$ are actually chosen by the learning algorithm depending on the dictionaries built over time, thus building a strong dependency between the stream of points and the dictionary itself. How to analyze such dependency and whether the accuracy of the inducing points is preserved in this case remains as an open question. Finally, notice that despite being more elegant and efficient, such incremental dictionary update would not significantly reduce the computational complexity, since computing the posterior variance for each arm would still dominate the overall runtime.

4. Regret Analysis

We are now ready to present the second main contribution of this paper, a bound on the regret achieved by BKB. To prove our result we additionally assume that the reward function f has bounded norm, i.e., $\|f\|_{\mathcal{H}}^2 = \langle f, f\rangle < \infty$. We use an upper-bound $\|f\|_{\mathcal{H}} \leq F$ to properly tune $\widetilde{\beta}_t$ to the "range" of the reward. If F is not known in advance, standard guess-and-double techniques can be applied.

Theorem 2 *Assume $\|f\|_{\mathcal{H}} \leq F < \infty$. For any desired $0 < \varepsilon < 1$, $0 < \delta < 1$, $0 < \lambda$, let $\alpha = (1+\varepsilon)/(1-\varepsilon)$ and $\overline{q} \geq 6\alpha\log(4T/\delta)/\varepsilon^2$. If we run* BKB *with*

$$\widetilde{\beta}_t = 2\xi\sqrt{\alpha\log(\kappa^2 t)\left(\sum\nolimits_{s=1}^{t}\widetilde{\sigma}_t^2(\widetilde{\mathbf{x}}_s)\right) + \log(1/\delta)} + \left(1 + \tfrac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F,$$

then, with prob. $1-\delta$, BKB*'s regret R_T is bounded by*

$$R_T \leq 2(2\alpha)^{3/2}\sqrt{T}\left(\xi d_{\mathit{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T) + \sqrt{\lambda F^2 d_{\mathit{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T)} + \xi\log(1/\delta)\right).$$

Theorem 2 shows that BKB achieves exactly the same regret as (exact) GP-UCB up to small α constant and $\log(\kappa^2 T)$ multiplicative factor.[4] For instance, setting $\varepsilon = 1/2$ results in a bound only $3\log(T)$ times larger than GP-UCB. At the same time, the choice $\varepsilon = 1/2$ only accounts for a constant factor 12 in the per-step computational complexity, which is still dramatically reduced from $t^2 A$ to $d_{\text{eff}}^2 A$. Note also that even if we send ε to 0, in the worst case we will include all

[4] Here we derive a *frequentist* regret bound and thus we compare with the result of Chowdhury and Gopalan (2017) rather than the original *Bayesian* analysis of Srinivas et al. (2010).

arms selected so far in $\mathcal{S}_t$, i.e. $\mathcal{S}_t = \{\widetilde{\mathbf{X}}_t\}$. Therefore even in this case BKB's runtime does not grow unbounded, but BKB simply transform back into exact GP-UCB. Moreover, it is easy to show that $d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T) \leq \log\det(\mathbf{K}_T/\lambda + \mathbf{I})$ (see Proposition 5 in the appendix), so any bound on $\log\det(\mathbf{K}_T/\lambda + \mathbf{I})$ available for GP-UCB can be directly applied to BKB. This means that up to an extra $\log(T)$ factor we also match GP-UCB's $\widetilde{\mathcal{O}}(\log(T)^{2d})$ rate for Gaussian kernel, $\widetilde{\mathcal{O}}(T^{\frac{1}{2}\frac{2\nu+3d^2}{2\nu+d^2}})$ rate for Matérn kernel, and $\widetilde{\mathcal{O}}(d\sqrt{T})$ for linear kernel. While these bounds are not minimax optimal, they closely follow the lower bounds derived in Scarlett et al. (2017). On the other hand, in the case of linear kernel (i.e., the linear bandit problem) we match the lower bound of Dani et al. (2008) up to logarithmic factors.

Another interesting aspect of BKB is that computing the trade-off parameter $\widetilde{\beta}_t$ can be done efficiently. Previous methods bounded this quantity with a loose (deterministic) upper bound (e.g., $\mathcal{O}(\log(T)^d)$ for Gaussian kernels) to avoid the large cost of computing $\log\det(\mathbf{K}_T/\lambda + \mathbf{I})$. In our $\widetilde{\beta}_t$, we bound the $\log\det$ by d_{eff}, which is then bounded by $\sum_{s=1}^{t} \widetilde{\sigma}_t^2(\mathbf{x}_s)$ (see Thm. 1), where all $\widetilde{\sigma}_t^2$ are already efficiently computed at each step. While this is up to $\log(t)$ larger than the exact $\log\det$, it is still data adaptive and much smaller than the worst case upper bounds.

This regret guarantee is crucially achieved without requiring an *increasing accuracy* in our approximation. One would expect that to obtain a sublinear regret the error induced by the approximation should decrease as $1/T$. Instead, in BKB the constants ε and λ that govern the accuracy level are fixed and thus it is not possible to guarantee that $\widetilde{\mu}_t$ will ever get close to μ_t everywhere. Adaptivity is key here: we can afford the same approximation level at every step because accuracy is actually increased only on a specific part of the arm set. For example, if a suboptimal arm is selected too often due to bad approximation, it will be eventually included in $\mathcal{S}_t$. After the inclusion, the approximation accuracy in the region of the suboptimal arm increases, and it would not be selected anymore. As the set of inducing points is updated *fast enough*, the impact of inaccurate approximations is limited in time, thus preventing large regret to accumulate. Note that this is a significant divergence from existing results. In particular approximation bounds that are uniformly accurate for all $\mathbf{x}_i \in \mathcal{A}$, such as those obtained with Quadrature FF (Mutny and Krause, 2018), rely on packing arguments. Due to the nature of packing, this usually causes the runtime or regret to scale exponentially in the input dimension d, and requires the kernel k to posses specific structure, e.g. to be statonary. Our new analysis avoids both of these problem.

Finally, we point out that the adaptivity of BKB allows drawing an interesting connection between learning and computational complexity. In fact, both the regret and the computation of BKB scale with the log-determinant and effective dimension of $\mathbf{K}_T$, which is related to the effective dimension of the sequence of arms selected over time. As a result, if the problem is difficult from a learning point of view (i.e., the regret is large because of large log-determinant), then BKB automatically adapts the set $\mathcal{S}_t$ by including many more inducing points to guarantee the level of accuracy needed to solve the problem. Conversely, if the problem is simple (i.e., small regret), then BKB can greatly reduce the size of $\mathcal{S}_t$ and achieve the derived level of accuracy.

4.1. Proof sketch

We build on the analysis of GP-UCB by Chowdhury and Gopalan (2017). Their analysis relies on a confidence interval formulation of GP-UCB that is more conveniently expressed using an explicit feature-based representation of the GP. For any GP with covariance k, there is a corresponding RKHS $\mathcal{H}$ with k as its kernel function. Furthermore, any kernel function k is associated to a non-linear feature map $\phi(\cdot) : \mathbb{R}^d \rightarrow \mathcal{H}$ such that $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}')^\mathsf{T}\phi(\mathbf{x}')$. As a result, any reward function $f \in \mathcal{H}$ can be written as $f(\mathbf{x}) = \phi(\mathbf{x})^\mathsf{T}\mathbf{w}_*$, where $\mathbf{w}_* \in \mathcal{H}$.

Confidence interval view of GP-UCB. Let $\boldsymbol{\Phi}(\mathbf{X}_t) = [\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_t)]^\mathsf{T}$ be the matrix $\mathbf{X}_t$ after the application of $\phi(\cdot)$ to each row. The regularized design matrix is then defined as $\mathbf{A}_t = \boldsymbol{\Phi}(\mathbf{X}_t)^\mathsf{T}\boldsymbol{\Phi}(\mathbf{X}_t) + \lambda\mathbf{I}$, while the regularized least-squares estimate is computed as

$$\widehat{\mathbf{w}}_t = \operatorname*{arg\,min}_{\mathbf{w}\in\mathcal{H}} \sum\nolimits_{i=1}^{t}(y_i - \phi(\mathbf{x}_i)^\mathsf{T}\mathbf{w})^2 + \lambda\|\mathbf{w}\|_2^2 = \mathbf{A}_t^{-1}\boldsymbol{\Phi}(\mathbf{X}_t)^\mathsf{T}\mathbf{y}_t.$$

We define the *confidence interval* C_t as the ellipsoid induced by $\mathbf{A}_t$ with center $\widehat{\mathbf{w}}_t$ and radius β_t

$$C_t = \{\mathbf{w} : \|\mathbf{w} - \widehat{\mathbf{w}}_t\|_{\mathbf{A}_t} \leq \beta_t\}, \qquad \beta_t = \lambda^{1/2}F + R\sqrt{2(\log\det(\mathbf{A}_t/\lambda) + \log(1/\delta)}, \tag{7}$$

where the radius β_t is such that $\mathbf{w}_* \in C_t$ w.h.p. (Chowdhury and Gopalan, 2017). Finally, using Lagrange multipliers we can reformulate the GP-UCB scores as

$$u_t(\mathbf{x}_i) = \max_{\mathbf{w}\in C_t} \phi(\mathbf{x}_i)^\mathsf{T}\mathbf{w} = \overbrace{\phi(\mathbf{x}_i)^\mathsf{T}\widehat{\mathbf{w}}_t}^{\mu_t(\mathbf{x}_i)} + \beta_t \overbrace{\sqrt{\phi(\mathbf{x}_i)^\mathsf{T}\mathbf{A}_t^{-1}\phi(\mathbf{x}_i)}}^{\sigma_t(\mathbf{x}_i)}. \tag{8}$$

Approximating the confidence ellipsoid. Let us focus now on the subset of arm $\mathcal{S}_t = \{\mathbf{x}_i\}_{i=1}^m$ chosen by BKB at each step, and denote by $\mathbf{X}_{\mathcal{S}_t} \in \mathbb{R}^{m\times d}$ the matrix with all arms in $\mathcal{S}_t$ as rows. We denote by $\widetilde{\mathcal{H}}_t = \operatorname{Im}(\boldsymbol{\Phi}(\mathbf{X}_{\mathcal{S}_t}))$ the smaller m-rank RKHS spanned by $\boldsymbol{\Phi}(\mathbf{X}_{\mathcal{S}_t})$, and by $\mathbf{P}_t$ the symmetric orthogonal projection operator on $\widetilde{\mathcal{H}}_t$. We can now define an *approximate* feature map $\widetilde{\phi}_t(\cdot) = \mathbf{P}_t\phi(\cdot) : \mathbb{R}^d \rightarrow \widetilde{\mathcal{H}}_t$ and associated approximations of $\mathbf{A}_t$ and $\widehat{\mathbf{w}}_t$

$$\widetilde{\mathbf{A}}_t = \widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t)^\mathsf{T}\widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t) + \lambda\mathbf{I} \tag{9}$$

$$\widetilde{\mathbf{w}}_t = \operatorname*{arg\,min}_{\mathbf{w}\in\mathcal{H}} \sum_{i=1}^{t}(y_i - \widetilde{\phi}(\mathbf{x}_i)^\mathsf{T}\mathbf{w})^2 + \lambda\|\mathbf{w}\|_2^2 = \widetilde{\mathbf{A}}_t^{-1}\widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t)^\mathsf{T}\mathbf{y}_t. \tag{10}$$

This leads to an approximate confidence ellipsoid $\widetilde{C}_t = \left\{\mathbf{w} : \|\mathbf{w} - \widetilde{\mathbf{w}}_t\|_{\widetilde{\mathbf{A}}_t} \leq \widetilde{\beta}_t\right\}$. A subtle element in these definitions is that while $\widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t)^\mathsf{T}\widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t)$ and $\widetilde{\mathbf{w}}_t$ are now restricted to $\widetilde{\mathcal{H}}_t$, the identity operator $\lambda\mathbf{I}$ in the regularization of $\widetilde{\mathbf{A}}_t$ still acts over the whole $\mathcal{H}$, and therefore $\widetilde{\mathbf{A}}_t$ does not belong to $\widetilde{\mathcal{H}}_t$ and remains full-rank and invertible. This immediately leads to using $k(\mathbf{x}_i, \mathbf{x}_i)$ in the definition of $\widetilde{\sigma}$ in Eq. 4, instead of the its approximate version using the Nyström embeddings.

Bounding the regret. To find an appropriate $\widetilde{\beta}_t$ we follow an approach similar to Abbasi-Yadkori et al. (2011). Exploiting the relationship $y_t = \widetilde{\phi}(\widetilde{\mathbf{x}}_t)^\mathsf{T}\mathbf{w}_* + \eta_t$, we bound

$$\|\mathbf{w}_* - \widetilde{\mathbf{w}}_t\|_{\widetilde{\mathbf{A}}_t}^2 \leq \overbrace{\lambda^{1/2}\|\mathbf{w}_*\|}^{(a)} + \overbrace{\|\widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t)\boldsymbol{\eta}_t\|_{\widetilde{\mathbf{A}}_t^{-1}}}^{(b)} + \overbrace{\|\boldsymbol{\Phi}(\mathbf{X}_t)^\mathsf{T}\|_{\mathbf{I}-\mathbf{P}_t} \cdot \|\mathbf{w}_*\|}^{(c)}.$$

Both (a) and (b) are present in GP-UCB and OFUL's analysis. Term (a) is due to the bias introduced in the least-square estimator $\widetilde{\mathbf{w}}_t$ by the regularization λ. Term (b) is due to the noisy in the reward observations. Note that the same term (b) appears in GP-UCB's analysis as $\|\boldsymbol{\Phi}(\mathbf{X}_t)\boldsymbol{\eta}_t\|_{\mathbf{A}_t^{-1}}$ and it is bounded by $\log\det(\mathbf{A}_t/\lambda)$ using self-normalizing concentration inequalities Chowdhury and Gopalan (2017). However our $\|\widetilde{\boldsymbol{\Phi}}_t(\mathbf{X}_t)\boldsymbol{\eta}_t\|_{\widetilde{\mathbf{A}}_t^{-1}}$ is a more complex object, since the projection $\mathbf{P}_t$

contained in $\widetilde{\mathbf{\Phi}}_t(\mathbf{X}_t) = \mathbf{P}_t\mathbf{\Phi}(\mathbf{X}_t)$ depends on the whole process up to time time t, and therefore $\widetilde{\mathbf{\Phi}}_t(\mathbf{X}_t)$ also depends on the whole process, losing its martingale structure. To avoid this, we use Sylvester's identity and the projection operator $\mathbf{P}_t$ to bound

$$\log\det(\widetilde{\mathbf{A}}_t/\lambda) = \log\det\left(\frac{\mathbf{\Phi}(\mathbf{X}_t)\mathbf{P}_t\mathbf{\Phi}(\mathbf{X}_t)^\mathsf{T}}{\lambda} + \mathbf{I}\right) \leq \log\det\left(\frac{\mathbf{\Phi}(\mathbf{X}_t)\mathbf{\Phi}(\mathbf{X}_t)^\mathsf{T}}{\lambda} + \mathbf{I}\right) = \log\det(\mathbf{A}_t/\lambda).$$

In other words, restricting the problem to $\widetilde{\mathcal{H}}_t$ acts as a regularization and reduces the variance of the martingale. Unfortunately, $\log\det(\mathbf{A}_t/\lambda)$ is too expensive to compute, so we first bound it with $d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_t)\log(\kappa^2 t)$, and then we bound $d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_t) \leq \alpha\sum_{s=1}^{t}\widetilde{\sigma}_t^2(\mathbf{x}_s)$ (Theorem 1), which can be computed efficiently. Finally, a new bias term (c) appears. Combining Theorem 1 with results from Calandriello and Rosasco (2018) on a projection $\mathbf{P}_t$ obtained using RLSs sampling we show that

$$\mathbf{I} - \mathbf{P} \preceq \frac{\lambda}{1-\varepsilon}\mathbf{A}_t^{-1}.$$

The combination of (a), (b), and (c) finally leads to the definition of $\widetilde{\beta}_t$ and the final regret bound $R_T \leq \sqrt{\widetilde{\beta}_T}\sqrt{\sum_{t=1}^{T}\boldsymbol{\phi}(\mathbf{x}_t)^\mathsf{T}\widetilde{\mathbf{A}}_t^{-1}\boldsymbol{\phi}(\mathbf{x}_t)}$. To conclude the proof we bound $\sum_{t=1}^{T}\boldsymbol{\phi}(\mathbf{x}_t)^\mathsf{T}\widetilde{\mathbf{A}}_t^{-1}\boldsymbol{\phi}(\mathbf{x}_t)$ with the following corollary to Theorem 1.

Corollary 3 *Under the same conditions as Theorem 2, for all $t \in T$ we have $\mathbf{A}_t/\alpha \preceq \widetilde{\mathbf{A}}_t \preceq \alpha\mathbf{A}_t$.*

Remarks. The novel bound $\|\mathbf{\Phi}(\mathbf{X}_t)^\mathsf{T}\|_{\mathbf{I}-\mathbf{P}_t} \leq \frac{\lambda}{1-\varepsilon}\|\mathbf{\Phi}(\mathbf{X}_t)^\mathsf{T}\|_{\mathbf{A}_t^{-1}}$ has a crucial role in controlling the bias due to the projection $\mathbf{P}_t$. Note that the second term measures the error with the same metric $\mathbf{A}_t^{-1}$ used by the variance martingale. In other words, the bias introduced by BKB's approximation can be seen as a *self-normalizing* bias. It is larger along directions that have been sampled less frequently, and smaller along directions correlated with arms selected often (e.g., the optimal arm). Our analysis bears some similarity with the one recently and independently developed by Kuzborskij et al. (2019). Nonetheless, our proof improves their result along two dimensions. First, we consider the more general (and challenging) GP optimization setting. Second, we do not fix the rank of our approximation in advance. While their analysis also exploits a self-normalized bias argument, this applies only to the k largest components. If the problem has an effective dimension larger than k, their radius and regret become essentially linear. In BKB we use our adaptive sampling scheme to include all necessary directions and to achieve the same regret rate as exact GP-UCB.

5. Discussion

As the literature in Bayesian optimization is vast and a complete review is out of the scope of this paper, we do not compare to alternative GP acquisition function, such as GP-EI or GP-PI, and we focus on approximation techniques with theoretical guarantees. Similarly, we exclude scalable variational inference based methods, even when their approximate posterior is provably accurate such as pF-DTC (Huggins et al., 2019), since they only provide guarantees for GP regression and not the harder optimization setting. We also do not discuss KERNELUCB (Valko et al., 2013), which has a tighter analysis than GP-UCB, since the algorithm construction is not efficient in practice.

Infinite arm sets. Looking at the proof of Theorem 1, the guarantees on $\widetilde{u}_t$ hold for all $\mathcal{H}$, and in Theorem 2 we only require that the maximum $\widetilde{\mathbf{x}}_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{A}}\max_{\mathbf{w}\in\widetilde{C}_t}\boldsymbol{\phi}(\mathbf{x})^\mathsf{T}\mathbf{w}$ is returned. Therefore, the accuracy and regret guarantees hold also for an infinite set of arms $\mathcal{A}$. However, the search over $\mathcal{A}$ can be difficult. In the general case maximization of a GP posterior is an NP-hard problem, with algorithms that usually scale exponentially with the input dimension d and are not practical. We focused instead on the easier case of finite sets, where enumeration is sufficient. Note that this automatically introduces an $\Omega(A)$ runtime dependency, which could be removed if the user can provide an efficient method to solve the maximization problem on a specific infinite set $\mathcal{A}$. As an example, Mutny and Krause (2018) prove that a GP posterior approximated using QFF can be optimized efficiently in low dimensions, and we expect similar results can hold for BKB and low effective dimension. Finally, note that recomputing a new set $\mathcal{S}_t$ still requires $\min\{A, t\}d_{\text{eff}}^2$ at each step. This represents a bottleneck in BKB independent from the arm selection problem, and must be addressed separately in future work.

Linear bandit with matrix sketching. Our analysis is closely related to CBRAP (Yu et al., 2017) and SOFUL (Kuzborskij et al., 2019). CBRAP uses Gaussian projections to embed all arms in a lower dimensional space for efficiency. Unfortunately their approach must either use an embedded space at least $\Omega(T)$ large, which in most cases would be even slower than exact OFUL, or it incurs linear regret w.h.p. Another approach for Euclidean spaces based on matrix approximation is SOFUL, introduced by Kuzborskij et al. (2019). It uses Frequent Direction (Ghashami et al., 2016), a method similar to incremental PCA, to embed the arms into $\mathbb{R}^m$, where m is fixed in advance. When discussing SOFUL, we must distinguish between SOFUL-UCB and SOFUL-TS, a variant based on Thompson sampling.
SOFUL-UCB achieves a $\widetilde{\mathcal{O}}(TAm^2)$ runtime, and $\widetilde{\mathcal{O}}((1+\varepsilon_m)^{3/2}(d+m)\sqrt{T})$ regret, where ε_m is the sum of the $d-m$ smallest eigenvalues of $\mathbf{A}_T$. However, notice that if the tail do not decrease quickly enough this algorithm may suffer linear regret and no adaptive way to tune m is provided. On the same task BKB does achieve a $\widetilde{\mathcal{O}}(d\sqrt{T})$ regret, since it adaptively chooses the size of the embedding. Computationally, directly instantiating BKB to use a linear kernel would achieve a $\widetilde{\mathcal{O}}(TAm_t^2)$ runtime, matching Kuzborskij et al. (2019)'s.
Compared to SOFUL-TS, BKB achieves better regret, but is potentially slower. Since Thompson sampling does not need to compute all confidence intervals, but can instead solve a simpler optimization problem, SOFUL-TS requires only $\widetilde{\mathcal{O}}(TAm)$ time against BKB's $\widetilde{\mathcal{O}}(TAm_t^2)$. It remains an open problem to see if a variant of BKB can match this complexity.

Approximate GP with RFF. Traditionally, RFF approaches have been popular to transform GP optimization in a finite-dimensional problem and allow for scalability. Unfortunately GP-UCB with traditional RFF is not low-regret, as RFF are well known to suffer from variance starvation (Wang et al., 2018) and unfeasibly large RFF embeddings would be necessary to prevent it. Recently Mutny and Krause (2018) proposed an alternative approach based on QFF, a specialized approach to random features for stationary kernels. They achieve the same regret rate as GP-UCB and BKB, with a near-optimal $\mathcal{O}(TA\log(T)^{d+1})$ runtime, and present additional variations based on Thompson sampling and exact posterior maximization. However Quadrature based approaches apply to stationary kernel only, and require to ε-cover $\mathcal{A}$, hence they cannot escape an exponential dependency on the dimensionality d. Conversely BKB can be applied to any kernel function, and while not specifically designed for this task it also achieve a close $\widetilde{\mathcal{O}}(TA\log(T)^{3(d+1)})$ runtime. Moreover in practice the size of $\mathcal{S}_T$ can be much less than exponential in d.

Acknowledgements This material is based upon work supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216. L. R. acknowledges the financial support of the AFOSR projects FA9550-17-1-0390 and BAA-AFRL-AFOSR-2016-0007 (European Office of Aerospace

Research and Development), and the EU H2020-MSCA-RISE project NoMADS - DLV-777826. The research presented was also supported by European CHIST-ERA project DELTA, French Ministry of Higher Education and Research, Nord-Pas-de-Calais Regional Council, Inria and Otto-von-Guericke-Universität Magdeburg associated-team north-European project Allocate, and French National Research Agency projects ExTra-Learn (n.ANR-14-CE24-0010-01) and BoB (n.ANR-16-CE23-0003). This research has benefited from the support of the FMJH Program PGMO and from the support to this program from Criteo.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

Ahmed El Alaoui and Michael W. Mahoney. Fast randomized kernel methods with statistical guarantees. In *Neural Information Processing Systems*, 2015.

Daniele Calandriello and Lorenzo Rosasco. Statistical and computational trade-offs in kernel k-means. In *Neural Information Processing Systems*, 2018.

Daniele Calandriello, Alessandro Lazaric, and Michal Valko. Second-order kernel online convex optimization with adaptive sketching. In *International Conference on Machine Learning*, 2017a.

Daniele Calandriello, Alessandro Lazaric, and Michal Valko. Distributed adaptive sampling for kernel matrix approximation. In *AISTATS*, 2017b.

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

Sayak Ray Chowdhury and Aditya Gopalan. Online learning in kernelized Markov decision processes. In *International Conference on Artificial Intelligence and Statistics*, may 2019. URL http://arxiv.org/abs/1805.08052.

Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

Mina Ghashami, Edo Liberty, Jeff M Phillips, and David P Woodruff. Frequent directions: Simple and deterministic matrix sketching. *SIAM Journal on Computing*, 45(5):1762–1792, 2016.

Avishek Ghosh, Sayak Ray Chowdhury, and Aditya Gopalan. Misspecified linear bandits. In *AAAI*, apr 2017. URL https://arxiv.org/abs/1704.06880.

Elad Hazan, Adam Kalai, Satyen Kale, and Amit Agarwal. Logarithmic regret algorithms for online convex optimization. In *Conference on Learning Theory*. Springer, 2006.

Jonathan H Huggins, Trevor Campbell, Mikołaj Kasprzak, and Tamara Broderick. Scalable gaussian process inference with finite-data mean and variance guarantees. *International Conference on Artificial Intelligence and Statistics*, 2019.

Ilja Kuzborskij, Leonardo Cella, and Nicolò Cesa-Bianchi. Efficient linear bandits through matrix sketching. In *International Conference on Artificial Intelligence and Statistics*, 2019.

Tor Lattimore and Csaba Szepesvári. Bandit algorithms.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. *International World Wide Web Conference*, 2010. URL http://rob.schapire.net/papers/www10.pdf.

Haitao Liu, Yew-Soon Ong, Xiaobo Shen, and Jianfei Cai. When gaussian process meets big data: A review of scalable gps. *arXiv preprint arXiv:1807.01065*, 2018.

Jonas Mockus. *Global optimization and the Bayesian approach*. 1989. doi: 10.1007/978-94-009-0909-0_1. URL http://www.springerlink.com/index/10.1007/978-94-009-0909-0{_}1.

Mojmir Mutny and Andreas Krause. Efficient High Dimensional Bayesian Optimization with Additivity and Quadrature Fourier Features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 9019–9030. Curran Associates, Inc., 2018.

Martin Pelikan. Hierarchical bayesian optimization algorithm. In *Hierarchical Bayesian Optimization Algorithm*, pages 105–129. Springer, 2005.

Joaquin Quinonero-Candela, Carl Edward Rasmussen, and Christopher KI Williams. Approximation methods for gaussian process regression. *Large-scale kernel machines*, pages 203–224, 2007.

Ali Rahimi and Ben Recht. Random features for large-scale kernel machines. In *Neural Information Processing Systems*, 2007.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. ISBN 978-0-262-18253-9. OCLC: ocm61285753.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematics Society*, 58:527–535, 1952.

Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. In *Conference on Learning Theory*, pages 1723–1742, 2017.

Matthias Seeger, Christopher Williams, and Neil Lawrence. Fast forward selection to speed up sparse gaussian process regression. In *Artificial Intelligence and Statistics 9*, number EPFL-CONF-161318, 2003.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

Niranjan Srinivas, Andreas Krause, Matthias Seeger, and Sham M Kakade. Gaussian process optimization in the bandit setting: No regret and experimental design. In *International Conference on Machine Learning*, pages 1015–1022, 2010.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. ISSN 1935-8237. doi: 10.1561/2200000048. URL http://dx.doi.org/10.1561/2200000048.

Michal Valko, Nathan Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 654–663. AUAI Press, 2013.

Grace Wahba. *Spline models for observational data*, volume 59. Siam, 1990.

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

Xiaotian Yu, Michael R. Lyu, and Irwin King. CBRAP: Contextual bandits with random projection. In *AAAI Conference on Artificial Intelligence*, 2017.

Appendix A. Relaxing assumptions

In the derivation so far we applied several assumptions. While some are necessary, others can be relaxed.

Assumptions on the noise. Throughout the paper, we assumed that the noise η_t is i.i.d. Gaussian. Since Chowdhury and Gopalan's results hold for any ξ-sub-Gussian noise that is measurable based on the previous iterations, this assumption can be easily relaxed.

Assumptions on the arms. So far we considered a set of arms that is (a) in $\mathbb{R}^d$, (b) fixed for all t, and (c) finite. Relaxing (a) is easy, since we do not make any assumption beyond boundedness on the kernel function k, and there are many bounded kernel function for non-Euclidean spaces, e.g. strings or graphs. Relaxing (b) is also trivial, we just need to embed the changing arm sets as they are provided, and store and re-embed previously selected arms as necessary. The per-step time complexity will now depend on the size of the set of arms available at each step. Relaxing (c) is straightforward from a theoretical perspective, but has varying computational consequences. In particular, looking at the proof of Theorem 1, the guarantees on $\widetilde{u}_t$ hold for all $\mathcal{H}$, and in Theorem 2 we only require that the maximum $\widetilde{\mathbf{x}}_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{A}} \max_{\mathbf{w}\in\widetilde{C}_t} \boldsymbol{\phi}(\mathbf{x})^\mathsf{T}\mathbf{w}$ is returned. Therefore, at least from the regret point of view, everything hold also for infinite $\mathcal{A}$. However, while the inner maximization over $\widetilde{C}_t$ can be solved in closed form for a fixed $\mathbf{x}$, the same cannot be said of the search over $\mathcal{A}$. If the designer can provide an efficient method to solve the maximization problem on an infinite $\mathcal{A}$, e.g., linear bandit optimization over compact subsets or $\mathbb{R}^d$, than all BKB guarantees hold. But in the general case maximization of a GP posterior, or of a general function over $\mathcal{H}$, is an NP-hard problem and we focused on the easier case of finite sets, where enumeration is sufficient. Note however that recomputing a new set $\mathcal{S}_t$ still requires $\min\{A, t\}d_{\text{eff}}^2$ at each step. This represents a bottleneck independent from the arm selection problem, and must be addressed separately.

Appendix B. Preliminary results on posterior variance

For simplicity we will collect here several results we will use on the posterior variance $\sigma_t^2(\cdot)$. While most of these hold for generic RLS, we will adapt them to our notation.

Proposition 4 (Calandriello et al., 2017b)

$$\frac{1}{\kappa^2/\lambda + 1}\sigma_{t-1}^2(\widetilde{\mathbf{x}}_t) \leq \frac{1}{\sigma_{t-1}^2(\widetilde{\mathbf{x}}_t) + 1}\sigma_{t-1}^2(\widetilde{\mathbf{x}}_t) \leq \sigma_t^2(\widetilde{\mathbf{x}}_t) \leq \sigma_{t-1}^2(\widetilde{\mathbf{x}}_t)$$

Proof The leftmost inequality follows from $\kappa^2/\lambda \geq \sigma_0^2(x)$ and $\sigma_a^2(x) \geq \sigma_b^2(x), \forall a \leq b$, the others from Calandriello et al., 2017b ■

Proposition 5 (Hazan et al., 2006; Calandriello et al., 2017a)

$$
\begin{aligned}
d_{\mathit{eff}}(\lambda, \widetilde{\mathbf{X}}_T) = \operatorname{Tr}(\mathbf{K}_T(\mathbf{K}_T + \lambda\mathbf{I})^{-1}) &= \sum\nolimits_{t=1}^{T} \sigma_T^2(\widetilde{\mathbf{x}}_t) \\
&\overset{(1)}{\leq} \sum\nolimits_{t=1}^{T} \sigma_t^2(\widetilde{\mathbf{x}}_t) \\
&\overset{(2)}{\leq} \log\det(\mathbf{K}_T/\lambda + \mathbf{I}) \\
&\overset{(3)}{\leq} \operatorname{Tr}(\mathbf{K}_T(\mathbf{K}_T + \lambda\mathbf{I})^{-1})\Big(1 + \log\Big(\tfrac{\|\mathbf{K}_T\|}{\lambda} + 1\Big)\Big).
\end{aligned}
$$

Proof Inequality (1) is due to Proposition 4. Inequality (2) is due to Hazan et al. (2006). Inequality (3) is due to Calandriello et al. (2017a). ■

Appendix C. Proof of Theorem 1

Let B_t be the bad event where the guarantees of Theorem 1 do not hold. Our goal in this proof is to prove that B_t happens at most with probability δ for all $t \in [T]$.

C.1. Notation

For notational simplicity, in the following we refer to $\mathbf{\Phi}(\widetilde{\mathbf{X}}_t)$ as $\mathbf{\Phi}_t$, $\widetilde{\mathbf{\Phi}}(\widetilde{\mathbf{X}}_t)$ as $\widetilde{\mathbf{\Phi}}_t$ and $\boldsymbol{\phi}(\widetilde{\mathbf{x}}_t)$ as $\boldsymbol{\phi}_t$. When the subscript is clear from the context, we will omit it. Since we leverage several results of Calandriello et al. (2017a), we report here some necessary additional notation.

First we extend our notation for the subset $\mathcal{S}_t$ to include a possible reweighing of the inducing points. We denote with $\mathcal{S}_t = \{(\boldsymbol{\phi}_j, s_j)\}_{j=1}^{m_t}$ a *weighted* subset, i.e. a weighted *dictionary*, of columns from $\mathbf{\Phi}_t$, with positive weights $s_j > 0$ that must be appropriately chosen.

Denote with $i_j \in [t]$ the index of the sample $\boldsymbol{\phi}_j$ as a column in $\mathbf{\Phi}_t$. Using a standard approach (Alaoui and Mahoney, 2015), we choose $s_j = 1/\sqrt{\widetilde{p}_{t,i_j}}$, where $\widetilde{p}_{t,i} = \overline{q}\widetilde{\sigma}_{t-1}^2(\widetilde{\mathbf{x}}_i)$ is the probability[5] used by Algorithm 1 when sampling $\boldsymbol{\phi}_{i_j}$ from $\mathbf{\Phi}_t$.

Let us now denote with $\mathbf{S}_t \in \mathbb{R}^{t \times t}$ the diagonal matrix with $q_{t,i}/\sqrt{\widetilde{p}_{t,i}}$ on the diagonal, where $q_{t,i}$ are the $\{0,1\}$ r.v. selected by Algorithm 1. Then we can see that

$$
\sum_{j=1}^{m_t} \frac{1}{\widetilde{p}_{t,i_j}} \boldsymbol{\phi}_{i_j}\boldsymbol{\phi}_{i_j}^\mathsf{T} = \sum_{i=1}^{t} \frac{q_{t,i}}{\widetilde{p}_{t,i}} \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\mathsf{T} = \mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T}. \tag{11}
$$

Calandriello et al. (2017b) define $\mathcal{S}_t$ to be an ε-accurate dictionary of $\mathbf{\Phi}_t$ if it satisfies the following condition

$$
(1-\varepsilon)\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} - \varepsilon\lambda\mathbf{I} \preceq \mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T} \preceq (1+\varepsilon)\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} + \varepsilon\lambda\mathbf{I}. \tag{12}
$$

We can also now fully define the projection operator at time t (see Section 4.1 for more details) as

$$
\mathbf{P}_t = \mathbf{\Phi}_t\mathbf{S}_t(\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T}\mathbf{\Phi}_t\mathbf{S}_t)^+\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T}
$$

the projection matrix spanned by the dictionary.

[5] Note that $\widetilde{p}_{t,i}$ might be larger than 1, but with a small abuse of notation and w.l.o.g. we will still refer to it as a probability.

C.2. Event decomposition

We can decompose Theorem 1 into an accuracy part, i.e. $\mathcal{S}_t$ must induce accurate $\widetilde{\sigma}_t$, and an efficiency part, i.e. $m_t \leq d_{\text{eff}}(t)$. We also connect the accuracy of $\widetilde{\sigma}_t$ to the definition of ε-accuracy.

Lemma 6 *Let $\alpha = \frac{1+\varepsilon}{1-\varepsilon}$. If $\mathcal{S}_t$ is ε-accurate w.r.t. $\mathbf{\Phi}_t$, then*

$$\mathbf{A}_t/\alpha \preceq \widetilde{\mathbf{A}}_t \preceq \alpha\mathbf{A}_t$$

and $\sigma_t^2(\mathbf{x})/\alpha \leq \min\{\widetilde{\sigma}_t^2(\mathbf{x}), 1\} \leq \alpha\sigma_t^2(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{A}$.

Proof Inverting the bound in Equation 12, and using the fact that $\mathbf{P}_t\mathbf{\Phi}_t\mathbf{S}_t = \mathbf{\Phi}_t\mathbf{S}_t$ we have

$$\begin{aligned}\mathbf{P}_t\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T}\mathbf{P}_t &\preceq \frac{1}{1-\varepsilon}(\mathbf{P}_t\mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T}\mathbf{P}_t + \varepsilon\lambda\mathbf{P}_t) \preceq \frac{1}{1-\varepsilon}(\mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T} + \varepsilon\lambda\mathbf{P}_t)\\ &\preceq \frac{1}{1-\varepsilon}((1+\varepsilon)\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} + \varepsilon\lambda\mathbf{I} + \varepsilon\lambda\mathbf{P}_t) \preceq \frac{1+\varepsilon}{1-\varepsilon}\left(\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} + \frac{2\varepsilon}{1+\varepsilon}\lambda\mathbf{I}\right).\end{aligned}$$

Repeating for the other side we obtain

$$\frac{1-\varepsilon}{1+\varepsilon}\left(\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} - \frac{2\varepsilon}{1-\varepsilon}\lambda\mathbf{I}\right) \preceq \mathbf{P}_t\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T}\mathbf{P}_t \preceq \frac{1+\varepsilon}{1-\varepsilon}\left(\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} + \frac{2\varepsilon}{1+\varepsilon}\lambda\mathbf{I}\right).$$

Applying this to $\widetilde{\mathbf{A}}_t$ we have

$$\widetilde{\mathbf{A}}_t = \mathbf{P}_t\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T}\mathbf{P}_t + \lambda\mathbf{I} \succeq \frac{1-\varepsilon}{1+\varepsilon}\left(\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} - \frac{2\varepsilon}{1-\varepsilon}\lambda\mathbf{I}\right) + \lambda\mathbf{I} = \frac{1-\varepsilon}{1+\varepsilon}(\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} + \lambda\mathbf{I}) = \frac{1-\varepsilon}{1+\varepsilon}\mathbf{A}_t,$$

which can again be applied on the other side to obtain our result. To prove the accuracy of the approximate posterior variance $\widetilde{\sigma}_t^2(\mathbf{x}_i)$ we simply apply the definition

$$\frac{1-\varepsilon}{1+\varepsilon}\overbrace{\boldsymbol{\phi}_i^\mathsf{T}\mathbf{A}_t\boldsymbol{\phi}_i}^{\sigma_t^2(\mathbf{x}_i)} \preceq \overbrace{\boldsymbol{\phi}_i^\mathsf{T}\widetilde{\mathbf{A}}_t\boldsymbol{\phi}_i}^{\widetilde{\sigma}_t^2(\mathbf{x}_i)} \preceq \frac{1+\varepsilon}{1-\varepsilon}\overbrace{\boldsymbol{\phi}_i^\mathsf{T}\mathbf{A}_t\boldsymbol{\phi}_i}^{\sigma_t^2(\mathbf{x}_i)},$$

■

Using Lemma 6 we can now decompose our bad event $B_t = A_t \cup E_t$ where A_t is the event where $\mathcal{S}_t$ is not ε-accurate w.r.t. $\mathbf{\Phi}_t$. and E_t is the event where m_t is much larger than $d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_t)$. We can further decompose the event A_t

$$\begin{aligned}A_t &= (A_t \cap A_{t-1}) \cup (A_t \cap A_{t-1}^\complement)\\ &\subseteq A_{t-1} \cup (A_t \cap A_{t-1}^\complement) = A_0 \cup \left(\bigcup_{s=1}^{t}(A_s \cap A_{s-1}^\complement)\right) = \bigcup_{s=1}^{t}(A_s \cap A_{s-1}^\complement),\end{aligned}$$

where A_0 is the empty event since $\mathbf{\Phi}_0$ is empty and it is well approximated by the empty $\mathcal{S}_0$. Moreover, we can simplify a part of the problem by noting

$$B_t = A_t \cup E_t = A_t \cup (E_t \cap A_{t-1}^\complement) \cup (E_t \cap A_{t-1}) \subseteq A_t \cup A_{t-1} \cup (E_t \cap A_{t-1}^\complement),$$

which will help us when bounding the event E_t, where we will directly act as if A_t does not hold. Putting it all together

$$\begin{aligned}\bigcup_{t=1}^{T} B_t &= \bigcup_{t=1}^{T} (A_t \cup E_t) \subseteq \bigcup_{t=1}^{T} \Big(A_t \cup A_{t-1} \cup (E_t \cap A_{t-1}^{\complement}) \Big) \\ &= \left(\bigcup_{t=1}^{T} A_t \right) \cup \left(\bigcup_{t=1}^{T} (E_t \cap A_{t-1}^{\complement}) \right) = \left(\bigcup_{t=1}^{T} A_t \right) \cup \left(\bigcup_{t=1}^{T} (E_t \cap A_{t-1}^{\complement}) \right) \\ &\subseteq \left(\bigcup_{t=1}^{T} \left(\bigcup_{s=1}^{t} (A_s \cap A_{s-1}^{\complement}) \right) \right) \cup \left(\bigcup_{t=1}^{T} (E_t \cap A_{t-1}^{\complement}) \right) \\ &= \left(\bigcup_{t=1}^{T} (A_t \cap A_{t-1}^{\complement}) \right) \cup \left(\bigcup_{t=1}^{T} (E_t \cap A_{t-1}^{\complement}) \right)\end{aligned}$$

C.3. Bounding $\Pr(A_t \cap A_{t-1}^{\complement})$

We can now bound the probability of event $A_t \cap A_{t-1}^{\complement}$ happening. Our first step is to formally define A_t using Equation 12. In particular we can rewrite the ε-accuracy condition as

$$\begin{aligned}&(1-\varepsilon)\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} - \varepsilon\lambda\mathbf{I} \preceq \mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^{\mathsf{T}}\mathbf{\Phi}_t^{\mathsf{T}} \preceq (1+\varepsilon)\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \varepsilon\lambda\mathbf{I} \\ \iff\ & -\varepsilon(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I}) \preceq \mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^{\mathsf{T}}\mathbf{\Phi}_t^{\mathsf{T}} - \mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} \preceq \varepsilon(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I}) \\ \iff\ & -\varepsilon\mathbf{I} \preceq (\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-1/2}(\mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^{\mathsf{T}}\mathbf{\Phi}_t^{\mathsf{T}} - \mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}})(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-1/2} \preceq \varepsilon\mathbf{I} \\ \iff\ & \|(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-1/2}(\mathbf{\Phi}_t\mathbf{S}_t\mathbf{S}_t^{\mathsf{T}}\mathbf{\Phi}_t^{\mathsf{T}} - \mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}})(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-1/2}\| \leq \varepsilon,\end{aligned}$$

where $\|\cdot\|$ is the spectral norm. We can now focus on this last reformulation, and frame it as a random matrix concentration problem in a RKHS $\mathcal{H}$. Let $\psi_{t,i} = (\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-\frac{1}{2}}\boldsymbol{\phi}_i$ and $\mathbf{\Psi}_t = \mathbf{\Phi}_t(\mathbf{\Phi}_t^{\mathsf{T}}\mathbf{\Phi}_t + \lambda\mathbf{I})^{-\frac{1}{2}} = [\psi_{t,1}, \dots, \psi_{t,t}]^{\mathsf{T}}$, and define the operator $\mathbf{G}_{t,i} = \Big(\frac{q_{t,i}}{\widetilde{p}_{t,i}} - 1\Big)\psi_{t,i}\psi_{t,i}^{\mathsf{T}}$. Then we can rewrite ε-accuracy as

$$\|(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-\frac{1}{2}}\mathbf{\Phi}_t(\mathbf{S}_t\mathbf{S}_t^{\mathsf{T}} - \mathbf{I})\mathbf{\Phi}_t^{\mathsf{T}}(\mathbf{\Phi}_t\mathbf{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-\frac{1}{2}}\| = \left\| \sum_{i=1}^{t} \left(\frac{q_{t,i}}{\widetilde{p}_{t,i}} - 1 \right) \psi_{t,i}\psi_{t,i}^{\mathsf{T}} \right\| = \left\| \sum_{i=1}^{t} \mathbf{G}_{t,i} \right\| \leq \varepsilon,$$

and the event A_t as the event where $\left\|\sum_{i=1}^{t} \mathbf{G}_{t,i}\right\| \geq \varepsilon$, Note that this reformulation exploits the fact that $q_{t,i} = 0$ encodes the column that are not selected in $\mathcal{S}_t$ (see Equation 11). To study this random object, we begin by defining the filtration $\mathcal{F}_t = \{q_{s,i}, \eta_s\}_{s=1}^{t}$ at time t containing all the randomness coming from the construction of the various $\mathcal{S}_s$ and the noise on the function η_t. In particular, note that the $\{0,1\}$ r.v. $q_{t,i}$ used by Algorithm 1 are not Bernoulli r.v., since the probability $\widetilde{p}_{t,i}$ used to select 0 or 1 is itself random. However they become well defined Bernoulli when conditioned on

$\mathcal{F}_{t-1}$. Let $\mathbb{I}\{\cdot\}$ indicate the indicator function of an event. We can rewrite

$$\begin{aligned}\Pr(A_t \cap A_{t-1}^{\complement}) &= \Pr\left(\left\|\sum_{i=1}^{t}\mathbf{G}_{t,i}\right\| \geq \varepsilon \cap \left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\| \leq \varepsilon\right)\\ &= \mathop{\mathbb{E}}_{\mathcal{F}_t}\left[\mathbb{I}\left\{\left\|\sum_{i=1}^{t}\mathbf{G}_{t,i}\right\| \geq \varepsilon \cap \left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\| \leq \varepsilon\right\}\right]\\ &= \mathop{\mathbb{E}}_{\mathcal{F}_{t-1}}\left[\mathop{\mathbb{E}}_{\eta_t,\{q_{t,i}\}}\left[\mathbb{I}\left\{\left\|\sum_{i=1}^{t}\mathbf{G}_{t,i}\right\| \geq \varepsilon \cap \left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\| \leq \varepsilon\right\} \;\middle|\; \mathcal{F}_{t-1}\right]\right]\\ &= \mathop{\mathbb{E}}_{\mathcal{F}_{t-1}}\left[\mathop{\mathbb{E}}_{\{q_{t,i}\}}\left[\mathbb{I}\left\{\left\|\sum_{i=1}^{t}\mathbf{G}_{t,i}\right\| \geq \varepsilon \cap \left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\| \leq \varepsilon\right\} \;\middle|\; \mathcal{F}_{t-1}\right]\right],\end{aligned}$$

where the last passage is due to the fact that $\mathbf{G}_{t,i}$ is independent from η_t. Notice now that conditioned on $\mathcal{F}_{t-1}$ the event $A_{t-1}^{\complement}$ is not random anymore, and we can restrict our expectations to the outcomes where $\left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\| \leq \varepsilon$,

$$\Pr(A_t \cap A_{t-1}^{\complement}) = \mathop{\mathbb{E}}_{\mathcal{F}_{t-1}:\left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\|\leq\varepsilon}\left[\mathop{\mathbb{E}}_{\{q_{t,i}\}}\left[\mathbb{I}\left\{\left\|\sum_{i=1}^{t}\mathbf{G}_{t,i}\right\| \geq \varepsilon\right\} \;\middle|\; \mathcal{F}_{t-1}\right]\right].$$

Moreover, conditioned on $\mathcal{F}_{t-1}$ all the $q_{t,i}$ become independent r.v., and we can use the following result of Tropp (2015).

Proposition 7 *Let $\mathbf{G}_1, \ldots, \mathbf{G}_n$ be a sequence of independent self-adjoint random operators such that $\mathbb{E}\,\mathbf{G}_i = 0$ and $\|\mathbf{G}_i\| \leq R$ a.s.*
Denote $\sigma^2 = \left\|\sum_{i=1}^{t}\mathbb{E}\,\mathbf{G}_i^2\right\|$. Then for any $\varepsilon \geq 0$

$$\Pr\left(\left\|\sum_{i=1}^{t}\mathbf{G}_i\right\| \geq \varepsilon\right) \leq 4t\exp(\frac{\varepsilon^2/2}{\sigma^2 + R\varepsilon/3})$$

We begin by computing the mean of $\mathbf{G}_{t,i}$

$$\begin{aligned}\mathop{\mathbb{E}}_{q_{t,i}}[\mathbf{G}_{t,i} \mid \mathcal{F}_{t-1}] &= \mathop{\mathbb{E}}_{q_{t,i}}\left[\left(\frac{q_{t,i}}{\widetilde{p}_{t,i}} - 1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}} \;\middle|\; \mathcal{F}_{t-1}\right]\\ &= \left(\frac{\mathbb{E}_{q_{t,i}}[q_{t,i} \mid \mathcal{F}_{t-1}]}{\widetilde{p}_{t,i}} - 1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}} = \left(\frac{\widetilde{p}_{t,i}}{\widetilde{p}_{t,i}} - 1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}} = \mathbf{0},\end{aligned}$$

where we use the fact that $\widetilde{p}_{t,i}$ is fixed conditioned on $\mathcal{F}_{t-1}$ and it is the (conditional) expectation of $q_{t,i}$. Since $\mathbf{G}$ is zero-mean, we can use Proposition 7. First we find R. We can upper bound

$$\|\mathbf{G}_{t,i}\| = \left\|\left(\frac{q_{t,i}}{\widetilde{p}_{t,i}} - 1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\right\| \leq \left|\left(\frac{q_{t,i}}{\widetilde{p}_{t,i}} - 1\right)\right|\|\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\| \leq \frac{1}{\widetilde{p}_{t,i}}\|\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\|.$$

Note that due to the definition of $\psi_{t,i}$

$$\|\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\| = \psi_{t,i}^{\mathsf{T}}\psi_{t,i} = \boldsymbol{\phi}_i^{\mathsf{T}}(\boldsymbol{\Phi}_t\boldsymbol{\Phi}_t^{\mathsf{T}} + \lambda\mathbf{I})^{-1}\boldsymbol{\phi}_i = \sigma_t^2(\widetilde{\mathbf{x}}_i).$$

Moreover, we are only considering outcomes of $\mathcal{F}_{t-1}$ where $\left\|\sum_{i=1}^{t}\mathbf{G}_{t-1,i}\right\| \leq \varepsilon$, which implies that $\mathcal{S}_{t-1}$ is ε accurate, and by Lemma 6 we have $\widetilde{\sigma}_{t-1}(\widetilde{\mathbf{x}}_i) \geq \sigma_{t-1}(\widetilde{\mathbf{x}}_i)/\alpha$. Finally, due to Proposition 4 we have $\sigma_{t-1}(\widetilde{\mathbf{x}}_i) \geq \sigma_t(\widetilde{\mathbf{x}}_i)$. Putting this all together we can bound

$$\frac{1}{\widetilde{p}_{t,i}}\|\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\|. = \frac{1}{\overline{q}\widetilde{\sigma}_{t-1}(\widetilde{\mathbf{x}}_i)}\sigma_t(\widetilde{\mathbf{x}}_i) \leq \frac{\alpha}{\overline{q}} := R.$$

For the variance term, we expand

$$\begin{aligned}
\sum_{i=1}^{t}\mathop{\mathbb{E}}_{q_{t,i}}\left[\mathbf{G}_{t,i}^2 \mid \mathcal{F}_{t-1}\right] &= \sum_{i=1}^{t}\mathop{\mathbb{E}}_{q_{t,i}}\left[\left(\frac{q_{t,i}}{\widetilde{p}_{t,i}}-1\right)^2 \,\middle|\, \mathcal{F}_{t-1}\right]\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\\
&= \sum_{i=1}^{t}\left(\mathop{\mathbb{E}}_{q_{t,i}}\left[\frac{q_{t,i}^2}{\widetilde{p}_{t,i}^2}\,\middle|\,\mathcal{F}_{t-1}\right] - \mathop{\mathbb{E}}_{q_{t,i}}\left[2\frac{q_{t,i}}{\widetilde{p}_{t,i}}\,\middle|\,\mathcal{F}_{t-1}\right]+1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\\
&= \sum_{i=1}^{t}\left(\mathop{\mathbb{E}}_{q_{t,i}}\left[\frac{q_{t,i}}{\widetilde{p}_{t,i}^2}\,\middle|\,\mathcal{F}_{t-1}\right]-1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\psi_{t,i}\psi_{t,i}^{\mathsf{T}} = \sum_{i=1}^{t}\left(\mathop{\mathbb{E}}_{q_{t,i}}\left[\frac{q_{t,i}}{\widetilde{p}_{t,i}^2}\,\middle|\,\mathcal{F}_{t-1}\right]-1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\\
&= \sum_{i=1}^{t}\left(\frac{1}{\widetilde{p}_{t,i}}-1\right)\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\psi_{t,i}\psi_{t,i}^{\mathsf{T}} \preceq \sum_{i=1}^{t}\frac{1}{\widetilde{p}_{t,i}}\|\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\|\psi_{t,i}\psi_{t,i}^{\mathsf{T}} \preceq \sum_{i=1}^{t}R\psi_{t,i}\psi_{t,i}^{\mathsf{T}}
\end{aligned}$$

where we used the fact that $q_{t,i}^2 = q_{t,i}$ and $\mathbb{E}_{q_{t,i}}[q_{t,i}|\mathcal{F}_{t-1}] = \widetilde{p}_{t,i}$. We can now bound this quantity as

$$\left\|\sum_{i=1}^{t}\mathop{\mathbb{E}}_{q_{t,i}}\left[\mathbf{G}_{t,i}^2 \mid \mathcal{F}_{t-1}\right]\right\| \leq \left\|\sum_{i=1}^{t}R\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\right\| = R\left\|\sum_{i=1}^{t}\psi_{t,i}\psi_{t,i}^{\mathsf{T}}\right\| = R\|\mathbf{\Psi}_t^{\mathsf{T}}\mathbf{\Psi}_t\| \leq R := \sigma^2.$$

Therefore, we have $\sigma^2 = R$ and $R = 1/\overline{q}$, and applying Proposition 7 and a union bound we conclude the proof.

C.4. Bounding $\Pr(E_t \cap A_{t-1}^{\complement})$

We will use the following concentration for independent Bernoulli r.v.

Proposition 8 (Calandriello et al., 2017b, App. D.4) *Let $\{q_s\}_{s=1}^t$ be independent Bernoulli random variables, each with success probability p_s, and denote their sum as $d = \sum_{s=1}^t p_s \geq 1$. Then,*[6]

$$\mathbb{P}\left(\sum_{s=1}^{t}q_s \geq 3d\right) \leq \exp\{-3d(3d-(\log(3d)+1))\} \leq \exp\{-2d\}$$

We can also rigorously define the event E_t as the event where

$$\sum_{i=1}^{t}q_{t,i} \geq 3\alpha(1+\kappa^2/\lambda)\log(t/\delta)\sum_{i=1}^{t}\sigma_t^2(\widetilde{\mathbf{x}}_i) = 3\alpha(1+\kappa^2/\lambda)d_{\text{eff}}(\lambda,\widetilde{\mathbf{X}}_t)\log(t/\delta).$$

[6] This is a simple variant of Chernoff bound where the Bernoulli random variables are not identically distributed.

Once again we will use conditioning, in particular

$$\Pr(E_t \cap A_t^{\complement}) = \mathop{\mathbb{E}}_{\mathcal{F}_{t-1}:\left\|\sum_{i=1}^t \mathbf{G}_{t-1,i}\right\|\leq\varepsilon}\left[\mathop{\mathbb{E}}_{\{q_{t,i}\}}\left[\mathbb{I}\left\{\sum_{i=1}^t q_{t,i} \geq 3\alpha(1+\kappa^2/\lambda)\log(t/\delta)\sum_{i=1}^t \sigma_t^2(\widetilde{\mathbf{x}}_i)\right\} \,\middle|\, \mathcal{F}_{t-1}\right]\right].$$

Conditioned on $\mathcal{F}_{t-1}$ the r.v. $q_{t,i}$ become independent Bernoulli with probability $\widetilde{p}_{t,i} = \overline{q}\widetilde{\sigma}_{t-1}(\widetilde{\mathbf{x}}_i)$. Since we restrict the outcomes to $A_{t-1}^{\complement}$, we can exploit Lemma 6 and the guarantees of ε accuracy to bound $\widetilde{p}_{t,i} \leq \alpha\sigma_{t-1}^2(\widetilde{\mathbf{x}}_i)$. Then we can use Proposition 4 to bound $\sigma_{t-1}^2(\widetilde{\mathbf{x}}_i) \leq (1+\kappa^2/\lambda)\sigma_t^2(\widetilde{\mathbf{x}}_i)$. Therefore, $q_{t,i}$ are conditionally independent Bernoulli with probability at most $\overline{q}(1+\kappa^2/\lambda)\sigma_t^2(\widetilde{\mathbf{x}}_i)$. Applying a simple stochastic dominance argument, and Proposition 8, we obtain our result.

Appendix D. Proof of Theorem 2

Following Abbasi-Yadkori et al. (2011), we divide the proof in two parts, bounding the approximate confidence ellipsoid, and bounding the regret.

D.1. Bounding the confidence ellipsoid

We begin by proving an intermediate result on the confidence ellipsoid

Theorem 9 *Under the same assumptions as Theorem 2 with probability at least $1-\delta$ and for all $t \geq 0$ $\mathbf{w}_*$ lies in the set*

$$\widetilde{C}_t := \left\{\mathbf{w} : \|\mathbf{w} - \widetilde{\mathbf{w}}_t\|_{\widetilde{\mathbf{A}}_t} \leq \widetilde{\beta}_t\right\}$$

with

$$\widetilde{\beta}_t := 2\xi\sqrt{\alpha\log(\kappa^2 t)\left(\sum_{s=1}^t \widetilde{\sigma}_t^2(\mathbf{x}_s)\right) + \log\left(\frac{1}{\delta}\right)} + \left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F$$

Proof For simplicity, we now omit the subscript t. We begin noticing that

$$\begin{aligned}
(\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}(\widetilde{\mathbf{w}} - \mathbf{w}_*) &= (\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}(\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{y} - \mathbf{w}_*) \\
&= (\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}(\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}(\mathbf{\Phi}\mathbf{w}_* + \eta - \mathbf{w}_*) \\
&= (\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}(\underbrace{\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}\mathbf{w}_* - \mathbf{w}_*}_{\text{bias}}) + (\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}^{1/2}\underbrace{\widetilde{\mathbf{A}}^{-1/2}\widetilde{\mathbf{\Phi}}^\mathsf{T}\eta}_{\text{variance}}.
\end{aligned}$$

Bounding the bias. We first focus on the first term, which is made complicated by the mismatch $\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}$. We have

$$\begin{aligned}
\widetilde{\mathbf{A}}(\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}\mathbf{w}_* - \mathbf{w}_*) &= \widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}\mathbf{w}_* - \widetilde{\mathbf{\Phi}}^\mathsf{T}\widetilde{\mathbf{\Phi}}\mathbf{w}_* - \lambda\mathbf{w}_* \\
&= \widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{w}_* + \widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}\mathbf{P}\mathbf{w}_* - \widetilde{\mathbf{\Phi}}^\mathsf{T}\widetilde{\mathbf{\Phi}}^{\mathbf{w}}_* - \lambda\mathbf{w}_* \\
&= \widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{w}_* - \lambda\mathbf{w}_*
\end{aligned}$$

Therefore

$$\begin{aligned}(\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}(\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}\mathbf{w}_* - \mathbf{w}_*) &= (\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{w}_* - \lambda(\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\mathbf{w}_* \\ &\leq \|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}}\left(\|\widetilde{\mathbf{A}}^{-1/2}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{w}_*\| + \lambda\|\mathbf{w}^*\|_{\widetilde{\mathbf{A}}^{-1}}\right) \\ &\leq \|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}}\left(\|\widetilde{\mathbf{A}}^{-1/2}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{w}_*\| + \tfrac{\lambda}{\sqrt{\lambda}}\|\mathbf{w}^*\|\right).\end{aligned}$$

We have

$$\begin{aligned}\|\widetilde{\mathbf{A}}^{-1/2}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{w}_*\| &\leq \|\widetilde{\mathbf{A}}^{-1/2}\widetilde{\mathbf{\Phi}}^\mathsf{T}\|\|\mathbf{\Phi}(\mathbf{I} - \mathbf{P})\|\|\mathbf{w}_*\| \\ &\leq \sqrt{\lambda_{\max}(\widetilde{\mathbf{\Phi}}\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T})}\sqrt{\lambda_{\max}(\mathbf{\Phi}(\mathbf{I} - \mathbf{P})^2\mathbf{\Phi}^\mathsf{T})}\|\mathbf{w}_*\|.\end{aligned}$$

It is easy to see that

$$\lambda_{\max}(\widetilde{\mathbf{\Phi}}\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}) = \lambda_{\max}(\widetilde{\mathbf{\Phi}}(\widetilde{\mathbf{\Phi}}^\mathsf{T}\widetilde{\mathbf{\Phi}} + \lambda\mathbf{I})^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}) \leq 1.$$

To bound the other term we use the following result from Calandriello and Rosasco (2018).

Proposition 10 *If $\mathcal{S}_t$ is ε-accurate w.r.t. $\mathbf{\Phi}_t$, then*

$$\mathbf{I} - \mathbf{P}_t \preceq \mathbf{I} - \mathbf{\Phi}_t\mathbf{S}_t(\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T}\mathbf{\Phi}_t\mathbf{S}_t + \lambda\mathbf{I})^{-1}\mathbf{S}_t^\mathsf{T}\mathbf{\Phi}_t^\mathsf{T} \preceq \frac{\lambda}{1 - \varepsilon}(\mathbf{\Phi}_t\mathbf{\Phi}_t^\mathsf{T} + \lambda\mathbf{I})^{-1}.$$

Since from Theorem 1 we have that $\mathcal{S}_t$ is ε-accurate, we have from Theorem 10

$$\mathbf{\Phi}(\mathbf{I} - \mathbf{P})^2\mathbf{\Phi}^\mathsf{T} = \mathbf{\Phi}(\mathbf{I} - \mathbf{P})\mathbf{\Phi}^\mathsf{T} \preceq \frac{\lambda}{1 - \varepsilon}\mathbf{\Phi}(\mathbf{\Phi}^\mathsf{T}\mathbf{\Phi} + \lambda\mathbf{I})^{-1}\mathbf{\Phi}^\mathsf{T} \preceq \frac{\lambda}{1 - \varepsilon}\mathbf{I}$$

Putting it all together

$$(\widetilde{\mathbf{w}} - \mathbf{w}_*)^\mathsf{T}\widetilde{\mathbf{A}}(\widetilde{\mathbf{A}}^{-1}\widetilde{\mathbf{\Phi}}^\mathsf{T}\mathbf{\Phi}\mathbf{w}_* - \mathbf{w}_*) \leq \left(1 + \frac{1}{\sqrt{1 - \varepsilon}}\right)\|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}}\sqrt{\lambda}\|\mathbf{w}_*\|$$

Bounding the variance. We will need this self-normalized martingale concentration inequality from Abbasi-Yadkori et al. (2011). It can be trivially extended to RKHSs in the case of finite sets such as our $\mathcal{A}$. Note that if the reader is interested in infinite sets, Chowdhury and Gopalan (2017) provide a generalization with slightly worst constants.

Proposition 11 (Abbasi-Yadkori et al. (2011)) *Let $\{\mathcal{F}_t\}_{t=0}^\infty$ be a filtration. Let $\{\eta_t\}_{t=1}^\infty$ be a real-valued stochastic process such that η_t is $\mathcal{F}_t$-measurable and zero-mean ξ-subgaussian. Let $\{\mathbf{\Phi}_t\}_{t=1}^\infty$ be an $\mathcal{H}$-valued stochastic process such that $\mathbf{\Phi}_t$ is $\mathcal{F}_{t-1}$-measurable. Let $\mathbf{I}$ be the identity operator on $\mathcal{H}$. For any $t \geq 1$, define*

$$\mathbf{A}_t = \mathbf{\Phi}_t^\mathsf{T}\mathbf{\Phi}_t + \lambda\mathbf{I}, \qquad \mathbf{V}_t = \mathbf{\Phi}_t^\mathsf{T}\eta_t.$$

Then, for any $\delta > 0$, with probability at least $1 - \delta$, for all $t \geq 0$,

$$\|\mathbf{V}_t\|_{\mathbf{A}_t^{-1}}^2 \leq 2\xi^2\log(\mathrm{Det}(\mathbf{A}_t/\lambda)/\delta).$$

Remembering the definition of $\alpha \geq 1$ from Theorem 1, we can reformulate

$$\begin{aligned}(\widetilde{\mathbf{w}} - \mathbf{w}_*)^\intercal \widetilde{\mathbf{A}}^{1/2} \widetilde{\mathbf{A}}^{-1/2} \widetilde{\boldsymbol{\Phi}} \eta &\leq \|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}} \|\widetilde{\boldsymbol{\Phi}} \eta\|_{\widetilde{\mathbf{A}}^{-1}} \\ &= \|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}} \|\widetilde{\boldsymbol{\Phi}}^\intercal \eta\|_{(\widetilde{\boldsymbol{\Phi}}^\intercal \widetilde{\boldsymbol{\Phi}} + \lambda \mathbf{I})^{-1}} \\ &= \|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}} \|\widetilde{\boldsymbol{\Phi}}^\intercal \eta / \lambda\|_{(\widetilde{\boldsymbol{\Phi}}^\intercal \widetilde{\boldsymbol{\Phi}}/\lambda + \mathbf{I})^{-1}}\end{aligned}$$

We need now to make a remark that requires temporal notation. Note that we cannot directly apply Theorem 11 to $\widetilde{\boldsymbol{\Phi}}_t \eta_t = \mathbf{P}_t \boldsymbol{\Phi}_t \eta_t$. In particular, for $s < t$ we have that $\widetilde{\boldsymbol{\Phi}}_s \eta_s = \mathbf{P}_t \boldsymbol{\Phi}_s \eta_s$ is not $\mathcal{F}_{s-1}$ measurable, since $\mathbf{P}_t$ depends on all randomness up to time t. However, since $\mathbf{P}_t$ is always a projection matrix we know that the variance of the projected process is bounded by the variance of the original process

$$\begin{aligned}\|\widetilde{\boldsymbol{\Phi}}^\intercal \eta/\lambda\|_{(\widetilde{\boldsymbol{\Phi}}^\intercal \widetilde{\boldsymbol{\Phi}}/\lambda + \mathbf{I})^{-1}} &= \sqrt{\eta^\intercal \widetilde{\boldsymbol{\Phi}} (\widetilde{\boldsymbol{\Phi}}^\intercal \widetilde{\boldsymbol{\Phi}}/\lambda + \mathbf{I})^{-1} \widetilde{\boldsymbol{\Phi}}^\intercal \eta / \lambda} = \sqrt{\eta^\intercal \widetilde{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\Phi}}^\intercal (\widetilde{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\Phi}}^\intercal/\lambda + \mathbf{I})^{-1} \eta / \lambda} \\ &\overset{(a)}{=} \sqrt{\eta^\intercal (\mathbf{I} - \lambda(\widetilde{\boldsymbol{\Phi}} \widetilde{\boldsymbol{\Phi}}^\intercal/\lambda + \mathbf{I})^{-1}) \eta/\lambda} = \sqrt{\eta^\intercal (\mathbf{I} - \lambda(\boldsymbol{\Phi} \mathbf{P} \boldsymbol{\Phi}^\intercal/\lambda + \mathbf{I})^{-1}) \eta/\lambda} \\ &\overset{(b)}{\leq} \sqrt{\eta^\intercal (\mathbf{I} - \lambda(\boldsymbol{\Phi} \boldsymbol{\Phi}^\intercal/\lambda + \mathbf{I})^{-1}) \eta/\lambda} \overset{(c)}{=} \|\boldsymbol{\Phi}^\intercal \eta/\lambda\|_{(\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi}/\lambda + \mathbf{I})^{-1}},\end{aligned}$$

where in (a) we added and subtracted $\lambda\mathbf{I}$ from $\widetilde{\boldsymbol{\Phi}}\widetilde{\boldsymbol{\Phi}}^\intercal$, in (b) we used the fact that $\|\mathbf{P}\| \leq 1$ for all projection matrices, and in (c) we reversed the reformulation from (a). We can finally use Theorem 11

$$\begin{aligned}\|\boldsymbol{\Phi}^\intercal \eta/\lambda\|_{(\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi}/\lambda + \mathbf{I})^{-1}} &\leq \sqrt{2\xi^2 \log(\mathrm{Det}(\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi}/\lambda + \mathbf{I})/\delta)} \\ &= \sqrt{2\xi^2 \log(\mathrm{Det}(\mathbf{A}/\lambda)/\delta)}.\end{aligned}$$

While this is now a valid bound on the radius of the confidence interval, it is still not satisfactory. In particular, we can use Sylvester's identity to reformulate

$$\log\det(\mathbf{A}/\lambda) = \log\det(\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi}/\lambda + \mathbf{I}) = \log\det(\boldsymbol{\Phi}\boldsymbol{\Phi}^\intercal/\lambda + \mathbf{I}) = \log\det(\mathbf{K}/\lambda + \mathbf{I}).$$

Computing the radius would require constructing the matrix $\mathbf{K} \in \mathbb{R}^{t \times t}$ and is too expensive. Instead, we obtain a cheap but still small upper bound as follow

$$\begin{aligned}\log\det(\mathbf{K}_t/\lambda + \mathbf{I}) &\leq \mathrm{Tr}(\mathbf{K}_t(\mathbf{K}_t + \lambda\mathbf{I})^{-1})(1 + \log(\|\mathbf{K}_t\| + 1)) \\ &\leq \mathrm{Tr}(\mathbf{K}_t(\mathbf{K}_t + \lambda\mathbf{I})^{-1})(1 + \log(\mathrm{Tr}\,\mathbf{K}_t + 1)) \\ &\leq \mathrm{Tr}(\mathbf{K}_t(\mathbf{K}_t + \lambda\mathbf{I})^{-1})(1 + \log(\kappa^2 t + 1)) \\ &= (1 + \log(\kappa^2 t + 1)) \sum_{s=1}^{t} \sigma_t^2(\mathbf{x}_s) \\ &\leq \alpha(1 + \log(\kappa^2 t + 1)) \sum_{s=1}^{t} \widetilde{\sigma}_t^2(\mathbf{x}_s) \\ &\leq 2\alpha \log(\kappa^2 t) \sum_{s=1}^{t} \widetilde{\sigma}_t^2(\mathbf{x}_s),\end{aligned}$$

where $\widetilde{\sigma}_t^2(\mathbf{x}_s)$ can be computed efficiently, and is already done so by the algorithm at every step. Putting it all together

$$\begin{aligned}\|\widetilde{\mathbf{w}} - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}} &\leq 2\xi\sqrt{\alpha \log(\kappa^2 t)\left(\sum_{s=1}^{t} \widetilde{\sigma}_t^2(\mathbf{x}_s)\right) + \log(1/\delta)} + \left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}\|\mathbf{w}_*\| \\ &\leq 2\xi\sqrt{\alpha \log(\kappa^2 t)\left(\sum_{s=1}^{t} \widetilde{\sigma}_t^2(\mathbf{x}_s)\right) + \log(1/\delta)} + \left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F := \widetilde{\beta}_t\end{aligned}$$

■

D.2. Bounding the regret

The regret analysis is straightforward. Assume that $\mathbf{w}_* \in \widetilde{C}_t$ is satisfied (i.e. the event from Theorem 9 holds) and remember that by the definition $\boldsymbol{\phi}_t = \arg\max_{\mathbf{x}_i in \mathcal{A}} \max_{\mathbf{w}\in\widetilde{\mathbf{C}}_t} \boldsymbol{\phi}_i^{\mathsf{T}}\mathbf{w}$. We can also define $\overline{\mathbf{w}}_{t,i} = \arg\max_{\mathbf{w}\in\widetilde{\mathbf{C}}_t} \boldsymbol{\phi}_i^{\mathsf{T}}\mathbf{w}$ as the auxiliary vector which encodes the optimistic behaviour of the algorithm. With a slight abuse of notation we will also use $*$ as a subscript to indicate the (unknown) index of the optimal arm, so that $\overline{\mathbf{w}}_{t,*} = \arg\max_{\mathbf{w}\in\widetilde{\mathbf{C}}_t} \boldsymbol{\phi}_*^{\mathsf{T}}\mathbf{w}$. Since $\mathbf{w}_* \in \widetilde{C}_t$

$$\boldsymbol{\phi}_t^{\mathsf{T}}\overline{\mathbf{w}}_{t,t} \geq \boldsymbol{\phi}_*\overline{\mathbf{w}}_{t,*} \geq \boldsymbol{\phi}_*\mathbf{w}_*.$$

We can now bound the instantaneous regret r_t as

$$\begin{aligned}r_t &= \boldsymbol{\phi}_*^{\mathsf{T}}\mathbf{w}_* - \boldsymbol{\phi}_t^{\mathsf{T}}\mathbf{w}_* \leq \boldsymbol{\phi}_t^{\mathsf{T}}\overline{\mathbf{w}}_{t,t} - \boldsymbol{\phi}_t^{\mathsf{T}}\mathbf{w}_* \\ &= \boldsymbol{\phi}_t^{\mathsf{T}}(\overline{\mathbf{w}}_{t,t} - \widehat{\mathbf{w}}_t) + \boldsymbol{\phi}_t^{\mathsf{T}}(\widehat{\mathbf{w}}_t - \mathbf{w}_*) \\ &= \boldsymbol{\phi}_t^{\mathsf{T}}\widetilde{\mathbf{A}}_t^{-1/2}\widetilde{\mathbf{A}}_t^{1/2}(\overline{\mathbf{w}}_{t,t} - \widehat{\mathbf{w}}_t) + \boldsymbol{\phi}_t^{\mathsf{T}}\widetilde{\mathbf{A}}_{t-1}^{-1/2}\widetilde{\mathbf{A}}_t^{1/2}(\widehat{\mathbf{w}}_t - \mathbf{w}_*) \\ &\leq \sqrt{\boldsymbol{\phi}_t^{\mathsf{T}}\widetilde{\mathbf{A}}_t^{-1}\boldsymbol{\phi}_t}(\|\overline{\mathbf{w}}_{t,t} - \widehat{\mathbf{w}}_t\|_{\widetilde{\mathbf{A}}_t} + \|\widehat{\mathbf{w}}_t - \mathbf{w}_*\|_{\widetilde{\mathbf{A}}_t}) \\ &\leq 2\widetilde{\beta}_t\sqrt{\boldsymbol{\phi}_t^{\mathsf{T}}\widetilde{\mathbf{A}}_t^{-1}\boldsymbol{\phi}_t}\end{aligned}$$

Summing over t and taking the max over $\widetilde{\beta}_t$

$$R_t \leq 2\widetilde{\beta}_T\sum_{t=1}^{T}\sqrt{\boldsymbol{\phi}_t^{\mathsf{T}}\widetilde{\mathbf{A}}_t^{-1}\boldsymbol{\phi}_t} \leq 2\widetilde{\beta}_T\sqrt{T}\sqrt{\sum_{t=1}^{T}\boldsymbol{\phi}_t^{\mathsf{T}}\widetilde{\mathbf{A}}_t^{-1}\boldsymbol{\phi}_t} \leq 2\widetilde{\beta}_T\sqrt{T}\sqrt{\alpha\sum_{t=1}^{T}\boldsymbol{\phi}_t^{\mathsf{T}}\mathbf{A}_t^{-1}\boldsymbol{\phi}_t}$$

We can now use once again Proposition 5 to obtain

$$R_T \leq 2\widetilde{\beta}_T\sqrt{\alpha T\sum_{t=1}^{T}\boldsymbol{\phi}_t^{\mathsf{T}}\mathbf{A}_t^{-1}\boldsymbol{\phi}_t} = 2\widetilde{\beta}_T\sqrt{\alpha T\sum_{t=1}^{T}\sigma_t^2(\widetilde{\mathbf{x}}_t)} \leq 2\widetilde{\beta}_T\sqrt{2\alpha T d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T)}.$$

We can also further upper bound $\widetilde{\beta}_T$ as

$$\begin{aligned}
\widetilde{\beta}_T &= 2\xi\sqrt{\alpha\log(\kappa^2 T)\left(\sum_{s=1}^{T}\widetilde{\sigma}_t^2(\mathbf{x}_s)\right) + \log(1/\delta)} + \left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F\\
&\leq 2\xi\sqrt{\alpha^2\log(\kappa^2 T)\left(\sum_{s=1}^{T}\sigma_t^2(\mathbf{x}_s)\right) + \log(1/\delta)} + \left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F\\
&\leq 2\xi\alpha\sqrt{d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T) + \log(1/\delta)} + \left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F
\end{aligned}$$

Putting it together we obtain

$$\begin{aligned}
R_T \leq& 2\left(2\xi\alpha\sqrt{d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T) + \log(1/\delta)}\right)\sqrt{2\alpha T d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T)}\\
&+ 2\left(\left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\sqrt{\lambda}F\right)\sqrt{2\alpha T d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T)}\\
\leq& 2\xi(2\alpha)^{3/2}\left(d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T) + \log(1/\delta)\right)\\
&+ 2\left(2\sqrt{\alpha}\sqrt{\lambda}F\right)\sqrt{2\alpha T d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T)}\\
\leq& 2(2\alpha)^{3/2}\left(\sqrt{T}\xi d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T) + \sqrt{T}\log(1/\delta) + \sqrt{T\lambda F^2 d_{\text{eff}}(\lambda, \widetilde{\mathbf{X}}_T)\log(\kappa^2 T)}\right)
\end{aligned}$$