Nonstochastic Contextual Combinatorial Bandits

Lukas Zierahn
Università degli Studi di Milano
Milan, Italy

Dirk van der Hoeven
Korteweg-de Vries Institute for Mathematics
University of Amsterdam
Amsterdam, The Netherlands

Nicolò Cesa-Bianchi
Università degli Studi di Milano and Politecnico di Milano
Milan, Italy

Gergely Neu
Universitat Pompeu Fabra
Barcelona, Spain

Abstract

We study a contextual version of online combinatorial optimisation with full and semi-bandit feedback. In this sequential decision-making problem, an online learner has to select an action from a combinatorial decision space after seeing a vector-valued context in each round. As a result of its action, the learner incurs a loss that is a bilinear function of the context vector and the vector representation of the chosen action. We consider two natural versions of the problem: semi-bandit where the losses are revealed for each component appearing in the learner's combinatorial action, and full-bandit where only the total loss is observed. We design computationally efficient algorithms based on a new loss estimator that takes advantage of the special structure of the problem, and show regret bounds order $\sqrt{T}$ with respect to the time horizon. The bounds demonstrate polynomial scaling with the relevant problem parameters which is shown to be nearly optimal. The theoretical results are complemented by a set of experiments on simulated data.

1 INTRODUCTION

The theory of multi-armed bandits has inspired several practical applications and extensions to the basic setup (Lattimore and Szepesvári, 2020). The two most fundamental extensions to the standard multi-armed bandit setup are *contextual bandits*, which allow taking contextual information into account during decision making, and *combinatorial bandits*, which allow the formulation of large-scale decision making problems with combinatorial decision spaces. Both of these aspects are important to handle in the key application areas of bandit algorithms, including online advertising and recommendation systems (Li et al., 2010), and sequential treatment allocation (Tewari and Murphy, 2017). For instance, recommendation systems often need to produce structured lists of recommendations (providing a combinatorial aspect), while also taking into account the unique preferences of the user (providing a contextual aspect). The two aspects have been successfully addressed in the framework of *contextual combinatorial bandits* in a sequence of works initiated by Qin et al. (2014). All of these previous works have focused on the relatively simple scenario where the losses associated with the learner's actions are generated independently at random from a fixed distribution throughout the decision-making process. In the present work, we study the *nonstochastic* version of the contextual combinatorial bandit problem, where the sequence of losses incurred by the learning agent does not necessarily come from a fixed distribution, but can be possibly influenced by an external (even malicious) force. Since the real world is rarely stationary, this extension is of key practical importance as it significantly broadens the scope of the existing theory.

The setting we consider unifies many previous problem settings, and presents a new level of challenges that have not been encountered in previous work. In particular, handling the nonstochastic setting requires a drastically different set of tools than needed in the i.i.d. case considered in all past work on contextual combinatorial bandits: while all known approaches in this latter scenario are based on the principle of optimism in the face of uncertainty (Auer, 2002; Auer et al., 2002a), this idea is known to fail when the losses can be generated by an adversarial external process—see, e.g., Cesa-Bianchi and Lugosi, 2006, Section 4.1. A natural alternative route that we follow in this paper is to adapt the classic Exp3 algorithm of Auer et al. (2002b) to deal with the potential nonstationarity of the losses via the use of an importance-weighted loss estimator. This method has been adapted to deal with combinatorial action spaces by Cesa-Bianchi and Lugosi (2012) and Audibert et al. (2014), and to deal with contextual information by Neu and Olkhovskaya (2020). Both of these exten-

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

sions are based on generalizing the standard scalar-valued importance-weighted estimator of Auer et al. (2002b) to be able to directly estimate an unknown vector-valued problem parameter. A direct combination of these techniques to tackle our problem is far from straightforward due to the fact that our scenario requires the estimation of a *matrix-valued* parameter. Our main contribution is addressing this challenge via designing a range of new estimation procedures suitable for estimating such parameter matrices based on limited observations, and using them in conjunction with online decision making algorithms.

Following the terminology of Audibert et al. (2014), we consider two different feedback models: *semi-bandit* feedback, where the learner gets to observe the feedback associated with each component of its combinatorial action, and *full-bandit* feedback, where the learner only observes the total loss associated with its decision (for formal definitions, see Section 2). For both of these scenarios, we design new loss estimators based on the geometric resampling method proposed by Neu and Olkhovskaya (2020), which itself is a generalization of the geometric resampling method of Neu and Bartók (2013, 2016). The most challenging full-bandit scenario requires rather sophisticated treatment: here, our estimators are based on calculating and manipulating certain linear operators over matrices, which we represent by tensors of appropriate size. The estimation procedure used in this setting as well as the control over the resulting estimator is probably the most advanced technical tool we develop in this paper.

The concrete results we achieve in this work are the following. We suppose that the context vectors are d-dimensional, that the actions can be represented by K-dimensional binary vectors with at most m components being equal to 1, and that the losses suffered by the learner are linear in both the contexts and the actions, parametrized by a $K \times d$ matrix specifying the loss function. In this setting, we prove regret bounds of order $\sqrt{mKT \max\{d, m/\lambda_{\min}\}}$ in the semi-bandit setting and $m^{3/2}\sqrt{KT \max\{d, m/\lambda_{\min}\}}$ in the full-bandit setting (neglecting minor logarithmic factors). Here, $\lambda_{\min}$ is a lower bound on the smallest eigenvalue of the covariance matrix of the contexts. These bounds are achieved by combining the estimators mentioned in the previous paragraph with appropriately chosen extensions of the classic Exp3 and FTRL algorithms adapted to the combinatorial setting (Cesa-Bianchi and Lugosi, 2012; Audibert et al., 2014). The best known results are recovered for both adversarial contextual bandits[1] when $m = 1$ (Neu and Olkhovskaya, 2020) and for combinatorial bandits and semi-bandits when $d = 1$ (Audibert et al., 2014). Our algorithms can be implemented with polynomial runtime whenever the decision space allows an efficient implementation of the FTRL/Exp3 variants our methods are based on—more details are given in Sections 3 and 4 presenting the two methods.

Similar results have been achieved previously for the simpler i.i.d. setting. Qin et al. (2014) consider a scenario where the loss function is determined by a single d-dimensional parameter vector and the context can be represented by a $K \times d$ matrix. They propose an algorithm based on the principle of optimism in the face of uncertainty and achieve a regret guarantee of order $d\sqrt{mT \log(KT)}$ for this setting[2]. Similar results have been achieved by Li et al. (2016) (and a sequence of follow-up works) who consider a slightly different observation model generalizing semi-bandit feedback. On the side of non-stochastic losses, the only relevant works we are aware of are those of Kale et al. (2010) and Krishnamurthy et al. (2016), who both consider the semi-bandit setting with loss functions that are potentially non-linear with respect to the contexts, but are restricted to work with a finite policy class that maps contexts to combinatorial actions. A naïve instantiation of their bounds roughly[3] results in a regret bound of order $K\sqrt{dmT \log(T)}$. Implementing these latter algorithms requires either a full enumeration of the exponentially-sized policy space or access to a non-standard optimization oracle. In comparison, the computational steps required by our algorithms are relatively standard, and our methods can be implemented efficiently in a range of practically interesting problem settings.

The rest of the paper is organised as follows. In Section 2 we formally introduce the setting and corresponding assumptions. In Section 3 we the algorithm and analysis of the algorithm for the semi-bandit setting and in Section 4 we do the same for the full-bandit setting. In Section 5 we provide lower bounds and finally, in Section 6 we empirically evaluate our algorithms.

2 PRELIMINARIES

As outlined in the introduction, we are considering a non-stochastic bandit problem with combinatorial actions and contexts provided in each timestep. Given an action set $\mathcal{A} \subseteq \{0,1\}^K$, a context space $\mathcal{X} \subseteq \mathbb{R}^d$, and a distribution $\mathcal{D}$ over $\mathcal{X}$, our learning protocol can be described as

[1] The bounds stated by Neu and Olkhovskaya (2020) do not explicitly feature the $1/\lambda_{\min}$ factor, although a careful inspection of their proofs reveal that their bounds should indeed increase with this quantity.

[2] Their dependence on K is much milder due to the number of parameters to estimate being only d as opposed to Kd in our setting. The dependence on $\sqrt{m}$ we claim here follows from instantiating their bound with $C = m$ which is required when considering linear losses. Their bounds actually hold with slightly greater generality, allowing generalized linear loss functions.

[3] This follows from discretizing the space of loss matrices at a resolution of order $1/T$, and considering the class of greedy policies with respect to this cover. Details of how such an argument can be fully worked out are non-trivial, and a fully rigorous argument may likely lead to a worse regret bound.

follows. In each round t:

1. The (nonoblivious) adversary picks a loss matrix $\Theta_t \in \mathbb{R}^{d\times K}$
2. The environment draws an independent context vector $X_t \in \mathcal{X}$ from distribution $\mathcal{D}$
3. The learner observes X_t and picks an action from the action set $A_t \in \mathcal{A} \subseteq \{0,1\}^K$
4. The learner incurs the loss $X_t^\top \Theta_t A_t$
5. The learner observes:
   full-bandit: $X_t^\top \Theta_t A_t$
   semi-bandit: $X_t^\top \Theta_t \odot A_t$

where $\odot$ is the elementwise multiplication of two vectors. Without loss of generality, the matrix $\mathcal{A}$ is full rank.

Additional notation. We denote by $\mathbb{E}_X[\cdot]$ the expectation over random variable X. We denote by $\mathbb{E}_X[\cdot|Y]$ the expectation over random variable X conditioned on Y. When we write $\mathbb{E}[X]$ ($\mathbb{E}[X|Y]$) we take the expectation (conditioned on Y) with respect to all sources of randomness in X. Furthermore we will define the filtration $\mathcal{F}_t = \sigma(X_1, \xi_1, A_1, \ldots, X_t, \xi_t, A_t)$, where ξ_t captures any randomness employed by the learner in timesteps up to and including t. Each element A of the action set $\mathcal{A}$ is called an action. Each one of the K dimensions that make up the action set is called a sub-action, and a single action A has at most m active sub-actions, $\left\|A\right\|_1 \leq m$. We use $\lambda_{\min}(\cdot)$ to denote the smallest eigenvalue of a matrix or tensor, $\left\|\cdot\right\|_{\text{op}}$ to denote the operator norm of a matrix given by the largest eigenvalue $\lambda_{\max}(\cdot)$ of that matrix, and $\boldsymbol{e}_i$ to denote the basis vector in direction i.

Our results rely on the following assumptions that are rather standard in the combinatorial and contextual bandit literature.

- The distribution $\mathcal{D}$ from which the contexts are independently drawn is known and satisfies $\mathbb{E}[XX^\top] = \Sigma \succ 0I$;
- there exists a $\sigma > 0$ such that $\left\|X\right\|_2 \leq \sigma$ holds $\mathcal{D}$-almost surely;
- $\max_t \left\|\Theta_t\right\|_F \leq G$ for some $G > 0$, where $\left\|\cdot\right\|_F$ is the Frobenius norm;
- $\max_t \|(\Theta_t)_{\cdot,k}\|_2 \leq R$ for all $k \in K$, where $(\cdot)_{\cdot,k}$ is the $k-$th row of a matrix;
- $\max_t \max_i (|X^\top \Theta_t|)_i \leq 1$ holds $\mathcal{D}$-almost surely, where $(\cdot)_i$ is the $i-$th element of a vector.

Estimators and Matrix Geometric Resampling. In the paper we introduce two new unbiased estimators and two new biased estimators. Both unbiased estimators require us to invert a matrix of an expectation of outer products. While computing the expectation explicitly is possible, it might be computationally prohibitive. In order to reduce the computational burden we use the Matrix Geometric Resampling (MGR) method of Neu and Olkhovskaya (2020) to obtain a biased estimate of the inverse. The definition of MGR can be found in Algorithm 1, which is a more general version than MGR as introduced by Neu and Olkhovskaya (2020). The algorithm needs to be supplied by a sampling scheme and will output an estimate of the inverse of the expected matrix. Throughout the paper we will repeatedly make use of Lemma 1 below, which contains several crucial properties of MGR. The proof of Lemma 1 can be found in Appendix A.

Algorithm 1 MGR

Require: Sampling Scheme S, $\beta > 0$, $M > 0$
1: **for** $k = 1, \ldots, M$: **do**
2: Draw $\widehat{P}_k$ according to S
3: Compute $C_k = \prod_{j=1}^{k} (I - \beta \widehat{P}_j)$
4: **end for**
5: **Output** $\widehat{P}^+ = \beta \sum_{k=0}^{M} C_k$

Lemma 1. Let $\widehat{P}^+$ be defined by the MGR procedure (Algorithm 1) run for M iterations where each $\widehat{P}_k \in \mathbb{R}^{b\times b}$ drawn in Step 2 of Algorithm 1 is symmetric, positive semi-definite, and such that $\mathbb{E}[\widehat{P}_k] = P$, where P is also symmetric and positive semi-definite. Choose $\beta \leq \frac{1}{\lambda_{\max}(P)}$, then

$$\operatorname{tr}\left(\mathbb{E}_{\text{MGR}}[P\widehat{P}^{+\top} P \widehat{P}^+]\right) < 2b$$
$$\mathbb{E}_{\text{MGR}}[\widehat{P}^+]P = I - (I - \beta P)^M$$
$$\left\|\widehat{P}^+\right\|_{\text{op}} \leq (M+1)\beta\,.$$

Computational efficiency of MGR. If i.i.d. samples from the context distribution $\mathcal{D}$ and from the distibution $\pi_t(\cdot|X)$ over $\mathcal{A}$ are both available through sampling oracles, then MGR can be run in time of order $MKd + Kd^2$ (Neu and Olkhovskaya, 2020), where we assume both oracles can return a random draw from their corresponding distributions in unit time. Conditions on $\mathcal{A}$ enabling an efficient implementation of the sampling oracle for $\pi_t(\cdot|X)$ are discussed in Sections 3 and 4.

Regret Decomposition. The regret in our setting is defined by the best context-to-action mapping $\pi : \mathcal{X} \to \mathcal{A}$ in hindsight

$$\mathcal{R}_T = \max_{\pi\in\Pi} \mathbb{E}\left[\sum_{t=1}^{T} \left(X_t^\top \Theta_t A_t - X_t^\top \Theta_t \pi(X_t)\right)\right]$$

Now we can define the regret $\widehat{\mathcal{R}}_T(x)$ that the algorithm incurs at any context $x \in \mathcal{X}$ by using an unbiased estimator

$\widehat{\Theta}_t$ as follows

$$\widehat{\mathcal{R}}_T(x) = \mathbb{E}\left[\sum_{t=1}^{T}\left(x^\top \widehat{\Theta}_t A_t - x^\top \widehat{\Theta}_t \pi_T^*(x)\right)\right] \quad (1)$$
$$\pi_T^*(x) = \min_{A\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\left(x^\top \Theta_t A\right)\right]$$

Borrowing an idea from Neu and Olkhovskaya (2020), we introduce a ghost context $X_0 \sim \mathcal{D}$ drawn independently of all other contexts X_t. A crucial property of the regret is stated in Lemma 2 below.

Lemma 2 (Neu and Olkhovskaya (2020, Equation (6))). Let $\widetilde{\Theta}_t$ be some estimator of Θ_t with bias $B_t = \Theta_t - \widetilde{\Theta}_t$, then for any $X_0 \sim \mathcal{D}$

$$\mathcal{R}_T \le \mathbb{E}_{X_0}\left[\widehat{\mathcal{R}}_T(X_0)\right] + 2\mathbb{E}\left[\sum_{t=1}^{T}\max_{A\in\mathcal{A}}\left|\mathbb{E}\left[X_0^\top B_t A \middle| \mathcal{F}_{t-1}\right]\right|\right]$$

Because of Lemma 2 we only need to control the regret of any algorithm against a fixed context $x \in \mathcal{X}$ and any bias introduced by MGR. We will repeatedly use this fact throughout the paper, both for the upper and lower bounds we provide. Crucially, X_0 is independent from $X_1, \dots, X_T$ used to construct the loss estimators.

3 SEMI-BANDIT SETTING

In this section we introduce CO_2-FTRL (Algorithm 2), our algorithm for the contextual combinatorial setting with semi-bandit feedback. CO_2-FTRL is a instance of Follow The Regularized Leader run with a suitable estimator.

Recall that in the semi-bandit setting we are able to observe the losses of all the individual sub-actions by observing the vector $X_t^\top \Theta_t \odot A_t$. This allows us to construct an estimator of the loss matrix Θ_t as follows. The columns of estimator $\widehat{\Theta}_t \in \mathbb{R}^{d\times K}$ are defined to be

$$(\widehat{\Theta}_t)_{\cdot,k} = \Sigma_{t,k}^{-1} X_t (X_t^\top \Theta_t)_k (A_t)_k \quad (2)$$

where $\Sigma_{t,k} = \mathbb{E}_{A_t,X}\left[(A_t)_k X X^\top \mid \mathcal{F}_{t-1}\right]$. A crucial property is that the columns of $\widehat{\Theta}_t$ are unbiased estimators of the columns of Θ_t:

$$\mathbb{E}_{A_t X_t}\left[(\widehat{\Theta}_t)_{\cdot,k} \mid \mathcal{F}_{t-1}\right] = \Sigma_{t,k}^{-1}\,\mathbb{E}_{A_t,X_t}\left[X_t X_t^\top (A_t)_k \mid \mathcal{F}_{t-1}\right](\Theta_t)_{\cdot,k} = (\Theta_t)_{\cdot,k}$$

where the last equality follows from the definition of $\Sigma_{t,k}$.

However, note that $\widehat{\Theta}_t$ requires us to compute the inverse of the $d \times d$ covariance matrix $\Sigma_{t,k}$ K times, which can be computationally intensive. As discussed in Section 2, we make use of the MGR method (Neu and Olkhovskaya, 2020) to construct a potentially computationally cheaper estimator of $\Sigma_{t,k}$. We denote the estimated version of $\Sigma_{t,k}^{-1}$ by $\widehat{\Sigma}_{t,k}^{+} = \mathrm{MGR}(S(\mathcal{D}, \pi_t, k), \beta, M)$, where S is Sampling Scheme 4 defined in Appendix B. The corresponding estimator of Θ_t is denoted by $\widetilde{\Theta}_t$ and has columns

$$(\widetilde{\Theta}_t)_{\cdot,k} = \widehat{\Sigma}_{t,k}^{+} X_t (X_t^\top \Theta_t)_k (A_t)_k \,. \quad (3)$$

An apparent drawback of using the above estimator rather than the estimator defined in (2) is that the estimator (3) is biased. Fortunately, Lemma 1 gives us the tools to control the bias. In particular, by using the fact that $\widehat{P}_k$ is unbiased, we can see that

$$\mathbb{E}\left[\Theta_t - \widetilde{\Theta}_t \mid \mathcal{F}_{t-1}\right] = (I - \beta\Sigma_{t,k})^M \,.$$

Now, by Hölder's inequality and our assumptions on $\mathcal{X}$, Θ_t, and $\mathcal{A}$, we have that for any $A \in \mathcal{A}$ and any $x \in \mathcal{X}$

$$\mathbb{E}\left[x^\top(\Theta_t - \widetilde{\Theta}_t)A \mid \mathcal{F}_{t-1}\right] \le R\sqrt{m}\sigma \left\|(I - \beta\Sigma_{t,k})^M\right\|_{\mathrm{op}}$$

Since $\Sigma_{t,k} \succeq I\lambda_{\min}(\Sigma_{t,k})$, we have that

$$\|(I - \beta\Sigma_{t,k})^M\|_{\mathrm{op}} \le \left(1 - \beta\lambda_{\min}(\Sigma_{t,k})\right)^M$$

and so, using $1 + x \le \exp(x)$, the bias can be bounded by $R\sqrt{m}\sigma \exp\left(-M\beta\lambda_{\min}(\Sigma_{t,k})\right)$.

To control the bias of the estimator, we thus need to control $\lambda_{\min}(\Sigma_{t,k})$. Our solution is quite straightforward. We first construct an exploration set $E \subseteq \mathcal{A}$ such that there is at least one $A \in E$ satisfying $(A)_k = 1$ for each $k \in [K]$. To ensure that E always exists, we assume that for each $k \in K$ there exists at least one $A \in \mathcal{A}$ with $(A)_k = 1$. If this is not the case, one can trivially reduce the problem to a lower dimension. Given context X, our predictions are sampled from $\pi_t(\cdot|X)$, defined in Step 5 of Algorithm 2, which is a mixture with parameter $\gamma \in (0,1)$ between distribution $p_t(\cdot|X)$ over $\mathcal{A}$ and the uniform distribution over E. Since this π_t guarantees that $\mathbb{P}_t\big((A_t)_k = 1\big) \ge \gamma|E|^{-1}$, it is straightforward to see that $\lambda_{\min}(\Sigma_{t,k}) \ge \gamma\lambda_{\min}(\Sigma)|E|^{-1}$. The formal result can be found in Lemma 3 and its proof in Appendix B.

Lemma 3. Let $\beta \le \min_{t,k} \frac{1}{\lambda_{\max}(\Sigma_{t,k})}$. For any $A \in \mathcal{A}$ and all $x \in \mathcal{X}$ Algorithm 2 guarantees

$$\mathbb{E}\left[x^\top(\widehat{\Theta}_t - \widetilde{\Theta}_t)A \,\middle|\, \mathcal{F}_{t-1}\right] \le R\sqrt{m}\sigma\, e^{-\frac{M\beta\gamma}{|E|}\lambda_{\min}(\Sigma)}$$

simultaneously for all $t = 1, \dots, T$.

We now specify distribution $p_t(\cdot|X_t)$, which is inspired by the non-contextual algorithms of Koolen et al. (2010) and Audibert et al. (2014). First we compute $\bar{A}_t(X_t)$, defined in Step 3 of Algorithm 2, which is the prediction of FTRL on the convex hull of $\mathcal{A}$ with cumulative loss estimate $X_t \sum_{s=1}^{t-1} \widetilde{\Theta}_s A$. Distribution $p_t(\cdot|X_t)$ is then chosen

such that $\mathbb{E}_{A\sim p_t(\cdot|X_t)}[A] = \bar{A}_t(X_t)$. Since any A sampled from $p_t(\cdot|X_t)$ is equal to $\bar{A}_t(X_t)$ in expectation, we can analyse the algorithm as if the predictions are $\bar{A}_t(X_t)$.

The last missing piece is the FTRL regularizer φ. We use the unnormalized negative entropy (Koolen et al., 2010; Audibert et al., 2014) defined as

$$\varphi(A) = \frac{1}{\eta}\sum_{k=1}^{K}\big((A)_k \ln(A)_k - (A)_k\big). \tag{4}$$

The next result is a regret bound for the predictions $\bar{A}_t(X_t)$. The proof can be found in Appendix B and follows from standard arguments.

Lemma 4. Let $\beta = \frac{1}{\sigma^2}$ and let $\eta \le \frac{\ln(2)}{M+1}$. For any $x \in \mathcal{X}$, $\bar{A}_t(x)$ defined by (5) and any $u \in \mathcal{A}$ with φ as in equation (4) guarantees

$$\sum_{t=1}^{T} x^\top \widetilde{\Theta}_t(\bar{A}_t(x) - u) \le \frac{m\left(1+\ln\left(\frac{K}{m}\right)\right)}{\eta} + \eta\sum_{t=1}^{T}\sum_{k=1}^{K}(x^\top\widetilde{\Theta}_t)_k^2(\bar{A}_t(x))_k\,.$$

Next, we need to control the $\sum_{k=1}^{K}(x^\top\widetilde{\Theta}_t)_k^2(\bar{A}_t(x))_k$ terms in Lemma 4. First we show that $\mathbb{E}\left[(x^\top\widetilde{\Theta}_t)_k^2(\bar{A}_t(x))_k \,\middle|\, \mathcal{F}_{t-1}\right]$ is upper bounded by $2\mathbb{E}\left[\mathrm{tr}\left(\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+,\top}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+}\right) \,\middle|\, \mathcal{F}_{t-1}\right]$, after which we can use Lemma 1 to control the bias. The result can be found in Lemma 5 and the complete proof in Appendix B.

Lemma 5. For any $\gamma \in (0,1)$ and for all $t \in [T]$ we have that

$$\mathbb{E}\left[\sum_{k=1}^{K}\left(X_0^\top\widetilde{\Theta}_t\right)_k^2\left(\bar{A}_t(X_0)\right)_k \,\middle|\, \mathcal{F}_{t-1}\right] \le \frac{3Kd}{1-\gamma}$$

By combining Lemmas 3, 4, 5 we arrive at the final regret bound in Theorem 6, whose proof is implied by Theorem 16 in Appendix B.

Theorem 6. *Algorithm 2 with appropriate tuning satisfies*

$$\mathcal{R}_T \in O\left(\sqrt{mKT\left(1+\ln\frac{K}{m}\right)\max\left\{d, \frac{m\sigma^2\ln(T)}{\lambda_{\min}(\Sigma)}\right\}}\right)$$

Computational efficiency. Given a sampling oracle for $\mathcal{D}$, the running time of CO$_2$-FTRL on a pair $(\mathcal{D}, \mathcal{A})$ is essentially the same as the running time of the OSMD algorithm on $\mathcal{A}$. Audibert et al. (2014, Section 2) discuss the conditions on $\mathcal{A}$ that allow an efficient implementation of OSMD.

Algorithm 2 CO$_2$-FTRL

Require: learning rate $\eta > 0$, exploration rate $\gamma \in (0,1)$
Require: exploration set $E \subseteq \mathcal{A}$
1: **for** t in $[T]$ **do**
2: Observe X_t
3: Compute

$$\bar{A}_t(X_t) = \operatorname*{arg\,min}_{A\in\mathrm{Conv}(\mathcal{A})}\sum_{s=1}^{t-1} X_t^\top\widetilde{\Theta}_s A + \varphi(A) \tag{5}$$

4: Find probability distribution $p_t(\cdot|X_t)$ such that $\mathbb{E}_{A\sim p_t(\cdot|X_t)}[A] = \bar{A}_t(X_t)$
5: Set

$$\pi_t(A|X) = (1-\gamma)p_t(A|X) + \gamma\frac{\mathbb{1}[A\in E]}{|E|} \tag{6}$$

6: Draw and play $A_t \sim \pi_t(\cdot|X_t)$
7: Observe loss $X_t^\top\Theta_t \odot A_t$ and compute $\widetilde{\Theta}_t$ using (3).
8: **end for**

4 FULL-BANDIT SETTING

In this section we describe our results in the full-bandit setting, where we only have access to the total loss $X_t^\top\Theta_t A_t$ incurred at each timestep t rather than the loss components $X_t^\top\Theta_t\odot A_t$ that we had access to in the semi-bandit setting.

We start by describing how to construct an unbiased estimator for Θ_t. In order to do so we need to introduce several definitions related to tensors.

Tensor definitions. For a more extensive background on tensors we refer the reader to Appendix D. Here we introduce the definitions necessary to understand the ideas behind our algorithm. Let $\Phi \in \mathbb{R}^{d\times d\times K\times K}$ be a tensor and let $\Theta \in \mathbb{R}^{d\times K}$ be a matrix. Tensor Φ acting on Θ is denoted by $\Phi(\Theta) = B$, where $B \in \mathbb{R}^{d\times K}$ is a matrix with elements

$$B_{i,k} = \sum_{a}^{d}\sum_{b}^{K}\Phi_{i,a,b,k}\Theta_{a,b}$$

The definition of a tensor acting on a tensor is given in Definition D.1 in Appendix D. Tensors are associative in the sense that $\Phi(\Psi(\Theta)) = \Phi(\Psi)(\Theta)$ where $\Psi \in \mathbb{R}^{d\times d\times K\times K}$, see Lemma 20.

The tensor product between matrices Θ and $\Theta' \in \mathbb{R}^{d\times K}$ is defined as $(\Theta\otimes\Theta') = W \in \mathbb{R}^{d\times d\times K\times K}$, which has elements $W_{i,i',k,k'} = \Theta_{i,k}\Theta'_{i',k'}$. We define an identity tensor $\mathcal{I}$, which satisfies $\mathcal{I}(\Theta) = \Theta$ for all Θ. If the inverse exists, then the inverse of tensor Φ is denoted by Φ^{-1}, which satisfies $\Phi^{-1}(\Phi) = \mathcal{I}$. The central equality, which we use in our novel estimator, can be found in Lemma 7 whose proof is in Appendix D.

Lemma 7. $DCB^\top = (D \otimes B)(C)$, where B, C, D are matrices of appropriate size.

With the above definitions and equalities, we are ready to construct our unbiased estimator defined as

$$\widehat{\Theta}_t = \Psi_t^{-1}\Big(X_t X_t^\top \Theta_t A_t A_t^\top\Big) \tag{7}$$

where $\Psi_t = \mathbb{E}_{X_t, A_t}\Big[\big(X_t X_t^\top \otimes A_t A_t^\top\big)\Big|\mathcal{F}_{t-1}\Big]$.

In Lemma 41 in Appendix E we show that the inverse of tensor Ψ_t indeed exists. To see that $\widehat{\Theta}_t$ is unbiased, we use Lemma 7 to see that, conditioned on $\mathcal{F}_{t-1}$,

$$\begin{aligned}\mathbb{E}_{X_t,A_t}\big[\widehat{\Theta}_t\big] &= \mathbb{E}_{X_t,A_t}\Big[\Psi_t^{-1}\big(X_t X_t^\top \Theta_t A_t A_t^\top\big)\Big] \\ &= \mathbb{E}_{X_t,A_t}\Big[\Psi_t^{-1}\big((X_t X_t^\top \otimes A_t A_t^\top)(\Theta_t)\big)\Big] \\ &= \Psi_t^{-1}\big(\Psi_t(\Theta_t)\big)\,.\end{aligned}$$

Now, using the associative property of tensors and the definition of inverses of tensors, we can see that

$$\Psi_t^{-1}\big(\Psi_t(\Theta_t)\big) = \Psi_t^{-1}\big(\Psi_t\big)(\Theta_t) = \mathcal{I}(\Theta) = \Theta.$$

As in the semi-bandit setting, we make use of the MGR algorithm (Algorithm 1) to construct a computationally more efficient version of the unbiased estimator. In particular, we aim at constructing Ψ_t^{-1} more efficiently. Unfortunately, the MGR algorithm as stated in Section 2 only works for matrices. To resolve this issue, we temporarily flatten the tensor to a matrix (Definition D.8). It turns out that the MGR algorithm applied to the flattened version of Ψ_t, which we denote by Ψ_t^F, returns a matrix which we can unflatten (Definition D.9). We denote by Θ^U the unflattening of a matrix Θ.

The estimator that makes use of the MGR algorithm to estimate Ψ_t^{-1} is defined by

$$\widetilde{\Theta}_t = \widehat{\Psi}_t^{+}\big(X_t X_t^\top \Theta_t A_t A_t^\top\big) \tag{8}$$

where $\widehat{\Psi}_t^{+} = \mathrm{MGR}(S(\mathcal{D}, \pi_t), \beta, M)^U$ and S is the Sampling Scheme 5, defined in Appendix E. The number M of iterations the MGR is run for and β are hyper-parameters of the algorithm set according to Theorem 43. Using the sampling scheme with the MGR like this means that we do not need to compute any expectation explicitly to run the algorithm.

As in the semi-bandit setting we need to ensure that the eigenvalues of the the tensor Ψ_t^{-1} are not too large, see Definition D.10 for the definition of eigenvalues of tensors. We ensure this by employing an exploration distribution μ over $\mathcal{A}$ based on the Kiefer-Wolfowitz theorem (see also Theorem 40 in the appendix). And the result is shown by arguing that the smallest eigenvalue of the flattened tensor Ψ_t^F is properly bounded, see Lemma 11 below.

The exploration scheme is mixed with a version of Exp3 (Auer et al., 2002b). In particular, we simply run Exp3 on all the actions in $\mathcal{A}$, an approach also used by Cesa-Bianchi and Lugosi (2012). The full algorithm is specified in Algorithm 3 and is aptly named Exp3-Tensor. Its regret bound can be found in Theorem 8, whose proof is implied by Theorem 43 in Appendix E.

Theorem 8. *Algorithm 3 with appropriate tuning satisfies*

$$\mathcal{R}_T \in O\left(m^{3/2}\sqrt{KT\ln(K)\max\Big\{d, \frac{m\sigma^2\ln(T)}{\lambda_{\min}(\Sigma)}\Big\}}\right).$$

In the remainder of this section we provide a sketch of the proof of Theorem 43.

As a first step, observe that we need to control the bias of our estimator: the $\mathbb{E}\big[X_0^\top B_t A \mid \mathcal{F}_{t-1}\big]$ term of Lemma 2. We do precisely this in the following Lemma.

Lemma 9. Suppose that $\beta \le \frac{1}{\lambda_{\max}(\Psi_t^F)}$. Then for $\widetilde{\Theta}_t$ defined in equation (8)

$$\mathbb{E}\big[X_0^\top(\Theta_t - \widetilde{\Theta}_t)A \mid \mathcal{F}_{t-1}\big] \le \sigma G\sqrt{me}^{-\frac{M\beta\gamma m}{K}\lambda_{\min}(\Sigma)}$$

The proof of Lemma 9 can be found in Appendix E. The proof is very similar to the proof of Lemma 3, with the main difference being the fact that we need to carefully track the effect of the flattening and unflattening operations.

Another part of the proof is bounding the regret of Exp3. The following result can be derived from the standard Exp3 analysis (Auer et al., 2002b) and is provided in Lemma 10 (recall the definition (1) of auxiliary game).

Lemma 10. *Fix any $x \in \mathcal{X}$ and suppose that $\widetilde{\Theta}_t$ and $\eta > 0$ are such that $\max_t \eta|x^\top \widetilde{\Theta}_t A| < 1$ for all $A \in \mathcal{A}$. Then the regret of Algorithm 3 in the auxiliary game at x satisfies*

$$\begin{aligned}\widehat{\mathcal{R}}_T(x) \le& \frac{\ln(|\mathcal{A}|)}{\eta} + \gamma\, U_T(x) \\ &+ \eta\mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_{A\sim\pi_t(\cdot|x)}\big[(x^\top\widetilde{\Theta}_t A)^2|\mathcal{F}_{t-1}\big]\right]\end{aligned}$$

where $U_T(x) = \sum_{t=1}^T \sum_{A\in\mathcal{A}} \mu_A x^\top \widetilde{\Theta}_t\big(A - \pi_T^*(x)\big)$ and μ is the distribution on $\mathcal{A}$ defined by the Kiefer-Wolfowitz theorem.

To ensure that we can apply Lemma 10, we only need to show that our learning rate is chosen correctly and that our estimator $\widetilde{\Theta}_t$ behaves nicely enough, which essentially boils down to controlling the smallest eigenvalue of the flattened tensor Ψ_t^F. This is shown in Lemma 11.

Lemma 11. *For all $t \ge 1$,*

$$\lambda_{\min}(\Psi_t^F) \ge \frac{\gamma K \lambda_{\min}(\Sigma)}{m}$$

Moreover, for $\eta \le \frac{1}{m(M+1)}$, any $A \in \mathcal{A}$, and any x in the support of $\mathcal{D}$ it also holds that $\eta\big|x^\top\widetilde{\Theta}_t A\big| < 1$.

While the regret of Exp3 is relatively straightforward to control with the standard importance weighted estimator, here we face a complicated term $\mathbb{E}_{A\sim\pi_t(\cdot,x)}\left[(x^\top\widetilde{\Theta}_t A)^2|\mathcal{F}_{t-1}\right]$ due to our choice of estimator $\widetilde{\Theta}_t$. Not only is $\widetilde{\Theta}_t$ biased, but we also need to significantly manipulate the $(x^\top\widetilde{\Theta}_t A)^2$ term in order to recover an expression resembling the term we would have if we were to use the standard importance weighted estimator. We do so in Appendix E, which leads to the following result.

Lemma 12. *Fix a $t \in [T]$ and let $A_0 \sim \pi_t(\cdot|X_0)$. Then*

$$\mathbb{E}\left[(X_0^\top\widetilde{\Theta}_t A_0)^2 \,\middle|\, \mathcal{F}_{t-1}\right] \le 2m^2Kd$$

To finish the proof of Theorem 8, we only need to assemble the pieces we have collected so far. Applying Lemma 12 to the right-hand side of Lemma 10 gives.

$$\mathbb{E}_{X_0}[\widehat{\mathcal{R}}_T(X_0)] \le \frac{\ln(|\mathcal{A}|)}{\eta} + \eta T m^2 Kd + \gamma U_T(X_0)$$

The remaining steps are applying this result to Lemma 2, applying Lemma 9, and tuning η, γ, M and β accordingly. All details can be found in Theorem 43 in Appendix E.

Computational efficiency. The crucial steps in Exp3-Tensor are the computation of μ via the Kiefer-Wolfowitz theorem and the computation of (9) and (10). Computing μ exactly is not efficient in general. However, there are efficient algorithms that compute approximations to μ (Lattimore and Szepesvári, 2020, Section 21.2, Note 3). Using this approximation to μ does not deteriorate the order of regret. The running time for executing the remaining steps is essentially equivalent to the time it takes to run the corresponding steps in CombBand plus the runtime of the MGR procedure. Cesa-Bianchi and Lugosi (2012) show various concrete examples of action sets $\mathcal{A}$ on which ComBand can be run efficiently. Since the MGR procedure can be run efficiently, this implies that Exp3-Tensor is efficient on a pair $(\mathcal{D}, \mathcal{A})$ whenever a sampling oracle for $\mathcal{D}$ is available and CombBand can be run efficiently on $\mathcal{A}$.

5 LOWER BOUNDS

In this section we provide lower bounds for both the full- and semi-bandit settings. All details related to the results in this section can be found in Appendix C. Our lower bounds hold for a large class of algorithms which we call orthogonal algorithms. Informally, if two contexts x_s and x_t are orthogonal, then orthogonal algorithms do not use information from round $s < t$ to compute a prediction for round t. Essentially all algorithms using the estimators in Sections 3 and 4 are orthogonal algorithms. The lower bound for the semi-bandit setting is provided below.

Algorithm 3 Exp3-Tensor

Require: $\eta > 0, \gamma \in (0,1), M > 0, \beta > 0$

1: Find probability distribution μ over $\mathcal{A}$ as defined by the Kiefer-Wolfowitz theorem (Theorem 40)
2: **for** t in $[T]$ **do**
3: Observe X_t and for all $A \in \mathcal{A}$ set

$$w_t(X_t, A) = \exp\left(-\eta \sum_{s=1}^{t-1} X_t^\top \widetilde{\Theta}_s A\right) \quad (9)$$

4: Draw A_t from

$$\pi_t(A|X_t) = \frac{(1-\gamma)w_t(X_t, A)}{\sum_{A'\in\mathcal{A}} w_t(X_t, A')} + \gamma\mu_A \quad (10)$$

5: Observe loss $X_t^\top \Theta_t A$ and compute $\widetilde{\Theta}_t$ using (8) and Sampling Scheme 5 in Appendix E.
6: **end for**

Theorem 13. *In the semi-bandit setting, any orthogonal algorithm must suffer $\Omega(\sqrt{dmKT})$ regret.*

Theorem 13 is implied by Theorem 18, whose proof follows from a reduction to online learning with stochastic feedback graphs (Esposito et al., 2022).

For the full-bandit setting the result can be found below. The result is implied by Theorem 19, whose proof follows from carefully adapting the lower bound of Cohen et al. (2017) to work in our setting.

Theorem 14. *In the full-bandit setting, any orthogonal algorithm must suffer $\widetilde{\Omega}(m^{3/2}\sqrt{dKT})$ regret.*

Ignoring factors logarithmic in K, our upper bounds in both settings have an extra $\max\left\{d, \frac{m\sigma^2\ln(T)}{\lambda_{\min}(\Sigma)}\right\}$ factor. Although the term d is captured by our lower bounds, the $\frac{m\sigma^2\ln(T)}{\lambda_{\min}(\Sigma)}$ term is not. In particular, in the construction of our lower bounds $\sigma = 1$ and $\lambda_{\min}(\Sigma) = 1/d$. Thus, with the same set of losses as in the lower bound, our algorithms suffer $\widetilde{O}(m^2\sqrt{dKT})$ regret in the full-bandit setting and $\widetilde{O}(m\sqrt{dKT})$ regret in the semi-bandit setting. This implies that for $m = 1$ our algorithms as well as the algorithm of Neu and Olkhovskaya (2020) have a tight regret bound. However, a gap of $\sqrt{m}$ exists when m is a parameter of the problem.

6 EXPERIMENTS

The full code for the experiments can be found here[4].

To the best of our knowledge, our work is the first one in this setting, and thus there are no natural strong base-

[4] https://github.com/LukasZierahn/Combinatorial-Contextual-Bandits

lines to compare against in experiments. Our baselines are RealLinExp3 (Neu and Olkhovskaya, 2020) and two versions of CombBand (Cesa-Bianchi and Lugosi, 2012). RealLinExp3 uses contextual information but ignores any structure in the action set, implying that each action is treated independently from the others. The first version of ComBand that we compare with ignores contexts but is able to exploit the combinatorial nature of actions. Although our algorithms can handle arbitrary context distributions (no expectation needs to be computed thanks to the MGR procedure), to accommodate the reasonable baseline of running one instance of ComBand algorithm per context, we run our experiments on a finitely supported distribution $\mathcal{D}$.

Let $\mathcal{B}_{K,m} = \{x : x \in \{0,1\}^K, \|x\|_1 = m\}$ be the set containing all m-sized subsets of a base set of K elements. We use $\mathcal{B}_{d,m}$ with $(d,m) \in \{(3,1),(5,2),(12,3)\}$ to define context spaces $\mathcal{X}$, and $\mathcal{B}_{K,m}$ with $(K,m) \in \{(3,1),(5,2),(8,3)\}$ to define action sets $\mathcal{A}$. Our experiments are run over a length of 10^5 timesteps and averaged over 10 repetitions. The losses Θ_t are generated as follows: for each $i \in [d]$ we choose m subactions that are "good actions". This means that $(\Theta_t)_{i,j}$ is drawn from a Bernoulli(0.4) distribution if j is a good action in i and from a Bernoulli(0.5) distribution otherwise. The contexts are drawn uniformly at random.

In our experiments, we did not use the MGR procedure for any of the algorithms. While MGR enjoys a $O(MKd + Kd^2)$ scaling in computational cost (Neu and Olkhovskaya, 2020), due to the nature of the context distribution using the unbiased estimator turns out to be significantly faster, as the M factor in the scaling of the runtime of MGR is of order $1/\gamma$ and there exists a highly optimised implementation of matrix inversion in NumPy (Harris et al., 2020). We stress that this direct computation is only possible due to the simple setting and if the expectation cannot be computed then the MGR is necessary.

We run experiments with Exp3-Tensor with theoretical tuning, and using uniform random exploration instead of the Kiefer-Wolfowitz based exploration to ease implementation. All the baseline algorithm also use theoretical tuning. While the exploration rate for Exp3-Tensor is very large (equal to $\gamma = 51.66\%$ in the case where $d = 12$ and $K = 8$), the algorithm is only marginally falling behind the other algorithms (see Figure 6 and the figures in Appendix F) with exploration rate of 1.32% for RealLinExp3, 5.38% for Comband, and 18.64% for the Comband One-Per-Context. We conjecture that the theoretical tuning might be too restrictive in these artificial scenarios with a finite number of contexts, which we designed to accommodate our baselines. More experiments, using more general context spaces and improved tunings, are necessary to gain a better understanding of the algorithms' behaviour.

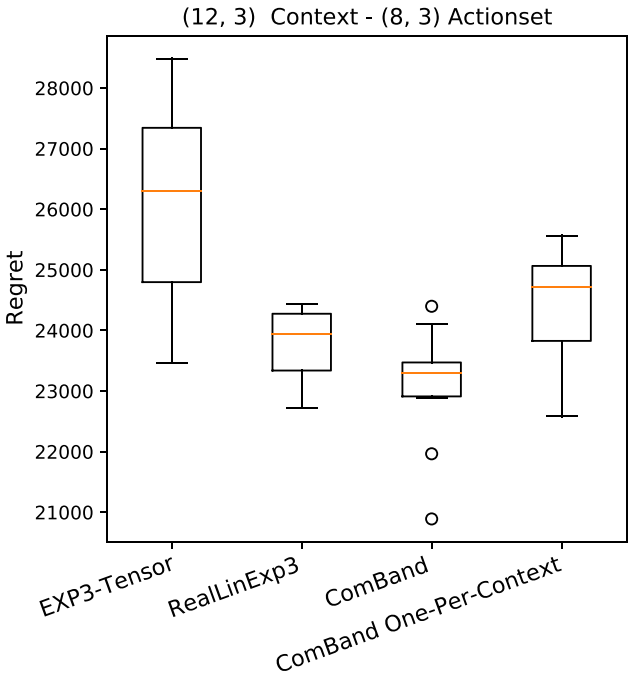

Figure 1: Boxplots over 10 repetitions showing the regret of the four algorithms (lower is better).

7 DISCUSSION

As already discussed in Section 5, our bounds are tight with respect to all parameters except m. More precisely, there is an extra factor $\sqrt{m}$ in the upper bounds for both semi and full bandit settings. We leave open the question whether this extra term is necessary or not.

One direction for future work is to empirically evaluate our algorithm as well as the baselines we specified in Section 6 in more general experimental settings. The experiments in Section 6 were specified to accommodate the baselines and it would be interesting to understand how all algorithms fare under different circumstances. One such circumstance could be replacing the discrete distribution over context in our experiments with a continuous distribution, even though it is not clear how to accommodate all baselines for such a context distribution.

Acknowledgements

This work was mostly done while DvdH was at the University of Milan partially supported by the MIUR PRIN grant Algorithms, Games, and Digital Markets (ALGADIMAR) and partially supported by Netherlands Organization for Scientific Research (NWO), grant number VI.Vidi.192.095. LZ and NCB were partially supported by the EU Horizon 2020 ICT-48 research and in-

novation action under grant agreement 951847, project ELISE (European Learning and Intelligent Systems Excellence) and the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme. GN has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 950180).

References

Audibert, J.-Y., Bubeck, S., and Lugosi, G. (2014). Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45.

Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002a). Finite-time analysis of the multiarmed bandit problem. *Machine Learning Journal*, 47(2-3):235–256.

Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. (2002b). The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77.

Bubeck, S., Cesa-Bianchi, N., and Kakade, S. M. (2012). Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41.1–41.14.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA.

Cesa-Bianchi, N. and Lugosi, G. (2012). Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422. JCSS Special Issue: Cloud Computing 2011.

Cohen, A., Hazan, T., and Koren, T. (2017). Tight bounds for bandit combinatorial optimization. In *Conference on Learning Theory*, pages 629–642.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354(7):769–822.

Esposito, E., Fusco, F., Van der Hoeven, D., and Cesa-Bianchi, N. (2022). Learning on the edge: Online learning with stochastic feedback graphs. *arXiv preprint arXiv:2210.04229*.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.

Kale, S., Reyzin, L., and Schapire, R. E. (2010). Non-stochastic bandit slate problems. *Advances in Neural Information Processing Systems*, 23.

Knuth, D. E. (1997). *The Art of Computer Programming: Volume 1: Fundamental Algorithms (3rd ed.)*. Addison-Wesley.

Koolen, W. M., Warmuth, M. K., Kivinen, J., et al. (2010). Hedging structured concepts. In *Conference on Learning Theory*, pages 93–105.

Krishnamurthy, A., Agarwal, A., and Dudik, M. (2016). Contextual semibandits via supervised learning oracles. *Advances in Neural Information Processing Systems*, 29.

Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *International World Wide Web Conference*, pages 661–670.

Li, S., Wang, B., Zhang, S., and Chen, W. (2016). Contextual combinatorial cascading bandits. In *International Conference on Machine Learning*, pages 1245–1253.

Neu, G. and Bartók, G. (2013). An efficient algorithm for learning with semi-bandit feedback. In *International Conference on Algorithmic Learning Theory*, pages 234–248.

Neu, G. and Bartók, G. (2016). Importance weighting without importance weights: An efficient algorithm for combinatorial semi-bandits. *Journal of Machine Learning Research*, 17:1–21.

Neu, G. and Olkhovskaya, J. (2020). Efficient and robust algorithms for adversarial linear contextual bandits. In *Conference on Learning Theory*, pages 3049–3068.

Orabona, F. (2019). A modern introduction to online learning. *CoRR*, abs/1912.13213.

Qin, L., Chen, S., and Zhu, X. (2014). Contextual combinatorial bandit and its application on diversified online recommendation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 461–469.

Tewari, A. and Murphy, S. A. (2017). From ads to interventions: Contextual bandits in mobile health. In *Mobile Health - Sensors, Analytic Methods, and Applications*, pages 495–517.

Van der Hoeven, D., Van Erven, T., and Kotłowski, W. (2018). The many faces of exponential weights in online learning. In *Conference on Learning Theory*, pages 2067–2092.

Warmuth, M. K. and Kuzmin, D. (2008). Randomized online pca algorithms with regret bounds that are logarithmic in the dimension. *Journal of Machine Learning Research*, 9(75):2287–2320.

A DETAILS OF SECTION 2 (PRELIMINARIES)

Lemma 1. Let $\widehat{P}^+$ be defined by the MGR procedure (Algorithm 1) run for M iterations where each $\widehat{P}_k \in \mathbb{R}^{b\times b}$ drawn in Step 2 of Algorithm 1 is symmetric, positive semi-definite, and such that $\mathbb{E}[\widehat{P}_k] = P$, where P is also symmetric and positive semi-definite. Choose $\beta \leq \frac{1}{\lambda_{\max}(P)}$, then

$$\operatorname{tr}\left(\mathbb{E}_{\mathrm{MGR}}[P\widehat{P}^{+\top}P\widehat{P}^{+}]\right) < 2b$$
$$\mathbb{E}_{\mathrm{MGR}}[\widehat{P}^{+}]P = I - (I-\beta P)^M$$
$$\left\|\widehat{P}^{+}\right\|_{\mathrm{op}} \leq (M+1)\beta\,.$$

Proof. The second and third statement of the lemma are proven in Section 4.2 of Neu and Olkhovskaya (2020). While a similar statement to the first statement of the lemma can be found in Neu and Olkhovskaya (2020), there is a transpose missing from their statement compared to ours. In particular, since $\widehat{P}^+$ might not be symmetric we can not conclude that $\widehat{P}^{+\top} = \widehat{P}^+$. Therefore, we need to prove the first statement of the lemma.

A central part of our consideration of that will be $\widehat{P}^{+\top}$, which we explore now. For that we will employ the definitions of $\widehat{P}^{+\top}$ and C_k as given by the MGR procedure in Algorithm 1 and the fact that $(I-\beta\widehat{P}_k)$ is symmetric as $\widehat{P}_k$ is symmetric by assumption.

$$\begin{aligned}
\widehat{P}^{+\top} &= \left(\beta\sum_{k=0}^{M} C_k\right)^{\top} \\
&= \beta\sum_{k=0}^{M} C_k^{\top} \\
&= \beta\sum_{k=0}^{M}\left(\prod_{j=1}^{k}(I-\beta\widehat{P}_k)\right)^{\top} \\
&= \beta\sum_{k=0}^{M}\prod_{j=1}^{k}(I-\beta\widehat{P}_{k-j+1})^{\top} \\
&= \beta\sum_{k=0}^{M}\prod_{j=1}^{k}\underbrace{(I-\beta\widehat{P}_{k-j+1})}_{D_{k-j+1}},
\end{aligned}$$

where we also introduced the notation $D_j = (I-\beta\widehat{P}_j)$.

Now we are equipped to focus on $\widehat{P}^{+\top}P\widehat{P}^{+}$, which we do by using the above equation for $\widehat{P}^{+\top}$. We then multiply out to obtain

$$\begin{aligned}
\widehat{P}^{+\top}P\widehat{P}^{+} &= \left(\beta\sum_{k=0}^{M}\prod_{j=1}^{k} D_{k-j+1}\right) P \left(\beta\sum_{k=0}^{M}\prod_{j=1}^{k} D_j\right) \\
&= \beta^2\sum_{k=0}^{M}\sum_{k'=0}^{M}\left(\prod_{j=1}^{k} D_{k-j+1}\right) P \left(\prod_{j=1}^{k'} D_j\right).
\end{aligned}$$

Throughout the remainder of the proof we will use that

$$\mathbb{E}\left[\left(\prod_{j=1}^{k} D_{k_{\min}-j+1}\right) P \left(\prod_{j=1}^{k'} D_j\right)\right] = \mathbb{E}\left[D_{k'}D_{k'-1}\dots D_2 D_1 P D_1 D_2 \dots D_{k-1}D_k\right].$$

As a next step we take the expectation over $\widehat{P}_j$ for all $j \in [M]$. We start by looking at the individual terms of the sum. Define $k_{\min} = \min(k,k')$ and $k_{\max} = \max(k,k')$. For $k \leq k'$ we have that in the term $\left(\prod_{j=1}^{k'} D_j\right)$ there are $k'-k$ D_j

terms that do not appear in $\left(\prod_{j=1}^{k} D_{k-j+1}\right)$. Similarly, for $k \geq k'$ we have that in the term $\left(\prod_{j=1}^{k} D_{k-j+1}\right)$ there are $k-k$ D_j terms that do not appear in $\left(\prod_{j=1}^{k'} D_j\right)$. Therefore, by linearity of the expectation and the tower rule we have that

$$\begin{aligned}
&\mathbb{E}_{\mathrm{MGR}}\left[\widehat{P}^{+\top} P \widehat{P}^{+}\right] \\
&= \mathbb{E}_{\mathrm{MGR}}\left[\beta^2 \sum_{k=0}^{M}\sum_{k'=0}^{M}\left(\prod_{j=1}^{k} D_{k-j+1}\right) P \left(\prod_{j=1}^{k'} D_j\right)\right] \\
&= \beta^2 \sum_{k=0}^{M}\sum_{k'=0}^{M} \mathbb{E}_{\mathrm{MGR}}\left[\left(\prod_{j=1}^{k} D_{k-j+1}\right) P \left(\prod_{j=1}^{k'} D_j\right)\right] \\
&= \beta^2 \sum_{k=0}^{M}\sum_{k'=0}^{M} \mathbb{E}_{\widehat{P}_1,\ldots,\widehat{P}_{k_{\min}}}\left[\mathbb{E}_{\widehat{P}_{k_{\min}},\ldots,\widehat{P}_{k_{\max}}}\left[\left(\prod_{j=1}^{k} D_{k-j+1}\right) P \left(\prod_{j=1}^{k'} D_j\right) \middle| \widehat{P}_1,\ldots,\widehat{P}_{k_{\min}}\right]\right] \\
&= \beta^2 \sum_{k=0}^{M}\sum_{k'=0}^{M} \mathbb{E}_{\widehat{P}_1,\ldots,\widehat{P}_{k_{\min}}}\left[(I-\beta P)^{\max\{k_{\max}-k,0\}}\left(\prod_{j=1}^{k_{\min}} D_{k_{\min}-j+1}\right) P \left(\prod_{j=1}^{k_{\min}} D_j\right)(I-\beta P)^{\max\{k_{\max}-k',0\}}\right].
\end{aligned} \tag{11}$$

Fix some $k \in [M]$, now we will inspect $\mathbb{E}_{\widehat{P}_1,\ldots,\widehat{P}_k}\left[\left(\prod_{j=1}^{k_{\min}} D_{k_{\min}-j+1}\right) P \left(\prod_{j=1}^{k_{\min}} D_j\right)\right]$. Pick some $j \in [k]$, some matrix H that is commutative with P, i.e., $HP = PH$. By using $\widehat{P}_j \preceq \lambda_{\max}(\widehat{P}_j) I \preceq \beta^{-1} I$ we can see that

$$\begin{aligned}
\mathbb{E}_{\widehat{P}_j}[D_j H D_j] &= \mathbb{E}_{\widehat{P}_j}[(I-\beta\widehat{P}_j)H(I-\beta\widehat{P}_j)] \\
&= \mathbb{E}_{\widehat{P}_j}[H - \beta\widehat{P}_j H - \beta H\widehat{P}_j + \beta^2 \widehat{P}_j H \widehat{P}_j] \\
&\preceq \mathbb{E}_{\widehat{P}_j}[H - \beta\widehat{P}_j H - \beta H\widehat{P}_j + \beta H \widehat{P}_j] \\
&= H - \beta P H \\
&= (I-\beta P)H
\end{aligned}$$

It is also clear that if H and P commute then so do $H - \beta PH$ and P. Thus we can now use the above idea recursively $k_{\min}$ times in total, starting with $H = P$.

$$\mathbb{E}_{\widehat{P}_1,\ldots,\widehat{P}_k}\left[\left(\prod_{j=1}^{k_{\min}} D_{k_{\min}-j+1}\right) P \left(\prod_{j=1}^{k_{\min}} D_j\right)\right] \preceq P(I-\beta P)^{k_{\min}}$$

Plugging this into equation (11) to find

$$\begin{aligned}
&\mathbb{E}_{\mathrm{MGR}}\left[\widehat{P}^{+\top} P \widehat{P}^{+}\right] \\
&= \mathbb{E}_{\mathrm{MGR}}\left[\beta^2 \left(\prod_{j=1}^{k} D_{k-j+1}\right) P \left(\prod_{j=1}^{k'} D_j\right)\right] \\
&= \beta^2 \sum_{k=0}^{M}\sum_{k'=0}^{M} \mathbb{E}_{\widehat{P}_1,\ldots,\widehat{P}_{k_{\min}}}\left[(I-\beta P)^{\max\{k_{\max}-k,0\}}\left(\prod_{j=1}^{k_{\min}} D_{k_{\min}-j+1}\right) P \left(\prod_{j=1}^{k_{\min}} D_j\right)(I-\beta P)^{\max\{k_{\max}-k',0\}}\right] \\
&\preceq \beta^2 \sum_{k=0}^{M}\sum_{k'=0}^{M} (I-\beta P)^{\max\{k_{\max}-k,0\}} P (I-\beta P)^{k_{\min}} (I-\beta P)^{\max\{k_{\max}-k',0\}} \\
&= \beta^2 P \sum_{k=0}^{M}\sum_{k'=0}^{M} (I-\beta P)^{k_{\max}}.
\end{aligned} \tag{12}$$

Let $a_{k',k} = (I - \beta P)^{k_{\max}}$. We can order the double sum as $\sum_{k=0}^{M}\sum_{k'=0}^{M} a_{k,k'} = 2\sum_{k=0}^{M}\sum_{k'=k}^{M} a_{k,k'} - \sum_{k=0}^{M} a_{k,k}$ since $a_{k,k'} = a_{k',k}$. Thus, by using that $(I - \beta P) \preceq I$, we can see that

$$\begin{aligned}
\sum_{k=0}^{M}\sum_{k'=0}^{M}(I-\beta P)^{k_{\max}} &= 2\sum_{k=0}^{M}\sum_{k'=k}^{M}(I-\beta P)^{k_{\max}} - \sum_{k=0}^{M}(I-\beta P)^{k_{\max}} \\
&\preceq 2\sum_{k=0}^{M}\sum_{k'=k}^{M}(I-\beta P)^{k_{\max}} \\
&= 2\sum_{k=0}^{M}\sum_{k'=k}^{M}(I-\beta P)^{k'} \\
&= 2\sum_{k=0}^{M}(I-\beta P)^{k}\sum_{k'=k}^{M}(I-\beta P)^{k'-k}
\end{aligned}$$

where in the third equality we replaced $k_{\max}$ by k'. By using the equality $P^{-1} = \beta\sum_{k=0}^{\infty}(I-\beta)^k$ (equation (3) of Neu and Olkhovskaya (2020)) and the fact that $I - \beta P$ and P are positive semi-definite we can see that

$$\begin{aligned}
2\sum_{k=0}^{M}(I-\beta P)^{k}\sum_{k'=k}^{M}(I-\beta P)^{k'-k} &= 2\sum_{k=0}^{M}(I-\beta P)^{k}(\beta^{-1}P^{-1} - (I-\beta P)^{M-k}\beta^{-1}P^{-1}) \\
&\preceq 2\sum_{k=0}^{M}(I-\beta P)^{k}\beta^{-1}P^{-1} \\
&= 2(\beta^{-1}P^{-1} - (I-\beta P)^{M}\beta^{-1}P^{-1})\beta^{-1}P^{-1} \\
&\preceq 2\beta^{-2}P^{-2}. \qquad (13)
\end{aligned}$$

Using equations (13) and (12) we may conclude the proof as

$$\begin{aligned}
\operatorname{tr}\left(\mathbb{E}_{\mathrm{MGR}}[P^{\top}\widehat{P}^{+\top}P\widehat{P}^{+}]\right) &\le \operatorname{tr}\left(P^{\top}\beta^{2}P\sum_{k=0}^{M}\sum_{k'=0}^{M}(I-\beta P)^{k_{\max}}\right) \\
&\le 2\operatorname{tr}\left(P^{\top}\beta^{2}P(\beta^{-2}P^{-2})\right) \\
&= 2\operatorname{tr}\left(P\beta^{2}P\beta^{-2}P^{-2}\right) \\
&= 2b.
\end{aligned}$$

□

Lemma 2 (Neu and Olkhovskaya (2020, Equation (6)))**.** Let $\widetilde{\Theta}_t$ be some estimator of Θ_t with bias $B_t = \Theta_t - \widetilde{\Theta}_t$, then for any $X_0 \sim \mathcal{D}$

$$\begin{aligned}
\mathcal{R}_T \le\ & \mathbb{E}_{X_0}\left[\widehat{\mathcal{R}}_T(X_0)\right] \\
&+ 2\mathbb{E}\left[\sum_{t=1}^{T}\max_{A\in\mathcal{A}}\left|\mathbb{E}\left[X_0^{\top}B_tA\middle|\mathcal{F}_{t-1}\right]\right|\right]
\end{aligned}$$

Proof. The result is stated in equation (6) of Neu and Olkhovskaya (2020). □

B DETAILS OF SECTION 3 (SEMI-BANDIT SETTING)

We explicitly define a sampling scheme usable by the MGR for the semi-bandit case in Sampling Scheme 4.

Lemma 3. Let $\beta \le \min_{t,k}\frac{1}{\lambda_{\max}(\Sigma_{t,k})}$. For any $A \in \mathcal{A}$ and all $x \in \mathcal{X}$ Algorithm 2 guarantees

$$\mathbb{E}\left[x^{\top}(\widehat{\Theta}_t - \widetilde{\Theta}_t)A \,\middle|\, \mathcal{F}_{t-1}\right] \le R\sqrt{m}\sigma\, e^{-\frac{M\beta\gamma}{|E|}\lambda_{\min}(\Sigma)}$$

simultaneously for all $t = 1, \ldots, T$.

Sampling Scheme 4 CO_2-FTRL Sampling Scheme

Require: Context distribution $\mathcal{D}$, current policy π_t, k

1: Draw $X \sim \mathcal{D}$
2: Draw $A \sim \pi_t(\cdot|X)$
3: Output $(A)_k X X^\top$

Proof. By using the fact that $\widehat{P}_k$ is unbiased we can see that

$$\mathbb{E}\left[\Theta_t - \widetilde{\Theta}_t \mid \mathcal{F}_{t-1}\right] = (I - \beta\Sigma_{t,k})^M$$

Now, by Hölder's inequality and our assumptions on $\mathcal{X}$, Θ_t, and $\mathcal{A}$, we have that for any $A \in \mathcal{A}$ and any $x \in \mathcal{X}$

$$\begin{aligned}&\mathbb{E}\left[x^\top(\Theta_t - \widetilde{\Theta}_t)A \mid \mathcal{F}_{t-1}\right]\\ &\leq R\sqrt{m}\sigma \left\|(I - \beta\Sigma_{t,k})^M\right\|_{\mathrm{op}}\end{aligned}$$

Since $\Sigma_{t,k} \succeq I\lambda_{\min}(\Sigma_{t,k})$, we have that $\|(I - \beta\Sigma_{t,k})^M\|_{\mathrm{op}} \leq \big(1 - \beta\lambda_{\min}(\Sigma_{t,k})\big)^M$ and therefore

$$\mathbb{E}\left[x^\top(\Theta_t - \widetilde{\Theta}_t)A \mid \mathcal{F}_{t-1}\right] \leq \sqrt{m}\sigma R \exp\big(-M\beta\lambda_{\min}(\Sigma_{t,k})\big),$$

where we used $1 + x \leq \exp(x)$.

By construction of $\mathbb{E}\ \pi_t$ guarantees that $\mathbb{P}_t\big((A_t)_k = 1\big) \geq \gamma|E|^{-1}$. Therefore we may conclude that $\lambda_{\min}(\Sigma_{t,k}) \geq \gamma\lambda_{\min}(\Sigma)|E|^{-1}$. □

Lemma 4. Let $\beta = \frac{1}{\sigma^2}$ and let $\eta \leq \frac{\ln(2)}{M+1}$. For any $x \in \mathcal{X}$, $\bar{A}_t(x)$ defined by (5) and any $u \in \mathcal{A}$ with φ as in equation (4) guarantees

$$\begin{aligned}&\sum_{t=1}^T x^\top \widetilde{\Theta}_t(\bar{A}_t(x) - u)\\ &\qquad\leq \frac{m\left(1 + \ln\left(\frac{K}{m}\right)\right)}{\eta} + \eta\sum_{t=1}^T\sum_{k=1}^K (x^\top\widetilde{\Theta}_t)_k^2(\bar{A}_t(x))_k\,.\end{aligned}$$

Proof. In order to decompose the regret into auxiliary games, we fix a context $x \in \mathcal{X}$.

We define $F_t(A) = \sum_{s=1}^t x^\top\widetilde{\Theta}_s A + \varphi(A)$. We will use Orabona (2019, Lemma 7.13) to control the regret. We restate a less general version in our notation here.

Lemma 15 (Orabona (2019, Lemma 7.13))**.** *Let $\mathcal{A}$ to be non-empty and closed and $\arg\min_{A\in\mathcal{A}} F_t(A)$ exists and is non-empty. Assume $\varphi(\cdot)$ is twice differentiable with a positive definite Hessian in the interior of its domain. Then, for all $t = 1, \ldots, T$ there exists a z_t on the line segment between $\bar{A}_t(x) = \arg\min_{A\in\mathrm{Conv}(\mathcal{A})} F_{t-1}$ and $\widetilde{A}_{t+1}(x) = \arg\min_{A\in\mathbb{R}^K} F_t$ such that the following holds for any $u \in \mathcal{A}$*

$$\sum_{t=1}^T x^\top\widetilde{\Theta}_t(\bar{A}_t(x) - u) \leq \varphi(u) - \varphi(\bar{A}_1(x)) + \frac{1}{2}\sum_{t=1}^T \left\|x^\top\widetilde{\Theta}_t\right\|^2_{\left(\frac{\partial^2}{(\partial z_t)^2}\varphi(z_t)\right)^{-1}},$$

where $\left\|x^\top\widetilde{\Theta}_t\right\|^2_{\left(\frac{\partial^2}{(\partial z_t)^2}\varphi(z_t)\right)^{-1}} = x^\top\left(\frac{\partial^2}{(\partial z_t)^2}\varphi(z_t)\right)^{-1}x.$

Since the Hessian of φ at A is a diagonal matrix with diagonal elements $1/(A)_1, \ldots, 1/(A)_K$ it is clear to see that this Hessian is positive definite for all $A \in \mathrm{int}\big(\mathrm{Conv}(\mathcal{A})\big)$ and thus the requirements of the lemma are met.

We proceed to bound $-\varphi(A)$ and for any $A \in \mathrm{Conv}(\mathcal{A})$ we have that

$$\begin{aligned}
-\varphi(A) &= -\frac{1}{\eta}\sum_{k=1}^{K}\big((A)_k \ln(A)_k - (A)_k\big) \\
&= \frac{1}{\eta}\|A\|_1 \sum_{k=1}^{K}\left(\frac{(A)_k}{\|A\|_1}\ln\frac{1}{(A)_k}\right) + \frac{\|A\|_1}{\eta} \\
&\le \frac{1}{\eta}\|A\|_1 \ln\left(\sum_{k=1}^{K}\frac{(A)_k}{\|A\|_1}\frac{1}{(A)_k}\right) + \frac{\|A\|_1}{\eta} \\
&\le \frac{m\big(1+\ln(\frac{K}{m})\big)}{\eta}, \qquad (14)
\end{aligned}$$

where in the second inequality we used Jensen's inequality and in the last inequality we used that $x\ln(K/x) + x$ is increasing in x for $x \in [1, K]$ and the assumption that $\|A\|_1 \le m$.

Now, to control the $\big\|x^\top \widetilde{\Theta}_t\big\|^2_{\left(\frac{\partial^2}{(\partial z_t)^2}\varphi(z_t)\right)^{-1}}$ term we need to control z_t. Recall that by assumption $|(x^\top \Theta_t)_k| \le 1$ for any $x \in \mathcal{X}$. Therefore we have that

$$\begin{aligned}
\max_{k\in[K]}\big|\eta x^\top(\widetilde{\Theta}_t)_k\big| &= \max_{k\in[K]}\big|\eta x^\top(\widetilde{\Theta}_t)_k\big| \\
&= \max_{k\in[K]}\Big|\eta x^\top \widehat{\Sigma}^+_{t,k} X_t (X_t^\top \Theta_t)_k (A_t)_k\Big| \\
&\le \max_{k\in[K]}\big|\eta x^\top \widehat{\Sigma}^+_{t,k} X_t\big| \\
&\le \eta\sigma^2 \max_{k\in[K]}\big\|\widehat{\Sigma}^+_{t,k}\big\|_{\mathrm{op}} \\
&\le \eta\sigma^2\beta(M+1) \\
&\le \ln(2),
\end{aligned}$$

where the second inequality is Hölder's inequality, the third inequality is due to Lemma 1, and the last equality holds since by assumption $\eta \le \frac{\ln(2)}{(M+1)}$ and $\beta = \frac{1}{\sigma^2}$. Since $(\widetilde{A}_{t+1}(x))_k = (\bar{A}_t(x))_k \exp(-\eta(x^\top\widetilde{\Theta}_t)_k)$ this means that $(\widetilde{A}_{t+1}(x))_k(x) \in [0.5\bar{A}_t(x), 2\bar{A}_t(x)]$. Now, since z_t is on the line segment between $\widetilde{A}_{t+1}(x)$ and $\bar{A}_t(x)$, it follows that $(z_t)_k \le 2(\bar{A}_t(x))_k$ for all $k \in [K]$. This in turn implies that

$$\big\|x^\top\widetilde{\Theta}_t\big\|^2_{\left(\frac{\partial^2}{(\partial z_t)^2}\varphi(z_t)\right)^{-1}} \le 2\big\|x^\top\widetilde{\Theta}_t\big\|^2_{\left(\frac{\partial^2}{(\partial \bar{A}_t(x))^2}\varphi(\bar{A}_t(x))\right)^{-1}},$$

which, when combined with Lemma 15 and equation (14) completes the proof. □

Lemma 5. For any $\gamma \in (0,1)$ and for all $t \in [T]$ we have that

$$\mathbb{E}\left[\sum_{k=1}^{K}\big(X_0^\top\widetilde{\Theta}_t\big)_k^2\big(\bar{A}_t(X_0)\big)_k\,\Big|\,\mathcal{F}_{t-1}\right] \le \frac{3Kd}{1-\gamma}$$

Proof. By using that $X_t^\top(\Theta_t)_k \le 1$ we can see that

$$\begin{aligned}
&\mathbb{E}_{X_0}\left[\mathbb{E}\Big[\sum_{k=1}^{K}(X_0^\top\widetilde{\Theta}_t)_k^2(\bar{A}_t(X_0))_k\Big|X_0,\mathcal{F}_{t-1}\Big]\Big|\mathcal{F}_{t-1}\right] \\
&= \mathbb{E}_{X_0}\left[\mathbb{E}\Big[\sum_{k=1}^{K}\Big(X_0^\top\widehat{\Sigma}^+_{t,k}X_t(X_t^\top\Theta_t)_k(A_t)_k\Big)^2(\bar{A}_t(X_0))_k\Big|X_0,\mathcal{F}_{t-1}\Big]\Big|\mathcal{F}_{t-1}\right] \\
&\le \mathbb{E}_{X_0}\left[\mathbb{E}\Big[\sum_{k=1}^{K}(X_0^\top\widehat{\Sigma}^+_{t,k}X_t)^2(A_t)_k(\bar{A}_t(X_0))_k\Big|X_0,\mathcal{F}_{t-1}\Big]\Big|\mathcal{F}_{t-1}\right].
\end{aligned}$$

Now, by writing out the square we can see that

$$
\begin{aligned}
&(1-\gamma)\mathbb{E}_{X_0}\left[\mathbb{E}\Big[\sum_{k=1}^{K}(X_0^\top\widehat{\Sigma}_{t,k}^{+}X_t)^2(A_t)_k(\bar{A}_t(X_0))_k\Big|X_0,\mathcal{F}_{t-1}\Big]\Big|\mathcal{F}_{t-1}\right]\\
&=(1-\gamma)\mathbb{E}_{X_0}\left[\mathbb{E}\Big[\sum_{k=1}^{K}X_0^\top\widehat{\Sigma}_{t,k}^{+}X_tX_t^\top(A_t)_k\widehat{\Sigma}_{t,k}^{+\top}X_0(\bar{A}_t(X_0))_k\Big|X_0,\mathcal{F}_{t-1}\Big]\Big|\mathcal{F}_{t-1}\right]\\
&=(1-\gamma)\mathbb{E}_{X_0}\left[\sum_{k=1}^{K}X_0^\top\widehat{\Sigma}_{t,k}^{+}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+\top}X_0(\bar{A}_t(X_0))_k\Big|\mathcal{F}_{t-1}\right],
\end{aligned}
$$

where in the last equality we used the definition of $\Sigma_{t,k}$. Now, using the definition of π_t we can see that

$$
\begin{aligned}
&(1-\gamma)\mathbb{E}_{X_0}\left[\sum_{k=1}^{K}X_0^\top\widehat{\Sigma}_{t,k}^{+}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+\top}X_0(\bar{A}_t(X_0))_k\Big|\mathcal{F}_{t-1}\right]\\
&\le\mathbb{E}_{X_0,A\sim\pi_t(\cdot|X_0)}\left[\sum_{k=1}^{K}X_0^\top\widehat{\Sigma}_{t,k}^{+}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+\top}X_0(A)_k\Big|\mathcal{F}_{t-1}\right]\\
&=\mathbb{E}_{X_0,A\sim\pi_t(\cdot|X_0)}\left[\sum_{k=1}^{K}\mathrm{tr}(\widehat{\Sigma}_{t,k}^{+}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+\top}X_0X_0^\top(A)_k\Big|\mathcal{F}_{t-1}\right]\\
&=\mathbb{E}_{\mathrm{MGR_t}}\left[\sum_{k=1}^{K}\mathrm{tr}(\widehat{\Sigma}_{t,k}^{+}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+\top}\Sigma_{t,k})\Big|\mathcal{F}_{t-1}\right]\\
&=\mathbb{E}_{\mathrm{MGR_t}}\left[\sum_{k=1}^{K}\mathrm{tr}(\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+\top}\Sigma_{t,k}\widehat{\Sigma}_{t,k}^{+})\Big|\mathcal{F}_{t-1}\right]\le 3Kd,
\end{aligned}
$$

where the last inequality follows from Lemma 1. Dividing by $(1-\gamma)$ completes the proof. □

Theorem 16. *Let $\beta=\frac{1}{\sigma^2}$, let $\eta\le\frac{\ln(2)}{(M+1)}$, and let $\gamma\in(0,\frac{1}{2})$. Algorithm 2 guarantees*

$$
\mathcal{R}_T\le\frac{2m\left(1+\ln\left(\frac{K}{m}\right)\right)}{\eta}+3\eta TKd+2\gamma Tm+2Tm\sigma Re^{-M\beta\gamma\frac{\lambda_{\min}(\Sigma)}{|E|}}
$$

Furthermore, if $M=\frac{|E|\ln(T)}{\gamma\beta\lambda_{\min}(\Sigma)}$, $\gamma=\min\left\{1,\sqrt{\frac{(1+\ln(K/m))|E|\ln(T)}{T\beta\lambda_{\min}(\Sigma)}}\right\}$, and $\eta=\min\left\{\frac{\ln(2)}{(M+1)},\sqrt{\frac{m\left(1+\ln\left(\frac{K}{m}\right)\right)}{3KdT}}\right\}$ then Algorithm 2 guarantees

$$
\begin{aligned}
\mathcal{R}_T\le{}&2\sqrt{3dmKT(1+\ln(K/m))}+6m\sqrt{\sigma^2(1+\ln(K/m))\frac{T|E|\ln(T)}{\lambda_{\min}(\Sigma)}}+2m\sigma R+\frac{m(1+\ln(K/m))}{\ln(2)}\\
&+\frac{2|E|\ln(T)m\sigma^2(1+\ln(K/m))}{\lambda_{\min}(\Sigma)}.
\end{aligned}
$$

Proof. Define

$$
\widetilde{\mathcal{R}}_T(x)=\mathbb{E}\left[\sum_{t=1}^{T}\left(x^\top\widetilde{\Theta}_t\bar{A}_t(x)-x^\top\widetilde{\Theta}_t\pi_T^*(x)\right)\right]
$$

and $\pi_T^*(x)=\min_{A\in\mathcal{A}}\mathbb{E}_{\mathcal{F}_T}\left[\sum_{t=1}^{T}\left(x^\top\Theta_tA\right)\right]$. We start from a slightly modified version of Lemma 2:

$$
\mathcal{R}_T\le(1-\gamma)\mathbb{E}_{X_0}\left[\widetilde{\mathcal{R}}_T(X_0)+2\mathbb{E}\Big[\sum_{t=1}^{T}\max_{A\in\mathcal{A}}\Big|\mathbb{E}\big[X_0^\top(\widehat{\Theta}_t-\widetilde{\Theta}_t)A|\mathcal{F}_{t-1}\big]\Big|\Big]\right]+2\gamma Tm
$$

where we used that $\mathbb{E}_{A_t}[A_t|\mathcal{F}_{t-1},X_0]=(1-\gamma)\bar{A}_t(X_0)+\frac{\gamma}{|E|}\sum_{A\in E}A$ and that $\frac{\gamma}{|E|}\sum_{A\in E}X_0^\top\Theta_tA\le\gamma m$.

Note that since $\Sigma_{t,k} \preceq \Sigma \preceq \sigma^2 I$ setting $\beta = \frac{1}{\sigma^2} \leq \frac{1}{\lambda_{\max}(\Sigma_{t,k})}$ is a valid choice to use in Lemma 3. By Lemma 3 we have that

$$2\mathbb{E}\left[\sum_{t=1}^{T} \max_{A\in\mathcal{A}} \left|\mathbb{E}\big[X_0^\top(\widehat{\Theta}_t - \widetilde{\Theta}_t)|\mathcal{F}_{t-1}\big]\right|\right] \leq 2T\sqrt{m}\sigma R e^{-M\beta\gamma\frac{\lambda_{\min}(\Sigma)}{|E|}}.$$

Now, by Lemmas 4 and 5 we have that

$$(1-\gamma)\mathbb{E}_{X_0}\left[\widetilde{\mathcal{R}}_T(X_0)\right] \leq \frac{m\left(1+\ln\left(\frac{K}{m}\right)\right)}{\eta} + 3\eta TKd$$

Combining the above we find

$$\mathcal{R}_T \leq \frac{2m\left(1+\ln\left(\frac{K}{m}\right)\right)}{\eta} + 3\eta TKd + 2\gamma Tm + 2T\sqrt{m}\sigma R e^{-M\beta\gamma\frac{\lambda_{\min}(\Sigma)}{|E|}}.$$

Now, setting $M = \frac{|E|\ln(T)}{\gamma\beta\lambda_{\min}(\Sigma)}$ we find

$$\mathcal{R}_T \leq \frac{m\left(1+\ln\left(\frac{K}{m}\right)\right)}{\eta} + 3\eta TKd + 2\gamma Tm + 2\sqrt{m}\sigma R.$$

Set $\gamma = \min\left\{1, \sqrt{\frac{(1+\ln(K/m))|E|\ln(T)}{T\beta\lambda_{\min}(\Sigma)}}\right\}$ to find that $M = \max\left\{\frac{|E|\ln(T)}{\beta\lambda_{\min}(\Sigma)}, \sqrt{T\frac{|E|\ln(T)}{(1+\ln(K/m))\beta\lambda_{\min}(\Sigma)}}\right\}$ and

$$\mathcal{R}_T \leq \frac{m\left(1+\ln\left(\frac{K}{m}\right)\right)}{\eta} + 3\eta TKd + 2m\sqrt{(1+\ln(K/m))\frac{T|E|\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + 2\sqrt{m}\sigma R.$$

Finally, setting $\eta = \min\left\{\frac{\ln(2)}{(M+1)}, \sqrt{\frac{m\left(1+\ln\left(\frac{K}{m}\right)\right)}{3KdT}}\right\}$ and replacing M by its value we find

$$\begin{aligned}
\mathcal{R}_T &\leq 2\sqrt{3dmKT(1+\ln(K/m))} + 2m\sqrt{(1+\ln(K/m))\frac{T|E|\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + 2m\sigma R + \frac{(M+1)m(1+\ln(K/m))}{\ln(2)} \\
&\leq 2\sqrt{3dmKT(1+\ln(K/m))} + 6m\sqrt{(1+\ln(K/m))\frac{T|E|\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + 2m\sigma R + \frac{m(1+\ln(K/m))}{\ln(2)} \\
&\quad + \frac{2|E|\ln(T)m(1+\ln(K/m))}{\beta\lambda_{\min}(\Sigma)},
\end{aligned}$$

after which replacing $\beta = \frac{1}{\sigma^2}$ completes the proof. □

C DETAILS OF SECTION 5 (LOWER BOUNDS)

Before we prove the lower bound we first describe a peculiar property of our estimators. Suppose $\mathcal{X}$ consists of only basis vectors. Also suppose that in round t the context was a basis vector in direction i. Then the feedback obtained at round t does not affect the algorithm's prediction at all subsequent rounds where the context is a basis vector in direction $i' \neq i$. More formally if $X_t = e_i \neq e_j = X_{t'}$ then for all A

$$X_{t'}^\top \widetilde{\Theta}_t A = 0.$$

The proof of this statement can be found in Lemma 17. This implies that equation (5) in Algorithm 2 reduces to

$$\bar{A}_t(\boldsymbol{e}_i) = \operatorname*{arg\,min}_{A\in\mathrm{Conv}(\mathcal{A})} \sum_{s=1}^{t-1} \boldsymbol{e}_i^\top \widetilde{\Theta}_s A + \varphi(A) = \operatorname*{arg\,min}_{A\in\mathrm{Conv}(\mathcal{A})} \sum_{s<t\,:\,X_s=\boldsymbol{e}_i} \boldsymbol{e}_i^\top \widetilde{\Theta}_s A + \varphi(A)$$

Similarly equation (9) of Tensor-Exp3 (Algorithm 3) reduces to

$$w_t(X_t, A) = \exp\left(-\eta \sum_{s=1}^{t-1} X_t^\top \widetilde{\Theta}_s A\right) = \exp\left(-\eta \sum_{s<t\,:\,X_s=\boldsymbol{e}_i} X_t^\top \widetilde{\Theta}_s A\right).$$

Hence, both algorithms ignore feedback from any previous round s in which $X_t \neq X_s$.

Lemma 17. *Let $\mathcal{X}$ consist of only basis vectors and pick some $t \in [T]$. Let $X_{t'} \neq X_t$ and let $A \in \mathcal{A}$, then*

$$X_{t'}^\top \widetilde{\Theta}_t A = 0$$

holds for the biased and unbiased estimators of CO$_2$-FTRL (equation (3) *and equation* (2)*) as well as the biased and unbiased estimators of Tensor-EXP3 (equation* (8) *and equation* (7)*).*

Proof. First we introduce the concept of a n-sparse matrix which we define as a matrix such that $A_{i,j} = 0$ if $i \not\equiv j \mod n$.

Now let D and E be n-sparse, then so is $E + D$ as well as ED, which we can recognise by spelling out

$$ED = \{E_a{}^b D_b{}^c\}_a{}^c.$$

Then ED at the index a, c can only be non-zero if there exist some b such that $a \equiv b \mod n$ and $c \equiv b \mod n$, which can only exist if $a \equiv c \mod n$.

Let $\widehat{P}^+$ be a sample of the MGR process, we can conclude that it is n-sparse if all samples $\widehat{P}_j$ are also n-sparse by writing out

$$\widehat{P}^+ = \beta \sum_{k=0}^{M} \prod_{j=1}^{k} (I - \beta \widehat{P}_j).$$

Let $\widehat{P} \in \mathbb{R}^{d \times d}$ be a sample generated by the Sampling Scheme 4, used by CO$_2$-FTRL. Then $\widehat{P}$ is diagonal and thus d-sparse. It follows that $\Sigma_{t,k}^{-1}$ is also diagonal for all k. Let $X_{t'} = \boldsymbol{e}_i$ and $X_t = \boldsymbol{e}_j$ and pick $k \in [K]$. We now consider the kth entry of the product $X_{t'}^\top \widetilde{\Theta}_t$, given by the biased estimator for CO$_2$-FTRL as defined in equation (3). First we recognise that $X_{t'}$ selects the ith row in the first equality. In the second equality we pull out the scalars $(X_t^\top \Theta_t)_k (A_t)_k$ and we finish in the last equality be recognising that X_t selects the jth column.

$$\begin{aligned}
(X_{t'}^\top \widetilde{\Theta}_t)_k &= (\widetilde{\Theta}_t)_{i,k} \\
&= (\Sigma_{t,k}^{-1} X_t (X_t^\top \Theta_t)_k (A_t)_k)_i \\
&= (\Sigma_{t,k}^{-1} X_t)_i (X_t^\top \Theta_t)_k (A_t)_k \\
&= (\Sigma_{t,k}^{-1})_{i,j} (X_t^\top \Theta_t)_k (A_t)_k
\end{aligned}$$

From here it is clear that $(\Sigma_{t,k}^{-1})_{i,j}$ can only be non-zero if $i = j$ and if $j \neq i$, we conclude that

$$X_{t'}^\top \widetilde{\Theta}_t A = 0,$$

showing the first result.

For the biased estimator of Tensor-EXP3, as given in equation (8), we investigate $(B \otimes C)^F$ for some matrices B, C. By the definition of $\otimes$ (Definition D.6) and flattening (Definition D.8), we have that

$$\begin{aligned}
(B \otimes C)^F &= (\{B_a{}^b C^\top{}_c{}^d\}_a{}^b{}_c{}^d)^F \\
&= \{B_{(a \mod m)+1}{}^{(b \mod m)+1} C^\top{}_{\lfloor b/m \rfloor + 1}{}^{\lfloor a/m \rfloor + 1}\}_a{}^b.
\end{aligned}$$

If B is now a diagonal matrix of dimension n, then it is clear that any entry of $(B \otimes C)^F$ at the index a, b must be zero if $a \not\equiv b \mod n$, we conclude that $(B \otimes C)^F$ is n-sparse. It follows that any $\widehat{P}$ drawn using Sampling Scheme 5, the sampling scheme associated with Tensor-Exp3, is d-sparse. As a conclusion $\widehat{\Psi}_t^{+\,F}$ is also d-sparse.

Unflattening (Definition D.9) some n-sparse matrix $C \in \mathbb{R}^{mn \times mn}$ can only be non-zero for some indices a, b, c, d if $a \equiv b \mod n$ as $C^U{}_a{}^b{}_c{}^d = C_{(a-1)+(d-1)n}{}^{(b-1)+(c-1)n}$.

We first apply the definition of $\widetilde{\Theta}_t$ (equation (8)) and then apply Lemma 7

$$\begin{aligned}
X_{t'}^\top \widetilde{\Theta}_t A &= X_{t'} \widehat{\Psi}_t^+ \big(X_t X_t^\top \Theta_t A_t A_t^\top\big) A \\
&= X_{t'}^\top \widehat{\Psi}_t^+ \big((X_t X_t^\top \otimes A_t A_t^\top)\big)(\Theta_t) A \\
&= X_{t'}^\top \{\widehat{\Psi}_t^+{}_a{}^b{}_c{}^d X_t{}_b X_t^\top{}^e A_t{}_f A_t^\top{}^c)\}_a{}^e{}_f{}^d (\Theta_t) A \\
&= \{X_{t'}^\top{}^a \widehat{\Psi}_t^+{}_a{}^b{}_c{}^d X_t{}_b X_t^\top{}^e A_t{}_f A_t^\top{}^c)\}_a{}^e{}_f{}^d (\Theta_t) A = 0,
\end{aligned}$$

where in the third equality we used definitions of the tensor product (Definition D.6) and of $\Psi(\Phi)$ (Definition D.1). The final equality is due to the fact that $\widehat{\Psi}_t^{+}{}_a{}^b{}_c{}^d = 0$ if $a \neq b$ that for any $X_{t'} \neq X_t$ and any A.

For the unbiased estimator of CO_2-FTRL, as given in equation (2), it is enough to simply recognise that $\mathbb{E}_{A_t,X}\left[(A_t)_k X X^\top \mid \mathcal{F}_{t-1}\right]$ is always diagonal if all X are basis vectors. The statement follows by the same arguments as above. For the unbiased estimator of Tensor-EXP3, as given in equation (8), we recognise that applying the MGR with Sampling Scheme 5 and $M = \infty$ yields Ψ_t^{-1F} as shown in Section 3.2 of Neu and Olkhovskaya (2020). To find the i, j coordinate of Ψ_t^{-1F} we can write

$$(\Psi_t^{-1F})_{i,j} = \left(\beta \sum_{k=0}^{\infty} C_k\right)_{i,j} = \beta \sum_{k=0}^{\infty} (C_k)_{i,j}\,,$$

where all C_k, as defined by the MGR, are d-sparse as shown above. The rest of the argument follows as above by thus recognising Ψ_t^{-1F} as d-sparse. □

Theorem 18. *In the semi-bandit setting, for all $T \geq 0.0064\, dK^3$ and for any possibly randomized orthogonal algorithm, there exists a sequence of losses $\Theta_1, \ldots, \Theta_T$ such that $\mathcal{R}_T \geq 0.017\sqrt{dmKT}$.*

Proof. In the construction of the lower bound we consider sequences of losses that are independent of the actions of the learner and contexts. The contexts are basis vectors sampled uniformly at random. For any sequence of (randomized) context to action mapping π_t we have that

$$\begin{aligned}\mathcal{R}_T &= \mathbb{E}\Big[\sum_{t=1}^{T} \boldsymbol{e}_i^\top \Theta_t \pi_t(\boldsymbol{e}_i) - \min_{A\in\mathcal{A}} \sum_{t=1}^{T} \boldsymbol{e}_i^\top \Theta_t A\Big] \\ &= \mathbb{E}\Big[\sum_{t=1}^{T} (\Theta_t)_{i,\cdot}\, \pi_t(\boldsymbol{e}_i) - \min_{A\in\mathcal{A}} \sum_{t=1}^{T} (\Theta_t)_{i,\cdot} A\Big].\end{aligned}$$

Note that introduced the ghost sample $X_0 = \boldsymbol{e}_i$ as in Lemma 2.

Since we assume that π_t is a orthogonal algorithm, it does not use information from rounds in which $X_s \neq \boldsymbol{e}_i$ for $s < t$ to compute π_t.

To prove the lower bound we will use Yao's minimax principle, which tells us that it is sufficient to provide a stochastic strategy for the adversary on which the expected regret of any deterministic algorithm is lower bounded. In the construction of the lower bound the action set $\mathcal{A}$ is the set of basis vectors. The losses are generated as follows. We sample Z from the uniform distribution over $[K]$. Conditioned on $Z = k$, the loss $(\Theta_t)_{i,k'}$ is sampled from an independent Bernoulli distribution with mean $\frac{1}{2}$ if $k' \neq k$ and it is sampled from Bernoulli distribution with mean $\frac{1}{2} - \epsilon$ for some $\epsilon \in [0, \frac{1}{4}]$.

We follow the proof of Theorem 7 by Esposito et al. (2022). Esposito et al. (2022) construct a lower bound for online learning with stochastic feedback graphs, where the only edges in the feedback graphs are self loops which realise with probability $\frac{1}{d}$ (the lower bound constructed by Esposito et al. (2022) is more general, but for our results we only require this particular instance).

Random variables $T_1, \ldots, T_K$ denote the number of times that the learner played an $\pi_t(\boldsymbol{e}_i) = A_t^i$ such that $(A_t^i)_k = 1$. For each $k \in [K]$ we introduce notations $\mathbb{P}_k$ and $\mathbb{E}_k$ to denote the probability and expectation with respect to the marginal distributions under which $Z = k$. From equation (16) in (Esposito et al., 2022) we have that for any deterministic algorithm

$$\mathcal{R}_T \geq \epsilon\left(T - \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_k[T_k]\right)$$

We also consider auxiliary distribution $\mathbb{P}_0$, which is equivalent to distribution $\mathbb{P}_k$ as specified above but with $\epsilon = 0$ for all k. We denote the corresponding expectation by $\mathbb{E}_0$. Denote by λ_t the feedback set in round t. Denote by $\lambda^t = (\lambda_1, \ldots, \lambda_t)$ the tuple of feedback sets the learner has access to in round $t + 1$.

Since the action $\pi_t(\boldsymbol{e}_i)$ is fully determined by λ^{t-1}, the central object of interest is the distribution over λ^{t-1} and in particular the information the learner gains from observing certain losses. Observe that if the action in round t given λ^{t-1} is not equal to $\boldsymbol{e}_k$ then the learner does not obtain any information. If the action of the learner given the history of losses is

equal to $\boldsymbol{e}_k$ the learner only obtains information if $X_t = \boldsymbol{e}_i$. To formalise this idea, we use equation (17) of Esposito et al. (2022):

$$\mathbb{E}_k[T_k] - \mathbb{E}_0[T_k] \leq \sqrt{\frac{1}{2}\sum_{t=1}^{T}\sum_{\lambda^{t-1}} \mathrm{KL}(P_{0,t}\|P_{k,t})},$$

where $P_{k,t}(\lambda_t) = \mathbb{P}_k(\lambda_t|\lambda^{t-1})$ and KL is the KL-divergence. Since the distribution of λ_t given λ^{t-1} is the same under $\mathbb{P}_0$ and $\mathbb{P}_k$ when $\pi_t(\boldsymbol{e}_i) \neq \boldsymbol{e}_k$ the KL-divergence is 0. If $\pi_t(\boldsymbol{e}_i) \neq \boldsymbol{e}_k$ then with probability $\frac{1}{d}$ the learner observes the loss and with probability $1 - \frac{1}{d}$ the learner does not observe anything. Thus, by equation (18) of Esposito et al. (2022) we have that $\mathrm{KL}(P_{0,t}\|P_{k,t}) \leq 8\ln(4/3)\epsilon^2\frac{1}{d}$. Thus, for all $T \geq 0.0064\, dK^3$, we can now simply follow the remainder of the proof of Theorem 7 of Esposito et al. (2022) and use the same parameters to arrive at

$$\left(\mathbb{E}\left[\sum_{t=1}^{T}(\Theta_t)_{i,\cdot}\pi_t(\boldsymbol{e}_i)\right] - \mathbb{E}\left[\min_{A}\sum_{t=1}^{T}(\Theta_t)_{i,\cdot}A\right]\right) \geq 0.017\sqrt{dKT}\,.$$

Therefore, we may conclude that any orthogonal algorithm satisfies

$$\mathcal{R}_T \geq 0.017\sqrt{dKT}$$

which, after observing that $m = 1$, completes the proof.

□

Theorem 19. *Suppose $T \geq dmK$ and that $K \geq 2m$. In the full-bandit setting, any orthogonal algorithm satisfies*

$$\mathcal{R}_T \geq \frac{m^{3/2}\sqrt{dKT}}{16(192 + 96\ln(T))}$$

Proof. As in the proof of Theorem 18 the proof heavily relies on the following. In the construction of the lower bound we consider sequences of losses that are independent of the actions of the learner and contexts. The context are basis vectors sampled uniformly at random. For any sequence of (randomized) context to action mapping π_t we have that

$$\begin{aligned}\mathcal{R}_T &= \mathbb{E}\Big[\sum_{t=1}^{T}\boldsymbol{e}_i^\top\Theta_t\pi_t(\boldsymbol{e}_i) - \min_{A\in\mathcal{A}}\sum_{t=1}^{T}\boldsymbol{e}_i^\top\Theta_t A\Big] \\ &= \mathbb{E}\Big[\sum_{t=1}^{T}(\Theta_t)_{i,\cdot}\pi_t(\boldsymbol{e}_i) - \min_{A\in\mathcal{A}}\sum_{t=1}^{T}(\Theta_t)_{i,\cdot}A\Big].\end{aligned}$$

Note that introduced the ghost sample $X_0 = \boldsymbol{e}_i$ as in Lemma 2. Since we assume that π_t is an orthogonal algorithm, it does not use information from rounds in which $X_s \neq \boldsymbol{e}_i$ for $s < t$ to compute π_t. For simplicity we write $A_t^i = \pi_t(\boldsymbol{e}_i)$ and assume that $n = K/m$ is an integer. The set of actions we consider is

$$\mathcal{A} = \left\{A \in \{0,1\}^K : \forall j \in [m] \sum_{k=(j-1)n+1}^{jn} A_k = 1\right\}.$$

In other words, we consider m instances of the n-armed bandit problem.

As in to the proof of Theorem 18, to prove the lower bound we use Yao's minimax principle. The sequence of stochastic losses that we use is almost exactly the same as the sequence of stochastic losses chosen by Cohen et al. (2017).

As in the proof of the lower bound by Cohen et al. (2017), we first construct an environment which generates unbounded losses, after which we adapt the lower bound to the bounded loss setting.

Let $\varepsilon = \sigma\sqrt{dmK/4T}$ for some $\sigma > 0$. In each of the m bandit problems, the environment samples the best action uniformly at random. Denote by $A^\star \in \mathcal{A}$ the vector of the best composite action. In every round t, the environment samples $\zeta_t \sim \mathcal{N}(0,\sigma)$ and sets $(\Theta_t)_{i,k} = \frac{1}{2} - \varepsilon(A^\star)_k + \zeta_t$.

We now follow the proof by Cohen et al. (2017, Lemma 4). We denote by $a_1^\star, \ldots, a_n^\star$ the locations of the non-zero coordinates of $A^\star$, arranged in increasing order. Random variables $T_1, \ldots, T_m$ denote the number of times that the learner

played an A_t^i such that $(A_t^i)_{a_j^\star} = 1$. For each $A \in \mathcal{A}$ we introduce notations $\mathbb{P}_A$ and $\mathbb{E}_A$ to denote the probability and expectation with respect to the marginal distributions under which $A = A^\star$. By Cohen et al. (2017, equation (5)) we have that

$$\mathbb{E}\left[\sum_{t=1}^{T}(A_t^i - A^\star)(\Theta_t)_{\cdot,i}\right] = \varepsilon\left(mT - \sum_{j=1}^{m}\frac{1}{n^m}\sum_{A\in\mathcal{A}}\mathbb{E}_A[T_j]\right) . \tag{15}$$

For every $A \in \mathcal{A}$ we also define the auxillary distribution $\mathbb{P}_{A,-j}$ and corresponding expectation $\mathbb{E}_{A,-j}$. This is the same distribution as $\mathbb{P}_A$ except with $(\Theta_t)_{i,k} = \frac{1}{2} + \zeta_t$. Denote by λ_t the loss observed in round t and by $\lambda^t = (\lambda_1, \ldots, \lambda_t)$ the tuple of losses observed up to and including round t. Crucially, λ_t might be empty since the learner does not observe $(\Theta_t)_{\cdot,i}$ whenever $X_t \neq \boldsymbol{e}_i$. Since the sequence λ^T determines the actions of the algorithm over the game we have that by Pinsker's inequality

$$\begin{aligned}\mathbb{E}_A[T_j] - \mathbb{E}_{A,-j}[T_j] &\leq T\sqrt{\tfrac{1}{2}\mathrm{KL}\big(\mathbb{P}_A[\lambda^T] \| \mathbb{P}_{A,-j}[\lambda^T]\big)} \\ &= T\sqrt{\tfrac{1}{2}\mathbb{E}_{\lambda^{t-1}\sim\mathbb{P}_{A,-j}}\Big[\mathrm{KL}\big(\mathbb{P}_A[\lambda_t|\lambda^{t-1}] \| \mathbb{P}_{A,-j}[\lambda_t|\lambda^{t-1}]\big)\Big]}\end{aligned} \tag{16}$$

We now shift our attention to the single terms in the sum. If $X_t \neq \boldsymbol{e}_i$ then λ_t is the same under $\mathbb{P}_A$ and $\mathbb{P}_{A,-j}$ irrespective of A_t^i and thus the KL divergence is 0. Similarly, if $(A_t^i)_{a_j^\star} = 0$ then λ_t is the same under $\mathbb{P}_A$ and $\mathbb{P}_{A,-j}$. Otherwise, if $X_t \neq \boldsymbol{e}_i$ and $(A_t^i)_{a_j^\star} = 1$ then $\mathbb{P}_A$ and $\mathbb{P}_{A,-j}$ are Gaussian distributions with the same variance $\sigma^2 m^2$ whose means are ε apart. Therefore, by the log-sum inequality we have that

$$\begin{aligned}&\mathrm{KL}\big(\mathbb{P}_A[\lambda_t|\lambda^{t-1}] \| \mathbb{P}_{A,-j}[\lambda_t|\lambda^{t-1}]\big) \\ &= \mathrm{KL}\Big((1-\tfrac{1}{d})\,\mathbb{P}_A[\lambda_t|\lambda^{t-1}, X_t \neq \boldsymbol{e}_i] + \tfrac{1}{d}\,\mathbb{P}_A[\lambda_t|\lambda^{t-1}, X_t = \boldsymbol{e}_i] \,\|\, (1-\tfrac{1}{d})\,\mathbb{P}_{A,-j}[\lambda_t|\lambda^{t-1}, X_t \neq \boldsymbol{e}_i] \\ &\qquad + \tfrac{1}{d}\,\mathbb{P}_{A,-j}[\lambda_t|\lambda^{t-1}, X_t = \boldsymbol{e}_i]\Big) \\ &\leq \frac{1}{d}\mathrm{KL}\Big(\mathbb{P}_{A,-j}[\lambda_t|\lambda^{t-1}, X_t = \boldsymbol{e}_i] \,\|\, \mathbb{P}_{A,-j}[\lambda_t|\lambda^{t-1}, X_t = \boldsymbol{e}_i]\Big) \\ &= \frac{\varepsilon^2}{d2m^2\sigma^2}\end{aligned}$$

Using the above inequality in equation (16) we can see that

$$\begin{aligned}\mathbb{E}_A[T_j] &\leq \mathbb{E}_{A,-j}[T_j] + \frac{\varepsilon T}{2m\sigma^2}\sqrt{\tfrac{1}{d}\sum_{t=1}^{T}\mathbb{P}_{A,-j}[(A_t^i)_{a_j^\star} = 1]} \\ &= \mathbb{E}_{A,-j}[T_j] + \frac{\varepsilon T}{2m\sigma^2}\sqrt{\tfrac{1}{d}\mathbb{E}_{A,-j}[T_j]}.\end{aligned}$$

Thus, by Jensen's inequality we have that

$$\begin{aligned}\frac{1}{n^m}\sum_{A\in\mathcal{A}}\mathbb{E}_A[T_j] &\leq \frac{1}{n^m}\sum_{A\in\mathcal{A}}\mathbb{E}_{A,-j}[T_j] + \frac{\varepsilon T}{2m\sigma^2}\frac{1}{n^m}\sum_{A\in\mathcal{A}}\sqrt{\tfrac{1}{d}\mathbb{E}_{A,-j}[T_j]} \\ &\leq \frac{1}{n^m}\sum_{A\in\mathcal{A}}\mathbb{E}_{A,-j}[T_j] + \frac{\varepsilon T}{2m\sigma^2}\frac{1}{n^m}\sum_{A\in\mathcal{A}}\sqrt{\frac{1}{dn^m}\sum_{A\in\mathcal{A}}\mathbb{E}_{A,-j}[T_j]} \\ &\leq \frac{T}{2} + \frac{\varepsilon T}{2\sigma}\sqrt{\frac{T}{dmK}}\end{aligned}$$

where the last inequality follows from Lemma 7 by Cohen et al. (2017) and the assumption that $K \geq 2m$.

Returning to equation (15) we can see that

$$\mathbb{E}\left[\sum_{t=1}^{T}(A_t^i - A^\star)(\Theta_t)_{\cdot,i}\right] \geq \varepsilon m\left(\tfrac{1}{2} - \frac{\varepsilon}{2\sigma}\sqrt{\frac{T}{dmK}}\right) = \frac{\sigma}{8}m^{3/2}\sqrt{dKT}$$

where the equality follows from $\varepsilon = \sigma\sqrt{dmK/4T}$.

As a final step we have to convert the regret on the unconstrained sequence of losses to the regret on a constrained sequence of losses. Luckily the steps in the proof of Cohen et al. (2017, Theorem 5) apply to our setting too, and we can see that as long as $T \geq dmK$ we can simply set $\sigma^2 = \frac{1}{192+96\ln(T)}$ and choose losses $(\Theta_t)'_{i,k} = \max\left\{\min\{(\Theta_t)_{i,k}, 1\}, 0\right\}$ to show that

$$\mathcal{R}_T \geq \frac{m^{3/2}\sqrt{dKT}}{16(192 + 96\ln(T))}$$

which completes the proof. □

D TENSOR BACKGROUND

In this section of the appendix we will rigorously introduce tensors. While tensors enjoy a wide employment in physics and other fields, to the best of our knowledge this is their first usage in bandit literature which justifies some background on tensors. Nevertheless, we only define some narrow concepts which may or may not align with how tensors have been used before.

Let $\mathbb{R}^d$ be a vectorspace made up of (column) vectors $x \in \mathbb{R}^d$ equipped with the standard basis. Covectors are now (row) vectors, which are elements of the dual space that are linear functions that map vectors to the reals, i.e., covectors are elements of the form $f : \mathbb{R}^d \to \mathbb{R}$. Thus we can combine a vector and a covector to a single real number or we could combine two covectors and get a function that takes two vectors as arguments and then returns a real number.

Tensors are now made up of vectors and covectors and we will write the rank of a vector as rank(a, b), where a is the number of vector and b the number of covector elements. A vector is of rank$(1, 0)$, a covector is rank$(0, 1)$, a matrix is of rank$(1, 1)$. We will primarily be interested in tensors of rank$(2, 2)$. While vectors and matrices are also tensors, from here on out we will usually only refer to rank$(2, 2)$ tensors as tensors, which we denote by $\Phi, \Psi \in \mathbb{R}^{m\times m\times n\times n}$. It is not required for general tensors that the first two and last two dimension agree but that will be the case for all tensors we will consider. Furthermore, we will also follow the notation introduced by Einstein (1916) and index vector elements by a lower index like this x_i and covectors with an upper index like this x^i. Since a matrix is made up of a vector and covector part, we will index it as follows $A_i{}^j$. The Einstein notation also seeks to omit a lot of the sums and clutter associated with writing tensors usually. Instead of writing all sums explicitly, when an indexing variable appears once in a vector and once in a covector component then we are implicitly summing over the variable, if it only appears once then it is an index, let $B, C \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$:

$$B_i{}^j x_j = \sum_j^m A_i{}^j x_j = (Bx)_i$$
$$B_i{}^k C_k{}^j = \sum_k^m B_i{}^k C_k{}^j = (BC)_{i,j}.$$

Finally, we will extend the notation by curly brackets that collect all free parameters and compile them to a single object, which gives greater clarity on which parameters act as free indices and which are being summed over:

$$Bx = \sum_j B_i{}^j x_j = \{B_i{}^j x_j\}_i$$
$$BC = \sum_k B_i{}^k C_k{}^j = \{B_i{}^k C_k{}^j\}_i{}^j$$
$$x^\top Bx = \sum_i \sum_j x^{\top\, i} B_i{}^j x_j = \{x^{\top\, i} B_i{}^j x_j\}_{\mathbb{R}},$$

where $\{\cdot\}_{\mathbb{R}}$ is a real number, i.e., all indices are being summed over. To reiterate, $\{\Psi_a{}^b{}_c{}^d\}_a{}^b{}_c{}^d$ is a tensor, $\Psi_a{}^b{}_c{}^d \in \mathbb{R}$ is an element of Ψ at the position a, b, c, d, and $\{\Psi_a{}^b{}_c{}^d\}_{\mathbb{R}} \in \mathbb{R}$ is obtained by summing over all indices of the tensor.

Definition D.1. Let $\Psi, \Phi \in \mathbb{R}^{m\times m\times n\times n}$ and $B, C \in \mathbb{R}^{m\times n}$, then we will define the following basic operations

$$\begin{aligned}
\Psi(B) &= \{\Psi_a{}^b{}_c{}^d B_b{}^c\}_a{}^d \\
\Psi \cdot B &= \{\Psi_a{}^b{}_c{}^d B_d{}^e\}_a{}^b{}_c{}^e \\
\Psi(\Phi) &= \{\Psi_a{}^b{}_c{}^d \Phi_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d \\
B(C) &= \{B_a{}^b C_b{}^a\}_{\mathbb{R}} \\
B \cdot \Psi &= \{B_e{}^a \Psi_a{}^b{}_c{}^d\}_e{}^b{}_c{}^d.
\end{aligned}$$

Here it is important to pay close attention to the dimensions of each element. $\Psi(A)$ is rank$(1,1)$ tensor, i.e., a matrix while $\Psi \cdot A$ is a rank$(2,2)$ tensor. All of the above operations are linear.

We denote by $\delta_{a,b}$ the Kronecker delta defined as $\delta_{a,b} = \mathbb{1}[a = b]$ with the indicator function $\mathbb{1}$ which is 1 if the condition is true and 0 otherwise. The neutral element in our tensor space is given in the following definition.

Definition D.2. We will call the neutral element of the tensor space $\mathcal{I}$ and define it as follows $\mathcal{I} = \{\delta_{a,b}\delta_{c,d}\}_a{}^b{}_c{}^d = \{\mathbb{1}[a = b \wedge c = d]\}_a{}^b{}_c{}^d$.

Thus, $\mathcal{I}_a{}^b{}_c{}^d$ is one if $a = b$ and $c = d$ and zero otherwise, in a sense the elements of any tensor where $\Psi_a{}^a{}_b{}^b$ for some a, b can be seen as the diagonal of the tensor and we will define the trace in terms of these elements later. $\mathcal{I}$ is then the tensor with ones on the diagonal and zeros everywhere else.

Observe that $\mathcal{I}$ in fact is the identity for all operations defined above:

$$\begin{aligned}
\mathcal{I}(B) &= \{\mathcal{I}_a{}^b{}_c{}^d B_b{}^c\}_a{}^d = B \\
\mathcal{I}(\Psi) &= \{\mathcal{I}_a{}^b{}_c{}^d \Psi_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d = \Psi.
\end{aligned}$$

We will define the inverse Φ of a tensor Ψ in terms of this neutral element $\mathcal{I}$ as follows.

Definition D.3. Let Φ be a tensor, we will call Φ an inverse of Φ if

$$\Phi(\Psi) = \mathcal{I} \quad \Longleftrightarrow \quad \{\Psi_a{}^b{}_c{}^d \Phi_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d = \{\delta_{a,e}\delta_{f,d}\}_a{}^e{}_f{}^d.$$

If such a Φ exists, it will also be denoted by Ψ^{-1} and we will call Φ invertible if Ψ^{-1} exists.

Lemma 20. *Let Ψ, Φ be tensors of equal dimension and let B be a matrix of appropriate dimension, then tensors are associative:*

$$\Phi(\Psi(B)) = (\Phi(\Psi))(B).$$

Proof. The proof follows from repeatedly applying the definition of $\Psi(B)$ and $\Phi(\Psi)$ in D.1.

$$\begin{aligned}
\Phi(\Psi(B)) &= \Phi(\{\Psi_a{}^b{}_c{}^d B_b{}^c\}_a{}^d) \\
&= \{\Phi_e{}^a{}_d{}^f \Psi_a{}^b{}_c{}^d B_b{}^c\}_e{}^f) \\
&= \{(\Phi(\Psi))_e{}^b{}_c{}^f B_b{}^c\}_e{}^f \\
&= (\Phi(\Psi))(B).
\end{aligned}$$

□

We now introduce our definition of the Frobenius inner product.

Definition D.4. Let Ψ, Φ be tensors. The Frobenius inner product between Ψ and Φ is defined as

$$\langle \Psi, \Phi \rangle = \{\Psi_a{}^b{}_c{}^d \Phi_b{}^a{}_d{}^c\}_{\mathbb{R}}.$$

For matrices B, C the Frobenius inner product is defined as

$$\langle B, C \rangle = \{B_a{}^b C_b{}^a\}_{\mathbb{R}} = B(C).$$

The following Lemma characterises how the inner product acts on inverses.

Lemma 21. *Let $\Psi \in \mathbb{R}^{m\times m\times n\times n}$, be invertible, then*

$$\langle \Psi, \Psi^{-1} \rangle = mn.$$

For this proof, it is more intuitive to write the sums explicitly, first we apply the definition of $\langle \cdot, \cdot \rangle$ (Definition D.4), then reordering the sums, then applying the definition of $\Psi(\Phi)$ (Definition D.1), while the last steps only consist of applying the definition of Ψ^{-1} (Definition D.3) and $\mathcal{I}$ (Definition D.2).

Proof.

$$\begin{aligned}
\langle \Psi, \Psi^{-1} \rangle &= \{\Psi_a{}^b{}_c{}^d \Psi^{-1}{}_b{}^a{}_d{}^c\}_{\mathbb{R}} \\
&= \sum_a^m \sum_b^m \sum_c^n \sum_d^n \Psi_a{}^b{}_c{}^d \Psi^{-1}{}_b{}^a{}_d{}^c \\
&= \sum_a^m \sum_d^n \sum_c^n \sum_b^m \Psi_a{}^b{}_c{}^d \Psi^{-1}{}_b{}^a{}_d{}^c \\
&= \sum_a^m \sum_d^n \Psi(\Psi^{-1})_a{}^a{}_d{}^d \\
&= \sum_a^m \sum_d^n \mathcal{I}_a{}^a{}_d{}^d \\
&= mn.
\end{aligned}$$

□

We will also need to define the transpose of the tensor.

Definition D.5. Let Ψ be a tensor, then the transpose of Ψ is defined as

$$\Psi^\top{}_a{}^b{}_c{}^d = \Psi_b{}^a{}_d{}^c.$$

A tensor is called *symmetric* iff it is invariant under transpose, i.e., $\Psi^\top = \Psi$.

Lemma 22. *Let Φ be an invertible tensor, then transposing and inverting can be interchanged, i.e.,*

$$(\Psi^{-1})^\top = (\Psi^\top)^{-1}.$$

Proof. We will show the claim be showing that $(\Psi^{-1})^\top$ is the inverse of $\Psi^\top$ by first applying the definition of $\Phi(\Psi)$ (Definition D.1), applying the definition of the transpose (Definition D.5) twice, reordering and applying the transpose to the entire expression and finally applying $\Phi(\Psi)$ again:

$$\begin{aligned}
(\Psi^{-1})^\top(\Psi^\top) &= \{(\Psi^{-1})^\top{}_a{}^b{}_c{}^d \Psi^\top{}_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d \\
&= \{\Psi^{-1}{}_b{}^a{}_d{}^c \Psi_e{}^b{}_c{}^f\}_a{}^e{}_f{}^d \\
&= \{\Psi_e{}^b{}_c{}^f \Psi^{-1}{}_b{}^a{}_d{}^c\}_a{}^e{}_f{}^d \\
&= \left(\{\Psi_e{}^b{}_c{}^f \Psi^{-1}{}_b{}^a{}_d{}^c\}_e{}^a{}_d{}^f\right)^\top \\
&= (\Psi(\Psi^{-1}))^\top.
\end{aligned}$$

□

We write the tensor product as $\otimes$. It will take two arbitrary tensors as input and output another tensor, we will only concretely define the following for two vectors x, y and two matrices B, C:

Definition D.6. Let x, y be vectors and B, C be matrices, then we define the tensor product $\otimes$ as follows

$$\begin{aligned}
(x \otimes y) &= x_a y^\top{}^b = xy^\top \\
(B \otimes C) &= B_a{}^b C^\top{}_c{}^d.
\end{aligned}$$

It is important to notice that $x \otimes y$ is a matrix of rank$(1,1)$ and $(B \otimes C)$ is a tensor of rank$(2,2)$. The following Lemma details the relationship between the tensor product and tensor matrix operations.

Lemma 7. $DCB^\top = (D \otimes B)(C)$, where B, C, D are matrices of appropriate size.

Proof. We will proof by starting from the other side and applying the definition of the tensor product (Definition D.6), followed by the definition of $\Phi(\Psi)$ (Definition D.1) and rearranging.

$$\begin{aligned}(D \otimes B)(C) &= (\{D_a{}^b B^\top{}_c{}^d\}_a{}^b{}_c{}^d)(C)\\ &= \{D_a{}^b B^\top{}_c{}^d C_b{}^c\}_a{}^d\\ &= \{D_a{}^b C_b{}^c B^\top{}_c{}^d\}_a{}^d\\ &= DCB^\top.\end{aligned}$$

□

Lemma 23. *Let B and C be symmetric matrices such that $B = B^\top$ and $C = C^\top$, now $(B \otimes C)$ is a symmetric tensor, i.e.,*

$$(B \otimes C) = (B \otimes C)^\top$$

Proof. Fix some a, b, c, d. First we use the definition of the tensor product (Definition D.6), and then we use the fact that both B and C are symmetric. Then we use the definition of transposing for matrices and finally recognise the tensor product again

$$\begin{aligned}(B \otimes C)_a{}^b{}_c{}^d &= B_a{}^b C_c{}^d\\ &= B^\top{}_a{}^b C^\top{}_c{}^d\\ &= B_b{}^a C_d{}^c\\ &= (B \otimes C)^\top{}_a{}^b{}_c{}^d.\end{aligned}$$

□

Next we introduce summing over tensors.

Definition D.7. Let Ψ, Φ be tensors, then $\Psi + \Phi = \{\Psi_a{}^b{}_c{}^d + \Phi_a{}^b{}_c{}^d\}_a{}^b{}_c{}^d$.

Lemma 24. *Let $C, B_1, \ldots, B_N$ be matrices, then*

$$\sum_{n=1}^{N}(C \otimes B_n) = (C \otimes \sum_{n=1}^{N} B_n).$$

Proof. We will immediately apply the definition of the tensor outer product (Definition D.6), applying the definition of summing tensors (Definition D.7) N times, using the associative property of $C_a{}^b$ for any given a, b and finally the definition of the tensor product (Definition D.6) again.

$$\begin{aligned}\sum_{n=1}^{N}(C \otimes B_n) &= \sum_{n=1}^{N}\{C_a{}^b B_n^\top{}_c{}^d\}_a{}^b{}_c{}^d\\ &= \{\sum_{n=1}^{N} C_a{}^b B_n^\top{}_c{}^d\}_a{}^b{}_c{}^d\\ &= \{C_a{}^b \left(\sum_{n=1}^{N} B_n^\top\right)_c{}^d\}_a{}^b{}_c{}^d\\ &= (C \otimes \sum_{n=1}^{N} B_n)\end{aligned}$$

□

Lemma 25. *Let Ψ be a tensor and let x, y, v, w be vectors, then*

$$x^\top \Psi(yw^\top)v = \langle \Psi, (yx^\top \otimes vw^\top)\rangle \qquad \textit{and} \qquad \langle wx^\top, \Psi(yv^\top)\rangle = \langle \Psi, (yx^\top \otimes vw^\top)\rangle$$

Proof. First we will write $x^\top \Psi(yw^\top)v$ in tensor notation and then apply the Frobenius inner product (Definition D.4) alongside the definition of the tensor product (Definition D.6)

$$\begin{aligned}
x^\top \Psi(yw^\top)v &= \{x^{\top\, a} \Psi_{a\ c}^{\ b\ \ d} y_{\ b} w^{\top\, c} v_{\ d}\}_{\mathbb{R}} \\
&= \{\Psi_{a\ c}^{\ b\ \ d} x^{\top\, a} y_{\ b} w^{\top\, c} v_{\ d}\}_{\mathbb{R}} \\
&= \langle \Psi, \{x^{\top\, a} y_{\ b} w^{\top\, c} v_{\ d}\}_{b\ d}^{\ a\ c}\rangle \\
&= \langle \Psi, \{(yx^\top)_b^{\ a} (vw^\top)_d^{\ c}\}_{b\ d}^{\ a\ c}\rangle \\
&= \langle \Psi, (yx^\top \otimes vw^\top)\rangle.
\end{aligned}$$

We will show the second statement by applying the common definition of the Frobenius inner product for matrices, followed by applying the first half of the lemma

$$\begin{aligned}
\langle wx^\top, \Psi(yv^\top)\rangle &= \{w_{\ b} x^{\top\, a} \Psi(yv^\top)_{a\ }^{\ b}\}_{\mathbb{R}} \\
&= x^\top \Psi(yv^\top) w \\
&= \langle \Psi, (yx^\top \otimes vw^\top)\rangle.
\end{aligned}$$

□

Next we relate previous definitions and operations to standard linear algebra. In order to do so we need the following definitions, the first of which is the flattening operation that can act on a tensor or matrix.

Definition D.8. Let $\Psi \in \mathbb{R}^{m\times m\times n\times n}$ be a tensor and $A \in \mathbb{R}^{m\times n}$ be a matrix, then

$$\begin{aligned}
\Psi^F{}_{a\ }^{\ b} &= \Psi_{(a \mod m)+1\ \ \lfloor b/m\rfloor+1}^{\ \ (b \mod m)+1\ \ \lfloor a/m\rfloor+1} \\
A^F{}_{a} &= A_{(a \mod m)+1}^{\ \ \lfloor a/m\rfloor+1}
\end{aligned}$$

Furthermore $\Psi^F \in \mathbb{R}^{mn\times mn}$ and $A \in \mathbb{R}^{mn}$.

Note that in the definition above there is a $+1$ is the indexes. These are necessary since our indices start counting from 1 and not from 0.

Similarly, we define how to unflatten a matrix or a vector.

Definition D.9. Let $A \in \mathbb{R}^{mn\times mn}$ be a matrix and $x \in \mathbb{R}^{mn}$ be a vector, then we will define

$$\begin{aligned}
A^{U\ \ \ \ \ \ }_{\ a\ b\ c\ d} &= A_{(a-1)+(d-1)m}^{\ \ \ \ \ \ (b-1)+(c-1)m} \\
x^{U\ \ }_{\ a\ \ b} &= x_{(a-1)+(b-1)m}
\end{aligned}$$

Furthermore $A^U \in \mathbb{R}^{m\times m\times n\times n}$ and $x \in \mathbb{R}^{m\times n}$.

We will not show this but to gain some intuition, observe that flattening the tensor product of two matrices is equal to the Kronecker product of those matrices.

Next, we show that unflattening a flattened tensor recovers the original tensor.

Lemma 26. *Let Ψ be a tensor, then unflattening is the inverse of flattening*

$$(\Psi^F)^U = \Psi.$$

Proof. Fix some a, b, c, d, then we first apply the definition of unflattening (Definition D.9) followed by the definition of flattening (Definition D.8)

$$\begin{aligned}
(\Psi^F)^U{}_{a\ \ c}^{\ b\ \ d} &= \Psi^F{}_{(a-1)+(d-1)m}^{\ \ \ \ \ \ (b-1)+(c-1)m} \\
&= \Psi_{((a-1)+(d-1)m \mod m)+1\ \ (\lfloor (b-1)+(c-1)m/m\rfloor)+1}^{\ \ ((b-1)+(c-1)m \mod m)+1\ \ (\lfloor (a-1)+(d-1)m/m\rfloor)+1} \\
&= \Psi_{a\ \ c}^{\ b\ \ d}.
\end{aligned}$$

□

Lemma 27. *Let Ψ be a tensor, then Ψ^F is symmetric if and only if Ψ is symmetric.*

Proof. Let Ψ be symmetric, we will now show that Ψ^F is symmetric. For that we use the definition of flattening (Definition D.8) alongside the symmetry of Ψ.

$$\Psi^F{}_a{}^b = \Psi_{(a \mod m)+1}{}^{(b \mod m)+1}{}_{\lfloor b/m \rfloor+1}{}^{\lfloor a/m \rfloor+1} = \Psi_{(b \mod m)+1}{}^{(a \mod m)+1}{}_{\lfloor a/m \rfloor+1}{}^{\lfloor b/m \rfloor+1} = \Psi^F{}_b{}^a$$

Now let Ψ^F be symmetric, now we show that Ψ is too. First we use the definition of unflatteing (Definition D.9) alongside the symmetry of Ψ^F to show

$$\Psi_a{}^b{}_c{}^d = \Psi^F{}_{(a-1)+(d-1)m}{}^{(b-1)+(c-1)m} = \Psi^F{}_{(b-1)+(c-1)m}{}^{(a-1)+(d-1)m} = \Psi_b{}^a{}_d{}^c.$$

□

Lemma 28. *Let $\Psi, \Phi \in \mathbb{R}^{m\times m\times n\times n}$ be tensors and $B \in \mathbb{R}^{n\times n}$ a matrix, then Ψ acting on B or Φ is equivalent in the higher or lower dimensional space*

$$\Psi(\Phi) = (\Psi^F \Phi^F)^U$$
$$\Psi(B) = (\Psi^F B^F)^U$$

where $\Psi^F \Phi^F$ is employing the usual matrix multiplication.

Proof. We start with the first claim by using the definition $\Psi(B)$ and $\Phi(\Psi)$ (Definition D.1), writing the sums explicitly, reindexing, applying the definition of flattening (Definition D.8) and recognising a matrix product before finally using the definition of unflattening (Definition D.9)

$$\begin{aligned}
\Psi(\Phi) &= \{\Psi_a{}^b{}_c{}^d \Phi_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d \\
&= \{\sum_{b=1}^{m}\sum_{c=1}^{n} \Psi_a{}^b{}_c{}^d \Phi_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d \\
&= \{\sum_{i=1}^{mn} \Psi_a{}^{(i \mod m)+1}{}_{\lfloor i/m \rfloor+1}{}^d \Phi_{(i \mod m)+1}{}^e{}_f{}^{\lfloor i/m \rfloor+1}\}_a{}^e{}_f{}^d \\
&= \{\sum_{i=1}^{mn} \Psi^F{}_{(a-1)+(d-1)m}{}^i \Phi^F{}_i{}^{(e-1)+(f-1)m)}\}_a{}^e{}_f{}^d \\
&= \{(\Psi^F \Phi^F)_{(a-1)+(d-1)m)}{}^{(e-1)+(f-1)m)}\}_a{}^e{}_f{}^d \\
&= (\Psi^F \Phi^F)^U.
\end{aligned}$$

The second part of the proof follows the exact same steps except recognising a matrix vector product instead of a matrix product in the second to last step

$$\begin{aligned}
\Psi(B) &= \{\Psi_a{}^b{}_c{}^d B_b{}^c\}_a{}^d \\
&= \{\sum_{b=1}^{m}\sum_{c=1}^{n} \Psi_a{}^b{}_c{}^d B_b{}^c\}_a{}^d \\
&= \{\sum_{i=1}^{mn} \Psi_a{}^{(i \mod m)+1}{}_{\lfloor i/m \rfloor+1}{}^d B_{(i \mod m)+1}{}^{\lfloor i/m \rfloor+1}\}_a{}^d \\
&= \{\sum_{i=1}^{mn} \Psi^F{}_{(a-1)+(d-1)m)}{}^i B^F{}_i\}_a{}^d \\
&= \{(\Psi^F B^F)_{(a-1)+(d-1)m)}\}_a{}^d \\
&= (\Psi^F B^F)^U.
\end{aligned}$$

□

Lemma 29. *Let Φ be an invertible tensor, then flattening and inverting can be interchanged, i.e.*

$$(\Psi^{-1})^F = (\Psi^F)^{-1}$$

Proof. We will show that $(\Psi^{-1})^F$ is the inverse of Ψ^F by first using the fact that flattening and unflattening cancel another (Lemma 26), followed by applying $\Psi(\Phi) = (\Psi^F \Phi^F)^U$ (Lemma 28) and the definition of Ψ^{-1} (Definition D.3)

$$\begin{aligned}
\Psi^F(\Psi^{-1})^F &= ((\Psi^F(\Psi^{-1})^F)^U)^F \\
&= (\Psi(\Psi^{-1}))^F \\
&= (\mathcal{I})^F \\
&= I \\
\Rightarrow (\Psi^{-1})^F &= (\Psi^F)^{-1}
\end{aligned}$$

□

Lemma 30. *Let Ψ be a tensor, then transposing and flattening can be interchanged, i. e.*

$$\Psi^{\top F} = \Psi^{F\top}$$

Proof. Fix some a, b. First we apply the definition of flattening (Definition D.8) and then the definition of transposing (Definition D.5). We then proceed using the definition of flattening followed by the definition of transposing one more time.

$$\begin{aligned}
\Psi^{\top F}{}_a{}^b &= \Psi^{\top}{}_{(a \mod m)+1}{}^{(b \mod m)+1}{}_{\lfloor b/m \rfloor + 1}{}^{\lfloor a/m \rfloor + 1} \\
&= \Psi{}_{(b \mod m)+1}{}^{(a \mod m)+1}{}_{\lfloor a/m \rfloor + 1}{}^{\lfloor b/m \rfloor + 1} \\
&= \Psi{}_{(b \mod m)+1}{}^{(a \mod m)+1}{}_{\lfloor a/m \rfloor + 1}{}^{\lfloor b/m \rfloor + 1} \\
&= \Psi^F{}_b{}^a \\
&= \Psi^{F\top}{}_a{}^b
\end{aligned}$$

□

Next we will show two quick facts about how flattening and the tensor product interact

Lemma 31. *Let Ψ be a tensor and let x, y, v, w be vectors, then we can perform the tensor product on x, y, v, w in a lower dimension*

$$(xy^\top \otimes wv^\top) = \left((xw^\top)^F \otimes (vy^\top)^F\right)^U$$

Proof. We start by applying the definition of the tensor product (Definition D.6), followed by the fact that flattening and unflattening cancel another (Lemma 26), then the definition of flattening (Definition D.8), some rearranging before applying the same flattening definition for matrices twice, lastly we apply the tensor product one more time

$$\begin{aligned}
(xy^\top \otimes wv^\top) &= \{x_a y^{\top b} v_c w^{\top d}\}_a{}^b{}_c{}^d \\
&= ((\{x_a y^{\top b} v_c w^{\top d}\}_a{}^b{}_c{}^d)^F)^U \\
&= \left(\{x_{(a \mod m)+1} y^{\top (b \mod m)+1} v_{\lfloor b/m \rfloor + 1} w^{\top \lfloor a/m \rfloor + 1}\}_a{}^b\right)^U \\
&= \left(\{x_{(a \mod m)+1} w^{\top \lfloor a/m \rfloor + 1} v_{\lfloor b/m \rfloor + 1} y^{\top (b \mod m)+1}\}_a{}^b\right)^U \\
&= \left(\{(xw^\top)^F{}_a ((vy^\top)^F)^{\top b}\}_a{}^b\right)^U \\
&= \left((xw^\top)^F \otimes (vy^\top)^F\right)^U.
\end{aligned}$$

□

Lemma 32. *Let Ψ be a tensor and let x, y, v, w be vectors, then we also perform the Frobenius inner product between Ψ and $(xy^\top \otimes wv^\top)$ in a lower dimension*

$$\langle \Psi, (xy^\top \otimes wv^\top) \rangle = ((yv^\top)^F)^\top \Psi^F (xw^\top)^F$$

Proof. First we apply the definition of the Frobenius inner product (Definition D.4), then the definition of the tensor product, (Definition D.6), some rearranging, then we change the indexing by using the definition of flattening (Definition D.8) on Ψ and then twice again on the vectors, finally we recognise the product between two tensors and a matrix

$$\begin{aligned}
\langle \Psi, (xy^\top \otimes wv^\top) \rangle &= \{\Psi_a{}^b{}_c{}^d (xy^\top \otimes wv^\top)_b{}^a{}_d{}^c\}_{\mathbb{R}} \\
&= \{\Psi_a{}^b{}_c{}^d x_b y^{\top a} v_d w^{\top c}\}_{\mathbb{R}} \\
&= \{\Psi_a{}^b{}_c{}^d y^{\top a} x_b w^{\top c} v_d\}_{\mathbb{R}} \\
&= \{\Psi^F{}_a{}^b y^{\top (a \mod m)+1} x_{(b \mod m)+1} w^{\top \lfloor b/m \rfloor + 1} v_{\lfloor a/m \rfloor + 1}\}_{\mathbb{R}} \\
&= \{\Psi^F{}_a{}^b ((yv^\top)^F)^{\top a} (xw^\top)^F{}_b\}_{\mathbb{R}} \\
&= ((yv^\top)^F)^\top \Psi^F (xw^\top)^F.
\end{aligned}$$

□

Lemma 33. *Let B be a matrix and x, y be vectors, then*

$$x^\top B y = (xy^\top)^{F\top} B^F$$

Proof. First we write $x^\top B$ in the tensor notation and regroup y and $x^\top$ to a single matrix. Then we re-index by writing the sums explicitly. Now we can transpose $yx^\top$ and use the flatten operator (Definition D.8)). Finally we write in Einstein notation again and conclude by recognising the quantity on the left-hand side.

$$\begin{aligned}
x^\top B y &= \{x^{\top a} B_a{}^b y_b\}_{\mathbb{R}} \\
&= \{(yx^\top)_b{}^a B_a{}^b\}_{\mathbb{R}} \\
&= \sum_i (yx^\top)_{\lfloor i/m \rfloor + 1}{}^{(i \mod m)+1} B_{(i \mod m)+1}{}^{\lfloor i/m \rfloor + 1} \\
&= \sum_i (xy^\top)_{(i \mod m)+1}{}^{\lfloor i/m \rfloor + 1} B_{(i \mod m)+1}{}^{\lfloor i/m \rfloor + 1} \\
&= \sum_i (xy^\top)^F{}_i B^F{}_i \\
&= \{(xy^\top)^{F\top}{}^i B^F{}_i\}_{\mathbb{R}} \\
&= (xy^\top)^{F\top} B^F.
\end{aligned}$$

□

Next we will define eigenmatrices and eigenvalues for our definition of tensors.

Definition D.10. We will call a scalar $\lambda \in \mathbb{R}$ an eigenvalue of a tensor Ψ if there exists a matrix B such that

$$\Psi(B) = \lambda B.$$

Such a matrix B is then called an eigenmatrix of Ψ.

In the next Lemma we will show that flattened tensors and tensors have the same eigenvalues, i.e. Ψ and Ψ^F have the same eigenvalues.

Lemma 34.
Let $\lambda'_1, \ldots, \lambda'_{k'}$ be the eigenvalues of Ψ with eigenmatrices $B'_1, \ldots, B'_{k'}$ and let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of Ψ^F with eigenvectors $B_1, \ldots, B_k$, then $k = k'$ and there exists a permutation $\sigma \in S_k$ such that

$$\begin{aligned}
\lambda_1, \ldots, \lambda_k &= \lambda'_{\sigma(1)}, \ldots, \lambda'_{\sigma(k)} \\
B_1, \ldots, B_k &= B'^F_{\sigma(1)}, \ldots, B'^F_{\sigma(k)}.
\end{aligned}$$

Proof. Let all variables be defined as in the lemma. We now show that all eigenmatrices of Ψ are eigenvectors of Ψ^F, thus let B' be an eigenmatrix of Ψ with eigenvalue λ'

$$\Psi^F(B'^F) = (\Psi(B'))^F = (\lambda' B')^F = \lambda'(B')^F,$$

where we used the fact that an operation can be performed in lower or higher dimensional space (Lemma 28). Next we will use the same lemma again to conclude that all eigenvectors of Ψ^F are eigenmatrices of Ψ if unflattened and thus let B be an eigenvector of Ψ^F and λ the corresponding eigenvalue

$$\Psi(B) = (\Psi^F(B^F))^U = (\lambda B^F)^U = \lambda B.$$

Since all eigenvectors of Ψ^F correspond to an eigenmatrix of Ψ and all eigenmatrices of Ψ^F correspond to an eigenvector of Ψ^F, it is clear that $k = k'$ and that there exists an $\sigma \in S_k$ such that $\lambda_1, \dots, \lambda_k = \lambda'_{\sigma(1)}, \dots, \lambda'_{\sigma(k)}$ and $B_1, \dots, B_k = B'^F_{\sigma(1)}, \dots, B'^F_{\sigma(k)}$. □

Lemma 35. *Let $D \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{n\times n}$ be matrices with eigenvalues $\lambda_1, \dots, \lambda_m$ and $\lambda'_1, \dots, \lambda'_n$ respectively, then the eigenvalues of $(D \otimes B)$ will be*

$$\lambda_i \lambda'_j, \forall i \in [m], j \in [n].$$

Proof. By Lemma 34 we can conclude that $(D \otimes B)$ has mn eigenvalues in total and we will show that $\lambda_i \lambda'_j$ for each of the mn combinations of $i \in [n]$ and $j \in [m]$ is an eigenvalue of $(D \otimes B)$.
Fix some $i \in [n]$ and $j \in [m]$ with eigenvectors x_i and x'_j, then use the fact that $ACB^\top = (D \otimes B)(C)$ (Lemma 7) to show

$$\begin{aligned}(D \otimes B)((x_i \otimes x'_j)) &= D(x_i \otimes x'_j)B^\top \\ &= Dx_i {x'_j}^\top B^\top \\ &= Dx_i (Bx'_j)^\top \\ &= \lambda_i x_i (\lambda'_j x'_j)^\top \\ &= \lambda_i \lambda'_j (x_i \otimes x'_j).\end{aligned}$$

□

Definition D.11. We will define a notion of the trace for tensors as follows

$$\operatorname{tr}(\Psi) = \{\Psi_a{}^a{}_c{}^c\}_{\mathbb{R}}$$

Similarly, the trace for a matrix B can be written in tensor notation as

$$\operatorname{tr}(B) = \{B_a{}^a\}_{\mathbb{R}}$$

Lemma 36. *Let $\Psi \in \mathbb{R}^{m\times m\times n\times n}$ be a tensor, then the trace of a flattened tensor is equivalent to the trace of the tensor*

$$\operatorname{tr}(\Psi) = \operatorname{tr}(\Psi^F).$$

And if $A \in \mathbb{R}^{m\times m}$ and $C \in \mathbb{R}^{n\times n}$ are some matrices then the trace of the outer product is the product of the traces of the matrices.

$$\operatorname{tr}((B \otimes C)) = \operatorname{tr}(B)\operatorname{tr}(C)$$

where the trace on the right-hand side of either equation is the classic trace for matrices.

Proof. The first result can be shown by writing out the definition of the trace (Definition D.11) followed by reindexing and applying the definition of flattening (Definition D.8) alongside the definition of the trace for matrices

$$\operatorname{tr}(\Psi) = \sum_{a=1}^{m}\sum_{c=1}^{n} \Psi_a{}^a{}_c{}^c = \sum_{a=1}^{mn} \Psi_{(a \mod m)+1}{}^{(a \mod m)+1}{}_{\lfloor a/m \rfloor+1}{}^{\lfloor a/m \rfloor+1} = \operatorname{tr}(\Psi^F).$$

The second result follows by applying the definition of the trace (Definition D.11) alongside the definition of the tensor product (definitionD.6), followed by applying the definition of the trace for matrices twice

$$\begin{aligned}\operatorname{tr}((B \otimes C)) &= \sum_{a=1}^{m}\sum_{c=1}^{n} A_a{}^a B_c{}^c \\ &= \operatorname{tr}(B)\sum_{c=1}^{n} B_c{}^c \\ &= \operatorname{tr}(B)\operatorname{tr}(C).\end{aligned}$$

□

Next we will see how the Frobenius inner product and the trace interact

Lemma 37. *Let Ψ, Φ be tensors, then Frobenius product can be written as a trace*

$$\langle \Psi, \Phi \rangle = \operatorname{tr}(\Psi(\Phi)).$$

Proof. We start from the trace and apply the definition of one tensor being applied to another tensor (Definition D.1), then the definition of the trace (Definition D.11), followed by the definition of the Frobenius product (Definition D.4)

$$\begin{aligned}\operatorname{tr}(\Psi(\Phi)) &= \operatorname{tr}(\{\Psi_a{}^b{}_c{}^d \Phi_b{}^e{}_f{}^c\}_a{}^e{}_f{}^d) \\ &= \{\Psi_a{}^b{}_c{}^d \Phi_b{}^a{}_d{}^c\}_{\mathbb{R}} \\ &= \langle \Psi, \Phi \rangle.\end{aligned}$$

□

Lemma 38. *Let B, C be matrices, then*

$$\operatorname{tr}(B^T C) = \operatorname{tr}(BC^T)$$

Proof. First we apply the definition of matrix multiplication (Definition D.1), then the common definition of the trace for matrices. We then transpose both B and C which does not change anything and then we reassemble by executing the first two steps backwards.

$$\begin{aligned}\operatorname{tr}(B^T C) &= \operatorname{tr}(\{B_a^T{}^b B_b{}^c\}_a{}^c) \\ &= \{B_a^T{}^b B_b{}^a\}_{\mathbb{R}} \\ &= \{B_b{}^a C_a^T{}^b\}_{\mathbb{R}} \\ &= \operatorname{tr}(\{B_b{}^a C_a^T{}^b\}_a{}^c) \\ &= \operatorname{tr}(BC^T).\end{aligned}$$

□

Lemma 39. *Let Ψ be a tensor and B, C be matrices, then*

$$\operatorname{tr}(\Psi(B) \cdot C) = \langle B^\top, \Psi^\top(C^\top) \rangle$$

where $\cdot$ is used to express matrix multiplication between B and C.

Proof. We start by using the definition of $\Psi(B)$ (Definition D.1), immediately followed by the definitions for matrix multiplication and the trace (Definition D.11). Then we transpose, recognise the definition of $\Psi(B)$ again and finally apply

Sampling Scheme 5 Tensor-Exp3 Sampling Scheme

Require: Context distribution $\mathcal{D}$, current policy π_t
1: Draw $X \sim \mathcal{D}$
2: Draw $A \sim \pi_t(\cdot|X)$
3: **Output** $(XX^\top \otimes AA^\top)^F$

the definition of the Frobenius product (Definition D.4)

$$\begin{aligned}
\operatorname{tr}(\Psi(B) \cdot C) &= \operatorname{tr}(\{\Psi_a{}^b{}_c{}^d B_b{}^c\}_a{}^d \cdot C) \\
&= \operatorname{tr}(\{\Psi_a{}^b{}_c{}^d B_b{}^c C_d{}^e\}_a{}^e) \\
&= \{\Psi_a{}^b{}_c{}^d B_b{}^c C_d{}^a\}_{\mathbb{R}} \\
&= \{\Psi^\top{}_b{}^a{}_d{}^c B^\top{}_c{}^b C^\top{}_a{}^d\}_{\mathbb{R}} \\
&= \{(\Psi^\top(C^\top))_b{}^c B^\top{}_c{}^b\}_{\mathbb{R}} \\
&= \langle B^\top, \Psi^\top(C^\top)\rangle.
\end{aligned}$$

□

E FULL-BANDIT EXP3 PROOF DETAILS

For the exploration we will be using the Kiefer-Wolfowitz theorem used in Bubeck et al. (2012), which we will restate in slightly changed notation here.

Theorem 40 (Kiefer–Wolfowitz Theorem, Lattimore and Szepesvári (2020, Theorem 21.1))**.** *Let $\mathcal{G} \subset \mathbb{R}^K$ be a convex set with span$(\mathcal{G}) = \mathbb{R}^K$. Then there exists a probability distribution over the points of $\mathcal{G}$, $\mu_g \in \mathbb{R}$ for all $g \in \mathcal{G}$ with $\mu \in \Delta_K$ such that*

$$K = \max_{g \in \mathcal{G}} \left\|g\right\|^2_{V^{-1}},$$

where $V = \sum_{g \in \mathcal{G}} \mu_g g g^\top$. Furthermore, $\lambda_{\min}(V) \geq \frac{K}{m}$.

Proof. A μ_g such that $K = \max_{g \in \mathcal{G}} \left\|g\right\|^2_{V^{-1}}$ with V as above exists by Lattimore and Szepesvári (2020, Theorem 21.1). Left to prove is $\lambda_{\min}(V) \geq \frac{K}{m}$. We apply $K = \max_{g \in \mathcal{G}} \left\|g\right\|^2_{V^{-1}}$ and continue as follows

$$K = \max_{g \in \mathcal{G}} g^\top V^{-1} g = m \max_{g \in \mathcal{G}} \frac{g^\top}{\sqrt{m}} V^{-1} \frac{g}{\sqrt{m}} \leq m \lambda_{\max}\left(V^{-1}\right) = m \lambda_{\min}\left(V\right),$$

where we used the fact that $\lambda_{\max}(B) = \max_{||x||_2 = 1} x^\top B x$. Dividing by m provides the desired result. □

Lemma 41. $\Psi_t = \mathbb{E}_{X_t, A_t}\left[\left(X_t X_t^\top \otimes A_t A_t^\top\right) \middle| \mathcal{F}_{t-1}\right]$ *is invertible.*

Proof. We know from Lemma 11 that $\lambda_{\min}(\Psi_t^F) \geq \frac{\gamma K \lambda_{\min}(\Sigma)}{m} > 0$. It thus follows that Ψ_t^F is invertible and we conclude the proof by using the fact that Ψ_t is invertible if Ψ_t^F is (Lemma 29). □

We explicitly define a sampling scheme usable by the MGR for the full-bandit case in Sampling Scheme 5.

Lemma 42. *Samples generated by the sampling method detailed in Sampling Scheme 5 are unbiased samples of Ψ_t.*

Proof. To show that $(XX^\top \otimes AA^\top)$ is indeed an unbiased sample of Ψ_t it is sufficient to take the expectation over $(XX^\top \otimes AA^\top)$ explicitly

$$\mathbb{E}_{X \sim \mathcal{D}, A \sim \pi_t(\cdot|X)}\left[(XX^\top \otimes AA^\top)\right] = \Psi_t.$$

□

Lemma 10. *Fix any $x \in \mathcal{X}$ and suppose that $\widetilde{\Theta}_t$ and $\eta > 0$ are such that $\max_t \eta |x^\top \widetilde{\Theta}_t A| < 1$ for all $A \in \mathcal{A}$. Then the regret of Algorithm 3 in the auxiliary game at x satisfies*

$$\widehat{\mathcal{R}}_T(x) \leq \frac{\ln(|\mathcal{A}|)}{\eta} + \gamma\, U_T(x) + \eta \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_{A \sim \pi_t(\cdot|x)}\left[(x^\top \widetilde{\Theta}_t A)^2 | \mathcal{F}_{t-1}\right]\right]$$

where $U_T(x) = \sum_{t=1}^T \sum_{A \in \mathcal{A}} \mu_A x^\top \widetilde{\Theta}_t \big(A - \pi_T^*(x)\big)$ and μ is the distribution on $\mathcal{A}$ defined by the Kiefer-Wolfowitz theorem.

Proof. The proof will follow the classical Exp3 analysis. By recognising that $p_t(A) \propto w_t(x, A)$ is the exponential weights distribution we can apply Van der Hoeven et al. (2018, Lemma 1) to find

$$\begin{aligned}
&\sum_{t=1}^T x^\top \widetilde{\Theta}_t \Big(\Big(\sum_{A \in \mathcal{A}} \pi_t(A|x) A \Big) - \pi^\star(x) \Big) \\
&= (1-\gamma) \sum_{t=1}^T x^\top \widetilde{\Theta}_t \Big(\sum_{A \in \mathcal{A}} p_t(A) A \Big) - \pi^\star(x)) + \gamma \underbrace{\sum_{t=1}^T \sum_{A \in \mathcal{A}} \mu_A x^\top \widetilde{\Theta}_t \left(A - \pi_T^*(x)\right)}_{U_T(x)} \\
&\leq (1-\gamma) \Big(\frac{\ln(|\mathcal{A}|)}{\eta} + \sum_{t=1}^T x^\top \widetilde{\Theta}_t \Big(\sum_{A \in \mathcal{A}} p_t(A) A \Big) + \frac{1}{\eta} \ln \Big(\sum_{A \in \mathcal{A}} p_t(A) \exp\big(-\eta x^\top \widetilde{\Theta}_t A \big) \Big) \Big) + \gamma U_T(x). \qquad (17)
\end{aligned}$$

Since by assumption $\eta |x \widetilde{\Theta}_t A| \leq 1$ we may apply $\exp(-z) \leq 1 - z + z^2$ for $|z| \leq 1$ to find

$$\begin{aligned}
&x^\top \widetilde{\Theta}_t \Big(\sum_{A \in \mathcal{A}} p_t(A) A \Big) + \frac{1}{\eta} \ln \Big(\sum_{A \in \mathcal{A}} p_t(A) \exp\big(-\eta x^\top \widetilde{\Theta}_t A \big) \Big) \\
&\leq x^\top \widetilde{\Theta}_t \Big(\sum_{A \in \mathcal{A}} p_t(A) A \Big) + \frac{1}{\eta} \ln \Big(1 - \sum_{A \in \mathcal{A}} p_t(A) \eta x^\top \widetilde{\Theta}_t A + \eta^2 \sum_{A \in \mathcal{A}} p_t(A) (x^\top \widetilde{\Theta}_t A)^2 \Big) \Big) \\
&\leq \eta \sum_{A \in \mathcal{A}} p_t(A) (x^\top \widetilde{\Theta}_t A)^2,
\end{aligned}$$

where the last inequality is because $\ln(1+z) \leq z$ for $|z| \leq 1$. Using the above inequality in equation (17) we find

$$\begin{aligned}
\sum_{t=1}^T x^\top \widetilde{\Theta}_t \Big(\Big(\sum_{A \in \mathcal{A}} \pi_t(A|x) A \Big) - \pi^\star(x) \Big) &\leq \frac{\ln(|A|)}{\eta} + \eta \sum_{t=1}^T \sum_{A \in \mathcal{A}} (1-\gamma) p_t(A) (x^\top \widetilde{\Theta}_t A)^2 + \gamma U_T(x) \\
&\leq \frac{\ln(|A|)}{\eta} + \eta \sum_{t=1}^T \mathbb{E}_{A \sim \pi_t(\cdot|x)} (x^\top \widetilde{\Theta}_t A)^2 + \gamma U_T(x),
\end{aligned}$$

which completes the proof after taking expectations.

□

Lemma 11. *For all $t \geq 1$,*

$$\lambda_{\min}(\Psi_t^F) \geq \frac{\gamma K \lambda_{\min}(\Sigma)}{m}$$

Moreover, for $\eta \leq \frac{1}{m(M+1)}$, any $A \in \mathcal{A}$, and any x in the support of $\mathcal{D}$ it also holds that $\eta \big| x^\top \widetilde{\Theta}_t A \big| < 1$.

Proof. Let Ψ_t be as defined in equation (7), then

$$\begin{aligned}
\lambda_{\min}(\Psi_t^F) &= \lambda_{\min}(\Psi_t) \\
&= \lambda_{\min}\left(\mathbb{E}_{X_t, A_t}[(X_t X_t^\top \otimes A_t A_t^\top) | \mathcal{F}_{t-1}] \right) \\
&= \lambda_{\min}\left(\mathbb{E}_{X_t}[\mathbb{E}_{A_t}[(X_t X_t^\top \otimes A_t A_t^\top) | X_t, \mathcal{F}_{t-1}]] \right)
\end{aligned}$$

where we first used the fact that Ψ_t and Ψ_t^F agree on eigenvalues (Lemma 34) and plugged in the definition of Ψ_t (equation (7)).

Next we will write the expectation over A_t explicitly, plug in the definition of π_t (equation (10)) and use the fact that

$$\sum_{A_t \in A} (1-\gamma) \frac{w_t(X_t, A_t)}{\sum_{A' \in \mathcal{A}} w_t(X_t, A')} \geq 0$$

alongside the fact that $(X_t X_t^\top \otimes A_t A_t^\top)$ only has positive eigenvalues to continue like follows

$$\begin{aligned}
\lambda_{\min}(\Psi_t^F) &= \lambda_{\min}\left(\mathbb{E}_{X_t}\left[\sum_{A \in A} \pi_t(A|X_t)(X_t X_t^\top \otimes A A^\top)\right]\right) \\
&= \lambda_{\min}\left(\mathbb{E}_{X_t}\left[\sum_{A \in A}\left((1-\gamma)\frac{w_t(X_t, A)}{\sum_{A' \in \mathcal{A}} w_t(X_t, A')} + \gamma \mu_A\right)(X_t X_t^\top \otimes A A^\top)\right]\right) \\
&\geq \lambda_{\min}\left(\mathbb{E}_{X_t}\left[\sum_{A \in A} \gamma \mu_A (X_t X_t^\top \otimes A A^\top)\right]\right)
\end{aligned}$$

Now the only thing that is left to do is to apply Lemma 24, use Lemma 35 to recognise that $\lambda_{\min}((A \otimes B)) = \lambda_{\min}(A)\lambda_{\min}(B)$ and use the Kiefer-Wolfowitz theorem (Theorem 40).

$$\begin{aligned}
\lambda_{\min}(\Psi_t^F) &\geq \lambda_{\min}\left(\mathbb{E}_{X_t}\left[\sum_{A \in \mathcal{A}} \gamma \mu_A (X_t X_t^\top \otimes A A^\top)\right]\right) \\
&= \gamma \lambda_{\min}\left(\mathbb{E}_{X_t}[X_t X_t^\top] \otimes \sum_{A \in \mathcal{A}} \mu_A A A^\top\right) \\
&= \gamma \lambda_{\min}(\Sigma) \lambda_{\min}\left(\sum_{A \in \mathcal{A}} \mu_A A A^\top\right) \\
&= \frac{\gamma K \lambda_{\min}(\Sigma)}{m}
\end{aligned}$$

We now prove the second claim of the lemma. Let $A \in \mathcal{A}$ and x in the support of $\mathcal{D}$. We can start by writing the in the definition of $\widetilde{\Theta}_t$, then upper bounding $X_t^\top \Theta_t A_t$ by m and then using Lemma 25.

$$\begin{aligned}
|x^\top \widetilde{\Theta}_t A| &= \left|x^\top \widehat{\Psi}_t^+ (X_t X_t^\top \Theta_t A_t A_t^\top) A\right| \\
&\leq m \left|x^\top \widehat{\Psi}_t^+ (X_t A_t^\top) A\right| \\
&= m \left|((x^\top A)^F)^\top (\widehat{\Psi}_t^+)^F (X_t A_t^\top)^F\right|.
\end{aligned}$$

Now, observe that for any x in the support of $\mathcal{D}$ and $A \in \mathcal{A}$

$$\left\|(x A^\top)^F\right\|_2 = \sqrt{\sum_{i=1}^{d}\sum_{k=1}^{K}(x)_i^2 (A)_k^2} = \sqrt{\left(\sum_{i=1}^{d}(x)_i^2\right)\left(\sum_{k=1}^{K}(A)_k^2\right)} \leq \sigma\sqrt{m}. \tag{18}$$

By using that $\left\|(\widehat{\Psi}_t^+)^F\right\|_{\text{op}} \leq (M+1)\beta$ by Lemma 1, equation (18), and $\beta \leq \frac{1}{m\sigma^2}$ we can see that

$$\begin{aligned}
|x^\top \widetilde{\Theta}_t A| &\leq m \left|((x^\top A)^F)^\top (\widehat{\Psi}_t^+)^F (X_t A_t^\top)^F\right| \\
&\leq m \left\|(x A^\top)^F\right\|_2 \left\|(\widehat{\Psi}_t^+)^F\right\|_{\text{op}} \left\|(X_t A_t^\top)^F\right\|_2 \\
&\leq m^2 \sigma^2 \beta (M+1) \\
&\leq m(M+1).
\end{aligned}$$

Using the fact that $\eta \leq \frac{2}{m(M+1)}$ allows us to conclude that $\eta |X^\top \widetilde{\Theta}_t A| \leq 1$ which finishes the proof $\square$

Lemma 12. *Fix a $t \in [T]$ and let $A_0 \sim \pi_t(\cdot|X_0)$. Then*

$$\mathbb{E}\left[(X_0^\top \widetilde{\Theta}_t A_0)^2 \,\middle|\, \mathcal{F}_{t-1}\right] \leq 2m^2 Kd$$

Proof. Most of this proof will be technical calculations, starting from the beginning by plugging in the definition of $\widetilde{\Theta}_t$ and upper bounding $(X_t^\top \Theta_t A_t)^2$ by m^2

$$\begin{aligned}
&\mathbb{E}\left[(X_0^\top \widetilde{\Theta}_t A_0)^2 | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[(X_0^\top \widehat{\Psi}_t^+(X_t X_t^\top \Theta_t A_t A_t^\top) A_0)^2 | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[(X_t^\top \Theta_t A_t)^2 (X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0)^2 | \mathcal{F}_{t-1}\right] \\
&\leq m^2 \mathbb{E}\left[(X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0)^2 | \mathcal{F}_{t-1}\right]
\end{aligned}$$

Next we simplify $(X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0)^2$ in isolation. We do so by first expanding the square and then using the fact that $x^\top y = \mathrm{tr}(xy^\top)$. Then we use the fact that $DCB^\top = (D \otimes B)(C)$ by Lemma 7 to obtain

$$\begin{aligned}
(X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0)^2 &= X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0 A_0^\top (\widehat{\Psi}_t^+(X_t A_t^\top))^\top X_0 \\
&= \mathrm{tr}\left(X_0 X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0 A_0^\top (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right) \\
&= \mathrm{tr}\left((X_0 X_0^\top \otimes A_0 A_0^\top)(\widehat{\Psi}_t^+(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right).
\end{aligned}$$

Here $\cdot$ denotes the classic matrix multiplication and is used to emphasis that the tensor $(X_0 X_0^\top \otimes A_0 A_0^\top)$ is acting on $\widehat{\Psi}_t^+(X_t A_t^\top)$.

We now use this result together with the fact that X_t, A_t and X_0, A_0 are independent alongside the definition of Ψ_t as follows

$$\begin{aligned}
&\mathbb{E}\left[(X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top) A_0)^2 | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\mathrm{tr}\left((X_0 X_0^\top \otimes A_0 A_0^\top)(\widehat{\Psi}_t^+(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right) | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\mathrm{tr}\left(\mathbb{E}_{X_0, A_0}\left[(X_0 X_0^\top \otimes A_0 A_0^\top)\right](\widehat{\Psi}_t^+(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right) | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\mathrm{tr}\left(\Psi_t(\widehat{\Psi}_t^+(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right) | \mathcal{F}_{t-1}\right].
\end{aligned}$$

This is only possible as $X_0 \sim \mathcal{D}$ and $A_0 \sim \pi_t(\cdot|X_0)$, as per assumption on A_0.

We isolate the term inside the expectation again, $\mathrm{tr}\left(\Psi_t(\widehat{\Psi}_t^+(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right)$, and use the fact that $\mathrm{tr}(BC^\top) = \mathrm{tr}(B^\top C)$ (Lemma 38). We then use $\mathrm{tr}(\Psi(B) \cdot C) = \langle B^\top, \Psi^\top(C^\top)\rangle$ (Lemma 39). We finish by applying the fact that $\langle wx^\top, \Psi(yv^\top)\rangle = \langle \Psi, (yx^\top \otimes vw^\top)\rangle$ (Lemma 25) alongside the fact that tensors are associative (Lemma 20)

$$\begin{aligned}
\mathrm{tr}\left(\Psi_t(\widehat{\Psi}_t^+(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right) &= \mathrm{tr}\left((\Psi_t(\widehat{\Psi}_t^+(X_t A_t^\top)))^\top \cdot \widehat{\Psi}_t^+(X_t A_t^\top)\right) \\
&= \langle (A_t X_t^\top), \widehat{\Psi}_t^{+\top}(\Psi_t(\widehat{\Psi}_t^+(X_t A_t^\top)))\rangle \\
&= \langle \widehat{\Psi}_t^{+\top}(\Psi_t(\widehat{\Psi}_t^+)), (X_t X_t^\top \otimes A_t A_t^\top)\rangle.
\end{aligned}$$

By using the above equality we can thus see that

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{tr}\left(\Psi_t(\widehat{\Psi}_t^{+\top}(X_t A_t^\top)) \cdot (\widehat{\Psi}_t^+(X_t A_t^\top))^\top\right) | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\langle \widehat{\Psi}_t^{+\top}(\Psi_t(\widehat{\Psi}_t^+)), (X_t X_t^\top \otimes A_t A_t^\top)\rangle | \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}_{\mathrm{MGR}_t}\left[\left\langle \widehat{\Psi}_t^{+\top}(\Psi_t(\widehat{\Psi}_t^+)), \Psi_t \right\rangle \,\middle|\, \mathcal{F}_{t-1}\right],
\end{aligned}$$

where in the last equality we used the linearity of the inner product. Observe that by Lemma 23 Ψ_t is symmetric and thus $\Psi_t = \Psi_t^\top$. Now, using that $\mathrm{tr}(\Psi(\Phi)) = \langle \Psi, \Phi \rangle$ for any tensors of appropriate dimension Φ, Ψ by Lemma 37 and by Lemma 30 we have that

$$\begin{aligned}
&\mathbb{E}_{\mathrm{MGR_t}}\left[\left\langle \widehat{\Psi}_t^{+\top}(\Psi_t(\widehat{\Psi}_t^+)), \Psi_t \right\rangle \middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}_{\mathrm{MGR_t}}\left[\mathrm{tr}(\Psi_t^\top(\widehat{\Psi}_t^{+\top}(\Psi_t(\widehat{\Psi}_t^+))))\middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}_{\mathrm{MGR_t}}\left[\mathrm{tr}(\Psi_t^{\top F}\widehat{\Psi}_t^{+\top F}\Psi_t^F\widehat{\Psi}_t^{+F})\middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}_{\mathrm{MGR_t}}\left[\mathrm{tr}(\Psi_t^{F\top}\widehat{\Psi}_t^{+F\top}\Psi_t^F\widehat{\Psi}_t^{+F})\middle| \mathcal{F}_{t-1}\right] \leq 2Kd
\end{aligned}$$

where the last equality is due to Lemma 30, which states that we may switch flattening and transpose operations, and the inequality is due to Lemma 1.

By collecting the results we can bound

$$\mathbb{E}\left[(X_0^\top \widetilde{\Theta}_t A_0)^2 | \mathcal{F}_{t-1}\right] \leq m^2 \mathbb{E}\left[(X_0^\top \widehat{\Psi}_t^+(X_t A_t^\top)A_0)^2 | \mathcal{F}_{t-1}\right] \leq 2m^2 Kd,$$

which concludes the proof. □

Theorem 43. *For any positive $\eta \leq \frac{2}{m(M+1)}$, $\beta \leq \frac{1}{\sigma^2 m}$ and any $\gamma \in (0,1)$ the expected regret of the algorithm satisfies*

$$R_T \leq 2\gamma T m + \eta T m^2 dK + \frac{\ln(|\mathcal{A}|)}{\eta} + 3T\sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma\lambda_{\min}(\Sigma)}{K}\right).$$

Furthermore, let $\beta = \frac{1}{\sigma^2 m}$, $\gamma = \min\left\{1, \sqrt{K\ln(T)\frac{\ln(|\mathcal{A}|)}{T\beta\lambda_{\min}(\Sigma)}}\right\}$, $M = \max\left\{\frac{K\ln(T)}{\beta m\lambda_{\min}(\Sigma)}, \sqrt{\frac{TK\ln(T)}{\ln(|\mathcal{A}|)\beta\lambda_{\min}(\Sigma)}}\right\}$, and $\eta = \min\left\{\frac{1}{m(M+1)}, \sqrt{\frac{\ln(|\mathcal{A}|)}{Tm^2Kd}}\right\}$. Then

$$\begin{aligned}
\mathcal{R}_T \leq\ & 2\sqrt{\ln(eK)Tm^3Kd} + 2\sqrt{m^4TK\frac{\sigma^2\ln(eK)\ln(T)}{\lambda_{\min}(\Sigma)}} + \sqrt{m^4\sigma^2\frac{TK\ln(T)\ln(eK)}{\lambda_{\min}(\Sigma)}} \\
&+ 3\sigma\sqrt{m}G + m^2\ln(eK) + \frac{m^3 2\sigma^2 K\ln(T)\ln(eK)}{\lambda_{\min}(\Sigma)}.
\end{aligned}$$

Proof. Most of the work has been done in the previous lemmas already, now we only need to assemble them correctly, control the bias and check the conditions on the hyperparameters. Starting from Lemma 2

$$\mathcal{R}_T = \mathbb{E}_{\mathcal{F}_T, X_0}\left[\widehat{\mathcal{R}}_T(X_0)\right] + 2\sum_{t=1}^T \max_{A\in\mathcal{A}} \left|\mathbb{E}_{X_0, X_t, A_t}[X_0 B_t A | \mathcal{F}_{t-1}]\right|$$

Lemma 9. Suppose that $\beta \leq \frac{1}{\lambda_{\max}(\Psi_t^F)}$. Then for $\widetilde{\Theta}_t$ defined in equation (8)

$$\mathbb{E}\big[X_0^\top(\Theta_t - \widetilde{\Theta}_t)A \mid \mathcal{F}_{t-1}\big] \leq \sigma G\sqrt{m}e^{-\frac{M\beta\gamma m}{K}\lambda_{\min}(\Sigma)}$$

Proof. We will need to find the exact expectation of $\widehat{\Psi}_t^+$ and we will do that here

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_t}[\widetilde{\Theta}_t | \mathcal{F}_{t-1}] &= \mathbb{E}_{\mathcal{F}_t}[\widehat{\Psi}_t^+\big(X_t X_t^\top \Theta_t A_t A_t^\top\big) | \mathcal{F}_{t-1}] \\
&= \mathbb{E}_{\mathrm{MGR_t}}[\widehat{\Psi}_t^+(\Psi_t(\Theta_t)) | \mathcal{F}_{t-1}] \\
&= (\mathbb{E}_{\mathrm{MGR_t}}[\widehat{\Psi}_t^{+F} | \mathcal{F}_{t-1}]\Psi_t^F\Theta_t^F)^U \\
&= (\Theta_t^F - (I - \beta\Psi_t^F)^M\Theta_t^F)^U
\end{aligned}$$

Next we plug in the definition of B_t, use the above equation, use Lemma 33 and finally use $\left\|(X_0 A^\top)^F\right\|_2 \leq \sigma\sqrt{m}$ (equation (18)) alongside $\left\|\Theta_t\right\|_F \leq G$.

$$\begin{aligned}
\max_{A\in\mathcal{A}} \left|\mathbb{E}[X_0^\top B_t A|\mathcal{F}_{t-1}]\right| &= \max_{A\in\mathcal{A}} \left|\mathbb{E}[X_0^\top(\Theta_t - \widetilde{\Theta}_t)A|\mathcal{F}_{t-1}]\right| \\
&= \max_{A\in\mathcal{A}} \left|\mathbb{E}[X_0^\top(\Theta_t - (\Theta_t^F - (I-\beta\Psi_t^F)^M\Theta_t^F)^U)A]\right| \\
&= \max_{A\in\mathcal{A}} \left|\mathbb{E}[X_0^\top((I-\beta\Psi_t^F)^M\Theta_t^F)^U A]\right| \\
&= \max_{A\in\mathcal{A}} \left|\mathbb{E}[(X_0A^\top)^{F\top}(I-\beta\Psi_t^F)^M\Theta_t^F]\right| \\
&\leq \max_{A\in\mathcal{A}} \left|\mathbb{E}\left[\left\|(X_0A^\top)^F\right\|_2\left\|(I-\beta\Psi_t^F)^M\right\|_{\mathrm{op}}\left\|\Theta_t^F\right\|_2\right]\right| \\
&\leq \sigma\sqrt{m}G\left\|(I-\beta\Psi_t^F)^M\right\|_{\mathrm{op}}
\end{aligned}$$

Next we need to bound $\left\|(I-\beta\Psi_t^F)^M\right\|_{\mathrm{op}}$ for which we will first use $\lambda_{\min}(\Psi_t^F) \geq \frac{\gamma K\lambda_{\min}(\Sigma)}{m}$ (Lemma 11) alongside the fact that Ψ_t^F is positive semi-definite. Then we apply $1-z \leq e^{-z}$ (which holds for all $z\in\mathbb{R}$).

$$\left\|(I-\beta\Psi_t^F)^M\right\|_{\mathrm{op}} \leq \left(1-\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right)^M \leq \exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right)$$

Thus we can now follow that

$$\max_{A\in\mathcal{A}} \left|\mathbb{E}[X_0 B_t A|\mathcal{F}_{t-1}]\right| \leq \sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right) \tag{19}$$

□

Next we need to bound $U_T(x) = \sum_{t=1}^T \sum_{A\in\mathcal{A}} \mu_A x^\top \widetilde{\Theta}_t (A - \pi_T^*(x))$ in expectation. First we will use the definition of the bias $B_t = \Theta_t - \widetilde{\Theta}_t$, then multiply out and use the triangle inequality and finally upper bound $x^\top\Theta_t A \leq m$ which holds for all A and apply equation (19).

$$\begin{aligned}
\mathbb{E}[U_T(X_0)] &= \mathbb{E}_{\mathcal{F}_T,X_0}\left[\sum_{t=1}^T\sum_{A\in\mathcal{A}} \mu_A X_0^\top\widetilde{\Theta}_t(A-\pi_T^*(X_0))\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T\sum_{A\in\mathcal{A}} \mu_A X_0^\top(\Theta_t - B_t)(A-\pi_T^*(X_0))\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^T\sum_{A\in\mathcal{A}} \mu_A\left(\left|X_0^\top\Theta_t A\right| + \left|X_0^\top\Theta_t\pi_T^*(X_0)\right| + \left|X_0^\top B_t A\right| + \left|X_0^\top B_t\pi_T^*(X_0)\right|\right)\right] \\
&\leq \sum_{t=1}^T\sum_{A\in\mathcal{A}} \mu_A\left(2m + \sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right)\right) \\
&= 2Tm + T\sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right)
\end{aligned}$$

Now we are in a position to apply Lemma 10 as $|\eta x^\top\widetilde{\Theta}_t A| < 1$ by Lemma 11 since $\eta \leq \frac{1}{m(M+1)}$ per assumption. Next

we apply Lemma 12 and then use the upper bound we found for $\mathbb{E}_{\mathcal{F}_T,X_0}[U_T(X_0)]$ above as well as use the fact that $\gamma \leq 1$:

$$\begin{aligned}
\mathcal{R}_T &= \mathbb{E}_{X_0}\left[\widehat{\mathcal{R}}_T(X_0)\right] + 2\mathbb{E}\Big[\sum_{t=1}^{T}\max_{A\in\mathcal{A}}\left|\mathbb{E}[X_0 B_t A|\mathcal{F}_{t-1}]\right|\Big] \\
&\leq \mathbb{E}_{X_0}\left[\widehat{\mathcal{R}}_T(X_0)\right] + 2T\sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right) \\
&\leq \frac{\ln(|\mathcal{A}|)}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_{A\sim\pi_t(\cdot,X_0)}\left[(X_0^\top\widetilde{\Theta}_t A)^2|\mathcal{F}_{t-1}\right] + \gamma U_T(X_0)\right] + 2T\sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right) \\
&\leq \frac{\ln(|\mathcal{A}|)}{\eta} + 2T\eta m^2 Kd + 2T\gamma m + 3T\sigma\sqrt{m}G\exp\left(-M\beta\frac{\gamma K\lambda_{\min}(\Sigma)}{m}\right),
\end{aligned}$$

which proves the first result of the theorem. For the next result of the theorem, we first set $M = \frac{m\ln(T)}{\beta\gamma K\lambda_{\min}(\Sigma)}$ to find

$$\mathcal{R}_T \leq \frac{\ln(|\mathcal{A}|)}{\eta} + 2T\eta m^2 Kd + 2T\gamma m + 3\sigma\sqrt{m}G$$

Next, set $\gamma = \min\left\{1, \sqrt{K\ln(T)\frac{\ln(|\mathcal{A}|)}{T\beta\lambda_{\min}(\Sigma)}}\right\}$ to find

$$\mathcal{R}_T \leq \frac{\ln(|\mathcal{A}|)}{\eta} + 2T\eta m^2 Kd + 2m\sqrt{TK\frac{\ln(|\mathcal{A}|)\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + 3\sigma\sqrt{m}G.$$

Finally, set $\eta = \min\left\{\frac{1}{m(M+1)}, \sqrt{\frac{\ln(|\mathcal{A}|)}{Tm^2Kd}}\right\}$ to find

$$\begin{aligned}
\mathcal{R}_T &\leq 3\sqrt{\ln(|\mathcal{A}|)Tm^2Kd} + mM\ln(|\mathcal{A}|) + m\ln(|\mathcal{A}|) + 2m\sqrt{TK\frac{\ln(|\mathcal{A}|)\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + 3\sigma\sqrt{m}G \\
&\leq 3\sqrt{\ln(|\mathcal{A}|)Tm^2Kd} + 2m\sqrt{TK\frac{\ln(|\mathcal{A}|)\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + m\sqrt{\frac{TK\ln(T)\ln(|\mathcal{A}|)}{\beta\lambda_{\min}(\Sigma)}} \\
&\quad + 3\sigma\sqrt{m}G + m\ln(|\mathcal{A}|) + \frac{mK\ln(T)\ln(|\mathcal{A}|)}{\beta\lambda_{\min}(\Sigma)}.
\end{aligned}$$

We examine $\ln(|\mathcal{A}|)$ for the final result.

$$\begin{aligned}
\ln(|\mathcal{A}|) &\leq \ln\left(\sum_{j=1}^{m}\binom{K}{j}\right) \\
&\leq \ln\left(\sum_{j=1}^{m}\left(\frac{eK}{j}\right)^j\right) \\
&\leq \ln\left(m^m\left(\frac{eK}{m}\right)^m\right) \\
&= m\ln(eK) \qquad (20)
\end{aligned}$$

where we used the well known fact that $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ (Knuth, 1997, §1.2.6 : Binomial Coefficients: Exercise 67) where e is Euler's number. This allows us to conclude

$$\begin{aligned}
\mathcal{R}_T &\leq 3\sqrt{\ln(eK)Tm^3Kd} + 2\sqrt{m^3TK\frac{\ln(eK)\ln(T)}{\beta\lambda_{\min}(\Sigma)}} + \sqrt{m^3\frac{TK\ln(T)\ln(eK)}{\beta\lambda_{\min}(\Sigma)}} \\
&\quad + 3\sigma\sqrt{m}G + m^2\ln(eK) + \frac{m^2K\ln(T)\ln(eK)}{\beta\lambda_{\min}(\Sigma)}.
\end{aligned}$$

which completes the proof of the second result of the theorem after using $\beta = \frac{1}{m\sigma^2}$.

□

CO_2-FTRL	**Exp3-Tensor**	NC FTRL	ComBand	RealLinExp3
$O\left(\lvert\mathcal{D}\rvert d^2K^2 + d^3K\right)$	$O\left(\lvert\mathcal{A}\rvert\lvert\mathcal{D}\rvert d^2K^2 + d^3K^3\right)$	$O\left(K^2\right)$	$O\left(\lvert\mathcal{A}\rvert K^2 + K^3\right)$	$O\left(\lvert\mathcal{A}\rvert(\lvert\mathcal{D}\rvert d^2 + d^3)\right)$

Table 1: The theoretical runtime of the algorithms presented in this paper (bold) and the baselines in each timestep on the m-set problem and without the use of MGR.

F DETAILS OF THE EXPERIMENTS

The experiments were ran using Python 3.10.7, on an Intel(R) Xeon(R) CPU E7-4870 (2.40GHz) and we considered the full- and semi-bandit setting.

In total the runtime of the experiments on this hardware is around 80 hours. The implementation for none of the algorithms is particularly optimised. Unsurprisingly, the algorithms proposed in this paper with their more sophisticated estimators are somewhat more computationally demanding. A table with the theoretical runtimes of the algorithms can be found in Table F.

We will use One-Per-Context (OPC) as suffix to denote running an algorithm independently for each context. Since we draw uniformly from $\mathcal{B}_{K,m}$, we know that each context appears $T/|\mathcal{B}_{K,m}|$ times in expectation. Thus, we tune all instances of the sub-algorithm for length $T' = T/|\mathcal{B}_{K,m}|$. All other parameters are unchanged.

We ran two versions of each algorithm. The first version is the algorithm with tuning suggested by theory. The second version is the algorithm tuned with all parameters set to 1, except for T. We refer to this last choice of tuning as $1/\sqrt{T}$ tuning, as most parameters of the algorithms reduce to $1/\sqrt{T}$ with this tuning.

F.1 Full-Bandit Setting

The results for the full-bandit setting with theoretical tuning can be found in Figure 2. As expected the comparative performance of RealLinExp3 deteriorates with a more complicated combinatorial element and improves with a more difficult contextual element. Curiously, non-contextual ComBand sometimes outperforms ComBand OPC. It is possible that some actions are by random chance better across multiple contexts and thus ignoring contexts can lead to better results in the short term as the algorithm is able to exploit those better actions. This phenomenon appears most prevalent when the number of contexts is large, i.e., in the $(5, 2)$- and $(12, 3)$-context cases. EXP3-Tensor seems to be performing on par or somewhat worse than the other algorithms.

The results for the full-bandit setting with $1/\sqrt{T}$ tuning can be found in Figure 3. The results are comparable to the ones obtained with theoretical tuning. Unfortunately, this means that the results for Exp3-Tensor are on par with or slightly worse than the results of other the other algorithms. Interestingly, RealLinExp3 seems to be performing even better in the settings without a combinatorial aspect but continues to struggle with a strong combinatorial aspect.

Unfortunately, with both theoretical tuning and $1/\sqrt{T}$ tuning Exp3-Tensor was at best on par with the other algorithms. We conjecture that the artificial scenarios with a finite number of contexts, which we designed to accommodate our baselines, are not sufficiently expressive for Exp3-Tensor to exploit.

F.2 Semi-Bandit Setting

To efficiently implement Line 4 of Algorithm 2 we use Warmuth and Kuzmin (2008, Algorithm 4).

One of the algorithms we compare with in the semi-bandit setting is a variant of the Online Stochastic Mirror Descent (OSMD) algorithm presented in Audibert et al. (2014, Figure 3). Specifically, our variant uses the same estimator as that algorithm. However, we use FTRL instead of OSMD. We use exploration parameter $\gamma = \sqrt{\frac{K}{Tm}}$ and learning rate $\eta = \sqrt{\frac{m \log(\frac{K}{m})}{TK}}$. From here on, we refer to this algorithm as Non-Contextual (NC) FTRL. The list of algorithms in the semi-bandit setting is thus CO_2-FTRL, RealLinExp3, NC FTRL, and NC FTRL OPC.

The results for the theoretical learning rates can be found in Figure 4. RealLinExp3 overlaps with the full-bandit case and the results relative to the other algorithms are similar: RealLinExp3's relative performance decreases as the problem becomes more combinatorial. NC FTRL OPC does well in general and especially so if the number of contexts is small. However, NC FTRL OPC is outperformed by NC FTRL in the $(12, 3)$-context case. CO_2-FTRL is on par with the other algorithms in most experiments, although in some experiments it is outperformed and sometimes it outperforms other

algorithms. Similar to the high γ in the full-bandit case, CO_2-FTRL uses $\gamma = 24.11\%$ in the most complicated case which might contribute to regret.

The results for $1/\sqrt{T}$ tuning can be found in Figure 5. When using the $1/\sqrt{T}$ tuning CO_2-FTRL is on par with the best competitors in the simpler cases and outperforming in the more complicated problem instances. This also suggests that CO_2-FTRL is less sensitive to miss-specified tunings and perhaps performs better with more aggressive tuning.

Unlike in the full-bandit setting the algorithm we designed, CO_2-FTRL, does outperform the baseline algorithms. In the full bandit setting we conjectured that the simplified experimental setting was not sufficiently expressive. We believe the the semi-bandit setting is a simpler setting, which is the reason why ,even though the experiments are not very expressive, CO_2-FTRL was able to outperform the baseline algorithms.

Figure 2: Boxplots over 10 repetitions of the regret (in thousands) of the algorithms in the full-bandit setting using theoretical tuning (lower is better).

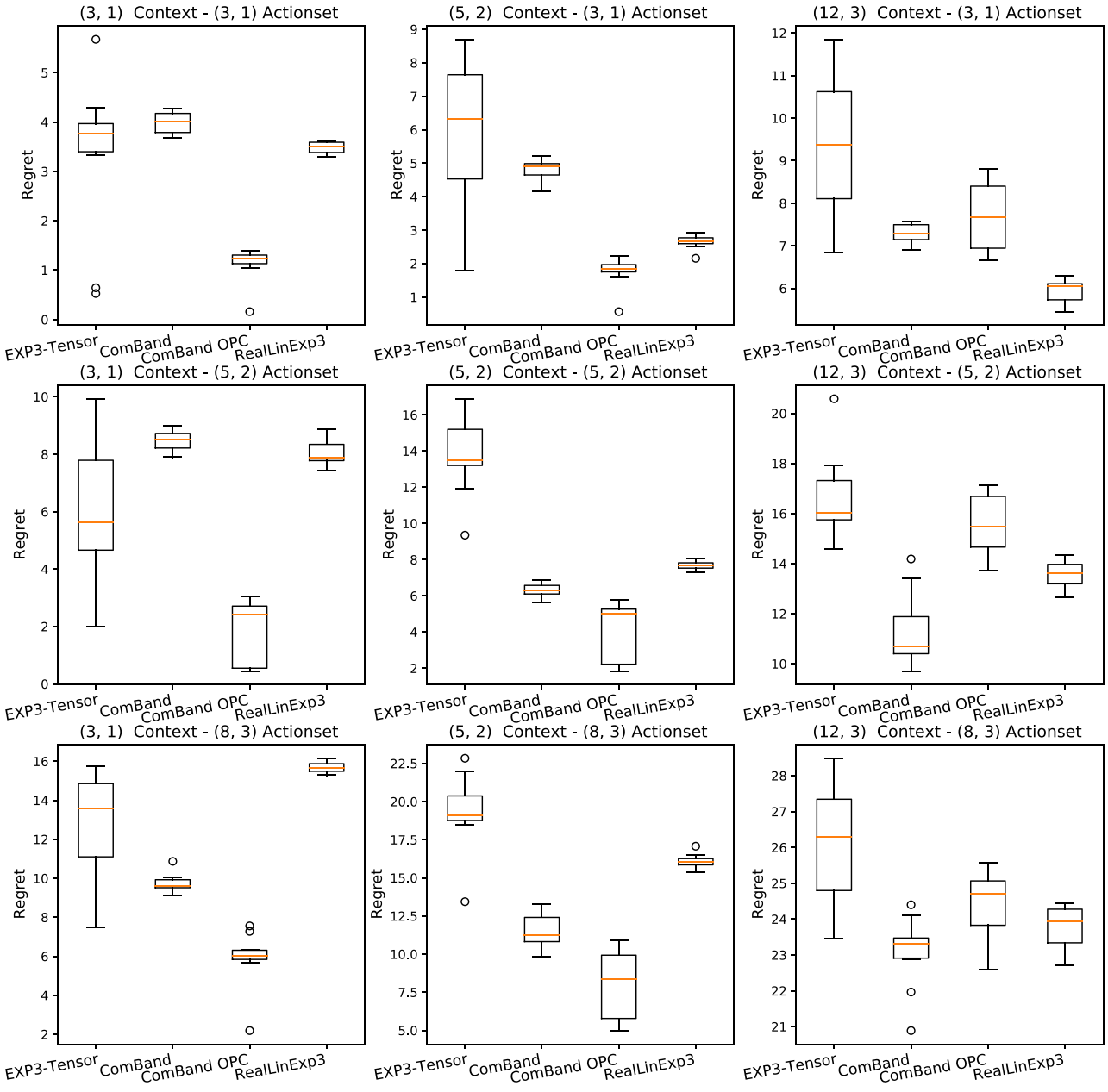

Figure 3: Boxplots over 10 repetitions of the regret (in thousands) of the algorithms in the full-bandit setting using $1/\sqrt{T}$ tuning (lower is better).

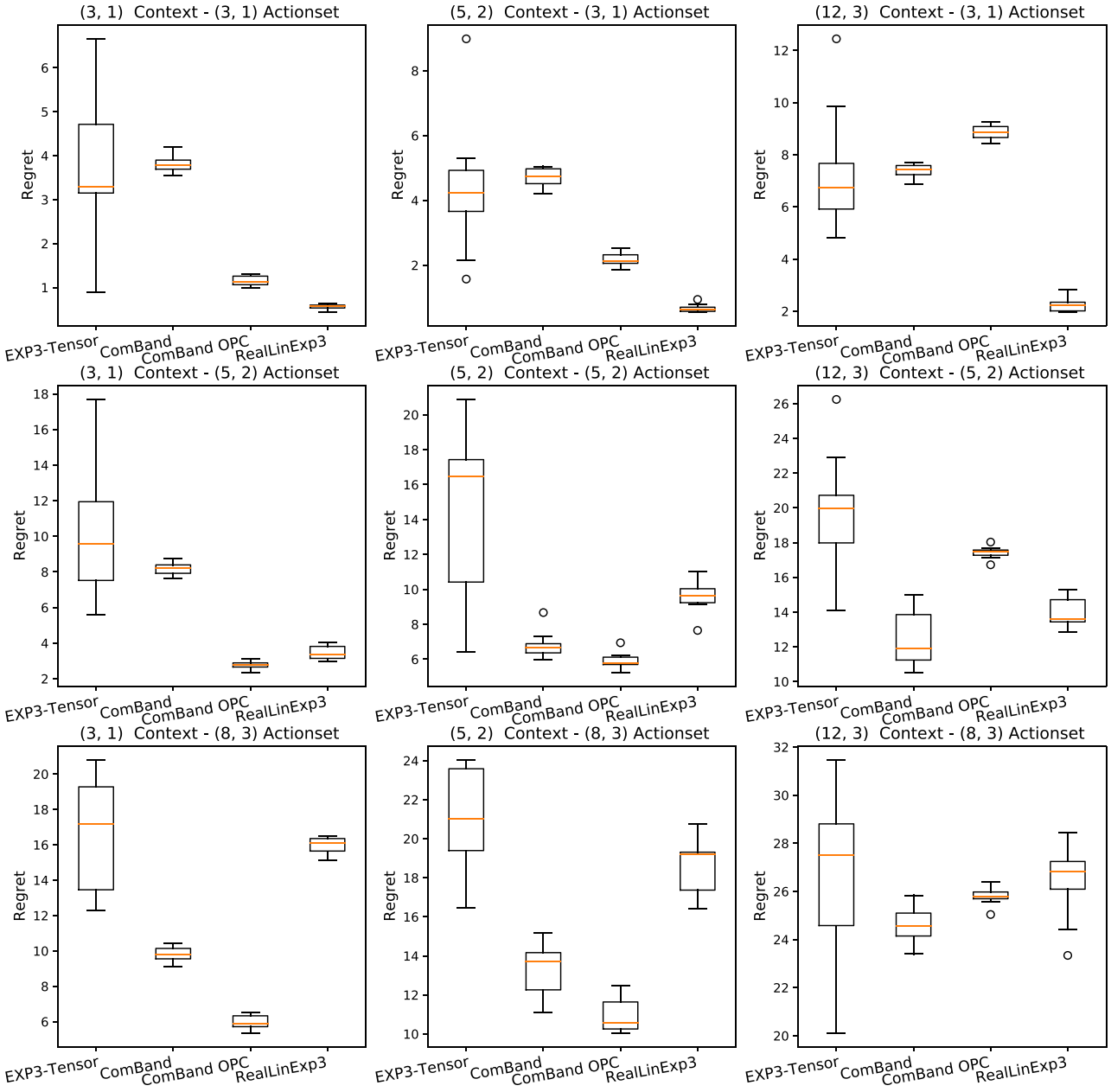

Figure 4: Boxplots over 10 repetitions of the regret (in thousands) of the algorithms in the semi-bandit setting using theoretical tuning (lower is better).

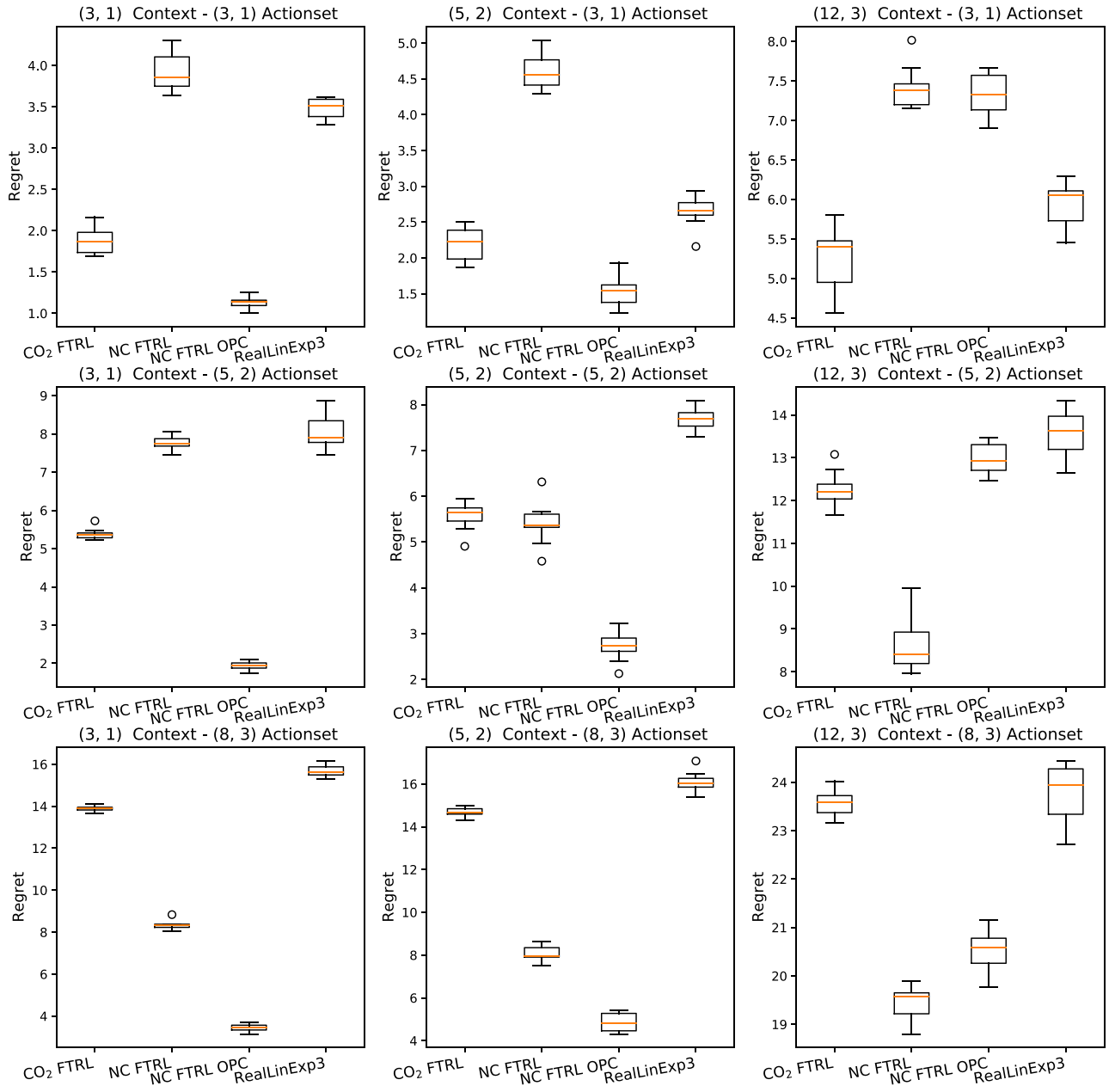

Figure 5: Boxplots over 10 repetitions of the regret (in thousands) of the algorithms in the semi-bandit setting using $1/\sqrt{T}$ (lower is better).