Generalized Kernelized Bandits: Self-Normalized Bernstein-Like Dimension-Free Inequality and Regret Bounds

Alberto Maria Metelli
Politecnico di Milano
albertomaria.metelli@polimi.it

Simone Drago
Politecnico di Milano
simone.drago@polimi.it

Marco Mussi
Politecnico di Milano
marco.mussi@polimi.it

Abstract

We study the regret minimization problem in the novel setting of *generalized kernelized bandits* (GKBs), where we optimize an unknown function f^* belonging to a *reproducing kernel Hilbert space* (RKHS) having access to samples generated by an *exponential family* (EF) noise model whose mean is a non-linear function $\mu(f^*)$. This model extends both *kernelized bandits* (KBs) and *generalized linear bandits* (GLBs). We propose an optimistic algorithm, GKB-UCB, and we explain why existing self-normalized concentration inequalities do not allow to provide tight regret guarantees. For this reason, we devise a novel self-normalized Bernstein-like dimension-free inequality resorting to Freedman's inequality and a stitching argument, which represents a contribution of independent interest. Based on it, we conduct a regret analysis of GKB-UCB, deriving a regret bound of order $\widetilde{O}(\gamma_T \sqrt{T/\kappa_*})$, being T the learning horizon, γ_T the maximal information gain, and κ_* a term characterizing the magnitude the reward nonlinearity. Our result matches, up to multiplicative constants and logarithmic terms, the state-of-the-art bounds for both KBs and GLBs and provides a *unified view* of both settings.

1 Introduction

Multi-Armed Bandits [MABs, 15] have been extensively studied and extended over the years. One key research direction involves expanding the MAB framework to continuous action spaces. Doing this requires introducing some notion of similarity or structure in the expected rewards relative to the distance between arms. Without such structure, information gathered from explored actions/arms cannot be transferred to unexplored ones, making learning infeasible [4]. The most known and studied structure over the arms is the *linear* one, and led to the design of *linear bandits* [LBs, 1, 6]. In LBs, the expected reward is modeled as the inner product between the action and an unknown parameter vector (i.e., $\mathbb{E}[y_t|\mathbf{x}_t; \boldsymbol{\theta}^*] = \langle \mathbf{x}_t, \boldsymbol{\theta}^* \rangle$). This setting strictly generalizes the finite-arms MABs [15, 23] that can be retrieved considering arms as in an $\mathbb{R}^d$ canonical basis.

LBs, in turn, have been extended in parallel in two directions: *generalized linear bandits* [GLBs, 10] and *kernelized bandits* [KBs, 5, 29]. On the one hand, GLBs employ a *generalized linear model* [GLM, 19] to allow for the representation of different noise models (including Gaussian and Bernoulli). This is achieved with the use of a real-valued non-linear *inverse link function* $\mu(\cdot)$, such that the expected payoff is defined as $\mathbb{E}[y_t|\mathbf{x}_t; \boldsymbol{\theta}^*] = \mu(\langle \mathbf{x}_t, \boldsymbol{\theta}^* \rangle)$. On the other hand, KBs focus on the optimization of an unknown expected reward function belonging to a *reproducing kernel Hilbert*

Preprint. Under review.

space (RKHS) induced by a known kernel function $k(\mathbf{x}, \mathbf{x}')$, often resorting to Gaussian processes for designing algorithms [22]. We observe that GLBs fall back to LBs when the identity link function $\mu = I$ is considered, and KBs fall back to LBs when a linear kernel $k(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$ is considered.

In this work, we propose the novel *generalized kernelized bandit* (GKB) setting, which unifies GLBs and KBs (Figure 1). This setting enables learning in the scenarios in which the unknown function f^* comes from an RKHS and the samples come from an exponential family model whose mean is obtained by applying an inverse link function μ to function f^*. This allows accounting for a variety of noise models, including Gaussian and Bernoulli [3].

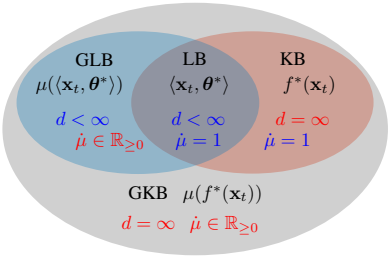

Figure 1: Inclusion of the settings ($f(\cdot)$ is assumed to belong to a RKHS).

As established by the literature [1, 9, 17], when designing *optimistic* regret minimization algorithms for either GLBs and KBs, a fundamental technical tool are *self-normalized* concentration inequalities [7]. When targeting regret minimization in the novel setting of GKBs, it is necessary to employ a concentration inequality that combines the requirements of GLBs and KBs, i.e., it should avoid dependencies on the minimum slope $\dot{\mu}$ of the inverse link function (as in GLBs) and on the dimensionality of the feature representation (as in KBs). The seminal work [1] provides a self-normalized concentration inequality for least square estimators under subgaussian noise, exploiting theoretical advancements in self-normalized processes and pseudo-maximization of [7, 8]. However, this inequality does not conveniently manage the case in which the samples come from an exponential family model where the variances depend on inverse link function μ, ultimately leading to a dependence on its minimum slope. To cope with this issue, [9] derive a concentration inequality via a pseudo-maximization technique that results in a tight regret bound for GLBs, accounting for the heteroscedastic characteristics of the noise (i.e., Bernstein-like). However, their concentration inequality presents a dependency on the dimensionality of the feature vector (i.e., dimension-dependent). While not being problematic for GLBs, this hinders a direct application to GKBs, where the feature representation (induced by the kernel function) can be infinite-dimensional. Additionally, [5] design a self-normalized bound for martingales which provides tight concentration results for the KB setting, directly operating with kernels. However, this result can be considered the counterpart of [1] in the dual (kernel) space and, for this reason, it shares the same limitation when using an inverse link function, generating a dependence on the minimum value of $\dot{\mu}$ when applied to GKBs.[1] It appears now necessary to derive a novel concentration result that is both dimension-free and Bernstein-like to properly address the GKB setting.

Outline and Contributions. We start by introducing the setting of the GKBs, the assumptions, and the learning problem (Section 3). Then, we design `GKB-UCB`, an optimistic regret minimization algorithm (Section 4) and we introduced some preliminary results (Section 5). The key contributions of this work are contained in Sections 6 and 7. In Section 6, we discuss more formally the limitations of the existing inequalities and derive a novel self-normalized Bernstein-like dimension-free inequality via the application of Freedman's inequality together with a stitching argument. In Section 7, we analyze the `GKB-UCB` with a confidence set defined in terms of the previously derived inequality and show that it achieves regret of order $\widetilde{O}(\gamma_T \sqrt{T/\kappa_*})$, being T the learning horizon, γ_T the maximal information gain, and κ_* a term characterizing the slope of the inverse link function in the optimal decision (an efficient implementation is reported in Appendix A). This result matches the state-of-the-art of both GLBs and KBs up to multiplicative constants and logarithmic terms.

2 Preliminaries

Notation. Let $a, b \in \mathbb{N}$ with $a \leq b$, we denote with $[\![a, b]\!] \coloneqq \{a, a+1, \dots, b\}$ and with $[\![b]\!] \coloneqq [\![1, b]\!]$. Let $d \in \mathbb{N}$, $\mathbf{I}_d$ denotes the identity matrix of order d and $\mathbf{0}_d$ the column vector of all zeros of size d (d is omitted when clear from the context). $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multi-variate Gaussian distribution.

Reproducing Kernel Hilbert Space. Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a decision set and $\mathcal{H}$ be a Hilbert space endowed with the inner product $\langle \cdot, \cdot \rangle$ (and induced norm $\| \cdot \|$). $\mathcal{H}$ is a *reproducing kernel Hilbert*

[1] We refer to Table 1 for an overview of the properties of concentration inequalities present in the literature.

Self-normalized Concentrations	Properties				
	Condition	Dim-free	Empirical	Heterosc.	Technique
Dani et al., 2008 [6]	Hoeffding	✗	✗	✗	Freedman
Abbasi-Yadkori et al., 2011 [1]	Hoeffding	✓	✓	✗	Pseudo-Max
Chowdhury and Gopalan, 2017 [5]	Hoeffding	✓	✓	✗	Pseudo-Max
Faury et al., 2020 [9]	Bernstein	✗	✓	✓	Pseudo-Max
Zhou et al., 2021 [36]	Bernstein	✗	✗	✗	Freedman
Ziemann, 2024 [37]	Bernstein	✗	✓	✗	PAC-Bayes
Our work	**Bernstein**	✓	✓	✓	**Freedman**

Table 1: Summary of the properties of self-normalized concentrations.

space [28] if there exists a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, called *kernel*, such that it satisfies the *reproducing property*, i.e., for every function $f \in \mathcal{H}$ it holds that $f(\mathbf{x}) = \langle f, k(\mathbf{x}, \cdot) \rangle$ for every $\mathbf{x} \in \mathcal{X}$. It follows that the kernel k is symmetric and satisfies the conditions for positive semi-definiteness. We denote with I the identity operator on $\mathcal{H}$. From Mercer's theorem [20, 14], there exists a (possibly infinite-dimensional) feature mapping $\phi : \mathcal{X} \to \mathbb{R}^{\mathbb{N}}$ such that for every function $f \in \mathcal{H}$ there exists a (possibly infinite-dimensional) vector of coefficients $\alpha \in \mathbb{R}^{\mathbb{N}}$ such that for every $\mathbf{x} \in \mathcal{X}$, we have $f(\mathbf{x}) = \sum_{i \in \mathbb{N}} \alpha_i \phi_i(\mathbf{x}) = \langle \alpha, \phi(\mathbf{x}) \rangle$, where α depends on f but not on $\mathbf{x}$ and for every $i \in \mathbb{N}$, we have that $\phi_i : \mathcal{X} \to \mathbb{R}$ depends on $\mathbf{x}$ but not on f and the series converges absolutely and uniformly for almost all $\mathbf{x}$. Moreover, for every $i, j \in \mathbb{N}$ with $i \neq j$, we have $\|\phi_i\| = \langle \phi_i, \phi_i \rangle = 1$ and $\langle \phi_i, \phi_j \rangle = 0$, i.e., $(\phi_i)_{i \in \mathbb{N}}$ forms an orthonormal basis. Thus, if $f = \langle \alpha, \phi(\mathbf{x}) \rangle$, we have $\|f\| = \|\alpha\|$. Furthermore, for every $\mathbf{x} \in \mathcal{X}$, we have that $|f(\mathbf{x})| \leq \|f\| \|k(\cdot, \mathbf{x})\| = \|f\| \sqrt{k(\mathbf{x}, \mathbf{x})}$.

Information Gain. Let k be a kernel, let $t \in \mathbb{N}$, and let $\mathbf{x}_1, \dots, \mathbf{x}_t \in \mathcal{X}$ be a sequence of decisions, the *information gain* Γ_t and the *maximal information gain* γ_t are defined, respectively as [29]: $\Gamma_t \coloneqq \frac{1}{2} \log \det(\mathbf{I} + \lambda^{-1} \mathbf{K}_t)$ and $\gamma_t \coloneqq \max_{\mathbf{x}_1, \dots, \mathbf{x}_t \in \mathcal{X}} \Gamma_t$, where $\lambda > 0$ and $\mathbf{K}_t \in \mathbb{R}^{(t-1) \times (t-1)}$ is the Kernel matrix $(\mathbf{K}_t)_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ for $i, j \in [\![t-1]\!]$. Γ_t is the *mutual information* between the random vectors $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \nu^2 \mathbf{K}_t)$ and $\mathbf{y}_t = \mathbf{f}_t + \boldsymbol{\epsilon}_t$ where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_t, v^2 \lambda \mathbf{I}_t)$, for arbitrary $v > 0$. We use the abbreviation $\mathbf{K}_t(\lambda) \coloneqq \lambda \mathbf{I} + \mathbf{K}_t$, so that, $\Gamma_t \coloneqq \frac{1}{2} \log \det(\lambda^{-1} \mathbf{K}_t(\lambda))$.[2]

Covariance Operators. Let $\mathcal{H}$ be a RKHS with kernel k inducing the feature mapping ϕ, let $t \in \mathbb{N}$ and $\mathbf{x}_1, \dots, \mathbf{x}_t \in \mathcal{X}$ be a sequence of decisions, the *covariance operator* is defined as: $V_t(\lambda) \coloneqq V_t + \lambda I = \sum_{s=1}^{t-1} \phi(\mathbf{x}_s) \phi(\mathbf{x}_s)^\top + \lambda I$. The following identity was shown in [32]:

$$\det(\lambda^{-1} V_t(\lambda)) = \det(\lambda^{-1} \mathbf{K}_t(\lambda)). \tag{1}$$

Canonical Exponential Family Models. Let $f : \mathcal{X} \to \mathbb{R}$, a real-valued random variable y belongs to the *canonical exponential family* [EF, 3] if it has density:

$$p(y|\mathbf{x}; f) = \exp\left(\frac{y f(\mathbf{x}) - m(f(\mathbf{x}))}{\mathfrak{g}(\tau)} + h(y, \tau) \right), \tag{2}$$

where $\tau > 0$ is a temperature parameter and $\mathfrak{g}, m : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ are suitably defined functions [17]. This EF model allows representing a variety of distributions, including Gaussian, Bernoulli, exponentials, and Poisson. Function m is called *log-partition function* and fulfills the following assumptions. As customary [17, 25], m is assumed to be three times differentiable and convex. We define the *inverse link function* $\mu = m'$, that, since m is convex, is monotonically non-decreasing. Thus, the following hold [17]: $\mathbb{E}[y|\mathbf{x}; f] = m'(f(\mathbf{x})) = \mu(f(\mathbf{x}))$ and $\mathbb{V}\mathrm{ar}[y|\mathbf{x}; f] = \mathfrak{g}(\tau) \dot{\mu}(f(\mathbf{x}))$. When f is a linear function, the model in Equation (2) is also called *generalized linear model* [GLM, 19]. We also define the maximum slope of μ, i.e., $R_{\dot{\mu}} \coloneqq \sup_{f \in \mathcal{H}, \mathbf{x} \in \mathcal{X}} \dot{\mu}(f(\mathbf{x}))$.

3 Problem Formulation

We define the novel *generalized kernelized bandit* (GKB) setting and the learning problem.

Setting. Let $f^* \in \mathcal{H}$ be an unknown function belonging to the RKHS $\mathcal{H}$. At every round $t \in [\![T]\!]$, being $T \in \mathbb{N}$ the learning horizon, the learner chooses a decision $\mathbf{x}_t \in \mathcal{X}$ by means of a policy

[2] Known bounds for γ_t are available for commonly used kernels [29, 31].

$\pi_t : \mathcal{F}_{t-1} \to \mathcal{X}$, being $\mathcal{F}_{t-1} = \sigma(\mathbf{x}_1, y_1, \ldots, \mathbf{x}_{t-1}, y_{t-1})$ the filtration of all random variables realized so far, and observes a reward $y_t \sim p(\cdot|\mathbf{x}_t; f^*)$. The goal of the agent is to find a decision $\mathbf{x}^* \in \mathcal{X}$ maximizing the expected reward: $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathcal{X}} \mu(f^*(\mathbf{x}))$. Since μ is monotonically non-decreasing, maximizing $\mu(f^*(\cdot))$ is equivalent to maximizing $f^*(\cdot)$. It is worth noting that the GKB generalizes two well-known settings: (i) *generalized linear bandits* [GLBs, 18] when the kernel is linear $k(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$ and (ii) *kernelized bandits* [KBs, 5] when the inverse link function is the identity function, i.e., $\mu = I$.

Learning Problem. We evaluate the performance of a learner, i.e., a $\pi = (\pi_t)_{t\in[\![T]\!]}$, with *cumulative regret*: $R(\pi, T) \coloneqq \sum_{t\in[\![T]\!]} (\mu(f^*(\mathbf{x}^*)) - \mu(f^*(\mathbf{x}_t)))$, where $\mathbf{x}_t = \pi_t(\mathcal{F}_{t-1})$ for all $t \in [\![T]\!]$.

Assumptions. We make the following assumptions about function f^* and the RKHS $\mathcal{H}$.

Assumption 3.1 (Bounded Norm)**.** *It exists a known constant $B < +\infty$ such that $\|f^*\| \leq B$.*

Assumption 3.2 (Bounded Kernel)**.** *It exists a known constant $K < +\infty$ such that $\sup_{\mathbf{x}\in\mathcal{X}} k(\mathbf{x},\mathbf{x}) \leq K^2$.*

Assumptions 3.1 and 3.2 are widely employed in the KB literature [5], where, in particular, Assumption 3.2 is enforced with $K = 1$ and it is fulfilled by commonly used kernels (e.g., Gaussian and Matérn kernels). Assumptions 3.1 and 3.2 are the analogous in GLBs of requiring the boundedness of the parameter vector (since if $f = \langle \alpha, \phi \rangle$, then, $\|f\| = \|\alpha\|$) and requiring the boundedness of the norm of the decisions (since when $k(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$ we have that $k(\mathbf{x}, \mathbf{x}) = \|\mathbf{x}\|^2$), respectively [2, 17]. The combination of the two allows bounding the L_∞-norm of f^* as $\|f^*\|_\infty \leq BK$.

Concerning the EF noise model, we make the following assumptions.

Assumption 3.3 (Bounded noise)**.** *Let $\mathbf{x} \in \mathcal{X}$, $y \sim p(\cdot|\mathbf{x}; f^*)$, let $\epsilon = y - \mu(f^*(\mathbf{x}))$. There exists a known constant $R < +\infty$ such that $|\epsilon| \leq R$ almost surely.*

This assumption is widely used in the GLB literature [2, 25]. If we deal with ν^2-subgaussian noise (instead of bounded), we can take $R = \nu\sqrt{2\log(2T/\delta)}$ to ensure that $|\epsilon_t| \leq R$ uniformly for $t \in [\![T]\!]$ w.p. $1 - \delta$.[3] Finally, we introduce the *generalized self-concordance* property [24].

Assumption 3.4 ((Generalized) Self-concordance)**.** *There exists a known constant $R_s < +\infty$ such that for every function $f \in \mathcal{H}$ and decision $\mathbf{x} \in \mathcal{X}$, it holds that $|\ddot{\mu}(f(\mathbf{x}))| \leq R_s \dot{\mu}(f(\mathbf{x}))$.*

In [25], the authors show (Lemma 2.1) that if the EF model generates random variables that are bounded by $|y| \leq Y$ a.s., Assumption 3.4 hold with $R_s = Y$. Moreover, it holds for Bernoulli noise with $R_s = 1$ and Gaussian with $R_s = 0$ [17].

Problem Characterization. We define the following characterizing the difficulty of the problem: $\kappa_* = \frac{1}{\dot{\mu}(f^*(\mathbf{x}^*))}$ and $\kappa_{\mathcal{X}} = \sup_{\mathbf{x}\in\mathcal{X}} \frac{1}{\dot{\mu}(f^*(\mathbf{x}))}$. We have that $\kappa_* \leq \kappa_{\mathcal{X}}$. Our goal is to devise algorithms for which the dominating term in the regret bound depends on κ_* only.

4 Algorithm

In this section, we introduce `Generalized Kernelized Bandits-Upper Confidence Bounds` (`GKB-UCB`), a regret minimization optimistic algorithm for the GKB setting (Algorithm 1). `GKB-UCB` is composed of two steps: *maximum likelihood* (ML) estimation and *optimistic decision selection*. We provide a computationally tractable version in Appendix A.

Input: Decision set $\mathcal{X}$, confidence sets $\mathcal{C}_t(\delta)$
for $t \in [\![T]\!]$ **do**
  `//Maximum Likelihood Estimate`
  $\widehat{f}_t \in \arg\min_{f\in\mathcal{H}} \mathcal{L}_t(f)$ (Equation 3)
  `//Optimistic Decision Selection`
  $(\widetilde{f}_t, \mathbf{x}_t) \in \arg\max_{f\in\mathcal{C}_t(\delta),\, \mathbf{x}\in\mathcal{X}} \mu(f(\mathbf{x}))$ (Equation 6)
  Play $\mathbf{x}_t$ and observe y_t
end

Algorithm 1: `GKB-UCB`.

Maximum Likelihood Estimate. At each round $t \in [\![T]\!]$, we employ the samples collected so far $\{(\mathbf{x}_s, y_s)\}_{s\in[\![t-1]\!]}$, to obtain an estimate $\widehat{f}_t$ of f^*. Starting from the EF model, we minimize

[3] This will result in an additional logarithmic term in the final regret bound only.

the *Ridge-regularized log-likelihood*:

$$\mathcal{L}_t(f) \coloneqq \sum_{s=1}^{t-1} \frac{-y_s f(\mathbf{x}_s) + m(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)} + \frac{\lambda}{2}\|f\|^2, \quad \forall f \in \mathcal{H},\, t \in [\![T]\!], \tag{3}$$

where $\lambda \geq 0$ is the Ridge regularization parameter. The ML estimate is denoted as $\widehat{f}_t \in \arg\min_{f\in\mathcal{H}} \mathcal{L}_t(f)$. Since, for Mercer's theorem, when $f \in \mathcal{H}$, we can write $f = \langle \alpha, \phi \rangle$ with a fixed feature function ϕ, with little abuse of notation, we can look at $\mathcal{L}_t$ as a function of the parameters α, i.e., $\mathcal{L}_t(\alpha) \equiv \mathcal{L}_t(f)$. With this in mind, we introduce the operator $g_t(f) \in \mathbb{R}^{\mathbb{N}}$ related to the gradient of the loss $\mathcal{L}_t(f)$ w.r.t. the parameters α and the *weighted covariance operator* $\widetilde{V}_t(\lambda; f) \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ corresponding to the Hessian of the loss $\mathcal{L}_t(f)$ w.r.t. parameters α:

$$g_t(f) \coloneqq \sum_{s=1}^{t-1} \frac{\mu(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)}\phi(\mathbf{x}_s) + \lambda\alpha, \qquad \nabla\mathcal{L}_t(f) = -\sum_{s=1}^{t-1} \frac{y_s\phi(\mathbf{x}_s)}{\mathfrak{g}(\tau)} + g_t(f), \tag{4}$$

$$\widetilde{V}_t(\lambda; f) \coloneqq \nabla^2\mathcal{L}_t(f) = \widetilde{V}_t(f) + \lambda I = \sum_{s=1}^{t-1} \frac{\dot{\mu}(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)}\phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top + \lambda I. \tag{5}$$

The loss function $\mathcal{L}_t$ and the operators g_t and $\widetilde{V}_t$ defined above reduce to the ones employed for GLBs under the assumption that the kernel k is the linear one [2, 9, 17]. Furthermore, if $\mu = I$, we have that $\widetilde{V}_t(\lambda; f) = V_t(\lambda)$, i.e., the covariance operator.

Optimistic Decision Selection. Once the ML function $\widehat{f}_t$ is computed, the algorithm chooses an optimistic function $\widetilde{f}_t \in \mathcal{H}$ in a suitable confidence set $\mathcal{C}_t(\delta)$, together with the optimistic choice $\mathbf{x}_t$:

$$(\widetilde{f}_t, \mathbf{x}_t) \in \underset{f\in\mathcal{C}_t(\delta),\, \mathbf{x}\in\mathcal{X}}{\arg\max}\ \mu(f(\mathbf{x})). \tag{6}$$

It is worth noting that since μ is non-decreasing, we can ignore μ in the maximization. We will consider a confidence set, defined for every round $t \in [\![T]\!]$ and confidence $\delta \in (0, 1)$ as follows:[4]

$$\mathcal{C}_t(\delta) = \left\{ f \in \mathcal{H} \,:\, \left\| g_t(f) - g_t(\widehat{f}_t) \right\|_{\widetilde{V}_t^{-1}(\lambda; f)} \leq B_t(\delta; f) \right\}, \tag{7}$$

where the confidence ratio $B_t(\delta; f)$ will be specified later with the goal of guaranteeing optimism, i.e., that the true unknown function f^* belongs to $\mathcal{C}_t(\delta)$ in high probability, and limiting the regret.

5 Weighted Kernel

We discuss how the combination between a function $f \in \mathcal{H}$ with an inverse link function μ induced another RKHS space that can be characterized by its *weighted kernel*. Let $f \in \mathcal{H}$, we define the weighted feature mapping (now dependent on f) for every $\mathbf{x} \in \mathcal{X}$ as: $\widetilde{\phi}(\mathbf{x}; f) = \sqrt{\dot{\mu}(f(\mathbf{x}))\mathfrak{g}(\tau)^{-1}}\phi(\mathbf{x})$. In the primal (feature) space, this allows looking at the weighted covariance operator $\widetilde{V}_t(\lambda; f)$ as the covariance operator induced by the feature mapping $\widetilde{\phi}(\cdot; f)$, i.e., $\widetilde{V}_t(\lambda; f) = \sum_{s=1}^{t-1} \widetilde{\phi}(\mathbf{x}_s; f)\widetilde{\phi}(\mathbf{x}_s; f)^\top + \lambda I$. Passing to the dual (kernel) space, we define the weighted kernel as:

$$\widetilde{k}(\mathbf{x}, \mathbf{x}'; f) \coloneqq \langle \widetilde{\phi}(\mathbf{x}; f), \widetilde{\phi}(\mathbf{x}'; f) \rangle = \mathfrak{g}(\tau)^{-1}\sqrt{\dot{\mu}(f(\mathbf{x}))}k(\mathbf{x}, \mathbf{x}')\sqrt{\dot{\mu}(f(\mathbf{x}'))}, \quad \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}. \tag{8}$$

This is, in all regards, a valid kernel since it is obtained starting from a valid kernel and performing a legal transformation [28]. This way, we can define the weighted kernel matrix as $\widetilde{\mathbf{K}}_t(\lambda; f) = \lambda\mathbf{I} + \widetilde{\mathbf{K}}_t(f)$, where $\widetilde{\mathbf{K}}_t(f) = (\widetilde{k}(\mathbf{x}_i, \mathbf{x}_j; f))_{i,j\in[\![t-1]\!]}$. Using the identity in Equation (1), we can also deduce that $\det(\lambda^{-1}\widetilde{V}_t(\lambda; f)) = \det(\lambda^{-1}\widetilde{\mathbf{K}}_t(\lambda; f))$. We also define the *weighted information gain* $\widetilde{\Gamma}_t(f)$ and the *weighted maximal information gain* $\widetilde{\gamma}_t(f)$ as $\widetilde{\Gamma}_t(f) \coloneqq \frac{1}{2}\log\det(\lambda^{-1}\widetilde{\mathbf{K}}_t(\lambda; f))$ and $\widetilde{\gamma}_t(t) \coloneqq \max_{\mathbf{x}_1,\dots,\mathbf{x}_t\in\mathcal{X}} \widetilde{\Gamma}_t(f)$. Finally, we consider the maximum value of the (maximal) information gain by varying the function f in $\mathcal{H}$, i.e., $\widetilde{\Gamma}_t(\mathcal{H}) = \sup_{f\in\mathcal{H}} \widetilde{\Gamma}_t(f)$ and $\widetilde{\gamma}_t(\mathcal{H}) = \sup_{f\in\mathcal{H}} \widetilde{\gamma}_t(f)$. The following result relates weighted and unweighted information gains.

[4] Assessing whether a function $f \in \mathcal{H}$ belongs to the confidence set $\mathcal{C}_t(\delta)$ is clearly intractable since it requires computing norms of operators. In Appendix A, we provide an efficient alternative confidence set that will lead to analogous regret guarantees.

Lemma 5.1. *Let $\mathcal{H}$ be a RKHS induced by kernel k. Let $t \in \mathbb{N}$ and let $\mathbf{x}_1, \dots, \mathbf{x}_t \in \mathcal{X}$ be a sequence of decisions. It holds that $\widetilde{\Gamma}_t(\mathcal{H}) \leq \max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\Gamma_t$.*

Notice that the bound introduces just a dependence on the maximum slope of the inverse link function $R_{\dot{\mu}}$ and no dependence on the minimum slope $\kappa_{\mathcal{X}}$. This result will play a significant role in the derivation of an efficient implementation for `GKB-UCB` (Appendix A).

6 Challenges and New Technical Tools

In this section, we discuss the main challenges for achieving sensible regret guarantees for GKBs. We start discussing the limitations of existing *self-normalized* concentration bounds (see Table 1) to control the error in the ML estimate (Section 6.1). This motivates the need for a novel self-normalized inequality that represents a key contribution of this work (Section 6.2).

6.1 Limitations of Existing Self-Normalized Concentration Inequalities

To understand the need for a novel concentration bound, we need to anticipate some key passages of the regret analysis. We recall that the confidence radius $B_t(\delta; f)$ should be designed to guarantee that: (i) the true unknown function $f^* = \langle \alpha^*, \phi \rangle$ belongs to $\mathcal{C}_t(\delta)$ (Equation 7) and (ii) the regret is as small as possible. For point (i), we can conveniently express the difference between the operators g_t evaluated in the true function f^* and in the ML estimate $\widehat{f}_t$ (see Lemma 7.1):

$$g_t(f^*) - g_t(\widehat{f}_t) = \mathfrak{g}(\tau)^{-1} \sum_{s=1}^{t-1} \epsilon_s \phi(\mathbf{x}_s) + \lambda \alpha^*, \tag{9}$$

where $\epsilon_s = y_s - \mu(f^*(\mathbf{x}_s))$ is the noise. Thus, since since α^* is bounded in norm under Assumption 3.1, to suitably design $B_t(\delta; f)$, we need to control the martingale $S_t = \sum_{s=1}^{t-1} \epsilon_s \phi(\mathbf{x}_s)$. For point (ii), in the regret analysis, we need to bound the difference between optimistic function $\widetilde{f}_t$ and true unknown function f^*, both evaluated in the played decision $\mathbf{x}_t$, i.e., $\widetilde{f}_t(\mathbf{x}_t) - f^*(\mathbf{x}_t)$ with the martingale S_t. Similarly to [2, 9], this is done by decomposing both functions as an inner product (Mercer's theorem) and then applying a Cauchy-Schwarz inequality by making a *specific* choice of operator $W_t(f^*)$, possibly depending on the unknown function f^*:

$$\widetilde{f}_t(\mathbf{x}_t) - f^*(\mathbf{x}_t) = \langle \widetilde{\alpha}_t - \alpha^*, \phi(\mathbf{x}_t) \rangle \leq \underbrace{\|\widetilde{\alpha}_t - \alpha^*\|_{W_t(f^*)}}_{\text{(A)}} \underbrace{\|\phi(\mathbf{x}_t)\|_{W_t(f^*)^{-1}}}_{\text{(B)}}. \tag{10}$$

The choice of operator $W_t(f^*)$ has two effects: (i) by relating term (A) with the confidence set $\mathcal{C}_t(\delta)$ definition, it determines the multiplicative coefficient and the norm under which martingale S_t has to be controlled and (ii) it allows bounding (B) by means of an *elliptic potential lemma* [16, Lemma 19.4]. We now discuss two choices of operators $W_t(f^*)$ leading to different concentration bounds and, consequently, confidence sets, and discuss their advantages and disadvantages.

Covariance Operator ($W_t(f^*) = V_t(\lambda)$). We start considering the case in which $W_t(f^*) = V_t(\lambda)$, where V_t is the usual covariance operator. In this case, we can link the term (A) with the confidence set as follows (see Lemma C.4):

$$\text{(A)} = \|\widetilde{\alpha}_t - \alpha^*\|_{V_t(\lambda)} \leq (1 + 2R_s BK) \max\{1, \mathfrak{g}(\tau)\boldsymbol{\kappa_{\mathcal{X}}}\} \left\| g_t(\widetilde{f}_t) - g_t(f^*) \right\|_{V_t^{-1}(\lambda)}, \tag{11}$$

introducing an inconvenient multiplicative dependence on $\max\{1, \mathfrak{g}(\tau)\kappa_{\mathcal{X}}\}$, i.e., on the minimum slope $\kappa_{\mathcal{X}}$ of the inverse link function. At this point, we have to control the martingale S_t under the norm weighted by $V_t^{-1}(\lambda)$, as $\left\| g_t(f^*) - g_t(\widehat{f}_t) \right\|_{V_t^{-1}(\lambda)} \leq \|S_t\|_{V_t^{-1}(\lambda)} + \frac{B}{\sqrt{\lambda}}$. The quantity $\|S_t\|_{V_t^{-1}(\lambda)}$ can be conveniently bounded by using a self-normalized concentration bound for sub-gaussian[5] martingales (i.e., *Hoeffding-like*), as in the seminal work [1]:

$$\|S_t\|_{V_t^{-1}} \leq R\sqrt{2\log(\delta^{-1}) + \log\det(\lambda^{-1}V_t(\lambda))} = R\sqrt{2\log(\delta^{-1}) + \log\det(\lambda^{-1}\mathbf{K}_t(\lambda))}, \tag{12}$$

[5]We recall that since $|\epsilon_s| \leq R$ a.s., it is also R^2-subgaussian.

where the equality is obtained by Equation (1). We recall that the second bound is also obtained in Theorem 1 of [5] where the quantity $\|S_t\|_{V_t^{-1}}$ is controlled in the dual (kernel) space. The advantage of these bounds is that they do not exhibit a dependence on the dimensionality d of the feature space ϕ, which in GKBs is infinite. Nevertheless, in this way, the dependence on the minimum slope of the inverse link function $\kappa_{\mathcal{X}}$ (as in Equation 11) becomes unavoidable in the regret. This suggests that we should prefer a different choice of operator $W_t(f^*)$.

Weighted Covariance Operator ($W_t(f^*) = \widetilde{V}_t(\lambda; f^*)$). The presence of the multiplicative factor $\kappa_{\mathcal{X}}$ depends on the covariance operator and emerges also in the GLB setting when making the choice $W_t(f^*) = V_t(\lambda)$ [2, 9]. The solution, in the GLB case, consists of choosing the weighted covariance operator $W_t(f^*) = \widetilde{V}_t(\lambda; f^*)$, where each outer product $\phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top$ is weighted by the variance $\frac{\dot{\mu}(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)}$ of the noise random variable ϵ_s. This allows relating the distance of the parameters with the confidence set $\mathcal{C}_t(\delta)$, avoiding the inconvenient dependence on $\kappa_{\mathcal{X}}$ (see Lemma C.4 with $f'' = f$):

$$\text{(A)} = \|\widetilde{\alpha}_t - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} \leq (1 + 2R_s BK) \left\| g_t(\widetilde{f}_t) - g_t(f^*) \right\|_{\widetilde{V}_t^{-1}(\lambda; f^*)}. \tag{13}$$

Proceeding analogously as above, we should now control the quantity $\|S_t\|_{\widetilde{V}_t^{-1}(\lambda; f^*)}$. Since the weighted covariance operator $\widetilde{V}_t(\lambda; f^*)$ contains the variance of each sample, we need to resort to a *Bernstein-like* self-normalized concentration bound in order to make effective use of such information. The fundamental result in the GLB literature is the bound of [9, Theorem 1]:

$$\|S_t\|_{\widetilde{V}_t^{-1}(\lambda; f^*)} \leq \frac{\sqrt{\lambda}}{2} + \frac{2}{\sqrt{\lambda}} \boldsymbol{d} \log 2 + \frac{2}{\sqrt{\lambda}} \log \frac{1}{\delta} + \frac{1}{\sqrt{\lambda}} \log \det(\lambda^{-1} \widetilde{V}_t(\lambda; f^*)), \tag{14}$$

where d is the dimensionality of the feature map ϕ, which is infinite-dimensional in our GKB setting, making the bound vacuous.[6]

6.2 A Novel Bernstein-like Dimension-Free Self-Normalized Inequality

From the above discussion, it should now appear clear why we need a *novel self-normalized concentration bound* that combines two desired properties:

- *Bernstein-like*: it should account for a weighted covariance operator $\widetilde{V}_t(\lambda; f^*)$ where the weights correspond to the variance of the samples to avoid the inconvenient multiplicative factor $\kappa_{\mathcal{X}}$;
- *Dimension-free*: it should avoid any dependence on the dimensionality of the feature space ϕ, in order to make it applicable to our GKB setting, where ϕ can be infinite-dimensional.

With this goal, we deviate from the two traditional approaches to derive self-normalized concentrations, i.e., *pseudomaximization* via method of mixtures [1, 7, 9] and *PAC Bayes* [17, 37]. Instead, we follow the path of [36] that, in turn, extends [6], by directly decomposing the norm $\|S_t\|_{\widetilde{V}_t^{-1}(\lambda; f^*)}$ and bounding individual terms by means of Freedman's inequality [11]. In addition to the requirements above, we aim to obtain a *data-driven* bound in which, just like in Equations (12) and (14), the bound depends on the sequence of the actual decisions, i.e., on the weighted information gain $\widetilde{\Gamma}_t(f^*) = \frac{1}{2} \log \det(\lambda^{-1} \widetilde{V}_t(\lambda; f^*))$ instead of the *maximal* information gain $\widetilde{\gamma}_t(f^*)$. This is clearly desirable since $\widetilde{\Gamma}_t(f^*) \leq \widetilde{\gamma}_t(f^*)$.[7] However, this is not straightforward when following the technique of [6, 36], that necessitates deterministic bounds to the cumulative variance for the application of Freedman's inequality. For this reason, we provide a first result that extends Freedman's inequality allowing for bounds of the cumulative variance that are not deterministic but, instead, predictable processes. This will represent the core for deriving our self-normalized concentration bound.

Theorem 6.1 (A data-driven Freedman's inequality)**.** *Let $(z_t)_{t \geq 1}$ be a real-valued martingale difference sequence adapted to the filtration $\mathcal{F}_t$ such that $z_t \leq R$ a.s. for all $t \geq 1$. Let $(v_t)_{t \geq 1}$ be a process predictable by the filtration $\mathcal{F}_t$ such that for every $t \geq 1$, we have that $\sum_{s=1}^{t} \mathbb{E}[z_s^2 | \mathcal{F}_{s-1}] \leq v_t$*

[6] One could attempt to operate as in [9, Theorem 1] for deriving but directly in the dual (kernel) space. Although this is possible, it would make appear a dependence on the order of the weighted kernel matrix $\widetilde{\mathbf{K}}_t(\lambda; f)$, i.e., t in replacement of d. This is not of any help since it will make the regret degenerate to linear.

[7] Indeed, in [36], the bound depends on an upper bound of γ_t obtained by bounding the maximum value of $\log \det(\lambda^{-1} V_t)$ considering the worst-case sequence of decisions [see Lemma B.2 of 36].

a.s.. Then, for every $\eta > 1$ and $v_0 > 0$, with probability at least $1 - \delta$, it holds that:

$$\forall t \geq 1: \qquad \sum_{s=1}^{t} z_s \leq \sqrt{2 \max\{v_0, \eta v_t\} \log \frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}} + \frac{R}{3} \log \frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}, \tag{15}$$

where $\widehat{\ell} = \max\left\{0, \left\lceil \log_\eta(v_t/v_0) \right\rceil\right\}$.

The inequality of Theorem 6.1, compared to the standard Freedman's inequality (see Lemma B.1), allows obtaining a bound that depends on the predictable process v_t that we can think to as a proxy (upper bound) of the variance that, however, does not need to be deterministic. This allows us to obtain bounds that depend on the actual sequence of decisions $\mathbf{x}_1, \dots, \mathbf{x}_t$ and their weighted information gain $\widetilde{\Gamma}_t(f^*)$ rather than on the maximal weighted information gain $\widetilde{\gamma}_t(f^*)$, with an improvement over previous inequalities like [36]. From a technical perspective, Theorem 6.1 is obtained using a *stitching* argument [13] that brings two beneficial effects. First, it allows to accurately perform *union bounds* considering the values that the predictable process can take over a geometric grid $\{\eta^\ell v_0 \,:\, \ell \in \mathbb{N}\}$ enabling the use of the data-driven quantity v_t, where the parameters $\eta > 1$ and $v_0 > 0$ can be selected to tighten the bound. Second, it allows replacing a $\log t$ term in the bound with a $\log \log t$ at the price of a larger multiplicative constant $\eta > 1$. A similar data-driven result has been provided in [12, Theorem 12]. However, our result allows tuning the parameters η and v_0 to tighten the bound, ultimately leading to an improvement of the constants. We can now use Theorem 6.1 to derive our novel *self-normalized Bernstein-like dimension-free* concentration inequality.

Theorem 6.2 (Bernstein-Like Dimension-Free Self-Normalized Concentration)**.** *Let $(\mathbf{x}_t)_{t\geq 1}$ be a discrete-time stochastic process predictable by the filtration $\mathcal{F}_t$ and let $(\epsilon_t)_{t\geq 1}$ be a real-valued stochastic process adapted to the $\mathcal{F}_t$ such that $\mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0$, $\mathbb{V}\mathrm{ar}[\epsilon_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \sigma^2(\mathbf{x}_t)$, and $|\epsilon_t| \leq R$ a.s. for every $t \geq 1$. Let $\phi : \mathcal{X} \to \mathbb{R}^{\mathbb{N}}$ be the feature mapping induced by kernel k such that $\|\phi(\mathbf{x})\|_2 \leq K$ for every $\mathbf{x} \in \mathcal{X}$. Let:*

$$S_t \coloneqq \sum_{s=1}^{t-1} \epsilon_s \phi(\mathbf{x}_s), \qquad \widetilde{V}_t(\lambda) \coloneqq \sum_{s=1}^{t-1} \sigma_s^2 \phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top + \lambda I. \tag{16}$$

Then, for every $\delta \in (0,1)$ and $t \geq 1$, with probability at least $1 - \delta$ it holds that:

$$\|S_t\|_{\widetilde{V}_t^{-1}(\lambda)} \leq \left(\sqrt{73 \log\det(\lambda^{-1}\widetilde{V}_t)} + \sqrt{3}\right) \sqrt{\log \frac{\pi^2(\rho+1)^2}{3\delta}} + \frac{3RK}{\sqrt{\lambda}} \log \frac{\pi^2(\rho+1)^2}{3\delta}, \tag{17}$$

where $\rho = \max\left\{0, \left\lceil \log\left(\frac{8R^2K^2(t-1)^3}{\lambda} \log\left(1 + \frac{K^2R^2}{\lambda}\right)\right)\right\rceil\right\}$.

The concentration bound, as desired, displays no dependence on the dimensionality d of the feature map ϕ and no explicit dependence on t (apart from sub-logarithmic ones). We succeeded to remove the dependence from d by replacing it with the norm of the feature map, which is bounded by K under Assumption 3.1. It is worth noting that, thanks to the data-driven bound of Theorem 6.1, we have a dependence on the term $\log\det(\lambda^{-1}\widetilde{V}_t(\lambda))$ that, thanks to the identity in Equation (1), can be expressed in the dual (kernel) space by means of the information gain $2\widetilde{\Gamma}_t = \log\det(\lambda^{-1}\widetilde{\mathbf{K}}_t(\lambda))$, where the weighted kernel matrix $\widetilde{\mathbf{K}}_t(\lambda)$ is obtained by means of the weighted kernel $\widetilde{k}(\mathbf{x}, \mathbf{x}') = \sigma(\mathbf{x})k(\mathbf{x}, \mathbf{x}')\sigma(\mathbf{x}')$ that induces the modified feature map $\widetilde{\phi}(\mathbf{x}) = \sigma(\mathbf{x})\phi(\mathbf{x})$. By denoting with $\widetilde{\gamma}_t = \max_{\mathbf{x}_1,\dots,\mathbf{x}_t \in \mathcal{X}} \widetilde{\Gamma}_t$, we can write the non-data-driven bound, holding with probability $1 - \delta$:

$$\forall t \geq 1: \qquad \|S_t\|_{\widetilde{V}_t^{-1}} \leq \left(\sqrt{146\widetilde{\gamma}_t} + \sqrt{3}\right) \sqrt{\log \frac{\pi^2(\rho+1)^2}{3\delta}} + \frac{3RK}{\sqrt{\lambda}} \log \frac{\pi^2(\rho+1)^2}{3\delta}. \tag{18}$$

7 Regret Analysis

In this section, we provide the regret analysis of `GKB-UCB` (Algorithm 1). We start with a lemma to show that f^* belongs to the confidence set $\mathcal{C}_t(\delta)$ (in high probability) with a proper choice of the confidence radius $B_t(\delta; f)$ (Lemma 7.1). Then, we move to the regret analysis (Theorem 7.2).

Lemma 7.1 (Good Event). *Let $t \in \mathbb{N}$, $f \in \mathcal{H}$, and $\delta \in (0,1)$, define the confidence radius as:*

$$B_t(\delta; f) \coloneqq \sqrt{\lambda}B + \frac{1}{\mathfrak{g}(\tau)}\left(\sqrt{73\log\det(\lambda^{-1}\widetilde{V}_t(\lambda; f))} + \sqrt{3}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{3\delta}} + \frac{3RK}{\mathfrak{g}(\tau)\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{3\delta},$$

where $\rho = \max\left\{0, \left\lceil \log\left(\frac{8R^2K^2(t-1)^3}{\lambda}\log\left(1+\frac{K^2R^2}{\lambda}\right)\right)\right\rceil\right\}$. *Let* $\mathcal{E}_\delta \coloneqq \{\forall t \geq 1 \,:\, f^* \in \mathcal{C}_t(\delta)\}$. *Under Assumptions 3.1, 3.2, and 3.3, it holds that* $\Pr(\mathcal{E}_\delta) \geq 1-\delta$.

Lemma 7.1 resorts to our novel self-normalized bound (Theorem 6.2), together with Assumption 3.1, to provide a form to the confidence radius $B_t(\delta; f)$. It is worth noting that, differently from the majority of existing works [1, 2, 17], $B_t(\delta; f)$ explicitly depends on function f since the operator $\widetilde{V}_t(\lambda; f)$ necessitates f to compute the weights $\mathfrak{g}(\tau)^{-1}\dot{\mu}(f(\mathbf{x}_s))$. By exploiting the identity in Equation (1), we can move to the dual (kernel) space in order to operate with finite-dimensional objects: $\log\det(\lambda^{-1}\widetilde{V}_t(\lambda; f)) = \log\det(\widetilde{\mathbf{K}}_t(\lambda; f)) = 2\widetilde{\Gamma}_t(f)$. Let us also define its worst-case version w.r.t. the choice of function $f \in \mathcal{H}$, i.e., $B_t(\delta; \mathcal{H}) = \sup_{f\in\mathcal{H}} B_t(\delta; f)$. Although GKB-UCB makes use of the confidence radius $B_t(\delta; f)$, for analysis purposes, we also define a non-data-driven confidence radius, where the information gain $\widetilde{\Gamma}_t(f)$ is replaced by its maximal version:

$$\beta_t(\delta; f) \coloneqq \sqrt{\lambda}B + \mathfrak{g}(\tau)^{-1}\left(\sqrt{146\widetilde{\gamma}_t(f)} + \sqrt{3}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{3\delta}} + \frac{3\mathfrak{g}(\tau)^{-1}RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{3\delta},$$

and, finally, we introduce its worst-case version w.r.t. the choice of function $f \in \mathcal{H}$, i.e., $\beta_t(\delta; \mathcal{H}) = \sup_{f\in\mathcal{H}} \beta_t(\delta; f)$, i.e., obtained from $\beta_t(\delta; f)$ by replacing $\widetilde{\gamma}_t(f)$ with $\widetilde{\gamma}_t(\mathcal{H})$.

We are now ready to present the regret bound of GKB-UCB.

Theorem 7.2 (Regret Bound of GKB-UCB). *Under Assumptions 3.1, 3.2, 3.3, and 3.4,* GKB-UCB *with the confidence radius $B_t(\delta; f)$ as defined in Lemma 7.1 and $\lambda > 0$, for every $\delta \in (0,1)$, with probability at least $1-\delta$, suffers regret bounded as $R(\texttt{GKB-UCB}, T) = R_{\text{perm}}(T) + R_{\text{trans}}(T)$, where:*

$$R_{\text{perm}}(T) \leq 8(1+2R_sBK)\beta_T(\delta; \mathcal{H})\sqrt{\max\left\{\mathfrak{g}(\tau), \lambda^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}}, \tag{19}$$

$$R_{\text{trans}}(T) \leq 32R_s(1+R_{\dot{\mu}}\kappa_{\mathcal{X}})(1+2R_sBK)^2\beta_T(\delta; \mathcal{H})^2\max\left\{\mathfrak{g}(\tau), \lambda^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*). \tag{20}$$

The proof schema of Theorem 7.2 follows similar steps to [2] and the result, indeed, displays an analogous regret decomposition into a *permanent* term $R_{\text{perm}}(T)$ and a *transient* term $R_{\text{trans}}(T)$. Regarding the dependence on explicit T and κ_*, $R_{\text{perm}}(T)$ is the dominating term that displays the desired dependence on $\sqrt{T/\kappa_*}$, whereas $R_{\text{trans}}(T)$ exhibits a dependence on the minimum slope of the inverse link function $\kappa_{\mathcal{X}}$, but has only logarithmic dependence on T and, for this reason, it is negligible. To highlight the dependence on the information gain, we explicit the form of the individual terms in the case $\lambda \geq \Omega(K^2)$:[8] $\beta_T(\delta; \mathcal{H}) = \widetilde{O}(\sqrt{\lambda}B + \sqrt{\widetilde{\gamma}_T(\mathcal{H})\log(\delta^{-1})} + RK\log(\delta^{-1}))$. Thus, we obtain a regret bound of order:

$$R(\texttt{GKB-UCB}, T) \leq \widetilde{O}\left((1+R_sBK)\left(\sqrt{\lambda}B + \sqrt{\widetilde{\gamma}_T(\mathcal{H})\log(\delta^{-1})} + RK\log(\delta^{-1})\right)\sqrt{\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}}\right).$$

We have two terms related to the weighted information gain, i.e., $\widetilde{\gamma}_T(\mathcal{H})$ and $\widetilde{\gamma}_T(f^*)$. This is due to the fact that our weighted kernel $\widetilde{k}(\cdot,\cdot; f)$ explicitly depends on the evaluated function f. It is worth noting that, thanks to Lemma 5.1, we can bound both with the (unweighted) information gain as $\widetilde{\gamma}_T(f^*) \leq \widetilde{\gamma}_T(\mathcal{H}) \leq \max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\gamma_T$ at the mild price of a multiplicative term.

Let us now comment on the tightness of the bound in the particular cases of KBs and GLBs. For KBs, we are in the presence of ν^2-subgaussian noise and, thus, we need to set $R = O(\nu\sqrt{\log(T/\delta)})$. Furthermore, we have that $R_s = 0$ and $\mu = I$ (consequently, $\dot{\mu} = 1$, $\kappa_* = 1$, and $\widetilde{\gamma}_T(f^*) = \widetilde{\gamma}_T(\mathcal{H}) = \gamma_T$). This allows recovering the bound of order

[8] With the $O(\cdot)$ notation, we suppress multiplicative constants and dependencies on $\mathfrak{g}(\tau)$ and $R_{\dot{\mu}}$. With the $\widetilde{O}(\cdot)$ notation, we also suppress logarithmic dependencies on all variables, except for δ.

$\widetilde{O}\left(\left(\sqrt{\lambda}B + \sqrt{\gamma_T \log(\delta^{-1})} + K\nu \log(\delta^{-1})^{3/2}\right)\sqrt{\gamma_T T}\right)$, matching the regret order of [5] up to logarithmic terms. For GLBs, we can bound the information gain as (see Lemma 11 of [2]):

$$\widetilde{\gamma}_T(\mathcal{H}) \leq \max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\gamma_T \leq \max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\} d \log\left(\lambda + \frac{TK^2}{d}\right). \tag{21}$$

This leads to bound of order $\widetilde{O}((1 + R_s BK)(\sqrt{\lambda}B + \sqrt{d\log(\delta^{-1})} + RK\log(\delta^{-1}))\sqrt{dT/\kappa_*})$, matching the result of [2] up to logarithmic terms.

8 Conclusions

In this paper, we have introduced the novel setting of GKBs, unifying KBs and GLBs. We have provided a novel Bernstein-like dimension-free self-normalized bound of independent interest. We employed it to analyze the regret of `GKB-UCB` showing tight regret bounds. Future works include investigating the use of the techniques from [17] in order to remove the multiplicative dependence on the norm and kernel bounds $(1 + R_s BK)$ in the regret bound as well as the study of the inherent complexity of regret minimization in the GLB setting by conceiving regret lower bounds [26].

References

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

[2] Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 3691–3699. PMLR, 2021.

[3] Ole Barndorff-Nielsen. *Information and exponential families in statistical theory*. John Wiley & Sons, 1978.

[4] Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and Csaba Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(5), 2011.

[5] Sayak R. Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning (ICML)*, pages 844–853. PMLR, 2017.

[6] Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic linear optimization under bandit feedback. In *Annual Conference on Learning Theory (COLT)*, pages 355–366. Omnipress, 2008.

[7] Victor H. De la Peña, Michael J. Klass, and Tze L. Lai. Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *The Annals of Probability*, pages 1902–1933, 2004.

[8] Victor H. De la Peña, Tze L. Lai, and Qi-Man Shao. *Self-normalized processes: Limit theory and Statistical Applications*. Springer, 2009.

[9] Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning (ICML)*, pages 3052–3060. PMLR, 2020.

[10] Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems (NIPS)*, pages 586–594. Curran Associates, Inc., 2010.

[11] David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, pages 100–118, 1975.

[12] Pierre Gaillard, Gilles Stoltz, and Tim van Erven. A second-order bound with excess losses. In *Conference on Learning Theory (COLT)*, volume 35 of *JMLR Workshop and Conference Proceedings*, pages 176–196. JMLR.org, 2014.

[13] Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021.

[14] Hermann König. *Eigenvalue distribution of compact operators*, volume 16. Birkhäuser, 1986.

[15] Tze L. Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[16] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[17] Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[18] Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning (ICML)*, pages 2071–2080. PMLR, 2017.

[19] Peter Mccullagh and John A. Nelder. *Generalized linear models*. Chapman and Hall, 1989.

[20] James Mercer. Functions of positive and negative type, and their connection the theory of integral equations. *Philosophical transactions of the royal society of London*, 209:415–446, 1909.

[21] Marco Rando, Luigi Carratino, Silvia Villa, and Lorenzo Rosasco. Ada-bkb: Scalable gaussian process optimization on continuous domains by adaptive discretization. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 7320–7348. PMLR, 2022.

[22] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, 2006.

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

[24] Yoan Russac, Louis Faury, Olivier Cappé, and Aurélien Garivier. Self-concordant analysis of generalized linear bandits with forgetting. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 658–666. PMLR, 2021.

[25] Ayush Sawarni, Nirjhar Das, Siddharth Barman, and Gaurav Sinha. Generalized linear bandits with limited adaptivity. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[26] Jonathan Scarlett, Ilija Bogunovic, and Volkan Cevher. Lower bounds on regret for noisy gaussian process bandit optimization. In *Annual Conference on Learning Theory (COLT)*, volume 65 of *Proceedings of Machine Learning Research*, pages 1723–1742. PMLR, 2017.

[27] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. A generalized representer theorem. In *International conference on computational learning theory*, pages 416–426. Springer, 2001.

[28] Alex J. Smola and Bernhard Schölkopf. *Learning with kernels*. Citeseer, 1998.

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

[30] Terence Tao. 254a, notes 3a: Eigenvalues and sums of hermitian matrices. *Terence Tao's blog*, 2010.

[31] Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, pages 82–90. PMLR, 2021.

[32] Justin A. Whitehouse, Aaditya Ramdas, and Zhiwei S. Wu. On the sublinear regret of GP-UCB. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 35266–35276, 2023.

[33] Max A. Woodbury. *Inverting modified matrices*. Department of Statistics, Princeton University, 1950.

[34] Xingzhi Zhan. *Matrix inequalities*. Springer, 2004.

[35] Tong Zhang. *Mathematical analysis of machine learning algorithms*. Cambridge University Press, 2023.

[36] Dongruo Zhou, Quanquan Gu, and Csaba Szepesvari. Nearly minimax optimal reinforcement learning for linear mixture markov decision processes. In *Annual Conference on Learning Theory (COLT)*, pages 4532–4576. PMLR, 2021.

[37] Ingvar Ziemann. A vector bernstein inequality for self-normalized martingales. *Transactions on Machine Learning Research*, 2025.

A Efficient Implementation

In this section, we show how to make Algorithm 1 computationally tractable with negligible effects in the final regret bound. Indeed, Algorithm 1 is based on a confidence set $\mathcal{C}_t(\delta)$ (Equation 7) that requires evaluating the norm of a difference of operators g_t, which is clearly computationally infeasible. In the following, we discuss how to make both steps of the algorithm computationally tractable. The resulting algorithm, `Eff-GKB-UCB`, is provided in Algorithm 2.

Input: Decision set $\mathcal{X}$, confidence level δ, confidence sets $\mathcal{D}_t(\delta)$
for $t \in [\![T]\!]$ **do**
  `/* Step 1: Efficient Maximum Likelihood Estimate */`
  $\widehat{\boldsymbol{\alpha}}_t = \arg\min_{\boldsymbol{\alpha} \in \mathbb{R}^{t-1}} \mathcal{L}_t\left(\langle \boldsymbol{\alpha}, \mathbf{k}_t(\cdot)\rangle\right)$ (Equation 3)
  `/* Step 2: Efficient Optimistic Decision Selection */`
  $\mathbf{x}_t \in \arg\max_{\mathbf{x} \in \mathcal{X}} \max_{\boldsymbol{\alpha} \in \mathbb{R}^{t-1}} \langle \boldsymbol{\alpha}, \mathbf{k}_t(\mathbf{x})\rangle$
    s.t. $\mathcal{L}_t(\langle \boldsymbol{\alpha}, \mathbf{k}_t(\cdot)\rangle) \leq \mathcal{L}_t(\langle \widehat{\boldsymbol{\alpha}}_t, \mathbf{k}_t(\cdot)\rangle) + D_t(\delta; \mathcal{H})$ (Equation 23)
  Play $\mathbf{x}_t$ and observe y_t
end

Algorithm 2: `Eff-GKB-UCB`.

Efficient Maximum Likelihood Estimation. Since function m is convex, loss function $\mathcal{L}_t(f)$ is convex in $f = \langle \alpha, \phi \rangle \in \mathcal{H}$ and, consequently, also in the parameter vector $\alpha \in \mathbb{R}^{\mathbb{N}}$. However, optimizing over either f or α is infeasible, being both infinite-dimensional. Nevertheless, thanks to the *generalized representer theorem* [27, Theorem 1], we can restrict the optimization to the functions of the form $f(\cdot) = \sum_{s=1}^{t-1} \alpha_s k(\cdot, \mathbf{x}_s) = \langle \boldsymbol{\alpha}, \mathbf{k}_t(\cdot)\rangle$, where $\boldsymbol{\alpha} = (\alpha_s)_{s \in [\![t-1]\!]}^\top$ and $\mathbf{k}_t(\cdot) = (k(\cdot, \mathbf{x}_s))_{s \in [\![t-1]\!]}^\top$. This allows limiting the problem to the minimization of a convex function on a vector of $t-1$ real variables $\boldsymbol{\alpha} \in \mathbb{R}^{t-1}$.

Efficient Optimistic Decision Selection. To make the choice of the optimistic function, we propose a different (looser) confidence set based on the evaluation of the loss function only [2], defined for every round $t \in [\![T]\!]$ and confidence $\delta \in (0, 1)$:[9]

$$\mathcal{D}_t(\delta) \coloneqq \left\{ f \in \mathcal{H} \,:\, \mathcal{L}_t(f) - \mathcal{L}_t(\widehat{f}_t) \leq D_t(\delta; \mathcal{H}) \coloneqq (1 + 2R_s BK) B_t(\delta; \mathcal{H}) \right\}, \tag{22}$$

We prove in Lemma B.2 that the choice of the confidence radius ensures the inclusion property between the confidence sets $\mathcal{C}_t(\delta) \subseteq \mathcal{D}_t(\delta)$. Having fixed a decision $\widehat{\mathbf{x}} \in \mathcal{X}$,[10] the optimistic decision selection can be formulated, thanks to the generalized representer theorem [27] as the following constrained convex program:[11]

$$\begin{aligned} \min_{\boldsymbol{\alpha} \in \mathbb{R}^{t-1}} \quad & -\langle \boldsymbol{\alpha}, \mathbf{k}_t(\widehat{\mathbf{x}})\rangle \\ \text{subject to} \quad & \mathcal{L}_t(\langle \boldsymbol{\alpha}, \mathbf{k}_t(\cdot)\rangle) \leq \mathcal{L}_t(\langle \widehat{\boldsymbol{\alpha}}_t, \mathbf{k}_t(\cdot)\rangle) + (1 + 2R_s BK) B_t(\delta; \mathcal{H}), \end{aligned} \tag{23}$$

where $\widehat{\boldsymbol{\alpha}}_t$ are the parameters of the ML function computed in the previous step. Thus, the program has a linear objective function and a convex constraint, being $\mathcal{L}_t(\langle \boldsymbol{\alpha}, \mathbf{k}_t(\cdot)\rangle)$ convex in $\boldsymbol{\alpha}$.

[9] Computing $B_t(\delta; \mathcal{H})$ can be not straightforward for specific choices of kernel k and inverse link function μ. In such a case, we can upper bound it using Lemma 5.1 by replacing $\sup_{f \in \mathcal{H}} \log\det(\lambda^{-1}\widetilde{V}_t(\lambda; f))$ with $\max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\} \log\det(\lambda^{-1}V_t(\lambda))$. This has the effect of replacing the terms $\widetilde{\gamma}_T(\mathcal{H})$ with $\max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\gamma_T$ in the final regret bound.

[10] As customary in this literature [29, 5], we do not address the issue of optimizing over the decision space efficiently. This can surely be done efficiently when $\mathcal{X}$ is finite. When $\mathcal{X}$ is continuous, we can resort to a *discretization* based on the regularity properties of the kernel function, with a controllable effect on the final regret performances [21].

[11] Even if the representer theorem is formulated for unconstrained minimization, it admits costs functions that take $+\infty$ as value [27]. Thus, we can convert a constrained minimization into an unconstrained one by bringing the constraint into the objective function and making it take value $+\infty$ when the constraint is violated.

We now show that the choice of the new confidence set $\mathcal{D}_t(\delta)$ does not degrade the dependence on the relevant quantities compared to using $\mathcal{C}_t(\delta)$.

Theorem A.1 (Regret Bound of `Eff-GKB-UCB`)**.** *Under Assumptions 3.1, 3.2, 3.3, and 3.4,* `GKB-UCB` *with confidence radius* $(1+2R_sBK)B_t(\delta;\mathcal{H})$ *and* $\lambda>0$*, for every* $\delta\in(0,1)$*, with probability at least* $1-\delta$*, suffers regret bounded as:* $R(\texttt{Eff-GKB-UCB},T)=R_{\text{perm}}(T)+R_{\text{trans}}(T)$*, where:*

$$R_{\text{perm}}(T)\leq 4\sqrt{\max\{\mathfrak{g}(\tau),\lambda^{-1}R_{\dot{\mu}}K^2\}}(2+2R_sBK)\sqrt{\beta_T(\delta;\mathcal{H})}\left(\sqrt{\beta_T(\delta;\mathcal{H})}+2\right)\sqrt{\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}},$$

$$R_{\text{trans}}(T)\leq 8R_s(1+R_{\dot{\mu}}\kappa_{\mathcal{X}})\max\left\{\mathfrak{g}(\tau),\lambda^{-1}R_{\dot{\mu}}K^2\right\}(2+2R_sBK)^2\beta_T(\delta;\mathcal{H})\left(\sqrt{\beta_T(\delta;\mathcal{H})}+2\right)^2\widetilde{\gamma}_T(f^*).$$

The bounds of Theorem 7.2 and Theorem A.1 exhibit the same order dependence on the relevant quantities, but Theorem A.1 has a larger constant, approximately 3 times larger than Theorem 7.2 for $R_{\text{perm}}(T)$ and 9 times larger for $R_{\text{trans}}(T)$.

B Proofs

B.1 Proofs of Section 5

Lemma 5.1. *Let* $\mathcal{H}$ *be a RKHS induced by kernel* k*. Let* $t\in\mathbb{N}$ *and let* $\mathbf{x}_1,\dots,\mathbf{x}_t\in\mathcal{X}$ *be a sequence of decisions. It holds that* $\widetilde{\Gamma}_t(\mathcal{H})\leq\max\{1,R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\Gamma_t$*.*

Proof. A direct application of Lemma C.5. □

B.2 Proofs of Section 6

Lemma B.1 (Freedman's Inequality)**.** *Let* $(z_t)_{t\geq1}$ *be a real-valued martingale difference sequence adapted to the filtration* $\mathcal{F}_t$ *such that* $z_t\leq R$ *a.s. for all* $t\geq1$*. Then, for every* $\lambda\in(0,3/R)$ *it holds that with probability at least* $1-\delta$*:*

$$\forall t\geq1:\qquad\sum_{s=1}^{t}z_s\leq\frac{\lambda}{2(1-\lambda R/3)}\sum_{s=1}^{t}\mathbb{E}[z_s^2|\mathcal{F}_{s-1}]+\frac{\log\delta^{-1}}{\lambda}.\tag{24}$$

This implies that for every $\nu>0$*, with probability at least* $1-\delta$*:*

$$\forall t\geq1:\qquad\sum_{s=1}^{t}z_s\leq\nu\sqrt{2\log\delta^{-1}}+\frac{R\log\delta^{-1}}{3}\quad\textit{or}\quad\sum_{s=1}^{t}\mathbb{E}[z_s^2|\mathcal{F}_{s-1}]>\nu^2.\tag{25}$$

Proof. Refer to Theorem 13.6 of [35]. □

Theorem 6.1 (A data-driven Freedman's inequality)**.** *Let* $(z_t)_{t\geq1}$ *be a real-valued martingale difference sequence adapted to the filtration* $\mathcal{F}_t$ *such that* $z_t\leq R$ *a.s. for all* $t\geq1$*. Let* $(v_t)_{t\geq1}$ *be a process predictable by the filtration* $\mathcal{F}_t$ *such that for every* $t\geq1$*, we have that* $\sum_{s=1}^{t}\mathbb{E}[z_s^2|\mathcal{F}_{s-1}]\leq v_t$ *a.s.. Then, for every* $\eta>1$ *and* $v_0>0$*, with probability at least* $1-\delta$*, it holds that:*

$$\forall t\geq1:\qquad\sum_{s=1}^{t}z_s\leq\sqrt{2\max\{v_0,\eta v_t\}\log\frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}}+\frac{R}{3}\log\frac{\pi^2(\widehat{\ell}+1)^2}{6\delta},\tag{15}$$

where $\widehat{\ell}=\max\left\{0,\left\lceil\log_\eta(v_t/v_0)\right\rceil\right\}$*.*

Proof. The proof makes use of classical Freedman's inequality [11] combined with a *stitching* argument [13]. We start from the version of Freedman's inequality of Lemma B.1 taken from [35]:

$$\Pr\left(\exists t\geq1:\sum_{s=1}^{t}z_s>\nu\sqrt{2\log\delta^{-1}}+\frac{R\log\delta^{-1}}{3},\ \sum_{s=1}^{t}\mathbb{E}[z_s^2|\mathcal{F}_{s-1}]\leq\nu^2\right)\leq\delta.\tag{26}$$

Since $v_t \geq \sum_{s=1}^{t} \mathbb{E}[z_s^2|\mathcal{F}_{s-1}]$ a.s. for every $t \geq 1$, it immediately follows that:

$$\Pr\left(\exists t \geq 1 : \sum_{s=1}^{t} z_s > \nu\sqrt{2\log\delta^{-1}} + \frac{R\log\delta^{-1}}{3},\ v_t \leq \nu^2\right) \leq \delta. \tag{27}$$

We now proceed by performing a stitching argument with a geometric grid over the values of $\nu \geq 0$ defined as $\{\eta^\ell v_0 \ : \ \ell \in \mathbb{N}\}$ for any choice of $\eta > 1$ and $v_0 > 0$. Thus, we have:

$$\Pr\left(\exists \ell \in \mathbb{N},\ \exists t \geq 1 : \sum_{s=1}^{t} z_s > \sqrt{2\eta^\ell v_0 \log\frac{\pi^2(\ell+1)^2}{6\delta}} + \frac{R}{3}\log\frac{\pi^2(\ell+1)^2}{6\delta},\ v_t \leq \eta^\ell v_0\right) \tag{28}$$

$$\leq \sum_{\ell\in\mathbb{N}} \Pr\left(\exists t \geq 1 : \sum_{s=1}^{t} z_s > \sqrt{2\eta^\ell v_0 \log\frac{\pi^2(\ell+1)^2}{6\delta}} + \frac{R}{3}\log\frac{\pi^2(\ell+1)^2}{6\delta},\ v_t \leq \eta^\ell v_0\right) \tag{29}$$

$$\leq \sum_{\ell\in\mathbb{N}} \frac{6\delta}{\pi^2(\ell+1)^2} \leq \delta, \tag{30}$$

where line (29) follows from a union bound, line (30) is an application of Equation (27) with $\nu = \eta^\ell v_0$ and by observing that $\sum_{\ell\in\mathbb{N}} \frac{1}{(\ell+1)^2} = \frac{\pi^2}{6}$. Let us now consider the smallest value of $\widehat{\ell} \in \mathbb{N}$ such that $v_t \leq \eta^\ell v_0$:

$$\widehat{\ell} = \min\left\{\ell \in \mathbb{N} \ : \ v_t \leq \eta^\ell v_0\right\} = \max\left\{0, \left\lceil \log_\eta \frac{v_t}{v_0}\right\rceil\right\}. \tag{31}$$

For this value of $\widehat{\ell}$, we have:

$$\eta^{\widehat{\ell}} v_0 \leq \eta^{\max\left\{0,\left\lceil\log_\eta \frac{v_t}{v_0}\right\rceil\right\}} v_0 \leq \eta^{\max\left\{0,\log_\eta \frac{v_t}{v_0}+1\right\}} v_0 \leq \max\left\{v_0, \eta v_t\right\}. \tag{32}$$

Finally, we prove the inequality:

$$\Pr\left(\exists t \geq 1 : \sum_{s=1}^{t} z_s > \sqrt{2\max\left\{v_0,\eta v_t\right\}\log\frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}} + \frac{R}{3}\log\frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}\right) \tag{33}$$

$$\leq \Pr\left(\exists t \geq 1 : \sum_{s=1}^{t} z_s > \sqrt{2\eta^{\widehat{\ell}} v_0 \log\frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}} + \frac{R}{3}\log\frac{\pi^2(\widehat{\ell}+1)^2}{6\delta}\ , v_t \leq \eta^{\widehat{\ell}} v_0\right) \tag{34}$$

$$\leq \Pr\left(\exists \ell \in \mathbb{N},\ \exists t \geq 1 : \sum_{s=1}^{t} z_s > \sqrt{2\eta^\ell v_0 \log\frac{\pi^2(l+1)^2}{6\delta}} + \frac{R}{3}\log\frac{\pi^2(l+1)^2}{6\delta}\ , v_t \leq \eta^\ell v_0\right) \tag{35}$$

$$\leq \delta,$$

where line (34) follows from Equation (32) and line (35) from line (30). □

Theorem 6.2 (Bernstein-Like Dimension-Free Self-Normalized Concentration)**.** *Let $(\mathbf{x}_t)_{t\geq 1}$ be a discrete-time stochastic process predictable by the filtration $\mathcal{F}_t$ and let $(\epsilon_t)_{t\geq 1}$ be a real-valued stochastic process adapted to the $\mathcal{F}_t$ such that $\mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0$, $\mathbb{V}\mathrm{ar}[\epsilon_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \sigma^2(\mathbf{x}_t)$, and $|\epsilon_t| \leq R$ a.s. for every $t \geq 1$. Let $\phi : \mathcal{X} \to \mathbb{R}^{\mathbb{N}}$ be the feature mapping induced by kernel k such that $\|\phi(\mathbf{x})\|_2 \leq K$ for every $\mathbf{x} \in \mathcal{X}$. Let:*

$$S_t \coloneqq \sum_{s=1}^{t-1} \epsilon_s \phi(\mathbf{x}_s), \qquad \widetilde{V}_t(\lambda) \coloneqq \sum_{s=1}^{t-1} \sigma_s^2 \phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top + \lambda I. \tag{16}$$

Then, for every $\delta \in (0,1)$ and $t \geq 1$, with probability at least $1-\delta$ it holds that:

$$\|S_t\|_{\widetilde{V}_t^{-1}(\lambda)} \leq \left(\sqrt{73\log\det(\lambda^{-1}\widetilde{V}_t)} + \sqrt{3}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{3\delta}} + \frac{3RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{3\delta}, \tag{17}$$

where $\rho = \max\left\{0, \left\lceil \log\left(\frac{8R^2K^2(t-1)^3}{\lambda}\log\left(1+\frac{K^2R^2}{\lambda}\right)\right)\right\rceil\right\}$.

Proof. The proof follows similar steps as [6, 36], using Theorem 6.1 as base inequality. For the sake of this derivation, we will suppress the dependence on λ, simply writing $\widetilde{V}_t(\lambda) = \widetilde{V}_t$.[12] Let us introduce the notation $Z_t \coloneqq \|S_t\|_{\widetilde{V}_t^{-1}}$, $w_t \coloneqq \|\phi(\mathbf{x}_t)\|_{\widetilde{V}_t^{-1}}$, and $\widetilde{w}_t \coloneqq \sigma_t \|\phi(\mathbf{x}_t)\|_{\widetilde{V}_t^{-1}}$. We denote with $K = \sup_{\mathbf{x}\in\mathcal{X}} \|\phi(\mathbf{x})\|_2$. From the matrix inversion lemma [33], we have:

$$\widetilde{V}_t^{-1} = \widetilde{V}_{t-1}^{-1} - \frac{\widetilde{V}_{t-1}^{-1}\phi(\mathbf{x}_{t-1})\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}\sigma_{t-1}^2}{1 + \|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}_{t-1}^{-1}}\sigma_{t-1}^2} \tag{36}$$

$$= \widetilde{V}_{t-1}^{-1} - \frac{\widetilde{V}_{t-1}^{-1}\phi(\mathbf{x}_{t-1})\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}\sigma_{t-1}^2}{1 + \widetilde{w}_{t-1}^2}. \tag{37}$$

Let us decompose Z_t:

$$Z_t^2 \coloneqq \|S_t\|^2_{\widetilde{V}_t^{-1}} = S_t^\top \widetilde{V}_t^{-1} S_t \tag{38}$$

$$= (S_{t-1} + \epsilon_{t-1}\phi(\mathbf{x}_{t-1}))^\top \widetilde{V}_t^{-1}(S_{t-1} + \epsilon_{t-1}\phi(\mathbf{x}_{t-1})) \tag{39}$$

$$= S_{t-1}\widetilde{V}_t^{-1}S_{t-1} + 2\epsilon_{t-1}\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_t^{-1}S_{t-1} + \epsilon_{t-1}^2\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_t^{-1}\phi(\mathbf{x}_{t-1}) \tag{40}$$

$$\leq S_{t-1}\widetilde{V}_{t-1}^{-1}S_{t-1} + \underbrace{2\epsilon_{t-1}\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_t^{-1}S_{t-1}}_{\text{(A)}} + \underbrace{\epsilon_{t-1}^2\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_t^{-1}\phi(\mathbf{x}_{t-1})}_{\text{(B)}}, \tag{41}$$

having exploited the fact that $\widetilde{V}_t \succeq \widetilde{V}_{t-1}$. We analyze terms (A) and (B) separately.

Analysis of Term (A). From the matrix inversion lemma, we have:

$$2\epsilon_{t-1}\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_t^{-1}S_{t-1} = 2\epsilon_{t-1}\Bigg(\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}S_{t-1} \tag{42}$$

$$- \frac{\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}\phi(\mathbf{x}_{t-1})\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}S_{t-1}\sigma_{t-1}^2}{1 + \widetilde{w}_{t-1}^2}\Bigg) \tag{43}$$

$$= 2\epsilon_{t-1}\left(\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}S_{t-1} - \frac{\widetilde{w}_{t-1}^2}{1 + \widetilde{w}_{t-1}^2}\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}S_{t-1}\right) \tag{44}$$

$$= 2\epsilon_{t-1}\frac{\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}S_{t-1}}{1 + \widetilde{w}_{t-1}^2} =: \ell_t. \tag{45}$$

Consider now the event $\mathcal{E}_t = \mathbb{1}\{0 \leq s \leq t : Z_s \leq \beta_t\}$, being β_t a non-negative non-decreasing predictable process, whose expression will be defined later. Furthermore, let us define $\widetilde{\beta}_t = \min\left\{\beta_t, \frac{(t-1)RK}{\sqrt{\lambda}}\right\}$ which is non-decreasing as well. Under event $\mathcal{E}_t$, we know that $Z_s \leq \widetilde{\beta}_t$ thanks to Lemma C.7. Under $\mathcal{E}_t$, we bound the maximum value and the variance of ℓ_t. Let us start with the maximum value:

$$\ell_t\mathcal{E}_t \leq |\ell_t\mathcal{E}_t| \leq \left|2\epsilon_{t-1}\frac{\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1}S_{t-1}}{1 + \widetilde{w}_{t-1}^2}\mathcal{E}_t\right| \tag{46}$$

$$\leq \frac{2R}{1 + \widetilde{w}_{t-1}^2}\|\phi(\mathbf{x}_{t-1})\|_{\widetilde{V}_{t-1}^{-1}}\|S_{t-1}\|_{\widetilde{V}_{t-1}^{-1}}\mathcal{E}_t \tag{47}$$

$$\leq \frac{2R\|\phi(\mathbf{x}_{t-1})\|_{\lambda^{-1}I}\beta_{t-1}}{1 + \widetilde{w}_{t-1}^2} \tag{48}$$

$$\leq \frac{2RK}{\sqrt{\lambda}}\beta_t, \tag{49}$$

where line (47) follows from the application of Cauchy-Schwarz inequality and recalling that $|\epsilon_{t-1}| \leq R$ a.s., line (48) is obtained by observing that $\widetilde{V}_{t-1} \succeq \lambda I$ and by exploiting event $\mathcal{E}_t$, and line (49)

[12] With little abuse, we will ignore the fact that ϕ is an infinite-dimensional feature mapping to avoid excessive technicalities. We refer the interested reader to [32] that shows that all passages we do are indeed legal when ϕ is the feature mapping induced by an RKHS.

comes from the bound on $\|\phi(\mathbf{x}_{t-1})\| \leq K$ and the monotonicity of β_t. Let us move to the variance, recalling that ℓ_t is zero mean, i.e., $\mathbb{E}[\ell_t|\mathcal{F}_{t-1}] = 0$:

$$\mathbb{E}\left[\ell_t^2|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\left(2\epsilon_{t-1}\frac{\phi(\mathbf{x}_{t-1})^\top \widetilde{V}_{t-1}^{-1} S_{t-1}}{1+\widetilde{w}_{t-1}^2}\right)^2 \mathcal{E}_t|\mathcal{F}_{t-1}\right] \tag{50}$$

$$\leq \frac{4\sigma_{t-1}^2\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}_{t-1}^{-1}}\|S_{t-1}\|^2_{\widetilde{V}_{t-1}^{-1}}}{(1+\widetilde{w}_{t-1}^2)^2}\mathcal{E}_t \tag{51}$$

$$\leq \left(\frac{2\widetilde{w}_{t-1}}{1+\widetilde{w}_{t-1}^2}\right)^2 \widetilde{\beta}_{t-1}^2 \tag{52}$$

$$\leq \min\{1, 2\widetilde{w}_{t-1}\}^2\widetilde{\beta}_{t-1}^2, \tag{53}$$

where line (51) follows from Cauchy-Schwarz inequality and recalling that $\mathbb{E}[\epsilon_{t-1}|\mathcal{F}_{t-1}] = \sigma_{t-1}^2$, line (53) follows from the inequality $\frac{2x}{1+x^2} \leq \min\{1, 2x\}$ for $x \geq 0$. Summing, we obtain:

$$\sum_{s=1}^{t}\mathbb{E}\left[\ell_s^2|\mathcal{F}_{s-1}\right] \leq \sum_{s=1}^{t}\min\{1, 2\widetilde{w}_{s-1}\}^2\widetilde{\beta}_{t-1}^2 \tag{54}$$

$$\leq 4\widetilde{\beta}_t^2\sum_{s=1}^{t}\min\{1, \widetilde{w}_{s-1}\}^2, \tag{55}$$

where we bounded $\widetilde{\beta}_{t-1} \leq \widetilde{\beta}_t$ and $\min\{1, 2\widetilde{w}_{s-1}\}^2 \leq 4\min\{1, \widetilde{w}_{s-1}\}^2$. From a standard elliptical potential lemma (Lemma C.6 with $M = 1$), we obtain:

$$\sum_{s=1}^{t}\min\{1, \widetilde{w}_{s-1}\}^2 \leq 2\log\frac{\det(\widetilde{V}_t)}{\det(\widetilde{V}_0)} = 2\log\det(\lambda^{-1}\widetilde{V}_t), \tag{56}$$

where $\widetilde{V}_0 = \lambda I$. By Theorem 6.1, setting $\eta = e$, $v_0 = 1$, $v_t = 8\beta_t^2\log\det(\lambda^{-1}\widetilde{V}_t)$, we have that with probability at least $1-\delta$:

$$\forall t \geq 1: \sum_{s=1}^{t}\ell_s \leq \sqrt{2\max\left\{1, 8e\beta_t^2\log\det(\lambda^{-1}\widetilde{V}_t)\right\}\log\frac{\pi^2(\widehat{\rho}+1)^2}{6\delta}} + \frac{2RK}{3\sqrt{\lambda}}\beta_t\log\frac{\pi^2(\widehat{\rho}+1)^2}{6\delta}, \tag{57}$$

with $\widehat{\rho} = \max\left\{0, \left\lceil\log\left(8\frac{(t-1)^2R^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\right)\right\rceil\right\}$, having bounded $\widetilde{\beta}_t \leq \beta_t$ in the inequality and $\widetilde{\beta}_t \leq \frac{(t-1)RK}{\sqrt{\lambda}}$ in the expression of $\widehat{\rho}$.

Analysis of Term (B). We proceed again by using the matrix inversion lemma:

$$\epsilon_{t-1}^2\phi(\mathbf{x}_{t-1})^\top\widetilde{V}_t^{-1}\phi(\mathbf{x}_{t-1}) = \epsilon_{t-1}^2\Bigg(\phi(\mathbf{x}_{t-1})^\top\widetilde{V}_{t-1}^{-1}\phi(\mathbf{x}_{t-1}) \tag{58}$$

$$- \frac{\phi(\mathbf{x}_{t-1})^\top\widetilde{V}_{t-1}^{-1}\phi(\mathbf{x}_{t-1})\phi(\mathbf{x}_{t-1})^\top\widetilde{V}_{t-1}^{-1}\phi(\mathbf{x}_{t-1})\sigma_{t-1}^2}{1+\widetilde{w}_{t-1}^2}\Bigg) \tag{59}$$

$$= \frac{\epsilon_{t-1}^2\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}_{t-1}^{-1}}}{1+\widetilde{w}_{t-1}^2}. \tag{60}$$

Let us define:

$$\ell_t \coloneqq \frac{\epsilon_{t-1}^2\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}_{t-1}^{-1}}}{1+\widetilde{w}_{t-1}^2} - \mathbb{E}\left[\frac{\epsilon_{t-1}^2\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}_{t-1}^{-1}}}{1+\widetilde{w}_{t-1}^2}|\mathcal{F}_{t-1}\right].$$

Let us start bounding the maximum value:

$$\ell_t \leq \frac{\epsilon_{t-1}^2\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}_{t-1}^{-1}}}{1+\widetilde{w}_{t-1}^2} \leq \frac{R^2K^2}{\lambda}, \tag{61}$$

where we bounded $\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}} \leq \|\phi(\mathbf{x}_{t-1})\|^2_{\lambda^{-1}I} \leq \frac{K^2}{\lambda}$.

Concerning the variance, we have:

$$\mathbb{V}\mathrm{ar}[\ell_t|\mathcal{F}_{t-1}] = \mathbb{V}\mathrm{ar}\left[\frac{\epsilon^2_{t-1}\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}}}{1+\widetilde{w}^2_{t-1}}|\mathcal{F}_{t-1}\right] \tag{62}$$

$$\leq \mathbb{E}\left[\left(\frac{\epsilon^2_{t-1}\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}}}{1+\widetilde{w}^2_{t-1}}\right)^2|\mathcal{F}_{t-1}\right] \tag{63}$$

$$\leq \frac{R^2K^2}{\lambda}\,\mathbb{E}\left[\frac{\epsilon^2_{t-1}\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}}}{1+\widetilde{w}^2_{t-1}}|\mathcal{F}_{t-1}\right] \tag{64}$$

$$= \frac{R^2K^2}{\lambda}\frac{\sigma^2_{t-1}\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}}}{1+\widetilde{w}^2_{t-1}} \tag{65}$$

$$= \frac{R^2K^2}{\lambda}\frac{\widetilde{w}^2_{t-1}}{1+\widetilde{w}^2_{t-1}} \tag{66}$$

$$\leq \frac{R^2K^2}{\lambda}\min\{1,\widetilde{w}_{t-1}\}^2, \tag{67}$$

where line (64) derives from applying Equation (61), line (67) follows from the inequality $\frac{x}{1+x} \leq \min\{1,x\}$. Summing and applying the elliptic potential lemma (Lemma C.6 with $M=1$), we have:

$$\sum_{s=1}^{t}\mathbb{V}\mathrm{ar}[\ell_s|\mathcal{F}_{s-1}] \leq \frac{R^2K^2}{\lambda}\sum_{s=1}^{t}\min\{1,\widetilde{w}_{s-1}\}^2 \leq \frac{2R^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t). \tag{68}$$

Furthermore, following the same steps from Equation (64), we obtain:

$$\sum_{s=1}^{t}\mathbb{E}[\ell_s|\mathcal{F}_{s-1}] = \sum_{s=1}^{t}\mathbb{E}\left[\frac{\epsilon^2_{t-1}\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}}}{1+\widetilde{w}^2_{t-1}}|\mathcal{F}_{t-1}\right] \leq 2\log\det(\lambda^{-1}\widetilde{V}_t). \tag{69}$$

We now apply Theorem 6.1, setting $\eta = e$, $v_0 = 1$, $v_t = \frac{2R^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)$, we have that with probability at least $1-\delta$:

$$\forall t \geq 1 : \sum_{s=1}^{t}\frac{\epsilon^2_{t-1}\|\phi(\mathbf{x}_{t-1})\|^2_{\widetilde{V}^{-1}_{t-1}}}{1+\widetilde{w}^2_{t-1}} \leq 2\log\det(\lambda^{-1}\widetilde{V}_t) \tag{70}$$

$$+\sqrt{2\max\left\{1,\frac{2eR^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\right\}\log\frac{\pi^2(\widetilde{\rho}+1)^2}{6\delta}} + \frac{R^2K^2}{3\lambda}\log\frac{\pi^2(\widetilde{\rho}+1)^2}{6\delta}, \tag{71}$$

with $\widetilde{\rho} = \max\left\{0, \left\lceil\log\left(\frac{2R^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\right)\right\rceil\right\}$.

Putting All Together. We observe that $\widehat{\rho} \geq \widetilde{\rho}$, and that $\log\det(\lambda^{-1}\widetilde{V}_t) \leq (t-1)\log\left(1+\frac{R^2K^2}{\lambda}\right)$ from Lemma C.7, we define $\rho := \max\left\{0, \left\lceil\log\left(\frac{8R^2K^2(t-1)^3}{\lambda}\log\left(1+\frac{K^2R^2}{\lambda}\right)\right)\right\rceil\right\}$. Putting together the two bounds, we have to find β_t in order to satisfy the following condition:

$$\text{(A)} + \text{(B)} \leq \sqrt{2\max\left\{1, 8e\beta_t^2\log\det(\lambda^{-1}\widetilde{V}_t)\right\}\log\frac{\pi^2(\rho+1)^2}{6\delta}} + \frac{2RK}{3\sqrt{\lambda}}\beta_t\log\frac{\pi^2(\rho+1)^2}{6\delta} \tag{72}$$

$$+ 2\log\det(\lambda^{-1}\widetilde{V}_t) + \sqrt{2\max\left\{1,\frac{2eR^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\right\}\log\frac{\pi^2(\rho+1)^2}{6\delta}} \tag{73}$$

$$+\frac{R^2K^2}{3\lambda}\log\frac{\pi^2(\rho+1)^2}{6\delta}\leq\beta_t^2. \tag{74}$$

We proceed by bounding the maxima in the left-hand-side as $\max\{a,b\}\leq a+b$ for $a,b\geq 0$ and using the subadditivity of the square root to get a stricter condition:

$$\sqrt{2\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\sqrt{16e\beta_t^2\log\det(\lambda^{-1}\widetilde{V}_t)\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\frac{2RK}{3\sqrt{\lambda}}\beta_t\log\frac{\pi^2(\rho+1)^2}{6\delta} \tag{75}$$

$$+2\log\det(\lambda^{-1}\widetilde{V}_t)+\sqrt{2\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\sqrt{\frac{4eR^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\log\frac{\pi^2(\rho+1)^2}{6\delta}} \tag{76}$$

$$+\frac{R^2K^2}{3\lambda}\log\frac{\pi^2(\rho+1)^2}{6\delta}\leq\beta_t^2. \tag{77}$$

This is a second-degree inequality in the variable β_t and, thus, we have to find the minimum value of β_t fulfilling such an inequality. Using the polynomial inequality of Proposition 7 of [2] (i.e., $x^2\leq bx+c=0\implies x\leq b+\sqrt{c}$ when $b,c\geq 0$), we have:

$$\beta_t\leq\sqrt{16e\log\det(\lambda^{-1}\widetilde{V}_t)\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\frac{2RK}{3\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{6\delta} \tag{78}$$

$$+\left(2\sqrt{2\log\frac{\pi^2(\rho+1)^2}{6\delta}}+2\log\det(\lambda^{-1}\widetilde{V}_t)\right. \tag{79}$$

$$\left.+\sqrt{\frac{4eR^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\frac{R^2K^2}{3\lambda}\log\frac{\pi^2(\rho+1)^2}{6\delta}\right)^{\frac{1}{2}} \tag{80}$$

$$\leq\left((\sqrt{16e}+\sqrt{2})\sqrt{\log\det(\lambda^{-1}\widetilde{V}_t)}+\sqrt{2\sqrt{2}}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{6\delta}} \tag{81}$$

$$+\left(\frac{2}{3}+\frac{1}{\sqrt{3}}\right)\frac{RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{6\delta} \tag{82}$$

$$+\left(\frac{4eR^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\log\frac{\pi^2(\rho+1)^2}{6\delta}\right)^{\frac{1}{4}} \tag{83}$$

$$\leq\left(\left(\sqrt{16e}+\sqrt{2}+\frac{1}{2}\right)\sqrt{\log\det(\lambda^{-1}\widetilde{V}_t)}+\sqrt{2\sqrt{2}}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{6\delta}} \tag{84}$$

$$+\left(\frac{2}{3}+\frac{1}{\sqrt{3}}+\sqrt{e}\right)\frac{RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{6\delta}, \tag{85}$$

where line (82) follows from the subadditivity of the square root and recalling that $\log\frac{6(\rho+1)^2}{\pi^2\delta}\geq 1$ for $t\geq 1$, to get line (85), we apply Young's inequality for products as $ab\leq a^2/2+b^2/2$ for $a,b\geq 0$ to get:

$$\left(\frac{4eR^2K^2}{\lambda}\log\det(\lambda^{-1}\widetilde{V}_t)\log\frac{\pi^2(\rho+1)^2}{6\delta}\right)^{\frac{1}{4}}\leq\sqrt{e}\frac{RK}{\sqrt{\lambda}}\sqrt{\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\frac{1}{2}\sqrt{\log\det(\lambda^{-1}\widetilde{V}_t)}. \tag{86}$$

To obtain more manageable constant, we write:

$$\beta_t\leq\left(\sqrt{73\log\det(\lambda^{-1}\widetilde{V}_t)}+\sqrt{3}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\frac{3RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{6\delta}. \tag{87}$$

A simple inductive argument allows to conclude that, with probability at least $1-2\delta$:

$$Z_t^2\leq\left(\sqrt{73\log\det(\lambda^{-1}\widetilde{V}_t)}+\sqrt{3}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{6\delta}}+\frac{3RK}{\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{6\delta}. \tag{88}$$

Notice that, as requested, β_t is a non-decreasing sequence of t, since ρ is non-decreasing with t and $\det(\lambda^{-1}\widetilde{V}_t)$ is non-decreasing as well. Indeed, since $\widetilde{V}_t=\widetilde{V}_{t-1}+\sigma_{t-1}\phi(\mathbf{x}_{t-1})\phi(\mathbf{x}_{t-1})^\top$, we have that thanks to Weyl's inequality for eigenvalues $\lambda_i(\widetilde{V}_t)\geq\lambda_i(\widetilde{V}_{t-1})$ for all $i\in\mathbb{N}$, being λ_i the i-th eigenvalue [30]. It follows that $\det(\lambda^{-1}\widetilde{V}_t)\geq\det(\lambda^{-1}\widetilde{V}_{t-1})$. Rescaling $\delta\leftarrow\delta/2$, we get the result. □

B.3 Proofs of Section 7

Lemma 7.1 (Good Event)**.** *Let $t \in \mathbb{N}$, $f \in \mathcal{H}$, and $\delta \in (0,1)$, define the confidence radius as:*

$$B_t(\delta; f) \coloneqq \sqrt{\lambda}B + \frac{1}{\mathfrak{g}(\tau)}\left(\sqrt{73\log\det(\lambda^{-1}\widetilde{V}_t(\lambda; f))} + \sqrt{3}\right)\sqrt{\log\frac{\pi^2(\rho+1)^2}{3\delta}} + \frac{3RK}{\mathfrak{g}(\tau)\sqrt{\lambda}}\log\frac{\pi^2(\rho+1)^2}{3\delta},$$

where $\rho = \max\left\{0, \left\lceil \log\left(\frac{8R^2K^2(t-1)^3}{\lambda}\log\left(1+\frac{K^2R^2}{\lambda}\right)\right)\right\rceil\right\}$. Let $\mathcal{E}_\delta \coloneqq \{\forall t \geq 1 \,:\, f^* \in \mathcal{C}_t(\delta)\}$. Under Assumptions 3.1, 3.2, and 3.3, it holds that $\Pr(\mathcal{E}_\delta) \geq 1-\delta$.

Proof. First of all, we observe that $\mathcal{E}_\delta = \left\{\forall t \geq 1 \,:\, \left\|g_t(f^*) - g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda;f)} \leq B_t(\delta; f)\right\}$. Let $t \in \mathbb{N}$ and let us define $\epsilon_t \coloneqq -y_t + \mu(f^*(\mathbf{x}_t))$. We have:

$$\begin{aligned}
&g_t(f^*) - g_t(\widehat{f}_t) && (89)\\
&\quad= \sum_{s=1}^{t-1}\mathfrak{g}(\tau)^{-1}\mu(f^*(\mathbf{x}_s))\phi(\mathbf{x}_s) + \lambda\alpha^* - \sum_{s=1}^{t-1}\mathfrak{g}(\tau)^{-1}\mu(\widehat{f}_r(\mathbf{x}_s))\phi(\mathbf{x}_s) - \lambda\widehat{\alpha}_t && (90)\\
&\quad= \sum_{s=1}^{t-1}\mathfrak{g}(\tau)^{-1}(-y_s + \mu(f^*(\mathbf{x}_s))\phi(\mathbf{x}_s)) + \lambda\alpha^* && (91)\\
&\qquad- \left(\underbrace{\sum_{s=1}^{t-1}\mathfrak{g}(\tau)^{-1}(-y_s + \mu(\widehat{f}_t(\mathbf{x}_s)))\phi(\mathbf{x}_s) + \lambda\widehat{\alpha}_t}_{\nabla\mathcal{L}_t(\widehat{f}_t)=0}\right) && (92)\\
&\quad= -\mathfrak{g}(\tau)^{-1}\sum_{s=1}^{t-1}\epsilon_s\phi(\mathbf{x}_s) + \lambda\alpha^*, && (93)
\end{aligned}$$

having exploited the first-order optimality condition for the loss evaluated in the maximum-likelihood estimate, i.e., $\nabla\mathcal{L}_t(\widehat{f}_t) = 0$ and the definition of $\epsilon_s = y_s - \mu(f^*(\mathbf{x}_s))$. Now, by computing the norm, we have:

$$\left\|g_t(f^*) - g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda;f^*)} \leq \mathfrak{g}(\tau)^{-1}\left\|\sum_{s=1}^{t-1}\epsilon_s\phi(\mathbf{x}_s)\right\|_{\widetilde{V}_t^{-1}(\lambda;f^*)} + \lambda\left\|\alpha^*\right\|_{\widetilde{V}_t^{-1}(\lambda;f^*)}. \tag{94}$$

We can immediately bound the second term under Assumption 3.1:

$$\|\alpha^*\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)} = (\alpha^*)^\top\widetilde{V}_t^{-1}(\lambda; f^*)\alpha^* \leq \lambda^{-1}\|\alpha^*\|^2 \leq \lambda^{-1}B^2, \tag{95}$$

since $\widetilde{V}_t^{-1}(\lambda; f^*) \succeq \lambda I$. For the first term, we resort to the self-normalized concentration inequality of Theorem 6.2, recalling that the variance of the noise is $\mathbb{V}\mathrm{ar}[\epsilon_s|\mathcal{F}_{s-1}] = \dot{\mu}(f^*(\mathbf{x}_s))\mathfrak{g}(\tau)^{-1}$. □

Theorem 7.2 (Regret Bound of `GKB-UCB`)**.** *Under Assumptions 3.1, 3.2, 3.3, and 3.4, `GKB-UCB` with the confidence radius $B_t(\delta; f)$ as defined in Lemma 7.1 and $\lambda > 0$, for every $\delta \in (0,1)$, with probability at least $1-\delta$, suffers regret bounded as $R(\texttt{GKB-UCB}, T) = R_{\text{perm}}(T) + R_{\text{trans}}(T)$, where:*

$$R_{\text{perm}}(T) \leq 8(1+2R_sBK)\beta_T(\delta;\mathcal{H})\sqrt{\max\left\{\mathfrak{g}(\tau), \lambda^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}}, \tag{19}$$

$$R_{\text{trans}}(T) \leq 32R_s(1+R_{\dot{\mu}}\kappa_{\mathcal{X}})(1+2R_sBK)^2\beta_T(\delta;\mathcal{H})^2\max\left\{\mathfrak{g}(\tau), \lambda^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*). \tag{20}$$

Proof. We start by performing a second-order Taylor's expansion of the regret:

$$\sum_{t=1}^{T}(\mu(f^*(\mathbf{x}^*)) - \mu(f^*(\mathbf{x}_t))) = \underbrace{\sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))\left(f^*(\mathbf{x}^*) - f^*(\mathbf{x}_t)\right)}_{=:R_1(T)} \tag{96}$$

$$+\underbrace{\sum_{t=1}^{T}\left(\int_{v=0}^{1}(1-v)\ddot{\mu}((1-v)f^*(\mathbf{x}_t)+vf^*(\mathbf{x}^*))\mathrm{d}v\right)\left(f^*(\mathbf{x}^*)-f^*(\mathbf{x}_t)\right)^2}_{=:R_2(T)}. \quad (97)$$

We know that $\widetilde{f}_t \in \mathcal{C}_t(\delta)$ and if the good event $\mathcal{E}_\delta$ holds, we also have $f^* \in \mathcal{C}_t(\delta)$. Using the optimism, we know that $\widetilde{f}_t(\mathbf{x}_t) \geq f^*(\mathbf{x}^*)$. We start by analyzing $R_1(T)$, recalling that $\dot{\mu}(f^*(\mathbf{x}_t)) \geq 0$:

$$\begin{aligned}
R_1(T) &= \sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))\left(f^*(\mathbf{x}^*)-f^*(\mathbf{x}_t)\right) && (98)\\
&= \sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))\left(f^*(\mathbf{x}^*)-f^*(\mathbf{x}_t)\pm\widetilde{f}_t(\mathbf{x}_t)\right) && (99)\\
&\leq \sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))(\widetilde{f}_t(\mathbf{x}_t)-f^*(\mathbf{x}_t)) && (100)\\
&= \sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))\langle\widetilde{\alpha}_t-\alpha^*,\phi(\mathbf{x}_t)\rangle && (101)\\
&\leq \sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))\underbrace{\|\widetilde{\alpha}_t-\alpha^*\|_{\widetilde{V}_t(\lambda;f^*)}}_{\text{(a)}}\underbrace{\|\phi(\mathbf{x}_t)\|_{\widetilde{V}_t^{-1}(\lambda;f^*)}}_{\text{(b)}}, && (102)
\end{aligned}$$

where we decompose the functions as inner products and the Cauchy-Schwarz's inequality. For term (a), we apply Lemma C.4 with $f \leftarrow \widetilde{f}_t, f' \leftarrow f^*, f'' \leftarrow \widehat{f}_t$ and exploit the good event:

$$\begin{aligned}
\|\widetilde{\alpha}_t-\alpha^*\|_{\widetilde{V}_t(\lambda;f^*)} &\leq (1+2R_sBK) && (103)\\
&\quad\cdot\left(\left\|g_t(\widetilde{f}_t)-g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda;\widetilde{f}_t)}+\left\|g_t(f^*)-g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda;f^*)}\right) && (104)\\
&\leq (1+2R_sBK)(B_t(\delta;\widetilde{f}_t)+B_t(\delta;f^*)) && (105)\\
&\leq 2(1+2R_sBK)\beta_T(\delta;\mathcal{H}), && (106)
\end{aligned}$$

having observed that $\beta_T(\delta;\mathcal{H}) \geq \beta_t(\delta;\mathcal{H}) \geq \max\{B_t(\delta;\widetilde{f}_t), B_t(\delta;f^*)\}$. For term (b), we apply Cauchy-Schwarz's inequality:

$$\begin{aligned}
&\sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))\|\phi(\mathbf{x}_t)\|_{\widetilde{V}_t^{-1}(\lambda;f^*)} && (107)\\
&\qquad\leq\sqrt{\mathfrak{g}(\tau)}\sqrt{\sum_{t=1}^{T}\dot{\mu}(f^*(\mathbf{x}_t))}\sqrt{\sum_{t=1}^{T}\mathfrak{g}(\tau)^{-1}\dot{\mu}(f^*(\mathbf{x}_t))\|\phi(\mathbf{x}_t)\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)}}. && (108)
\end{aligned}$$

Recalling that $\mathfrak{g}(\tau)^{-1}\dot{\mu}(f^*(\mathbf{x}_t))\|\phi(\mathbf{x}_t)\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)} = \|\widetilde{\phi}(\mathbf{x}_t;f^*)\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)}$, we can apply an elliptic potential lemma (Lemma C.6 with $M=\max\left\{1,\lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}$), where $\lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2$ is a bound to the maximum value $\|\widetilde{\phi}(\mathbf{x}_t;f^*)\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)}$ can take as:

$$\|\widetilde{\phi}(\mathbf{x}_t;f^*)\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)} = \mathfrak{g}(\tau)^{-1}\dot{\mu}(f^*(\mathbf{x}_t))\|\phi(\mathbf{x}_t)\|^2_{\widetilde{V}_t^{-1}(\lambda;f^*)} \leq \mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\lambda^{-1}, \quad (109)$$

as $\widetilde{V}_t^{-1}(\lambda;f^*) \succeq \lambda I$. Thus, we have:

$$\begin{aligned}
\sum_{t=1}^{T}\|\widetilde{\phi}(\mathbf{x}_t;f^*)\|^2_{\widetilde{V}_t^{-1}(f^*)} &\leq 2\max\left\{1,\lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\log\det(\lambda^{-1}\widetilde{V}_t(f^*)) && (110)\\
&\leq 4\max\left\{1,\lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*). && (111)
\end{aligned}$$

The remaining term can be treated as follows, by means of a Taylor expansion:

$$\sum_{t=1}^{T} \dot{\mu}(f^*(\mathbf{x}_t)) = \sum_{t=1}^{T} \dot{\mu}(f^*(\mathbf{x}^*)) + \sum_{t=1}^{T} \left(\int_{v=0}^{1} \ddot{\mu}((1-v)f^*(\mathbf{x}^*) + vf^*(\mathbf{x}_t)) \right) (f^*(\mathbf{x}_t) - f^*(\mathbf{x}^*)) \tag{112}$$

$$\leq T\dot{\mu}(f^*(\mathbf{x}^*)) + R_s \sum_{t=1}^{T} \left(\int_{v=0}^{1} \dot{\mu}((1-v)f^*(\mathbf{x}^*) + vf^*(\mathbf{x}_t)) \right) (f^*(\mathbf{x}^*) - f^*(\mathbf{x}_t)) \tag{113}$$

$$= \frac{T}{\kappa_*} + R_s \sum_{t=1}^{T} (\mu(f^*(\mathbf{x}^*)) - \mu(f^*(\mathbf{x}_t))) \tag{114}$$

$$= \frac{T}{\dot{\mu}(f^*(\mathbf{x}^*))} + R_s R(\texttt{GKB-UCB}, T) \tag{115}$$

$$= \frac{T}{\kappa_*} + R_s R(\texttt{GKB-UCB}, T). \tag{116}$$

where we exploited $f^*(\mathbf{x}^*) \geq f^*(\mathbf{x}_t)$, the self-concordance property (Assumption 3.4) and mean-value theorem. Putting all together, we get:

$$R_1(T) \leq 4\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H}) \sqrt{\frac{T}{\kappa_*} + R_s R(\texttt{GKB-UCB}, T)} \tag{117}$$

$$\cdot \sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1} R_{\dot{\mu}} K^2\right\} \widetilde{\gamma}_T(f^*)} \tag{118}$$

Let us move to the second term, using optimism and proceeding with the same rationale as before:

$$R_2(T) \leq R_{\dot{\mu}} R_s \sum_{t=1}^{T} (f^*(\mathbf{x}^*) - f^*(\mathbf{x}_t))^2 \tag{119}$$

$$\leq R_{\dot{\mu}} R_s \sum_{t=1}^{T} (\widetilde{f}_t(\mathbf{x}_t) - f^*(\mathbf{x}_t))^2 \tag{120}$$

$$\leq R_{\dot{\mu}} \sum_{t=1}^{T} \|\widetilde{\alpha}_t - \alpha^*\|^2_{\widetilde{V}_t(\lambda; f^*)} \|\phi(\mathbf{x}_t)\|^2_{\widetilde{V}_t^{-1}(\lambda; f^*)} \tag{121}$$

$$\leq 4R_{\dot{\mu}} R_s (1 + 2R_s BK)^2 \beta_T(\delta; \mathcal{H})^2 \sum_{t=1}^{T} \|\phi(\mathbf{x}_t)\|^2_{\widetilde{V}_t^{-1}(\lambda; f^*)} \tag{122}$$

$$\leq 4\mathfrak{g}(\tau) R_{\dot{\mu}} R_s \kappa_{\mathcal{X}} (1 + 2R_s BK)^2 \beta_T(\delta; \mathcal{H})^2 \sum_{t=1}^{T} \mathfrak{g}(\tau)^{-1} \dot{\mu}(f^*(\mathbf{x}_t)) \|\phi(\mathbf{x}_t)\|^2_{\widetilde{V}_t^{-1}(\lambda; f^*)} \tag{123}$$

$$\leq 4\mathfrak{g}(\tau) R_{\dot{\mu}} R_s \kappa_{\mathcal{X}} (1 + 2R_s BK)^2 \beta_T(\delta; \mathcal{H})^2 \sum_{t=1}^{T} \|\widetilde{\phi}(\mathbf{x}_t; f^*)\|^2_{\widetilde{V}_t^{-1}(\lambda; f^*)} \tag{124}$$

$$\leq 16\mathfrak{g}(\tau) R_{\dot{\mu}} R_s \kappa_{\mathcal{X}} (1 + 2R_s BK)^2 \beta_T(\delta; \mathcal{H})^2 \max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1} R_{\dot{\mu}} K^2\right\} \widetilde{\gamma}_T(f^*), \tag{125}$$

having, in addition, exploited the fact that $\kappa_{\mathcal{X}} \geq \dot{\mu}(f^*(\mathbf{x}_t))^{-1}$. Putting all together, we have:

$$R(\texttt{GKB-UCB}, T) = R_1(T) + R_2(T) \tag{126}$$

$$\leq 4\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H}) \sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1} R_{\dot{\mu}} K^2\right\} \widetilde{\gamma}_T(f^*)} \tag{127}$$

$$\cdot \left(\sqrt{\frac{T}{\kappa_*}} + \sqrt{R_s R(\texttt{GKB-UCB}, T)} \right) + R_2(T). \tag{128}$$

Using the polynomial inequality of Proposition 7 of [2] (i.e., $x^2 \leq bx + c = 0 \implies x \leq b + \sqrt{c}$ when $b, c \geq 0$), we have:

$$\sqrt{R(\texttt{GKB-UCB}, T)} \leq 4\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H})\sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{R_s} \tag{129}$$

$$+\sqrt{4\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H})\sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}} + R_2(T)}. \tag{130}$$

Squaring both sides and bounding the square as $(a + b)^2 \leq 2a^2 + 2b^2$, we obtain:

$$R(\texttt{GKB-UCB}, T) \tag{131}$$

$$\leq 2\left(4\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H})\sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{R_s}\right)^2 \tag{132}$$

$$+ 2\left(4\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H})\sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}} + R_2(T)\right) \tag{133}$$

$$\leq 8\sqrt{\mathfrak{g}(\tau)}(1 + 2R_s BK)\beta_T(\delta; \mathcal{H})\sqrt{\max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*)}\sqrt{\frac{T}{\kappa_*}} \tag{134}$$

$$+ 32R_s(1 + R_{\dot{\mu}}\kappa_{\mathcal{X}})\mathfrak{g}(\tau)(1 + 2R_s BK)^2\beta_T(\delta; \mathcal{H})^2 \max\left\{1, \lambda^{-1}\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}K^2\right\}\widetilde{\gamma}_T(f^*). \tag{135}$$

We get the result by defining $R_{\text{perm}}(T)$ and $R_{\text{trans}}(T)$ as in the statement. □

B.4 Proofs of Appendix A

Lemma B.2 (Confidence Set)**.** *Let $t \in \mathbb{N}$, $f \in \mathcal{H}$, and $\delta \in (0, 1)$. Then, it holds that $\mathcal{C}_t(\delta) \subseteq \mathcal{D}_t(\delta)$. Furthermore, under the good event $\mathcal{E}_\delta$, for every $f = \langle \alpha, \phi \rangle \in \mathcal{D}_t(\delta)$, we have:*

$$\|\alpha - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} \leq (2 + 2R_s BK)\sqrt{\beta_t(\delta; \mathcal{H})}\left(\sqrt{\beta_t(\delta; \mathcal{H})} + \sqrt{2}\right). \tag{136}$$

Proof. Following the same derivation of Lemma 2 of [2], based on Taylor's expansion and using the definitions of G_t and $\widetilde{G}_t$ in Appendix C. We have:

$$\mathcal{L}_t(f) - \mathcal{L}_t(\widehat{f}_t) \tag{137}$$

$$= (\alpha - \widehat{\alpha}_t)^\top \underbrace{\nabla\mathcal{L}_t(\widehat{f}_t)}_{=0} + (\alpha - \widehat{\alpha}_t)^\top \left(\int_{v=0}^{1}(1 - v)\widetilde{V}_t(\lambda; \widehat{f}_t + v(f - \widehat{f}_t))\mathrm{d}v\right)(\alpha - \widehat{\alpha}_t) \tag{138}$$

$$= \|\alpha - \widehat{\alpha}_t\|^2_{\widetilde{G}_t(\widehat{f}_t, f)} \tag{139}$$

$$\leq \|\alpha - \widehat{\alpha}_t\|^2_{G_t(\widehat{f}_t, f)} \tag{140}$$

$$= \left\|g_t(f) - g_t(\widehat{f}_t)\right\|^2_{\widetilde{G}_t^{-1}(\widehat{f}_t, f)} \tag{141}$$

$$\leq (1 + 2R_s BK)\left\|g_t(f) - g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda; f)}, \tag{142}$$

where we used Equation (162) and (167). Thus, let $f \in \mathcal{C}_t(\delta)$, we have that $\left\|g_t(f) - g_t(\widehat{f}_t)\right\|_{\widetilde{V}_t^{-1}(\lambda; f)} \leq B_t(\delta; f) \leq B_t(\delta; \mathcal{H})$ and, consequently, $f \in \mathcal{D}_t(\delta)$.

For the second part, suppose the good event $\mathcal{E}_\delta$ holds and consider $f \in \mathcal{D}_t(\delta)$, we have via Taylor's expansion:

$$\mathcal{L}_t(f) - \mathcal{L}_t(f^*) \tag{143}$$

$$= (\alpha - \alpha^*)^\top \nabla \mathcal{L}_t(f^*) + (\alpha - \alpha^*)^\top \left(\int_{v=0}^{1} (1-v) \widetilde{V}_t(f^* + v(f - f^*); \lambda) \mathrm{d}v \right) (\alpha - \alpha^*) \tag{144}$$

$$= (\alpha - \alpha^*)^\top \nabla \mathcal{L}_t(f^*) + \|\alpha - \alpha^*\|^2_{\widetilde{G}_t(f^*, f)} \tag{145}$$

$$\geq (\alpha - \alpha^*)^\top \nabla \mathcal{L}_t(f^*) + (2 + 2R_s BK)^{-1} \|\alpha - \alpha^*\|^2_{\widetilde{V}_t(\lambda; f^*)}, \tag{146}$$

where we used Equation (168). Thus, we have:

$$\|\alpha - \alpha^*\|^2_{\widetilde{V}_t(f^*; \lambda)} \tag{147}$$

$$\leq (2 + 2R_s BK)(\mathcal{L}_t(f) - \mathcal{L}_t(f^*)) + (2 + 2R_s BK)(\alpha - \alpha^*)^\top \nabla \mathcal{L}_t(f^*) \tag{148}$$

$$\leq (2 + 2R_s BK)(\mathcal{L}_t(f) - \mathcal{L}_t(\widehat{f}_t)) + (2 + 2R_s BK)(\mathcal{L}_t(f^*) - \mathcal{L}_t(\widehat{f}_t)) \tag{149}$$

$$+ (2 + 2R_s BK) \|\alpha - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} \|\nabla \mathcal{L}_t(f^*)\|_{\widetilde{V}_t^{-1}(\lambda; f^*)} \tag{150}$$

$$\leq 2(2 + 2R_s BK)(1 + 2R_s BK) B_t(\delta; \mathcal{H}) + (2 + 2R_s BK) \|\alpha - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} B_t(\delta; f^*). \tag{151}$$

where we used the fact that $\mathcal{L}_t(f) \geq \mathcal{L}_t(\widehat{f}_t) \wedge \mathcal{L}_t(f^*) \geq \mathcal{L}_t(\widehat{f}_t)$, that $f^* \in \mathcal{C}_t(\delta) \subseteq \mathcal{D}_t(\delta)$ under the good event and $f \in \mathcal{D}_t(\delta)$, and that:

$$\|\nabla \mathcal{L}_t(f^*)\|_{\widetilde{V}_t^{-1}(\lambda; f^*)} = \|g_t(f^*) - g_t(\widehat{f}_t)\|_{\widetilde{V}_t^{-1}(\lambda; f^*)} \leq B_t(\delta; f^*). \tag{152}$$

holding under the good event. By the choice of confidence radius and bounding $B_t(\delta; f^*) \leq B_t(\delta; \mathcal{H}) \leq \beta_t(\delta; \mathcal{H})$, we have the second-degree inequality:

$$\|\alpha - \alpha^*\|^2_{\widetilde{V}_t(\lambda; f^*)} \leq 2(2 + 2R_s BK)(1 + 2R_s BK) \beta_t(\delta; \mathcal{H}) \tag{153}$$

$$+ (2 + 2R_s BK) \|\alpha - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} \beta_t(\delta; \mathcal{H}). \tag{154}$$

Using the polynomial inequality of Proposition 7 of [2] (i.e., $x^2 \leq bx + c = 0 \implies x \leq b + \sqrt{c}$ when $b, c \geq 0$), we have:

$$\|\alpha - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} \leq \sqrt{2(2 + 2R_s BK)(1 + 2R_s BK) \beta_t(\delta; \mathcal{H})} + (2 + 2R_s BK) \beta_t(\delta; \mathcal{H}) \tag{155}$$

$$\leq (2 + 2R_s BK) \sqrt{\beta_t(\delta; \mathcal{H})} \left(\sqrt{\beta_t(\delta; \mathcal{H})} + \sqrt{2} \right). \tag{156}$$

having bounded $1 + 2R_s BK \leq 2 + 2R_s BK$. □

Theorem A.1 (Regret Bound of `Eff-GKB-UCB`)**.** *Under Assumptions 3.1, 3.2, 3.3, and 3.4,* `GKB-UCB` *with confidence radius* $(1 + 2R_s BK) B_t(\delta; \mathcal{H})$ *and* $\lambda > 0$*, for every* $\delta \in (0, 1)$*, with probability at least* $1 - \delta$*, suffers regret bounded as:* $R(\texttt{Eff-GKB-UCB}, T) = R_{\text{perm}}(T) + R_{\text{trans}}(T)$*, where:*

$$R_{\text{perm}}(T) \leq 4\sqrt{\max\{\mathfrak{g}(\tau), \lambda^{-1} R_{\dot{\mu}} K^2\}} (2 + 2R_s BK) \sqrt{\beta_T(\delta; \mathcal{H})} \left(\sqrt{\beta_T(\delta; \mathcal{H})} + 2 \right) \sqrt{\widetilde{\gamma}_T(f^*)} \sqrt{\frac{T}{\kappa_*}},$$

$$R_{\text{trans}}(T) \leq 8R_s (1 + R_{\dot{\mu}} \kappa_{\mathcal{X}}) \max\left\{\mathfrak{g}(\tau), \lambda^{-1} R_{\dot{\mu}} K^2\right\} (2 + 2R_s BK)^2 \beta_T(\delta; \mathcal{H}) \left(\sqrt{\beta_T(\delta; \mathcal{H})} + 2 \right)^2 \widetilde{\gamma}_T(f^*).$$

Proof. The proof follows the same steps as Theorem 7.2, with the only difference that we exploit the bound of Lemma B.2:

$$\|\alpha - \alpha^*\|_{\widetilde{V}_t(\lambda; f^*)} \leq (2 + 2R_s BK) \sqrt{\beta_t(\delta; \mathcal{H})} \left(\sqrt{\beta_t(\delta; \mathcal{H})} + \sqrt{2} \right). \tag{157}$$

□

C Technical Lemmas

In this section, we introduce some technical concepts and lemmas to be used in the analysis. We consider $\mathbf{x} \in \mathcal{X}$ and $f = \langle \alpha, \phi \rangle, f' = \langle \alpha, \phi' \rangle \in \mathcal{H}$, we define the following quantities, analogous to those of [2]:

$$\xi(\mathbf{x}, f, f') \coloneqq \int_{v=0}^{1} \dot{\mu}((1 - v) f(\mathbf{x}) + v f'(\mathbf{x})) \mathrm{d}v, \tag{158}$$

$$\widetilde{\xi}(\mathbf{x}, f, f') \coloneqq \int_{v=0}^{1} (1-v)\dot{\mu}((1-v)f(\mathbf{x}) + vf'(\mathbf{x}))\mathrm{d}v, \tag{159}$$

$$G_t(f, f') \coloneqq \sum_{s=1}^{t-1} \frac{\xi(\mathbf{x}, f, f')}{\mathfrak{g}(\tau)} \phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top + \lambda I, \tag{160}$$

$$\widetilde{G}_t(f, f') \coloneqq \sum_{s=1}^{t-1} \frac{\widetilde{\xi}(\mathbf{x}, f, f')}{\mathfrak{g}(\tau)} \phi(\mathbf{x}_s)\phi(\mathbf{x}_s)^\top + \lambda I. \tag{161}$$

We have that $\xi(\mathbf{x}, f, f') \geq \widetilde{\xi}(\mathbf{x}, f, f')$ and, consequently, we have that $G_t(f, f') \succeq \widetilde{G}_t(f, f')$. Thanks to the mean-value theorem and the definition of function $g_t(f)$, we have that:

$$g_t(f) - g_t(f') = G_t(f, f')(\alpha - \alpha'). \tag{162}$$

Using Assumption 3.4, we can easily extend Lemmas 7 and 8 of [2].

Lemma C.1 (Extension of Lemma 7 of [2])**.** *Let $\mathcal{Z} \subset \mathbb{R}$ be any bounded interval of $\mathbb{R}$ and let $f : \mathcal{Z} \to \mathbb{R}$ be a monothonically non-decreasing function such that $|\ddot{f}| \leq R_s \dot{f}$. Then, for every $z_1, z_2 \in \mathcal{Z}$:*

$$\int_{v=0}^{1} \dot{f}(z_1 + v(z_2 - z_1))\mathrm{d}v \geq \frac{\dot{f}(z)}{1 + R_s|z_1 - z_2|}, \quad \forall z \in \{z_1, z_2\}. \tag{163}$$

Proof. Immediately follows from the same steps of [2, Lemma 7]. □

Lemma C.2 (Extension of Lemma 8 of [2])**.** *Let $\mathcal{Z} \subset \mathbb{R}$ be any bounded interval of $\mathbb{R}$ and let $f : \mathcal{Z} \to \mathbb{R}$ be a monothonically non-decreasing function such that $|\ddot{f}| \leq R_s \dot{f}$. Then, for every $z_1, z_2 \in \mathcal{Z}$:*

$$\int_{v=0}^{1} (1-v)\dot{f}(z_1 + v(z_2 - z_1))\mathrm{d}v \geq \frac{\dot{f}(z_1)}{2 + R_s|z_1 - z_2|}. \tag{164}$$

Proof. See [17, Lemma D.1]. □

From Lemma C.1 and Lemma C.2, we immediatly have:

$$\xi(\mathbf{x}, f, f') \coloneqq \int_{v=0}^{1} \dot{\mu}((1-v)f(\mathbf{x}) + vf'(\mathbf{x}))\mathrm{d}v \geq \frac{\dot{\mu}(\overline{f})}{1 + R_s|f(\mathbf{x}) - f'(\mathbf{x})|}, \text{ for } \overline{f} \in \{f(\mathbf{x}), f'(\mathbf{x})\}. \tag{165}$$

$$\widetilde{\xi}(\mathbf{x}, f, f') \coloneqq \int_{v=0}^{1} (1-v)\dot{\mu}((1-v)f(\mathbf{x}) + vf'(\mathbf{x}))\mathrm{d}v \geq \frac{\dot{\mu}(f(\mathbf{x}))}{2 + R_s|f(\mathbf{x}) - f'(\mathbf{x})|}. \tag{166}$$

Moreover, under Assumptions 3.2 and 3.1, we have that $|f(\mathbf{x}) - f'(\mathbf{x})| \leq 2\|f\|_\infty \leq 2BK$. This allows us to write:

$$G_t(f, f') \succeq (1 + 2R_s BK)^{-1} \widetilde{V}_t(\lambda; \overline{f}), \qquad \text{for } \overline{f} \in \{f, f'\}, \tag{167}$$

$$\widetilde{G}_t(f, f') \succeq (2 + 2R_s BK)^{-1} \widetilde{V}_t(\lambda; f). \tag{168}$$

Lemma C.3. *Let $f \in \mathcal{H}$, $\widetilde{V}_t(\lambda; f)$ and $V_t(\lambda)$ defined as in the main paper. The following semidefinite inequalities holds:*

$$\min\{1, \mathfrak{g}(\tau)R_{\dot{\mu}}^{-1}\}\widetilde{V}_t(\lambda; f) \preceq V_t(\lambda) \preceq \max\{1, \mathfrak{g}(\tau)\kappa_{\mathcal{X}}(f)\}\widetilde{V}_t(\lambda; f), \tag{169}$$

where $\kappa_{\mathcal{X}}(f) = \sup_{\mathbf{x} \in \mathcal{X}} \frac{1}{\dot{\mu}(f(\mathbf{x}))}$

Proof. For one inequality, we have:

$$\widetilde{V}_t(\lambda; f) = \sum_{s=1}^{t-1} \frac{\dot{\mu}(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)} \phi(\mathbf{x})\phi(\mathbf{x})^\top + \lambda I \tag{170}$$

$$\succeq \mathfrak{g}(\tau)^{-1}\kappa_{\mathcal{X}}(f)^{-1}\sum_{s=1}^{t-1}\phi(\mathbf{x})\phi(\mathbf{x})^{\top}+\lambda I \tag{171}$$

$$\succeq \min\left\{1,\mathfrak{g}(\tau)^{-1}\kappa_{\mathcal{X}}(f)^{-1}\right\}\left(\sum_{s=1}^{t-1}\phi(\mathbf{x})\phi(\mathbf{x})^{\top}+\lambda I\right) \tag{172}$$

$$= \min\left\{1,\mathfrak{g}(\tau)^{-1}\kappa_{\mathcal{X}}(f)^{-1}\right\}V_t(\lambda). \tag{173}$$

For the other inequality, we have:

$$\widetilde{V}_t(\lambda;f)=\sum_{s=1}^{t-1}\frac{\dot{\mu}(f(\mathbf{x}_s))}{\mathfrak{g}(\tau)}\phi(\mathbf{x})\phi(\mathbf{x})^{\top}+\lambda I \tag{174}$$

$$\preceq \mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}\sum_{s=1}^{t-1}\phi(\mathbf{x})\phi(\mathbf{x})^{\top}+\lambda I \tag{175}$$

$$\preceq \max\left\{1,\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}\right\}\left(\sum_{s=1}^{t-1}\phi(\mathbf{x})\phi(\mathbf{x})^{\top}+\lambda I\right) \tag{176}$$

$$= \max\left\{1,\mathfrak{g}(\tau)^{-1}R_{\dot{\mu}}\right\}V_t(\lambda). \tag{177}$$

□

Lemma C.4. *Let $f=\langle\alpha,\phi\rangle, f'=\langle\alpha',\phi\rangle\in\mathcal{H}$, then for every $f''=\langle\alpha'',\phi\rangle\in\mathcal{H}$, it holds that:*

- $\|\alpha-\alpha'\|_{\widetilde{V}_t(\lambda;f')}\leq(1+2R_sBK)\left(\|g_t(f)-g_t(f'')\|_{\widetilde{V}_t^{-1}(\lambda;f)}+\|g_t(f')-g_t(f'')\|_{\widetilde{V}_t^{-1}(\lambda;f')}\right)$;
- $\|\alpha-\alpha'\|_{V_t(\lambda)}\leq(1+2R_sBK)\max\{1,\mathfrak{g}(\tau)\kappa_{\mathcal{X}}(f')\}\cdot$
$\cdot\left(\|g_t(f)-g_t(f'')\|_{V_t^{-1}(\lambda)}+\|g_t(f')-g_t(f'')\|_{V_t^{-1}(\lambda)}\right)$.

Proof. From the mean-value theorem (Equation 162), we have:

$$g_t(f)-g_t(f')=G_t(f,f')(\alpha-\alpha'). \tag{178}$$

The first statement follows the same derivation of Proposition 4 of [2], with the only care of applying Equation (167). The second statement starts from the following intermediate passage of the proof of Proposition 4 of [2]:

$$\|\alpha-\alpha'\|_{\widetilde{V}_t(\lambda;f')}\leq\sqrt{1+2R_sBK}\left(\|g_t(f)-g_t(f'')\|_{G_t^{-1}(f,f')}+\|g_t(f')-g_t(f'')\|_{G_t^{-1}(f,f')}\right) \tag{179}$$

$$\leq(1+2R_sBK)\left(\|g_t(f)-g_t(f'')\|_{\widetilde{V}_t^{-1}(\lambda;f')}+\|g_t(f')-g_t(f'')\|_{\widetilde{V}_t^{-1}(\lambda;f')}\right). \tag{180}$$

Then, we use the semidefinite inequality $\widetilde{V}_t(\lambda;f')\succeq\max\{1,\mathfrak{g}(\tau)\kappa_{\mathcal{X}}(f')\}^{-1}V_t(\lambda)$ (Lemma C.3):

$$\|\alpha-\alpha'\|_{\widetilde{V}_t(\lambda;f')}\geq\max\{1,\mathfrak{g}(\tau)\kappa_{\mathcal{X}}(f')\}^{-1/2}V_t\|\alpha-\alpha'\|_{V_t(\lambda)} \tag{181}$$

$$\|g_t(f)-g_t(f'')\|_{\widetilde{V}_t^{-1}(\lambda;f')}\leq\max\{1,\mathfrak{g}(\tau)\kappa_{\mathcal{X}}(f')\}^{1/2}\|g_t(f)-g_t(f'')\|_{V_t^{-1}(\lambda)}. \tag{182}$$

□

Lemma C.5. *Let $t\in\mathbb{N}$, $f\in\mathcal{H}$, $\mathbf{K}_t$ and $\widetilde{\mathbf{K}}_t(f)$ defined as in the main paper. It holds that:*

$$\log\det(\mathbf{I}_t+\lambda^{-1}\widetilde{\mathbf{K}}_t(f))\leq\log\det(\mathbf{I}_t+\lambda^{-1}R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\mathbf{K}_t) \tag{183}$$

$$\leq\max\{1,R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\log\det(\mathbf{I}_t+\lambda^{-1}\mathbf{K}_t). \tag{184}$$

Proof. We can look at matrix $\widetilde{\mathbf{K}}_t(f)$ as follows:

$$\widetilde{\mathbf{K}}_t(f)=\mathfrak{g}(\tau)^{-1}\mathbf{M}(f)^{1/2}\mathbf{K}_t\mathbf{M}(f)^{1/2}, \tag{185}$$

where $\mathbf{M}(f) = \operatorname{diag}((\dot{\mu}(f(\mathbf{x}_s))_{s\in[\![t-1]\!]})$ is a diagonal matrix. Using Horn's inequality for eigenvalues [34], we have that for every $i \in [\![t-1]\!]$:

$$\lambda_i(\widetilde{\mathbf{K}}_t(f)) \leq \lambda_i(\mathbf{K}_t) \max_{s\in[\![t-1]\!]} \dot{\mu}(f(\mathbf{x}_s))\mathfrak{g}(\tau)^{-1} \leq \lambda_i(\mathbf{K}_t) R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}. \tag{186}$$

Furthermore, using Weyl's inequality for eigenvalues, we have for $i \in [\![t-1]\!]$:

$$\begin{aligned}
\lambda_i(\mathbf{I}_t + \lambda^{-1}\widetilde{\mathbf{K}}_t(f)) &\leq 1 + \lambda^{-1}\lambda_i(\widetilde{\mathbf{K}}_t(f)) && (187)\\
&\leq 1 + \lambda^{-1} R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\lambda_i(\mathbf{K}_t) && (188)\\
&\leq 1 + \lambda^{-1}\max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}\lambda_i(\mathbf{K}_t) && (189)\\
&\leq \left(1 + \lambda^{-1}\lambda_i(\mathbf{K}_t)\right)^{\max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}} && (190)\\
&\leq \lambda_i(1 + \lambda^{-1}\mathbf{K}_t)^{\max\{1, R_{\dot{\mu}}\mathfrak{g}(\tau)^{-1}\}}, && (191)
\end{aligned}$$

where we exploited the inequality $1 + ab \leq (1+b)^a$ for $b \geq 0$ and $a \geq 1$. The statement is obtained passing to the determinant and to its logarithm. □

Lemma C.6 (Elliptic Potential Lemma (slightly extended))**.** *Let $(y_t)_{t\geq 1}$ be a sequence, let $M \geq 1$, and $V_t(\lambda) = \sum_{s=1}^{t-1} y_s y_s^\top + \lambda I$. For every $T \geq 1$, it holds that:*

$$\sum_{t=1}^{T} \min\{M, \|y_t\|_{V_t^{-1}(\lambda)}\}^2 \leq 2M \log\det(\lambda^{-1}V_t(\lambda)). \tag{192}$$

Proof. We follow the steps of Lemma 12 of [2]. Using the inequality $\min\{1, u\} \leq 2\log(1+u)$ for every $u \geq 0$, we have:

$$\begin{aligned}
\sum_{t=1}^{T} \min\{M, \|y_t\|^2_{V_t^{-1}(\lambda)}\} &= M\sum_{t=1}^{T} \min\{1, M^{-1}\|y_t\|^2_{V_t^{-1}(\lambda)}\} && (193)\\
&\leq 2M\sum_{t=1}^{T} \log\left(1 + M^{-1}\|y_t\|^2_{V_t^{-1}(\lambda)}\right) && (194)\\
&\leq 2M\sum_{t=1}^{T} \log\left(1 + \|y_t\|^2_{V_t^{-1}(\lambda)}\right), && (195)
\end{aligned}$$

having exploited that $M \geq 1$. Now the last equation can be bounded following the usual steps of [2], to obtain:

$$\sum_{t=1}^{T} \log\left(1 + \|y_t\|^2_{V_t^{-1}(\lambda)}\right) \leq \log\det(\lambda^{-1}V_t(\lambda)). \tag{196}$$

□

Lemma C.7. *Let S_t and $\widetilde{V}_t(\lambda)$ defined as in Theorem 6.2. The following inequalities hold:*

$$\|S_t\|^2_{\widetilde{V}_t^{-1}(\lambda)} \leq \frac{(t-1)^2K^2R^2}{\lambda}, \qquad \log\det(\lambda^{-1}\widetilde{V}_t(\lambda)) \leq (t-1)\log\left(1 + \frac{K^2R^2}{\lambda}\right). \tag{197}$$

Proof. For the first inequality, we proceed as follows:

$$\|S_t\|^2_{\widetilde{V}_t^{-1}(\lambda)} \leq \|S_t\|^2_{\lambda^{-1}I} \leq \lambda^{-1}\left(\sum_{s=1}^{t-1} |\epsilon_s|\|\phi(\mathbf{x}_s)\|\right)^2 \leq \frac{(t-1)^2K^2R^2}{\lambda}. \tag{198}$$

For the second inequality, we proceed as follows:

$$\begin{aligned}
\det(\lambda^{-1}\widetilde{V}_t(\lambda)) &= \lambda^{-(t-1)}\det(\widetilde{\mathbf{K}}_t(\lambda)) && (199)\\
&\leq \lambda^{-(t-1)}\left(\frac{1}{t-1}\operatorname{tr}(\widetilde{\mathbf{K}}_t(\lambda))\right)^{t-1} && (200)
\end{aligned}$$

$$\leq \lambda^{-(t-1)}(\lambda + K^2R^2)^{(t-1)} \tag{201}$$

$$= \left(1 + \frac{K^2R^2}{\lambda}\right)^{(t-1)} . \tag{202}$$

having applied the identity of Equation (1), the determinant-trace inequality and bounded $\mathrm{tr}(\widetilde{\mathbf{K}}_t(\lambda)) \leq (t-1)(\lambda + K^2R^2)$, since the diagonal elements of $\widetilde{\mathbf{K}}_t(\lambda)$ are of the form $\lambda + \sigma(\mathbf{x})k(\mathbf{x}, \mathbf{x}')\sigma(\mathbf{x}') \leq \lambda + R^2K^2$, being the variance bounded by the square of the range. □