Efficient Linear Bandits through Matrix Sketching

Ilja Kuzborskij **Leonardo Cella** **Nicolò Cesa-Bianchi**

Dipartimento di Informatica & DSRC, Università degli Studi di Milano, 20133 Milano, Italy

Abstract

We prove that two popular linear contextual bandit algorithms, OFUL and Thompson Sampling, can be made efficient using Frequent Directions, a deterministic online sketching technique. More precisely, we show that a sketch of size m allows a $\mathcal{O}(md)$ update time for both algorithms, as opposed to $\Omega(d^2)$ required by their non-sketched versions in general (where d is the dimension of context vectors). This computational speedup is accompanied by regret bounds of order $(1+\varepsilon_m)^{3/2}d\sqrt{T}$ for OFUL and of order $\big((1+\varepsilon_m)d\big)^{3/2}\sqrt{T}$ for Thompson Sampling, where ε_m is bounded by the sum of the tail eigenvalues not covered by the sketch. In particular, when the selected contexts span a subspace of dimension at most m, our algorithms have a regret bound matching that of their slower, non-sketched counterparts. Experiments on real-world datasets corroborate our theoretical results.

1 Introduction

The stochastic contextual bandit is a sequential decision-making problem where an agent interacts with an unknown environment in a series of rounds. In each round, the environment reveals a set of feature vectors (called contexts, or actions) to the agent. The agent chooses an action from the revealed set and observes the stochastic reward associated with that action (bandit feedback). The strategy used by the agent for choosing actions based on past observations is called a policy. The goal of the agent is to learn a policy minimizing the regret, defined as the difference between the total reward of the optimal policy (i.e., the policy choosing the action with highest expected reward at each round) and the total reward of the agent's policy.

Contextual bandits are a popular modelling tool in many interactive machine learning tasks. A typical area of application is personalized recommendation, where a recommender system selects a product for a given user from a set of available products (each described by a feature vector) and receives a feedback (purchase or non-purchase) for the selected product.

We focus on the *stochastic linear bandit* model (Auer, 2002; Dani et al., 2008), where the set of actions (or decision set) is a finite[1] set $D_t \subset \mathbb{R}^d$, and the reward for choosing action $\boldsymbol{x}_t \in D_t$ is given by $Y_t = \boldsymbol{x}_t^\top \boldsymbol{w}^\star + \eta_t$ where $\boldsymbol{w}^\star \in \mathbb{R}^d$ is a fixed and unknown vector of real coefficients and η_t is a zero-mean random variable. The regret in this setting is defined by

$$R_T = \sum_{t=1}^{T} \boldsymbol{x}_t^{\star\top} \boldsymbol{w}^\star - \sum_{t=1}^{T} \boldsymbol{x}_t^\top \boldsymbol{w}^\star \tag{1}$$

where $\boldsymbol{x}_t^\star = \arg\max_{\boldsymbol{x}\in D_t} \boldsymbol{x}^\top \boldsymbol{w}^\star$ is the optimal action at round t. Bounds on the regret typically apply to any individual sequence of decision sets D_t and depend on quantities arising from the interplay between $\boldsymbol{w}^\star$, the sequence of decision sets, and the randomness of the rewards. Note that R_T is a random variable because the actions $\boldsymbol{x}_t \in D_t$ selected by the policy are functions of the past observed rewards. For this reason, our regret bounds only hold with probability at least $1-\delta$, where δ is a confidence parameter. By choosing $\delta = T^{-1}$, we can instead bound the expected regret $\mathbb{E}\left[R_T\right]$ by paying only a $\ln T$ extra factor in the bound.

We consider two of the most popular algorithms for stochastic linear bandits: OFUL (Abbasi-Yadkori et al., 2011) and linear Thompson Sampling (Agrawal and Goyal, 2013) (linear TS for short). While exhibiting good theoretical and empirical performances, both algorithms require $\Omega\big(d^2\big)$ time to update their model after each round. In this work we investigate whether it is possible to significantly reduce this update time while ensuring that the regret remains nicely bounded.

Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

[1] Note that our regret bounds do not actually depend on the cardinality of the sets D_t.

The quadratic dependence on d is due to the computation of the inverse correlation matrix of past actions (a cubic dependence is avoided because each new inverse is a rank-one perturbation of the previous inverse). The occurrence of this matrix is caused by the linear nature of rewards: to compute their decisions, both algorithms essentially solve a regularized least squares problem at every round. In order to improve the running time, we sketch the correlation matrix using a specific technique —Frequent Directions, (Ghashami et al., 2016)— that works well in a sequential learning setting. While matrix sketching is a well-known approach (Woodruff, 2014), to the best of our knowledge this is the first work that applies sketching to linear contextual bandits while providing rigorous performance guarantees.

With a sketch size of m, a rank-one update of the correlation matrix takes only time $\mathcal{O}(md)$, which is linear in d for a constant sketch size. However, this speed-up comes at a price, as sketching reduces the matrix rank causing a loss of information which —in turn— affects the least squares estimates used by the algorithms. Our main technical contribution shows that when OFUL and linear TS are run with a sketched correlation matrix, their regret blows up by a factor which is controlled by the spectral decay of the correlation matrix of selected actions. More precisely, we show that the sketched variant of OFUL, called SOFUL, achieves a regret bounded by

$$R_T \overset{\widetilde{\mathcal{O}}}{=} \big(1+\varepsilon_m\big)^{\frac{3}{2}}\Big(m + d\ln\big(1+\varepsilon_m\big)\Big)\sqrt{T} \qquad (2)$$

where m is the sketch size and ε_m is upper bounded by the spectral tail (sum of the last $d-m+1$ eigenvalues) of the correlation matrix for all T rounds. In the special case when the selected actions span a number of dimensions equal or smaller than the sketch size, then $\varepsilon_m = 0$ implying a regret of order $m\sqrt{T}$. Thus, we have a regret bound matching that of the slower, non-sketched counterpart.[2] When the correlation matrix has rank larger than the sketch size, the regret of SOFUL remains small to the extent the spectral tail of the matrix grows slowly with T. In the worst case of a spectrum with heavy tails, SOFUL may incur linear regret. In this respect, sketching is only justified when the computational cost of running OFUL cannot be afforded. Similarly, we prove that the efficient sketched formulation of linear TS enjoys a regret bound of order

$$R_T \overset{\widetilde{\mathcal{O}}}{=} \Big(m + d\ln(1+\varepsilon_m)\Big)\big(1+\varepsilon_m\big)^{\frac{3}{2}}\sqrt{dT}\ . \qquad (3)$$

Once again, for $\varepsilon_m = 0$ our bound is of order $m\sqrt{dT}$, which matches the regret bound for linear TS. When the rank of the correlation matrix is larger than the sketch size, the bound for linear TS behaves similarly to the bound for SOFUL.

Finally, we show a problem-dependent regret bound for SOFUL. This bound, which exhibits a logarithmic dependence on T, depends on the smallest gap Δ between the expected reward of the best and the second best action across the T rounds,

$$R_T \overset{\widetilde{\mathcal{O}}}{=} \frac{1}{\Delta}\big(1+\varepsilon_m\big)^3\Big(m + d\ln\big(1+\varepsilon_m\big)\Big)^2(\ln T)^2\ . \qquad (4)$$

When $\varepsilon_m(T) = 0$ this bound is of order $\frac{m^2}{\Delta}(\ln T)^2$ which matches the corresponding bound for OFUL. Experiments on six real-world datasets support our theoretical results.

Additional related work. For an introduction to contextual bandits, we refer the reader to the recent monograph of Lattimore and Szepesvári (2018). The idea of applying sketching techniques to linear contextual bandits was also investigated by Yu et al. (2017), where they used random projections to preliminarly draw a random m-dimensional subspace which is then used in every round of play. However, the per-step computation time of their algorithm is cubic in m rather than quadratic like ours. Morover, random projection introduces an additive error ε in the instantaneous regret which becomes of order $m^{-1/2}$ for any value of the confidence parameter δ bounded away from 1. A different notion of compression in contextual bandits is explored by Jun et al. (2017), where they use hashing algorithms to obtain a computation time sublinear in the number K of actions. An application of sketching (including Frequent Directions) to speed up 2nd order algorithms for online learning is studied by Luo et al. (2016), in a RKHS setting by Calandriello et al. (2017), and in stochastic optimization by Gonen et al. (2016).

2 Notation and preliminaries

Let $\mathcal{B}(\boldsymbol{z}, r) \subset \mathbb{R}^d$ be the Euclidean ball of center $\boldsymbol{z}$ and radius $r > 0$ and let $\mathcal{B}(r) = \mathcal{B}(\mathbf{0}, r)$. Given a positive definite $d \times d$ matrix $\boldsymbol{A}$, we define the inner product $\langle \boldsymbol{x}, \boldsymbol{z}\rangle_{\boldsymbol{A}} = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}$ and the induced norm $\|\boldsymbol{x}\|_{\boldsymbol{A}} = \sqrt{\boldsymbol{x}^\top A \boldsymbol{x}}$, for any $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d$. Throughout the paper, we write $f \overset{\widetilde{\mathcal{O}}}{=} g$ to denote $f = \widetilde{\mathcal{O}}(g)$. The contextual bandit protocol is described in Algorithm 1.

Algorithm 1 (Contextual Bandit)

1: **for** $t = 1, 2, \ldots$ **do**
2: Get decision set $D_t \subset \mathbb{R}^d$
3: Use current policy to select action $\boldsymbol{x}_t \in D_t$
4: Observe reward $Y_t \in \mathbb{R}$
5: Use pair $(\boldsymbol{x}_t, Y_t)$ to update the current policy
6: **end for**

[2] The regret bound of OFUL in (Abbasi-Yadkori et al., 2011, Theorem 3) is stated as $\mathcal{O}(d\sqrt{T})$, however, it can be improved for low-rank problems by using the "log-det" formulation of the confidence ellipsoid.

We introduce some standard assumptions for the linear contexual bandit setting. At any round $t = 1, 2, \dots$ the decision set $D_t \subset \mathbb{R}^d$ is finite and such that $\|\boldsymbol{x}\| \leq L$ for all $\boldsymbol{x} \in D_t$ and for all $t \geq 1$. The noise sequence $\eta_1, \eta_2 \dots, \eta_T$ is conditionally R-subgaussian for some fixed constant $R \geq 0$. Formally, for all $t \geq 1$ and all $\lambda \in \mathbb{R}$, $\mathbb{E}\left[e^{\lambda \eta_t} \,\middle|\, \eta_1, \dots, \eta_{t-1}\right] \leq \exp\left(\lambda^2 R^2/2\right)$. Note that this implies $\mathbb{E}[\eta_t \mid \eta_1, \dots, \eta_{t-1}] = 0$ and $\mathrm{Var}[\eta_t \mid \eta_1, \dots, \eta_{t-1}] \leq R^2$. Finally, we assume that a known upper bound S on $\|\boldsymbol{w}^\star\|$ is available.

Both OFUL and Linear TS operate by computing a confidence ellipsoid to which $\boldsymbol{w}^\star$ belongs with high probability. Let $\boldsymbol{X}_t = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_t]^\top$ be the $t \times d$ matrix of all actions selected up to round t by an arbitrary policy for linear contextual bandits. For $\lambda > 0$, define the regularized correlation matrix of actions $\boldsymbol{V}_t$ and the regularized least squares (RLS) estimate $\widehat{\boldsymbol{w}}_t$ as

$$\boldsymbol{V}_t = \boldsymbol{X}_t^\top \boldsymbol{X}_t + \lambda \boldsymbol{I} \quad \text{and} \quad \widehat{\boldsymbol{w}}_t = \boldsymbol{V}_t^{-1} \sum_{s=1}^{t} \boldsymbol{x}_s Y_s \ . \quad (5)$$

The following theorem (Abbasi-Yadkori et al., 2011, Theorem 2) bounds in probability the distance, in terms of the norm $\|\cdot\|_{\boldsymbol{V}_t}$, between the optimal parameter $\boldsymbol{w}^\star$ and the RLS estimate $\widehat{\boldsymbol{w}}_t$.

Theorem 1 (Confidence Ellipsoid)**.** *Let $\widehat{\boldsymbol{w}}_t$ be the RLS estimate constructed by an arbitrary policy for linear contextual bandits after t rounds of play. For any $\delta \in (0, 1)$, the optimal parameter $\boldsymbol{w}^\star$ belongs to the set $C_t \equiv \left\{\boldsymbol{w} \in \mathbb{R}^d \,:\, \|\boldsymbol{w} - \widehat{\boldsymbol{w}}_t\|_{\boldsymbol{V}_t} \leq \beta_t(\delta)\right\}$ with probability at least $1 - \delta$, where*

$$\beta_t(\delta) = R\sqrt{d \ln\left(1 + \frac{tL^2}{\lambda d}\right) + 2 \ln\left(\frac{1}{\delta}\right)} + S\sqrt{\lambda} \ . \quad (6)$$

OFUL. The actions selected by OFUL are solutions to the following constrained optimization problem

$$\boldsymbol{x}_t = \operatorname*{arg\,max}_{\boldsymbol{x} \in D_t} \max_{\boldsymbol{w} \in \mathbb{R}^d} \boldsymbol{x}^\top \boldsymbol{w}$$
$$\text{such that} \quad \|\boldsymbol{w} - \widehat{\boldsymbol{w}}_{t-1}\|_{\boldsymbol{V}_{t-1}} \leq \beta_{t-1}(\delta) \ .$$

Using Lemma 5 (in the appendix), OFUL can be formulated as Algorithm 2. Note that $\boldsymbol{x}_t$ maximizes the expected reward estimate $\widehat{\boldsymbol{w}}_{t-1}^\top \boldsymbol{x}$ plus a term $\beta_{t-1}(\delta)\|\boldsymbol{x}\|_{\boldsymbol{V}_{t-1}^{-1}}$ that provides an upper confidence bound for the RLS estimate in the direction of $\boldsymbol{x}$.

Linear TS. The linear Thompson Sampling algorithm of Agrawal and Goyal (2013) is Bayesian in nature: the selected actions and the observed rewards are used to update a Gaussian prior over the parameter space. Each action $\boldsymbol{x}_t$ is selected by maximixing $\boldsymbol{x}^\top \widehat{\boldsymbol{w}}_t^{\text{TS}}$ over $\boldsymbol{x} \in D_t$, where $\widehat{\boldsymbol{w}}_t^{\text{TS}}$ is a random vector drawn from the posterior. As shown by Abeille and Lazaric (2017),

Algorithm 2 (OFUL)

Input: $\delta, \lambda > 0$
$\widehat{\boldsymbol{w}}_0 = \boldsymbol{0}, \boldsymbol{V}_0^{-1} = \frac{1}{\lambda}\boldsymbol{I}$.
for $t = 1, 2, \dots$ **do**
  Get decision set D_t
  Play $\boldsymbol{x}_t \leftarrow \operatorname*{arg\,max}_{\boldsymbol{x} \in D_t} \left\{\widehat{\boldsymbol{w}}_{t-1}^\top \boldsymbol{x} + \beta_{t-1}(\delta)\|\boldsymbol{x}\|_{\boldsymbol{V}_{t-1}^{-1}}\right\}$
  Observe reward Y_t
  Compute $\boldsymbol{V}_t^{-1}$ and $\widehat{\boldsymbol{w}}_t$ using (5)
end for

linear TS can be equivalently defined as a randomized algorithm based on the RLS estimate (see Algorithm 3). The random vectors $\boldsymbol{Z}_t$ are drawn i.i.d. from a suitable

Algorithm 3 (Linear TS)

Input: $\delta, \lambda > 0, m \in \{1, \dots, d-1\}$, $\mathcal{D}^{\text{TS}}$ (sampling distribution)
$\widehat{\boldsymbol{w}}_0 = \boldsymbol{0}, \boldsymbol{V}_0^{-1} = \frac{1}{\lambda}\boldsymbol{I}_{d\times d}, \delta' = \delta/(4T)$
for $t = 1, 2, \dots$ **do**
  Get decision set D_t
  Sample $\boldsymbol{Z}_t \sim \mathcal{D}^{\text{TS}}$
  Play $\boldsymbol{x}_t \leftarrow \operatorname*{arg\,max}_{\boldsymbol{x} \in D_t} \boldsymbol{x}^\top \left(\widehat{\boldsymbol{w}}_{t-1} + \widetilde{\beta}_t(\delta')\boldsymbol{V}_{t-1}^{-\frac{1}{2}}\boldsymbol{Z}_t\right)$
  Observe reward Y_t
  Compute $\boldsymbol{V}_t^{-\frac{1}{2}}$ and $\widehat{\boldsymbol{w}}_t$ using (5)
end for

multivariate distribution $\mathcal{D}^{\text{TS}}$ that need not be related to the posterior. In order to prove regret bounds, it is sufficient that the law of $\boldsymbol{Z}_t$ satisfies certain properties.

Definition 1 (TS-sampling distribution)**.** *A multivariate distribution $\mathcal{D}^{\text{TS}}$ on $\mathbb{R}^d$, absolutely continuous w.r.t. the Lebesgue measure, is TS-sampling if it satisfies the following two properties:*

- *(Anti-concentration) There exists $p > 0$ such that for any $\boldsymbol{u}$ with $\|\boldsymbol{u}\| = 1$, $\mathbb{P}\big(\boldsymbol{u}^\top \boldsymbol{Z} \geq 1\big) \geq p$.*
- *(Concentration) There exist $c, c' > 0$ such that for all $\delta \in (0, 1)$,*

$$\mathbb{P}\left(\|\boldsymbol{Z}\| \leq \sqrt{cd \ln\left(\frac{c'd}{\delta}\right)}\right) \geq 1 - \delta \ .$$

Similarly to OFUL, linear TS uses the notion of confidence ellipsoid. However, due to the properties of the sampling distribution $\mathcal{D}^{\text{TS}}$, the ellipsoid used by linear TS is larger by a factor of order $\sqrt{d}$ than the ellipsoid used by OFUL. This causes an extra factor of $\sqrt{d}$ in the regret bound, which is not known to be necessary.

Note that both OFUL and linear TS need to maintain $\boldsymbol{V}_t^{-1}$ (or $\boldsymbol{V}_t^{-\frac{1}{2}}$), which requires time $\Omega\big(d^2\big)$ to update. In the next section, we show how this update time can be improved by sketching the regularized correlation matrix $\boldsymbol{V}_t$.

3 Sketching the correlation matrix

The idea of sketching is to maintain an approximation of $\boldsymbol{X}_t$, denoted by $\boldsymbol{S}_t \in \mathbb{R}^{m\times d}$, where $m \ll d$ is a small constant called the sketch size. If we choose m such that $\boldsymbol{S}_t^\top \boldsymbol{S}_t$ approximates $\boldsymbol{X}_t^\top \boldsymbol{X}_t$ well, we could use $\boldsymbol{S}_t^\top \boldsymbol{S}_t + \lambda \boldsymbol{I}$ in place of $\boldsymbol{V}_t$. In the following we use the notation $\widetilde{\boldsymbol{V}}_t = \boldsymbol{S}_t^\top \boldsymbol{S}_t + \lambda \boldsymbol{I}$ to denote the sketched regularized correlation matrix. The RLS estimate based upon it is denoted by

$$\widetilde{\boldsymbol{w}}_t = \widetilde{\boldsymbol{V}}_t^{-1} \sum_{s=1}^{t} \boldsymbol{x}_s Y_s \ . \tag{7}$$

A trivial replacement of $\boldsymbol{V}$ with $\widetilde{\boldsymbol{V}}$ does not yield an efficient algorithm. On the other hand, using the Woodbury identity we may write

$$\widetilde{\boldsymbol{V}}_t^{-1} = \frac{1}{\lambda}\left(\boldsymbol{I}_{d\times d} - \boldsymbol{S}_t^\top \boldsymbol{H}_t \boldsymbol{S}_t\right)$$

where $\boldsymbol{H}_t = \left(\boldsymbol{S}_t \boldsymbol{S}_t^\top + \lambda \boldsymbol{I}_{m\times m}\right)^{-1}$. Here matrix-vector multiplications involving $\boldsymbol{S}_t$ require time $\mathcal{O}(md)$, while matrix-matrix multiplications involving $\boldsymbol{H}_t$ require time $\mathcal{O}(m^2)$. So, as long as $\boldsymbol{S}_t$ and $\boldsymbol{H}_t$ can be efficiently maintained, we obtain an algorithm for linear stochastic bandits where $\widetilde{\boldsymbol{V}}_t^{-1}$ can be updated in time $\mathcal{O}(md + m^2)$. Next, we focus on a concrete sketching algorithm that ensures efficient updates of $\boldsymbol{S}_t$ and $\boldsymbol{H}_t$.

Frequent Directions. Frequent Directions (FD) (Ghashami et al., 2016) is a deterministic sketching algorithm that maintains a matrix $\boldsymbol{S}_t$ whose last row is invariably $\boldsymbol{0}$. On each round, we insert $\boldsymbol{x}_t^\top$ into the last row of $\boldsymbol{S}_{t-1}$, perform an eigendecomposition $\boldsymbol{S}_{t-1}^\top \boldsymbol{S}_{t-1} + \boldsymbol{x}_t \boldsymbol{x}_t^\top = \boldsymbol{U}_t \boldsymbol{\Sigma}_t \boldsymbol{U}_t^\top$, and then set $\boldsymbol{S}_t = \left(\boldsymbol{\Sigma}_t - \rho_t \boldsymbol{I}_{m\times m}\right)^{\frac{1}{2}} \boldsymbol{U}_t$, where ρ_t is the smallest eigenvalue of $\boldsymbol{S}_t^\top \boldsymbol{S}_t$. Observe that the rows of $\boldsymbol{S}_t$ form an orthogonal basis, and therefore $\boldsymbol{H}_t$ is a diagonal matrix which can be updated and stored efficiently. Now, the only step in question is an eigendecomposition, which can also be done in time $\mathcal{O}(md)$ —see (Ghashami et al., 2016, Section 3.2). Hence, the total update time per round is $\mathcal{O}(md)$. The updates of matrices $\boldsymbol{S}_t$ and $\boldsymbol{H}_t$ are summarized in Algorithm 4.

Algorithm 4 (FD Sketching)

Input: $\boldsymbol{S}_{t-1} \in \mathbb{R}^{m\times d}, \boldsymbol{x}_t \in \mathbb{R}^d, \lambda > 0$

1: Compute eigendecomposition $\boldsymbol{U}^\top \mathrm{diag}\{\rho_1, \ldots, \rho_m\} \boldsymbol{U} = \boldsymbol{S}_{t-1}^\top \boldsymbol{S}_{t-1} + \boldsymbol{x}_t \boldsymbol{x}_t^\top$

2: $\boldsymbol{S}_t \leftarrow \mathrm{diag}\{\sqrt{\rho_1 - \rho_m}, \ldots, \sqrt{\rho_{m-1} - \rho_m}, 0\} \boldsymbol{U}$

3: $\boldsymbol{H}_t \leftarrow \mathrm{diag}\left\{\frac{1}{\rho_1 - \rho_m + \lambda}, \ldots, \frac{1}{\lambda}\right\}$

Output: $\boldsymbol{S}_t, \boldsymbol{H}_t$

It is not hard to see that FD sketching sequentially identifies the top-m eigenvectors of the matrix $\boldsymbol{X}_T^\top \boldsymbol{X}_T$. Thus, whenever we use a sketched estimate, we lose a part of the spectrum tail. This loss is captured by the following notion of *spectral error*,

$$\varepsilon_m = \min_{k=0,\ldots,m-1} \frac{\lambda_{d-k} + \lambda_{d-k+1} + \cdots + \lambda_d}{\lambda(m-k)} \tag{8}$$

where $\lambda_1 \geq \ldots \geq \lambda_d$ are the eigenvalues of the correlation matrix $\boldsymbol{X}_T^\top \boldsymbol{X}_T$. Note that $\varepsilon_m \leq (\lambda_m + \cdots + \lambda_d)/\lambda$. For matrices with low rank or light-tailed spectra we expect this spectral error to be small. In the following, we use $\widetilde{m}$ to denote the quantity $m + d\ln(1+\varepsilon_m)$ which occurs often in our bounds involving sketching. Note that $\widetilde{m} \geq m$ and $\widetilde{m} \to m$ as the spectral error vanishes.

Since the matrix $\boldsymbol{V}_t$ is used to compute both the RLS estimate $\widehat{\boldsymbol{w}}_t$ and the norm $\|\cdot\|_{\boldsymbol{V}_t}$, the sketching of $\boldsymbol{V}_t$ clearly affects the confidence ellipsoid. The next theorem quantifies how much the confidence ellipsoid must be blown up in order to compensate for the sketching error. Let ρ_t be the smallest eigenvalue of the FD-sketched correlation matrix $\boldsymbol{S}_t^\top \boldsymbol{S}_t$ and let $\bar{\rho}_t = \rho_1 + \cdots + \rho_t$. The following proposition due to Ghashami et al. (2016) (see the proof of Thm. 3.1, bound on Δ) relates $\bar{\rho}_t$ to ε_m defined in (8).

Proposition 1. *For any $t = 0, \ldots, T$, any $\lambda > 0$, and any sketch size $m = 1, \ldots, d$, it holds that $\bar{\rho}_t/\lambda \leq \varepsilon_m$.*

A key lemma in the analysis of regret is the following sketched version of (Abbasi-Yadkori et al., 2011, Lemma 11), which bounds the sum of the ridge leverage scores. Although sketching introduces the spectral error ε_m, it also improves the dependence on the dimension from d to m whenever ε_m is sufficiently small.

Lemma 1 (Sketched leverage scores)**.**

$$\sum_{t=1}^{T} \min\left\{1, \|\boldsymbol{x}_t\|^2_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}}\right\} \leq 2\,(1+\varepsilon_m)\left(\widetilde{m} + m\ln\left(1 + \frac{TL^2}{m\lambda}\right)\right) \ . \tag{9}$$

We can now state the main result of this section.

Theorem 2 (Sketched confidence ellipsoid)**.** *Let $\widetilde{\boldsymbol{w}}_t$ be the RLS estimate constructed by an arbitrary policy for linear contextual bandits after t rounds of play. For any $\delta \in (0,1)$, the optimal parameter $\boldsymbol{w}^\star$ belongs to the set $\widetilde{C}_t \equiv \left\{\boldsymbol{w} \in \mathbb{R}^d \ : \ \|\boldsymbol{w} - \widetilde{\boldsymbol{w}}_t\|_{\widetilde{\boldsymbol{V}}_t} \leq \widetilde{\beta}_t(\delta)\right\}$ with probability at least $1-\delta$, where*

$$\widetilde{\beta}_t(\delta) = R\sqrt{m\ln\left(1 + \frac{tL^2}{m\lambda}\right) + 2\ln\frac{1}{\delta} + d\ln\left(1 + \frac{\bar{\rho}_t}{\lambda}\right)} \cdot \sqrt{1 + \frac{\bar{\rho}_t}{\lambda}} + S\sqrt{\lambda}\left(1 + \frac{\bar{\rho}_t}{\lambda}\right) \tag{10}$$

$$\overset{\widetilde{\mathcal{O}}}{=} R\sqrt{\widetilde{m}\,(1+\varepsilon_m)} + S\sqrt{\lambda}\,(1+\varepsilon_m) \ . \tag{11}$$

Note that (11) is larger than its non-sketched counterpart (6) due to the factors $1+\varepsilon_m$. However, when the spectral error ε_m vanishes, $\widetilde{\beta}_t(\delta)$ becomes of order $R\sqrt{m}+S\sqrt{\lambda}$, which improves upon (6) since we replace the dependence on the ambient space dimension d with the dependence on the sketch size m. In the following, we use the abbreviation $M_\lambda = \max\left\{1, 1/\sqrt{\lambda}\right\}$.

4 Sketched OFUL

Equipped with the sketched confidence ellipsoid and the sketched RLS estimate, we can now introduce SOFUL (Algorithm 5), the sketched version of OFUL. SOFUL

Algorithm 5 (SOFUL)

Input: $\delta, \lambda > 0, m \in \{1, \dots, d-1\}$

$\widetilde{\boldsymbol{w}}_0 = \boldsymbol{0}, \widetilde{\boldsymbol{V}}_0^{-1} = \frac{1}{\lambda}\boldsymbol{I}_{d\times d}, \boldsymbol{S}_0 = \boldsymbol{0}_{m\times d}$
for $t = 1, 2, \dots$ **do**
  Get decision set D_t
  Play $\boldsymbol{x}_t \leftarrow \arg\max_{\boldsymbol{x}\in D_t} \left\{ \widetilde{\boldsymbol{w}}_{t-1}^\top \boldsymbol{x} + \widetilde{\beta}_{t-1}(\delta) \left\|\boldsymbol{x}\right\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \right\}$
  Observe reward Y_t
  Compute $\boldsymbol{S}_t, \boldsymbol{H}_t$ using Alg. 4 given $\boldsymbol{S}_{t-1}, \boldsymbol{x}_t$
  $\widetilde{\boldsymbol{V}}_t^{-1} \leftarrow \frac{1}{\lambda}\left(\boldsymbol{I}_{d\times d} - \boldsymbol{S}_t^\top \boldsymbol{H}_t \boldsymbol{S}_t\right)$
  Compute $\widetilde{\boldsymbol{w}}_t$ using (7)
end for

enjoys the following regret bound, characterized in terms of the spectral error.

Theorem 3. *The regret of SOFUL with FD-sketching of size m w.h.p. satisfies*

$$R_T \overset{\widetilde{\mathcal{O}}}{=} M_\lambda \left(1+\varepsilon_m\right)^{\frac{3}{2}} \widetilde{m} \left(R + S\sqrt{\lambda}\right)\sqrt{T} \ .$$

Similarly to Abbasi-Yadkori et al. (2011), we also prove a distribution dependent regret bound for SOFUL. This bound is polylogarithmic in time and depends on the smallest difference Δ between the rewards of the best and the second best action in the decision sets,

$$\Delta = \min_{t=1,\dots,T} \max_{\boldsymbol{x}\in D_t \setminus \{\boldsymbol{x}_t^\star\}} \left(\boldsymbol{x}_t^\star - \boldsymbol{x}\right)^\top \boldsymbol{w}^\star \ .$$

Theorem 4. *The regret of SOFUL with FD-sketching of size m w.h.p. satisfies*

$$R_T \overset{\widetilde{\mathcal{O}}}{=} M_\lambda \left(1+\varepsilon_m\right)^3 \widetilde{m}^2 \left(R^2 + S^2\lambda\right) \frac{(\ln T)^2}{\Delta} \ .$$

Proofs of the regret bounds appear in the supplementary material (Section A.3).

5 Sketched linear TS

In this section we introduce a variant of linear TS (Algorithm 3) based on FD-sketching. Similarly to SOFUL, sketched linear TS (see Algorithm 6) uses the FD-sketched approximation $\widetilde{\boldsymbol{V}}_{t-1}$ of the correlation matrix $\boldsymbol{V}_{t-1}$ in order to select the action $\boldsymbol{x}_t$. Note

Algorithm 6 (Sketched linear TS)

Input: $\delta, \lambda > 0, m \in \{1, \dots, d-1\}$, $\mathcal{D}^{\text{TS}}$ (TS-sampling distribution)

$\widetilde{\boldsymbol{w}}_0 = \boldsymbol{0}, \widetilde{\boldsymbol{V}}_0^{-1} = \frac{1}{\lambda}\boldsymbol{I}_{d\times d}, \boldsymbol{S}_0 = \boldsymbol{0}_{m\times d}, \delta' = \delta/(4T)$
for $t = 1, 2, \dots$ **do**
  Get decision set D_t
  Sample $\boldsymbol{Z}_t \sim \mathcal{D}^{\text{TS}}$
  Play $\boldsymbol{x}_t \leftarrow \arg\max_{\boldsymbol{x}\in D_t} \boldsymbol{x}^\top \left(\widetilde{\boldsymbol{w}}_{t-1} + \widetilde{\beta}_t(\delta')\widetilde{\boldsymbol{V}}_{t-1}^{-\frac{1}{2}} \boldsymbol{Z}_t\right)$
  Observe reward Y_t
  Compute $\boldsymbol{S}_t, \boldsymbol{H}_t$ using Alg. 4 given $\boldsymbol{S}_{t-1}, X_t$
  $\widetilde{\boldsymbol{V}}_t^{-1} \leftarrow \frac{1}{\lambda}\left(\boldsymbol{I}_{d\times d} - \boldsymbol{S}_t^\top \boldsymbol{H}_t \boldsymbol{S}_t\right)$
  Compute $\widetilde{\boldsymbol{w}}_t$ using (7)
end for

that, in this case, we need both $\widetilde{\boldsymbol{V}}_{t-1}^{-1}$ and $\widetilde{\boldsymbol{V}}_{t-1}^{-\frac{1}{2}}$ to compute $\boldsymbol{x}_t$. Using the generalized Woodbury identity (Corollary 1 in Appendix A.2 for proofs), we can write

$$\widetilde{\boldsymbol{V}}_t^{-\frac{1}{2}} = \boldsymbol{S}_t^{'\top}\left(\boldsymbol{S}_t'\boldsymbol{S}_t^{'\top}\right)^{-1}\left(\frac{\lambda}{2}\boldsymbol{I} + \boldsymbol{S}_t'\boldsymbol{S}_t^{'\top}\right)^{-\frac{1}{2}} \boldsymbol{S}_t'$$

where $\boldsymbol{S}_t' = \left(\boldsymbol{\Sigma}_t + \left(\frac{\lambda}{2} - \rho_t\right)\boldsymbol{I}_{m\times m}\right)^{\frac{1}{2}} \boldsymbol{U}_t$. Note that $\widetilde{\boldsymbol{V}}_t^{-\frac{1}{2}}$ can still be computed in time $\mathcal{O}\left(md+m^2\right)$ because $\boldsymbol{S}_t'\boldsymbol{S}_t^{'\top}$ is a diagonal matrix.

The confidence ellipsoid stated in Theorem 2 applies to any contextual bandit policy, and so also to the $\widetilde{\boldsymbol{w}}_t$ constructed by sketched linear TS. However, as shown by Abeille and Lazaric (2017), the analysis needs a confidence ellipsoid larger by a factor equal to the bound on $\|\boldsymbol{Z}\|$ appearing in the concentration property of the TS-sampling distribution. More precisely, the *TS-confidence ellipsoid* is defined by $\widetilde{C}_t^{\text{TS}} \equiv \left\{\boldsymbol{w}\in\mathbb{R}^d \ : \ \|\boldsymbol{w} - \widetilde{\boldsymbol{w}}_t\|_{\widetilde{\boldsymbol{V}}_t} \le \widetilde{\gamma}_t\left(\delta/(4T)\right)\right\}$ where

$$\widetilde{\gamma}_t(\delta) = \widetilde{\beta}_t(\delta)\sqrt{cd\ln\left(\frac{c'd}{\delta}\right)} \ . \tag{12}$$

The quantity $\widetilde{\beta}_t(\delta)$ is defined in (10) and c, c' are the concentration constants of the TS-sampling distribution (Definition 1). We are now ready to prove a bound on the regret of linear TS with FD-sketching.

Theorem 5. *The regret of FD-sketched linear TS, run with sketch size m w.h.p. satisfies*

$$R_T \overset{\widetilde{\mathcal{O}}}{=} M_\lambda \left(1+\varepsilon_m\right)^{\frac{3}{2}} \widetilde{m} \left(R + S\sqrt{\lambda}\right)\sqrt{dT} \ .$$

The proof of Theorem 5 closely follows the analysis of Abeille and Lazaric (2017) with some key modifications due to the sketching operations. For completeness, we include the proof in the supplementary material.

6 Some proof sketches

Our regret analyses follow Abbasi-Yadkori et al. (2011); Abeille and Lazaric (2017) and related works. However, due to the sketching of the correlation matrix, some key components of the proofs now depend on the spectral error (8). Because of that, we need tools specific to the analysis of linear bandits with FD-sketching. These tools are used to bound the instantaneous regret $\left(\boldsymbol{x}^\star - \boldsymbol{x}_t\right)^\top \boldsymbol{w}^\star$ in terms of the norm $\|\boldsymbol{w}^\star - \widetilde{\boldsymbol{w}}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}}$ and the ridge leverage scores. Armed with these results, we then prove our regret bounds in Section 6.1. In this section we only present the core regret analysis of SOFUL, while we defer its complete analysis and analysis of the sketched linear TS to the supplementary material.

6.1 Proof of regret bounds for SOFUL

We start by introducting a basic relationship between the correlation matrix of actions $\boldsymbol{X}_s^\top \boldsymbol{X}_s$ and its FD-sketched estimate $\boldsymbol{S}_t^\top \boldsymbol{S}_t$ with sketch size $m \leq d$. Recall that ρ_t is the smallest eigenvalue of $\boldsymbol{S}_t^\top \boldsymbol{S}_t$ for $t = 1, \dots, T$ and $\bar{\rho}_t = \rho_1 + \cdots + \rho_t$.

Proposition 2. *Let $\boldsymbol{S}_s$ be the matrix computed by FD-sketching at time step $s = 1, \dots, t$ (where $\boldsymbol{S}_0 = \boldsymbol{0}$). Then $\boldsymbol{X}_s^\top \boldsymbol{X}_s = \boldsymbol{S}_s^\top \boldsymbol{S}_s + \bar{\rho}_s \boldsymbol{I}$.*

Proof. By construction, $\boldsymbol{S}_{s-1}^\top \boldsymbol{S}_{s-1} + \boldsymbol{x}_s \boldsymbol{x}_s^\top = \boldsymbol{U}_s \boldsymbol{\Sigma}_s \boldsymbol{U}_s^\top$ where $\boldsymbol{S}_s = \left(\boldsymbol{\Sigma}_s - \rho_s \boldsymbol{I}_{m \times m}\right)^{\frac{1}{2}} \boldsymbol{U}_s$. Thus,

$$\boldsymbol{S}_s^\top \boldsymbol{S}_s = \boldsymbol{U}_s \boldsymbol{\Sigma}_s \boldsymbol{U}_s^\top - \rho_s \boldsymbol{I} = \boldsymbol{S}_{s-1}^\top \boldsymbol{S}_{s-1} + \boldsymbol{x}_s \boldsymbol{x}_s^\top - \rho_s \boldsymbol{I}$$

Summing both sides of the above over $s = 1, \dots, t$, $\boldsymbol{S}_t^\top \boldsymbol{S}_t = \sum_{s=1}^t \boldsymbol{x}_s \boldsymbol{x}_s^\top - \sum_{s=1}^t \rho_s \boldsymbol{I}$. □

We will also use the following lemma of (Abbasi-Yadkori et al., 2011, Lemma 11).

Lemma 2. *For $\lambda \geq \max\left\{1, L^2\right\}$, we have that*

$$\sum_{t=1}^T \|\boldsymbol{x}_t\|^2_{\boldsymbol{V}_{t-1}^{-1}} \leq 2d \ln\left(1 + \frac{TL^2}{\lambda d}\right) \ . \tag{13}$$

The following lemma gives a sketch-specific version of the determinant-trace inequality, shown in the supplementary material (Lemma 8).

Lemma 3.

$$\ln\left(\frac{\det(\boldsymbol{V}_t)}{\det(\lambda \boldsymbol{I})}\right) \leq d \ln\left(1 + \frac{\bar{\rho}}{\lambda}\right) + m \ln\left(1 + \frac{tL^2}{m\lambda}\right) \ .$$

Next we prove Lemma 1, which is similar to Lemma 2. However, now the statement depends on the sketched matrix $\widetilde{\boldsymbol{V}}_{t-1}$ instead of $\boldsymbol{V}_{t-1}$.

Proof of Lemma 1. Throughout the proof, unless stated explicitly, we drop the subscripts containing t. Therefore, $\boldsymbol{V} = \boldsymbol{V}_{t-1}$, $\widetilde{\boldsymbol{V}} = \widetilde{\boldsymbol{V}}_{t-1}$, $\boldsymbol{x} = \boldsymbol{x}_t$, and $\bar{\rho} = \bar{\rho}_{t-1}$. Now suppose that $(\widetilde{\lambda}_i + \lambda, \widetilde{\boldsymbol{u}}_i)$ is an i-th eigenpair of $\widetilde{\boldsymbol{V}}$. Then, Proposition 2 implies that a corresponding eigenpair of $\boldsymbol{V}$ is $(\widetilde{\lambda}_i + \lambda + \bar{\rho}, \widetilde{\boldsymbol{u}}_i)$. Using this fact we have that

$$\begin{aligned}
&\|\boldsymbol{x}\|^2_{\boldsymbol{V}^{-1}} = \boldsymbol{x}^\top \widetilde{\boldsymbol{V}} \widetilde{\boldsymbol{V}}^{-1} \boldsymbol{V}^{-1} \boldsymbol{x} \\
&= \boldsymbol{x}^\top \left(\sum_{i=1}^d \widetilde{\boldsymbol{u}}_i \widetilde{\boldsymbol{u}}_i^\top \frac{1}{\widetilde{\lambda}_i + \lambda} \frac{\widetilde{\lambda}_i + \lambda}{\widetilde{\lambda}_i + \lambda + \bar{\rho}} \right) \boldsymbol{x} \\
&\geq \frac{\lambda}{\lambda + \bar{\rho}} \boldsymbol{x}^\top \left(\sum_{i=1}^d \widetilde{\boldsymbol{u}}_i \widetilde{\boldsymbol{u}}_i^\top \frac{1}{\widetilde{\lambda}_i + \lambda} \right) \boldsymbol{x} = \frac{\lambda}{\lambda + \bar{\rho}} \|\boldsymbol{x}\|^2_{\widetilde{\boldsymbol{V}}^{-1}} \ .
\end{aligned}$$

Furthermore, this implies that

$$\begin{aligned}
&\min\left\{1, \frac{\lambda}{\lambda + \bar{\rho}} \|\boldsymbol{x}\|^2_{\widetilde{\boldsymbol{V}}^{-1}}\right\} \leq \min\left\{1, \|\boldsymbol{x}\|^2_{\boldsymbol{V}^{-1}}\right\} \\
\Rightarrow \quad &\min\left\{1, \|\boldsymbol{x}\|^2_{\widetilde{\boldsymbol{V}}^{-1}}\right\} \leq \left(1 + \frac{\bar{\rho}}{\lambda}\right) \min\left\{1, \|\boldsymbol{x}\|^2_{\boldsymbol{V}^{-1}}\right\} \ .
\end{aligned}$$

Finally, combining the above with Lemma 2, equation and using the fact that $\bar{\rho}_{t-1} \leq \bar{\rho}_T$, we obtain

$$\begin{aligned}
&\sum_{t=1}^T \min\left\{1, \|\boldsymbol{x}_t\|^2_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}}\right\} \leq 2\left(1 + \frac{\bar{\rho}_T}{\lambda}\right) \ln\left(\frac{\det(\boldsymbol{V}_T)}{\det(\lambda I)}\right) \\
&\leq 2\left(1 + \frac{\bar{\rho}_T}{\lambda}\right) \left(d \ln\left(1 + \frac{\bar{\rho}_T}{\lambda}\right) + m \ln\left(1 + \frac{TL^2}{m\lambda}\right)\right) \\
&\leq 2\left(1 + \varepsilon_m\right) \left(d \ln\left(1 + \varepsilon_m\right) + m \ln\left(1 + \frac{TL^2}{m\lambda}\right)\right)
\end{aligned}$$

where the penultimate inequality follows from Lemma 3 and the last step follows from Proposition 1. □

Now we give a bound on the instantaneous regret.

Lemma 4. *For any $\delta > 0$, the instantaneous regret of SOFUL satisfies $(\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top \boldsymbol{w}^\star \leq 2\widetilde{\beta}_{t-1}(\delta) \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}}$ for $t = 1, \dots, T$.*

Proof. Let $\widetilde{\boldsymbol{w}}^{\text{SO}}_{t-1}$ be the FD-sketched RLS estimate of OFUL (Algorithm 5). Recall that the optimal action at time t is $\boldsymbol{x}_t^\star = \arg\max_{\boldsymbol{x} \in D_t} \boldsymbol{x}^\top \boldsymbol{w}^\star$, whereas

$$\left(\boldsymbol{x}_t, \widetilde{\boldsymbol{w}}^{\text{SO}}_{t-1}\right) = \arg\max \boldsymbol{x}^\top \boldsymbol{w} \quad (\boldsymbol{x}, \boldsymbol{w}) \in D_t \times \widetilde{C}_{t-1} \ .$$

We use these facts to bound the instantaneous regret,

$$\begin{aligned}
&\left(\boldsymbol{x}_t^\star - \boldsymbol{x}_t\right)^\top \boldsymbol{w}^\star \leq \boldsymbol{x}_t^\top \left(\widetilde{\boldsymbol{w}}^{\text{SO}}_{t-1} - \boldsymbol{w}^\star\right) \\
&= \boldsymbol{x}_t^\top \left(\widetilde{\boldsymbol{w}}^{\text{SO}}_{t-1} - \widetilde{\boldsymbol{w}}_{t-1}\right) + \boldsymbol{x}_t^\top \left(\widetilde{\boldsymbol{w}}_{t-1} - \boldsymbol{w}^\star\right)
\end{aligned}$$

$$\leq \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \left(\|\widetilde{\boldsymbol{w}}_{t-1}^{\mathrm{SO}} - \widetilde{\boldsymbol{w}}_{t-1}\|_{\widetilde{\boldsymbol{V}}_{t-1}} + \|\widetilde{\boldsymbol{w}}_{t-1} - \boldsymbol{w}^\star\|_{\widetilde{\boldsymbol{V}}_{t-1}} \right)$$
$$\leq 2\widetilde{\beta}_{t-1}(\delta) \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \qquad \text{(by Theorem 2)}$$

where we get penultimate inequality by Cauchy-Schwartz inequality. □

Proof of Theorem 3. Using Lemma 4 gives

$$\begin{aligned}
R_T &= \sum_{t=1}^{T} (\boldsymbol{x}_t^\star - \boldsymbol{x}_t)^\top \boldsymbol{w}^\star \\
&\leq 2 \sum_{t=1}^{T} \min \left\{ LS, \widetilde{\beta}_{t-1}(\delta) \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \right\} \qquad (14) \\
&\leq 2 \sum_{t=1}^{T} \widetilde{\beta}_{t-1}(\delta) \min \left\{ \frac{L}{\sqrt{\lambda}}, \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \right\} \qquad (15) \\
&\leq 2\widetilde{A} \sum_{t=1}^{T} \min \left\{ \frac{L}{\sqrt{\lambda}}, \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \right\} \\
&\leq 2\widetilde{A} \max \left\{ 1, \frac{L}{\sqrt{\lambda}} \right\} \sum_{t=1}^{T} \min \left\{ 1, \|\boldsymbol{x}_t\|_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \right\} \\
&\leq 2\widetilde{A} \max \left\{ 1, \frac{L}{\sqrt{\lambda}} \right\} \sqrt{T \sum_{t=1}^{T} \min \left\{ 1, \|\boldsymbol{x}_t\|^2_{\widetilde{\boldsymbol{V}}_{t-1}^{-1}} \right\}}
\end{aligned}$$

where we used $\widetilde{A} = \max_{t=0,\ldots,T-1} \widetilde{\beta}_t(\delta)$. Also, we get (14) since $\max_{t=1,\ldots,T} \max_{\boldsymbol{x} \in D_t} |\boldsymbol{x}^\top \boldsymbol{w}^\star| \leq LS$ by Cauchy-Schwartz, and (15) since $\min_{t=0,\ldots,T-1} \min_{\delta \in [0,1]} \widetilde{\beta}_t(\delta) \geq S\sqrt{\lambda}$. The last inequality is obtained via the Cauchy-Schwartz inequality. Now we finish by bounding $\widetilde{\beta}_t(\delta)$ using (11) and we bound the summation term using Lemma 1,

$$\begin{aligned}
R_T &\overset{\widetilde{\mathcal{O}}}{=} M_\lambda \sqrt{T} \left(R\sqrt{\widetilde{m}\,(1+\varepsilon_m)} + S\sqrt{\lambda}\,(1+\varepsilon_m) \right) \\
&\quad \cdot \sqrt{\widetilde{m}\,(1+\varepsilon_m)} \\
&\overset{\widetilde{\mathcal{O}}}{=} M_\lambda \sqrt{T} \left(R\,\widetilde{m}\,(1+\varepsilon_m) + S\sqrt{\lambda}\,(1+\varepsilon_m)^{\frac{3}{2}} \sqrt{\widetilde{m}} \right) \\
&\overset{\widetilde{\mathcal{O}}}{=} M_\lambda\,(1+\varepsilon_m)^{\frac{3}{2}}\,\widetilde{m} \left(R + S\sqrt{\lambda} \right) \sqrt{T} \qquad \square
\end{aligned}$$

7 Experiments

In this section we present a simple empirical evaluation of OFUL and linear TS against their sketched versions.

Setup. The idea of our experimental setup is similar to the one described by Cesa-Bianchi et al. (2013). We convert a K-class classification problem into a contextual bandit problem as follows: given a dataset of labeled instances $(\boldsymbol{x}, y) \in \mathbb{R}^d \times \{1, \ldots, K\}$, we partition it into K subsets according to the class labels. Then we create K sequences by drawing a random permutation of each subset. At each step t the decision set D_t is obtained by picking the t-th instance from each one of these K sequences. Finally, rewards are determined by choosing a class $y \in \{1, \ldots, K\}$ and then consistently assigning reward 1 to all instances labeled with y and reward 0 to all remaining instances. The reported mean and standard deviation of the cumulative reward are averaged over four random permutations of the K sequences. This procedure is used for all baselines.

Datasets. We perform experiments on six publicly available datasets for multiclass classification from the `openml` repository (Vanschoren et al., 2013), see the table below here for details.

Dataset	Examples	Features	Classes
Bank	45k	17	2
SatImage	6k	37	6
Spam	4k	58	2
Pendigits	11k	17	10
MFeat	2k	48	10
CMC	1.4k	10	3

Baselines. The hyperparameters β (confidence ellipsoid radius) and λ (RLS regularization parameter) are selected on a validation set of size 100 via grid search on $(\beta, \lambda) \in \{1, 10^2, 10^3, 10^4\} \times \{10^{-2}, 10^{-1}, 1\}$ for OFUL, and $\{1, 10^2, 10^3\} \times \{10^{-2}, 10^{-1}, 1, 10^2\}$ for linear TS.

Results The experiments we present in this section measure cumulative reward throughout time. In the first experiment we compare OFUL and linear TS to their sketched versions while varying the sketch size (in the plots the sketch size is expressed as a percentage of the context space dimension). Results for the three datasets are presented in Figure 1, while results for the remaining datasets can be found in the supplementary material. We observe that on two datasets out of three, sketched algorithms indeed do not suffer a substantial drop in performance when compared to the non-sketched ones, even when the sketch size amounts to 60% of the context space dimension. This demonstrates that sketching successfully captures relevant subspace information relatively to the goal of maximizing reward.

Because the FD-sketching procedure considered in this paper is essentially performing online PCA, it is natural to ask how our sketched algorithms would compare to their non-sketched version run on the best m-dimensional subspace (computed by running PCA on the entire dataset). In Figure 2, we compare OFUL and SOFUL (results for linear TS are included in the supplementary material). In particular, we keep 60%, 40%, and 20% of the top principal components, and notice that, like in Figure 1, there are cases with little or no loss in performance.

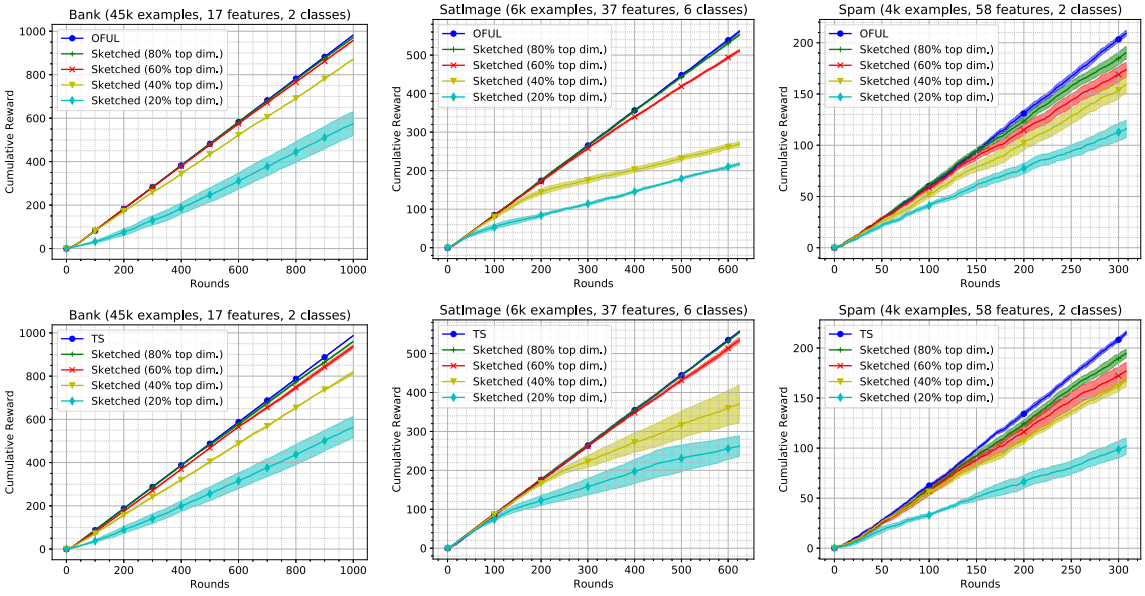

Figure 1: Comparison of SOFUL (first row) and sketched linear TS (second row) to their non-sketched variants on three real-world datasets and for different sketch sizes. Note that, in some cases, a sketch size equal to 80% and even 60% of the context space dimension does not significantly affect the perfomance.

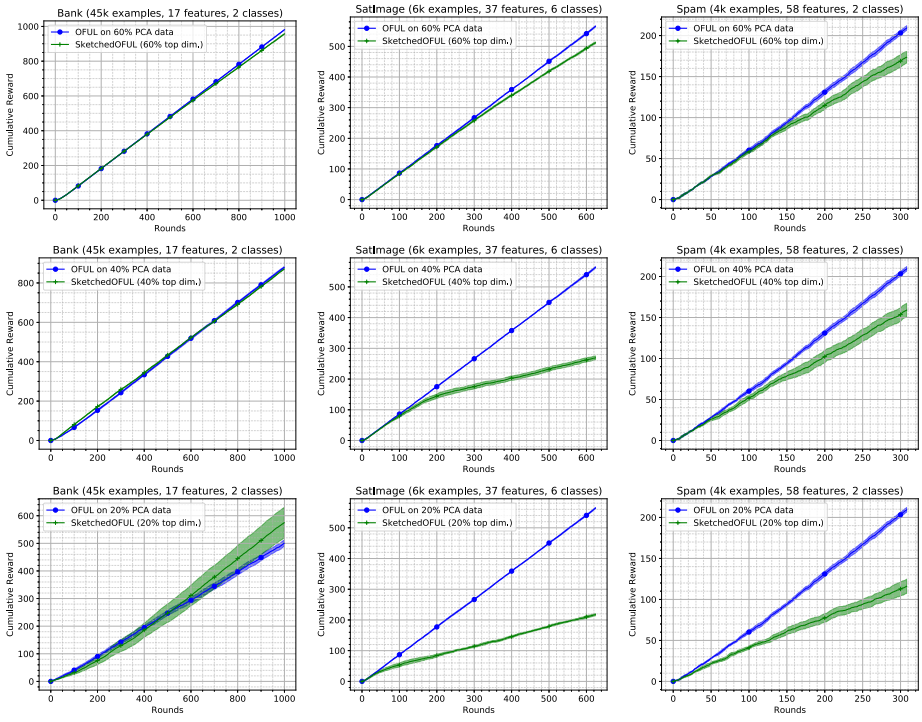

Figure 2: Comparison of OFUL run on the best m-dimensional subspace against SOFUL run with sketch size m. Rows show m as a fraction of the context space dimension: $60\%, 40\%, 20\%$, while columns correspond to different datasets. Note that, in some cases (with sketch size m of size at least 60%), SOFUL performs as well as if the best m-dimensional subspace had been known in hindsight.

References

Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In *Conference on Neural Information Processing Systems (NIPS)*, pages 2312–2320, 2011.

M. Abeille and A. Lazaric. Linear Thompson sampling revisited. *Electronic Journal of Statistics*, 11(2):5165–5197, 2017.

S. Agrawal and N. Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learing (ICML)*, pages 127–135, 2013.

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

D. Calandriello, A. Lazaric, and M. Valko. Efficient second-order online kernel learning with adaptive embedding. In *Conference on Neural Information Processing Systems (NIPS)*, pages 6140–6150, 2017.

N. Cesa-Bianchi, C. Gentile, and G. Zappella. A gang of bandits. In *Conference on Neural Information Processing Systems (NIPS)*, pages 737–745, 2013.

V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In *Conference on Computational Learning Theory (COLT)*, pages 355–366, 2008.

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

A. Gonen, F. Orabona, and S. Shalev-Shwartz. Solving ridge regression using sketched preconditioned SVRG. In *International Conference on Machine Learing (ICML)*, pages 1397–1405, 2016.

N. J. Higham. *Functions of matrices: theory and computation*, volume 104. Siam, 2008.

K.-S. Jun, A. Bhargava, R. Nowak, and R. Willett. Scalable generalized linear bandits: Online computation and hashing. In *Conference on Neural Information Processing Systems (NIPS)*, pages 99–109, 2017.

T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2018.

H. Luo, A. Agarwal, N. Cesa-Bianchi, and J. Langford. Efficient second order online learning by sketching. In *Conference on Neural Information Processing Systems (NIPS)*, pages 902–910, 2016.

J. Vanschoren, J. N. van Rijn, B. Bischl, and L. Torgo. OpenML: Networked Science in Machine Learning. *SIGKDD Explorations*, 15(2):49–60, 2013.

D. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1–2):1–157, 2014.

X. Yu, M. R. Lyu, and I. King. Cbrap: Contextual bandits with random projection. In *Conference on Artificial Intelligence (AAAI)*, pages 2859–2866, 2017.