Online learning with kernel losses

Aldo Pacchiano∗ PACCHIANO@BERKELEY.EDU
UC BERKELEY

Niladri S. Chatterji∗ CHATTERJI@BERKELEY.EDU
UC BERKELEY

Peter L. Bartlett PETER@BERKELEY.EDU
UC BERKELEY

Editor:

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 ($\mu_j \leq O(j^{-\beta})$), we find that the regret is bounded by $R_n \leq O(n^{3/2 - 1})$. While under the assumption of exponential eigen-decay ($\mu_j \leq O(e^{-\beta j})$) we get an even tighter bound on the regret $R_n \leq O(n^{1/2})$. 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 (Cesa-Bianchi and Lugosi, 2006; Hazan, 2016), a player interacts with an unknown and arbitrary 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

∗. A. Pacchiano and N. S. Chatterji contributed equally to this work.

© Aldo Pacchiano, Niladri S. Chatterji and Peter L. Bartlett.
License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/ Attribution requirements are provided at http://jmlr.org/papers/v/ .html
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 $a \in A \subset \mathbb{R}^d$. This action is mapped to an element in a reproducing kernel Hilbert space (RKHS) generated by a mapping $K(\cdot, \cdot)$. The function $K(\cdot, \cdot)$ 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 $K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}$, 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: $\langle a, w \rangle_{\mathcal{H}} = a^T w$. 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: $\langle \phi(a), \left(\begin{array}{c} W \\ b \end{array}\right) \rangle_{\mathcal{H}} = a^T Wa + b^T a$, 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 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: $\langle \Phi(a), \Phi(y) \rangle_{\mathcal{H}} = (1 + x^T y)^2$ 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 (Kleinberg, 2005; Flaxman et al., 2005). The problem most closely related to our work is the case when the losses are linear introduced earlier by (McMahan and Blum, 2004; Awerbuch and Kleinberg, 2004). In this setting (Dani et al., 2008; Cesa-Bianchi and Lugosi, 2012; Bubeck et al., 2012) proposed the EXP 2 (Expanded Exp) algorithm under different choices of exploration distributions. (Dani et al., 2008) worked with the uniform distribution over the barycentric spanner of the set, in (Cesa-Bianchi and Lugosi, 2012) this distribution was the uniform distribution over the set and in (Bubeck et al., 2012) they use the exploration distribution given by John’s theorem that leads to a regret bound of $O\left((dn \log(N))^{1/2}\right)$, 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 $A$ is convex and compact, (Abernethy et al.) analyzed Mirror descent to get a regret bound of $O\left(d\sqrt{n\log(n)}\right)$ for some $\theta > 0$. For the case with general convex losses with bandit feedback recently (Bubeck et al., 2016) proposed
a poly-time algorithm that has a regret guarantee of $\tilde{O}(d^{0.5} \sqrt{n})$, which is optimal in its dependence on the number of rounds $n$. Previous work on this problem includes, (Agarwal and Dekel, 2010; Saha and Tewari, 2011; Hazan and Levy, 2014; Dekel et al., 2015; Bubeck et al., 2015; Hazan and Li, 2016) in the adversarial setting under different assumptions on the structure of the convex losses and by (Agarwal et al., 2011) who studied this problem in the stochastic setting\footnote{For an extended bibliography of the work on online convex bandits see (Bubeck et al., 2016).}. (Valko et al., 2013) study stochastic kernelized contextual bandits with a modified UCB algorithm to obtain a regret bound similar to ours, $R_n \leq \tilde{d} \sqrt{n}$ 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 (Srinivas et al., 2009). Online learning under bandit feedback has also been studied when the losses are non-parametric, for example when the losses are Lipschitz (Cesa-Bianchi et al., 2017; Rakhlin and Sridharan, 2015).

In the full information case, the online optimization framework with convex losses was first introduced by (Zinkevich, 2003). 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 (Jaggi, 2011) and then improved subsequently by (Hazan and Kale, 2012). 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 (Allen-Zhu and Li, 2017) established a regret bound of $\tilde{O}(\sqrt{n})$ 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 (Arora and Kale, 2007; Tsuda et al., 2005; Warmuth and Kuzmin, 2006, 2008; Kalai and Vempala, 2005, Garber et al., 2015).

In the full information setting there has also been work on analyzing gradient descent and mirror descent in RKHS spaces (McMahan and Orabona, 2014; Balandat et al., 2016). 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 ($\mu_j \leq O(j^{-\beta})$) we can guarantee the regret is bounded as $R_n \leq O(n^{\frac{\beta}{2\beta - 1}})$. Under exponential eigendecay we can guarantee an even sharper bound on the regret of $R_n \leq \tilde{O}(n^{1/2})$. We also provide 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 4 we present two applications of our framework, and prove that our algorithms are computationally efficient in these settings. In Section 5 we present experimental evidence of our claims, while all the proofs and technical details 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 \( t = \{1, \ldots, n\} \), the player chooses an action vector \( \{a_t\}_{t=1}^n \in A \subset \mathbb{R}^d \). The underlying kernel function at each round is \( K(\cdot, \cdot) \) 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 (Scholkopf and Smola, 2001)). Let \( \Phi(\cdot) : \mathbb{R}^d \to \mathbb{R}^D \) denote a feature map of \( K(\cdot, \cdot) \) such that for every \( x, y \) we have \( K(x, y) = \langle \Phi(x), \Phi(y) \rangle_\mathcal{H} \). 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 \( \mathcal{H} \), \( w_t \in W \subset \mathbb{R}^D \) and at each round the loss incurred by the player is \( \langle a_t, w_t \rangle_\mathcal{H} \). We assume that the adversary is oblivious, that is, it is a function of the previous actions of the player \( (a_1, \ldots, a_{t-1}) \) but unaware of the randomness used to generate \( a_t \). We let the size of the sets \( A, W \) be bounded\(^2\) in kernel norm, that is,

\[
\sup_{a \in A} K(a, a) \leq G^2 \quad \text{and} \quad \sup_{w \in W} \langle w, w \rangle_\mathcal{H} \leq G^2.
\]

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

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

where \( a^* = \inf_{a \in A} \{ \sum_{t=1}^n \langle \Phi(a), w_t \rangle_\mathcal{H} \} \) 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

\(^2\) 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.
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.

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 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 \( K(\cdot, \cdot) \). This motivates the definition of \( \epsilon \)-approximate kernels.

**Definition 2 (\( \epsilon \)-approximate kernels)** Let \( K_1 \) and \( K_2 \) be two kernels over \( A \times A \) and let \( \epsilon > 0 \). We say \( K_2 \) is an \( \epsilon \)-approximation of \( K_1 \) if for all \( x, y \in A \), \(|K_1(x, y) - K_2(x, y)| \leq \epsilon \).

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 \( K(\cdot, \cdot) \), 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 \( \langle \Phi(a_t), w_t \rangle_H \) – and not the action of the adversary \( w_t \). We also assume that the player’s action set \( A \) has finite cardinality. 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 \( K \).

Whenever no approximate kernel is needed, for example when \( D < \infty \) we allow the adversary to be able to choose an action \( w_t \in W \subset \mathbb{R}^D \) without imposing extra requirements on the set \( W \) other than being bounded in \( H \) 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 \( A \subset \mathbb{R}^d \) and \( \mathbb{P} \) a probability measure supported over \( A \). We denote by \( L_2(A; \mathbb{P}) \) the space of square integrable functions over \( A \) and measure \( \mathbb{P} \), \( L_2(A; \mathbb{P}) := \left\{ f : A \to \mathbb{R} \left| \int_A f^2(x)d\mathbb{P}(x) < \infty \right. \right\} \).

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

---

3. This assumption can be relaxed to let \( A \) be a compact set when \( K \) is Lipschitz continuous. In this setting we can instead work with an appropriately fine approximating cover over the set \( A \).
Appendix C. If we have access to the eigenfunctions of the kernel $K$, then Mercer’s theorem yields a natural way to construct a feature map $\Phi(x)$ for $x \in A$. The proof of this lemma is a simple application of Mercer’s theorem and is relegated to Appendix C. 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).

**Theorem 5 (Mercer’s Theorem)** Let $A \subset \mathbb{R}^d$ be compact and $\mathbb{P}$ be a finite Borel measure with support $A$. Suppose $K$ is a continuous square integrable positive definite kernel on $A$, and define a positive definite operator $T_K : L_2(A; \mathbb{P}) \mapsto L_2(A; \mathbb{P})$ by

$$(T_K f)(\cdot) := \int_A K(\cdot, x) f(x) \, d\mathbb{P}.$$ 

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

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

where the convergence of the series holds uniformly.

Mercer’s theorem yields a natural way to construct a feature map $\Phi(x)$ for $K$ by defining the $i^{\text{th}}$ 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 $\langle \cdot, \cdot \rangle_\mathcal{H}$. Armed with Mercer’s theorem we first present a simple deterministic procedure to obtain a finite dimensional $\epsilon$-approximate kernel of $K$. Essentially when the eigenfunctions of the kernel are uniformly bounded, $\sup_{x \in A} |\phi_j(x)| \leq B$ 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 $K$ under a finite Borel measure $\mathbb{P}$ with support $A$ and eigenfunctions $\{\phi_j\}_{j=1}^\infty$ with $\mu_1 \geq \mu_2 \geq \cdots$. Further assume that $\sup_{j \in \mathbb{N}} \sup_{x \in A} |\phi_j(x)| \leq B$ for some $B < \infty$. Let $m(\epsilon)$ be such that $\sum_{j=m+1}^\infty \mu_j \leq \epsilon / 4B^2$. Then the kernel induced by a truncated feature map,

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

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

The Hilbert space induced by the $\hat{K}^\epsilon_m$ 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 $K$ we can construct and work with $\hat{K}^\epsilon_m$ because as Lemma 6 shows $\hat{K}^\epsilon_m$ is an $\epsilon / 4$-approximation to $K$. Additionally, $\hat{K}^\epsilon_m$ also has the same first $m$ Mercer eigenvalues and eigenfunctions under $\mathbb{P}$ as $K$. 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).

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 $i^{\text{th}}$ coordinate and zeros everywhere else. So for any two $v_i$ and $v_j$ with $i \neq j$ we have $\langle v_i, v_j \rangle_\mathcal{H} = 0$. 

6
Theorem 9 Let $K$ be the number of samples $p$ defined by projecting $\Phi(x)$ to the random subspace $S_m$. Define the stochastic feature map, $\Phi(x) = P_{S_m}(\Phi(x))$.

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

Let $x_1, x_2, \ldots, x_p \sim \mathbb{P}$ be $p$ i.i.d. samples and construct the sample (kernel) covariance matrix, $\Sigma := \frac{1}{p} \sum_{i=1}^p \Phi(x_i)\Phi(x_i)^\top$. 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 $\Phi_m(x), \hat{K}_m(x, y) = \langle \Phi_m(x), \Phi_m(y) \rangle_{\mathcal{H}}$ will also be an $\epsilon$-approximation for the original kernel $K$. 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 $B_m = \frac{2B}{\delta_m} (1 + \sqrt{\frac{\alpha}{2}})$, $2\delta_m > \sqrt{\epsilon} > 0$ and $2 \geq \frac{2B^2\epsilon}{\sqrt{\epsilon}}$. Then the finite dimensional kernels $\hat{K}_m^o$ and $\hat{K}_m$ satisfy the following properties with probability $1 - e^{-\alpha}$:

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

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

Theorem 9 shows that given $\epsilon > 0$ the finite dimensional proxy $\hat{K}_m$ is an $\epsilon$-approximation of $K$ with high probability as long as sufficiently large number of samples are used. Furthermore, the top $m$ eigenvalues of the second moment matrix of $K$ are at most $\sqrt{\epsilon}/2$-away from the eigenvalues of the second moment matrix of $\hat{K}_m$ 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 $\hat{K}$ 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 have been analyzed extensively applied to linear losses under bandit feedback (Dani et al., 2008; Cesa-Bianchi and Lugosi, 2006; Bubeck et al., 2012). 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

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 $L(A; \mathbb{P})$ 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 $\Phi_m(\cdot)$

1. sample $x_1, \ldots, x_p \sim \mathbb{P}$.
2. construct sample Gram matrix $\hat{K}_{i,j} = \frac{1}{p} \mathcal{K}(x_i, x_j)$.
3. calculate the top $m$ eigenvectors of $\hat{K} \to \{\omega_1, \omega_2, \ldots, \omega_m\}$.
4. for $j = 1, \ldots, m$ do
   5. Set $v_j = \sum_{k=1}^{p} \omega_{jk} \Phi(x_k)$ ( $\omega_{jk}$ is the $k^{th}$ component of $\omega_j$)
5. end
6. define the feature map

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

the feature map is infinite dimensional, this becomes challenging and we need to build an approximate feature map $\Phi_m(\cdot)$ using Algorithm 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 1. The rest of the algorithm is then almost identical to the exponential weights algorithm (EXP 2) studied for linear bandits in (Dani et al., 2008; Cesa-Bianchi and Lugosi, 2006; Bubeck et al., 2012). In Algorithm 3 we set the exploration distribution $\nu_{\mathcal{A}J}$ to be such that it induces John's distribution ($\nu_J$) over $\Phi_m(\mathcal{A}) := \{\Phi_m(a) \in \mathbb{R}^m : a \in \mathcal{A}\}$ (first introduced as an exploration distribution in (Bubeck et al., 2012); also a short discussion is presented in Appendix F.1). Note that for finite sets it is possible to build minimal volume ellipsoid containing $\text{conv}(\Phi_m(\mathcal{A}))$–John’s ellipsoid and John’s distribution in polynomial time (Grötschel et al., 2012;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

---

6. It is thus possible to construct $\nu_J$ over $\Phi_m(\mathcal{A})$ in polynomial time. However, as $\mathcal{A}$ is a finite set, using $\Phi_m(\cdot)$ and $\nu_J$ it is also possible to construct $\nu_J$ 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 $\nu^A_J$ over $\mathcal{A}$, kernel map $\mathcal{K}$, effective dimension $m(\epsilon)$, number of samples $p$.

1. Build kernel $\hat{K}_m$ with feature map $\Phi_m(\cdot)$ using Algorithm 1 with kernel $\mathcal{K}$, dimension $m$, distribution $\mu$, bias tolerance $\epsilon$ and number of samples $p$.

2. set $q_1(a) = \nu^A_J$.

3. for $t = 1, \ldots, n$ do

4. set $p_t = \gamma \nu^A_J + (1 - \gamma) q_t$

5. choose $a_t \sim p_t$

6. observe $\langle \Phi(a_t), w_t \rangle_H$

7. build the covariance matrix

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

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

9. update $q_{t+1}(a) \propto q_t(a) \cdot \exp (-\eta \cdot \langle \hat{w}_t, \Phi_m(a) \rangle_H)$

end

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 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 $\gamma = \eta G^4 m$. Then Algorithm 3 with probability $1 - e^{-\alpha}$ has regret bounded by

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

We prove this theorem in Appendix A. Note that this is similar to the regret rate attained for adversarial linear bandits in (Dani et al., 2008; Cesa-Bianchi and Lugosi, 2012; Bubeck et al., 2012) 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 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 $\mu_j \leq C j^{-\beta}$.

2. $\mathcal{K}$ is said to have $(C, \beta)$-exponential eigenvalue decay if for all $j \in \mathbb{N}$ we have $\mu_j \leq C e^{-\beta j}.$
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 $\mathcal{K}_m$. We establish bounds on $m$ in Proposition 34 presented in Appendix C.1.

**Corollary 12** Let the conditions stated in Theorem 10 hold. Then Algorithm 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 $\beta > 1$. Then by choosing $\eta = \frac{1}{3 \beta^2 - 1} \cdot \left( \frac{\beta - 1}{4CB^2} \right)^{1/2} \cdot \left( \frac{\log(|A|)}{(e-1)G^4} \right)^{\frac{2\beta}{\beta-1}} \cdot m = \left[ \frac{4CB^2}{(\beta - 1)\epsilon} \right]^{1/\beta - 1}$ and $m = \left[ \frac{4CB^2}{(\beta - 1)\epsilon} \right]^{1/\beta - 1}$, the expected regret is bounded by
   
   $$R_n \leq 3^{\frac{2\beta - 2}{\beta - 1}} \left[ \frac{4CB^2}{\beta - 1} \right]^{\frac{1}{\beta - 1}} \left( (e - 1)G^4 \log(|A|) \right)^{\frac{\beta - 1}{\beta - 1}} \cdot n^{\frac{\beta}{\beta - 1}}.$$  

2. If $\mathcal{K}$ has $(C, \beta)$-exponential eigen-decay under measure $\mu$. Then by choosing
   
   $$\eta = \left( \frac{\beta \log(|A|)}{(e-1)G^4 \log \left( \frac{4CB^2}{\beta} \right) n} \right)^{1/2} \quad \text{and} \quad m = \frac{1}{\beta \log \left( \frac{4CB^2}{\beta} \right)} \left( \frac{4CB^2}{\beta \epsilon} \right) \quad \text{where} \quad \epsilon = \frac{(e - 1)G^4}{3\beta \log \left( \frac{4CB^2}{\beta} \right)},$$
   
   with $n$ large enough so that $\epsilon < 1$, the expected regret is bounded by
   
   $$R_n \leq \left[ \frac{9(e - 1)G^4 \log(|A|) \cdot n}{\beta \log \left( \frac{4CB^2}{\beta} \right)} \right]^{1/2} \left\{ \log \left( \frac{12CB^2}{\beta \log(|A|) \log \left( \frac{4CB^2}{\beta \epsilon} \right) G^4}^{1/2} \right) \right\}.$$  

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

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 a small estimate of the adversarial action $\hat{w}_t$ which has low bias, however this estimate would have large variance ($O(m)$). 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 $m = O(\log(n))$ 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 $m = O(n^{1/2})$ 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 $\beta \to 1$ (the effective dimension $m \to \infty$) the regret bound becomes close to linear in $n$. 

10
Algorithm 3: Full Information: Exponential Weights

\textbf{Input} : Set $\mathcal{A}$, learning rate $\eta > 0$, number of rounds $n$.

1. Set $p_1(a)$ uniform distribution over $\mathcal{A}$.
2. for $t = 1, \ldots, n$ do
   3. choose $a_t \sim p_t$
   4. observe $\langle \Phi(a_t), w_t \rangle_H$
   5. update $p_{t+1}(a) \propto p_t(a) \cdot \exp \left( -\eta \cdot \langle w_t, \Phi(a) \rangle_H \right)$
3. end

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, $\langle \Phi(a_t), w_t \rangle_H$ we update the distribution by a multiplicative factor, $\exp(-\eta \langle w_t, \Phi(a) \rangle_H)$. 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 $O(\sqrt{n})$ for the exponential weights algorithm.

**Theorem 14** Assume that in Algorithm 3 the step size $\eta$ is chosen to be, $\eta = \sqrt{\frac{\log(\text{vol}(\mathcal{A}))}{e-2}}$. Let $n$ large enough such that $\sqrt{\frac{\log(\text{vol}(\mathcal{A}))}{e-2}} \frac{1}{n^{1/2}} \leq 1$. Then the expected regret after $n$ rounds is bounded by,

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

We prove this regret bound in Appendix E.1.

3.2 Conditional Gradient Descent

Next we present an online conditional gradient (Frank-Wolfe) method (Hazan and Kale, 2012) 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 Hazan (2016)). 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 $O(n^{3/4})$ (presented in Appendix E.2).

**Theorem 15** 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 $R_n \leq 8G^2 n^{3/4}$.
Algorithm 4: Full Information: Conditional Gradient Method

**Input**: Set $A$, number of rounds $n$, initial action $a_1 \in A$, inner product $\langle \cdot, \cdot \rangle_H$, learning rate $\eta$, mixing rates $\left\{ \gamma_t \right\}_{t=1}^n$.

1. $X_1 = \Phi(a_1)$
2. choose $D_1$ such that $\mathbb{E}_{x \sim D_1} \Phi(x) = X_1$
3. for $t = 1, 2, \ldots, n$ do
   4. sample $a_t \sim D_t$
   5. observe the loss $\langle \Phi(a_t), w_t \rangle_H$
   6. define $F_t(Y) \triangleq \eta \sum_{s=1}^{t-1} \langle w_s, Y \rangle_H + \|Y - X_1\|_H^2$
   7. compute $v_t = \text{argmin}_{a \in A} \langle \nabla F_t(X_t), \Phi(a) \rangle_H$
   8. update mean $X_{t+1} = (1 - \gamma_t)X_t + \gamma_t \Phi(v_t)$
   9. choose $D_{t+1}$ s.t. $\mathbb{E}_{x \sim D_{t+1}} [\Phi(x)] = X_{t+1}$.
10. end

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 $\|A\|_F^2 + \|b\|_2^2 \leq G^2$. If we embed this pair into a Hilbert space defined by the feature map $(A, b)$ we get a kernel loss defined as $\langle \Phi(x), (A, b) \rangle_H = x^\top Ax + b^\top x$, where $\Phi(x) = (xx^\top, x)$ is the associated feature map for any $x \in A$ 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 (Warmuth and Kuzmin, 2006, 2008; Garber et al., 2015; Allen-Zhu and Li, 2017). 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 16 (Proof in Appendix E.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 $\|a\|_2 \leq 1$, $a \in \mathbb{R}^d$ is tractable in $\tilde{O}(d^4)$ 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, $v_t = \text{argmin}_{a \in A} \langle \nabla F_t(X_t), \Phi(a) \rangle_H$. When the set of actions is $A = a \in \mathbb{R}^d : \|a\|_2 \leq 1$ then under quadratic losses this problem
becomes,

$$v_t = \arg\min_{a \in A} a^\top Ba + b^\top a,$$

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 $A$. The dual problem is the (semi-definite program) SDP,

$$\max - t - \mu$$

s. t.

$$\begin{bmatrix} B + \mu I & b/2 \\ b/2 & t \end{bmatrix} \succ 0.$$  

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 Boyd and Vandenberghe (2004)).

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 17 (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 18 (Posynomial) A function $f : \mathbb{R}_+^d \mapsto \mathbb{R}$ defined as

$$f(x) = \sum_{k=1}^{m} c_kx_1^{\alpha_{1k}}x_2^{\alpha_{2k}}\cdots x_d^{\alpha_{dk}},$$

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,

$$x = (\underbrace{x_1, x_2, \cdots, x_{d-2}, x_{d-1}}_{s_1}, \underbrace{x_d}_{s_m}),$$
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 (Boyd et al., 2007) 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 \(A = \{a \in \mathbb{R}^d : \|a\|_2 \leq 1\}\). Additionally we all choose the initial action to be the solution to the optimization problem,
\[
a_1 = \arg\min_{a \in A} \sum_{k=1}^{d} \Phi(a)_i.
\]

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 Boyd et al. (2007)). At each round of the conditional gradient descent algorithm requires us to solve the optimization problem,
\[
v_t = \arg\min_{a \in A} \left( \sum_{s=1}^{t-1} w_t + 2(X_t - \Phi(a_1)), \Phi(a) \right)_{\mathcal{H}}.
\]

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.

5. 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 show plots of the exponential weights performance on Gaussian losses. In the bandit feedback setting we again study quadratic and Gaussian losses.

Full Information

Exponential weights requires to sample from a distribution of the form \(p(x) \propto \exp(\mu \sum_{i=1}^{t} K(x, w_i))\). Although in general sampling from these distributions is possibly intractable, they present
Online learning with kernel losses

Figure 1: Quadratic with linear term Full Information.

Figure 2: Gaussian Losses Full Information.

Figure 3: Quadratics with linear term Bandit Feedback.

Figure 4: Gaussian Losses Bandit Feedback.

good empirical performance. The following plot shows a diffusion MCMC algorithm sampling from a distribution proportional to $\propto \exp\left(-\eta \sum_{i=1}^{3} K(x, z_i)\right)$ where $K$ 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**

The kernel exponential weights algorithm presents also a sublinear regret profile. The Gaussian experiments involved the construction of the finite dimensional kernel $K_m$ via kernel PCA.

**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. Under bandit
feedback it would also be interesting to explore if it is possible to establish lower bounds on the regret under the eigen-decay conditions stated.

ACKNOWLEDGMENTS

References

Jacob D Abernethy, Elad Hazan, and Alexander Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization.

Alekh Agarwal and Ofer Dekel. Optimal algorithms for online convex optimization with multi-point bandit feedback. In COLT, pages 28–40, 2010.

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.

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

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.

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.

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.

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.

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

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

Stephen Boyd and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

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

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.
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.

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

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

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

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.

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

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.

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.

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.

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

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

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

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

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

Elad Hazan and Yuanzhi Li. An optimal algorithm for bandit convex optimization. arXiv preprint arXiv:1603.04350, 2016.
Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Alan J Hoffan and Helmut W Wielandt. The variation of the spectrum of a normal matrix. 1953.

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

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

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

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

H Brendan McMahan and Avrim Blum. Online geometric optimization in the bandit setting against an adaptive adversary. Springer, 2004.

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.

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

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.

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.

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

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.

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

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

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.

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

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

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.

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.

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 A. Bandits Exponential Weights Regret Bound

In this section we prove the regret bound stated in Section 2.3. Here we borrow all the notation from Section 2. As defined before the expected regret for Algorithm 3 after \( n \) rounds is

\[
R_n = E \left[ n \sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle_H - \langle \Phi(a^*), w_t \rangle_H \right],
\]

where \( p_t \) is the exponential weights distribution described in Algorithm 3, \( a^* \) is the optimal action and \( F_{t-1} \) is the sigma field that conditions on the events up to the end of round \( t - 1 \). Now we expand this and have,

\[
R_n = E \left[ \sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t} \left[ \langle \Phi(a_t), w_t \rangle_H - \langle \Phi(a^*), w_t \rangle_H \biggm| F_{t-1} \right] \right]
\]

\[
+ E \left[ \sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t} \left[ \langle \Phi_m(a^*), w_t \rangle_H - \langle \Phi(a^*), w_t \rangle_H \biggm| F_{t-1} \right] \right]
\]

\[
+ E \left[ \sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t} \left[ \langle \Phi_m(a_t), w_t \rangle_H - \langle \Phi_m(a^*), w_t \rangle_H \biggm| F_{t-1} \right] \right],
\]

where \( R_m \) is the regret when we play the distribution is Algorithm 3 but are hit with losses that are governed by the kernel \( \tilde{K}_m(\cdot, \cdot) \) (with the same \( a^* \) as before). Also note that as the proxy kernel \( \tilde{K}_m \) is close uniformly by Theorem 9 we have,

\[
R_n \leq 2\varepsilon n + R_m.
\]

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 some terms in the regret bound being zero \( (\varepsilon = 0) \). 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 \( y_t \in \mathbb{R}^d \). Also observe that in \( R_m \) only the component of \( w_t \) in the \( \tilde{S}_m \) contributes to the inner product. Thus every term of the form \( \langle \Phi_m(a_t), w_t \rangle = \langle \Phi_m(a_t), \Phi_m(y_t) \rangle = \tilde{K}_m(y_t, a_t) \).

A.1 Regret Bound for finite dimensional proxy

In this section we will now bound \( R_m \). In this section we will assume the following,

\[
\mathbb{E}_{p_t} [\tilde{w}_t|\mathcal{F}_t] = \Phi_m(y_t) + \mathcal{B}_{\text{Bias}}.
\]
as \( \hat{w}_t \) depends on draws from \( p_t \). We will control the bias \( \xi_t \) in Appendix A.2. Given this assumption we have the following lemma.

**Lemma 19 (Some estimates involving \( \hat{w}_t \))** Conditioned on \( F_{t-1} \) for any fixed \( a \in A \) we have,

\[
\mathbb{E}[\langle \hat{w}_t, \Phi_m(a) \rangle] = \hat{K}_m(y_t, a) + \langle \xi_t, \Phi_m(a) \rangle.
\]

We also have,

\[
\mathbb{E}[\hat{K}_m(y_t, a_t)] = \mathbb{E}\left[ \left. \sum_{a \in A} p_t(a) \langle \hat{w}_t, \Phi_m(a) \rangle \right| F_{t-1} \right] - \mathbb{E}\left[ \left. \sum_{a \in A} p_t(a) \langle \xi_t, \Phi_m(a) \rangle \right| F_{t-1} \right].
\]

**Proof** The first claim follows by the assumption (8) and the linearity of expectation. Conditioned on \( F_{t-1} \) we have

\[
\mathbb{E}[\langle \hat{w}_t, \Phi_m(a) \rangle] = \langle \mathbb{E}[\hat{w}_t], \Phi_m(a) \rangle = \hat{K}_m(y_t, a) + \langle \xi_t, \Phi_m(a) \rangle.
\]

where the expectation is taken over \( p_t \). Now to prove the other part of the theorem statement we will use tower property. Observe that conditioned on \( F_{t-1} \), \( p_t \) and \( a_t \) become measureable.

\[
\mathbb{E}[\hat{K}_m(y_t, a_t)] = \mathbb{E}\left[ \left. \hat{K}_m(y_t, a_t) \right| F_{t-1} \right] = \mathbb{E}\left[ \left. \sum_{a \in A} p_t(a) \langle \hat{w}_t, \Phi_m(a) \rangle - \langle \xi_t, \Phi_m(a) \rangle \right| F_{t-1} \right] - \mathbb{E}\left[ \left. \sum_{a \in A} p_t(a) \langle \xi_t, \Phi_m(a) \rangle \right| F_{t-1} \right].
\]

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

**Proof** [Proof of Theorem 10]
Using Lemma 19 we can write the cumulative loss, the first term in $R_{m}^{n}$ as

$$
E \left[ \sum_{t=1}^{n} \hat{K}_{m}(a_{t}, y_{t}) \right] = E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} p_{t}(a) \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right] 
- E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} p_{t}(a) \langle \xi_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right] 
= (1 - \gamma)E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} q_{t}(a) \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right]
$$

\[\text{Exploitation}\]

$$
+\gamma \cdot E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} \nu_{t}^{A}(a) \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right]
- E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} p_{t}(a) \langle \xi_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right]
$$

\[\text{Exploration} \]

\[= B_{e, bias} \]  

where the second line follows by the definition of $p_{t}$. Under our choice of $\gamma$ by Lemma 22 we know that $\eta \langle \hat{w}_{t}, \Phi(a_{t}) \rangle > -1$. Therefore by Hoeffding’s inequality (Lemma 51) we get,

$$
E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} q_{t}(a) \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right] 
\leq -\frac{1}{\eta} E \left[ \sum_{t=1}^{n} \log \left( E_{q_{t}} \left[ \exp \left( -\eta \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \right) \right] \right) \right]
\leq \Gamma_{1}

+ (e - 2)\eta E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} q_{t}(a) \left( \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \right)^{2} \left| F_{t-1} \right. \right] \right]
\leq \Gamma_{2}

(10)

We will see that $\Gamma_{1}$ is a telescoping series and is controlled in Lemma 20. While the second term $\Gamma_{2}$ is the variance of the estimated loss is bounded in Lemma 21. Plugging in appropriate bounds on $\Gamma_{1}$ and $\Gamma_{2}$ from these lemmas we see that the exploitation term is bounded as,

$$
(1 - \gamma) \cdot E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} q_{t}(a) \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right] 
\leq E \left[ \sum_{i=1}^{n} \hat{K}_{m}(a^{*}, y_{i}) \right] + \frac{1}{\eta} \log (|A|) + (e - 2)\eta G_{mn}^{4}. 

(11)

Next observe that the exploration term is bounded above as

$$
\gamma \cdot E \left[ \sum_{t=1}^{n} \left[ \sum_{a \in A} \nu_{t}^{A}(a) \langle \hat{w}_{t}, \Phi_{m}(a) \rangle \left| F_{t-1} \right. \right] \right] \leq \gamma n.

(12)
Finally by Lemma 23 we can control the bias $B_c$ as
\[
|B_c| = \left| \mathbb{E} \left[ \sum_{t=1}^{n} \left( \sum_{a \in A} p_t(a) \langle \xi_t, \Phi_m(a) \rangle \right) \right] \right| \leq \epsilon n.
\]

Putting these all together into (9) we get the finite dimensional regret
\[
R_n^m = \mathbb{E} \left[ \sum_{t=1}^{n} \left( \hat{K}_m(a_t, y_t) - \hat{K}_m(a^*, y_t) \right) \right] \leq \gamma n + (e - 2)G^4mn + \epsilon n + \frac{1}{\eta} \log(|A|).
\]

Plugging this bound into (7) we get a bound on the expected regret as
\[
R_n \leq \gamma n + (e - 2)G^4mn + 3\epsilon n + \frac{1}{\eta} \log(|A|)
\]
completing the proof. $\blacksquare$

First let us focus on bounding $\Gamma_1$.

**Lemma 20** Let $\Gamma_1$ be as defined in (10) then we have
\[
\Gamma_1 \geq -\eta \mathbb{E} \left[ \sum_{i=1}^{n} \hat{K}_m(a^*, y_t) \right] - \log(|A|).
\]

**Proof** We ignore the outer expectation and the conditioning by $F_{t-1}$ inside the the sum to ease notation, then we have
\[
\Gamma_1 = \sum_{i=1}^{n} \log (\mathbb{E}_{q_t} [\exp (-\eta \langle \hat{w}_t, \Phi_m(a) \rangle)])
\]
\[
\overset{(i)}{=} \sum_{i=1}^{n} \log \left\{ \frac{\sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{t-1} \langle \hat{w}_i, \Phi_m(a) \rangle \right) \cdot \exp \left( -\eta \langle \hat{w}_t, \Phi_m(a) \rangle \right)}{\sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{t-1} \langle \hat{w}_i, \Phi_m(a) \rangle \right)} \right\}
\]
\[
\overset{(ii)}{=} \log \left( \sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{n-1} \langle \hat{w}_i, \Phi_m(a) \rangle \right) \cdot \exp \left( -\eta \langle \hat{w}_n, \Phi_m(a) \rangle \right) \right) - \log (|A|)
\]
\[
+ \sum_{t=1}^{n} \left\{ \log \left( \sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{t-1} \langle \hat{w}_i, \Phi_m(a) \rangle \right) \right) - \log \left( \sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{t} \langle \hat{w}_i, \Phi_m(a) \rangle \right) \right) \right\}
\]
\[
\overset{(iii)}{=} \log \left( \sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{n} \langle \hat{w}_i, \Phi_m(a) \rangle \right) \right) - \log (|A|)
\]
where $(i)$ follows by the definition of $q_t(a)$ and $(ii)$ is by expanding the sum. The $\log(|A|)$ term is because we start off with a uniform distribution over all elements. Finally $(iii)$ follows as the sum telescopes. Finally we have for a single element $a^*$,
\[
\log \left( \sum_{a \in A} \exp \left( -\eta \sum_{i=1}^{n} \langle \hat{w}_i, \Phi(a) \rangle \right) \right) \geq -\eta \sum_{i=1}^{n} \langle \hat{w}_i, \Phi(a^*) \rangle = \eta \sum_{i=1}^{n} \hat{K}_m(a^*, y_t).
\]
The next lemma controls of the variance of the expected loss – $\Gamma_2$.

**Lemma 21** Let $\Gamma_2$ be defined as in (10) and the choice of parameters as specified in Theorem 10 then we have

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

**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,

$$
\mathbb{E}_{q_t} \left[ \langle \hat{w}_t, \Phi_m(a) \rangle^2 \right] = \Phi_m(a)^\top \mathbb{E}_{a \sim p_t} \left[ K(a, y_t)^2 \Sigma_t^{-1} \Phi_m(a) \Phi_m(a)^\top \right] \Phi_m(a)
$$

$$
\leq G^4 \Phi_m(a)^\top \Sigma_t^{-1} \Phi_m(a).
$$

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

$$
\mathbb{E}_{a \sim p_t} \left[ \mathbb{E}_{a \sim p_t} \left[ \langle \hat{w}_t, \Phi_m(a) \rangle^2 \right] \right] \leq G^4 \mathbb{E}_{a \sim p_t} \left[ tr \left( \Phi_m(a)^\top \Sigma_t^{-1} \Phi_m(a) \right) \right]
$$

$$
= G^4 tr \left( \Sigma_t^{-1} \mathbb{E}_{a \sim p_t} \left[ \Phi_m(a) \Phi_m(a)^\top \right] \right) = G^4 tr \left( I_{m \times m} \right) = G^4 m.
$$

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

Note that while using Hoeffding’s bound 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, we can ensure that this condition holds.

**Lemma 22** For any $a \in A$ we have

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

for all $t = 1, \ldots, n$. Thus if the exploration parameter is $\gamma > \eta G^4 m$ then we have a bound on the estimated loss at each round

$$
\eta |\langle \hat{w}_t, \Phi_m(a) \rangle_{\mathcal{H}}| \leq 1
$$

for all $a \in A$.  

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

$$|\hat{w}_t^\top \Phi_m(a)| = |\mathcal{K}(a_t, w_t) (\Sigma_m^{-1}\Phi_m(a_t))^\top \Phi_m(a)| \leq \mathcal{G}^2 \sup_{c,d \in \mathcal{A}} \left|\Phi_m(c)^\top \Sigma_m^{-1}\Phi_m(d)\right|.$$  

(14)

Now we note that the matrix $\Sigma_m$ has its lowest eigenvalue lower bounded by $\gamma/m$ by Proposition 36 (see also discussion in (Bubeck et al., 2012). Thus we have that,

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

Combing this with (14) yields the desired claim.

Next 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] Under our choice of $\gamma = \eta \mathcal{G}^4 m$ the regret bound becomes,

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

Case 1: First we assume $(C, \beta)$-polynomial eigen-value decay. By the results of Proposition 34 we have a sufficient condition for the choice of $m$,

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

Under this choice of $m$ we equate the terms $n_1$ and $n_2$ to get,

$$n_1 = (e-1)\eta \left[ \frac{4CB^2}{(\beta-1)\epsilon} \right]^{1/\beta-1} \mathcal{G}^4 n = 3\epsilon n = n_2.$$  

Rearranging terms we get,

$$\epsilon = \left( \frac{(e-1)\eta \mathcal{G}^4}{3} \right)^{(\beta-1)/\beta} \left[ \frac{4CB^2}{\beta-1} \right]^{1/\beta}.$$  

Under this choice of $\epsilon$ we now equate $n_1$ and $n_3$ to get

$$n_1 = 3^{1/\beta} \left( (e-1)\mathcal{G}^4 \right)^{(\beta-1)/\beta} \left[ \frac{4CB^2}{\beta-1} \right]^{1/\beta} \cdot \eta^{(\beta-1)/\beta} \cdot n = \frac{1}{\eta} \log (|\mathcal{A}|) = n_3.$$  

25
This yields the choice,

\[ \eta = \frac{1}{3^{2^{\beta^{-1}}} \cdot \left[ \frac{\beta - 1}{4CB^2} \right]^{1/2}} \cdot \left[ \frac{\log(|A|)}{((e - 1)G^4)^{\beta^{-1}} \cdot n} \right]^{\frac{\beta}{2^{\beta^{-1}}} \cdot \left[ \log\left(\frac{4CB^2}{\beta e}\right) \right]} \cdot n^{\frac{\beta}{2^{\beta^{-1}}}}. \]

Under this choice we have the regret bounded as,

\[ \mathcal{R}_n \leq 3n_3 = 3^{2^{\beta^{-1}}} \cdot \left[ \frac{4CB^2}{\beta - 1} \right]^{1/2} \cdot \left[ \frac{\log(|A|)}{((e - 1)G^4)^{\beta^{-1}} \cdot n} \right]^{\frac{\beta}{2^{\beta^{-1}}} \cdot \left[ \log\left(\frac{4CB^2}{\beta e}\right) \right]} \cdot n^{\frac{\beta}{2^{\beta^{-1}}}}. \]

**Case 2:** Here we assume \((C, \beta)\)-exponential eigen-value decay. By the results of Proposition 34 we have a sufficient condition for the choice of \(m\),

\[ m = \frac{1}{\beta} \log \left( \frac{4CB^2}{\beta e} \right). \]

Under this choice of \(m\) we now equate \(n_1\) and \(n_2\) to get,

\[ n_1 = \frac{(e - 1)G^4}{\beta} \eta \log \left( \frac{4CB^2}{\beta e} \right) = 3\epsilon n = n_2. \]

The choice

\[ \epsilon = \frac{(e - 1)G^4}{3\beta} \eta \log \left( \frac{4CB^2}{\beta} \right) \]

ensures that \(n_1 > n_2\) when \(\epsilon < 1\). Next we equate \(n_2\) and \(n_3\) to get

\[ n_2 = \frac{(e - 1)G^4}{\beta} \eta \log \left( \frac{4CB^2}{\beta} \right) = \frac{1}{\eta} \log (|A|) = n_3. \]

Rearranging terms we get,

\[ \eta = \left( \frac{\beta \log (|A|)}{(e - 1)G^4 \log \left( \frac{4CB^2}{\beta} \right)} \cdot n \right)^{1/2}. \]

Thus with this choice of \(\eta\) and the fact that \(n_1 > n_2 = n_3\) we get

\[ \mathcal{R}_n \leq 3n_1 = \left[ 9(e - 1)G^4 \log (|A|) \cdot n \right]^{1/2} \left\{ \log \left( \frac{12CB^2 n^{1/2}}{(\beta \log (|A|) \log \left( \frac{4CB^2}{\beta} \right) G^4)^{1/2}} \right) \right\}. \]

This establishes the claimed result.
A.2 Controlling the Bias

In this section we show how to control the norm of the bias $\xi_t$ in the estimate of $\hat{w}_t$

$$E_{p_t} [\hat{w}_t | \mathcal{F}_{t-1}] = \Phi_m(y_t) + \xi_t. \quad (15)$$

First we define the Mercer basis of the finite dimensional kernel $\hat{K}_m(\cdot, \cdot)$ as the set of functions $(\phi_1^{(t),\mathcal{L}}(\cdot), \ldots, \phi_m^{(t),\mathcal{L}}(\cdot))$. Note that by definition this form an orthonormal basis of the space of functions in $L_2(\mathcal{A}, p_t)$. Also these functions form an orthogonal basis for a subspace of the entire Hilbert space $\mathcal{H}$ defined by $\mathcal{K}$. We have

$$\hat{K}_m(x, y) = \sum_{i=1}^{m} \hat{\mu}_i^{(t)} \phi_i^{(t),\mathcal{L}}(x) \phi_i^{(t),\mathcal{L}}(y).$$

Thus we can identify feature vectors for this kernel as,

$$\Phi_m(x) = \left[ \sqrt{\hat{\mu}_1^{(t)}} \phi_1^{(t),\mathcal{L}}(x), \ldots, \sqrt{\hat{\mu}_m^{(t)}} \phi_m^{(t),\mathcal{L}}(x) \right].$$

We assume that the Hilbert space $\mathcal{H}$ is separable. We then use the Gram-Schmidt procedure to extend this basis to an orthonormal basis of the entire Hilbert space $\mathcal{H}$ using first the $L_2(\mathcal{A}, p_t)$ inner product to get a basis,

$$\left( \phi_1^{(t),\mathcal{L}}(\cdot), \phi_2^{(t),\mathcal{L}}(\cdot), \ldots, \phi_m^{(t),\mathcal{L}}(\cdot), \phi_{m+1}^{(t),\mathcal{L}}(\cdot), \ldots \right).$$

We can also extend this basis to the entire Hilbert space using the Hilbert space inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ defined by the kernel $\mathcal{K}$. To get an alternate basis,

$$\left( \phi_1^{(t),\mathcal{H}}(\cdot), \phi_2^{(t),\mathcal{H}}(\cdot), \ldots, \phi_m^{(t),\mathcal{H}}(\cdot), \phi_{m+1}^{(t),\mathcal{H}}(\cdot), \ldots \right).$$

Note that the first $m$ basis functions are exactly the same, that is, $\phi_i^{(t),\mathcal{L}} = \phi_i^{(t),\mathcal{H}}$ for $i = 1, 2, \ldots, m$. While for $j > m$ we have

$$\phi_j^{(t),\mathcal{H}}(\cdot) = \sum_{k=1}^{\infty} C_{j,k} \phi_k^{(t),\mathcal{L}}(\cdot).$$

This follows as $\phi_j^{(t),\mathcal{H}}(\cdot)$ is a function in $L(\mathcal{A}, p_t)$. Note that by definition of the kernel $\hat{K}_m(\cdot, \cdot)$ and $\mathcal{K}$ we have,

$$\hat{K}_m(x, y) = \Phi_m(x)\Phi_m(y)$$

$$\mathcal{K}(x, y) = \Phi_m(x)\Phi_m(y) + \Phi_m^c(x)\Phi_m^c(y)$$

7. $(f, g) = \int_{\mathcal{A}} f(x)g(x)dp_t(x)$
where $\Phi_m^c(\cdot)$ is a feature vector built using the functions that are orthogonal to the subspace defined by $\hat{K}_m(\cdot, \cdot)$.

$$
\Phi_m^c(x) = \begin{bmatrix}
0 \\
\vdots \\
0 \\
\beta_{m+1}\phi_{m+1}(x) \\
\beta_{m+2}\phi_{m+2}(x) \\
\vdots 
\end{bmatrix}
$$

With this machinery setup, we now control the bias at each time step.

Lemma 23 Let the bias at each step be defined as in (9) then we have

$$
|B_\epsilon| = \left| \mathbb{E} \left[ \sum_{t=1}^{n} \mathbb{E} \left[ \langle \xi_n, \Phi_m(a) \rangle \big| F_{t-1} \right] \right] \right| \leq \epsilon n.
$$

Proof We lighten the notation to denote $\Sigma_m^{(t)}$ as $\Sigma_t$. We begin by expanding a single term in $B_\epsilon$ using its definition,

$$
|B_\epsilon| = \mathbb{E}_{a_t \sim p_t, a \sim p_t} \left[ \langle \xi_n, \Phi_m(a) \rangle \big| F_{t-1} \right]
$$

$$(i) = \mathbb{E}_{a_t \sim p_t, a \sim p_t} \left[ (K(a_t, y_t) - \hat{K}_m(a_t, y_t)) \Phi_m(a) \Sigma_t^{-1} \Phi_m(a) \big| F_{t-1} \right]
$$

$$
= \mathbb{E}_{a_t \sim p_t} \left[ \Phi_m^c(w_t) \Phi_m(a) \Sigma_t^{-1} \Phi_m(a) \big| F_{t-1} \right]
$$

$$
= \Phi_m(y_t) \left. \left( \mathbb{E}_{a \sim p_t} \left[ \Phi_m^c(a) \Phi_m(a) \Sigma_t^{-1} \Phi_m(a) \right] \right) \right|_{V_t}
$$

$$
\leq \|\Phi_m^c(y_t)\| \|V_t\| \\
(ii) \leq \sqrt{\epsilon} \cdot \|V_t\| \leq \epsilon.
$$

where $(i)$ follows as $\hat{K}_m(a_t, w_t)\Phi_m(a) \Sigma_t \Phi_m(a)$ is an unbiased estimate of $\Phi_m(y_t)$, $(ii)$ follows as $\hat{K}_m$ is an $\epsilon$-approximation of $K$. The last step $(iii)$ follows by Lemma 24 which establishes that $\|V\| \leq \sqrt{\epsilon}$. Summing over $t = 1$ to $n$ completes the proof.

We will now control the norm of $V_t$ where

$$
V_t := \mathbb{E}_{a \sim p_t} \left[ \underbrace{\mathbb{E}_{a_t \sim p_t} \left[ \Phi_m^c(a) \Phi_m(a) \right]}_{:=V'_t} \underbrace{\Sigma_t^{-1} \Phi_m(a)}_{:=V_t} \right].
$$

(16)
Lemma 24 Let $V_t$ be defined as in (16), then for any $t = 1, \ldots, n$ we have

$$\|V_t\| \leq \sqrt{\epsilon}.$$ 

Proof In the proof we drop the dependence on the index of the round $t$ to lighten the notation. We will now simplify to find the structure of both the vector $V'$ ($V'_t$) and the cross covariance matrix $\Sigma^{CR}_m$ ($\Sigma^{CR}_{m,t}$). First we have that, by construction the inverse covariance matrix $\Sigma^{-1}_m$ ($\Sigma^{-1}_{m,t}$) is of the form

$$\Sigma^{-1}_m = \begin{bmatrix} D_m & 0 \\ 0 & 0 \end{bmatrix}$$

with $D_m$ being a diagonal $m \times m$ matrix with entries,

$$D_m = \begin{bmatrix} 1/\hat{\mu}_1 & 0 & \cdots & 0 \\ 0 & 1/\hat{\mu}_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1/\hat{\mu}_m \end{bmatrix}.$$

We also have the feature map of $\tilde{\mathcal{K}}_m$ is

$$\Phi_m(a) = \begin{bmatrix} \sqrt{\hat{\mu}_1} \phi^L_1(a) \\ \sqrt{\hat{\mu}_2} \phi^L_2(a) \\ \vdots \\ \sqrt{\hat{\mu}_m} \phi^L_m(a) \\ 0 \\ \vdots \end{bmatrix}.$$ 

Thus we have by the definition of $V'$

$$V' = \Sigma^{-1}_m \Phi_m(a)$$

$$= \begin{bmatrix} \phi^L_1(a)/\sqrt{\hat{\mu}_1} \\ \phi^L_2(a)/\sqrt{\hat{\mu}_1} \\ \vdots \\ \phi^L_m(a)/\sqrt{\hat{\mu}_1} \\ 0 \\ \vdots \end{bmatrix}.$$
Next we explore the structure of the matrix $\Sigma_m^{CR}$,

\[
\Sigma_m^{CR} = \mathbb{E}_{a_t \sim p_t} \left[ \Phi_m^c(a_t) \Phi_m(a_t)^\top \right] = \mathbb{E}_{a_t \sim p_t} \begin{bmatrix}
0 \\
\vdots \\
0 \\
\beta_{m+1} \phi_{m+1}^H(a_t) \\
\beta_{m+2} \phi_{m+2}^H(a_t) \\
\vdots \\
\beta_i \phi_i^H(a_t) \\
\sqrt{\mu_i} \phi_i(a_t) \\
\sqrt{\mu_{i+m}} \phi_m(a_t) \\
0 \\
\vdots
\end{bmatrix} = \begin{bmatrix}
\beta_i \\
\sqrt{\mu_j} \phi_j^H(a_t) \\
\sqrt{\mu_m} \phi_m(a_t) \\
0 \\
\vdots
\end{bmatrix}.
\]

Here $\Sigma_{2,1}^{CR}$ is a matrix with infinite rows and $m$ columns. Note that the $(i, j)$ element of $\Sigma_m^{CR}$ with $i > m$ and $j \leq m$ can be identified as the $(i - m, j)$ element of $\Sigma_{2,1}^{CR}$,

\[
(\Sigma_m^{CR})_{i,j} := (\Sigma_{2,1}^{CR})_{i-m,j} = \mathbb{E}_{a_t \sim p_t} \left[ \sqrt{\mu_j} \beta_i \phi_i^H(a_t) \phi_j^H(a_t) \right].
\]

Note that for $j \leq m$ we have $\phi_j^H(a_t) = \phi_j^L(a_t)$ and for $i > m$ we have the relation, $\phi_i^H(a_t) = \sum_{k=1}^{\infty} C_{i,k} \phi_k^L(a_t)$. Substituting this into expression above we get,

\[
(\Sigma_m^{CR})_{i,j} = (\Sigma_{2,1}^{CR})_{i-m,j} = \mathbb{E}_{a_t \sim p_t} \left[ \sqrt{\mu_j} \beta_i \phi_i^H(a_t) \phi_j^H(a_t) \right] = \mathbb{E}_{a_t \sim p_t} \left[ \beta_i \phi_i^L(a_t) \sum_{k=1}^{\infty} C_{i,k} \phi_k^L(a_t) \right] = \sum_{k=1}^{\infty} \mathbb{E}_{a_t \sim p_t} \left[ \beta_i C_{i,k} \sqrt{\mu_j} \phi_j^L(a_t) \phi_k^L(a_t) \right] = \beta_i C_{i,j} \sqrt{\mu_j}. \tag{17}
\]
where the last equality follows as $\phi_j$ and $\phi_k$ are orthogonal functions in $L(A; p_t)$. Plugging this into the expression for $V$, we have that

$$
V = \mathbb{E}_{a \sim p_t} \left[ \Sigma_{m}^{CR} V' \right]
$$

$$
= \mathbb{E}_{a \sim p_t} \left[ \begin{array}{cc}
0 & 0 \\
\Sigma_{2,1}^{CR} & 0
\end{array} \right] \left[ \begin{array}{c}
\phi_1^c(a)/\sqrt{\mu_1} \\
\phi_2^c(a)/\sqrt{\mu_2} \\
\vdots \\
\phi_m^c(a)/\sqrt{\mu_m} \\
0 \\
\vdots
\end{array} \right]
$$

$$
= \mathbb{E}_{a \sim p_t} \left[ \begin{array}{c}
0 \\
\vdots \\
\beta_{m+2} \left( \sum_{k=1}^{m} C_{m+2,k} \phi_K^c(a) \right) \\
\beta_{m+1} \left( \sum_{k=1}^{m} C_{m+1,k} \phi_K^c(a) \right)
\end{array} \right]
$$

(18)

where the last equality again follows by the expression for $\Sigma_{2,1}^{CR}$ calculated in (17). Finally we have,

$$
\|V\|^2 \overset{(i)}{=} \sum_{p=m+1}^{\infty} \beta_p^2 \left( \sum_{k=1}^{m} C_{p,k} \mathbb{E}_{a \sim p_t} \left[ \phi_k^c(a) \right] \right)^2
$$

$$
\overset{(ii)}{\leq} \sum_{p=m+1}^{\infty} \beta_p^2 \mathbb{E}_{a \sim p_t} \left[ \sum_{k=1}^{m} C_{p,k} \phi_k^c(a) \right]^2
$$

$$
\overset{(iii)}{\leq} \sum_{p=m+1}^{\infty} \beta_p^2 \mathbb{E}_{a \sim p_t} \left[ \sum_{k=1}^{\infty} C_{p,k} \phi_k^c(a) \right]^2
$$

$$
= \phi_p^H(a)
$$

$$
\overset{(iv)}{=} \sum_{p=m+1}^{\infty} \beta_p^2 \mathbb{E}_{a \sim p_t} \phi_p^H(a)^2
$$

$$
\overset{(v)}{=} \mathbb{E}_{a \sim p_t} \left[ \sum_{p=m+1}^{\infty} \left( \beta_p \phi_p^H(a) \right)^2 \right]
$$

$$
\overset{(vi)}{\leq} \epsilon
$$

where (i) follows by the expression of $V$ calculated in (18), (ii) follows by Jensen’s inequality, (iii) follows by extending the inner sum to $\infty$ and by the orthogonality of $\phi_K^c$ under $L(A; p_t)$, (iv) is by the relation of $\phi_p^H(\cdot)$ and $\phi_p^c(\cdot)$, (v) is by switching the sum and the integral and finally (vi) follows as $\hat{K}_m$ is an $\epsilon$-approximation of $K$. Taking square roots we have that $\|V\|_2 \leq \sqrt{\epsilon}$. ■
Appendix B. Kernel principal component analysis

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

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 $K \in \mathbb{R}^{p \times p}$, where the sample Gram matrix is defined entry-wise as $K_{ij} = \frac{K(x_i, x_j)}{p}$.

**Proof** Let $X \in \mathbb{R}^{p \times D}$ be such that the $i^{th}$ row is $\frac{\Phi(x_i)}{p}$. The SVD decomposition of $X$ is $X = UD\Sigma V^\top$ with $U \in \mathbb{R}^{p \times p}$, $D \in \mathbb{R}^{p \times D}$ and $V \in \mathbb{R}^{D \times D}$. Therefore $X^\top X = V D^\top D V^\top$ and $X X^\top = UDD^\top U^\top$. We identify $X^\top X$ as the sample covariance matrix and $X X^\top$ 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 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 $\omega_j \in \mathbb{R}^p$). 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

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

**Appendix 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 $\mathbb{K}$ under measure $\mathbb{P}$ with eigenfunctions $\{\phi_j\}_{j=1}^{\infty}$, further 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 $\sum_{j=m+1}^{\infty} \mu_j \leq \frac{\epsilon}{4B^2}$ and the $m$th eigen-gap as $\delta_m = \frac{1}{2} (\mu_m - \mu_{m+1})$. $S_m$ and $\hat{S}_m$ are the subspaces spanned by the first $m$ eigenvectors of the covariance matrix $\mathbb{E}_{x \sim \mathbb{P}} [\Phi(x)\Phi(x)^\top]$ and a sample covariance matrix $\frac{1}{p} \sum_{i=1}^{p} \Phi(x_i)\Phi(x_i)^\top$. Define $P_{S_m}$ and $P_{\hat{S}_m}$ to be the projection operators to $S_m$ and $\hat{S}_m$. Let

$$\hat{\mathbb{K}}_{(m)}^o(x,y) = \langle P_{S_m}(\Phi(x)), P_{S_m}(\Phi(y)) \rangle_{\mathcal{H}} = \langle \Phi^o_m(x), \Phi^o_m(y) \rangle_{\mathcal{H}}$$

be a deterministic approximate kernel and define a stochastic proxy approximate kernel

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

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

**Lemma 26** Given $\epsilon > 0$, let $\{\mu_j\}_{j=1}^{\infty}$ be the Mercer operator eigenvalues of $\mathbb{K}$ 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 $\sum_{j=m+1}^{\infty} \mu_j \leq \frac{\epsilon}{4B^2}$. Then the kernel induced by a truncated feature map,

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

induces a kernel $\hat{\mathbb{K}}_m := \langle \Phi^o_m(x), \Phi^o_m(y) \rangle_{\mathcal{H}} = \sum_{j=1}^{m} \mu_j \phi_j(x) \phi_j(y)$, for all $(x,y) \in \mathcal{A} \times \mathcal{A}$ that is an $\epsilon/4$-approximation of $\mathbb{K}$. 33
Proof By definition, for all \( x, y \in A \)

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

The reverse inequality is also true therefore,

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

for all \( x, y \in A \).

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

**Theorem 27** Let \( m, p \) be as in Lemma 6 and let \( \epsilon > 0 \). Define the \( m \)-th level eigen-gap as \( \delta_m = \frac{1}{2} (\mu_m - \mu_{m+1}) \). Also let \( B_m = \frac{2G}{\delta_m} (1 + \sqrt{\frac{\epsilon}{2}}) \), \( \delta_m/2 > w > 0 \) and \( p \geq \frac{B^2 G^2}{w} \).

The finite dimensional proxies \( \hat{\mathcal{K}}_m^o \) and \( \hat{\mathcal{K}}_m \) satisfies 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. \( |\hat{\mathcal{K}}_m(x, y) - \hat{\mathcal{K}}_m^o(x, y)| \leq w \forall x, y \in A \).
3. 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 \( \sup_{i=1,\ldots,m} |\mu_i^{(m)} - \mu_i| \leq w \)

Theorem 27 shows that, as long as sufficiently samples \( p(m) \) are used, with high probability \( \hat{\mathcal{K}}_m \) is uniformly close to \( \hat{\mathcal{K}}_m^o \) and therefore to \( \mathcal{K} \). We prove this theorem via a series of lemmas and auxiliary theorems. We first prove part (1) and (2) of the theorem 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 via samples. We leverage a kernel PCA result from (Zwald and Blanchard, 2006) to construct \( \hat{\mathcal{K}}_m \).

**Theorem 28** *Adapted from Theorem 4 in (Zwald and Blanchard, 2006)*

If \( m, p, S_m, \hat{S}_m, \delta_m, B_m \) and \( \alpha \) be defined as stated in Theorem 27 then with probability \( 1 - \exp(-\alpha) \) we have

\[
\|P_{S_m} - P_{\hat{S}_m}\|_F \leq \frac{B_m}{\sqrt{p(m)}}. 
\]

In particular,

\[
\hat{S}_m \subset \left\{ g + h, g \in S_m, h \in S_m^\perp, \|h\|_\mathcal{H} \leq \frac{2B_m}{\sqrt{p}} \|g\|_\mathcal{H} \right\}.
\]
Now using this theorem we prove part (1) of Theorem 27.

**Lemma 29** With probability $1 - e^{-\alpha}$ we have,

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

for all $x, y \in \mathcal{A}$.

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

\[
\|\Phi(x) - P_{S_m}(\Phi(x))\|_\mathcal{H} \leq \|\Phi(x) - P_{S_m}(\Phi(x))\|_\mathcal{H} + \|P_{S_m}(\Phi(x)) - P_{\hat{S}_m}(\Phi(x))\|_\mathcal{H}
\]

\[
\leq \frac{\sqrt{\epsilon}}{2} + \|\Phi(x)\|_\mathcal{H} \|P_{S_m} - P_{\hat{S}_m}\|_{op}
\]

\[
\leq \frac{\sqrt{\epsilon}}{2} + G \frac{B_m}{\sqrt{p(m)}}
\]

\[
\leq \frac{\sqrt{\epsilon}}{2} + w
\]

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 28 and $(iv)$ is by the choice of $B_m$. Therefore with probability at least $1 - e^{-\alpha}$ for all $x \in \mathcal{A}$

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

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 $\|h_x\|_\mathcal{H} \leq \frac{\sqrt{\epsilon}}{2} + w$ and $\|h_y\|_\mathcal{H} \leq \frac{\sqrt{\epsilon}}{2} + w$. We know 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,

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

This establishes the claim.  

We now move on to the proof of claim (2) in Theorem 27.

**Lemma 30** If $p, B_m$ are chosen as stated in Theorem 27 we have

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

(21)
The feature map for $\hat{K}_m$ is $P_{S_m}(\Phi(x))$ for all $x \in A$ while $P_{\hat{S}_m}(\Phi(x))$ is the feature map for $\hat{K}_m$. We first show that for all $x \in A$,

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

where the second inequality follows by applying Theorem 28 and the last inequality follows by the choice of $B_m$. A similar argument as in the proof of Lemma 29 lets us then conclude that,

$$
|\hat{K}_m(x, y) - \hat{K}_o^m(x, y)| \leq w \ \forall x, y \in A.
$$

We now proceed to prove part (3) of Theorem 27. We will show that the Mercer operator eigenvalues of $\hat{K}_m$ are close to Mercer operator eigenvalues of $\hat{K}_o^m$. We first recall a useful result from (Mendelson et al., 2006).

**Theorem 31 (Adapted from Theorem 3.3 in (Mendelson et al., 2006))** Let $K$ be a kernel over $A \times A$ such that $\sup_{x \in A} K(x, x) \leq \mathcal{G}$. Also let $\hat{\mu}_1 \geq \hat{\mu}_2 \geq \cdots \geq \hat{\mu}_N$ be the eigenvalues of the Gram matrix $(K(x_i, x_j)/N)_{i,j=1}^N$ for $\{x_i\}_{i=1}^N \sim P$. Then there exists a universal constant $c$ such that for every $t > 0$

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

(22)

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

**Proposition 32** 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 $x_1, \cdots, x_N \sim P$, the Gram matrices of $\hat{K}_m^o$ ($\mathbb{K}_m^o(N)$) and $\hat{K}_m$ ($\mathbb{K}_m(N)$) are close in Frobenius norm as the matrices are close element-wise by part (2) of Theorem 27.

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

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 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 (Hoffman and Wielandt, 1953) this implies that the ordered eigenvalues are close,

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

Theorem 31 and Proposition 32 together imply that the same statement of Theorem 31 with the Gram matrix replaced by the sample covariance matrix holds.
Theorem 33  The Mercer operator eigenvalues $\mu_1^{(m)} \geq \cdots \geq \mu_m^{(m)}$ and $\mu_1 \geq \cdots \geq \mu_m$ of $\hat{K}_m$ and $K^0_m$ follow

$$\sup_{i=1,\ldots,m} |\mu_i^{(m)} - \mu_i| \leq w \quad (23)$$

Proof We will use the probabilistic method. By Theorem 31, 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 $K_m(N)$ and $K^0_m(N)$ 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 $t > 0$

$$\sup_{i=1,\ldots,m} |\mu_i^{(m)} - \mu_i| \leq \sup_{i_1} |\mu_{i_1}^{(m)} - \mu_{i_1}| + \sup_{i_2} |\mu_{i_2}^{(m)} - \mu_{i_2}^{(m,0)}| + \sup_{i_3} |\mu_{i_3}^{(m,0)} - \mu_{i_3}|$$

$$\leq t + w + t$$

$$= w + 2t$$

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

As a consequence of this theorem it follows that as long as $w < \frac{\delta m^2}{2}$, part (3) of Theorem 27 holds.

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 34 Let the conditions stated in Theorem 9 and Lemma 6 hold.

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

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

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

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

$$m \geq \frac{1}{\beta} \log \left( \frac{4CB^2}{\beta \epsilon} \right)$$

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

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

$$\sum_{j=m+1}^{\infty} \mu_j \leq \frac{\epsilon}{4B^2}.$$
We will prove the bound assuming a \((C, \beta)\)-polynomial eigenvalue decay, the calculation is similar when we have exponential eigenvalue decay. Note that,

\[
\sum_{j=m+1}^{\infty} \mu_j \leq \sum_{j=m+1}^{\infty} C j^{-\beta} \\
\leq \int_{m}^{\infty} C x^{-\beta} dx \\
= \frac{C}{\beta - 1} \frac{1}{m^{\beta - 1}}.
\]

We demand that

\[
\frac{C}{\beta - 1} \frac{1}{m^{\beta - 1}} \leq \frac{\epsilon}{4B^2}
\]

rearranging terms yields the desired claim. \(\blacksquare\)

### Appendix D. Properties of the Covariance matrix – \(\Sigma^{(t)}_m\)

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

**Lemma 35** Let \(\mu_m^{(t)}\) be the \(m\)-th (smallest) eigenvalue of \(\Sigma^{(t)}_m\). 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.

\[
v^\top \Sigma^{(t)}_m v = (1 - \gamma) \cdot v^\top E_{x \sim q_t} [\Phi_m(x)\Phi_m(x)^\top] v + \gamma \cdot v^\top E_{x \sim \nu} [\Phi_m(x)\Phi_m(x)^\top] v.
\]

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

Observe that by our discussion in Appendix F.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 \(\nu_{\mathcal{J}}^A\) is the exploration distribution over \(\mathcal{A}\) then \(\mu_m = 1/m\).

**Proposition 36** If \(\nu_{\mathcal{J}}^A\) be the exploration distribution then we have

\[
\mu_m^{(t)} \geq \frac{\gamma r}{m}.
\]

### D.1 Finite Sample Analysis

Next we analyze the sample complexity of the operation of building the second moment matrix in Algorithm 3 using samples. Let \(\hat{\Sigma}^{(t)}_m\) be the second moment matrix estimate built by using \(x_1, \cdots, x_r\) drawn i.i.d. from \(p_t\).

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

38
We will show how to choose $r$ 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 some observations:

**Remark 37 (Covariance eigenvalues are Mercer’s eigenvalues)** The eigenvalues $\mu_1^{(t)} \geq \cdots \geq \mu_m^{(t)}$ of $E_{x \sim p_t} [\Phi_m(x)(\Phi_m(x)^\top)$ are exactly Mercer operator eigenvalues for $\hat{K}_m$ under $p_t$.

**Remark 38 (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 $\hat{K}_m^{(t)}_{x,x} = (\hat{K}_m(x_i,x_j))_{i,j=1}^p$.

We formalize the above remark in Lemma 42. We will use an auxiliary lemma from Zwald and Blanchard (2006) which we present here for completeness.

**Lemma 39 (Lemma 1 in (Zwald and Blanchard, 2006))** Let $K'$ be a kernel over $X \times X$ such that $\sup_{x \in X} K'(x,x) \leq G'$. Let $\Sigma'$ be the covariance of $\Phi'(x), x \sim P$. If $\hat{\Sigma}'_m$ is the sample covariance built by using $r$ samples $x_1, \cdots, x_r \sim P$, with probability $1 - \exp(-\delta)$:

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

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 40** 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 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 41** 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

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

(where $c$ is the same constant as in Theorem 31) 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 31 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 31, this is true if \( \exp(-\zeta) \leq 2 \exp \left( -\frac{\gamma \mu_m^2}{2G} \sqrt{r \log(r)} \right) \). This yields the condition,

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

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

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 39, we have that

\[
\|\Sigma_m^{(t)} - \hat{\Sigma}_m^{(t)}\|_\text{op} \leq \frac{(\gamma \lambda_m)^2 \epsilon_1}{2},
\]

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

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

As the matrices \( A \) and \( B \) are close with high probability, Lemma 40 proves that the inverses are also close,

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

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

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 42 that formalizes the connection between the eigenvalues the Gram matrix and sample covariance matrix.
Lemma 42 For any \(x, y \in A\):

\[
\Phi_m(x) \top \left( \Sigma_m^{(t)} \right)^{-1} \Phi_m(y) = A_x \top \left( \hat{K}_{m,p}^{(t)} \right)^2 \top A_y
\]

Where \(A_x = (\hat{K}_m(x, x_1), \cdots, \hat{K}_m(x, x_p)) \top\) and \(A_y = (\hat{K}_m(y, x_1), \cdots, \hat{K}_m(y, x_p)) \top\).

Proof Run an SVD decomposition of both sides and check that they agree. \(\blacksquare\)

Given Lemma 41 we also prove a bound on the distance between the estimates of adversarial actions generated in Algorithm 3. Define \(\tilde{w}_t^{(2)} := \left( \Sigma_m^{(t)} \right)^{-1} \Phi_m(a_t) \hat{K}(a_t, w_t)\) and let \(\hat{w}_t := \tilde{w}_t^{(1)} = \left( \Sigma_m^{(t)} \right)^{-1} \Phi_m(a_t) \hat{K}(a_t, w_t)\).

Corollary 43 We have that

\[
\|\tilde{w}_t^{(2)} - \tilde{w}_t^{(1)}\|_H \leq \epsilon_1 \sqrt{G\mathcal{L}}.
\]

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.

Appendix E. Full Information Regret Bounds

E.1 Exponential Weights Regret Bound

In this section we prove a regret bound for exponential weights and present a proof of Theorem 14. The analysis of the regret is similar to the analysis of exponential weights for linear losses (see for example a review in Bartlett (2014)). 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 \((a_{t-1}, w_{t-1}, \ldots, a_1, w_1)\).

Proof [Proof of Theorem 14] By the tower property and by the definition of the regret we can write the cumulative loss as,

\[
\mathbb{E} \left[ \sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle \right] = \mathbb{E} \left[ \sum_{t=1}^{n} \mathbb{E}_{a_t \sim p_t} \left[ \langle \Phi(a_t), w_t \rangle \bigg| \mathcal{F}_{t-1} \right] \right]
= \mathbb{E} \left[ \sum_{t=1}^{n} \left[ \int_A p_t(a) \langle \Phi(a), w_t \rangle da \bigg| \mathcal{F}_{t-1} \right] \right].
\]
Observe that our choice of $\eta$ implies that $\eta \langle \Phi(a), w_t \rangle > -1$. By invoking Hoeffding’s inequality (stated as Lemma 51) we get
\[
E \left[ \sum_{t=1}^{n} \left[ \int_{A} p_t(a) \langle \Phi(a), w_t \rangle da \bigg| \mathcal{F}_{t-1} \right] \right] 
\leq -\frac{1}{\eta} E \left[ \sum_{t=1}^{n} \log \left( E_{a \sim p_t} \left[ \exp \left( -\eta \langle \Phi(a), w_t \rangle \right) \right] \bigg| \mathcal{F}_{t-1} \right) \right] + (e-2)\eta E \left[ \sum_{t=1}^{n} \int_{A} p_t(a) \langle \Phi(a), w_t \rangle^2 da \bigg| \mathcal{F}_{t-1} \right] 
\]
\[
\leq -\frac{\Gamma}{\eta} + (e-2)\eta G^4 n, 
\]
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
\[
E \left[ \sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle \right] \leq E \left[ \sum_{t=1}^{n} \langle \Phi(a^*), w_t \rangle \right] + \frac{\log(\text{vol}(\mathcal{A}))}{\eta} + (e-2)\eta G^4 \cdot n. 
\]
Rearranging terms we have the regret is bounded by
\[
R_n \leq (e-2)G^4 \eta n + \frac{\log(\text{vol}(\mathcal{A}))}{\eta}. 
\]
The choice of $\eta = \sqrt{\log(\text{vol}(\mathcal{A}))/\sqrt{(e-2)G^2 n^{1/2}}}$, 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 3. Also let $\Gamma$ be defined as follows
\[
\Gamma = E \left[ \sum_{t=1}^{n} \log \left( E_{a \sim p_t} \left[ \exp \left( -\eta \langle \Phi(a), w_t \rangle \right) \right] \bigg| \mathcal{F}_{t-1} \right) \right]. 
\]
Then we have that,
\[
\Gamma \geq -\eta E \left[ \sum_{i=1}^{n} \langle \Phi(a^*), w_i \rangle \right] - \log(\text{vol}(\mathcal{A})), 
\]
where $a^*$ is the optimal action in hindsight in the definition of regret and $\text{vol}(\mathcal{A})$ is the volume of the set $\mathcal{A}$.

**Proof** Expanding $\Gamma$ using the definition of $p_t$ we have that,
\[
\Gamma \overset{(i)}{=} E \left[ \sum_{t=1}^{n} \log \left( \frac{\int_{A} \exp \left( -\eta \sum_{i=1}^{t} \langle \Phi(a), w_i \rangle \right) da}{\int_{A} \exp \left( -\eta \sum_{i=1}^{t-1} \langle \Phi(a), w_i \rangle \right) da} \right) \right] 
\overset{(ii)}{=} E \left[ \log \left( \int_{A} \exp \left( -\eta \sum_{i=1}^{n} \langle \Phi(a), w_i \rangle \right) da \right) \right] - \log(\text{vol}(\mathcal{A})), 
\]
The second constraint can be rewritten as

\[ \eta \beta \lambda \]

Completing the squares (whenever \( \sum \lambda \) eigenvalues

Proof [Proof of Lemma 16] Let \( v_1, \ldots, v_d \) an orthonormal basis of eigenvectors of \( B \) with eigenvalues \( \lambda_1, \ldots, \lambda_d \) 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 \( \lambda_i \neq 0 \)),

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

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,

\[ t(\beta) \propto \exp \left( \sum_{i=1}^d \lambda_i \beta_i \right) \]

\[ 0 \leq \beta_i \]

\[ \sum_{i=1}^d \left( \sqrt{\beta_i - \frac{\gamma_i}{2 \lambda_i}} \right)^2 \leq 1. \]

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 \((\eta \beta_i + (1-\eta) \tilde{\beta}_i) \geq 0, \forall i\).

The second constraint can be rewritten as

\[ \sum_{i=1}^d \hat{\beta}_i - \frac{\gamma_i \sqrt{\hat{\beta}_i}}{\lambda_i} + \left( \frac{\gamma_i}{2 \lambda_i} \right)^2 \leq 1 \quad (24) \]

\[ \sum_{i=1}^d \tilde{\beta}_i - \frac{\gamma_i \sqrt{\tilde{\beta}_i}}{\lambda_i} + \left( \frac{\gamma_i}{2 \lambda_i} \right)^2 \leq 1. \quad (25) \]
These equations imply that,
\[
\eta \left[ \sum_{i=1}^{d} \hat{\beta}_i - \frac{\gamma_i \sqrt{\beta_i}}{\lambda_i} + \left( \frac{\gamma_i}{2\lambda_i} \right)^2 \right] + (1 - \eta) \left[ \sum_{i=1}^{d} \tilde{\beta}_i - \frac{\gamma_i \sqrt{\tilde{\beta}_i}}{\lambda_i} + \left( \frac{\gamma_i}{2\lambda_i} \right)^2 \right] \leq 1.
\]

By concavity of \(\sqrt{\cdot}\) we have
\[
\sum_{i=1}^{d} \eta \frac{\gamma_i \sqrt{\beta_i}}{\lambda_i} + (1 - \eta) \frac{\gamma_i \sqrt{\beta_i}}{\lambda_i} \leq \sum_{i=1}^{d} \frac{\gamma_i \sqrt{\eta \beta_i} + (1 - \eta) \tilde{\beta}_i}{\lambda_i},
\]
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 (Lovász and Vempala, 2007) to sample from \(t(\beta)\) in \(O(d^4)\) steps and convert to samples from \(q(\cdot)\) using the method described above. In case some eigenvalues are zero, say without loss of generality \(\lambda_1, \ldots, \lambda_R\). Then set \(\beta_i = \alpha_i^2\) for \(i \in \{R + 1, \ldots, d\}\) and sample from the distribution,
\[
t(\beta) \propto \exp \left( \sum_{i=1}^{R} \gamma_i \alpha_i + \sum_{i=R+1}^{d} \lambda_i \beta_i \right)
\]
\[
0 \leq \beta_i
\]
\[
\sum_{i=1}^{R} \alpha_i^2 + \sum_{i=R+1}^{d} \left( \sqrt{\beta_i} - \frac{\gamma_i}{2\lambda_i} \right)^2 \leq 1.
\]

The analysis follows as before for this case as well.

**E.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 (Hazan, 2016)). To prove this we first prove the regret bound of a different algorithm – follow the regularized leader.

**E.3 Follow the Regularized Leader**

We present a version of follow the regularized leader (Shalev-Shwartz and Singer, 2007) (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 \(\Phi(A)\). 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 \(u \in A\)
\[
\sum_{t=0}^{n} \langle X_t, \Phi(u) \rangle_H \geq \sum_{t=0}^{n} \langle X_t, X_{t+1} \rangle_H.
\]
Algorithm 5: Follow the Regularized leader (FTRL)

**Input**: Set $A$, number of rounds $n$, initial action $a_1 \in A$, inner product $\langle \cdot, \cdot \rangle_H$, learning rate $\eta > 0$.

1. Let $X_1 = \arg\min_{X \in \text{conv}(\Phi(A))} \frac{1}{\eta} \langle X, X \rangle_H$
2. choose $D_1$ such that $E_{x \sim D_1} [\Phi(x)] = X_1$
3. for $t = 1, 2, 3, \ldots, n$ do
4. choose $a_t \sim D_t$
5. observe $\langle \Phi(a_t), w_t \rangle_H$
6. update $X_{t+1} = \arg\min_{X \in \text{conv}(\Phi(A))} \eta \sum_{s=1}^t \langle w_s, X \rangle_H + \langle X, X \rangle_H$
7. choose $D_{t+1}$ s.t. $E_{x \sim D_{t+1}} [\Phi(x)] = X_{t+1}$
8. 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 (Hazan, 2016).

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

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

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

$$B_R(x||y) \triangleq R(x) - R(y) - \langle \nabla R(y), (x-y) \rangle_H.$$

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

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

$$\langle w_t, X_t - X_{t+1} \rangle_H \leq 2\eta\|w_t\|^2_H.$$

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

$$g_t(X_t) = g_t(X_{t+1}) + \langle X_t - X_{t+1}, \nabla g_t(X_{t+1}) \rangle_H + B_{g_t}(X_t||X_{t+1})$$

$$\geq g_t(X_{t+1}) + B_{g_t}(X_t||X_{t+1}),$$

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

$$B_{g_t}(X_t||X_{t+1}) \leq g_t(X_t) - g_t(X_{t+1})$$

$$= (g_{t-1}(X_t) - g_{t-1}(X_{t+1})) + \eta\langle w_t, X_t - X_{t+1} \rangle_H$$

$$\leq \eta\langle w_t, X_t - X_{t+1} \rangle_H, \quad (26)$$
where the last inequality follows because $X_{t-1}$ is the minimizer of the function $g_{t-1}(\cdot)$ over $\text{conv}(\Phi(X))$. Observe that $B_{g_t}(X_t\|X_{t+1}) = \frac{1}{2}\|X_t - X_{t+1}\|_H^2$. Thus by the Cauchy-Schwartz inequality we have,

$$\langle w_t, X_t - X_{t+1}\rangle_H \leq \|w_t(X_t)\| \|X_t - X_{t+1}\|_H$$

$$= \|w_t\| \sqrt{2B_{g_t}(X_t\|X_{t+1})}.$$

Substituting the upper bound from (26) we get,

$$\langle w_t, X_t - X_{t+1}\rangle_H \leq \|w_t\| \cdot \sqrt{2\eta \langle w_t, X_t - X_{t+1}\rangle_H}.$$

Rearranging terms establishes the result.

\[
\text{Theorem 49} \quad \text{Given a step size } \eta > 0, \text{ the regret suffered by Algorithm 5 after } n \text{ rounds is bounded by}
\]

$$R_n \leq 2n\eta G^2 + \frac{2G^2}{\eta}.$$

\textbf{Proof} \quad \text{By the definition of regret we have}

$$R_n = \mathbb{E} \left[ \sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle_H - \min_{a \in A} \mathbb{E} \left[ \sum_{t=1}^{n} \langle \Phi(a_t), w_t \rangle_H \right] \right]$$

$$\overset{(i)}{=} \mathbb{E} \left[ \sum_{t=1}^{n} \mathbb{E}_{a_t \sim D_t} \left[ \langle w_t, \Phi(a_t) - \Phi(a^*) \rangle_H \big| \mathcal{F}_{t-1} \right] \right]$$

$$\overset{(ii)}{=} \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t - \Phi(a^*) \rangle_H \right]$$

$$\overset{(iii)}{\leq} \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t - X_{t+1} \rangle_H \right] + \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_{t+1} - \Phi(a^*) \rangle_H \right] + \frac{1}{\eta} \langle X_1, X_1 \rangle_H - \langle X_0, X_0 \rangle_H$$

$$\overset{(iv)}{\leq} \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t - X_{t+1} \rangle_H \right] + \frac{2G^2}{\eta}.$$

The first equality follows as $a^*$ is the minimizer, (ii) is by evaluating the expectation with respect to $D_t$, (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 bound the first term in (27) to get a regret bound. To control the first term we now invoke Lemma 48

$$R_n \leq 2\eta \sum_{t=1}^{n} \|w_t\|_H^2 + \frac{2G^2}{\eta}$$

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

This establishes the stated result.
E.4 Regret Bound for Algorithm 4

In deploying Algorithm 4 we will at each round find distributions over the action space $A$ 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 $D_t$ such that,
\[
\mathbb{E}_{a \sim D_t} \Phi(a) = X_t,
\]
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 $R(X) = \|X - X_1\|_H^2$ and applied to the shifted loss function, $\langle w_t, X - (X_t^* - X_t) \rangle_H$. Notice that,
\[
|\langle X, w_t \rangle_H - \langle X - (X_t^* - X_t), w_t \rangle_H| \leq \|w_t\|_H \|X_t^* - X_t\|_H \leq G \|X_t^* - X_t\|_H. \tag{28}
\]
We are now ready to prove Theorem 15.

**Proof** [Proof of Theorem 15] We denote the filtration upto round $t$ by $F_{t-1}$, that is, we condition of 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,
\[
R_n = \mathbb{E} \left[ \sum_{t=1}^{n} \mathbb{E}_{a_t \sim D_t} \left[ \langle w_t, \Phi(a_t) \rangle_H - \langle w_t, \Phi(a^*) \rangle_H \bigg| F_{t-1} \right] \right]
\]
\[
\overset{(i)}{=} \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t - \Phi(a^*) \rangle_H \right]
\]
\[
= \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t - X_t^* \rangle_H \right] + \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t^* - \Phi(a^*) \rangle_H \right]
\]
\[
\overset{(ii)}{\leq} \mathbb{E} \left[ \sum_{t=1}^{n} \langle w_t, X_t - X_t^* \rangle_H \right] + 2n\eta G^2 + \frac{2G^2}{\eta}
\]
\[
\overset{(iii)}{\leq} \mathbb{E} \left[ \sum_{t=1}^{n} \|w_t\|_H \|X_t - X_t^*\|_H \right] + 2n\eta G^2 + \frac{2G^2}{\eta},
\]
where $(i)$ follows by taking expectation with respect to $D_t$, $(ii)$ follows by invoking Theorem 49 and $(iii)$ is by Cauchy-Schwartz inequality. We finally need to bound $\Xi$ to establish a bound on the regret.

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

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

\[
\mathcal{R}_n \leq 2 \sum_{t=1}^{n} G^2 \sqrt{\gamma_t} + 2 n \eta G^2 + \frac{2 G^2}{\eta},
\]

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

\[
\mathcal{R}_n \leq 4 G^2 n^{3/4} + 2 n \eta G^2 + \frac{2 G^2}{\eta},
\]

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 function 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 15, such that \(\eta G \sqrt{h_{t+1}} \leq G^2 \gamma_t^2\). The iterates \(X_t\) satisfy, \(h_t \leq 4 G^2 \gamma_t\).

**Proof** The functions \(F_t\) is 1-smooth therefore we have,

\[
h_t(X_{t+1}) = F_t(X_{t+1}) - F_t(X_t^*)
\]

\[
= F_t(X_t + \gamma_t(\Phi(v_t) - X_t)) - F_t(X_t^*)
\]

\[
\leq F_t(X_t) - F_t(X_t^*) + \gamma_t(\Phi(v_t) - X_t, \nabla F_t(X_t))_{\mathcal{H}} + \frac{\gamma_t^2}{2} \|\Phi(v_t) - X_t\|_{\mathcal{H}}^2
\]

\[
\leq (1 - \gamma_t)(F_t(X_t) - F_t(X_t^*)) + \gamma_t^2 G^2,
\]

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,

\[
h_{t+1}(X_{t+1}) = F_t(X_{t+1}) - F_t(X_{t+1}^*) + \eta \langle w_{t+1}, X_{t+1} - X_{t+1}^* \rangle_{\mathcal{H}}
\]

\[
\leq F_t(X_{t+1}) - F_t(X_t^*) + \eta \langle w_{t+1}, X_{t+1} - X_{t+1}^* \rangle_{\mathcal{H}}
\]

\[
\leq h_t(X_{t+1}) + \eta G \|X_{t+1} - X_{t+1}^*\|_{\mathcal{H}}, \tag{29}
\]

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,

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

\[
\leq h_t(X_{t+1}) + \eta G \sqrt{h_{t+1}(X_{t+1})}.
\]
Plugging in the bound on $h_t(X_{t+1})$ from (29) into the above inequality gives us the recursion relation,

$$h_{t+1} \leq h_t(1 - \gamma_t) + \frac{\gamma_t^2 G^2}{\gamma_t \sqrt{h_t}}$$

\[\leq h_t(1 - \gamma_t) + \frac{\gamma_t^2 G^2}{\gamma_t} + \eta G \sqrt{h_t} \leq h_t(1 - \gamma_t) + 2 \gamma_t^2 G^2,
\]

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

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

$$h_{t+1} \leq h_t(1 - \gamma_t) + 2 \gamma_t^2 G^2$$

\[\leq 4 G^2 (\gamma_t (1 - \gamma_t)) + 2 \gamma_t^2 G^2 = 4 G^2 \gamma_t \left(1 - \frac{\gamma_t}{2}\right) \leq 4 G^2 \gamma_{t+1},
\]

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

Appendix F. Technical Results

We present a version of Hoeffding’s inequality (Hoeffding, 1963) that is used in the regret bound analysis of exponential weights.

**Lemma 51 (Hoeffding’s Inequality)** Given a bounded random variable $X \geq -1$ with and $\lambda > 0$,

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

and hence

$$\mathbb{E} \left[ X \right] \leq -\frac{1}{\lambda} \log \left( \mathbb{E} \left[ e^{-\lambda X} \right] \right) + (e - 2) \lambda^2 \mathbb{E} \left[ X^2 \right]. \quad (30)$$

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

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

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

We also present John’s theorem (see (Ball, 1997)) that we use to construct an exploration distribution.

**Theorem 52 (John’s Theorem)** Let \( \mathcal{K} \subset \mathbb{R}^d \) be a convex set denote the ellipsoid of minimal volume containing it as,

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

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

\[
x = c + d \sum_{i=1}^{q} p_i (x - c, u_i - c)(u_i - c),
\]

where \( \langle \cdot, \cdot \rangle \) is the inner product for which the minimal ellipsoid is the unit ball about its center \( c : \langle x, y \rangle = x^\top H y \).

This shows that

\[
x - c = d \sum_{i} p_i (u_i - c)(u_i - c)^\top H(x - c)
\]

\[\iff\]

\[
\tilde{x} = d \sum_{i} p_i \tilde{u}_i \tilde{u}_i^\top \tilde{x}
\]

\[\iff\]

\[
\frac{1}{d} I_{d \times d} = \sum_{i} p_i \tilde{u}_i \tilde{u}_i^\top
\]

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

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

(31)