# Online learning with kernel losses

**Aldo Pacchiano** [*1] **Niladri S. Chatterji** [*1] **Peter L. Bartlett** [1]

## Abstract

We present a generalization of the adversarial linear bandits framework, where the underlying losses are kernel functions (with an associated reproducing kernel Hilbert space) rather than linear functions. We study a version of the exponential weights algorithm and bound its regret in this setting. Under conditions on the eigen-decay of the kernel we provide a sharp characterization of the regret for this algorithm. When we have polynomial eigen-decay ($\mu_j \leq \mathcal{O}(j^{-\beta})$), we find that the regret is bounded by $\mathcal{R}_n \leq \mathcal{O}(n^{\beta/2(\beta-1)})$. While under the assumption of exponential eigen-decay ($\mu_j \leq \mathcal{O}(e^{-\beta j})$) we get an even tighter bound on the regret $\mathcal{R}_n \leq \tilde{\mathcal{O}}(n^{1/2})$. When the eigen-decay is polynomial we also show a *non-matching* minimax lower bound on the regret of $\mathcal{R}_n \geq \Omega(n^{(\beta+1)/2\beta})$ and a lower bound of $\mathcal{R}_n \geq \Omega(n^{1/2})$ when the decay in the eigen-values is exponentially fast.

We also study the full information setting when the underlying losses are kernel functions and present an adapted exponential weights algorithm and a conditional gradient descent algorithm.

## 1. Introduction

In adversarial online learning, a player interacts with an unknown and arbitary adversary in a sequence of rounds. At each round, the player chooses an action from an action space and incurs a loss associated with that chosen action. The loss functions are determined by the adversary and are fixed at the beginning of each round. After choosing an action the player observes some feedback, which can help guide the choice of actions in subsequent rounds. The most common feedback model is the *full information* model, where the player has access to the entire loss function at the end of each round. Another, more challenging feedback model is the *partial information* or *bandit* feedback model where the player at the end of the round just observes the loss associated with the action chosen in that particular round. There are also other feedback models in between and beyond the full and bandit information models, many of which have also been studied in detail. A figure of merit that is often used to judge online learning algorithms is the notion of *regret*, which compares the players actions to the best single action in hindsight (defined formally in Section 1.2).

When the underlying action space is a continuous and compact (possibly convex) set and the losses are linear or convex functions over this set; there are many algorithms known that attain sub-linear and sometimes optimal regret in both these feedback settings. In this work we present a generalization of the well studied adversarial online linear learning framework. In our paper, at each round the player selects an action $a \in \mathcal{A} \subset \mathbb{R}^d$. This action is mapped to an element in a reproducing kernel Hilbert space (RKHS) generated by a mapping $\mathcal{K}(\cdot,\cdot)$. The function $\mathcal{K}(\cdot,\cdot)$ is a kernel map, that is, it can thought as an inner product of an appropriate Hilbert space $\mathcal{H}$. The kernel map can be expressed as $\mathcal{K}(x,y) = \langle \Phi(x), \Phi(y)\rangle_{\mathcal{H}}$, where $\Phi(\cdot) \in \mathbb{R}^D$ is the associated feature map.

Thus at each round the loss is $\langle \Phi(a), w\rangle_{\mathcal{H}}$, where $w \in \mathcal{H}$ is the adversary's action. In the full information setting, as feedback, the player has access to the entire adversarial loss function $\langle \cdot, w\rangle_{\mathcal{H}}$. In the bandit setting the player is just presented with the value of the loss, $\langle \Phi(a), w\rangle_{\mathcal{H}}$.

Notice that this class of losses is much more general than ordinary linear losses and includes potentially non-linear and non-convex losses like:

1. Linear Losses: $\langle a, w\rangle_{\mathcal{H}} = a^\top w$. This loss is well studied in both the bandit and full information setting. We shall see that our regret bounds will match the bounds established in the literature for these losses.
2. Quadratic Losses: $\left\langle \phi(a), \binom{W}{b} \right\rangle_{\mathcal{H}} = a^\top W a + b^\top a$, where $W$ is a symmetric matrix and $b$ is a vector. Convex quadratic losses have been well studied under full information feedback as the online eigenvector decomposition problem. Our work establishes regret bounds

[*]Equal contribution [1]University of California Berkeley. Correspondence to: Aldo Pacchiano <pacchiano@berkeley.edu>, Niladri S. Chatterji <chatterji@berkeley.edu>.

*Proceedings of the $36^{th}$ International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019. 

in the full information setting and also under the mostly unexplored bandit feedback.

3. Gaussian Losses: $\langle\Phi(a),\Phi(y)\rangle_{\mathcal{H}} = \exp\left(-\|a-y\|_2^2/2\sigma^2\right)$. We provide regret bounds of kernel losses not commonly studied before like Gaussian losses that provide a different loss profile than a linear or convex loss.

4. Polynomial Losses: $\langle\Phi(a),\Phi(y)\rangle_{\mathcal{H}} = (1+a^\top y)^2$ for example. We also provide regret bounds for polynomial kernel losses which are potentially (non-convex) under both partial and full information settings. Specifically in the full information setting we study posynomial losses (discussion in Section 4.3).

### 1.1. Related Work

Adversarial online convex bandits that was introduced and first studied by (33; 22). The problem most closely related to our work is the case when the losses are linear introduced earlier by (35; 7). In this setting (20; 18; 13) proposed the EXP 2 (Expanded Exp) algorithm under different choices of exploration distributions. (20) worked with the uniform distribution over the barycentric spanner of the set, in (18) this distribution was the uniform distribution over the set and in (13) they use the exploration distribution given by John's theorem that leads to a regret bound of $\mathcal{O}((dn\log(N))^{1/2})$, where $N$ is the number of actions, $n$ is the number of rounds and $d$ is the dimension of the losses. For this same problem when the set $\mathcal{A}$ is convex and compact, (1) analyzed Mirror descent to get a regret bound of $\mathcal{O}(d\sqrt{\theta n\log(n)})$ for some $\theta>0$. For the case with general convex losses with bandit feedback recently (15) proposed a poly-time algorithm that has a regret guarantee of $\tilde{\mathcal{O}}(d^{9.5}\sqrt{n})$, which is optimal in its dependence on the number of rounds $n$. Previous work on this problem includes, (2; 41; 27; 21; 14; 28) in the adversarial setting under different assumptions on the structure of the convex losses and by (3) who studied this problem in the stochastic setting[1]. (46) study stochastic kernelized contextual bandits with a modified UCB algorithm to obtain a regret bound similar to ours, $\mathcal{R}_n \le \sqrt{\tilde{d}n}$ where $\tilde{d}$ is the effective dimension dependent on the eigen-decay of the kernel. This problem was also studied previously for loss functions drawn from Gaussian processes in (44). Online learning under bandit feedback has also been studied when the losses are non-parametric, for example when the losses are Lipschitz (16; 40).

In the full information case, the online optimization framework with convex losses was first introduced by (49). The conditional gradient descent algorithm (a modification of which we study in this work) for convex losses in this setting was introduced and analyzed by (31) and then improved subsequently by (26). The exponential weights algorithm which we modify and use multiple times in this paper has a rich history and has been applied to various online as well as offline settings. The particular with the losses being convex quadratic functions has been well studied in the full information setting. This problem is also called online eigenvector decomposition or online PCA. Very recently (4) established a regret bound of $\tilde{\mathcal{O}}(\sqrt{n})$ for the problem by presented an efficient algorithm that achieves this rate – a modified exponential weights strategy, follow the compressed leader. Previous results for this problem were established in both adversarial and stochastic settings by modifications of exponential weights, gradient descent and follow the perturbed leader algorithms (6; 45; 47; 48; 32; 23).

In the full information setting there has also been work on analyzing gradient descent and mirror descent in RKHS spaces (36; 8). However, in these papers the player is allowed to play any action in a bounded set in Hilbert space, while in our paper the player is constrained to only play rank one actions, that is the player chooses an action in $\mathcal{A}$ which gets mapped to an action in the RKHS.

#### Contributions

Our primary contribution is to extend the linear bandits framework to more general classes of kernel losses. We present an algorithm in this setting and provide a regret bound for the same. We provide a more detailed analysis of the regret when we make assumptions on the eigen-decay of the kernel. Particularly when we assume the polynomial eigen-decay of the kernel ($\mu_j \le \mathcal{O}(j^{-\beta})$) we can guarantee the regret is bounded as $\mathcal{R}_n \le \mathcal{O}(n^{\frac{\beta}{2(\beta-1)}})$. Under exponential eigendecay we can guarantee an even sharper bound on the regret of $\mathcal{R}_n \le \tilde{\mathcal{O}}(n^{1/2})$. We also provide a minimax lower bound on the regret of $\mathcal{R}_n \ge \Omega(n^{(\beta+1)/2\beta})$ and $\mathcal{R}_n \ge \Omega(n^{1/2})$ under the polynomial and exponential decay eigen-decay assumptions respectively. We analyze an exponential weights algorithm and a conditional gradient algorithm for the full information case where we don't need to assume any conditions on the eigen-decay. Finally we provide a couple of applications of our framework – (i) general quadratic losses (not necessarily convex) with linear terms which we can solve efficiently in the full information setting, (ii) we provide a computationally efficient algorithm when the underlying losses are posynomial (special class of polynomials).

#### Organization of the Paper

In the next section we introduce the notation and definitions. In Section 2 we present our algorithm under bandit feedback and present regret bounds for this algorithm. In Section 3 we study the problem in the full information setting. In Section

[1] For an extended bibliography of the work on online convex bandits see (15).

4 we present two applications of our framework, and prove that our algorithms are computationally efficient in these settings. All the proofs, technical details and experiments are relegated to the appendix.

### 1.2. Notation, main definitions and setting

Here we introduce definitions and notational conventions used throughout the paper.

In each round $t = \{1, \dots, n\}$, the player chooses an action vector $\{a_t\}_{t=1}^n \in \mathcal{A} \subset \mathbb{R}^d$. The underlying kernel function at each round is $\mathcal{K}(\cdot,\cdot)$ which is a map from $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that it is a *kernel map* and has an associated separable reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ with an inner product $\langle \cdot,\cdot \rangle_{\mathcal{H}}$ (for more properties of kernel maps and RKHS see (42)). Let $\Phi(\cdot) : \mathbb{R}^d \mapsto \mathbb{R}^D$ denote a *feature map* of $\mathcal{K}(\cdot,\cdot)$ such that for every $x, y$ we have $\mathcal{K}(x,y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$. Note that the dimension of the RKHS, $D$ could be infinite (for example in the Gaussian kernel over $[0,1]^d$).

We let the adversary choose a vector in $\mathcal{H}$, $w_t \in \mathcal{W} \subset \mathbb{R}^D$ and at each round the loss incurred by the player is $\langle \Phi(a_t), w_t \rangle_{\mathcal{H}}$. We assume that the adversary is oblivious, that is, it is a function of the previous actions of the player $(a_1, \dots, a_{t-1})$ but unaware of the randomness used to generate $a_t$. We let the size of the sets $\mathcal{A}, \mathcal{W}$ be bounded[2] in *kernel norm*, that is,

$$\sup_{a \in \mathcal{A}} \mathcal{K}(a,a) \le \mathcal{G}^2 \qquad \text{and,} \ \sup_{w \in \mathcal{W}} \langle w, w \rangle_{\mathcal{H}} \le \mathcal{G}^2. \tag{1}$$

Throughout this paper we assume a *rank-one learner*, that is, in each round the player can pick a vector $v \in \mathcal{H}$, such that $v = \Phi(a)$ for some $a \in \mathbb{R}^d$. We now formally define the notion of expected regret.

**Definition 1 (Expected regret)** *The expected regret of an algorithm $\mathcal{M}$ after $n$ rounds is defined as*

$$\mathcal{R}_n = \mathbb{E}_{\mathcal{M}} \left[ \sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle_{\mathcal{H}} - \sum_{t=1}^{n} \langle \Phi(a^*), w_t \rangle_{\mathcal{H}} \right] \tag{2}$$

*where $a^* = \inf_{a \in \mathcal{A}} \{\sum_{t=1}^n \langle \Phi(a), w_t \rangle_{\mathcal{H}}\}$ and the expectation is over the randomness in the algorithm.*

Essentially this amounts to comparing against the *best single action* $a^*$ in hindsight. Our hope will be to find a randomized strategy such that the regret grows sub-linearly with the number of rounds $n$. In what follows we will omit the subscript $\mathcal{H}$ from the subscript of the inner product whenever it is clear from the context that it refers to the RKHS inner product.

[2]We set the bound on the size of both sets to be the same for ease of exposition, but they could be different and would only change the constants in our results.

To establish regret guarantees we will find that it is essential to work with finite dimensional kernels when working under bandit feedback (more details regarding this in the proof of the regret bound of Algorithm 2.3). General kernel maps can have infinite dimensional feature maps thus we will require the construction of a finite dimensional kernel that uniformly approximates the original kernel $\mathcal{K}(\cdot,\cdot)$. This motivates the definition of $\epsilon$-approximate kernels.

**Definition 2 ($\epsilon$-approximate kernels)** *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be two kernels over $\mathcal{A} \times \mathcal{A}$ and let $\epsilon > 0$. We say $\mathcal{K}_2$ is an $\epsilon$-approximation of $\mathcal{K}_1$ if for all $x, y \in \mathcal{A}$, $|\mathcal{K}_1(x,y) - \mathcal{K}_2(x,y)| \le \epsilon$.*

## 2. Bandit Feedback Setting

In this section we present our results on *kernel bandits*. In the bandit setting we assume the player knows the underlying kernel function $\mathcal{K}(\cdot,\cdot)$, however, at each round after the player plays a vector $a_t$ only the value of the loss associated with that action is revealed to the player – $\langle \Phi(a_t), w_t \rangle_{\mathcal{H}}$ – and not the action of the adversary $w_t$. We also assume that the player's action set $\mathcal{A}$ has finite cardinality[3].This is a generalization of the well studied adversarial linear bandits problem. As we will see in subsequent sections to guarantee a bound on the regret in the bandit setting our algorithm will build an estimate of adversary's action $w_t$. This becomes impossible if $w_t$ is infinite dimensional ($D \to \infty$). To circumvent this, we will construct a finite dimensional proxy kernel that is an $\epsilon$-approximation of $\mathcal{K}$.

Whenever no approximate kernel is needed, for example when $D < \infty$ we allow the adversary to be able to choose an action $w_t \in \mathcal{W} \subset \mathbb{R}^D$ without imposing extra requirements on the set $\mathcal{W}$ other than being bounded in $\mathcal{H}$ norm. When $D$ is infinite we impose an additional constraint on the adversary to also select *rank-one* actions at each round, that is, $w_t = \Phi(y_t)$ where $y_t \in \mathbb{R}^d$. Next we present a discussion of the procedure to construct a finite kernel that approximates the original kernel well.

### 2.1. Construction of the finite dimensional kernel

To construct the finite dimensional kernel we will rely crucially on Mercer's theorem. We first recall a couple of useful definitions.

**Definition 3** *Let $\mathcal{A} \subset \mathbb{R}^d$ and $\mathbb{P}$ a probability measure supported over $\mathcal{A}$. We denote by $L_2(\mathcal{A};\mathbb{P})$ the space of square integrable functions over $\mathcal{A}$ and measure $\mathbb{P}$, $L_2(\mathcal{A};\mathbb{P}) :=$*

[3]This assumption can be relaxed to let $\mathcal{A}$ be a compact set when $\mathcal{K}$ is Lipschitz continuous. In this setting we can instead work with an appropriately fine approximating cover over the set $\mathcal{A}$.

$\left\{ f : \mathcal{A} \to \mathbb{R} \middle| \int_{\mathcal{A}} f^2(x) d\mathbb{P}(x) < \infty \right\}.$

**Definition 4** *A kernel $\mathcal{K} : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$ is square integrable with respect to measure $\mathbb{P}$ over $\mathcal{A}$, if $\int_{\mathcal{A}\times\mathcal{A}} \mathcal{K}^2(x, y) d\mathbb{P}(x) d\mathbb{P}(y) < \infty$.*

Now we are ready to present Mercer's theorem (38) (see (19)).

**Theorem 5 (Mercer's Theorem)** *Let $\mathcal{A} \subset \mathbb{R}^d$ be compact and $\mathbb{P}$ be a finite Borel measure with support $\mathcal{A}$. Suppose $\mathcal{K}$ is a continuous square integrable positive definite kernel on $\mathcal{A}$, and define a positive definite operator $\mathcal{T}_{\mathcal{K}} : L_2(\mathcal{A}; \mathbb{P}) \mapsto L_2(\mathcal{A}; \mathbb{P})$ by*

$$(\mathcal{T}_{\mathcal{K}} f)(\cdot) := \int_{\mathcal{A}} \mathcal{K}(\cdot, x) f(x)\, \mathrm{d}\mathbb{P}.$$

*Then there exists a sequence of eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$ that form an orthonormal basis of $L_2(\mathcal{A}; \mathbb{P})$ consisting of eigenfunctions of $\mathcal{T}_{\mathcal{K}}$, and an associated sequence of non-negative eigenvalues $\{\mu_j\}_{j=1}^{\infty}$ such that $\mathcal{T}_{\mathcal{K}}(\phi_j) = \mu_j \phi_j$ for $j = 1, 2, \ldots$. Moreover the kernel function can be represented as*

$$\mathcal{K}(u, v) = \sum_{i=1}^{\infty} \mu_i \phi_i(u) \phi_i(v) \tag{3}$$

*where the convergence of the series holds uniformly.*

Mercer's theorem yields a natural way to construct a feature map $\Phi(x)$ for $\mathcal{K}$ by defining the $i^{th}$ component of the feature map to be $\Phi(x)_i := \sqrt{\mu_i}\phi_i(x)$. With this choice of feature map the eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$ are orthogonal under the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$[4]. Armed with Mercer's theorem we first present a simple deterministic procedure to obtain a finite dimensional $\epsilon$-approximate kernel of $\mathcal{K}$. Essentially when the eigenfunctions of the kernel are uniformly bounded, $\sup_{x\in\mathcal{A}} |\phi_j(x)| \le \mathcal{B}$ for all $j$, and if the eigenvalues decay at a suitable rate we can truncate the series in (3) to get a finite dimensional approximation.

**Lemma 6** *Given $\epsilon > 0$, let $\{\mu_j\}_{j=1}^{\infty}$ be the Mercer operator eigenvalues of $\mathcal{K}$ under a finite Borel measure $\mathbb{P}$ with support $\mathcal{A}$ and eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$ with $\mu_1 \ge \mu_2 \ge \cdots$. Further assume that $\sup_{j\in\mathbb{N}} \sup_{x\in\mathcal{A}} |\phi_j(x)| \le \mathcal{B}$ for some $\mathcal{B} < \infty$. Let $m(\epsilon)$ be such that $\sum_{j=m+1}^{\infty} \mu_j \le \frac{\epsilon}{4\mathcal{B}^2}$. Then the kernel induced by a truncated feature map,*

$$\Phi_m^o(x) := \begin{cases} \sqrt{\mu_i}\phi_i(x) & \text{if } i \le m \\ 0 & \text{o.w.} \end{cases} \tag{4}$$

*induces a kernel $\hat{\mathcal{K}}_m^o := \langle \Phi_m^o(x), \Phi_m^o(y) \rangle_{\mathcal{H}} = \sum_{j=1}^{m} \mu_j \phi_j(x) \phi_j(y)$, for all $(x, y) \in \mathcal{A} \times \mathcal{A}$ that is an $\epsilon/4$-approximation of $\mathcal{K}$.*

The Hilbert space induced by the $\hat{\mathcal{K}}_m^o$ is a subspace of the original Hilbert space $\mathcal{H}$. The proof of this lemma is a simple application of Mercer's theorem and is relegated to Appendix C. If we have access to the eigenfunctions of $\mathcal{K}$ we can construct and work with $\hat{\mathcal{K}}_m^o$ because as Lemma 6 shows $\hat{\mathcal{K}}_m^o$ is an $\epsilon/4$-approximation to $\mathcal{K}$. Additionally, $\hat{\mathcal{K}}_m^o$ also has the same first $m$ Mercer eigenvalues and eigenfunctions under $\mathbb{P}$ as $\mathcal{K}$. Unfortunately, in most applications of interest the analytical computation of the eigenfunctions $\{\phi_i\}_{i=1}^{\infty}$ is not possible. We can get around this by building an estimate of the eigenfunctions using samples from $\mathbb{P}$ by leveraging results from kernel principal component analysis (PCA).

**Definition 7** *Let $S_m$ be the subspace of $\mathcal{H}$ spanned by the first $m$ eigenvectors of the covariance matrix $\mathbb{E}_{x\sim\mathbb{P}}\left[\Phi(x)\Phi(x)^\top\right]$.*

This corresponds to the span of the eigenfunctions $\phi_1, \phi_2, \ldots, \phi_m$ in $\mathcal{H}$ [5] . Define the linear projection operator $P_{S_m} : \mathcal{H} \mapsto \mathcal{H}$ that projects onto the subspace $S_m$; where $P(S_m)(v + v^\perp) = v$, if $v \in S_m$ and $v^\perp \in S_m^\perp$.

**Remark 8** *The feature map $\Phi_m^o(x)$ is a projection of the complete feature map to this subspace, $\Phi_m^o(x) = P_{S_m}(\Phi(x))$.*

Let $x_1, x_2, \ldots, x_p \sim \mathbb{P}$ be $p$ i.i.d. samples and construct the sample (kernel) covariance matrix, $\hat{\Sigma} := \frac{1}{p}\sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top$. Let $\hat{S}_m$ be the subspace spanned by the $m$ top eigenvectors of $\hat{\Sigma}$. Define the stochastic feature map, $\Phi_m(x) := P_{\hat{S}_m}(\Phi(x))$, the feature map defined by projecting $\Phi(x)$ to the random subspace $\hat{S}_m$. Intuitively we would expect that if the number of samples $p$ is high enough, then the kernel defined by the feature map $\Phi_m(x)$, $\hat{\mathcal{K}}_m(x, y) = \langle \Phi_m(x), \Phi_m(y) \rangle_{\mathcal{H}}$ will also be an $\epsilon$-approximation for the original kernel $\mathcal{K}$. Formalizing this claim is the following theorem.

**Theorem 9** *Let $m, \mathbb{P}$ be defined as in Lemma 6. Define the $m$-th level eigen-gap as $\delta_m = \frac{1}{2}(\mu_m - \mu_{m+1})$. Also let $B_m = \frac{2\mathcal{G}}{\delta_m}\left(1 + \sqrt{\frac{\alpha}{2}}\right)$, $2\delta_m > \sqrt{\epsilon} > 0$ and $p \ge \frac{2B_m^2 \mathcal{G}^2}{\sqrt{\epsilon}}$. Then the finite dimensional kernels $\hat{\mathcal{K}}_m^o$ and $\hat{\mathcal{K}}_m$ satisfy the following properties with probability $1 - e^{-\alpha}$,*

1. $\sup_{x,y\in\mathcal{A}} |\mathcal{K}(x, y) - \hat{\mathcal{K}}_m(x, y)| \le \epsilon$.

[4] To see this observe that the function $\phi_i$ can be expressed as a vector in the RKHS as a vector $v_i$ with $\phi_i$ in the $i^{th}$ coordinate and zeros everywhere else. So for any two $v_i$ and $v_j$ with $i \neq j$ we have $\langle v_i, v_j \rangle_{\mathcal{H}} = 0$.

[5] This holds as the $i^{th}$ eigenvector of the covariance matrix has $\phi_i$ as the $i^{th}$ coordinate and zero everywhere else combined with the fact that $\{\phi_i\}_{i=1}^{\infty}$ are orthonormal under the $L(\mathcal{A}; \mathbb{P})$ inner product.

**Algorithm 1** Finite dimensional proxy construction

**Input** : Kernel $\mathcal{K}$, effective dimension $m$, set $\mathcal{A}$, measure $\mathbb{P}$, bias tolerance $\epsilon > 0$, number of samples $p$.

**Function :** Finite proxy feature map $\Phi_m(\cdot)$

sample $x_1, \cdots, x_p \sim \mathbb{P}$.

construct sample Gram matrix $\hat{\mathbb{K}}_{i,j} = \frac{1}{p}\mathcal{K}(x_i, x_j)$.

calculate the top $m$ eigenvectors of $\hat{\mathbb{K}} \to \{\omega_1, \omega_2, \dots, \omega_m\}$.

**for** $j = 1, \dots, m$ **do**

  Set $v_j = \sum_{k=1}^p \omega_{jk}\Phi(x_k)$, ($\omega_{jk}$ is the $k^{th}$ entry of $\omega_j$)

**end**

define the feature map

$$\Phi_m(\cdot) := \begin{bmatrix} \langle v_1, \Phi(x)\rangle_{\mathcal{H}} \\ \vdots \\ \langle v_m, \Phi(x)\rangle_{\mathcal{H}} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^p \omega_{1k}\mathcal{K}(x_k, x) \\ \vdots \\ \sum_{k=1}^p \omega_{mk}\mathcal{K}(x_k, x) \end{bmatrix}.$$

2. *The Mercer eigenvalues* $\mu_1^{(p)} \geq \cdots \geq \mu_m^{(p)}$ *and* $\mu_1 \geq \cdots \geq \mu_m$ *of* $\hat{\mathcal{K}}_m$ *and* $\hat{\mathcal{K}}_m^o$ *are close,* $\sup_{i=1,\cdots,m} |\mu_i^{(p)} - \mu_i| \leq \sqrt{\epsilon}/2$.

Theorem 9 shows that given $\epsilon > 0$ the finite dimensional proxy $\hat{\mathcal{K}}_m$ is a $\epsilon$-approximation of $\mathcal{K}$ with high probability as long as sufficiently large number of samples are used. Furthermore, the top $m$ eigenvalues of the second moment matrix of $\mathcal{K}$ are at most $\sqrt{\epsilon}/2$-away from the eigenvalues of the second moment matrix of $\hat{\mathcal{K}}_m$ under $\mathbb{P}$.

To construct $\Phi_m(\cdot)$ we need to calculate the top $m$ eigenvectors of the sample covariance matrix $\hat{\Sigma}$, however, it is equivalent to calculate the top $m$ eigenvectors of the sample Gram matrix $\mathbb{K}$ and use them to construct the eigenvectors of $\hat{\Sigma}$ (for more details see Appendix B where we review the basics of kernel PCA).

### 2.2. Bandits Exponential Weights

In this section we present a modified version of exponential weights adapted to work with kernel losses. Exponential weights has been analyzed extensively applied to linear losses under bandit feedback (20; 17; 13). Two technical challenges make it hard to directly adapt their algorithms to our setting.

The first challenge is that at each round we need to estimate the adversarial action $w_t$. If the feature map of the kernel is finite dimensional this is easy to handle, however when the feature map is infinite dimensional, this becomes challenging and we need to build an approximate feature map $\Phi_m(\cdot)$ using Algorithm 2.1. This introduces a bias in our estimate of the adversarial action $w_t$ and we will need to control the contribution of the bias in our regret analysis. The second challenge will be to lower bound the minimum eigenvalue of the kernel covariance matrix as we will need to invert this matrix to estimate $w_t$. For general kernels which are infinite dimensional, the minimum eigenvalue is zero. To resolve this we will again turn to our construction of a finite dimensional proxy kernel.

### 2.3. Bandit Algorithm and Regret Bound

In our exponential weights algorithm we first build the finite dimensional proxy kernel $\hat{\mathcal{K}}_m$ using Algorithm 2.1. The rest of the algorithm is then almost identical to the exponential weights algorithm (EXP 2) studied for linear bandits in (20; 17; 13). In Algorithm 2.3 we set the exploration distribution $\nu_{\mathcal{J}}^{\mathcal{A}}$ to be such that it induces John's distribution ($\nu_{\mathcal{J}}$) over $\Phi_m(\mathcal{A}) := \{\Phi_m(a) \in \mathbb{R}^m : a \in \mathcal{A}\}$ (first introduced as an exploration distribution in (13); also a short discussion is presented in Appendix H.1). Note that for finite sets it is possible to build minimal volume ellipsoid containing $conv(\Phi_m(\mathcal{A}))$–John's ellipsoid and John's distribution in polynomial time (24)[6]. We assume without loss of generality that the center of the set $\mathcal{A}$ is such that the John's ellipsoid is centered at the origin.

If we know beforehand the behavior of the eigen-decay of the Mercer eigenvalues of $\mathcal{K}$ under measure $\mu$ we will be able to choose our tuning parameters optimally. In our algorithm we also build and invert the exact covariance matrix $\Sigma_m^{(t)}$, however this can be relaxed and we can work with a sample covariance matrix instead. We analyze the required sample complexity and error introduced by this additional step in Appendix D. We now state the main result of this paper which is an upper bound on the regret of Algorithm 2.3.

**Theorem 10** *Let $\mu_i$ be the $i$-th Mercer operator eigenvalue of $\mathcal{K}$ for the uniform measure $\mu$ over $\mathcal{A}$. Let $m, p, \alpha$ and $\epsilon$ be chosen as specified by the conditions in Theorem 9. Let the mixing coefficient be chosen such that $\gamma = \eta\mathcal{G}^4 m$. Then Algorithm 2.3 with probability $1 - e^{-\alpha}$ has regret bounded by*

$$\mathcal{R}_n \leq \gamma n + (e-2)\mathcal{G}^4\eta m n + 3\epsilon n + \frac{1}{\eta}\log(|\mathcal{A}|).$$

We prove this theorem in Appendix A. Note that this is similar to the regret rate attained for adversarial linear bandits in (20; 18; 13) with an additional term $3\epsilon n$ that accounts for the bias in our loss estimates $\hat{w}_t$. In our regret bounds the parameter $m$ plays the role of the effective dimension and will be determined by the rate of the eigen-decay of the kernel. When the underlying Hilbert space is finite dimensional (as is the case when the losses are linear) our regret

[6] It is thus possible to construct $\nu_{\mathcal{J}}$ over $\Phi_m(\mathcal{A})$ in polynomial time. However, as $\mathcal{A}$ is a finite set, using $\Phi_m(\cdot)$ and $\nu_{\mathcal{J}}$ it is also possible to construct $\nu_{\mathcal{J}}^{\mathcal{A}}$ efficiently.

**Algorithm 2** Bandit Information: Exponential Weights

**Input :** Set $\mathcal{A}$, learning rate $\eta > 0$, mixing coefficient $\gamma > 0$, number of rounds $n$, uniform distribution $\mu$ over $\mathcal{A}$, exploration distribution $\nu_{\mathcal{J}}^{\mathcal{A}}$ over $\mathcal{A}$, kernel map $\mathcal{K}$, effective dimension $m(\epsilon)$, number of samples $p$.

Build kernel $\hat{\mathcal{K}}_m$ with feature map $\Phi_m(\cdot)$ using Algorithm 2.1 with kernel $\mathcal{K}$, dimension $m$, distribution $\mu$, bias tolerance $\epsilon$ and number of samples $p$.

set $q_1(a) = \nu_{\mathcal{J}}^{\mathcal{A}}$.

**for** $t = 1, \ldots, n$ **do**

set $p_t = \gamma \nu_{\mathcal{J}}^{\mathcal{A}} + (1-\gamma) q_t$

choose $a_t \sim p_t$

observe $\langle \Phi(a_t), w_t \rangle_{\mathcal{H}}$

build the covariance matrix

$$\Sigma_m^{(t)} = \mathbb{E}_{x \sim p_t}\left[\Phi_m(x)\Phi_m(x)^\top\right]$$

compute the estimate $\hat{w}_t = \Sigma_m^{-1}\Phi_m(a_t)\langle \Phi(a_t), w_t\rangle_{\mathcal{H}}$.

update $q_{t+1}(a) \propto q_t(a) \cdot \exp\left(-\eta \cdot \langle \hat{w}_t, \Phi_m(a)\rangle_{\mathcal{H}}\right)$

**end**

bound recovers exactly the results of previous work (that is, $\epsilon = 0$ and $m = d$). Next we state the following different characteristic eigenvalue decay profiles.

**Definition 11 (Eigenvalue decay)** *Let the Mercer operator eigenvalues of a kernel $\mathcal{K}$ with respect to a measure $\mathbb{P}$ over a set $\mathcal{A}$ be denoted by $\mu_1 \geq \mu_2 \geq \ldots$.*

1. *$\mathcal{K}$ is said to have $(C, \beta)$-**polynomial eigenvalue decay** (with $\beta > 1$) if for all $j \in \mathbb{N}$ we have $\mu_j \leq Cj^{-\beta}$.*
2. *$\mathcal{K}$ is said to have $(C, \beta)$-**exponential eigenvalue decay** if for all $j \in \mathbb{N}$ we have $\mu_j \leq Ce^{-\beta j}$.*

Under assumptions on the eigen-decay we can establish bounds on the *effective dimension* $m$ and $\mu_m$, so that the condition stated in Lemma 6 is satisfied and we are guaranteed to build an $\epsilon$-approximate kernel $\hat{\mathcal{K}}_m$. We establish bounds on $m$ in Proposition 33 presented in Appendix C.1.

**Corollary 12** *Let the conditions stated in Theorem 10 hold. Then Algorithm 2.3 has its regret bounded by the following rates with probability $1 - e^{-\alpha}$.*

1. *If $\mathcal{K}$ has $(C, \beta)$-polynomial eigenvalue decay under measure $\mu$, with $\beta > 2$. Then by choosing $\eta = \frac{1}{3^{\frac{1}{2\beta-1}}} \cdot \left[\frac{\beta-1}{4C\mathcal{B}^2}\right]^{1/2\beta-1} \cdot \left[\frac{\log(|\mathcal{A}|)}{((e-1)\mathcal{G}^4)^{\frac{\beta-1}{\beta}} n}\right]^{\frac{\beta}{2\beta-1}}$ and $m = \left[\frac{4C\mathcal{B}^2}{(\beta-1)\epsilon}\right]^{1/\beta-1}$ where $\epsilon = \left(\frac{(e-1)\eta\mathcal{G}^4}{3}\right)^{(\beta-1)/\beta}\left[\frac{4C\mathcal{B}^2}{\beta-1}\right]^{1/\beta}$, the expected regret is bounded by*
$$\mathcal{R}_n \leq 3\left[\frac{4C\mathcal{B}^2}{\beta-1}\right]^{\frac{1}{2\beta-1}}\left(e\mathcal{G}^4\log\left(|\mathcal{A}|\right)\right)^{\frac{\beta-1}{2\beta-1}} \cdot n^{\frac{\beta}{2(\beta-1)}}.$$
2. *If $\mathcal{K}$ has $(C, \beta)$-exponential eigenvalue decay under measure $\mu$. Then by choosing $\eta = \left(\frac{\beta\log(|\mathcal{A}|)}{(e-1)\mathcal{G}^4\log\left(\frac{4C\mathcal{B}^2}{\beta}\right)\cdot n}\right)^{1/2}$ and $m = \frac{1}{\beta}\log\left(\frac{4C\mathcal{B}^2}{\beta\epsilon}\right)$ where $\epsilon = \frac{(e-1)\mathcal{G}^4\eta}{3\beta}\log\left(\frac{4C\mathcal{B}^2}{\beta}\right)$, with $n$ large enough so that $\epsilon < 1$, the expected regret is bounded by*
$$\mathcal{R}_n \leq \tilde{\mathcal{O}}\left(\left[\frac{18\mathcal{G}^4\log\left(|\mathcal{A}|\right)\cdot n}{\beta\log\left(\frac{4C\mathcal{B}^2}{\beta}\right)}\right]^{1/2}\right).$$

**Remark 13** *Under $(C, \beta)$-polynomial eigen-decay condition we have that the regret is upper bounded by $\mathcal{R}_n \leq \mathcal{O}(n^{\frac{\beta}{2(\beta-1)}})$. While when we have $(C, \beta)$-exponential eigen-decay we almost recover the adversarial linear bandits regret rate (up to logarithmic factors), with $\mathcal{R}_n \leq \mathcal{O}(n^{1/2}\log(n))$.*

One way to interpret the results of Corollary 12 in contrast to the regret bounds obtained for linear losses is the following. We introduce additional parameters into our analysis to handle the infinite dimensionality of our feature vectors – the effective dimension $m$ and bias of our estimate $\epsilon$. When the effective dimension $m$ is chosen to be large we get can build an estimate of the adversarial action $\hat{w}_t$ which has low bias, however this estimate would have large variance ($\mathcal{O}(m)$). On the other hand if we choose $m$ to be small we can build a low variance estimate of the adversarial action but with high bias ($\epsilon$ is large). We trade these off optimally to get the regret bounds established above. In the case of exponential decay we obtain that the choice $m = \mathcal{O}(\log(n))$ is optimal, hence the regret bound only degrades by a logarithmic factor in terms of $n$ as compared to linear losses (where $m$ would be a constant). When we have polynomial decay, the effective dimension is higher $m = \mathcal{O}(n^{\frac{1}{2(\beta-1)}})$ which leads to worse bounds on the expected regret. Note that asymptotically as $\beta \to \infty$ the regret bound goes to $n^{1/2}$ which aligns well with the intuition that the effective dimension is small. While when $\beta \to 2$ (the effective dimension $m \to \infty$) the regret bound becomes close to linear in $n$.

We can also show a minimax lower bound for these two settings that are close to matching the upper bound.

**Proposition 14** *(informal) For any algorithm used by the player, there exist a strategy for the adversary such that $\mathcal{R}_n \geq \Omega\left(n^{\frac{\beta+1}{2\beta}}\right)$ whenever $\mu_j = \tilde{\mathcal{O}}(j^{-\beta})$, while when the decay is exponential $\mathcal{R}_n \geq \Omega\left(n^{1/2}\right)$.*

**Algorithm 3** Full Information: Exponential Weights

**Input** : Set $\mathcal{A}$, learning rate $\eta > 0$, number of rounds $n$.
Set $p_1(a)$ uniform distribution over $\mathcal{A}$.
**for** $t = 1, \ldots, n$ **do**
  choose $a_t \sim p_t$
  observe $w_t$
  update $p_{t+1}(a) \propto p_t(a) \cdot \exp\left(-\eta \cdot \langle w_t, \Phi(a) \rangle_{\mathcal{H}}\right)$
**end**

The lower bound follows by a modification of the arguments used to prove a lower bound linear bandits. For a complete proof see Appendix E.

# 3. Full Information Setting

## 3.1. Full information Exponential Weights

We begin by presenting a version of the exponential weights algorithm, Algorithm 3 adapted to our setup. In each round we sample an action vector $a_t \in \mathcal{A}$ from the exponential weights distribution $p_t$. After observing the loss, $\langle \Phi(a_t), w_t \rangle_{\mathcal{H}}$ we update the distribution by a multiplicative factor, $\exp(-\eta \langle w_t, \Phi(a) \rangle_{\mathcal{H}})$. In the algorithm presented we choose the initial distribution $p_1(a)$ to be uniform over the set $\mathcal{A}$, however we note that alternate initial distributions with support over the whole set could also be considered. We can establish a sub-linear regret of $\mathcal{O}(\sqrt{n})$ for the exponential weights algorithm.

**Theorem 15** *Assume that in Algorithm 3 the step size $\eta$ is chosen to be, $\eta = \sqrt{\frac{\log(vol(\mathcal{A}))}{e-2}} \cdot \frac{1}{\mathcal{G}^2 n^{1/2}}$, with $n$ large enough such that $\sqrt{\frac{\log(vol(\mathcal{A}))}{e-2}} \frac{1}{n^{1/2}} \leq 1$. Then the expected regret after $n$ rounds is bounded by,*

$$\mathcal{R}_n \leq \sqrt{(e-2)\log(vol(\mathcal{A}))}\mathcal{G}^2 n^{1/2}.$$

We prove this regret bound in Appendix F.1.

## 3.2. Conditional Gradient Descent

Next we present an online conditional gradient (Frank-Wolfe) method (26) adapted for kernel losses. The conditional gradient method is also a well studied algorithm studied in both the online and offline setting (for a review see (25)). The main advantage of the conditional gradient method is that as opposed to projected gradient descent and related methods, the projection step is avoided. At each round the conditional gradient method involves the optimization of a linear (kernel) objective function over $\mathcal{A}$ to get a point $v_t \in \mathcal{A}$. Next we update the *optimal mean* action $X_{t+1}$ by re-weighting the previous mean action $X_t$ by $(1 - \gamma_t)$ and weight our new action $v_t$ by $\gamma_t$. Note that this construction also automatically suggests a distribution over $a_1, v_1, v_2, \ldots, v_t \in \mathcal{A}$ such that, $X_{t+1}$ is a convex combination of $\Phi(a_1), \Phi(x_1), \ldots, \Phi(a_t)$. For this algorithm we can prove a regret bound of $\mathcal{O}(n^{3/4})$ (presented in Appendix F.2.).

**Algorithm 4** Full Information: Conditional Gradient

**Input** : Set $\mathcal{A}$, number of rounds $n$, initial action $a_1 \in \mathcal{A}$, inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, learning rate $\eta$, mixing rates $\{\gamma_t\}_{t=1}^n$.
$X_1 = \Phi(a_1)$
choose $\mathcal{D}_1$ such that $\mathbb{E}_{x \sim \mathcal{D}_1} \Phi(x) = X_1$
**for** $t = 1, 2, \ldots, n$ **do**
  sample $a_t \sim \mathcal{D}_t$
  observe the adversarial action $w_t$
  define $F_t(Y) \triangleq \eta \sum_{s=1}^{t-1} \langle w_s, Y \rangle_{\mathcal{H}} + \|Y - X_1\|_{\mathcal{H}}^2$
  compute $v_t = \mathrm{argmin}_{a \in \mathcal{A}} \langle \nabla F_t(X_t), \Phi(a) \rangle_{\mathcal{H}}$
  update mean $X_{t+1} = (1 - \gamma_t) X_t + \gamma_t \Phi(v_t)$
  choose $\mathcal{D}_{t+1}$ s.t. $\mathbb{E}_{x \sim \mathcal{D}_{t+1}}[\Phi(x)] = X_{t+1}$.
**end**

**Theorem 16** *Let the step size be $\eta = \frac{1}{2n^{3/4}}$. Also let the mixing rates be $\gamma_t = \min\{1, 2/t^{1/2}\}$, then Algorithm 4 attains regret of $\mathcal{R}_n \leq 8\mathcal{G}^2 n^{3/4}$.*

# 4. Applications

## 4.1. General Quadratic Losses

The first example of losses that we present are general quadratic losses. At each round the adversary can choose a symmetric (not necessarily positive semi-definite matrix) $A \in \mathbb{R}^{d \times d}$, and a vector $b \in \mathbb{R}^d$, with a constraint on the norm of the matrix and vector such that $\|A\|_F^2 + \|b\|_2^2 \leq \mathcal{G}^2$. If we embed this pair into a Hilbert space defined by the feature map $(A, b)$ we get a kernel loss defined as – $\langle \Phi(x), (A, b) \rangle_{\mathcal{H}} = x^\top A x + b^\top x$, where $\Phi(x) = (xx^\top, x)$ is the associated feature map for any $x \in \mathcal{A}$ and the inner product in the Hilbert space is defined as the concatenation of the trace inner product on the first coordinate and the Euclidean inner product on the second coordinate. The cumulative loss that the player would aspire to minimize is, $\sum_{t=1}^n x_t^\top A_t x_t + b_t^\top x_t$. The setting without the linear term, that is when $b_t = 0$ with positive semidefinite matrices $A_t$ is previously well studied in (47; 48; 23; 4). However when the matrix is not positive semi-definite (making the losses non-convex) and there is a linear term, regret guarantees and tractable algorithms have not been studied even in the full information case.

As this is a kernel loss we have regret bounds for these losses. We demonstrate in the subsequent sections in the full information case it is also possible to run our algorithms efficiently. First for exponential weights we show sampling is efficient for these losses.

**Lemma 17 (Proof in Appendix F.1)** *Let $B \in \mathbb{R}^{d\times d}$ be a symmetric matrix and $b \in \mathbb{R}^d$. Sampling from $q(a) \propto \exp(a^\top Ba + a^\top b)$ for $\|a\|_2 \le 1, a \in \mathbb{R}^d$ is tractable in $\tilde{\mathcal{O}}(d^4)$ time.*

### 4.2. Guarantees for Conditional Gradient Descent

We now demonstrate that conditional gradient descent also can be run efficiently when the adversary plays a general quadratic loss. At each round the conditional gradient descent requires the player to solve the optimization problem, $v_t = \text{argmin}_{a\in\mathcal{A}}\langle\nabla F_t(X_t), \Phi(a)\rangle_{\mathcal{H}}$. When the set of actions is $\mathcal{A} = a \in \mathbb{R}^d : \|a\|_2 \le 1$ then under quadratic losses this problem becomes,

$$v_t = \underset{a\in\mathcal{A}}{\text{argmin}}\, a^\top Ba + b^\top a, \tag{5}$$

for an appropriate matrix $B$ and $b$ that can be calculated by aggregating the adversary's actions up to step $t$. Observe that the optimization problem (5) is a quadratically constrained quadratic program (QCQP) given our choice of $\mathcal{A}$. The dual problem is the (semi-definite program) SDP,

$$\begin{aligned} & max - t - \mu \\ \text{s. t.} & \\ & \begin{bmatrix} B + \mu I & b/2 \\ b/2 & t \end{bmatrix} \succ 0. \end{aligned}$$

For this particular program with a norm ball constraint set it is known the duality gap is zero provided Slater's condition holds, that is, strong duality holds (see Annex B.1 (12)).

### 4.3. Posynomial Losses

In this section we will define a *posynomial game*, by introducing posynomial losses and prove that these losses can also be viewed as kernel inner products. We will use the connection between posynomials and *Geometric programs* to prove that conditional gradient descent can be run efficiently on this family of losses.

**Definition 18 (Monomial)** *A function $f : \mathbb{R}^d_+ \mapsto \mathbb{R}$ defined as*

$$f(x) = cx_1^{\alpha_1}x_2^{\alpha_2}\cdots x_d^{\alpha_d},$$

*where $c > 0$ and $\alpha_i \in \mathbb{R}$, is called a monomial function.*

A sum of monomials is a posynomial.

**Definition 19 (Posynomial)** *A function $f : \mathbb{R}^d_+ \mapsto \mathbb{R}$ defined as*

$$f(x) = \sum_{k=1}^{m} c_k x_1^{\alpha_{1k}} x_2^{\alpha_{2k}} \cdots x_d^{\alpha_{dk}}$$

*where $c_k > 0$ and $\alpha_{ik} \in \mathbb{R}$, is called a posynomial function.*

Note that posynomial functions are closed under addition, multiplication and non-negative scaling. If we assume the adversary at each round plays a vector of dimension $m$ with all non-negative entries, $w_t = (c_1, c_2, \cdots, c_m)$, while the player chooses a vector $x \in \mathbb{R}^d_+$. This vector is then partitioned into $m$ parts,

$$x = (\underbrace{x_1, x_2}_{s_1}, \dots, \underbrace{x_{d-2}, x_{d-1}, x_d}_{s_m}),$$

and the feature vector is defined as

$$\Phi(x) = \begin{bmatrix} x_1^{\alpha_1}x_2^{\alpha_2} \\ \vdots \\ x_{d-2}^{\alpha_{d-2}}x_{d-1}^{\alpha_{d-1}}x_d^{\alpha_d} \end{bmatrix}.$$

Where the $i^{th}$ component of $\Phi(\cdot)$ is only a function of the $i^{th}$ partition of the coordinates $s_i$. Then the loss obtained on the evaluation of the inner product between the adversary and player action is a posynomial loss function,

$$\langle w_t, \Phi(x)\rangle_{\mathcal{H}} = \sum_{k=1}^{m} c_k x_1^{\alpha_{k1}} \cdots x_d^{\alpha_{kd}}.$$

A number of scenarios can be modeled as a minimization/maximization problem over posynomial functions (see (11) for a detailed list of examples). We now show that conditional gradient descent can be run efficiently over posynomial losses. If we again assume that the set of actions $\mathcal{A} = \{a \in \mathbb{R}^d : \|a\|_2 \le 1\}$. Additionally we all choose the initial action to be the solution to the optimization problem,

$$a_1 = \underset{a\in\mathcal{A}}{\text{argmin}} \sum_{k=1}^{d} \Phi(a)_i.$$

Observe that the objective function is a posynomial subject to a posynomial inequality constraint. This is a geometric program that can be solved efficiently by changing variables and converting into a convex program (Section 2.5 in (11)). At each round of the conditional gradient descent algorithm requires us to solve the optimization problem,

$$v_t = \underset{a\in\mathcal{A}}{\text{argmin}}\langle \eta \sum_{s=1}^{t-1} w_t + 2(X_t - \Phi(a_1)), \Phi(a)\rangle_{\mathcal{H}}. \tag{6}$$

Given that posynomials are closed under addition, and given our choice of $a_1$, the objective function (6) is still a posynomial and the constraint is a posynomial inequality. This can again be cast as a geometric program that can be solved efficiently at each round.

### Conclusion

It would be interesting to explore and study more kernel losses for which we have regret guarantees and for which our algorithms are also computationally efficient.

## References


[1] Jacob Duncan Abernethy, Elad Hazan, and Alexander Rakhlin. An efficient algorithm for bandit linear optimization. In *21st Annual Conference on Learning Theory*, 2008.

[2] Alekh Agarwal and Ofer Dekel. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Conference on Learning Theory*, pages 28–40, 2010.

[3] Alekh Agarwal, Dean P Foster, Daniel J Hsu, Sham M Kakade, and Alexander Rakhlin. Stochastic convex optimization with bandit feedback. In *Advances in Neural Information Processing Systems*, pages 1035–1043, 2011.

[4] Zeyuan Allen-Zhu and Yuanzhi Li. Follow the compressed leader: Faster algorithms for matrix multiplicative weight updates. *arXiv preprint arXiv:1701.01722*, 2017.

[5] Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

[6] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 227–236. ACM, 2007.

[7] Baruch Awerbuch and Robert D Kleinberg. Adaptive routing with end-to-end feedback: Distributed learning and geometric approaches. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 45–53. ACM, 2004.

[8] Maximilian Balandat, Walid Krichene, Claire Tomlin, and Alexandre Bayen. Minimizing regret on reflexive Banach spaces and learning Nash equilibria in continuous zero-sum games. *arXiv preprint arXiv:1606.01261*, 2016.

[9] Keith Ball. An elementary introduction to modern convex geometry. *Flavors of geometry*, 31:1–58, 1997.

[10] Peter Bartlett. Learning in sequential decision problems. Lecture Notes Stat 260/CS 294-102, 2014.

[11] Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and engineering*, 8(1):67, 2007.

[12] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[13] Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Annual Conference on Learning Theory*, volume 23, pages 41–1. Microtome, 2012.

[14] Sébastien Bubeck, Ofer Dekel, Tomer Koren, and Yuval Peres. Bandit convex optimization: $\sqrt{T}$ regret in one dimension. In *Conference on Learning Theory*, pages 266–278, 2015.

[15] Sébastien Bubeck, Ronen Eldan, and Yin Tat Lee. Kernel-based methods for bandit convex optimization. *arXiv preprint arXiv:1607.03084*, 2016.

[16] Nicolò Cesa-Bianchi, Pierre Gaillard, Claudio Gentile, and Sébastien Gerchinovitz. Algorithmic chaining and the role of partial feedback in online nonparametric learning. In *Conference on Learning Theory*, pages 465–481, 2017.

[17] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[18] Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

[19] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[20] Varsha Dani, Sham M Kakade, and Thomas P Hayes. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems*, pages 345–352, 2008.

[21] Ofer Dekel, Ronen Eldan, and Tomer Koren. Bandit smooth convex optimization: Improving the bias-variance tradeoff. In *Advances in Neural Information Processing Systems*, pages 2926–2934, 2015.

[22] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.

[23] Dan Garber, Elad Hazan, and Tengyu Ma. Online learning of eigenvectors. In *International Conference on Machine Learning*, pages 560–568, 2015.

[24] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

[25] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

[26] Elad Hazan and Satyen Kale. Projection-free online learning. *arXiv preprint arXiv:1206.4657*, 2012.

[27] Elad Hazan and Kfir Levy. Bandit convex optimization: Towards tight bounds. In *Advances in Neural Information Processing Systems*, pages 784–792, 2014.

[28] Elad Hazan and Yuanzhi Li. An optimal algorithm for bandit convex optimization. *arXiv preprint arXiv:1603.04350*, 2016.

[29] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[30] Alan J Hoffman and Helmut W Wielandt. The variation of the spectrum of a normal matrix. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 118–120. World Scientific, 2003.

[31] Martin Jaggi. Convex optimization without projection steps. *arXiv preprint arXiv:1108.1170*, 2011.

[32] Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

[33] Robert D Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems*, pages 697–704, 2005.

[34] László Lovász and Santosh Vempala. The geometry of log-concave functions and sampling algorithms. *Random Structures & Algorithms*, 30(3):307–358, 2007.

[35] H Brendan McMahan and Avrim Blum. Online geometric optimization in the bandit setting against an adaptive adversary. In *International Conference on Computational Learning Theory*, pages 109–123. Springer, 2004.

[36] H Brendan McMahan and Francesco Orabona. Unconstrained online linear learning in Hilbert spaces: Minimax algorithms and normal approximations. In *Conference on Learning Theory*, pages 1020–1039, 2014.

[37] Shahar Mendelson, Alain Pajor, et al. On singular values of matrices with independent rows. *Bernoulli*, 12(5):761–773, 2006.

[38] James Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, 209:415–446, 1909.

[39] Jiazhong Nie, Wojciech Kotłowski, and Manfred K Warmuth. Online PCA with optimal regrets. In *International Conference on Algorithmic Learning Theory*, pages 98–112. Springer, 2013.

[40] Alexander Rakhlin and Karthik Sridharan. Online nonparametric regression with general loss functions. *arXiv preprint arXiv:1501.06598*, 2015.

[41] Ankan Saha and Ambuj Tewari. Improved regret guarantees for online smooth convex optimization with bandit feedback. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 636–642, 2011.

[42] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[43] Shai Shalev-Shwartz and Yoram Singer. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69(2-3):115–142, 2007.

[44] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*, 2009.

[45] Koji Tsuda, Gunnar Rätsch, and Manfred K Warmuth. Matrix exponentiated gradient updates for online learning and Bregman projection. *Journal of Machine Learning Research*, 6(Jun):995–1018, 2005.

[46] Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

[47] Manfred K Warmuth and Dima Kuzmin. Online variance minimization. In *International Conference on Computational Learning Theory*, pages 514–528. Springer, 2006.

[48] Manfred K Warmuth and Dima Kuzmin. Randomized online PCA algorithms with regret bounds that are logarithmic in the dimension. *Journal of Machine Learning Research*, 9(Oct):2287–2320, 2008.

[49] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936, 2003.

[50] Laurent Zwald and Gilles Blanchard. On the convergence of eigenspaces in kernel principal component analysis. In *Advances in neural information processing systems*, pages 1649–1656, 2006.

# Appendix

### Organization of the Appendix and Roadmap of the Proof

Here we describe the general organization of the proofs of the paper. We use the same notation for parameters as throughout the paper.

In Appendix A we provide a proof of Theorem 10 and Corollary 12. At a high level, the elements to prove this theorem are similar to that of proving regret bounds for linear losses. We first decompose the regret into an *approximation error* term that arises due to the construction of the finite dimensional proxy $\Phi_m(\cdot)$ and another term which corresponds to the regret of a finite dimensional linear loss game (see Equation 7). To prove Theorem 10 we then proceed in Appendix A.1 to control the regret of this finite dimensional linear bandit game by classical techniques. Crucially we also control terms that arise due to the bias in our estimators by invoking Lemma 23.

We prove minimax lower bounds for the bandit setting in Appendix E.

In Appendix B we introduce and discuss ideas related to kernel principal component analysis (PCA). While in Appendix C we prove Theorem 9. Recall that this theorem was vital in establishing that the finite dimensional feature map we construct in Algorithm 2.3 induces a kernel $\hat{\mathcal{K}}_m$ that is an $\epsilon$-approximation of $\mathcal{K}$. In Appendix D we establish bounds on the sample complexity and control the error if the sample covariance matrix is used instead of the full covariance matrix in Algorithm 2.3.

The results about the full information setting, specifically the proofs of Theorems 15 and 16 are provided in Appendix F. In Appendix 4.3 we apply our framework to posynomial losses, in Appendix H we discuss Hoeffding's inequality and John's Theorem. Finally in Appendix I we present experimental evidence to verify our claims.

## A. Bandits Exponential Weights Regret Bound

In this section we prove the regret bound stated in Section 2. We borrow all the notation from Section 2. As defined before the expected regret for Algorithm 2.3 after $n$ rounds is

$$\mathcal{R}_n = \mathbb{E}\left[\sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle_{\mathcal{H}} - \langle \Phi(a^*), w_t \rangle_{\mathcal{H}}\right] = \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t}\left[\langle \Phi(a_t), w_t \rangle_{\mathcal{H}} - \langle \Phi(a^*), w_t \rangle_{\mathcal{H}} \Big| \mathcal{F}_{t-1}\right]\right],$$

where $p_t$ is the exponential weights distribution described in Algorithm 2.3, $a^*$ is the optimal action and $\mathcal{F}_{t-1}$ is the sigma field that conditions on $(a_1, a_2, \ldots, a_{t-1}, y_1, y_2, \ldots, y_{t-1}, y_t)$, the events up to the end of round $t-1$. We will prove the regret bound for the case when the kernel is infinite dimensional, that is, the feature map $\Phi(a) \in \mathbb{R}^D$, where $D = \infty$. When $D$ is finite the proof is identical with $\epsilon = 0$. Recall that when $D$ is infinite we constrain the adversary to play rank-1 actions. We are going to refer to the adversarial action as $w_t =: \Phi(y_t)$ for some $y_t \in \mathbb{R}^d$. We now expand the definition of regret and get,

$$\begin{aligned}\mathcal{R}_n = \mathbb{E}&\left[\sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t}\left[\langle \Phi(a_t), w_t \rangle_{\mathcal{H}} - \langle \Phi_m(a_t), w_t \rangle_{\mathcal{H}} \Big| \mathcal{F}_{t-1}\right]\right] \\ &+ \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t}\left[\langle \Phi_m(a^*), w_t \rangle_{\mathcal{H}} - \langle \Phi(a^*), w_t \rangle_{\mathcal{H}} \Big| \mathcal{F}_{t-1}\right]\right] \\ &+ \underbrace{\mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t}\left[\langle \Phi_m(a_t), w_t \rangle_{\mathcal{H}} - \langle \Phi_m(a^*), w_t \rangle_{\mathcal{H}} \Big| \mathcal{F}_{t-1}\right]\right]}_{=:\mathcal{R}_n^m}.\end{aligned}$$

Here $\mathcal{R}_n^m$ is the regret when we play the distribution in Algorithm 2.3 but are hit with losses that are governed by the kernel – $\hat{\mathcal{K}}_m(\cdot, \cdot)$ (with the same $a^*$ as before). Observe that in $R_n^m$ only the component of $w_t$ in the subspace $\hat{S}_m$ contributes to the inner product, thus every term is of the form

$$\langle \Phi_m(a_t), w_t \rangle = \langle \Phi_m(a_t), \Phi_m(y_t) \rangle = \hat{\mathcal{K}}_m(y_t, a_t).$$

As the proxy kernel $\hat{\mathcal{K}}_m$ is uniformly close by Theorem 9 we have,

$$\mathcal{R}_n \leq 2\varepsilon n + \mathcal{R}_n^m. \tag{7}$$

### A.1. Proof of Theorem 10

We will now attempt to bound $\mathcal{R}_n^m$ and prove Theorem 10. First we define the unbiased estimator (conditioned on $\mathcal{F}_{t-1}$) of $\Phi_m(y_t)$ at each round,

$$\tilde{w}_t := \hat{\mathcal{K}}_m(a_t, y_t)\left((\Sigma_m^{(t)})^{-1}\Phi_m(a_t)\right), \qquad t \in \{1, \dots, n\}, \tag{8}$$

where $\Phi_m(y_t) = P_{\hat{S}_m}(\Phi(y_t)) = P_{\hat{S}_m}(w_t)$. We cannot build $\tilde{w}_t$ as we do not receive $\hat{\mathcal{K}}_m(a_t, y_t)$ as feedback. Thus we also have

$$\begin{aligned}\mathbb{E}_{a_t\sim p_t}\left[\hat{w}_t|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\mathcal{K}(a_t, y_t)\left((\Sigma_m^{(t)})^{-1}\Phi_m(a_t)\right)\Big|\mathcal{F}_{t-1}\right] \\ &= \Phi_m(y_t) + \mathbb{E}\left[\underbrace{\left(\mathcal{K}(a_t, y_t) - \hat{\mathcal{K}}_m(a_t, y_t)\right)\left((\Sigma_m^{(t)})^{-1}\Phi_m(a_t)\right)}_{=:\xi_t,\text{ the bias}}\Big|\mathcal{F}_{t-1}\right], t \in \{1, \dots, n\}.\end{aligned} \tag{9}$$

If $\Phi_m(\cdot) = \Phi(\cdot)$ then the bias $\xi_t$ would be zero. We now present some estimates involving $\tilde{w}_t$. In the following section we sometimes denote $\tilde{w}_t$ and $\hat{w}_t$ more explicitly as $\tilde{w}_t(a_t)$ and $\hat{w}_t(a_t)$ where there may be room for confusion.

**Lemma 20** *For any fixed $a \in \mathcal{A}$ we have,*

$$\mathbb{E}_{a_t\sim p_t}\left[\langle\tilde{w}_t(a_t), \Phi_m(a)\rangle\Big|\mathcal{F}_{t-1}\right] = \hat{\mathcal{K}}_m(y_t, a), \qquad t \in \{1, \dots, n\}.$$

*We also have for all $t \in \{1, \dots, t\}$,*

$$\mathbb{E}_{a_t\sim p_t}\left[\hat{\mathcal{K}}_m(y_t, a_t)\Big|\mathcal{F}_{t-1}\right] = \mathbb{E}_{a_t\sim p_t}\left[\sum_{a\in\mathcal{A}} p_t(a)\langle\tilde{w}_t(a_t), \Phi_m(a)\rangle\Big|\mathcal{F}_{t-1}\right].$$

**Proof** The first claim follows by Equation (8) and the linearity of expectation we have

$$\mathbb{E}_{a_t\sim p_t}\left[\langle\tilde{w}_t(a_t), \Phi_m(a)\rangle\Big|\mathcal{F}_{t-1}\right] = \langle\mathbb{E}\left[\tilde{w}_t(a_t)|\mathcal{F}_{t-1}\right], \Phi_m(a)\rangle = \hat{\mathcal{K}}_m(y_t, a).$$

where the expectation is taken over $p_t$. Now to prove the second part of the theorem statement we will use tower property. Observe that conditioned on $\mathcal{F}_{t-1}$, $p_t$ and $a_t$ are measurable.

$$\mathbb{E}\left[\hat{\mathcal{K}}_m(y_t, a_t)\Big|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\mathbb{E}_{a\sim p_t}\left[\langle\tilde{w}_t(a_t), \Phi_m(a)\rangle\Big|a_t\right]\Big|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\sum_{a\in\mathcal{A}} p_t(a)\langle\tilde{w}_t(a_t), \Phi_m(a)\rangle\Big|\mathcal{F}_{t-1}\right].$$

■

We are now ready to prove Theorem 10 and establish the claimed regret bound.

**Proof** [Proof of Theorem 10] The proof is similar to the regret bound analysis of exponential weights for linear bandits. We proceed in 4 steps. In the first step we decompose the cumulative loss in terms of an exploration cost and an exploitation cost. In Step 2 we control the exploitation cost by using Hoeffding's inequality as is standard in linear bandits literature, but additionally we need to control terms arising out of the bias of our estimate. In Step 3 we bound the exploration cost and finally in the fourth step we combine the different pieces and establish the claimed regret bound.

**Step 1:** Using Lemma 20 and the fact that $\tilde{w}_t$ is an unbiased estimate of $\Phi_m(y_t)$ we can decompose the cumulative loss, the first term in $R_n^m$ as

$$\mathbb{E}\left[\sum_{t=1}^{n}\hat{\mathcal{K}}_m(a_t,y_t)\right] = \mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}p_t(a)\left\langle\tilde{w}_t(a_t),\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right]$$

$$= \underbrace{(1-\gamma)\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left\langle\tilde{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right]}_{Exploitation} + \underbrace{\gamma\cdot\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}\nu_{\mathcal{J}}^{\mathcal{A}}(a)\left\langle\tilde{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right]}_{Exploration}, \quad (10)$$

where the second line follows by the definition of $p_t$.

**Step 2:** We first focus on the ‘Exploitation’ term.

$$♣ := (1-\gamma)\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left\langle\tilde{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right]$$

$$= (1-\gamma)\underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left\langle\hat{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right]}_{=:♠} + (1-\gamma)\underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left\langle\tilde{w}_t-\hat{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right]}_{=:\diamondsuit}. \quad (11)$$

Under our choice of $\gamma$ by Lemma 24 (proved in Appendix A.1.1) we know that $\eta\langle\hat{w}_t,\Phi(a_t)\rangle > -1$. Therefore by Hoeffding’s inequality (Lemma 53) we get,

$$♠ = \mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left\langle\hat{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right] \le -\frac{1}{\eta}\underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\log\left(\mathbb{E}_{a\sim q_t}\left[\exp\left(-\eta\left\langle\hat{w}_t(a_t),\Phi_m(a)\right\rangle\right)\right]\right)\right]}_{=:\Gamma_1}$$

$$+ \underbrace{(e-2)\eta\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left(\left\langle\hat{w}_t,\Phi_m(a)\right\rangle\right)^2\Big|\mathcal{F}_{t-1}\right]\right]}_{=:\Gamma_2}. \quad (12)$$

Both $\Gamma_1$ and $\Gamma_2$ can be bounded by standard techniques established in the literature of adversarial linear bandits. We will see that $\Gamma_1$ is a telescoping series and is controlled in Lemma 21. While the second term $\Gamma_2$ is the variance of the estimated loss is bounded in Lemma 22. We defer the proof of both Lemma 21 and Lemma 22 to Appendix A.1.1. Plugging in the bounds on $\Gamma_1$ and $\Gamma_2$ into Equation (12) we get,

$$(1-\gamma)♠ = (1-\gamma)\cdot\mathbb{E}\left[\sum_{t=1}^{n}\left[\sum_{a\in\mathcal{A}}q_t(a)\left\langle\hat{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right] \quad (13)$$

$$\le \mathbb{E}\left[\sum_{i=1}^{n}\hat{\mathcal{K}}_m(a^*,y_t)\right] + \frac{1}{\eta}\log(|\mathcal{A}|) + \mathbb{E}\left[\sum_{t=1}^{n}\langle\hat{w}_t-\tilde{w}_t,\Phi_m(a^*)\rangle\right] + (e-2)\eta\mathcal{G}^4 mn$$

$$\le \mathbb{E}\left[\sum_{i=1}^{n}\hat{\mathcal{K}}_m(a^*,y_t)\right] + \frac{1}{\eta}\log(|\mathcal{A}|) + \frac{\epsilon n}{\mathcal{G}^2\eta} + (e-2)\eta\mathcal{G}^4 mn,$$

where the last inequality follows by Lemma 23. Also by Lemma 23 we get, $\diamondsuit \le \epsilon n/\mathcal{G}^2\eta$. Combining this with Equation (11) we get,

$$♣ \le \mathbb{E}\left[\sum_{i=1}^{n}\hat{\mathcal{K}}_m(a^*,y_t)\right] + \frac{1}{\eta}\log(|\mathcal{A}|) + \frac{2\epsilon n}{\mathcal{G}^2\eta} + (e-2)\eta\mathcal{G}^4 mn. \quad (14)$$

**Step 3:** Next, observe that the exploration term is bounded above as

$$\gamma\cdot\mathbb{E}\left[\sum_{t=1}^{n}\mathbb{E}\left[\sum_{a\in\mathcal{A}}\nu_{\mathcal{J}}^{\mathcal{A}}(a)\left\langle\tilde{w}_t,\Phi_m(a)\right\rangle\Big|\mathcal{F}_{t-1}\right]\right] \le 4\gamma\mathcal{G}^2 n, \quad (15)$$

where the above inequality follows by Lemma 20 and Cauchy-Schwartz inequality along with the fact that $\epsilon \le \mathcal{G}^2$.

**Step 4:** Putting these all these together into Equation (10) we get the desired bound on the finite dimensional regret

$$\mathcal{R}_n^m = \mathbb{E}\left[\sum_{t=1}^n \left(\hat{\mathcal{K}}_m(a_t, y_t) - \hat{\mathcal{K}}_m(a^*, y_t)\right)\right] \le 4\gamma\mathcal{G}^2 n + (e-2)\mathcal{G}^4\eta mn + \frac{2\epsilon n}{\mathcal{G}^2\eta} + \frac{1}{\eta}\log(|\mathcal{A}|).$$

Plugging the above bound on $\mathcal{R}_n^m$ into Equation (7) we get a bound on the expected regret as

$$\mathcal{R}_n \le 4\gamma\mathcal{G}^2 n + (e-2)\mathcal{G}^4\eta mn + 2\epsilon n + \frac{2\epsilon n}{\mathcal{G}^2\eta} + \frac{1}{\eta}\log(|\mathcal{A}|),$$

completing the proof. ■

### A.1.1. Technical Results used in Proof of Theorem 10

First let us focus on bounding $\Gamma_1$. A term analogous to $\Gamma_1$ also appears in the regret bound analysis for exponential weights in the adversarial linear bandits setting; we adapt those proof techniques here to work with biased estimates ($\hat{w}_t$).

**Lemma 21** *Let $\Gamma_1$ be as defined in Equation (12) then we have*

$$\Gamma_1 \ge -\eta\mathbb{E}\left[\sum_{i=1}^n \hat{\mathcal{K}}_m(a^*, y_t)\right] - \log(|\mathcal{A}|) - \eta\mathbb{E}\left[\sum_{t=1}^n \langle \hat{w}_t - \tilde{w}_t, \Phi_m(a^*)\rangle\right].$$

**Proof** Expanding $\Gamma_1$ we get

$$\begin{aligned}\Gamma_1 &= \mathbb{E}\left[\sum_{t=1}^n \log\left(\mathbb{E}_{a\sim q_t}\left[\exp\left(-\eta\left\langle \hat{w}_t, \Phi_m(a)\right\rangle\right)\right]\right)\right]\\ &\stackrel{(i)}{=} \mathbb{E}\left[\sum_{t=1}^n \log\left\{\frac{\sum_{a\in\mathcal{A}}\exp\left(-\eta\sum_{i=1}^{t-1}\left\langle \hat{w}_i, \Phi_m(a)\right\rangle\right)\cdot\exp\left(-\eta\left\langle \hat{w}_t, \Phi_m(a)\right\rangle\right)}{\sum_{a\in\mathcal{A}}\exp\left(-\eta\sum_{i=1}^{t-1}\left\langle \hat{w}_i, \Phi_m(a)\right\rangle\right)}\right\}\right]\\ &\stackrel{(ii)}{=} \mathbb{E}\left[\log\left(\sum_{a\in\mathcal{A}}\exp\left(-\eta\sum_{i=1}^n\left\langle \hat{w}_i, \Phi_m(a)\right\rangle\right)\right) - \log(|\mathcal{A}|)\right],\end{aligned} \tag{16}$$

where $(i)$ follows by the definition of $q_t(a)$ and $(ii)$ follows as the sum telescopes and we start of with the uniform distribution over $\mathcal{A}$. We have for any element $a' \in \mathcal{A}$,

$$\begin{aligned}\mathbb{E}\left[\log\left(\sum_{a\in\mathcal{A}}\exp\left(-\eta\sum_{i=1}^n\left\langle \hat{w}_i, \Phi_m(a)\right\rangle\right)\right)\right] &\ge -\mathbb{E}\left[\eta\sum_{i=1}^n\left\langle \hat{w}_i, \Phi_m(a')\right\rangle\right]\\ &= -\eta\mathbb{E}\left[\sum_{t=1}^n\langle \tilde{w}_t, \Phi_m(a')\rangle\right] - \eta\mathbb{E}\left[\sum_{t=1}^n\langle \hat{w}_t - \tilde{w}_t, \Phi_m(a')\rangle\right]\\ &= -\eta\mathbb{E}\left[\sum_{t=1}^n \hat{\mathcal{K}}_m(a', y_t)\right] - \eta\mathbb{E}\left[\sum_{t=1}^n\langle \hat{w}_t - \tilde{w}_t, \Phi_m(a')\rangle\right],\end{aligned}$$

where the last equality by Lemma 20. Choosing $a' = a^*$ and plugging this lower bound into Equation (16) completes the proof. ■

The next lemma controls of the variance of the expected loss – $\Gamma_2$. A term analogous to $\Gamma_2$ appears in the regret bound analysis of exponential weights in adversarial linear bandits which we adapt to our setting.

**Lemma 22** *Let $\Gamma_2$ be defined as in Equation (12) and the choice of parameters as specified in Theorem 10 then we have*

$$\Gamma_2 = (e-2)\eta\mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}_{a\sim q_t}\left[\langle \hat{w}_t, \Phi_m(a)\rangle^2 \middle| \mathcal{F}_{t-1}\right]\right] \leq (e-2)\mathcal{G}^4\eta mn/(1-\gamma). \tag{17}$$

**Proof** Note that by definition of $q_t$, we have that $(1-\gamma)q_t(a) \leq p_t(a)$. To ease notation let $\Sigma_t := \Sigma_m^{(t)} = \mathbb{E}_{p_t}\left[\Phi_m(x)\Phi_m(x)^\top\right]$. Taking expectation over the randomness in $\hat{w}_t$ for any fixed $a$ we have,

$$\begin{aligned}
\mathbb{E}_{a_t\sim q_t}\left[\langle \hat{w}_t(a_t), \Phi_m(a)\rangle^2\right] &= \Phi_m(a)^\top \mathbb{E}\left[\hat{w}_t\hat{w}_t^\top\right]\Phi_m(a) \\
&= \Phi_m(a)^\top \mathbb{E}_{a_t\sim p_t}\left[\mathcal{K}(a_t,y_t)^2\Sigma_t^{-1}\Phi_m(a_t)\Phi_m(a_t)^\top\Sigma_t^{-1}\right]\Phi_m(a) \\
&\leq \mathcal{G}^4\Phi_m(a)^\top\Sigma_t^{-1}\Phi_m(a).
\end{aligned}$$

where the second equality follows by the definition of $\hat{w}_t$. Given this calculation we now also take expectation over the choice of $a$ so we have,

$$\begin{aligned}
\mathbb{E}_{a\sim p_t}\left[\mathbb{E}_{a_t\sim p_t}\left[\langle \hat{w}_t, \Phi_m(a)\rangle^2\right]\right] &\leq \mathcal{G}^4\mathbb{E}_{a\sim p_t}\left[tr\left(\Phi_m(a)^\top\Sigma_t^{-1}\Phi_m(a)\right)\right] \\
&= \mathcal{G}^4 tr\left(\Sigma_t^{-1}\mathbb{E}_{a\sim p_t}\left[\Phi_m(a)\Phi_m(a)^\top\right]\right) = \mathcal{G}^4 tr\left(I_{m\times m}\right) = \mathcal{G}^4 m.
\end{aligned}$$

Summing over $t=1$ to $n$ establishes the result. ∎

We now prove the bound on the terms that arise out of our biased estimates.

**Lemma 23** *Let $\gamma = 4\eta\mathcal{G}^4 m$ and let $\epsilon \leq \mathcal{G}^2$, then for all $a \in \mathcal{A}$ and for all $t \in \{1,\ldots,n\}$ we have*

$$|\langle \hat{w}_t - \tilde{w}_t, \Phi_m(a)\rangle| \leq \frac{\epsilon}{\eta\mathcal{G}^2}.$$

**Proof** By the definition of $\tilde{w}_t$ and $\hat{w}_t$,

$$\begin{aligned}
\|\tilde{w}_t - \hat{w}_t\|_2 &= \left\|\left(\hat{\mathcal{K}}_m(y_t,a_t) - \mathcal{K}(y_t,a_t)\right)\left(\Sigma_m^{(t)}\right)^{-1}\Phi_m(a_t)\right\|_2 \\
&= |\hat{\mathcal{K}}_m(y_t,a_t) - \mathcal{K}(y_t,a_t)|\left\|\left(\Sigma_m^{(t)}\right)^{-1}\Phi_m(a_t)\right\|_2 \\
&\leq \epsilon\cdot\frac{m}{\gamma}\cdot(\mathcal{G}+\sqrt{\epsilon}),
\end{aligned}$$

where the inequality follows as $\hat{\mathcal{K}}_m$ is an $\epsilon$-approximation of $\mathcal{K}$, the minimum eigenvalue of $\Sigma_m^{(t)}$ is $\gamma/m$ by Proposition 35. So by Cauchy-Schwartz we get,

$$|\langle \hat{w}_t - \tilde{w}_t, \Phi_m(a)\rangle| \leq \|\tilde{w}_t - \hat{w}_t\|_2\|\Phi_m\|_2 \leq \frac{\epsilon m(\mathcal{G}+\sqrt{\epsilon})^2}{\gamma},$$

the claim now follows by the choice of $\gamma$ and by the condition on $\epsilon$. ∎

While using Hoeffding's inequality to arrive at Inequality (12) we assume that the estimate of the loss is lower bounded by $-1/\eta$. The next lemma help us establish that under the choice of $\gamma$ the exploration parameter in Theorem 10 this condition holds.

**Lemma 24** *Let $\epsilon \leq \mathcal{G}^2$ then for any $a \in \mathcal{A}$ and for all $t = 1,\ldots,n$ we have*

$$|\langle \hat{w}_t, \Phi_m(a)\rangle| \leq \mathcal{G}^2 \sup_{c,d\in\mathcal{A}}\left|\Phi_m(c)^\top\left(\mathbb{E}_{a\sim p_t}\left[\Phi_m(a)\Phi_m(a)^\top\right]\right)^{-1}\Phi_m(d)\right|.$$

*Further if the exploration parameter is $\gamma > 4\eta\mathcal{G}^4 m$ then we have a bound on the estimated loss at each round*

$$\eta|\langle \hat{w}_t, \Phi_m(a)\rangle| \leq 1, \qquad \forall a \in \mathcal{A}.$$

**Proof** Recall the definition of $\Sigma_m^{(t)} = \mathbb{E}_{p_t}\left[\Phi_m(a)\Phi_m(a)^\top\right]$ (we drop the index $t$ to lighten notation in this proof). The proof follows by plugging in the definition of the loss estimate $\hat{w}_t$,

$$\begin{aligned}
|\hat{w}_t^\top \Phi_m(a)| &= \left|\mathcal{K}(a_t, y_t)\left(\Sigma_m^{-1}\Phi_m(a_t)\right)^\top \Phi_m(a)\right| \\
&\le \underbrace{|\mathcal{K}(a_t, y_t)|}_{\le \mathcal{G}^2}\left|\left(\Sigma_m^{-1}\Phi_m(a_t)\right)^\top \Phi_m(a)\right| \\
&\le 4\mathcal{G}^2 \sup_{c,d\in\mathcal{A}} \left|\Phi_m(c)^\top \Sigma_m^{-1}\Phi_m(d)\right|. \qquad (18)
\end{aligned}$$

Now note that the matrix $\Sigma_m$ has its lowest eigenvalue lower bounded by $\gamma/m$ by Proposition 35 (see also discussion by 13). Thus we have,

$$\sup_{c,d\in\mathcal{A}} \left|\Phi_m(c)^\top \Sigma_m^{-1}\Phi_m(d)\right| \le \frac{4\mathcal{G}^2 m}{\gamma},$$

where the inequality above follows by the assumption that $\epsilon \le \mathcal{G}^2$. Combing this with Equation (18) yields the desired claim. ■

### A.2. Proof of Corollary 12

In this section we present the proof of Corollary 12, which establishes the regret bound under particular conditions on the eigen-decay of the kernel.

**Proof** [Proof of Corollary 12] Given our assumption $\mathcal{G} = 1$ and by the choice of $\gamma = 4\eta\mathcal{G}^4 m$ the regret bound becomes,

$$\mathcal{R}_n \le \underbrace{20\eta m n}_{=:R_1} + \underbrace{\frac{2\epsilon n}{\eta}}_{=:R_2} + \underbrace{\frac{1}{\eta}\log\left(|\mathcal{A}|\right)}_{=:R_3} + \underbrace{2\epsilon n}_{=:R_4}.$$

**Case 1:** First we assume $(C,\beta)$-polynomial eigen-value decay. By the results of Proposition 33 we have a sufficient condition on the choice of $m$ for $\hat{\mathcal{K}}_m$ to be an $\epsilon$-approximation of $\mathcal{K}$,

$$m = \left[\frac{4C\mathcal{B}^2}{(\beta-1)\epsilon}\right]^{1/\beta-1}.$$

With this choice of $m$ we equate the terms $R_1, R_2$ and $R_3$ with each other. This yields the choice,

$$\epsilon = \frac{\log(|\mathcal{A}|)}{2n}, \qquad \text{and} \qquad \eta^2 = \frac{\epsilon}{10m}.$$

Note that under this choice there exists a constant $n_0(\beta, C, \mathcal{B}, \log(|\mathcal{A}|)) > \log(|\mathcal{A}|)/2$ such that when $n > n_0$ then, $R_4 < R_1$. Also note when $n > \log(|A|)/2$ then $\epsilon < 1 = \mathcal{G}^2$ so the conditions of Theorem 10 are indeed satisfied. Plugging in these choice of $\epsilon, m$ and $\eta$ for $n > n_0$ yields,

$$\mathcal{R}_n \le 4R_1 = \sqrt{160}\cdot\left[\frac{2^{\beta+2}C\mathcal{B}^2}{\beta-1}\right]^{\frac{1}{2(\beta-1)}}\cdot(\log(|A|))^{\frac{\beta-2}{2(\beta-1)}}\cdot n^{\frac{\beta}{2(\beta-1)}}.$$

**Case 2:** Here we assume $(C,\beta)$-exponential eigen-value decay. Again by the results of Proposition 33 we have a sufficient condition for the choice of $m$ for $\hat{\mathcal{K}}_m$ to be an $\epsilon$- approximation of $\mathcal{K}$,

$$m = \frac{1}{\beta}\log\left(\frac{4C\mathcal{B}^2}{\beta\epsilon}\right).$$

Again as before, by equating $R_1, R_2, R_3$ yields $\epsilon = \log(|\mathcal{A}|)/(2n)$ and $\eta^2 = \epsilon/10m$. Again as with Case 1, there exists a constant $n_0(\beta, C, \mathcal{B}, \log(|\mathcal{A}|)) > \log(|\mathcal{A}|)/2$ such that when $n > n_0$ then, $R_4 < R_1$. Plugging in these choice of $\epsilon, m$ and $\eta$ for $n > n_0$ yields,

$$\mathcal{R}_n \leq 4R_1 = \sqrt{\frac{320 \cdot \log(|\mathcal{A}|)}{\beta} \cdot \log\left(\frac{40C\mathcal{B}^2 n}{\beta \log(|\mathcal{A}|)}\right) \cdot n}.$$

■

## B. Kernel principal component analysis

We review the basic principles underlying kernel principal component analysis (PCA). Let $\mathcal{K}$ be some kernel defined over $\mathcal{A} \subset \mathbb{R}^d$ and $x_1, \cdots, x_p \sim \mathbb{P}$ a probability measure over $\mathcal{A}$. Let us denote a feature map of $\mathcal{K}$ by $\Phi : \mathbb{R}^d \to \mathbb{R}^D$.

The goal of PCA is to extract a set of eigenvalues and eigenvectors from a sample covariance matrix. In kernel PCA we want to calculate the eigenvectors and eigenvalues of the sample *kernel* covariance matrix,

$$\hat{\Sigma} = \frac{1}{p}\sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top.$$

When working in a reproducing kernel Hilbert space $\mathcal{H}$ in which no feature map is explicitly available, an alternative approach is taken by working instead with the sample Gram matrix.

**Lemma 25** *Let $\Phi(x_1), \cdots, \Phi(x_p)$ be $p$ points in $\mathcal{H}$. The eigenvalues of the sample covariance matrix, $\frac{1}{p}\sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top$ equal the eigenvalues of the sample Gram matrix $\mathbb{K} \in \mathbb{R}^{p\times p}$, where the sample Gram matrix is defined entry-wise as $\mathbb{K}_{ij} = \frac{\mathcal{K}(x_i,x_j)}{p}$.*

**Proof** Let $X \in \mathbb{R}^{p\times D}$ be such that the $i^{th}$ row is $\frac{\Phi(x_i)}{\sqrt{p}}$. The singular value decomposition (SVD) of $X$ is

$$X = UDV^\top,$$

with $U \in \mathbb{R}^{p\times p}$, $D \in \mathbb{R}^{p\times D}$ and $V \in \mathbb{R}^{D\times D}$. Therefore $X^\top X = VD^\top DV^\top$ and $XX^\top = UDD^\top U^\top$. We identify $X^\top X$ as the sample covariance matrix and $XX^\top$ as the sample Gram matrix. Since $DD^\top$ and $D^\top D$ are both diagonal and have the same nonzero values this establishes the claim. ■

Another insight used in kernel PCA procedures is the observation that the span of the eigenvectors corresponding to nonzero eigenvalues of the sample covariance matrix $\frac{1}{p}\sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top$ is a subspace of the span of the data-points $\{\Phi(x_i)\}_{i=1}^{p}$. This means that any eigenvector $v$ corresponding to a nonzero eigenvalue for the second moment sample covariance matrix can be written as a linear combination of the $p-$datapoints, $v_i = \sum_{j=1}^{p} \omega_{ij}\Phi(x_j)$ ($\omega_{ij}$ denotes the $j^{th}$ component of $\omega_i \in \mathbb{R}^p$). Observe that $v_i$ are the eigenvectors of the sample covariance matrix, so we have

$$\left[\frac{1}{p}\sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top\right]\left(\sum_{j=1}^{p} \omega_{ij}\Phi(x_j)\right) = \mu_i \sum_{j=1}^{p} \omega_{ij}\Phi(x_j).$$

This implies we may consider solving the equivalent system

$$\mu_i \left(\langle \Phi(x_k), v_i\rangle_{\mathcal{H}}\right) = \left\langle \Phi(x_k), \left(\frac{1}{p}\sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top\right) v_i \right\rangle_{\mathcal{H}} \quad \forall k = 1, \cdots, p. \tag{19}$$

Substituting $v_i = \sum_{j=1}^{p} \omega_{ij}\Phi(x_j)$ into Equation (19), and using the definition of $\mathbb{K}$ we obtain

$$\mu_i \mathbb{K}\omega_i = \mathbb{K}^2\omega_i.$$

To find the solution of this last equation we solve the eigenvalue problem,

$$\mathbb{K}\omega_i = \mu_i \omega_i.$$

Once we solve for $\alpha_i$ we can recover the eigenvector of the sample covariance matrix by setting $v_i = \sum_{j=1}^{p} \omega_{ij}\Phi(x_j)$.

## C. Proxy Kernel Properties

In this section we prove Theorem 9. We reuse the notation introduced in Section 2.1 which we recall here.

Let $\{\mu_j\}_{j=1}^{\infty}$ be the Mercer's eigenvalues of a kernel $\mathcal{K}$ under measure $\mathbb{P}$ with eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$, and we assume that $\sup_{j\in\mathbb{N}} \sup_{x\in\mathcal{A}} |\phi_j(x)| \leq \mathcal{B}$ for some $\mathcal{B} < \infty$. Let $m(\epsilon)$ be such that $\sum_{j=m+1}^{\infty} \mu_j \leq \frac{\epsilon}{4\mathcal{B}^2}$ and denote the $m^{th}$ eigen-gap as $\delta_m = \frac{1}{2}(\mu_m - \mu_{m+1})$. Denote by $S_m$ and $\hat{S}_m$ the subspaces spanned by the first $m$ eigenvectors of the covariance matrix $\mathbb{E}_{x\sim\mathbb{P}}\left[\Phi(x)\Phi(x)^{\top}\right]$ and a sample covariance matrix $\frac{1}{p}\sum_{i=1}^{p}\Phi(x_i)\Phi(x_i)^{\top}$ respectively. Define $P_{S_m}$ and $P_{\hat{S}_m}$ to be the projection operators to $S_m$ and $\hat{S}_m$. Recall the definition of

$$\hat{\mathcal{K}}_m^o(x,y) = \langle P_{S_m}(\Phi(x)), P_{S_m}(\Phi(y))\rangle_{\mathcal{H}} = \langle \Phi_m^o(x), \Phi_m^o(y)\rangle_{\mathcal{H}},$$

a deterministic approximate kernel and the stochastic proxy approximate kernel

$$\hat{\mathcal{K}}_m(x,y) = \langle P_{\hat{S}_m}(\Phi(x)), P_{\hat{S}_m}(\Phi(y))\rangle_{\mathcal{H}},$$

with associated feature map $\Phi_m(x) = P_{\hat{S}_m}\Phi(x)$. We first prove Lemma 6 restated here.

**Lemma 6** *Given $\epsilon > 0$, let $\{\mu_j\}_{j=1}^{\infty}$ be the Mercer operator eigenvalues of $\mathcal{K}$ under a finite Borel measure $\mathbb{P}$ with support $\mathcal{A}$ and eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$ with $\mu_1 \geq \mu_2 \geq \cdots$. Further assume that $\sup_{j\in\mathbb{N}} \sup_{x\in\mathcal{A}} |\phi_j(x)| \leq \mathcal{B}$ for some $\mathcal{B} < \infty$. Let $m(\epsilon)$ be such that $\sum_{j=m+1}^{\infty} \mu_j \leq \frac{\epsilon}{4\mathcal{B}^2}$. Then the kernel induced by a truncated feature map,*

$$\Phi_m^o(x) := \begin{cases} \sqrt{\mu_i}\phi_i(x) & \text{if } i \leq m \\ 0 & \text{o.w.} \end{cases} \tag{4}$$

*induces a kernel $\hat{\mathcal{K}}_m^o := \langle \Phi_m^o(x), \Phi_m^o(y)\rangle_{\mathcal{H}} = \sum_{j=1}^{m} \mu_j\phi_j(x)\phi_j(y)$, for all $(x,y) \in \mathcal{A}\times\mathcal{A}$ that is an $\epsilon/4$-approximation of $\mathcal{K}$.*

**Proof** By definition, for all $x, y \in \mathcal{A}$

$$\mathcal{K}(x,y) - \hat{\mathcal{K}}_m^o(x,y) = \sum_{j=m+1}^{\infty} \mu_j\phi_j(x)\phi_j(y) \leq \sum_{j=m+1}^{\infty} \mu_j|\phi_j(x)\phi_j(y)| \leq \sum_{j=m+1}^{\infty} \mu_j\mathcal{B}^2 \leq \frac{\epsilon}{4}.$$

The reverse inequality; $\hat{\mathcal{K}}_m^o(x,y) - \mathcal{K}(x,y) \leq \frac{\epsilon}{4}$, is also true therefore,

$$|\mathcal{K}(x,y) - \hat{\mathcal{K}}_m^o(x,y)| \leq \frac{\epsilon}{4},$$

for all $x, y \in \mathcal{A}$. ■

We now state and prove an expanded version of Theorem 9 (where $w = \min(\sqrt{\epsilon}/2, \delta_m/2)$) which is used to establish the $\epsilon$-approximability of the stochastic kernel $\hat{\mathcal{K}}_m$.

**Theorem 26** *Let $\epsilon, m, \mathbb{P}$ be as in Lemma 6. Define the $m$-th level eigen-gap as $\delta_m = \frac{1}{2}(\mu_m - \mu_{m+1})$. Also let $B_m = \frac{2\mathcal{G}^2}{\delta_m}\left(1 + \sqrt{\frac{\alpha}{2}}\right)$, $\delta_m/2 > w > 0$ and $p \geq \frac{B_m^2\mathcal{G}^2}{w^2}$. The finite dimensional proxies $\hat{\mathcal{K}}_m^o$ and $\hat{\mathcal{K}}_m$ satisfy the following properties with probability $1 - e^{-\alpha}$:*

1. $|\mathcal{K}(x,y) - \hat{\mathcal{K}}_m(x,y)| \leq \frac{\epsilon}{4} + \sqrt{\epsilon}w + w^2$.

2. $|\hat{\mathcal{K}}_m(x,y) - \hat{\mathcal{K}}^o_m(x,y)| \le w^2, \qquad \forall\, x, y \in \mathcal{A}.$

3. *The Mercer operator eigenvalues* $\mu_1^{(m)} \ge \cdots \ge \mu_m^{(m)}$ *and* $\mu_1 \ge \cdots \ge \mu_m$ *of* $\hat{\mathcal{K}}_m$ *and* $\hat{\mathcal{K}}^o_m$ *follow* $\sup_{i=1,\cdots,m} |\mu_i^{(m)} - \mu_i| \le w^2$.

Theorem 26 shows that, as long as sufficiently samples $p(m)$ are used, with high probability $\hat{\mathcal{K}}_m$ is uniformly close to $\hat{\mathcal{K}}^o_m$ and therefore to $\mathcal{K}$. We prove this theorem by a series of lemmas and auxiliary theorems. We first prove part (1) and (2) and establish that under mild conditions on $\mathcal{K}$ we can extract a finite dimensional proxy kernel $\hat{\mathcal{K}}_m$ by truncating the eigen-decomposition of $\mathcal{K}$ and estimating a feature map with samples. We leverage a kernel PCA result by (50) to construct $\hat{\mathcal{K}}_m$.

**Theorem 27** *(Adapted from Theorem 4 in 50) If* $m, p, S_m, \hat{S}_m, \delta_m, B_m$ *and* $\alpha$ *are defined as in Theorem 26 then with probability* $1 - \exp(-\alpha)$ *we have*

$$\|P_{S_m} - P_{\hat{S}_m}\|_F \le \frac{B_m}{\sqrt{p(m)}}. \tag{20}$$

*In particular,*

$$\hat{S}_m \subset \left\{ g + h, g \in S_m, h \in S_m^\perp, \|h\|_{\mathcal{H}} \le \frac{2B_m}{\sqrt{p}} \|g\|_{\mathcal{H}} \right\}.$$

Now using this theorem we prove Part (1) of Theorem 26.

**Lemma 28** *With probability* $1 - e^{-\alpha}$ *we have,*

$$|\mathcal{K}(x,y) - \hat{\mathcal{K}}_m(x,y)| \le \frac{\epsilon}{4} + \sqrt{\epsilon} w + w^2, \qquad \forall\, x, y \in \mathcal{A}.$$

**Proof** First we show this holds for $x = y$.

$$\begin{aligned}\|\Phi(x) - P_{\hat{S}_m}(\Phi(x))\|_{\mathcal{H}} &\overset{(i)}{\le} \|\Phi(x) - P_{S_m}(\Phi(x))\|_{\mathcal{H}} + \|P_{S_m}(\Phi(x)) - P_{\hat{S}_m}(\Phi(x))\|_{\mathcal{H}} \\ &\overset{(ii)}{\le} \frac{\sqrt{\epsilon}}{2} + \|\Phi(x)\|_{\mathcal{H}} \|P_{S_m} - P_{\hat{S}_m}\|_{op} \overset{(iii)}{\le} \frac{\sqrt{\epsilon}}{2} + \mathcal{G}\frac{B_m}{\sqrt{p(m)}} \overset{(iv)}{\le} \frac{\sqrt{\epsilon}}{2} + w,\end{aligned}$$

where $(i)$ follows by triangle inequality, $(ii)$ is by the fact that $P_{S_m}(\Phi(x))$ is an $\epsilon/4$ approximation of $\mathcal{K}$, $(iii)$ follows by Theorem 27 and $(iv)$ is by the choice of $p(m)$. Therefore with probability at least $1 - e^{\alpha}$ for all $x \in \mathcal{A}$

$$|\mathcal{K}(x,x) - \hat{\mathcal{K}}_m(x,x)| \le \frac{\epsilon}{4} + \sqrt{\epsilon} w + w^2.$$

Now we prove the statement for general $x, y \in \mathcal{A}$. We write $\Phi(x) = \Phi_m(x) + h_x$ and $\Phi(y) = \Phi_m(y) + h_y$. The above calculation implies that $\|h_x\|_{\mathcal{H}} \le \frac{\sqrt{\epsilon}}{2} + w$ and $\|h_y\|_{\mathcal{H}} \le \frac{\sqrt{\epsilon}}{2} + w$. We now expand $\mathcal{K}(x,y)$ to get

$$\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} = \langle \Phi_m(x), \Phi_m(y) \rangle_{\mathcal{H}} + \langle h_x, \Phi_m(y) \rangle_{\mathcal{H}} + \langle \Phi_m(x), h_y \rangle_{\mathcal{H}} + \langle h_x, h_y \rangle_{\mathcal{H}}.$$

Since $h_x$ and $h_y$ both live in $\hat{S}_m^\perp$:

$$\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} = \langle \Phi_m(x), \Phi_m(y) \rangle_{\mathcal{H}} + \langle h_x, h_y \rangle_{\mathcal{H}}.$$

Rearranging terms,

$$|\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} - \langle \Phi_m(x), \Phi_m(y) \rangle_{\mathcal{H}}| = |\langle h_x, h_y \rangle_{\mathcal{H}}| \le \frac{\epsilon}{4} + \sqrt{\epsilon} w + w^2.$$

This establishes the claim. ∎

We now move on to the proof of Part (2) in Theorem 26.

**Lemma 29** *If* $p, B_m$ *are chosen as stated in Theorem 26 we have*

$$|\hat{\mathcal{K}}_m(x,y) - \hat{\mathcal{K}}^o_m(x,y)| \le w^2 \qquad \forall x, y \in \mathcal{A} \tag{21}$$

**Proof** The feature map for $\hat{\mathcal{K}}_m^o$ is $P_{S_m}(\Phi(x))$ for all $x \in \mathcal{A}$ while $P_{\hat{S}_m}(\Phi(x))$ is the feature map for $\hat{\mathcal{K}}_m$. We first show that for all $x \in \mathcal{A}$,

$$\|P_{S_m}(\Phi(x)) - P_{\hat{S}_m}(\Phi(x))\|_{\mathcal{H}} \leq \|\Phi(x)\|_{\mathcal{H}} \|P_{S_m} - P_{\hat{S}_m}\|_{op} \leq \mathcal{G}\frac{B_m}{\sqrt{p(m)}} \leq w,$$

where the second inequality follows by applying Theorem 27 and the last inequality follows by the choice of $B_m$. A similar argument as the one used in the proof of Lemma 28 lets us then conclude that,

$$|\hat{\mathcal{K}}_m(x,y) - \hat{\mathcal{K}}_m^o(x,y)| \leq w^2, \qquad \forall x, y \in \mathcal{A}.$$

■

We now proceed to prove part (3) of Theorem 26. We will show that the Mercer operator eigenvalues of $\hat{\mathcal{K}}_m$ are close to Mercer operator eigenvalues of $\hat{\mathcal{K}}_m^o$. We first recall a useful result by (37).

**Theorem 30** *(Adapted from Theorem 3.3 in 37) Let $\mathcal{K}$ be a kernel over $\mathcal{A} \times \mathcal{A}$ such that $\sup_{x\in\mathcal{A}} \mathcal{K}(x,x) \leq \mathcal{G}^2$. Also let $\hat{\mu}_1 \geq \hat{\mu}_2 \geq \cdots \geq \hat{\mu}_N$ be the eigenvalues of the Gram matrix $(\mathcal{K}(x_i,x_j)/N)_{i,j=1}^N$ for $\{x_i\}_{i=1}^N \sim \mathbb{P}$. Then there exists a universal constant $c$ such that for every $t > 0$*

$$\mathbb{P}\left[\sup_{i\in 1,\ldots,N} |\hat{\mu}_i - \mu_i| \geq t\right] \leq 2\exp\left(-\frac{ct}{\mathcal{G}^2}\sqrt{\frac{N}{\log(N)}}\right), \tag{22}$$

*where for $i > N$ we define $\hat{\mu}_i = 0$.*

**Proposition 31** *The top $m$ eigenvalues of the sample kernel covariance matrix equal that of the Gram matrix.*

Recall that we established this proposition in Appendix B as Lemma 25. Further note that for any set of samples $x_1, \cdots, x_N \sim \mathbb{P}$, the Gram matrices of $\hat{\mathcal{K}}_m^o$ ($\mathbb{K}_m^o(N)$) and $\hat{\mathcal{K}}_m$ ($\mathbb{K}_m(N)$) are close in Frobenius norm as the matrices are close element-wise by part (2) of Theorem 26.

$$\|\mathbb{K}_m^o(N) - \mathbb{K}_m(N)\|_F \leq w^2.$$

Let $\hat{\mu}_1^{(m,o)} \geq \hat{\mu}_2^{(m,o)} \geq \cdots \hat{\mu}_N^{(m,o)}$ and $\hat{\mu}_1^{(m)} \geq \hat{\mu}_2^{(m)} \geq \cdots \hat{\mu}_N^{(m)}$ be the eigenvalues of $\mathbb{K}_m^o(N)$ and $\mathbb{K}_m(N)$ respectively. For both of these Gram matrices only the top $m$ out of $N$ eigenvalues will be nonzero, since both kernels are $m-$dimensional. By the Wielandt-Hoffman inequality (30) this implies that the *ordered eigenvalues* are close,

$$\sup_{i=1,\cdots N} |\hat{\mu}_i^{(m,o)} - \hat{\mu}_i^{(m)}| \leq w^2.$$

Theorem 30 and Proposition 31 together imply the statement of Theorem 30 with the Gram matrix replaced by the sample covariance matrix holds.

**Theorem 32** *The Mercer operator eigenvalues $\mu_1^{(m)} \geq \cdots \geq \mu_m^{(m)}$ and $\mu_1 \geq \cdots \geq \mu_m$ of $\hat{\mathcal{K}}_m$ and $\hat{\mathcal{K}}_m^o$ follow*

$$\sup_{i=1,\cdots,m} |\mu_i^{(m)} - \mu_i| \leq w^2. \tag{23}$$

**Proof** We will use the probabilistic method. By Theorem 30, for every $t > 0$ there is $N(t) \in \mathbb{N}$ large enough such that probability of the event – the eigenvalues of both sample Gram matrices $\mathbb{K}_m(N)$ and $\mathbb{K}_m^o(N)$ be uniformly close to the Mercer operator eigenvalues $\mu_1^{(m)} \geq \cdots \geq \mu_m^{(m)}$ and $\mu_1 \geq \cdots \geq \mu_m$ – is greater than zero. By triangle inequality this implies that for all $t > 0$

$$\begin{aligned}\sup_{i=1,\cdots,m} |\mu_i^{(m)} - \mu_i| &\leq \sup_{i_1} |\mu_{i_1}^{(m)} - \hat{\mu}_{i_1}^{(m)}| + \sup_{i_2} |\hat{\mu}_{i_2}^{(m)} - \hat{\mu}_{i_2}^{(m,0)}| + \sup_{i_3} |\hat{\mu}_{i_3}^{(m,0)} - \mu_{i_3}| \\ &\leq t + w^2 + t = w^2 + 2t.\end{aligned}$$

Taking the limit as $t \to 0$ yields the result. ■

### C.1. Bounds on the effective dimension $m$

In this section we establish bounds on the effective dimension $m$ under different eigenvalue decay assumptions.

**Proposition 33** *Let the conditions stated in Theorem 9 and Lemma 6 hold.*

1. *When the kernel $\mathcal{K}$ has $(C,\beta)$-polynomial eigenvalue decay then*

$$m \geq \left[\frac{4C\mathcal{B}^2}{(\beta-1)\epsilon}\right]^{1/\beta-1},$$

*suffices for $\hat{\mathcal{K}}^o_m$ to be an $\epsilon/4$-approximation of $\mathcal{K}$ and therefore for $\hat{\mathcal{K}}_m$ to be an $\epsilon$-approximation of $\mathcal{K}$.*

2. *When the kernel $\mathcal{K}$ has $(C,\beta)$-exponential eigenvalue decay then*

$$m \geq \frac{1}{\beta}\log\left(\frac{4C\mathcal{B}^2}{\beta\epsilon}\right),$$

*suffices for $\hat{\mathcal{K}}^o_m$ to be an $\epsilon/4$-approximation of $\mathcal{K}$ and therefore for $\hat{\mathcal{K}}_m$ to be an $\epsilon$-approximation of $\mathcal{K}$.*

**Proof** We need to ensure that the assumption in Lemma 6 holds. That is,

$$\sum_{j=m+1}^{\infty}\mu_j \leq \frac{\epsilon}{4\mathcal{B}^2}.$$

We will prove the bound assuming a $(C,\beta)$-polynomial eigenvalue decay, the calculation is similar when we have exponential eigenvalue decay. Note that,

$$\sum_{j=m+1}^{\infty}\mu_j \leq \sum_{j=m+1}^{\infty}Cj^{-\beta} \leq \int_m^{\infty}Cx^{-\beta}dx = \frac{C}{\beta-1}\frac{1}{m^{\beta-1}}.$$

We demand that,

$$\frac{C}{\beta-1}\frac{1}{m^{\beta-1}} \leq \frac{\epsilon}{4\mathcal{B}^2},$$

rearranging terms yields the desired claim. ■

## D. Properties of the Covariance matrix – $\Sigma_m^{(t)}$

We borrow the notation from Section 2.1. In this section we let $\mu_m$ be the smallest nonzero eigenvalue of $\mathbb{E}_{x\sim\nu}\left[\Phi_m(x)\Phi_m(x)^\top\right]$ where $\nu$ is the exploration distribution over $\mathcal{A}$.

**Lemma 34** *Let $\mu_m^{(t)}$ be the $m$-th (smallest) eigenvalue of $\Sigma_m^{(t)}$. Then we have*

$$\mu_m^{(t)} \geq \gamma\mu_m.$$

**Proof** Recall that in each step we set $p_t = (1-\gamma)q_t + \gamma\nu$. Let $v \in \mathcal{H}$ be a vector with norm 1.

$$v^\top\Sigma_m^{(t)}v = (1-\gamma)\cdot v^\top\mathbb{E}_{x\sim q_t}\left[\Phi_m(x)\Phi_m(x)^\top\right]v + \gamma\cdot v^\top\mathbb{E}_{x\sim\nu}\left[\Phi_m(x)\Phi_m(x)^\top\right]v.$$

Since both summands on the RHS are nonnegative, this quantity at least achieves a value of $\gamma\cdot v^\top\mathbb{E}_{x\sim\nu}\left[\Phi_m(x)\Phi_m(x)^\top\right]v \geq \gamma\mu_m$. ■

Observe that by our discussion in Appendix H.1, the minimum eigenvalue when the distribution is $\nu_{\mathcal{J}}$ (John's distribution) over $\Phi_m(\mathcal{A})$, then $\mu_m = 1/m$. That is, if $\nu_{\mathcal{J}}^{\mathcal{A}}$ is the exploration distribution over $\mathcal{A}$ then $\mu_m = 1/m$.

**Proposition 35** *If $\nu_{\mathcal{J}}^{\mathcal{A}}$ is the exploration distribution then we have*

$$\mu_m^{(t)} \geq \frac{\gamma}{m}.$$

### D.1. Finite Sample Analysis

Next we analyze the sample complexity of the operation of building the second moment matrix in Algorithm 2.3 using samples. Let $\hat{\Sigma}_m^{(t)}$ be the second moment matrix estimate built by using $x_1, \cdots, x_r$ drawn i.i.d. from $p_t$.

$$\hat{\Sigma}_m^{(t)} = \frac{1}{r}\sum_{i=1}^{r} \Phi_m(x_i)\Phi_m(x_i)^\top.$$

We will show how to chose $r$ appropriately to preserve the validity of the regret bound when we use $\hat{\Sigma}_m^{(t)}$ (built using finite samples) instead of $\Sigma_m^{(t)}$. First we present some observations.

**Remark 36 (Covariance eigenvalues are Mercer's eigenvalues)** *The eigenvalues* $\mu_1^{(t)} \geq \cdots \geq \mu_m^{(t)}$ *of* $\mathbb{E}_{x\sim p_t}\left[\Phi_m(x)\Phi_m(x)^\top\right]$ *are exactly Mercer operator eigenvalues for* $\hat{\mathcal{K}}_m$ *under* $p_t$.

**Remark 37 (Sample covariance and Gram matrix have the same eigenvalues)** *Assume* $r \geq m$. *Let* $x_1, \cdots, x_r \sim p_t$. *The eigenvalues of the sample covariance* $\hat{\Sigma}_m^{(t)}$ *coincide with the top* $m$ *eigenvalues of the Gram matrix* $\mathbb{K}_{m,p}^{(t)} = \left(\hat{\mathcal{K}}_m(x_i, x_j)\right)_{i,j=1}^{p}$.

We formalize the above remark in Lemma 41. We will use an auxiliary lemma by (50) which we present here for completeness.

**Lemma 38** *(Lemma 1 in 50) Let* $\mathcal{K}'$ *be a kernel over* $\mathcal{X}\times\mathcal{X}$ *such that* $\sup_{x\in\mathcal{X}} \mathcal{K}'(x,x) \leq \mathcal{G}'$. *Let* $\Sigma'$ *be the covariance of* $\Phi'(x)$, $x \sim \mathbb{P}$. *If* $\hat{\Sigma}_r'$ *is the sample covariance built by using* $r$ *samples* $x_1, \cdots, x_r \sim \mathbb{P}$, *with probability* $1-\exp(-\delta)$:

$$\|\Sigma' - \hat{\Sigma}_p'\|_{op} \leq \frac{2\mathcal{G}'}{\sqrt{r}}\left(1+\sqrt{\frac{\delta}{2}}\right).$$

The following lemma will allow us to derive an operator norm bound between the inverse matrices $\left(\Sigma_m^{(t)}\right)^{-1}$ and $\left(\hat{\Sigma}_m^{(t)}\right)^{-1}$ from an operator norm bound between the matrices $\Sigma_m^{(t)}$ and $\hat{\Sigma}_m^{(t)}$.

**Lemma 39** *If* $\|A-B\|_{op} \leq s$, *then* $\|A^{-1}-B^{-1}\|_{op} \leq \frac{s}{\lambda_{min}(A)\lambda_{min}(B)}$, *where* $\lambda_{min}(A)$ *and* $\lambda_{min}(B)$ *denote the minimum eigenvalues of* $A$ *and* $B$ *respectively.*

**Proof** The following equality holds:

$$A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}.$$

Applying Cauchy-Schwartz for spectral norms yields the desired result. ■

We are now ready to show that given enough samples $r$, the operator norm between the inverse covariance and the inverse sample covariance is small.

**Lemma 40** *Let* $g : \mathbb{R}^+_{(1,\infty)} \to \mathbb{R}$ *be defined as* $g(x)$ *is the value such that* $\frac{g(x)}{\log(g(x))} = x$. *If the number of samples*

$$r \geq \max\left( g\left(\left[\frac{(\ln(2)+\zeta)2\mathcal{G}}{c\gamma\mu_m}\right]^2\right), \left(\frac{4\mathcal{G}(1+\sqrt{\frac{\zeta}{2}})}{(\gamma\mu_m)^2\epsilon_1}\right)^2\right),$$

*where* $c$ *is the same constant as in Theorem 30, then with probability* $1-2e^{-\zeta}$:

$$\left\|\left(\Sigma_m^{(t)}\right)^{-1} - \left(\hat{\Sigma}_m^{(t)}\right)^{-1}\right\|_{op} \leq \epsilon_1.$$

**Proof** We start by showing that if $r$ follows the requirements stated in the lemma above, then the minimum eigenvalue of $\hat{\Sigma}_m^{(t)}$ is lower bounded by $\frac{\gamma\mu_m}{2}$ with probability $1-\exp(-\zeta)$. We invoke Theorem 30 to prove this. Let us denote by $\mu_1^{(t)} \geq \cdots \geq \mu_m^{(t)}$ and $\hat{\mu}_1^{(t)} \geq \cdots \geq \hat{\mu}_m^{(t)}$ the eigenvalues of $\Sigma_m^{(t)}$ and $\hat{\Sigma}_m^{(t)}$ respectively.

We want to ensure that the probability of $\sup_i |\mu_i^{(t)} - \hat{\mu}_i^{(t)}| \geq \frac{\gamma\mu_m}{2}$ be less than $e^{-\zeta}$. Again by invoking Theorem 30, this is true if $exp(-\zeta) \leq 2\exp\left(-\frac{c\gamma\mu_m}{2\mathcal{G}}\sqrt{\frac{r}{\log(r}}\right)$. This yields the condition,

$$\frac{r}{\log(r)} \geq \left[\frac{(\ln(2)+\zeta)2\mathcal{G}}{c\gamma\mu_m}\right]^2.$$

This together with triangle inequality (as $\mu_m^{(t)} \geq \mu_m \geq \gamma\mu_m$) ensures that if $r \geq g\left(\left[\frac{(\ln(2)+\zeta)2\mathcal{G}}{c\gamma\mu_m}\right]^2\right)$, then with probability $1-\exp(-\zeta)$,

$$\hat{\mu}_m \geq \frac{\gamma\mu_m}{2}.$$

Setting $A = \Sigma_m^{(t)}$ and $B = \hat{\Sigma}_m^{(t)}$ and invoking the concentration inequality Lemma 38, we have that

$$\left\|\Sigma_m^{(t)} - \hat{\Sigma}_m^{(t)}\right\|_{op} \leq \frac{(\gamma\lambda_m)^2\epsilon_1}{2},$$

with probability $1-\exp(-\zeta)$ as choose $r$ to satisfy

$$\frac{2\mathcal{G}}{2\sqrt{r}}\left(1+\sqrt{\frac{\zeta}{2}}\right) = \frac{(\gamma\mu_m)^2\epsilon_1}{2}.$$

As the matrices $A$ and $B$ are close with high probability, Lemma 39 proves that the inverses are also close,

$$\left\|(\Sigma_m(t))^{-1} - \left(\hat{\Sigma}_m^{(t)}\right)^{-1}\right\|_{op} \leq \epsilon_1.$$

with the same probability. By union bound as long as $r \geq \max\left(g\left(\left[\frac{(\ln(2)+\zeta)2\mathcal{G}}{c\gamma\lambda_m}\right]^2\right), \left(\frac{4\mathcal{G}(1+\sqrt{\frac{\zeta}{2}})}{(\gamma\lambda_m)^2\epsilon_1}\right)^2\right)$ the stated claim holds with probability $1-2\exp(-\zeta)$. ■

### D.1.1. Auxiliary Lemmas

Let us denote the pseudo-inverse of a symmetric matrix $A$ by $A^\dagger$. We now prove Lemma 41 that formalizes the connection between the eigenvalues the Gram matrix and sample covariance matrix.

**Lemma 41** *For any $x, y \in \mathcal{A}$:*

$$\Phi_m(x)^\top\left(\hat{\Sigma}_m^{(t)}\right)^{-1}\Phi_m(y) = A_x^\top\left(\mathbb{K}_{m,p}^{(t)}\right)^{2\dagger}A_y^\top,$$

*where* $A_x = \left(\hat{\mathcal{K}}_m(x,x_1),\cdots,\hat{\mathcal{K}}_m(x,x_p)\right)^\top$ *and* $A_y = \left(\hat{\mathcal{K}}_m(y,y_1),\cdots,\hat{\mathcal{K}}_m(y,x_p)\right)^\top$.

**Proof** The claim can be verified by a singular value decomposition of both sides. ■

Given Lemma 40 we also prove a bound on the distance between the estimates of adversarial actions generated in Algorithm 2.3. Define $\tilde{w}_t^{(2)} := \left(\hat{\Sigma}_m^{(t)}\right)^{-1}\Phi_m(a_t)\mathcal{K}(a_t,w_t)$ and let $\hat{w}_t := \tilde{w}_t^{(1)} = \left(\Sigma_m^{(t)}\right)^{-1}\Phi_m(a_t)\mathcal{K}(a_t,w_t)$.

**Corollary 42** *We have that*

$$\|\tilde{w}_t^{(2)} - \tilde{w}_t^{(1)}\|_{\mathcal{H}} \leq \epsilon_1 \mathcal{G}.$$

In other words, the bias resulting from using the sample covariance instead of the true covariance is of order $\epsilon_1$ as long as we take enough samples $p$ at each time step. We can drive $\epsilon_1$ to be as low as we like by choosing enough samples and hence this bias does not determine the rate in the regret bounds in Theorem 10.

## E. Proof of the lower bound

**Theorem 43** *Let $\tilde{\mathcal{A}} \subset \mathbb{R}^D$ with $\tilde{\mathcal{A}} = \{(A_j)_{j=1}^{\infty}$ s.t. $|A_j| \leq 1\ \forall j\}$ and $\tilde{\mathcal{W}} \subset \mathbb{R}^D$ with $\tilde{\mathcal{W}} = \{(w_j)_{j=1}^{\infty}$ s.t. $|w_j| \leq \mu_j\ \forall j\}$ be the action sets of a player and an adversary. For any integer $n$ and for any algorithm used by the player, there exist a strategy for the adversary such that $\mathcal{R}_n \geq \tilde{\Omega}\left(n^{\frac{\beta+1}{2\beta}}\right)$ whenever $\mu_j = \tilde{\mathcal{O}}(\frac{1}{j^\beta})$. When the decay is exponential $\mu_j = \mathcal{O}(e^{-\beta j})$ then $\mathcal{R}_n \geq \tilde{\Omega}\left(n^{1/2}\right)$.*

**Proof** We start by proving the result above for the following oblivious noisy feedback model:

1 At the beginning of time the adversary selects a vector in $\mathcal{W}$.

2 Every round the player selects an action vector $\tilde{a} \in \tilde{\mathcal{A}}$.

3 The player experiences $\langle \tilde{a}, \tilde{w}\rangle_{\mathcal{H}} + \eta$, where $\eta \sim \mathcal{N}(0, I)$.

Under this feedback model, the bounded-ness assumption of the losses is lost. We fix this at the end of the proof by considering a bounded loss model for an adversarial non oblivious adversary.

The adversarial case will be dealt with later by defining $\tilde{\mathcal{W}}^0 \subset \{c, c'\} \times \tilde{\mathcal{W}}$ for some $c, c' \in [0, 1]$ and $\tilde{\mathcal{A}}^0 = \{(1, \tilde{a}),\ \tilde{a} \in \tilde{\mathcal{A}}\}$. We then analyze an adversary that at time $t$ plays vectors $\tilde{w}_t^0 = (c_t, \tilde{w}^*)$ where $\tilde{w}^*$ is independent of $t$ and $c_t \in \{c, c'\}$ and is chosen by sampling independently from a Bernoulli distribution over $\{c, c'\}$. In this case $\langle \tilde{a}^0, \tilde{w}^0\rangle_{\mathcal{H}^0} = \langle \tilde{a}, \tilde{w}\rangle_{\mathcal{H}} + c$.

The lower bound we prove for the Gaussian noise model cannot be immediately turned into a result for an adversarial model without sacrificing the bounded-ness assumption for the adversarial vectors in $\tilde{\mathcal{W}}$. We fix this issue at the end by using a Bernoulli noise model instead. Consider the action sets for the adversary and the player:

$$\tilde{\mathcal{W}} = \{(\tilde{w}_j) \mid |\tilde{w}_j| \leq \mu_j\ \forall j\}$$
$$\tilde{\mathcal{A}} := \{(\tilde{a}_j) \mid |\tilde{a}_j| \leq 1\}.$$

We show that for any integer $n$, and any algorithm $\pi$, there exists $\tilde{w} \in \tilde{\mathcal{W}}$ such that

$$\mathcal{R}_n(\pi, \tilde{w}) \geq \Omega\left(n^{\frac{\beta+1}{2\beta}}\right),$$

where $\mathcal{R}_n(\pi, \tilde{w})$ denotes the regret incurred by the player's algorithm $\pi$ and an adversary's strategy "centered" around $\tilde{w}$.

Let $\tilde{w}, \tilde{w}' \in \tilde{\mathcal{W}}$. Denote by $P_{\tilde{w}}, P_{\tilde{w}'}$ the probability distributions induced by the interaction of a player using $\pi$ and an adversary using either $\tilde{w}$ or $\tilde{w}'$.

$$KL(P_{\tilde{w}}, P_{\tilde{w}'}) = \frac{1}{2}\sum_{t=1}^{n} E[\langle \tilde{a}_t, (\tilde{w} - \tilde{w}')\rangle_{\mathcal{H}}^2],$$

where $KL(p, q)$ is the Kullback-Leibler divergence between distributions $p$ and $q$. Let,

$$p_{\tilde{w},i} = \mathbb{P}_{\tilde{w}}(\tilde{a}_{t,i}\tilde{w}_i < 0 \text{ for at least } n/2 \text{ of indices } i).$$

Then we have,

$$\mathcal{R}_n(\pi, \tilde{w}) \geq \sum_i p_{\tilde{w},i} \frac{n}{2} |\tilde{w}_i| \tag{24}$$

**Averaging argument:** Let $r(\tilde{w}) = \sum_i p_{\tilde{w},i}|\tilde{w}_i|$, and

$$\tilde{\mathcal{W}}_d = \{(\pm\Delta_i)\}, \text{ where } \Delta_i = \min\left(\frac{1}{\sqrt{n}}, \mu_i\right).$$

Pinsker inequality: For $\tilde{w} \in \tilde{\mathcal{W}}_d$, let $\tilde{w}'_i$ be the vector equal to $\tilde{w}$ with the $i$th coordinate flipped, then $\tilde{w}_i - \tilde{w}'_i = \pm\Delta_i e_i$, and $KL(P_{\tilde{w}}, P_{\tilde{w}'}) = \frac{1}{2}\Delta_i^2 \sum_{t=1}^n \tilde{a}_{t,i}^2 \geq \frac{n}{2}\Delta_i^2$. Therefore

$$p_{i,\tilde{w}} + p_{i,\tilde{w}'_i} \geq \frac{1}{2}e^{-\frac{n}{2}\Delta_i^2}.$$

Let $\tilde{\mathcal{W}}_d^{(k)} = \{\tilde{w} \in \tilde{\mathcal{W}}_d \mid \tilde{w}_i = \Delta_i \forall i > k\}$, then:

$$\frac{1}{\tilde{\mathcal{W}}_d^{(k)}} \sum_{\tilde{w}\in\tilde{\mathcal{W}}_d^{(k)}} \Delta_i p_{i,\tilde{w}} = \frac{1}{\tilde{\mathcal{W}}_d^{(k)}} \sum_{\tilde{w}\in\tilde{\mathcal{W}}_d^{(k)}} \sum_i^k \Delta_i p_{i,\tilde{w}'_i} = \frac{1}{\tilde{\mathcal{W}}_d^{(k)}} \sum_{\tilde{w}\in\tilde{\mathcal{W}}_d^{(k)}} \sum_i^k \Delta_i \frac{p_{i,\tilde{w}} + p_{i,\tilde{w}'_i}}{2} \geq \frac{1}{2}\sum_{i=1}^k \Delta_i e^{-\frac{n}{2}\Delta_i^2}.$$

Therefore, for any $k$ there exists $\tilde{w}^{(k)}$ such that

$$\sum_i^k \delta_i p_{\tilde{w}^{(k)},i} \geq \frac{1}{2}\sum_{i=1}^k \Delta_i e^{-\frac{n}{2}\Delta_i^2}.$$

From (24) we conclude that

$$\mathcal{R}_n(\pi, \tilde{w}^{(k)}) \geq \frac{n}{2}\sum_i^k p_{\tilde{w}^{(k)},i}\Delta_i \geq \frac{n}{4}\sum_{i=1}^k \Delta_i e^{-\frac{n}{2}\Delta_i^2}.$$

By taking the supremum on both sides, we get

$$\sup_{\tilde{w}} \mathcal{R}_n(\pi, \tilde{w}) \geq \frac{n}{4}\sum_{i=1}^\infty \Delta_i e^{-\frac{n}{2}\Delta_i^2}.$$

Since

$$\Delta_i = \min\left(\frac{1}{\sqrt{n}}, \mu_i\right),$$

and let $M = \min\{i \text{ s.t. } \mu_i < \frac{1}{\sqrt{n}}\}$. For example if $\mu_i = \frac{1}{i^\beta}$, then $M = n^{\frac{1}{2\beta}}$. Then:

$$\sup_{\tilde{w}} \mathcal{R}_n(\pi, \tilde{w}) \geq \frac{n}{4}M\frac{1}{\sqrt{n}}e^{-\frac{1}{2}} + \frac{n}{4}\sum_{i>M} \mu_i e^{-\frac{n}{2}\mu_i^2} \geq \Omega(\sqrt{n}M)$$

For example if $\mu_i = \frac{1}{i^\beta}$, then $\sup_{\tilde{w}} \mathcal{R}_n(\pi, \tilde{w}) \geq \Omega\left(n^{\frac{1+\frac{1}{\beta}}{2}}\right)$ and therefore:

$$\sup_{\tilde{w}} \mathcal{R}_n(\pi, \tilde{w}) \geq \Omega\left(n^{\frac{\beta+1}{2\beta}}\right).$$

The same algorithm carries forward for an adversarial non obvlivious feedback model. We assume the action set of the player is $\{1\} \times \tilde{\mathcal{A}}$ and the action set of the adversary is $[-1, 1] \times \mathcal{W}$. We assume that $\max_{\tilde{a}\in\mathcal{A},\tilde{w}\in\mathcal{W}} |\langle\tilde{a}, \tilde{w}\rangle| \leq 1$.

1 At the beginning of time the adversary selects a vector $\tilde{w}$ in $\mathcal{W}$.

2 In every round the player selects an action vector $(1, \tilde{a}) \in \{1\} \times \tilde{\mathcal{A}}$.

3 The adversary computes $\langle\tilde{a}, \tilde{w}\rangle$ and samples value $v \in \{-1, 1\}$ from a Rademacher random variable with parameter $\frac{1}{2} + \frac{\langle\tilde{a},\tilde{w}\rangle}{2}$ and chooses as its action the vector $(v - \langle\tilde{a}, \tilde{w}\rangle, \tilde{w})$

4 The player experiences loss $\langle(1,\tilde{a}),(v-\langle\tilde{a},\tilde{w}\rangle,\tilde{w})\rangle = v$.

In this feedback model the same argument holds since the KL term has a quadratic bound as in the Gaussian case:

$$KL(P_{\tilde{w}},P_{\tilde{w}'}) \leq 2\sum_{t=1}^{n} E[\langle \tilde{a}_t,(\tilde{w}-\tilde{w}')\rangle_{\mathcal{H}}^2].$$

For some constant $C$. This is because the loss values are always either $1$ or $-1$. The upper bound follows from a $\chi^2$-squared bound on the KL of two Bernoulli random variables. The rest of the argument remains unchanged. ■

## F. Full Information Regret Bounds

### F.1. Exponential Weights Regret Bound

In this section we prove a regret bound for exponential weights and present a proof of Theorem 15. The analysis of the regret is similar to the analysis of exponential weights for linear losses (see for example a review in 10). In the proof below we denote the filtration at the end of round $t$ by $\mathcal{F}_t$, that is, it conditions on the past actions of the player and the adversary $(a_{t-1},w_{t-1},\dots,a_1,w_1)$.

**Proof** [Proof of Theorem 15] By the tower property and by the definition of the regret we can write the cumulative loss as,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{n}\langle\Phi(a_t),w_t\rangle\right] &= \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{E}_{a_t\sim p_t}\left[\langle\Phi(a_t),w_t\rangle\middle|\mathcal{F}_{t-1}\right]\right]\\ &= \mathbb{E}\left[\sum_{t=1}^{n}\left[\int_{\mathcal{A}} p_t(a)\langle\Phi(a),w_t\rangle da\middle|\mathcal{F}_{t-1}\right]\right].\end{aligned}$$

Observe that our choice of $\eta$ implies that $\eta\langle\Phi(a),w_t\rangle > -1$. By invoking Hoeffding's inequality (stated as Lemma 53) we get

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{n}\left[\int_{\mathcal{A}} p_t(a)\langle\Phi(a),w_t\rangle da\middle|\mathcal{F}_{t-1}\right]\right]\\ &\leq -\frac{1}{\eta}\underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\log\left(\mathbb{E}_{a\sim p_t}\left[\exp\left(-\eta\langle\Phi(a),w_t\rangle\right)|\mathcal{F}_{t-1}\right]\right)\right]}_{=:\Gamma}+(e-2)\eta\mathbb{E}\left[\sum_{t=1}^{n}\int_{\mathcal{A}}\left[p_t(a)\langle\Phi(a),w_t\rangle^2 da|\mathcal{F}_{t-1}\right]\right]\\ &\overset{(i)}{\leq} -\frac{\Gamma}{\eta}+(e-2)\eta\mathcal{G}^4 n,\end{aligned}$$

where $(i)$ follows by Cauchy-Schwartz and the bound on the adversarial and player actions. Next we bound $\Gamma$ using Lemma 44. Substituting this bound into the expression above we get

$$\mathbb{E}\left[\sum_{t=1}^{n}\langle\Phi(a_t),w_t\rangle\right] \leq \mathbb{E}\left[\sum_{t=1}^{n}\langle\Phi(a^*),w_t\rangle\right]+\frac{\log(vol(\mathcal{A}))}{\eta}+(e-2)\eta\mathcal{G}^4\cdot n.$$

Rearranging terms we have the regret is bounded by

$$\mathcal{R}_n \leq (e-2)\mathcal{G}^4\eta n+\frac{\log(vol(\mathcal{A}))}{\eta}.$$

The choice of $\eta = \sqrt{\log(vol(\mathcal{A}))}\Big/\sqrt{(e-2)}\mathcal{G}^2 n^{1/2}$, optimally trades of the two terms to establish a regret bound of $\mathcal{O}(n^{1/2})$. ■

Next we provide a proof of the bound on $\Gamma$ used above.

**Lemma 44** *Assume that $p_1(\cdot)$ is chosen as the uniform distribution in Algorithm 2.3. Also let $\Gamma$ be defined as follows*

$$\Gamma = \mathbb{E}\left[\sum_{t=1}^{n} \log\left(\mathbb{E}_{a\sim p_t}\left[\exp\left(-\eta\langle\Phi(a), w_t\rangle_{\mathcal{H}}\right)\middle|\mathcal{F}_{t-1}\right]\right)\right].$$

*Then we have that,*

$$\Gamma \geq -\eta\mathbb{E}\left[\sum_{i=1}^{n}\langle\Phi(a^*), w_i\rangle_{\mathcal{H}}\right] - \log\left(vol(\mathcal{A})\right),$$

*where $a^*$ is the optimal action in hindsight in the definition of regret and $vol(\mathcal{A})$ is the volume of the set $\mathcal{A}$.*

**Proof** Expanding $\Gamma$ using the definition of $p_t$ we have that,

$$\begin{aligned}
\Gamma &\overset{(i)}{=} \mathbb{E}\left[\sum_{t=1}^{n}\log\left\{\frac{\int_{\mathcal{A}}\exp\left(-\eta\sum_{i=1}^{t}\langle\Phi(a), w_i\rangle_{\mathcal{H}}\right)da}{\int_{\mathcal{A}}\exp\left(-\eta\sum_{i=1}^{t-1}\langle\Phi(a), w_i\rangle_{\mathcal{H}}\right)da}\right\}\right]\\
&\overset{(ii)}{=} \mathbb{E}\left[\log\left(\int_{\mathcal{A}}\exp\left(-\eta\sum_{i=1}^{n}\langle\Phi(a), w_i\rangle_{\mathcal{H}}\right)da\right)\right] - \log\left(vol(\mathcal{A})\right),
\end{aligned}$$

where $(i)$ follows by the definition of $p_t(a)$ and $(ii)$ is by expanding the sum and canceling the terms in a telescoping series. The $\log(vol(\mathcal{A}))$ term is because we start off with a uniform distribution over all elements. Lastly observe that by optimality of $a^*$ we have that,

$$\mathbb{E}\left[\log\left(\int_{\mathcal{A}}\exp\left(-\eta\sum_{i=1}^{n}\langle\Phi(a), w_i\rangle\right)da\right)\right] \geq -\eta\mathbb{E}\left[\sum_{i=1}^{n}\langle\Phi(a^*), w_i\rangle\right].$$

Plugging this into the above expression establishes the desired bound on $\Gamma$. ■

We now present a proof of Lemma 17 that guarantees that it is possible to sample efficiently from the exponential weights distribution when the losses are quadratics.

**Proof** [Proof of Lemma 17] Let $v_1, \cdots, v_d$ an orthonormal basis of eigenvectors of $B$ with eigenvalues $\lambda_1, \cdots, \lambda_d$ possibly negative. We express $b$ using the basis $\{v_i\}_{i=1}^d$ as $b = \sum_{i=1}^d \gamma_i v_i$. Also let $a = \sum_{i=1}^d \alpha_i v_i$. By the definition of the set $\mathcal{A}$ we have $\sum_{i=1}^d \alpha_i^2 \leq 1$. The distribution $q(\cdot)$ can be thus expressed as

$$q(a) \propto \exp\left(\sum_{i=1}^{d}(\lambda_i\alpha_i^2 + \gamma_i\alpha_i)\right).$$

Completing the squares (whenever $\lambda_i \neq 0$),

$$q(a) \propto \exp\left\{\sum_{i=1}^{d}\lambda_i\left(\alpha_i^2 + \frac{\gamma_i\alpha_i}{\lambda_i} + \left(\frac{\gamma_i}{2\lambda_i}\right)^2\right)\right\}.$$

Let us re-parametrize this distribution by setting $\beta_i = (\alpha_i + \frac{\gamma_i}{2\lambda_i})^2$. The inverse mapping is $\alpha_i = \sqrt{\beta_i} - \frac{\gamma_i}{2\lambda_i}$. To sample from $q(\cdot)$ it is enough to produce a sample from a surrogate distribution $\beta \sim t(\beta)$ and turn them into a sample of $q$ where,

$$\begin{aligned}
t(\beta) &\propto \exp\left(\sum_{i=1}^{d}\lambda_i\beta_i\right),\\
s.t. \quad & 0 \leq \beta_i,\\
\sum_{i=1}^{d}&\left(\sqrt{\beta_i} - \frac{\gamma_i}{2\lambda_i}\right)^2 \leq 1.
\end{aligned}$$

Let $\{\epsilon_i\}_{i=1}^d$ be independent Bernoulli $\{-1, 1\}$ variables, then $a = \sum_{i=1}^d \epsilon_i(\sqrt{\beta_i} - \frac{\gamma_i}{2\lambda_i})v_i$ is a sample from $q$. Note that the distribution $t(\beta)$ is log-concave. We now show that the constraint set $\mathcal{C}$ is convex, where $\mathcal{C} = \{\beta | \beta_i \geq 0, \sum_{i=1}^d (\sqrt{\beta_i} - \frac{\gamma_i}{2\lambda_i})^2 \leq 1\}$.

Let $\hat{\beta}$ and $\tilde{\beta}$ be two distinct points in $\mathcal{C}$. We show that for any $\eta \in [0, 1]$ the point $\eta\hat{\beta} + (1-\eta)\tilde{\beta} \in \mathcal{C}$. The non-negativity constraint is clearly satisfied $(\eta\hat{\beta} + (1-\eta)\tilde{\beta})_i \geq 0, \forall i$. The second constraint can be rewritten as

$$\sum_{i=1}^d \hat{\beta}_i - \frac{\gamma_i\sqrt{\hat{\beta}_i}}{\lambda_i} + \left(\frac{\gamma_i}{2\lambda_i}\right)^2 \leq 1 \tag{25}$$

$$\sum_{i=1}^d \tilde{\beta}_i - \frac{\gamma_i\sqrt{\tilde{\beta}_i}}{\lambda_i} + \left(\frac{\gamma_i}{2\lambda_i}\right)^2 \leq 1. \tag{26}$$

These equations imply that,

$$\eta\left[\sum_{i=1}^d \hat{\beta}_i - \frac{\gamma_i\sqrt{\hat{\beta}_i}}{\lambda_i} + \left(\frac{\gamma_i}{2\lambda_i}\right)^2\right] + (1-\eta)\left[\sum_{i=1}^d \tilde{\beta}_i - \frac{\gamma_i\sqrt{\tilde{\beta}_i}}{\lambda_i} + \left(\frac{\gamma_i}{2\lambda_i}\right)^2\right] \leq 1.$$

By concavity of the square root function we have

$$\sum_{i=1}^d \eta\frac{\gamma_i\sqrt{\hat{\beta}_i}}{\lambda_i} + (1-\eta)\frac{\gamma_i\sqrt{\tilde{\beta}_i}}{\lambda_i} \leq \sum_{i=1}^d \frac{\gamma_i\sqrt{\eta\hat{\beta}_i + (1-\eta)\tilde{\beta}_i}}{\lambda_i},$$

these two observations readily imply that $\eta\hat{\beta} + (1-\eta)\tilde{\beta}$ satisfies the constraint of $\mathcal{C}$ thus implying convexity of $\mathcal{C}$. We can thus use Hit-and-Run (34) to sample from $t(\beta)$ in $\tilde{O}(d^4)$ steps and convert to samples from $q(\cdot)$ using the method described above. In case some eigenvalues are zero, say without loss of generality $\lambda_1, \cdots, \lambda_R$. Then set $\beta_i = \alpha_i^2$ for $i \in \{R+1, \ldots, d\}$ and sample from the distribution,

$$\begin{aligned} t(\beta) \propto \exp&\left(\sum_{i=1}^R \gamma_i\alpha_i + \sum_{i=R+1}^d \lambda_i\beta_i\right), \\ s.t. \qquad 0 &\leq \beta_i, \\ \sum_{i=1}^R \alpha_i^2 + \sum_{i=R+1}^d &\left(\sqrt{\beta_i} - \frac{\gamma_i}{2\lambda_i}\right)^2 \leq 1. \end{aligned}$$

The analysis follows as before for this case as well. ■

### F.2. Conditional Gradient Method Analysis

The regret bound analysis for Algorithm 4, conditional gradient method over RKHSs follows by similar arguments to the analysis of the standard online conditional gradient descent (see for example review in 25). To prove this we first prove the regret bound of a different algorithm – follow the regularized leader.

### F.3. Follow the Regularized Leader

We present a version of follow the regularized leader (43) (FTRL, Algorithm 5) adapted to our setup. Note that this algorithm is not tractable in general as at each step we are required to perform an optimization problem over the convex hull of $\Phi(\mathcal{A})$. However, we provide a regret bound that we will use in our regret bound analysis for the conditional gradient method. Let us define $w_0 = X_1/\eta$. We first establish the following lemma.

**Lemma 45 (No regret strategy)** *For any $u \in \mathcal{A}$*

$$\sum_{t=0}^n \langle X_t, \Phi(u)\rangle_{\mathcal{H}} \geq \sum_{t=0}^n \langle X_t, X_{t+1}\rangle_{\mathcal{H}}.$$

**Algorithm 5** Follow the Regularized leader (FTRL)

**Input** : Set $\mathcal{A}$, number of rounds $n$, initial action $a_1 \in \mathcal{A}$, inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, learning rate $\eta > 0$.
Let $X_1 = \text{argmin}_{X \in conv(\Phi(\mathcal{A}))} \frac{1}{\eta} \langle X, X \rangle$
choose $\mathcal{D}_1$ such that $\mathbb{E}_{x \sim \mathcal{D}_1}[\Phi(x)] = X_1$
**for** $t = 1, 2, 3 \ldots, n$ **do**
  choose $a_t \sim \mathcal{D}_t$
  observe $\langle \Phi(a_t), w_t \rangle_{\mathcal{H}}$
  update $X_{t+1} = \text{argmin}_{X \in conv(\Phi(\mathcal{A}))} \eta \sum_{s=1}^{t} \langle w_s, X \rangle_{\mathcal{H}} + \langle X, X \rangle_{\mathcal{H}}$
  choose $\mathcal{D}_{t+1}$ s.t. $\mathbb{E}_{x \sim \mathcal{D}_{t+1}}[\Phi(x)] = X_{t+1}$
**end**

This is the crucial lemma needed to prove regret bounds for FTRL algorithms and its proof follows from standard arguments (see for example Lemma 5.3 25).

**Definition 46** *Define a function* $g_t(\cdot) : \mathbb{R}^D \mapsto R$ *as,*

$$g_t(X) \triangleq \left[ \eta \sum_{s=1}^{t} \langle w_s, X \rangle_{\mathcal{H}} + \langle X, X \rangle_{\mathcal{H}} \right].$$

**Definition 47** *Define the Bregman divergence as,*

$$B_R(x||y) \triangleq R(x) - R(y) - \langle \nabla R(y), (x - y) \rangle_{\mathcal{H}}.$$

Given these two definitions we now establish a lemma that will be used used to control the regret of FTRL.

**Lemma 48** *For any* $t \in \{1, 2, \ldots, n\}$ *we have the upper bound,*

$$\langle w_t, X_t - X_{t+1} \rangle_{\mathcal{H}} \leq 2\eta \|w_t\|_{\mathcal{H}}^2.$$

**Proof** By the definition of Bregman divergence we have,

$$\begin{aligned} g_t(X_t) &= g_t(X_{t+1}) + \langle X_t - X_{t+1}, \nabla g_t(X_{t+1}) \rangle_{\mathcal{H}} + B_{g_t}(X_t || X_{t+1}) \\ &\geq g_t(X_{t+1}) + B_{g_t}(X_t || X_{t+1}), \end{aligned}$$

where the inequality is because $X_{t+1}$ is the minimizer of $g_t(\cdot)$ over $conv(\Phi(\mathcal{X}))$. After rearranging terms we are left with an upper bound on the Bregman divergence,

$$\begin{aligned} B_{g_t}(X_t || X_{t+1}) &\leq g_t(X_t) - g_t(X_{t+1}) \\ &= \big( g_{t-1}(X_t) - g_{t-1}(X_{t+1}) \big) + \eta \langle w_t, X_t - X_{t+1} \rangle_{\mathcal{H}} \\ &\leq \eta \langle w_t, X_t - X_{t+1} \rangle_{\mathcal{H}}, \end{aligned} \tag{27}$$

where the last inequality follows because $X_{t-1}$ is the minimizer of the function $g_{t-1}(\cdot)$ over $conv(\Phi(\mathcal{X}))$. Observe that $B_{g_t}(X_t || X_{t+1}) = \frac{1}{2} \|X_t - X_{t+1}\|_{\mathcal{H}}^2$. Thus by the Cauchy-Schwartz inequality we have,

$$\begin{aligned} \langle w_t, X_t - X_{t+1} ) \rangle_{\mathcal{H}} &\leq \|w_t(X_t)\|_{\mathcal{H}} \|X_t - X_{t+1}\|_{\mathcal{H}} \\ &= \|w_t\|_{\mathcal{H}} \sqrt{2 B_{g_t}(X_t || X_{t+1})}. \end{aligned}$$

Substituting the upper bound from Equation (27) we get,

$$\langle w_t, X_t - X_{t+1} ) \rangle_{\mathcal{H}} \leq \|w_t\|_{\mathcal{H}} \cdot \sqrt{2\eta \langle w_t, X_t - X_{t+1} \rangle_{\mathcal{H}}}.$$

Rearranging terms establishes the result. ∎

**Theorem 49** *Given a step size $\eta > 0$, the regret suffered by Algorithm 5 after $n$ rounds is bounded by*

$$\mathcal{R}_n \leq 2n\eta\mathcal{G}^2 + \frac{2\mathcal{G}^2}{\eta}.$$

**Proof** By the definition of regret we have

$$\begin{aligned}
\mathcal{R}_n &= \mathbb{E}\left[\sum_{t=1}^{n}\langle \Phi(a_t), w_t\rangle_{\mathcal{H}} - \min_{a\in\mathcal{A}}\left[\sum_{t=1}^{n}\langle \Phi(a_t), w_t\rangle_{\mathcal{H}}\right]\right] \\
&\overset{(i)}{=} \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{E}_{a_t\sim\mathcal{D}_t}\left[\langle w_t, \Phi(a_t) - \Phi(a^*)\rangle_{\mathcal{H}}\middle|\mathcal{F}_{t-1}\right]\right] \overset{(ii)}{=} \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t - \Phi(a^*)\rangle_{\mathcal{H}}\right] \\
&\overset{(iii)}{\leq} \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t - X_{t+1}\rangle_{\mathcal{H}}\right] + \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_{t+1} - \Phi(a^*)\rangle_{\mathcal{H}}\right] + \frac{1}{\eta}(\langle X_1, X_1\rangle_{\mathcal{H}} - (\langle X_0, X_0\rangle_{\mathcal{H}}) \\
&\overset{(iv)}{\leq} \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t - X_{t+1}\rangle_{\mathcal{H}}\right] + \frac{2\mathcal{G}^2}{\eta}.
\end{aligned} \tag{28}$$

The first equality follows as $a^*$ is the minimizer, $(ii)$ is by evaluating the expectation with respect to $\mathcal{D}_t$, $(iii)$ is an algebraic manipulation and finally $(iv)$ follows by invoking Lemma 45 and using Cauchy-Schwartz to bound the last term. We need to control the first term in Equation (28) to get a regret bound. To control the first term we now invoke Lemma 48

$$\mathcal{R}_n \leq 2\eta\sum_{t=1}^{n}\|w_t\|_{\mathcal{H}}^2 + \frac{2\mathcal{G}^2}{\eta} \leq 2n\eta\mathcal{G}^2 + \frac{2\mathcal{G}^2}{\eta}.$$

This establishes the stated result. ■

### F.4. Regret Bound for Algorithm 4

In deploying Algorithm 4 we will at each round find distributions over the action space $\mathcal{A}$ as the player is only allowed play rank 1 actions in the Hilbert space at each round, while the action prescribed by the conditional gradient method might not be rank 1. Thus we find a distribution $\mathcal{D}_t$ such that,

$$\mathbb{E}_{a\sim\mathcal{D}_t}\Phi(a) = X_t,$$

where $X_t$ is the action prescribed by Algorithm 4. We will strive to match the optimal action in expectation by choosing an appropriate distribution and get bounds on expected regret. For all $t \in \{1, 2, \ldots, n\}$ let $X_t^*$ be defined as the iterates of the follow the regularized leader (Algorithm 5) with the regularization set to $R(X) = \|X - X_1\|_{\mathcal{H}}^2$ and applied to the shifted loss function, $\langle w_t, X - (X_t^* - X_t)\rangle_{\mathcal{H}}$. Notice that,

$$|\langle X, w_t\rangle_{\mathcal{H}} - \langle X - (X_t^* - X_t), w_t\rangle_{\mathcal{H}}| \leq \|w_t\|_{\mathcal{H}}\|X_t^* - X_t\|_{\mathcal{H}} \leq \mathcal{G}\|X_t^* - X_t\|_{\mathcal{H}}. \tag{29}$$

We are now ready to prove Theorem 16.

**Proof** [Proof of Theorem 16] We denote the filtration up to round $t$ by $\mathcal{F}_{t-1}$, that is, we condition on all past player and adversary actions. Also let us denote the optimal action in hindsight by $a^*$. We begin by expanding the definition of regret to get,

$$\begin{aligned}
\mathcal{R}_n &= \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{E}_{a_t\sim\mathcal{D}_t}\left[\langle w_t, \Phi(a_t)\rangle_{\mathcal{H}} - \langle w_t, \Phi(a^*)\rangle_{\mathcal{H}}\middle|\mathcal{F}_{t-1}\right]\right] \\
&\overset{(i)}{=} \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t - \Phi(a^*)\rangle_{\mathcal{H}}\right] = \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t - X_t^*\rangle_{\mathcal{H}}\right] + \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t^* - \Phi(a^*)\rangle_{\mathcal{H}}\right] \\
&\overset{(ii)}{\leq} \mathbb{E}\left[\sum_{t=1}^{n}\langle w_t, X_t - X_t^*\rangle_{\mathcal{H}}\right] + 2n\eta\mathcal{G}^2 + \frac{2\mathcal{G}^2}{\eta} \overset{(iii)}{\leq} \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\|w_t\|_{\mathcal{H}}\|X_t - X_t^*\|_{\mathcal{H}}\right]}_{=:\Xi} + 2n\eta\mathcal{G}^2 + \frac{2\mathcal{G}^2}{\eta},
\end{aligned}$$

where $(i)$ follows by taking expectation with respect to $\mathcal{D}_t$, $(ii)$ follows by invoking Theorem 49 and $(iii)$ by Cauchy-Schwartz inequality. We finally need to control $\Xi$ to establish a bound on the regret.

$$\Xi = \mathbb{E}\left[\sum_{t=1}^{n}\|w_t\|_{\mathcal{H}}\|X_t - X_t^*\|_{\mathcal{H}}\right] \overset{(i)}{\leq} \mathbb{E}\left[\sum_{t=1}^{n}\|w_t\|_{\mathcal{H}}\sqrt{F_t(X_t) - F_t(X_t^*)}\right] \overset{(ii)}{\leq} 2\sum_{t=1}^{n}\mathcal{G}^2\sqrt{\gamma_t},$$

here $(i)$ follows by the strong convexity of $F_t(\cdot)$ and $(ii)$ follows by the upper bound established in Lemma 50. Plugging this into the bound for regret we have

$$\mathcal{R}_n \leq 2\sum_{t=1}^{n}\mathcal{G}^2\sqrt{\gamma_t} + 2n\eta\mathcal{G}^2 + \frac{2\mathcal{G}^2}{\eta} \overset{(i)}{\leq} 4\mathcal{G}^2 n^{3/4} + 2n\eta\mathcal{G}^2 + \frac{2\mathcal{G}^2}{\eta},$$

where $(i)$ follows by summing the series $1/t^{1/4}$ ($\sqrt{\gamma_t}$). The choice $\eta = 1/n^{3/4}$ satisfies the conditions of Lemma 50 and we can plug in this choice to get,

$$\mathcal{R}_n \leq 4\mathcal{G}^2 n^{3/4} + 2\mathcal{G}^2 n^{1/4} + 2\mathcal{G}^2 n^{3/4} \leq 8\mathcal{G}^2 n^{3/4}.$$

This establishes the desired bound on the regret. ■

Finally we prove Lemma 50 used to establish the regret bound above. We introduce a new function,

$$h_t(X) \triangleq F_t(X) - F_t(X_t^*).$$

Also the shorthand that $h_t = h_t(X_t)$. These functions are defined conditioned on the filtration $\mathcal{F}_{t-1}$ and $f_t$.

**Lemma 50** *If the parameters $\eta$ and $\gamma_t$ are chosen as stated in Theorem 16, such that $\eta\mathcal{G}\sqrt{h_{t+1}} \leq \mathcal{G}^2\gamma_t^2$, the iterates $X_t$ satisfy, $h_t \leq 4\mathcal{G}^2\gamma_t$.*

**Proof** The functions $F_t$ is 1-smooth therefore we have,

$$\begin{aligned} h_t(X_{t+1}) &= F_t(X_{t+1}) - F_t(X_t^*) = F_t(X_t + \gamma_t(\Phi(v_t) - X_t)) - F_t(X_t^*) \\ &\overset{(i)}{\leq} F_t(X_t) - F_t(X_t^*) + \gamma_t\langle\Phi(v_t) - X_t, \nabla F_t(X_t)\rangle_{\mathcal{H}} + \frac{\gamma_t^2}{2}\|\Phi(v_t) - X_t\|_{\mathcal{H}}^2 \\ &\overset{(ii)}{\leq} (1-\gamma_t)\Big(F_t(X_t) - F_t(X_t^*)\Big) + \gamma_t^2\mathcal{G}^2, \end{aligned}$$

where $(i)$ follows by the strong convexity of $F_t$ and $(ii)$ follows as $\Phi(v_t)$ is the minimizer of $F_t(\cdot)$. By the definition of $F_{t+1}(\cdot)$ and $h_t$ we also have,

$$\begin{aligned} h_{t+1}(X_{t+1}) &= F_t(X_{t+1}) - F_t(X_{t+1}^*) + \eta\langle w_{t+1}, X_{t+1} - X_{t+1}^*\rangle_{\mathcal{H}} \\ &\overset{(i)}{\leq} F_t(X_{t+1}) - F_t(X_t^*) + \eta\langle w_{t+1}, X_{t+1} - X_{t+1}^*\rangle_{\mathcal{H}} \\ &\overset{(ii)}{\leq} h_t(X_{t+1}) + \eta\mathcal{G}\|X_{t+1} - X_{t+1}^*\|_{\mathcal{H}}, \end{aligned} \tag{30}$$

where $(i)$ follows as $X_t^*$ is the minimizer of $F_t$ and $(ii)$ is by Cauchy-Schwartz inequality. Again by leveraging the strong convexity of $F_t$ we have, $\|X - X_{t+1}^*\|_{\mathcal{H}}^2 \leq F_{t+1}(X) - F_{t+1}(X_{t+1}^*) = h_{t+1}$ which leads to the string of inequalities,

$$h_{t+1}(X_{t+1}) \leq h_t(X_{t+1}) + \eta\mathcal{G}\|X_{t+1} - X_{t+1}^*\|_{\mathcal{H}} \leq h_t(X_{t+1}) + \eta\mathcal{G}\sqrt{h_{t+1}(X_{t+1})}.$$

Plugging in the bound on $h_t(X_{t+1})$ from Equation (30) into the above inequality gives us the recursive relation,

$$h_{t+1} \leq h_t(1-\gamma_t) + \gamma_t^2\mathcal{G}^2 + \eta\mathcal{G}\sqrt{h_{t+1}} \overset{(i)}{\leq} h_t(1-\gamma_t) + 2\gamma_t^2\mathcal{G}^2,$$

where, the last step follows by our choice of the schedule for the mixing rate $\gamma_t$ such that $\eta\mathcal{G}\sqrt{h_{t+1}} \leq \mathcal{G}^2\gamma_t^2$. We now complete the proof by an induction over $t$.

For the base case $t = 1$, we have $h_1 = F_1(X_1) - F_1(X_1^*) = \|X_1 - X_1^*\|^2 \leq 4\gamma_1 \mathcal{G}^2$. Thus, by the induction hypothesis for the step $t+1$ we have,

$$h_{t+1} \leq h_t(1-\gamma_t) + 2\gamma_t^2\mathcal{G}^2 \overset{(i)}{\leq} 4\mathcal{G}^2(\gamma_t(1-\gamma_t)) + 2\gamma_t^2\mathcal{G}^2 = 4\mathcal{G}^2\gamma_t\left(1 - \frac{\gamma_t}{2}\right) \overset{(ii)}{\leq} 4\mathcal{G}^2\gamma_{t+1},$$

where $(i)$ follows by the upper bound on $h_t$, $(ii)$ is by the definition $\gamma_t = \min\left(1, \frac{2}{t^{1/2}}\right)$. ■

## G. Application: Posynomial Losses

In this section we will define a *posynomial game*, by introducing posynomial losses and prove that these losses can also be viewed as kernel inner products. We will use the connection between optimizing posynomials and *Geometric Programs* to prove that conditional gradient descent can be run efficiently on this family of losses.

**Definition 51 (Monomial)** *A function $f : \mathbb{R}_+^d \mapsto \mathbb{R}$ defined as*

$$f(x) = cx_1^{\alpha_1}x_2^{\alpha_2}\cdots x_d^{\alpha_d},$$

*where $c > 0$ and $\alpha_i \in \mathbb{R}$, is called a monomial function.*

A non-negative linear combination of monomials is a posynomial.

**Definition 52 (Posynomial)** *A function $f : \mathbb{R}_+^d \mapsto \mathbb{R}$ defined as*

$$f(x) = \sum_{k=1}^{m} c_k x_1^{\alpha_{1k}} x_2^{\alpha_{2k}} \cdots x_d^{\alpha_{dk}},$$

*where $c_k > 0$ and $\alpha_{ik} \in \mathbb{R}$, is called a posynomial function.*

Note that posynomial functions are closed under addition, multiplication and non-negative scaling. Assume the adversary at each round plays a vector of dimension $m$ with all non-negative entries, $w_t = (c_1, c_2, \cdots, c_m)$, while the player chooses a vector $x \in \mathbb{R}_+^d$. This vector is then partitioned into $m$ parts,

$$x = (\underbrace{x_1, x_2}_{s_1}, \ldots, \underbrace{x_{d-2}, x_{d-1}, x_d}_{s_m}),$$

and the feature vector is defined as

$$\Phi(x) = \begin{bmatrix} x_1^{\alpha_1}x_2^{\alpha_2} \\ \vdots \\ x_{d-2}^{\alpha_{d-2}}x_{d-1}^{\alpha_{d-1}}x_d^{\alpha_d} \end{bmatrix}.$$

Where the $i^{th}$ component of $\Phi(\cdot)$ is only a function of the $i^{th}$ partition of the coordinates $s_i$. Then the loss obtained on the evaluation of the inner product between the adversary and player action is a posynomial loss function,

$$\langle w_t, \Phi(x) \rangle_{\mathcal{H}} = \sum_{k=1}^{m} c_k x_1^{\alpha_{k1}} \cdots x_d^{\alpha_{kd}}.$$

A number of scenarios can be modeled as a minimization/maximization problem over posynomial functions (see 11, for a detailed list of examples). We now show that conditional gradient descent can be run efficiently over posynomial losses. If we again assume that the set of actions $\mathcal{A} = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}$. Additionally we choose the initial action to be the solution to the optimization problem,

$$a_1 = \operatorname*{argmin}_{a \in \mathcal{A}} \sum_{k=1}^{d} \Phi(a)_i.$$

The objective function is a posynomial subject to a posynomial inequality constraint. This is a geometric program that can be solved efficiently by changing variables and converting into a convex program (Section 2.5 in 11). At each round of the conditional gradient descent algorithm requires us to solve the optimization problem,

$$v_t = \underset{a\in\mathcal{A}}{\operatorname{argmin}}\langle \eta \sum_{s=1}^{t-1} w_t + 2(X_t - \Phi(a_1)), \Phi(a)\rangle_{\mathcal{H}}. \tag{31}$$

Given that posynomials are closed under addition, and given our choice of $a_1$, the objective function in Equation (31) is still a posynomial and the constraint is a posynomial inequality. This can again be cast as a geometric program that can be solved efficiently at each round.

## H. Technical Results

We present a version of Hoeffding's inequality (29) that is used in the regret bound analysis of exponential weights.

**Lemma 53 (Hoeffding's Inequality)** *Let $\lambda > 0$ and $X$ be a bounded random variable such that $\lambda X \geq -1$, then,*

$$\log\left(\mathbb{E}\left[e^{-\lambda X}\right]\right) \leq (e-2)\lambda^2 \mathbb{E}\left[X^2\right] - \lambda\mathbb{E}\left[X\right],$$

*and hence*

$$\mathbb{E}\left[X\right] \leq -\frac{1}{\lambda}\log\left(\mathbb{E}\left[e^{-\lambda X}\right]\right) + (e-2)\lambda^2\mathbb{E}\left[X^2\right]. \tag{32}$$

**Proof** We look at the log of the moment generating function to get,

$$\log\left(\mathbb{E}\left[\exp(-\lambda X)\right]\right) \overset{(i)}{\leq} \mathbb{E}\left[\exp(-\lambda X)\right] - 1 \overset{(ii)}{\leq} -\lambda\mathbb{E}\left[X\right] + (e-2)\lambda\mathbb{E}\left[X^2\right],$$

where $(i)$ follows by the inequality $\log(y) \leq y - 1$ for all $y > 0$ and $(ii)$ is by the bound $e^{-x} \leq 1 - x + (e-2)x^2$ for $x \geq -1$. ■

### H.1. John's Theorem

We present John's theorem (see 9) that we use to construct an exploration distribution.

**Theorem 54 (John's Theorem)** *Let $\mathcal{K} \subset \mathbb{R}^d$ be a convex set, denote the ellipsoid of minimal volume containing it as,*

$$\mathcal{E} := \left\{ x \in \mathbb{R}^d \middle| (x-c)^\top H(x-c) \leq 1 \right\}.$$

*Then there is a set $\{u_1, \ldots, u_q\} \subset \mathcal{E} \cap \mathcal{K}$ with $q \leq d(d+1)/2 + 1$ contact points and a distribution $p$ (John's distribution) on this set such that any $x \in \mathbb{R}^d$ can be written as*

$$x = c + d\sum_{i=1}^{q} p_i \langle x - c, u_i - c\rangle_J (u_i - c),$$

*where $\langle \cdot, \cdot \rangle_J$ is the inner product for which the minimal ellipsoid is the unit ball about its center $c$ : $\langle x, y\rangle_J = x^\top H y$ for all $x, y \in \mathbb{R}^d$.*

This shows that

$$\begin{aligned} x - c &= d\sum_i p_i (u_i - c)(u_i - c)^\top H(x-c) \\ \iff \tilde{x} &= d\sum_i p_i \tilde{u}_i \tilde{u}_i^\top \tilde{x} \\ \iff \frac{1}{d} I_{d\times d} &= \sum_i p_i \tilde{u}_i \tilde{u}_i^\top \end{aligned}$$

where $\tilde{u}_i = H^{1/2}(u_i - c)$, and similarly for $\tilde{x}$. We see that for any $a, b \in \mathcal{K}$,

$$\tilde{a}^\top \mathbb{E}_{u \sim p}\left[uu^\top\right]\tilde{b} = \frac{1}{d}\tilde{a}^\top \tilde{b}. \tag{33}$$

To use this theorem, we need to perform a preprocessing of the action set $\mathcal{A}$ following a similar procedure described in Section 3 by (13):

- First we map $\mathcal{A}$ onto the RKHS generated by the kernel $\hat{\mathcal{K}}_m$ to produce $\Phi_m(\mathcal{A})$.
- We assume that $\Phi_m(\mathcal{A})$ is full rank in $\mathbb{R}^m$. If not, we can redefine the feature map $\Phi_m$ as the projection onto a lower dimensional subspace.
- Find John's ellipsoid for $Conv(\Phi_m(\mathcal{A}))$ which we denote by $\mathcal{E} = \{x \in \mathbb{R}^m : (x - x_0)^\top H^{-1}(x - x_0) \leq 1\}$.
- Translate $\Phi_m(\mathcal{A})$ by $x_0$. In other words, assume that $\Phi_m(\mathcal{A})$ is centered around $x_0 = 0$ and define the inner product $\langle x, y\rangle_J = x^\top H y$.
- We now play on the set $\Phi_m^J(\mathcal{A}) := H^{-1}\Phi_m(\mathcal{A})$ in $\mathbb{R}^m$. Let the loss of playing an action $H^{-1}\Phi_m(a) \in \Phi_m^J(\mathcal{A})$ when the adversary plays $z$ be $\langle H^{-1}\Phi_m(a), z\rangle_J = \Phi_m(a)^\top z$.
- The contact points, $u_1, \ldots, u_q$ are in $\Phi_m^J(\mathcal{A})$ and are valid points to play. We now use $p$ – John's distribution – to be the exploration distribution.

Mimicking (13) it can be shown that Algorithm 2.3 works with a generic dot product and that all the steps in the regret bound in Appendix A go through.

## I. Experiments

We perform an empirical study of our algorithms in both the full information and the bandit settings and demonstrate their practicality. In the full information setting we conducted experiments with quadratic losses using exponential weights. We also plot the performance of exponential weights algorithm on Gaussian losses. In the bandit feedback setting we again study quadratic and Gaussian losses.

FULL INFORMATION

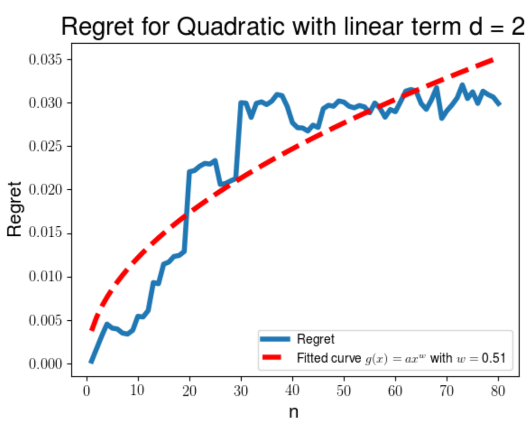


*Figure 1.* Quadratic with linear term Full Information.

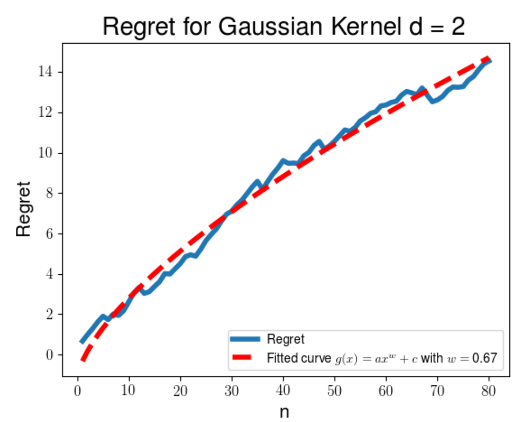


*Figure 2.* Gaussian Losses Full Information.

Exponential weights requires us to sample from a distribution of the form $p(x) \propto \exp(\mu \sum_{i=1}^{t} \mathcal{K}(x, w_i))$. In general sampling from these distributions is possibly intractable, however they present good empirical performance. The following plot shows a diffusion MCMC algorithm sampling from a distribution proportional to $\exp\left(-\eta \sum_{i=1}^{t} \mathcal{K}(x, z_i)\right)$ where $\mathcal{K}$ is the Gaussian kernel, $\eta = 10$, and $x$ is restricted to an $\ell_2$ ball of radius $10$. In practice using exponential weights in the full information setting and sampling using a diffusion MCMC algorithm yields sublinear regret profiles and tractable sampling even for Gaussian losses. We ran experiments generating random loss sequences and we plot the average regret over 60 runs of the algorithm.

BANDITS EXPERIMENTS

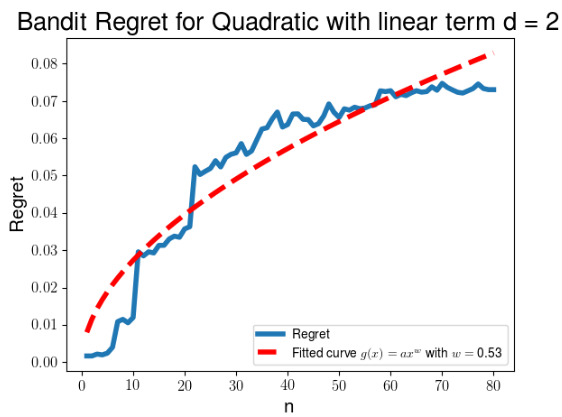


*Figure 3.* Quadratics with linear term Bandit Feedback.

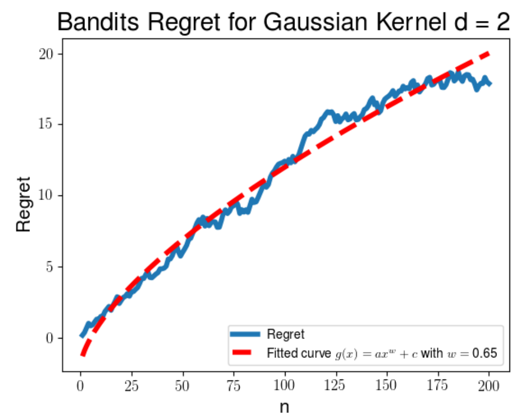


*Figure 4.* Gaussian Losses Bandit Feedback.

The kernel exponential weights algorithm presents also a sub-linear regret profile. The Gaussian experiments involved the construction of the finite dimensional kernel $\mathcal{K}_m$ by kernel PCA.