Jointly Efficient and Optimal Algorithms for Logistic Bandits

Louis Faury Criteo AI Lab

Marc Abeille Criteo AI Lab

Kwang-Sung Jun University of Arizona

Clément Calauzènes Criteo AI Lab

Abstract

Logistic Bandits have recently undergone careful scrutiny by virtue of their combined theoretical and practical relevance. This research effort delivered statistically efficient algorithms, improving the regret of previous strategies by exponentially large factors. Such algorithms are however strikingly costly as they require $\Omega(t)$ operations at each round. On the other hand, a different line of research focused on computational efficiency ($\mathcal{O}(1)$ per-round cost), but at the cost of letting go of the aforementioned exponential improvements. Obtaining the best of both world is unfortunately not a matter of marrying both approaches. Instead we introduce a new *learning* procedure for Logistic Bandits. It yields confidence sets which sufficient statistics can be easily maintained online without sacrificing statistical tightness. Combined with efficient *planning* mechanisms we design fast algorithms which regret performance still match the problem-dependent lower-bound of Abeille et al. (2021). To the best of our knowledge, those are the first Logistic Bandit algorithms that simultaneously enjoy statistical and computational efficiency.

1 INTRODUCTION

Logistic Bandit. The Logistic Bandit (**LogB**) framework describes sequential decision making problems in which an agent receives structured binary bandit feedback for her decisions. This namely allows to model numerous real-world situations where actions are evaluated by success/failure feedback (*e.g.* click/no-click in ad-recommandation problems). From a theoretical standpoint, the **LogB** framework allows a neat and concise study of the interactions between non-linearity and the exploration/exploitation trade-off. Recent research efforts on this front were conducted by Faury et al. (2020); Abeille et al. (2021); Jun et al. (2021), relying on improved confidence sets for the design and analysis of regret-minimizing **LogB** algorithms. This has led to significant improvement over the seminal work of Filippi et al. (2010), deflating the regret bounds by exponentially large factors. Their approach testifies of the importance of a careful handling of non-linearity in order to achieve optimal performances (*i.e* matching the regret lower-bound from Abeille et al. (2021, Theorem 2)). From a learning-theoretic standpoint, this line of work brings the understanding of **LogB** almost to a tie with the Linear Bandit (**LinB**). It highlights that some highly non-linear **LogB** instances are easier to solve (in some sense) than their **LinB** counterparts and brings forward algorithms with largely improved practical performances (see Abeille et al. (2021, Section H)).

Limitations. A severe drawback of those improved **LogB** algorithms resides in their tremendous *computational* cost. For instance, the `OFULog-r` algorithm of Abeille et al. (2021) requires to maintain batch maximum-likelihood estimators (which cannot be updated recursively) and to solve at every round expensive convex programs. The computational hardness of those tasks (respectively related to the *learning* and *planning* mechanisms of the algorithm) largely exceeds their **LinB** counterparts and lead to a painfully slow algorithm - prohibitively so for situations where decisions must be made on the fly. As a result, statistically efficient yet fast **LogB** algorithms are still missing - which is the topic of this paper.

Main Contributions. Our main contribution is **(1)** a new *learning* procedure for **LogB**. It yields **(2)** a new confidence set which sufficient statistics can be maintained at each round with $\widetilde{\mathcal{O}}(1)$ operations, without sacrificing statistical tightness. Furthermore **(3)** the shape of this set enables the deployment of efficient *planning* strategies as a plug-in. This enables the design of computationally efficient algorithms whose regret guarantees match the lower-bound of Abeille et al. (2021). To the best of our knowledge, those **LogB** algorithms are the first to enjoy both statistical and

Proceedings of the 25$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

computational efficiency simultaneously. We summarize our contributions in Table 1.

Organization. We formally introduce the learning problem in Section 2 and discuss previous works, their limitations and remaining challenges. In Section 3 we describe our new estimation method, coined Efficient Local Learning for Logistic Bandits (`ECOLog`). We then analyze an optimistic algorithm leveraging this procedure and claim that it enjoys the same regret guarantees obtained by Abeille et al. (2021) while being critically less computationally hungry. We exhibit the main technical arguments needed to obtain this result and discuss potential extensions as well as limitations of our approach. In Section 4 we detail a variant of our algorithm, more complex but better suited for deployment in real-life situations. We provide similar guarantees for this algorithm and illustrate its good practical behavior with numerical simulations.

2 PRELIMINARIES

2.1 The Learning Problem

Setting. The **LogB** framework describes a repeated game between an agent and her environment. At each round, the agent selects an action (a vector in some Euclidean space) and receives a binary, Bernoulli distributed reward. More precisely, given an arm-set[1] $\mathcal{A} \subset \mathbb{R}^d$ the agent plays at each round t an arm $a_t \in \mathcal{A}$ and receives a stochastic reward r_{t+1} following:

$$r_{t+1} \sim \text{Bernoulli}\left(\mu(a_t^\mathsf{T}\theta_\star)\right) \ , \tag{1}$$

where $\mu(z) = (1+\exp(-z))^{-1}$ is the logistic function. The parameter $\theta_\star$ is unknown to the agent. We will work under the following assumption, standard for the study of **LogB**.

Assumption 1 (Bounded Decision Set)**.** *For any $a \in \mathcal{A}$ we have $\|a\| \leq 1$. Also, $\|\theta_\star\| \leq S$ where S is known.*

Denote $a_\star := \arg\max_{a\in\mathcal{A}} a^\mathsf{T}\theta_\star$ the best action in hindsight. The goal of the agent is to minimize her cumulative pseudo-regret up to time T:

$$\text{Regret}(T) := T\mu\big(a_\star^\mathsf{T}\theta_\star\big) - \sum_{t=1}^{T} \mu\big(a_t^\mathsf{T}\theta_\star\big) \ .$$

Reward Sensitivity. Central to the analysis of **LogB** is the inverse minimal reward sensitivity κ. This *problem-dependent* constant is defined as:

$$\kappa := 1/\min_{a\in\mathcal{A}} \min_{\|\theta\|\leq S} \dot{\mu}(a^\mathsf{T}\theta) \ .$$

[1]For the sake of exposition we here only consider the static arm-set case. As later detailed, our results also apply to time-varying arm-sets and contextual settings.

Briefly, κ measures the level of non-linearity of the reward signal, usually high in **LogB** problems. As such κ is typically very large (numerically) even for reasonable configurations. We refer the reader to Faury et al. (2020, Section 2) for a detailed discussion on the importance of this quantity.

Additional Notations. For any $t \geq 1$ we denote the $\mathcal{F}_t := \sigma(a_1, r_2, .., a_t)$ the σ-algebra encoding the information acquired after playing a_t and before observing r_{t+1}. Throughout the paper time indexes reflect the measurability w.r.t $\mathcal{F}_t$ (for example, a_t is $\mathcal{F}_t$-measurable but not $\mathcal{F}_{t+1}$-measurable). For any pair $(x, y) \in \mathbb{R}\times\{0, 1\}$ we define:

$$\ell(x, y) = -y\log\mu(x) - (1-y)\log(1-\mu(x)) \ ,$$

and the log-loss associated with the pair (a_t, r_{t+1}) writes $\ell_{t+1}(\theta) := \ell(a_t^\mathsf{T}\theta, r_{t+1})$. Given a compact set $\Theta \subset \mathbb{R}^d$ its diameter under the arm-set $\mathcal{A}$ is:

$$\text{diam}_\mathcal{A}(\Theta) = \max_{a\in\mathcal{A}} \max_{\theta_1,\theta_2} |a^\mathsf{T}(\theta_1 - \theta_2)| \ .$$

We will use throughout the paper the symbol $\mathfrak{C}$ to denote universal constants (*i.e* independent of S, κ, d or T) which exact value can vary at each occurrence. Similarly, we use the generic notation $\gamma_t(\delta)$ to denote various *slowly* growing functions - more precisely such that $\gamma_t(\delta) = \mathfrak{C}\text{poly}(S)d\log(t/\delta)$. The exact values for the different occurrences of such functions are carefully reported in the supplementary materials.

2.2 Previous Work, Limitations and Remaining Challenges

Being a member of the Generalized Linear Bandit family, the first algorithm for **LogB** was given by Filippi et al. (2010). Their algorithm enjoys a regret scaling as $\widetilde{\mathcal{O}}(\kappa d\sqrt{T})$ - which although tight in d and T, suffers from a prohibitive dependency in κ. Further, it is computationally inefficient as it requires the computation of a batch estimator for $\theta_\star$ at each round (see Appendix E.2 for a detailed discussion). This efficiency issue was fixed by Zhang et al. (2016); Jun et al. (2017); Ding et al. (2021) who proposed fully online estimation procedures. Their approaches however still suffer from detrimental dependencies in κ.

Statistical Optimality. The κ dependency was trimmed by Faury et al. (2020) who introduced an algorithm enjoying $\widetilde{\mathcal{O}}(d\sqrt{T})$ regret. Their approach defers the effect of non-linearity (embodied by κ) to a second-order term in the regret, dominated for large values of T. Similar results were also achieved by Dong et al. (2019), but only for the Bayesian regret. From a statistical viewpoint the story was closed by Abeille et al. (2021) who proved a $\Omega(d\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_\star)T})$ regret lower-bound and matching regret upper-bounds (up

Algorithm	Regret Bound	Cost Per-Round	Minimax	Efficient
`GLM-UCB` Filippi et al. (2010)	$\widetilde{\mathcal{O}}\left(\kappa d\sqrt{T}\right)$	$\mathcal{O}\left(d^2K + d^2T\right)$	✘	✘
`GLOC`, `OL2M` Jun et al. (2017) Zhang et al. (2016)	$\widetilde{\mathcal{O}}\left(\kappa d\sqrt{T}\right)$	$\mathcal{O}\left(d^2K\right)$	✘	✔
`OFULog-r` Abeille et al. (2021)	$\widetilde{\mathcal{O}}\left(d\sqrt{T\dot{\mu}(a_\star^\intercal\theta_\star)}\right)$	$\mathcal{O}\left(d^2KT\right)$	✔	✘
`(ada-)OFU-ECOLog` (this paper)	$\widetilde{\mathcal{O}}\left(d\sqrt{T\dot{\mu}(a_\star^\intercal\theta_\star)}\right)$	$\widetilde{\mathcal{O}}\left(d^2K\right)$	✔	✔

Table 1: Comparison of frequentist regret guarantees and computational cost for different **LogB** algorithms, on instances where $|\mathcal{A}| = K < +\infty$. An algorithm is called minimax-optimal if it matches the regret lower-bound of (Abeille et al., 2021, Theorem 2) and efficient if it matches the computational cost of **LinB** algorithms (up to logarithmic factors).

to logarithmic factors) for their algorithm `OFULog-r`. Given the typical scalings of $\kappa \propto \exp(\|\theta_\star\|)$ and $\dot{\mu}(a_\star^\intercal\theta_\star) \propto \exp(-\|\theta_\star\|)$ this deflates the regret of previous approaches by exponentially large factors.

Computational Cost. Albeit statistically optimal, the algorithms proposed by Abeille et al. (2021) are strikingly computationally demanding and consequently prohibitively slow for practical situations. After inspection, two main computational bottlenecks of their approach emerge from their *learning* and *planning* mechanisms. From the learning side, they construct confidence regions of the form:

$$\left\{\theta,\ \|\theta - \hat{\theta}_t\|^2_{\mathbf{H}_{t-1}(\theta)} \leq \gamma_t(\delta)\right\}\ , \tag{2}$$

where $\hat{\theta}_t = \arg\min_{\mathbb{R}^d} \sum_{s=1}^{t-1} \ell_{s+1}(\theta) + \lambda\|\theta\|^2$ and

$$\mathbf{H}_t(\theta) = \sum_{s=1}^{t} \dot{\mu}(a_s^\intercal\theta) a_s a_s^\intercal + \lambda \mathbf{I}_d\ .$$

Those sufficient statistics are expensive to compute as both require a linear pass (at least) on the data. Note that simply testing whether a point lies in this set is costly - it requires $\Omega(t)$ operations. The planning mechanism which leverages this confidence region suffers from this downside; to find an optimistic arm it must solve one expensive convex program per arm at every round. This program involves the complete log-loss, which evaluation also takes $\Omega(t)$ operations. Furthermore, bypassing optimism through randomized exploration (*e.g.* Thompson Sampling) is particularly challenging as the results of Agrawal and Goyal (2013); Abeille and Lazaric (2017) do not apply to non-ellipsoidal confidence regions.

Challenges. Our goal is to develop an *efficient* algorithm (*i.e* with reduced per-round computational cost) which still enjoys statistical optimality (*i.e* matches the lower-bound of Abeille et al. (2021)). In light of the previous discussion, a crucial step is to derive an alternative to the confidence set from Equation (2) which sufficient statistics can be updated at little cost. This must be done without sacrificing the confidence set's appreciation of the *effective* reward sensitivity, captured by the matrix $\mathbf{H}_t(\theta)$ and central for optimal performance. In other words, we seek to develop an efficient estimation procedure that captures the *local* effects of non-linearity. This rules out merging the refined concentration tools of Faury et al. (2020) with the online approaches of Zhang et al. (2016); Jun et al. (2017) which explicitly ressorts to *global* quantities (*e.g.* κ) in their estimation routines.

3 MAIN RESULTS

In this section we present our approach to address the aforementioned challenges. We introduce `OFU-ECOLog`, an optimistic algorithm whose pseudo-code is provided in Algorithm 1. It is built on top on three building blocks; **(1)** a short warm-up phase (forced-exploration) of size τ described in Procedure 1, **(2)** the `ECOLog` estimation procedure described in Procedure 2 and **(3)** an optimistic planning mechanism.

We provide in Section 3.1 theoretical guarantees for the regret of `OFU-ECOLog` (Theorem 1) and quantify its per-round computational cost (Proposition 1). It demonstrates that `OFU-ECOLog` enjoys both statistical and computational efficiency.

Each building block (**1-3**) and their specific roles are detailed in subsequent sections. Section 3.2 is concerned with the initial forced-exploration phase and its length τ. Section 3.3 details the estimation procedure `ECOLog` and the confidence region it induces. Section 3.4 details the efficient deployment of the optimistic exploration strategy and describes the extension of `OFU-ECOLog` to `TS-ECOLog`, where optimism is replaced with randomization.

Algorithm 1 `OFU-ECOLog`

input: failure level δ, warm-up length τ.
Set $\Theta \leftarrow \texttt{WarmUp}(\tau)$ (see Procedure 1). ▷ *forced-exploration*
Initialize $\theta_{\tau+1} \in \Theta$, $\mathbf{W}_{\tau+1} \leftarrow \mathbf{I}_d$ and $\mathcal{C}_{\tau+1}(\delta) \leftarrow \Theta$.
for $t \geq \tau + 1$ **do**
  Play $a_t \in \arg\max_{a \in \mathcal{A}} \max_{\theta \in \mathcal{C}_t(\delta)} a^\mathsf{T}\theta$. ▷ *planning*
  Observe reward r_{t+1}, construct loss $\ell_{t+1}(\theta) = \ell(a_t^\mathsf{T}\theta, r_{t+1})$.
  Compute $(\theta_{t+1}, \mathbf{W}_{t+1}) \leftarrow \texttt{ECOLog}(1/t, \Theta, \ell_{t+1}, \mathbf{W}_t, \theta_t)$ (see Procedure 2). ▷ *learning*
  Compute $\mathcal{C}_{t+1}(\delta) \leftarrow \left\{ \|\theta - \theta_{t+1}\|^2_{\mathbf{W}_{t+1}} \leq \gamma_t(\delta) \right\}$.
end for

Procedure 1 `WarmUp`

input: length τ.
Set $\lambda \leftarrow \gamma_\tau(\delta)$, initialize $\mathbf{V}_0 \leftarrow \lambda \mathbf{I}_d$.
for $t \in [1, \tau]$ **do**
  Play $a_t \in \arg\max_{\mathcal{A}} \|a\|_{\mathbf{V}_{t-1}^{-1}}$, observe r_{t+1}.
  Update $\mathbf{V}_t \leftarrow \mathbf{V}_{t-1} + a_t a_t^\mathsf{T}/\kappa$.
end for
Compute $\hat{\theta}_{\tau+1} \leftarrow \arg\min_\theta \sum_{s=1}^{\tau} \ell_{s+1}(\theta) + \lambda\|\theta\|^2$.
output: $\Theta = \left\{ \theta, \left\| \theta - \hat{\theta}_{\tau+1} \right\|^2_{\mathbf{V}_\tau} \leq \gamma_\tau(\delta) \right\}$.

Procedure 2 `ECOLog`

input: accuracy ε, convex set Θ, ℓ_{t+1}, $\mathbf{W}_t$, θ_t.
Compute $D \leftarrow \mathrm{diam}_{\mathcal{A}}(\Theta)$, set $\eta \leftarrow (2 + D)^{-1}$.
Solve to precision ε:

$$\theta_{t+1} = \arg\min_{\theta \in \Theta} \left[\eta \|\theta - \theta_t\|^2_{\mathbf{W}_t} + \ell_{t+1}(\theta) \right] .$$

Update $\mathbf{W}_{t+1} \leftarrow \mathbf{W}_t + \dot{\mu}(a_t^\mathsf{T}\theta_{t+1}) a_t a_t^\mathsf{T}$.
output: $\theta_{t+1}, \mathbf{W}_{t+1}$

3.1 Statistical and Computational Efficiency

We claim the following result, which proof is deferred to Appendix D.1.

Theorem 1 (Regret Bound). *Let $\delta \in (0, 1]$. Setting $\tau = \kappa S^6 \gamma_T(\delta)^2$ ensures the regret of* `OFU-ECOLog`(δ, τ) *satisfies with probability at least $1 - 2\delta$:*

$$\mathrm{Regret}(T) \leq \mathfrak{C} S d \sqrt{T \dot{\mu}(a_\star^\mathsf{T}\theta_\star)} \log(T/\delta) + \mathfrak{C} S^6 \kappa d^2 \log(T/\delta)^2 .$$

As promised the dominating term in `OFU-ECOLog`'s regret-bound matches the lower-bound of Abeille et al. (2021) and scales with the reward sensitivity at the best action $a_\star$. Further, the second-order term identically matches its counterpart from previous work in its scaling w.r.t d, T and κ. This establishes the statistical efficiency and we now move up to the computational cost. For this, we claim the following bound on the complexity of `OFU-ECOLog`.

Proposition 1 (Computational Cost). *Let $|\mathcal{A}| = K < \infty$. Each round t of* `OFU-ECOLog` *can be completed within $\mathcal{O}(Kd^2 + d^2 \log(t)^2)$ operations.*

The proof is deferred to Appendix E.1. This result mainly relies on the fact that the `ECOLog` routine (Procedure 2) solves convex programs that are cheap (*i.e* for which gradients are inexpensive to compute) and that can be efficiently preconditioned. Furthermore `OFU-ECOLog` leverages ellipsoidal confidence sets, for which optimism can be efficiently enforced (at least for finite arm-sets). This fulfills our promise of computational efficiency.

3.2 Warm-Up

One of the main challenge to avoid prohibitive exponential dependencies in **LogB** is to tightly control the reward sensitivity across $\mathcal{A} \times \Theta$ - that is, without resorting to global problem-dependent constants (*e.g.* κ). Following Faury et al. (2020) a first useful step in that direction is to leverage the self-concordance property of the logistic function. It ensures that for any $a \in \mathcal{A}$:

$$\forall \theta_1, \theta_2 \in \Theta, \ \dot{\mu}(a^\mathsf{T}\theta_1) \leq \dot{\mu}(a^\mathsf{T}\theta_2) \exp(\mathrm{diam}_{\mathcal{A}}(\Theta)) . \quad (3)$$

The role of the warm-up phase is to identify a set Θ containing $\theta_\star$ (with high probability) and which diameter is a constant, independent of problem-dependent quantities (*e.g.* $\|\theta_\star\|$ or S). The warm-up mechanism described in Procedure 1 constructs such a set Θ which diameter is controlled through the length τ of this forced-exploration phase. In particular, we show that if $\tau \propto \kappa$ (Proposition 5 in the appendix):

$$\mathrm{diam}_{\mathcal{A}}(\Theta) \leq 1 \text{ for } \Theta \leftarrow \texttt{WarmUp}(\tau) . \quad (4)$$

Combining Equations (3) and (4) allows to control the reward sensitivity across the set Θ at little cost (*i.e* independent of problem-dependent constants). Theoret-

ically speaking, the regret incurred during the warm-up phase forms a second-order term, dominated in the overall regret bound. From a practical perspective however, resorting to forced-exploration is inconvenient - a downside we address in Section 4.

Remark (Optimal Design)**.** *There exists alternatives warm-up strategies which ensures similar guarantees - see for instance Jun et al. (2021) for a solution based on optimal design. It involves more complex mechanisms to reduce the length τ - we stick here to a simple strategy for the sake of exposition.*

3.3 Efficient Local Learning

We now describe `ECOLog`, a new estimation routine summarized in Procedure 2 which is at the core of the online construction of tight confidence sets. It operates on the convex set Θ returned by the warm-up procedure. It maintains estimates $\{\theta_t\}_t$ of $\theta_\star$ following the update rule:

$$\theta_{t+1} = \arg\min_{\theta\in\Theta} \left[\eta \left\|\theta - \theta_t\right\|^2_{\mathbf{W}_t} + \ell_{t+1}(\theta)\right] , \quad (5)$$

$$\text{where } \mathbf{W}_t = \sum_{s=1}^{t-1} \dot\mu(a_s^\mathsf{T}\theta_{s+1}) a_s a_s^\mathsf{T} + \lambda \mathbf{I}_d \ .$$

The learning rate η is tied to the diameter $\text{diam}_{\mathcal{A}}(\Theta)$ of the decision set $\mathcal{A}\times\Theta$. After round t, the next estimate θ_{t+1} minimizes an approximation of the true cumulative log-loss $\sum_{s=1}^t \ell_{s+1}(\theta)$ that is decomposed in two terms. The first consists in a quadratic proxy for the past losses constructed through the sequence $\{\theta_s\}_{s\le t}$. It is designed to incorporate the information acquired so far in the update since:

$$\arg\min_\theta \sum_{s=1}^{t-1} \ell_{s+1}(\theta) \approx \arg\min_\theta \left\|\theta - \theta_t\right\|^2_{\mathbf{W}_t} \ .$$

On the other hand, the second term is the instantaneous log-loss $\ell_{t+1}(\theta)$ which accounts for the novel information of the pair (a_t, r_{t+1}). The motivation behind the overall structure of the update is the following: while the cumulative log-loss is strongly convex and can therefore be well approximated by a quadratic function, the instantaneous loss ℓ_{t+1} has flat tails which cannot be captured by a quadratic shape.

Remark (Comparison with ONS)**.** *While at first glance it resembles the Online Newton Step (ONS) mechanisms used by Zhang et al. (2016); Jun et al. (2017) there are two important differences. First, the update is driven by the matrix $\mathbf{W}_t$ which relies on the* estimated *reward sensitivity, and not on its worst-case alternative κ. Second, we do not rely on (potentially loose) approximations for ℓ_{t+1}. This rules out having access to a closed-form for θ_{t+1}.*

Since the solution of Equation (5) does not admit a closed-form expression, one can only solve it up to an ε accuracy (*e.g.* with projected gradient descent). Formally, we compute estimators θ'_{t+1} such that $\|\theta'_{t+1} - \theta_{t+1}\| \le \varepsilon$. The following statement guarantees that this can be done at little cost.

Proposition 2 (Computational Cost)**.** *Running* `ECOLog` *up to $\varepsilon > 0$ accuracy requires $\mathcal{O}(d^2 \log(1/\varepsilon)^2)$ operations.*

For the sake of exposition, we ignore optimization errors in the following since ε can be arbitrarily small. The induced errors and their propagation are addressed in formal proofs in the supplementary.

Finally, the use of `ECOLog` at each round within Algorithm 1 yields a sequence $\{\theta_t, \mathbf{W}_t\}_t$ associated with the sets:

$$\mathcal{C}_t(\delta) := \left\{\theta, \|\theta - \theta_t\|^2_{\mathbf{W}_t} \le \gamma_t(\delta)\right\} \ ,$$

which are confidence regions for $\theta_\star$.

Proposition 3 (Confidence Set)**.** *Under the conditions of Theorem 1:*

$$\mathbb{P}(\forall t \ge \tau,\ \theta_\star \in \mathcal{C}_t(\delta)) \ge 1-\delta \ .$$

Proposition 3 emulates the original concentration results of Faury et al. (2020) (see Equation (2)). The matrix $\mathbf{W}_t$ stands as an on-policy proxy for the "correct" concentration metric $\mathbf{H}_t(\theta_\star)$. This ultimately preserves statistical tightness (Theorem 1) but with sufficient statistics that are now updated online.

Proof Sketch. We provide here the key technical arguments behind the derivation of Proposition 3. It is inspired and shares close connections with the work of Jézéquel et al. (2020) - which was conducted for an Online Convex Optimization setting.
A crucial ingredient for our analysis is a *local* quadratic lower-bound[2] for the logistic loss, stating that for any $\theta \in \Theta$:

$$\ell_{t+1}(\theta_\star) \gtrsim \ell_{t+1}(\theta) + \nabla\ell_{t+1}(\theta)^\mathsf{T}(\theta_\star - \theta) + \dot\mu(a_t^\mathsf{T}\theta)(a_t^\mathsf{T}(\theta_\star - \theta))^2 \ .$$

Notice how the above does not depend on any *global* quantities (*e.g.* S or $\|\theta_\star\|$). It allows to tie the parameters uncertainty $\|\theta_\star - \theta_{t+1}\|^2_{\mathbf{W}_{t+1}}$ to the excess cumulative loss in $\{\theta_{s+1}\}_s$. Indeed algebraic manipulations lead to:

$$\|\theta_\star - \theta_{t+1}\|^2_{\mathbf{W}_{t+1}} \lesssim \sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1}) \ .$$

[2] Similar bounds appear in Jézéquel et al. (2020); Abeille et al. (2021) but are used for different purposes.

We are therefore left to bound the r.h.s. To do so we introduce an intermediary parameter:

$$\bar{\theta}_s = \arg\min_{\Theta} \eta\|\theta-\theta_s\|^2_{\mathbf{W}_s} + \ell(a_s^\mathsf{T}\theta, 0) + \ell(a_s^\mathsf{T}\theta, 1) \ ,$$

and decompose the sum to control as follows:

$$\sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s) + \sum_{s=1}^{t} \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) \ . \tag{6}$$

The parameter $\bar{\theta}_s$ is a $\mathcal{F}_s$-measurable version of θ_{s+1}, regularized in the last direction a_s by two logistic losses fitting antipodal rewards ($r_{s+1} = 0$ and 1). The first term in Equation (6) is tied to the stochastic nature of the observations and is bounded using the concentration inequality of Faury et al. (2020, Theorem 1). With probability at least $1-\delta$,

$$\sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s) \lesssim \log(t/\delta) \ .$$

Bounding the second term requires quantifying the deviation between θ_{s+1} and its $\mathcal{F}_s$-measurable counterpart $\bar{\theta}_s$. Leveraging convexity leads to the sequence of inequalities:

$$\begin{aligned}\sum_{s=1}^{t} \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) &\leq \sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)\|a_s\|^2_{\mathbf{W}^{-1}_{t+1}} \\ &\lesssim \sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\theta_{s+1})\|a_s\|^2_{\mathbf{W}^{-1}_{t+1}} \\ &\lesssim d\log(t) \ .\end{aligned}$$

The second inequality is obtained by relating the reward sensitivities $\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)$ and $\dot{\mu}(a_s^\mathsf{T}\theta_{s+1})$. Both are comparable thanks to the warm-up procedure. Indeed from Equations (3) and (4),

$$\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s) \leq \exp(\mathrm{diam}_{\mathcal{A}}(\Theta))\dot{\mu}(a_s^\mathsf{T}\theta_{s+1}) \lesssim \dot{\mu}(a_s^\mathsf{T}\theta_{s+1}) \ . \tag{7}$$

The last inequality directly follows from the Elliptical Potential Lemma (see Lemma 9).

Remark (Warm-Up and Online Newton Step)**.** *It is natural to wonder whether the ONS-like approaches of Zhang et al. (2016); Jun et al. (2017) could also benefit from the refined parameter set returned by the warm-up procedure. As detailed in Appendix B.3 this is not the case. Their respective methods hard-code* global *quantities within their updates steps (such as the minimum curvature of the log-loss, or the exp-concavity constant). Those are related to κ and cannot be removed even when operating close to $\theta_\star$.*

3.4 Exploration Strategy

Optimistic Exploration. `OFU-ECOLog` builds on $\mathcal{C}_t(\delta)$ (the confidence set of Proposition 3) to find an optimistic arm. Formally, it prescribes playing:

$$a_t \in \arg\max_{a\in\mathcal{A}} \max_{\theta\in\mathcal{C}_t(\delta)} a^\mathsf{T}\theta \ .$$

A solution for this program might be expensive to compute in general. However, the ellipsoidal nature of $\mathcal{C}_t(\delta)$ may simplify this task as it allows for an equivalent definition of a_t as:

$$a_t \in \arg\max_{a\in\mathcal{A}} a^\mathsf{T}\theta_t + \sqrt{\gamma_t(\delta)}\|a\|_{\mathbf{W}_t^{-1}} \ .$$

For finite arm-sets ($|\mathcal{A}| = K < +\infty$) this program can be solved by enumerating over the arms - bringing the total cost of the optimistic planning to $\mathcal{O}(d^2K)$.

Thompson-Sampling extension. The shape of $\mathcal{C}_{t+1}(\delta)$ also enables the use of *randomized* exploration mechanisms in a principled fashion. For instance, Thompson Sampling (TS) replaces the burden to find an optimistic parameter by sampling in slightly inflated confidence sets (see Abeille and Lazaric (2017)). It is often preferred in practical applications for its simplicity and good empirical performances. It also allows to deal with infinite arm-sets, whenever an oracle for computing $a_\star(\theta) = \arg\max_{a\in\mathcal{A}} a^\mathsf{T}\theta$ is cheaply available for any θ (*e.g.* when the action space is the unit-ball $\mathcal{B}_d$). We introduce `TS-ECOLog` in Appendix D.2, a TS version of `OFU-ECOLog`. It enjoys similar regret bounds, but inflated by a $\sqrt{d}$ factor (as in the **LinB** case). The algorithm displays little conceptual novelty compared to its linear counterpart, but its analysis requires additional technical care to prove its statistical efficiency. Overall, this answers positively the question opened by Faury et al. (2020) about the extension of their approach to randomized strategies.

4 REMOVING THE WARM-UP

Practical Limitations. Despite being rather common in the Generalized Linear Bandit literature (*e.g.* (Li et al., 2017; Kveton et al., 2020; Jun et al., 2021; Ding et al., 2021)) the use of warm-up phases is concerning from a practical stand-point. Indeed **(1)** it *hard-codes* a forced-exploration regime lasting at least κ rounds at the beginning of any experiment. Given the typical scaling of κ in practical situations this implies that the algorithm selects actions at random for the first few thousand steps. While it only impacts low-order terms in the regret bound, it is problematic to suffer this price by design, even when not necessary (see Abeille et al. (2021, Section 4)). Furthermore, **(2)** generalizing warm-up phases to handle *contextual* arm-sets requires adopting strong distributional assumptions on the contexts - leaving out the case where an adversary picks context.

Algorithm 2 `ada-OFU-ECOLog`

input: failure level δ.
Initialize $\Theta_1 = \{\|\theta\| \leq S\}$, $\mathcal{C}_1(\delta) \leftarrow \Theta_1$, $\theta_1 \in \Theta$, $\mathbf{W}_1 \leftarrow \mathbf{I}_d$ and $\mathcal{H}_1 \leftarrow \emptyset$.
for $t \geq 1$ **do**
  Play $a_t \in \arg\max_{a\in\mathcal{A}} \max_{\theta\in\mathcal{C}_t(\delta)} a^\mathsf{T}\theta$, observe reward r_{t+1}.
  Compute the estimators θ_t^0, θ_t^1 (see Equation (8)) and $\bar{\theta}_t$.
  if $\dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t) \leq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^0)$ and $\dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t) \leq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^1)$ **then**
    Form the loss ℓ_{t+1} and compute $(\theta_{t+1}, \mathbf{W}_{t+1}) \leftarrow \texttt{ECOLog}(1/t, \Theta_t, \ell_{t+1}, \mathbf{W}_t, \theta_t)$.
    Compute $\mathcal{C}_{t+1}(\delta) \leftarrow \left\{\|\theta - \theta_{t+1}\|^2_{\mathbf{W}_{t+1}} \leq \gamma_t(\delta)\right\}$, set $\mathcal{H}_{t+1} \leftarrow \mathcal{H}_t$.
  else
    Set $\mathcal{H}_{t+1} \leftarrow \mathcal{H}_t \cup \{a_t, r_{t+1}\}$ and compute $\hat{\theta}^{\mathcal{H}}_{t+1} = \arg\min \sum_{(a,r)\in\mathcal{H}_{t+1}} \ell(a^\mathsf{T}\theta, r) + \gamma_t(\delta)\|\theta\|^2$.
    Update $\mathbf{V}_t^{\mathcal{H}} \leftarrow \sum_{a\in\mathcal{H}_{t+1}} aa^\mathsf{T}/\kappa + \gamma_t(\delta)\mathbf{I}_d$, $\theta_{t+1} \leftarrow \theta_t$ and $\mathbf{W}_{t+1} \leftarrow \mathbf{W}_t$.
    Compute $\Theta_{t+1} = \left\{\|\theta - \hat{\theta}^{\mathcal{H}}_{t+1}\|^2_{\mathbf{V}_t^{\mathcal{H}}} \leq \gamma_t(\delta)\right\}$.
  end if
end for

Data-Driven Alternative. We relied so far on warm-up phases to isolate the different challenges (locality, efficiency, statistical tightness). We now switch gears and propose a refined approach which addresses the issues raised by forced-exploration - however at the cost of a more intricate algorithm. At the heart of this refinement lies a *data-dependent* version of our confidence set. This allows **(1)** the design an *adaptive* mechanism which preserves statistical efficiency while ultimately removing the need for forced-exploration. Furthermore, it **(2)** extends the regret bound derived in Section 3 to the contextual case, without requiring any distributional assumptions on the exogenous contexts. To our knowledge, this is a first for approaches resorting to warm-ups.

4.1 An Adaptive Approach

Intuition. As highlighted by the proof sketch from Section 3.3, the warm-up phase allows to tightly control the radius of $\mathcal{C}_t(\delta)$. More precisely, it constructs a small admissible set Θ that constrains the reward sensitivities $\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)$ and $\dot{\mu}(a_s^\mathsf{T}\theta_{s+1})$ to be comparable for all s (see Equation (7)).

A naive way to remove the need for enforcing this property *a priori* would be to reject *on-the-fly* points that don't conform with the following condition:

$$\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s) \leq 2\dot{\mu}(a_s^\mathsf{T}\theta_{s+1}) \,, \tag{C_0}$$

This high-level idea is behind the design of our adaptive mechanism, detailed below.

Adaptive Mechanism. Given (a_s, r_{s+1}) if the associated θ_{s+1} breaks (C_0) we do not use it to update our current estimate. Instead we leverage this information to ensure that (C_0) is more likely to hold in the future. We maintain $\mathcal{H}_{s+1} = \{a_l, r_{l+1}\}_{l\leq s}$ formed by pairs rejected up to round s and compute:

$$\hat{\theta}^{\mathcal{H}}_{s+1} \in \arg\min \sum_{(a,r)\in\mathcal{H}_{s+1}} \ell(a^\mathsf{T}\theta, r) + \gamma_s(\delta)\|\theta\|^2 \,,$$

and $\mathbf{V}_s^{\mathcal{H}} = \sum_{a\in\mathcal{H}_s} aa^\mathsf{T}/\kappa + \gamma_s(\delta)\mathbf{I}_d$. We use this to build the parameter set:

$$\Theta_{s+1} = \left\{\theta,\ \left\|\theta - \hat{\theta}^{\mathcal{H}}_{s+1}\right\|^2_{\mathbf{V}_s^{\mathcal{H}}} \leq \gamma_s(\delta)\right\} \,.$$

We will use this convex set in the `ECOLog` procedure for subsequent rounds. As points are being added to $\{\mathcal{H}_s\}_s$ the sequence of $\{\Theta_s\}_s$ deflates. The downstream estimates $\{\theta_{s+1}, \bar{\theta}_s\}_s$ are therefore closer and (C_0) is more likely to hold. The key to assert the validity of this mechanism is to ensure that this sequential refinement does not occur too often.

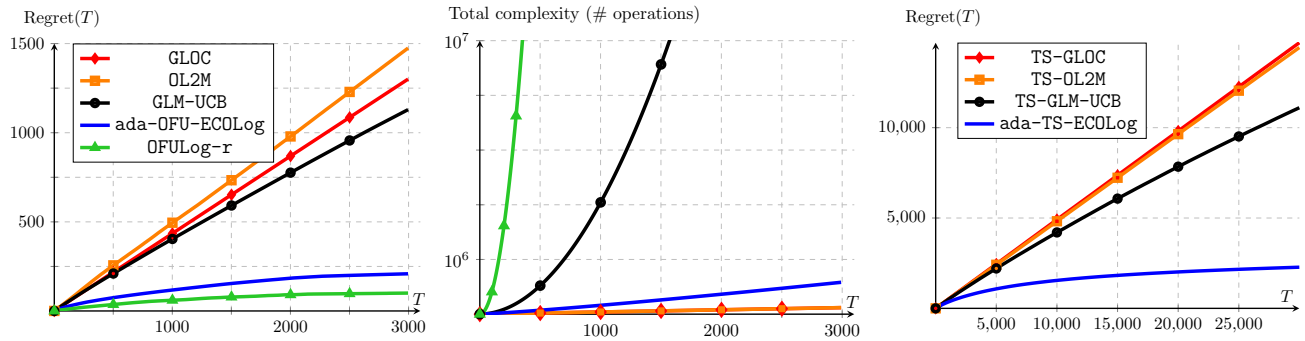
Technical Adjustment. The idea presented above needs a slight technical refinement to bear a principled algorithm. (C_0) prescribes filtering the arm a_s according to θ_{s+1}, an $\mathcal{F}_{s+1}$-adapted quantities. This breaches concentration properties we need to prove low-regret. To circumvent this issue we fall back on an $\mathcal{F}_s$-adapted condition covering the potential values of θ_{s+1} (depending on the realization of r_{s+1}). Let:

$$\theta_s^u = \arg\min_{\theta\in\Theta_s} \left[\eta\,\|\theta - \theta_s\|^2_{\mathbf{W}_s} + \ell(a_s^\mathsf{T}\theta, u)\right] \,, \tag{8}$$

for $u \in \{0, 1\}$. Note that θ_{s+1} is either θ_s^0 or θ_s^1. We replace (C_0) by the condition:

$$\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s) \leq 2\dot{\mu}(a_s^\mathsf{T}\theta_s^u), \quad \forall u \in \{0, 1\} \,. \tag{C_1}$$

The algorithm `ada-OFU-ECOLog` presented in Algorithm 2 combines this adjustment with the aforementioned adaptive mechanism.



$d = 2$, $|\mathcal{A}| = 20$, $\kappa = 400$ | $d = 2$, $|\mathcal{A}| = 20$, $\kappa = 400$ | $d = 5$, $\mathcal{A} = \mathcal{B}_5$, $\kappa = 400$.

Figure 1: Numerical simulations on **LogB** problems. We implement algorithms as prescribed by theory (*e.g.* we do not tune exploration) and average regret curves over 100 independent trajectories. *(left)* Regret curves on a two-dimensional **LogB** problem with 20 arms, sampled at random within the unit ball. We chose a small number of arms along with a short horizon to allow `OFULog-r` to run in a reasonable time. *(center)* Overall complexity of the different algorithms for this same instance. As hinted by the regret and complexity bounds, `ada-OFU-ECOLog` is the only one displaying good performances and at little computational cost. *(right)* Numerical simulations with infinite arm-set (5-dimensional unit-ball) for which we evaluate the TS version of each algorithm (this excludes `OFULog-r` which does not have a straightforward TS extension).

4.2 Theoretical Guarantees

Regret Bound. Thanks to its adaptivity, we can claim regret guarantees for `ada-OFU-ECOLog` in contextual settings - *i.e* that holds for *any* sequence of time-varying arm-set $\{\mathcal{A}_t\}_{t\geq 1}$.

Theorem 2. *Let $\delta \in [0,1)$. With probability at least $1-\delta$ the regret of* `ada-OFU-ECOLog`(δ) *satisfies:*

$$\mathrm{Regret}(T) \leq \mathfrak{C} S d \sqrt{\sum_{t=1}^{T} \dot{\mu}(a_{\star,t}^\mathsf{T}\theta_\star) \log(T/\delta)} + \mathfrak{C} S^6 \kappa d^2 \log(T/\delta)^2 \,,$$

where $a_{\star,t} = \arg\max_{a\in\mathcal{A}_t} a^\mathsf{T}\theta_\star$.

The proof is deferred to Appendix D.3. This result establishes similar (although more general) regret guarantees than Theorem 1. The two bounds are identical for constant arm-sets ($\mathcal{A}_t \equiv \mathcal{A}$). In the contextual case, the leading term is $\sqrt{T}\sqrt{\sum_{t=1}^{T} \dot{\mu}(a_{\star,t}^\mathsf{T}\theta_\star)/T}$, replacing the reward sensitivity at the optimal action by its on-trajectory average version.

Computational Cost. The per-round computational cost of Algorithm 2 is larger than `OFU-ECOLog` as it sometimes requires $\mathcal{O}(|\mathcal{H}_t|)$ extra operations to compute $\hat{\theta}_t^{\mathcal{H}}$. The computational overhead is small as we can prove that $|\mathcal{H}_t| \lesssim \kappa$.

Proposition 4. *The per-round computational cost of Algorithm 2 is bounded by $\mathcal{O}\left(\kappa + Kd^2 + d^2\log(T)^2\right)$.*

This extra computation is needed only when violating (C_1) which happens at most for κ rounds.

Adaptivity in Practice. Theorem 2 and Proposition 4 establish that hard-coding a warm-up phase can be avoided with little to no impact on the worst-case performance or computational cost. The adaptive nature of `ada-OFU-ECOLog` allows to enjoy stronger empirical performances in "nice" configurations. For instance, in all the numerical experiments that follow we found that the condition (C_1) was *never* triggered. In such cases, Algorithm 2 simply reduces to `OFU-ECOLog` *without* any forced-exploration. This is consistent with the analysis of Abeille et al. (2021, Section 4) which suggests that low-order κ dependencies (introduced here by the warm-up) can sometimes be avoided.

4.3 Numerical Simulations

The numerical illustrations presented in Figure 1 are consistent with our theoretical findings summarized in Table 1.[3] As predicted, `ada-OFU-ECOLog` enjoys best-of-both-worlds properties by displaying small regret and small computational cost. Additional numerical illustrations for **LogB** instances of higher dimensions can be found in Appendix G.

The value of κ for **LogB** instances we consider are reasonable (comparable to real-life situations). Still, it precludes the use of warm-up phase in practice for it will simply last longer than the horizon we consider (for which `OFULog-r` and `ada-OFU-ECOLog` already exhibits asymptotic behavior). Note that this is even worse for other approaches using forced-exploration (Kveton et al., 2020; Ding et al., 2021) as their re-

[3] For reproducing experiments, see https://github.com/criteo-research/logistic_bandit.

spective warm-ups are typically even longer ($\propto \kappa^2$). Finally, we report results for an infinite arm-set for which our approach yields the only tractable algorithm enjoying statistical efficiency.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved Algorithms for Linear Stochastic Bandits. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Abeille, M., Faury, L., and Calauzènes, C. (2021). Instance-Wise Minimax-Optimal Algorithms for Logistic Bandits. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 3691–3699. PMLR.

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

Agrawal, S. and Goyal, N. (2013). Thompson Sampling for Contextual Bandits with Linear Payoffs. In *Proceedings of the 30th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 127–135, Atlanta, Georgia, USA. PMLR.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Ding, Q., Hsieh, C.-J., and Sharpnack, J. (2021). An Efficient Algorithm For Generalized Linear Bandit: Online Stochastic Gradient Descent and Thompson Sampling . In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1585–1593. PMLR.

Dong, S., Ma, T., and Van Roy, B. (2019). On the Performance of Thompson Sampling on Logistic Bandits. In Beygelzimer, A. and Hsu, D., editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1158–1160, Phoenix, USA. PMLR.

Faury, L., Abeille, M., Calauzènes, C., and Fercoq, O. (2020). Improved Optimistic Algorithms for Logistic Bandits. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3052–3060, Virtual. PMLR.

Filippi, S., Cappe, O., Garivier, A., and Szepesvári, C. (2010). Parametric Bandits: The Generalized Linear Case. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc.

Golub, G. H. and van Loan, C. F. (2013). *Matrix Computations*. Johns Hopkins University Press, fourth edition.

Hazan, E. (2016). Introduction to Online Convex Optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325.

Jézéquel, R., Gaillard, P., and Rudi, A. (2020). Efficient Improper Learning for Online Logistic Regression. In *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2085–2108. PMLR.

Jun, K.-S., Bhargava, A., Nowak, R., and Willett, R. (2017). Scalable Generalized Linear Bandits: Online Computation and Hashing. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Jun, K.-S., Jain, L., Mason, B., and Nassif, H. (2021). Improved Confidence Bounds for the Linear Logistic Model and Applications to Bandits. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5148–5157. PMLR.

Kveton, B., Zaheer, M., Szepesvari, C., Li, L., Ghavamzadeh, M., and Boutilier, C. (2020). Randomized Exploration in Generalized Linear Bandits. In Chiappa, S. and Calandra, R., editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2066–2076. PMLR.

Li, L., Lu, Y., and Zhou, D. (2017). Provably Optimal Algorithms for Generalized Linear Contextual Bandits. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2071–2080. PMLR.

Valko, M., Munos, R., Kveton, B., and Kocák, T. (2014). Spectral Bandits for Smooth Graph Functions. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 46–54, Bejing, China. PMLR.

Zhang, L., Yang, T., Jin, R., Xiao, Y., and Zhou, Z.-h. (2016). Online Stochastic Linear Optimization under One-bit Feedback. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 392–401, New York, New York, USA. PMLR.

Jointly Efficient and Optimal Algorithms for Logistic Bandits Supplementary Material

ORGANIZATION OF THE APPENDIX

This appendix is organized as follows:

- In Appendix A we recall important notations and introduce some central inequalities.
- In Appendix B we link the length of the warm-up to the diameter of the set Θ it returns.
- In Appendix C we prove that $\mathcal{C}_t(\delta)$ is a confidence region for $\theta_\star$.
- In Appendix D we prove the different regret upper-bounds announced in the main paper
- In Appendix E we detail the computational cost of the different approaches discussed in the main paper.
- In Appendix F we list some auxiliary results, needed for the analysis.
- In Appendix G we provide additional numerical illustrations.

Table of Contents

A PRELIMINARIES

A.1 Notations

We detail below useful notations that will be used throughout the appendix. Below $T \in \mathbb{N}^+$, U is a set, $\Theta \subset \mathbb{R}^d$ is a compact set, $a \in \mathcal{A}$, $r \in \{0,1\}$ and $x, y \in \mathbb{R}^d$.

$[T]$	the set of integers from 1 to T.
$\lvert U\rvert$	cardinality of U.
$\text{diam}(\Theta) = \max_{\theta_1,\theta_2} \lVert\theta_1 - \theta_2\rVert$	diameter of Θ.
$\text{diam}_{\mathcal{A}}(\Theta) = \max_{a\in\mathcal{A}} \max_{\theta_1,\theta_2} \lvert a^\mathsf{T}(\theta_1 - \theta_2)\rvert$	diameter of Θ under $\mathcal{A}$.
$\mu(x) = (1+\exp(-x))^{-1}$	the logistic function at x.
$\ell(x,r) = -r\log\mu(x) - (1-r)\log(1-\mu(x))$	log-loss associated to (x,r).
$\ell_{t+1}(\theta) = \ell(a_t^\mathsf{T}\theta, r_{t+1})$	instantaneous log-loss of θ at round t.
$\bar{\ell}_{t+1}(\theta) = \ell(a_t^\mathsf{T}\theta, 1-r_{t+1})$	"reverse" instantaneous log-loss of θ at round t.
$\mathbf{H}_t(\theta) = \sum_{s=1}^t \dot{\mu}(a_s^\mathsf{T}\theta) a_s a_s^\mathsf{T} + \lambda \mathbf{I}_d$	Hessian of the cumulative-log loss at θ up to t.
$\mathbf{V}_t = \sum_{s=1}^t a_s a_s^\mathsf{T}/\kappa + \lambda \mathbf{I}_d$	linear-like design-matrix up to t.

Note that for all θ s.t $\lVert\theta\rVert \leq S$ we have $\dot{\mu}(a_s^\mathsf{T}\theta) \geq 1/\kappa$ by definition of κ. Therefore:

$$\forall \theta \text{ s.t } \lVert\theta\rVert \leq S, \quad \mathbf{H}_t(\theta) \succeq \mathbf{V}_t \ . \tag{9}$$

Below we define several "slowly growing" functions (uniformly denoted $\gamma_t(\delta)$ in the main paper). They will be used throughout the proofs.

$$\lambda_t(\delta) = d\log((4+t/4)/\delta) \ , \tag{10}$$
$$\gamma_t(\delta) = (S+3/2)^2\lambda_t(\delta) \ , \tag{11}$$
$$\beta_t(\delta) = \left(5/2 + (S+3/2)^2 + S\right)^2 \gamma_t(\delta) \ , \tag{12}$$
$$\nu_t(\delta) = 1/2 + 2\log\left(2\sqrt{t/4+1}/\delta\right) \ , \tag{13}$$
$$\sigma_t(\delta) = 8S^2 + 6 + 4\log(t) + 9\nu_t(\delta) + 18\exp(1)d\log(1+t/(4d)) \ , \tag{14}$$
$$\eta_t(\delta) = 4 + 4\log(t) + 16S^2 + (2+2S)^2\nu_t(\delta)/2 + 8(1+S)d\log(1+t/d) \ . \tag{15}$$

A.2 Useful Inequalities and Self-Concordant Control

A central idea when analyzing **LogB** is to tightly link *estimation* errors (*e.g.* between θ_1 and θ_2) to *prediction* errors (*e.g.* between $\mu(a^\mathsf{T}\theta_1)$ and $\mu(a^\mathsf{T}\theta_2)$). Exact Taylor expansion is a powerful tool to achieve this; as for previous works we will use it abundantly and in the following lines we introduce useful notations to this end. Specifically, for any $a \in \mathcal{A}$ and $x, y \in \mathbb{R}$ define:

$$\alpha(x,y) = \int_{v=0}^1 \dot{\mu}(x + v(y-x))dv = \alpha(y,x) \ , \tag{16}$$
$$\widetilde{\alpha}(x,y) = \int_{v=0}^1 (1-v)\dot{\mu}(x + v(y-x))dv \ . \tag{17}$$

After exact Taylor expansions we have the following identities for all $\theta_1, \theta_2 \in \mathbb{R}^d$:

$$\mu(a^\mathsf{T}\theta_2) - \mu(a^\mathsf{T}\theta_1) = \alpha(a^\mathsf{T}\theta_1, a^\mathsf{T}\theta_2)a^\mathsf{T}(\theta_2 - \theta_1) \ , \tag{18}$$
$$\ell_{t+1}(\theta_2) - \ell_{t+1}(\theta_1) = \nabla\ell_{t+1}(\theta_1)^\mathsf{T}(\theta_2 - \theta_1) + \tilde{\alpha}(a^\mathsf{T}\theta_1, a^\mathsf{T}\theta_2)(a^\mathsf{T}(\theta_2 - \theta_1))^2 \ . \tag{19}$$

Below are reminded some useful inequalities that stem from the self-concordance property of the logistic function (the fact that $\lvert\ddot{\mu}\rvert \leq \dot{\mu}$) The proofs can all be found in Appendix F of Abeille et al. (2021).

$$\alpha(x,y) \geq (1+\lvert x-y\rvert)^{-1}\dot{\mu}(z) \text{ for } z \in \{x,y\} \ , \tag{20}$$
$$\widetilde{\alpha}(x,y) \geq (2+\lvert x-y\rvert)^{-1}\dot{\mu}(x) \ , \tag{21}$$
$$\dot{\mu}(x) \leq \dot{\mu}(y)\exp\left(\lvert x-y\rvert\right) \ . \tag{22}$$

Procedure 1 `WarmUp` (detailed)

input: length τ.
Set $\lambda \leftarrow \lambda_\tau(\delta)$, initialize $\mathbf{V}_0 \leftarrow \lambda \mathbf{I}_d$. ▷ $\lambda_t(\delta)$ *is defined in Equation (10)*
for $t \in [1, \tau]$ **do**
Play $a_t \in \arg\max_{\mathcal{A}} \|a\|_{\mathbf{V}_{t-1}^{-1}}$, observe r_{t+1}.
Update $\mathbf{V}_t \leftarrow \mathbf{V}_{t-1} + a_t a_t^\mathsf{T}/\kappa$.
end for
Compute $\hat{\theta}_{\tau+1} \leftarrow \arg\min_\theta \sum_{s=1}^{\tau} \ell_{s+1}(\theta) + \lambda\|\theta\|^2/2$.
output: $\Theta = \left\{\theta, \left\|\theta - \hat{\theta}_{\tau+1}\right\|_{\mathbf{V}_\tau}^2 \le \beta_\tau(\delta)\right\}$. ▷ $\beta_t(\delta)$ *is defined in Equation (12)*

B WARM-UP PROCEDURE

We recall the warm-up procedure in Procedure 1 for which we now we give the exact values for the "slowly growing" functions that we use.

B.1 Warm-up Length and Parameter Set

The goal of this section is to prove the claim behind Equation (4), tying the length of the warm-up phase to the diameter of the induced parameter set Θ. The formal claim is made explicit in the following proposition.

Proposition 5. *Let $\delta \in (0, 1]$. Setting $\tau = \mathfrak{C}\kappa S^6 d^2 \log(T/\delta)^2$ ensures that Θ returned by* `WarmUp`(τ) *satisfies:*

$$(1)\ \ \mathbb{P}(\theta_\star \in \Theta) \ge 1 - \delta\ ,$$
$$(2)\ \ \mathrm{diam}_{\mathcal{A}}(\Theta) \le 1\ .$$

Proof. The set Θ returned by `WarmUp`(τ) is:

$$\Theta = \left\{\theta, \left\|\theta - \hat{\theta}_{\tau+1}\right\|_{\mathbf{V}_\tau}^2 \le \beta_\tau(\delta)\right\}\ . \tag{23}$$

where $\beta_t(\delta)$ is defined in Equation (12). It satisfies $\beta_t(\delta) \le \mathfrak{C}S^6 d \log(t/\delta)$.
To prove (1) we claim Lemma 1 which proof is deferred to Appendix C.1.

Lemma 1. *Let $\delta \in (0, 1]$. Then:*

$$\mathbb{P}\Big(\forall t \ge 1,\ \left\|\theta_\star - \hat{\theta}_{t+1}\right\|_{\mathbf{H}_t(\theta_\star)}^2 \le \beta_t(\delta)\Big) \ge 1 - \delta\ .$$

The proof of (1) directly follows:

$$\begin{aligned}
\mathbb{P}\left(\theta_\star \in \Theta\right) &= \mathbb{P}\left(\left\|\theta_\star - \hat{\theta}_{\tau+1}\right\|_{\mathbf{V}_\tau}^2 \le \beta_\tau(\delta)\right) && \text{(def. of } \Theta\text{)} \\
&\ge \mathbb{P}\left(\left\|\theta_\star - \hat{\theta}_{\tau+1}\right\|_{\mathbf{H}_\tau(\theta_\star)}^2 \le \beta_\tau(\delta)\right) && (\mathbf{V}_\tau \preceq \mathbf{H}_\tau(\theta_\star),\ \text{Equation (9)}) \\
&\ge 1 - \delta\ . && \text{(Lemma 1)}
\end{aligned}$$

To prove (2) we claim Lemma 2 which proof is provided in Appendix B.2:

Lemma 2. *Let $T \in \mathbb{N}^+$ and $\tau \in [T]$. Let Θ the set returned by* `WarmUp`(τ). *Then:*

$$\mathrm{diam}_{\mathcal{A}}(\Theta) \le 4\sqrt{\frac{\kappa \beta_T(\delta) d \log(1+T)}{\tau}}\ .$$

Therefore $\tau = 16\kappa d \beta_T(\delta) \log(1+T)$ ensures that $\mathrm{diam}_{\mathcal{A}}(\Theta) \le 1$. Since $\beta_T(\delta) \le \mathfrak{C}S^6 d \log(T/\delta)$ setting:

$$\tau = \mathfrak{C}\kappa S^6 d^2 \log(T/\delta)^2\ ,$$

yields $\mathrm{diam}_{\mathcal{A}}(\Theta) \le 1$ which finishes proving (2). □

B.2 Proof of Lemma 2

Lemma 2. *Let $T \in \mathbb{N}^+$ and $\tau \in [T]$. Let Θ the set returned by* `WarmUp`(τ). *Then:*

$$\mathrm{diam}_{\mathcal{A}}(\Theta) \leq 4\sqrt{\frac{\kappa\beta_T(\delta)d\log(1+T)}{\tau}} \ .$$

Proof. The proof is inspired by the demonstration of Lemma 8 of from Valko et al. (2014). Recall:

$$\Theta = \left\{\theta, \left\|\theta - \hat{\theta}_{\tau+1}\right\|^2_{\mathbf{V}_\tau} \leq \beta_\tau(\delta)\right\} \ ,$$

with $\mathbf{V}_\tau = \sum_{s=1}^{\tau} a_s a_s^\mathsf{T}/\kappa + \lambda_\tau(\delta)\mathbf{I}_d$ and for all $s \leq \tau$:

$$a_s \in \arg\max_{\mathcal{A}} \|a\|_{\mathbf{V}_{s-1}^{-1}} \ . \tag{24}$$

Therefore:

$$\begin{aligned}
\mathrm{diam}_{\mathcal{A}}(\Theta) &= \max_{a\in\mathcal{A}}\max_{\theta_1,\theta_2} |a^\mathsf{T}(\theta_1-\theta_2)| \ . \\
&\leq \max_{a\in\mathcal{A}}\max_{\theta_1,\theta_2} \|a\|_{\mathbf{V}_\tau^{-1}}\|\theta_1-\theta_2\|_{\mathbf{V}_\tau} && \text{(Cauchy-Schwarz)} \\
&\leq 2\sqrt{\beta_\tau(\delta)}\max_{a\in\mathcal{A}}\|a\|_{\mathbf{V}_\tau^{-1}} && (\theta_1,\theta_2\in\Theta) \\
&= 2\sqrt{\beta_\tau(\delta)}\sqrt{\max_{a\in\mathcal{A}}\|a\|^2_{\mathbf{V}_\tau^{-1}}} \\
&= 2\sqrt{\beta_\tau(\delta)}\tau^{-1/2}\sqrt{\sum_{s=1}^{\tau}\max_{a\in\mathcal{A}}\|a\|^2_{\mathbf{V}_\tau^{-1}}} \\
&\leq 2\sqrt{\beta_\tau(\delta)}\tau^{-1/2}\sqrt{\sum_{s=1}^{\tau}\max_{a\in\mathcal{A}}\|a\|^2_{\mathbf{V}_{s-1}^{-1}}} && (\mathbf{V}_\tau \succeq \mathbf{V}_{s-1}) \\
&\leq 2\sqrt{\beta_\tau(\delta)}\tau^{-1/2}\sqrt{\sum_{s=1}^{\tau}\|a_s\|^2_{\mathbf{V}_{s-1}^{-1}}} && \text{(Equation (24))} \\
&= 2\sqrt{\beta_\tau(\delta)}\tau^{-1/2}\sqrt{\kappa}\sqrt{\sum_{s=1}^{\tau}\|a_s/\sqrt{\kappa}\|^2_{\mathbf{V}_{s-1}^{-1}}} \\
&\leq 4\sqrt{\beta_\tau(\delta)}\tau^{-1/2}\sqrt{\kappa}\sqrt{d\log(1+\tau/d)} && \text{(Lemma 9)}
\end{aligned}$$

which yields the announced result since $T \geq \tau$. Notice the re-normalization of the action by κ so that we can apply the Elliptical Potential Lemma (Lemma 9) directly.

□

B.3 ONS and Warm-Up

The ONS-like approaches of Zhang et al. (2016); Jun et al. (2017) do not use a warm-up procedure and rely on a "crude" parameter set $\Theta_0 = \{\theta,\ \|\theta\| \leq S\}$. As discussed in the main paper (see Section 3.3) it is natural to wonder whether their mechanisms could be directly improved (by order of magnitude κ) by using the refined parameter Θ returned by the warm-up procedure. This is unfortunately not the case; both approaches hard-codes the κ-dependency in the size of their parameter updates. This dependency can only be *marginally* reduced when using Θ.

For instance Jun et al. (2017) rely the exp-concavity constant of the log-loss to design their update rule. Formally, for a parameter set Θ' it is defined as (see Hazan (2016, Definition 4.1)):

$$\rho(\Theta') := \sup_{r>0}\left\{r \text{ s.t } \nabla^2_\theta \ell(a^\mathsf{T}\theta, r) \succeq r\nabla_\theta\ell(a^\mathsf{T}\theta,r)\nabla_\theta\ell(a^\mathsf{T}\theta,r)^\mathsf{T},\ \forall\theta\in\Theta', \forall(a,r)\in\mathcal{A}\times\{0,1\}\right\} \ .$$

After some straight-forward manipulations it writes as:

$$\begin{aligned}\rho(\Theta') &= \sup_{r>0} \left\{ r \text{ s.t } r \leq \dot{\mu}(a^\mathsf{T}\theta)/(\mu(a^\mathsf{T}\theta) - r)^2,\ \forall \theta \in \Theta', \forall (a,r) \in \mathcal{A} \times \{0,1\} \right\} \\ &\leq 2 \min_{\theta \in \Theta'} \min_{a \in \mathcal{A}} \dot{\mu}(a^\mathsf{T}\theta) \ .\end{aligned}$$

The update rule designed by Jun et al. (2017) hard-codes a factor $\rho(\Theta_0)^{-1}$ in their update rule and therefore in the radius of the associated confidence regions. This induces exponentially inflated confidence sets as:

$$\rho(\Theta_0)^{-1} \geq \kappa/2 = \mathfrak{C} \exp(S) \ .$$

Refining this dependency by using a smaller Θ does not remove such exponential dependencies in problem-dependent constants (*e.g.* $\|\theta_\star\|$, S). Indeed if Θ is the set returned by $\texttt{WarmUp}(\tau)$ under the conditions of Proposition 5:

$$\rho(\Theta)^{-1} \geq \mathfrak{C} \exp(\|\theta_\star\|) \ .$$

A similar argument holds for the update mechanism Zhang et al. (2016), which rely on the strong-convexity constant of the log-loss.

C CONCENTRATION AND CONFIDENCE SETS

C.1 Refinement of Faury et al. (2020)

In the following, we consider that we have adaptively collected the dataset $\{a_t, r_{t+1}\}_t$. We denote:

$$\hat{\theta}_{t+1} := \arg\min_{\theta} \sum_{s=1}^{t} \ell_{s+1}(\theta) + \lambda_t(\delta)\|\theta\|^2/2 \,,$$

where $\lambda_t(\delta)$ is defined in Equation (10). Directly following the proof of Faury et al. (2020, Lemma 11):

$$\mathbb{P}\left(\forall t \geq 1,\ \left\|\theta_\star - \widetilde{\theta}_{t+1}\right\|^2_{\mathbf{H}_t(\theta_\star)} \leq 4(1+2S)^2\gamma_t(\delta)\right) \geq 1-\delta \,, \tag{25}$$

where $\widetilde{\theta}_{t+1}$ is obtained by "projecting" $\hat{\theta}_{t+1}$ on the ball $\{\|\theta\| \leq S\}$ through a *non-convex* minimization routine. The slowly growing function $\gamma_t(\delta)$ is obtained after applying simple upper-bounding operations to Faury et al. (2020, Theorem 1) and is formally defined in Equation (11). It checks:

$$\gamma_t(\delta) \leq \mathfrak{C} S^2 d \log(t/\delta) \,.$$

The following proposition establishes that Equation (25) still holds when $\tilde{\theta}_{t+1}$ is replaced by $\hat{\theta}_{t+1}$, at the price of only a minor degradation of the bound. This essentially removes the need to solve a non-convex program whenever $\|\hat{\theta}_{t+1}\| \geq S$. The function $\beta_t(\delta)$ is defined in Equation (12) and checks $\beta_t(\delta) \leq \mathfrak{C} S^6 d \log(t/\delta)$.

Lemma 1. *Let $\delta \in (0,1]$. Then:*

$$\mathbb{P}\Big(\forall t \geq 1,\ \left\|\theta_\star - \hat{\theta}_{t+1}\right\|^2_{\mathbf{H}_t(\theta_\star)} \leq \beta_t(\delta)\Big) \geq 1-\delta \,.$$

Remark 1. *Whenever $\|\hat{\theta}_{t+1}\| \leq S$ one can directly use the bound given in Equation (25), which is then valid for $\tilde{\theta}_{t+1} = \hat{\theta}_{t+1}$.*

Proof. The proof leverages the self-concordance property of the logistic function by using some intermediary results from Abeille et al. (2021). In the following, we denote for all θ:

$$g_t(\theta) := \sum_{s=1}^{t} \mu(a_s^\mathsf{T}\theta)a_s + \lambda\theta \quad \text{and} \quad \mathbf{G}_t(\theta) = \sum_{s=1}^{t} \alpha(a_s^\mathsf{T}\theta, a_s^\mathsf{T}\theta_\star)a_s a_s^\mathsf{T} \,,$$

where $\alpha(x,y)$ is defined in Appendix A.1. Further, define the event E_δ as follows:

$$E_\delta := \left\{\forall t \geq 1,\ \left\|g_t(\theta_\star) - g_t(\hat{\theta}_{t+1})\right\|^2_{\mathbf{H}_t(\theta_\star)^{-1}} \leq \gamma_t(\delta)\right\} \,.$$

By Lemma 1 of Faury et al. (2020) we have that $\mathbb{P}(E_\delta) \geq 1-\delta$. From the demonstration of Lemma 2 from Abeille et al. (2021) it can also be extracted that if E_δ holds then for any $t \geq 1$:

$$\begin{aligned}\mathbf{H}_t(\theta_\star) &\preceq \left(1 + \gamma_t(\delta)/\lambda_t(\delta) + \sqrt{\gamma_t(\delta)/\lambda_t(\delta)}\right)\mathbf{G}_t(\hat{\theta}_{t+1}) \\ &= \left(5/2 + (S+3/2)^2 + S\right)\mathbf{G}_t(\hat{\theta}_{t+1}) \,.\end{aligned} \tag{26}$$

Finally, recall that by the mean-value theorem we have the following identity for any θ:

$$g_t(\theta) - g_t(\theta_\star) = \mathbf{G}_t(\theta)(\theta - \theta_\star) \,. \tag{27}$$

We conclude by chaining inequalities, assuming that E_δ holds (which happens with probability at least $1-\delta$);

$$\begin{aligned}
\left\|\theta_\star - \hat{\theta}_{t+1}\right\|^2_{\mathbf{H}_t(\theta_\star)} &\leq \left(5/2 + (S+3/2)^2 + S\right)\left\|\theta_\star - \hat{\theta}_{t+1}\right\|^2_{\mathbf{G}_t(\hat{\theta}_{t+1})} && \text{(Equation (26))} \\
&= \left(5/2 + (S+3/2)^2 + S\right)\left\|g_t(\theta_\star) - g_t(\hat{\theta}_{t+1})\right\|^2_{\mathbf{G}_t(\hat{\theta}_{t+1})^{-1}} && \text{(Equation (27))} \\
&\leq \left(5/2 + (S+3/2)^2 + S\right)^2\left\|g_t(\theta_\star) - g_t(\hat{\theta}_{t+1})\right\|^2_{\mathbf{H}_t(\theta_\star)^{-1}} && \text{(Equation (26))} \\
&\leq \left(5/2 + (S+3/2)^2 + S\right)^2\gamma_t(\delta) = \beta_t(\delta) \,, && (E_\delta \text{ holds})
\end{aligned}$$

which proves the announced result. □

C.2 Statement of Theorem 3

The goal of this section is to justify the confidence sets used in the main paper through the statement of the more general Theorem 3 (see below). In particular, we deal here with the optimization errors introduced when running the ECOLog procedure.

Algorithm 4 Efficient Local Learning for Logistic Bandits (ECOLog, sequential form)

input: Compact convex sets $\{\Theta_t\}_t$, optimization accuracies $\{\varepsilon_t\}_t$.
  Let $\mathbf{W}_1 \leftarrow \mathbf{I}_d$, $\theta'_1 \in \Theta_1$. ▷ *initialization*
  Let $D \leftarrow \sup_{t\geq 1} \mathrm{diam}_{\mathcal{A}}(\Theta_t)$.
  for $t \geq 1$ **do**
    Receive the pair (a_t, r_{t+1}).
    Define θ_{t+1} as:

$$\theta_{t+1} = \operatorname*{arg\,min}_{\theta\in\Theta_t} \left(\frac{1}{2+D} \|\theta - \theta'_t\|^2_{\mathbf{W}_t} + \ell_{t+1}(\theta) \right) .$$

    Compute θ'_{t+1} by solving the above program to accuracy ε_t.
    Update $\mathbf{W}_{t+1} \leftarrow \mathbf{W}_t + \dot{\mu}(a_t^\mathsf{T}\theta'_{t+1}) a_t a_t^\mathsf{T}$.
  end for

We detail in Algorithm 4 the pseudo-code for ECOLog in its sequential form. It takes as input a sequence of compact convex sets $\{\Theta_t\}_t$ and a sequence $\{\varepsilon_t\}_t$ of optimization accuracy. Note the use of:

$$D := \sup_{t\geq 1} \mathrm{diam}_{\mathcal{A}}(\Theta_t)$$

If this quantity is unknown, D is replaced by an upper-bound on the supremum (the tighter, the better). Our use of ECOLog in both Algorithms 1 and 2 falls under this general description. For instance Algorithm 1 instantiates this procedure with $\Theta_t \equiv \Theta$ (the set returned by the warm-up) for which $D \leq 1$ (see Proposition 5).

We assume that at each round $t \geq 1$ the true minimizer:

$$\theta_{t+1} = \operatorname*{arg\,min}_{\theta\in\Theta_t} \left(\frac{1}{2+D} \|\theta - \theta_t\|^2_{\mathbf{W}_t} + \ell_{t+1}(\theta) \right) , \tag{28}$$

can computed up to accuracy ε_t. In other words, we have access to θ'_{t+1} such that:

$$\left\|\theta_{t+1} - \theta'_{t+1}\right\| \leq \varepsilon_t . \tag{29}$$

We discuss in Appendix E.1 how such θ'_{t+1} can be efficiently computed. We denote $\{(\theta'_{t+1}, \mathbf{W}_{t+1})\}_t$ the sequence of parameters maintained by ECOLog($\{\Theta_t\}_t$, $\{\varepsilon_t\}_t$) and claim the following concentration bound. The function $\nu_t(\delta)$ is defined in Equation (13). Numerical constants can be improved by a more careful analysis.

Theorem 3. *Let $\delta \in (0, 1]$ and assume that $\theta_\star \in \Theta_t$ for all $t \geq 1$. Then:*

$$\mathbb{P}\left(\forall t \geq 1,\ \left\|\theta'_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} \leq 8S^2 + 4\sum_{s=1}^{t} s\varepsilon_s^2 + 2D^2 + (2+D)^2\nu_t(\delta)/2 + 2(2+D)^2 \exp(D) d \log(1 + t/(4d)) \right) \geq 1 - \delta .$$

C.3 Proof of Theorem 3

An important technical piece of our analysis resides in the following Lemma, which derives a *local* quadratic lower-bound for the logistic loss $\ell_{t+1}(\theta)$. It is extracted from the self-concordance analysis of Abeille et al. (2021). A slightly stronger form, derived through other means, also appears in Jézéquel et al. (2020). The proof is deferred to Appendix C.3.1.

Proposition 6 (Local Quadratic Lower-Bound)**.** *For all $t \geq 1$ and any $\theta, \theta_r \in \Theta_t$:*

$$\ell_{t+1}(\theta) \geq \ell_{t+1}(\theta_r) + \nabla\ell_{t+1}(\theta_r)^\mathsf{T}(\theta - \theta_r) + \frac{\dot{\mu}(a_t^\mathsf{T}\theta_r)}{2 + \mathrm{diam}_{\mathcal{A}}(\Theta_t)} (a_t^\mathsf{T}(\theta - \theta_r))^2 .$$

Another important intermediary result is given by the following Lemma. It is obtained by directly leveraging the update rule. The proof is deferred to Appendix C.3.2.

Lemma 3. *At any round $t \geq 1$:*

$$\nabla \ell_{t+1}(\theta_{t+1})^\mathsf{T}(\theta_{t+1} - \theta_\star) \leq (1 + D/2)^{-1}(\theta_{t+1} - \theta'_t)^\mathsf{T}\mathbf{W}_t(\theta_\star - \theta_{t+1}) \,.$$

Combining Proposition 6 and Lemma 3 yields the following result, tying the deviation between θ'_{t+1} and $\theta_\star$ with the excess loss incurred by $\{\theta_{s+1}\}_{s=1}^t$. The proof is deferred to Appendix C.3.3.

Lemma 4. *For any $t \geq 1$ the following holds:*

$$\left\|\theta'_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} \leq 4S^2 + 4\sum_{s=1}^t s\varepsilon_s^2 + (4 + 2D)\left[\sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1})\right] \,.$$

To obtain a valid confidence set from Lemma 4 we are left to bound $\sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1})$. To do so, and inspired by the analysis of Jézéquel et al. (2020) in the online convex optimization setting, we introduce:

$$\bar{\theta}_t := \arg\min_{\Theta}\left(\frac{1}{2+D}\left\|\theta - \theta_t\right\|^2_{\mathbf{W}_t} + \ell(a_t^\mathsf{T}\theta, 0) + \ell(a_t^\mathsf{T}\theta, 1)\right) \,. \tag{30}$$

Note that $\bar{\theta}_t$ is $\mathcal{F}_t$-measurable (θ_{t+1} is $\mathcal{F}_{t+1}$ measurable). We rely on the following decomposition and bound each term of the r.h.s separately:

$$\sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1}) = \left[\sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s)\right] + \left[\sum_{s=1}^t \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1})\right] \,. \tag{31}$$

The first term is bounded with high probability as stated below. The proof is deferred to Appendix C.3.4 and uses a 1-dimensional version of a concentration result from Faury et al. (2020).

Lemma 5. *Let $\delta \in (0, 1]$. We have:*

$$\mathbb{P}\left(\forall t \geq 1,\ \sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s) \leq (2+D)\nu_t(\delta)/4 + D^2(2+D)^{-1}\right) \geq 1 - \delta' \,.$$

We now turn on bounding the second term in Equation (31) - that is:

$$\sum_{s=1}^t \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) \,.$$

We claim the following intermediary result, which proof is deferred to Appendix C.3.5.

Lemma 6. *The following result holds for any $t \geq 1$:*

$$\sum_{s=1}^t \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) \leq (1+D)\sum_{s=1}^t \dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} \,.$$

We finish the bound by the following result, a consequence of the self-concordance property of the logistic function. The proof is deferred to Appendix C.3.6.

Lemma 7. *The following result holds for any $t \geq 1$:*

$$\sum_{s=1}^t \dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} \leq \exp(D) d \log\left((1 + t/(4d)\right) \,.$$

Combining Lemmas 6 and 7 yields that:

$$\sum_{s=1}^{t} \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) \leq (1+D)\exp(D)d\log\left((1+t/(4d)\right) \ .$$

Assembling this result with Equation (31) yields that $\forall t \geq 1$:

$$\sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1}) \leq \left[\sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s)\right] + (1+D)\exp(D)d\log\left((1+t/(4d)\right) \ .$$

Thanks to Lemma 5 this further yields that with probability at least $1-\delta$:

$$\forall t \geq 1, \ \sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1}) \leq (2+D)\nu_t(\delta)/4 + D^2(2+D)^{-1} + (1+D)\exp(D)d\log\left((1+t/(4d)\right) \ .$$

Assembling this result with Lemma 4 along with simple bounding operations yield the announced result.

C.3.1 Proof of Proposition 6

Proposition 6 (Local Quadratic Lower-Bound)**.** *For all $t \geq 1$ and any $\theta, \theta_r \in \Theta_t$:*

$$\ell_{t+1}(\theta) \geq \ell_{t+1}(\theta_r) + \nabla\ell_{t+1}(\theta_r)^\mathsf{T}(\theta - \theta_r) + \frac{\dot{\mu}(a_t^\mathsf{T}\theta_r)}{2 + \mathrm{diam}_{\mathcal{A}}(\Theta_t)}(a_t^\mathsf{T}(\theta - \theta_r))^2 \ .$$

Proof. By a exact second-order Taylor of $\ell_{t+1}(\theta)$ decomposition around θ_r yields (see Equation (19)):

$$\ell_{t+1}(\theta) = \ell_{t+1}(\theta_r) + \nabla\ell_{t+1}(\theta_r)^\mathsf{T}(\theta - \theta_r) + \widetilde{\alpha}(a_t^\mathsf{T}\theta, a_t^\mathsf{T}\theta_r)(a_t^\mathsf{T}(\theta - \theta_r))^2 \ ,$$

Further by Equation (21) we have:

$$\begin{aligned}
\widetilde{\alpha}(a_t^\mathsf{T}\theta, a_t^\mathsf{T}\theta_r) &\geq \dot{\mu}(a_t^\mathsf{T}\theta_r)/(2 + |a_t^\mathsf{T}(\theta_r - \theta)|) \\
&\geq \dot{\mu}(a_t^\mathsf{T}\theta_r)/(2 + \mathrm{diam}_{\mathcal{A}}(\Theta_t)) \ , \qquad (\theta_r, \theta \in \Theta_t, \text{ def. of } \mathrm{diam}_{\mathcal{A}}(\Theta_t))
\end{aligned}$$

which concludes the proof. □

C.3.2 Proof of Lemma 3

Lemma 3. *At any round $t \geq 1$:*

$$\nabla\ell_{t+1}(\theta_{t+1})^\mathsf{T}(\theta_{t+1} - \theta_\star) \leq (1 + D/2)^{-1}(\theta_{t+1} - \theta'_t)^\mathsf{T}\mathbf{W}_t(\theta_\star - \theta_{t+1}) \ .$$

Proof. Denote $\widetilde{L}_{t+1}(\theta) := (2+D)^{-1}\left\|\theta - \theta'_t\right\|^2_{\mathbf{W}_t} + \ell_{t+1}(\theta)$ the function minimized by θ_{t+1} over Θ_t. Since Θ_t is a convex set and $\widetilde{L}_{t+1}$ a convex function, we have that for any $\theta \in \Theta_t$ (see Lemma 11);

$$\begin{aligned}
0 &\leq \nabla\widetilde{L}_{t+1}(\theta_{t+1})^\mathsf{T}(\theta - \theta_{t+1}) \\
&= \left((1+D/2)^{-1}\mathbf{W}_t(\theta_{t+1} - \theta'_t) + \nabla\ell_{t+1}(\theta_{t+1})\right)^\mathsf{T}(\theta - \theta_{t+1}) \\
&= (1+D/2)^{-1}(\theta_{t+1} - \theta'_t)^\mathsf{T}\mathbf{W}_t(\theta - \theta_{t+1}) + \nabla\ell_{t+1}(\theta_{t+1})^\mathsf{T}(\theta - \theta_{t+1})
\end{aligned}$$

Taking $\theta = \theta_\star \in \Theta_t$ (by assumption) in the above inequality yields the announced result. □

C.3.3 Proof of Lemma 4

Lemma 4. *For any $t \geq 1$ the following holds:*

$$\left\|\theta'_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} \leq 4S^2 + 4\sum_{s=1}^{t} s\varepsilon_s^2 + (4+2D)\left[\sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1})\right] \ .$$

Proof. By Proposition 6, and because $\theta_{t+1}, \theta_\star \in \Theta_t$ (by construction for θ_{t+1} and by assumption for $\theta_\star$) the following holds for any $s \geq 1$:

$$\begin{aligned}
\ell_{s+1}(\theta_\star) &\geq \ell_{s+1}(\theta_{s+1}) + \nabla\ell_{s+1}(\theta_{s+1})^\mathsf{T}(\theta_\star - \theta_{s+1}) + \frac{\mu(a_s^\mathsf{T}\theta_{s+1})}{2 + \mathrm{diam}_{\mathcal{A}}(\Theta_t)}(a_s^\mathsf{T}(\theta_\star - \theta_{s+1}))^2 \\
&\geq \ell_{s+1}(\theta_{s+1}) + \nabla\ell_{s+1}(\theta_{s+1})^\mathsf{T}(\theta_\star - \theta_{s+1}) + \frac{\mu(a_s^\mathsf{T}\theta_{s+1})}{2 + D}(a_s^\mathsf{T}(\theta_\star - \theta_{s+1}))^2 \,. \qquad (D \geq \mathrm{diam}_{\mathcal{A}}(\Theta))
\end{aligned}$$

After re-arranging this yields:

$$\ell_{s+1}(\theta_{s+1}) - \ell_{s+1}(\theta_\star) \leq \nabla\ell_{s+1}(\theta_{s+1})^\mathsf{T}(\theta_{s+1} - \theta_\star) - (2+D)^{-1}\dot\mu(a_s^\mathsf{T}\theta_{s+1})(a_s^\mathsf{T}(\theta_{s+1} - \theta_\star))^2 \,.$$

Using Lemma 3 in the above inequality gives:

$$\begin{aligned}
(1 + D/2)\left(\ell_{s+1}(\theta_{s+1}) - \ell_{s+1}(\theta_\star)\right) &\leq (\theta_{s+1} - \theta'_s)^\mathsf{T}\mathbf{W}_s(\theta_\star - \theta_{s+1}) - \frac{1}{2}\dot\mu(a_s^\mathsf{T}\theta_{s+1})(a_s^\mathsf{T}(\theta_{s+1} - \theta_\star))^2 \,, \\
&= -\frac{1}{2}\left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_s} + \frac{1}{2}\left\|\theta'_s - \theta_\star\right\|^2_{\mathbf{W}_s} - \frac{1}{2}\left\|\theta_{s+1} - \theta'_s\right\|^2_{\mathbf{W}_s} \\
&\quad - \frac{1}{2}\dot\mu(a_s^\mathsf{T}\theta_{s+1})(a_s^\mathsf{T}(\theta_{s+1} - \theta_\star))^2 \,, \\
&= -\frac{1}{2}\left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_{s+1}} + \frac{1}{2}\left\|\theta'_s - \theta_\star\right\|^2_{\mathbf{W}_s} - \frac{1}{2}\left\|\theta_{s+1} - \theta'_s\right\|^2_{\mathbf{W}_s} \\
&\leq -\frac{1}{2}\left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_{s+1}} + \frac{1}{2}\left\|\theta'_s - \theta_\star\right\|^2_{\mathbf{W}_s} \\
&\leq -\frac{1}{2}\left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_{s+1}} + \frac{1}{2}\left\|\theta_s - \theta_\star\right\|^2_{\mathbf{W}_s} + \frac{1}{2}\left\|\theta_s - \theta'_s\right\|^2_{\mathbf{W}_s} \\
&\leq -\frac{1}{2}\left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_{s+1}} + \frac{1}{2}\left\|\theta'_s - \theta_\star\right\|^2_{\mathbf{W}_s} + s\varepsilon_{s-1}^2
\end{aligned}$$

since $\left\|\theta'_s - \theta_\star\right\|^2_{\mathbf{W}_s} \leq \lambda_{\max}(\mathbf{W}_s)\left\|\theta'_s - \theta_s\right\|^2 \leq 2s\varepsilon_{s-1}^2$. By re-arranging:

$$(2+D)\left(\ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1})\right) - \left\|\theta_{s+1} - \theta_s\right\|^2_{\mathbf{W}_s} \geq \left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_{s+1}} - \left\|\theta_s - \theta_\star\right\|^2_{\mathbf{W}_s} - 2s\varepsilon_{s-1}^2$$

and summing from $s = 1$ to t:

$$\begin{aligned}
(2+D)\sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1}) &\geq \sum_{s=1}^t \left[\left\|\theta_{s+1} - \theta_\star\right\|^2_{\mathbf{W}_{s+1}} - \left\|\theta_s - \theta_\star\right\|^2_{\mathbf{W}_s}\right] - 2\sum_{s=1}^t s\varepsilon_{s-1}^2 \\
&= \left\|\theta_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} - \left\|\theta_1 - \theta_\star\right\|^2_{\mathbf{W}_1} - 2\sum_{s=1}^t s\varepsilon_{s-1}^2 \qquad \text{(telescopic sum)} \\
&= \left\|\theta_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} - \left\|\theta_1 - \theta_\star\right\|^2 - 2\sum_{s=1}^t s\varepsilon_{s-1}^2 \qquad (\mathbf{W}_1 = \mathbf{I}_d)
\end{aligned}$$

After re-arranging and setting $\varepsilon_0 = 0$ (there is no program to solve at $t = 0$);

$$\left\|\theta_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} \leq 4S^2 + 2\sum_{s=1}^{t-1} s\varepsilon_s^2 + (2+D)\left[\sum_{s=1}^t \ell_{s+1}(\theta_\star) - \ell_{s+1}(\theta_{s+1})\right] \,.$$

This concludes the proof as:

$$\begin{aligned}
\left\|\theta'_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} &\leq 2\left\|\theta_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} + 2\left\|\theta'_{t+1} - \theta_{t+1}\right\|^2_{\mathbf{W}_{t+1}} && ((a+b)^2 \leq 2(a^2+b^2)) \\
&\leq 2\left\|\theta_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} + 2(1+t)\left\|\theta'_{t+1} - \theta_{t+1}\right\|^2 && (\mathbf{W}_{t+1} \preceq (1+t)\mathbf{I}_d) \\
&\leq 2\left\|\theta'_{t+1} - \theta_\star\right\|^2_{\mathbf{W}_{t+1}} + 4t\varepsilon_t^2 \,.
\end{aligned}$$

□

C.3.4 Proof of Lemma 5

Lemma 5. *Let $\delta \in (0,1]$. We have:*

$$\mathbb{P}\left(\forall t \geq 1,\ \sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s) \leq (2+D)\nu_t(\delta)/4 + D^2(2+D)^{-1}\right) \geq 1-\delta' \ .$$

Proof. Using Proposition 6 with $\theta = \theta_\star$ and $\theta_r = \bar{\theta}_s$ yields:

$$\begin{aligned}
\sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s) &\leq \sum_{s=1}^{t} \nabla \ell_{s+1}(\theta_\star)^\mathsf{T}(\theta_\star - \bar{\theta}_s) - (2+D)^{-1}\sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\theta_\star)(a_s^\mathsf{T}(\theta_\star - \bar{\theta}_s))^2 \\
&= \sum_{s=1}^{t} (\mu(a_s^\mathsf{T}\theta_\star) - r_{s+1})a_s^\mathsf{T}(\theta_\star - \bar{\theta}_s) - (2+D)^{-1}\sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\theta_\star)(a_s^\mathsf{T}(\theta_\star - \bar{\theta}_s))^2 \\
&= D\sum_{s=1}^{t} \eta_{s+1}x_s - D^2(2+D)^{-1}X_t \ , \qquad (32)
\end{aligned}$$

where we denoted $x_s := a_s^\mathsf{T}(\theta_\star - \bar{\theta}_s)/D$, $X_t := \sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\theta_\star)x_s^2$ and $\eta_{s+1} := \mu(a_s^\mathsf{T}\theta_\star) - r_{s+1}$. We use a 1-dimensional version of the concentration result provided by Theorem 1 of Faury et al. (2020) to bound $\sum_{s=1}^{t} \eta_{s+1}x_s$. We remind its general form below for the sake of completeness.

Theorem 4 (Theorem 1 of Faury et al. (2020))**.** *Let $\{\mathcal{F}_t\}_{t=1}^{\infty}$ be a filtration. Let $\{x_t\}_{t=1}^{\infty}$ be a stochastic process in $\mathcal{B}_2(d)$ such that x_t is $\mathcal{F}_t$-measurable. Let $\{\eta_t\}_{t=2}^{\infty}$ be a martingale difference sequence such that η_{t+1} is $\mathcal{F}_{t+1}$ measurable. Furthermore, assume that conditionally on $\mathcal{F}_t$ we have $|\eta_{t+1}| \leq 1$ almost surely, and note $\sigma_t^2 := \mathbb{E}\left[\eta_{t+1}^2|\mathcal{F}_t\right]$. Let $\lambda > 0$ and for any $t \geq 1$ define:*

$$\mathbf{H}_t := \sum_{s=1}^{t} \sigma_s^2 x_s x_s^T + \lambda \mathbf{I}_d, \qquad S_{t+1} := \sum_{s=1}^{t} \eta_{s+1}x_s.$$

Then for any $\delta \in (0,1]$:

$$\mathbb{P}\left(\exists t \geq 1,\ \|S_{t+1}\|_{\mathbf{H}_t^{-1}} \geq \frac{\sqrt{\lambda}}{2} + \frac{2}{\sqrt{\lambda}}\log\left(\frac{\det(\mathbf{H_t})^{\frac{1}{2}}\lambda^{-\frac{d}{2}}}{\delta}\right) + \frac{2}{\sqrt{\lambda}}d\log(2)\right) \leq \delta.$$

Recall that we use the filtration $\mathcal{F}_t := \sigma(a_1, r_2, \ldots, a_t)$. In our case, x_s is 1-dimensional, is $\mathcal{F}_s$-measurable and satisfies $|x_s| \leq 1$ almost surely (by definition of D, and since both $\theta_\star, \bar{\theta}_s \in \Theta_t$). Further, η_{s+1} is $\mathcal{F}_{s+1}$-measurable and thanks to Equation (1) we have:

$$\mathbb{E}\left[\eta_{s+1}|\mathcal{F}_s\right] = 0 \quad \text{ and } \mathbb{E}\left[\eta_{s+1}^2|\mathcal{F}_s\right] = \dot{\mu}(a_s^\mathsf{T}\theta_\star) \ .$$

Furthermore, note that with the notations of Theorem 4 we have $\mathbf{H}_t = X_t + 1$. By a direct application of Theorem 4 we obtain that with probability at least $1-\delta$:

$$\begin{aligned}
\forall t \geq 1, \quad \sum_{s=1}^{t} \eta_{s+1}x_s &\leq \sqrt{X_t+1}\sqrt{1/2 + 2\log\left(\frac{2\sqrt{X_t+1}}{\delta}\right)} \\
&= \sqrt{X_t+\lambda}\sqrt{1/2 + 2\log\left(\frac{2\sqrt{\sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\theta_\star)z_s^2 + 1}}{\delta}\right)} \\
&\leq \sqrt{X_t+1}\sqrt{1/2 + 2\log\left(\frac{2\sqrt{t/4+1}}{\delta}\right)} && (\dot{\mu} \leq 1/4, |z_s| \leq 1) \\
&= \sqrt{\nu_t(\delta)}\sqrt{X_t+1} && (\text{def. of } \nu_t(\delta)) \\
&\leq \frac{\nu_t(\delta)}{4D(2+D)^{-1}} + D(2+D)^{-1}(X_t+1)
\end{aligned}$$

where in the second to last inequality we used the fact that $\forall a, b, \zeta > 0$ we have $\sqrt{ab} \leq a/(2\zeta) + \zeta b/2$ (applied with $a = \gamma_t(\delta)$, $b = X_t + \lambda$ and $\zeta = 2D(2+D)^{-1}$. Re-injecting in Equation (32) yields that with probability at least $1-\delta$:

$$\forall t \geq 1, \quad \sum_{s=1}^{t} \ell_{s+1}(\theta_\star) - \ell_{s+1}(\bar{\theta}_s) \leq (2+D)\nu_t(\delta)/4 + D^2(2+D)^{-1} \, .$$

□

C.3.5 Proof of Lemma 6

Lemma 6. *The following result holds for any $t \geq 1$:*

$$\sum_{s=1}^{t} \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) \leq (1+D) \sum_{s=1}^{t} \dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s) \, \|a_s\|^2_{\mathbf{W}_{s+1}^{-1}} \, .$$

Proof. By convexity of $\ell_{s+1}(\cdot)$ one has that for all $s \geq 1$:

$$\begin{aligned} \ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) &\leq \nabla \ell_{s+1}(\bar{\theta}_s)^\mathsf{T}(\bar{\theta}_s - \theta_{s+1}) \\ &\leq \left\|\nabla \ell_{s+1}(\bar{\theta}_s)\right\|_{\mathbf{W}_{s+1}^{-1}} \left\|\bar{\theta}_s - \theta_{s+1}\right\|_{\mathbf{W}_{s+1}} \qquad \text{(Cauchy-Schwarz)} \end{aligned} \tag{33}$$

Since $\ell(a_s^\mathsf{T}\theta, 0) + \ell(a_s^\mathsf{T}\theta, 1) = \ell_{s+1}(\theta) + \bar{\ell}_{s+1}(\theta)$, one can re-write the computation of $\bar{\theta}_s$ as:

$$\bar{\theta}_s = \operatorname*{arg\,min}_{\theta \in \Theta_s} \frac{1}{2+D} \|\theta - \theta_s\|^2_{\mathbf{W}_t} + \ell_{s+1}(\theta) + \bar{\ell}_{s+1}(\theta) \, .$$

By convexity of the objective function minimized by $\bar{\theta}_s$ and convexity of Θ_s we therefore have the following inequality (see Lemma 11) for any $s \geq 1$:

$$(1+D/2)^{-1}(\bar{\theta}_s - \theta_s)^\mathsf{T}\mathbf{W}_s(\theta_{s+1} - \bar{\theta}_s) + \nabla \ell_{s+1}(\bar{\theta}_s)^\mathsf{T}(\theta_{s+1} - \bar{\theta}_s) + \nabla \bar{\ell}_{s+1}(\bar{\theta}_s)^\mathsf{T}(\theta_{s+1} - \bar{\theta}_s) \geq 0 \, .$$

since $\theta_{s+1} \in \Theta_s$ by definition. By re-arranging this yields:

$$\begin{aligned} (1+D/2)\nabla \bar{\ell}_{s+1}(\bar{\theta}_s)^\mathsf{T}(\theta_{s+1} - \bar{\theta}_s) &\geq (\bar{\theta}_s - \theta_s)^\mathsf{T}\mathbf{W}_s(\bar{\theta}_s - \theta_{s+1}) + (1+D/2)\nabla \ell_{s+1}(\bar{\theta}_s)^\mathsf{T}(\bar{\theta}_s - \theta_{s+1}) \\ &= \left\|\bar{\theta}_s - \theta_{s+1}\right\|^2_{\mathbf{W}_s} + (\theta_{s+1} - \theta_s)^\mathsf{T}\mathbf{W}_s(\bar{\theta}_s - \theta_{s+1}) \\ &\quad + (1+D/2)\nabla \ell_{s+1}(\bar{\theta}_s)^\mathsf{T}(\bar{\theta}_s - \theta_{s+1}) \, . \end{aligned}$$

By the same argument, since $\bar{\theta}_s \in \Theta_s$ we also have the inequality:

$$(\theta_{s+1} - \theta_s)^\mathsf{T}\mathbf{W}_s(\bar{\theta}_s - \theta_{s+1}) \geq (1+D/2)\nabla \ell_{s+1}(\theta_{s+1})^\mathsf{T}(\theta_{s+1} - \bar{\theta}_s) \, .$$

Re-injecting above this yields that:

$$\begin{aligned} (1+D/2)\nabla \bar{\ell}_{s+1}(\bar{\theta}_s)^\mathsf{T}(\theta_{s+1} - \bar{\theta}_s) &\geq \left\|\bar{\theta}_s - \theta_{s+1}\right\|^2_{\mathbf{W}_s} + (1+D/2)(\bar{\theta}_s - \theta_{s+1})^\mathsf{T}(\nabla \ell_{s+1}(\bar{\theta}_s) - \nabla \ell_{s+1}(\theta_{s+1})) \\ &= \left\|\bar{\theta}_s - \theta_{s+1}\right\|^2_{\mathbf{W}_s} + (1+D/2)(\mu(a_s^\mathsf{T}\bar{\theta}_s) - \mu(a_s^\mathsf{T}\theta_{s+1}))a_s^\mathsf{T}(\bar{\theta}_s - \theta_{s+1}) \\ &= \left\|\bar{\theta}_s - \theta_{s+1}\right\|^2_{\mathbf{W}_s} + (1+D/2)\alpha(a_s^\mathsf{T}\bar{\theta}_s, a_s^\mathsf{T}\theta_{s+1})(a_s^\mathsf{T}(\bar{\theta}_s - \theta_{s+1}))^2 \\ &\geq \left\|\bar{\theta}_s - \theta_{s+1}\right\|^2_{\mathbf{W}_s} + (1+D/2)(1+D)^{-1}\dot{\mu}(a_s^\mathsf{T}\theta_{s+1})(a_s^\mathsf{T}(\bar{\theta}_s - \theta_{s+1}))^2 \end{aligned}$$

where in the second to last inequality we used Equation (20) to obtain $\alpha(a_s^\mathsf{T}\bar{\theta}_s, a_s^\mathsf{T}\theta_{s+1}) \geq (1+D)^{-1}\dot{\mu}(a_s^\mathsf{T}\theta_{s+1})$ (since $\theta_{s+1}, \bar{\theta}_s \in \Theta_s$). After easy manipulations this yields:

$$\begin{aligned} \left\|\bar{\theta}_s - \theta_{s+1}\right\|^2_{\mathbf{W}_{s+1}} &\leq (1+D)\nabla \bar{\ell}_{s+1}(\bar{\theta}_s)^\mathsf{T}(\theta_{s+1} - \bar{\theta}_s) \\ &\leq (1+D)\left\|\nabla \bar{\ell}_{s+1}(\bar{\theta}_s)\right\|_{\mathbf{W}_{s+1}^{-1}} \left\|\bar{\theta}_s - \theta_{s+1}\right\|_{\mathbf{W}_{s+1}} \end{aligned}$$

and therefore we obtain that $\left\|\bar{\theta}_s - \theta_{s+1}\right\|_{\widetilde{\mathbf{V}}_{s+1}} \leq (1+D)\left\|\nabla\bar{\ell}_{s+1}(\bar{\theta}_s)\right\|_{\mathbf{W}_{s+1}^{-1}}$. Assembling with Equation (33);

$$\begin{aligned}
\ell_{s+1}(\bar{\theta}_s) - \ell_{s+1}(\theta_{s+1}) &\leq (1+D)\left\|\nabla\ell_{s+1}(\bar{\theta}_s)\right\|_{\mathbf{W}_{s+1}^{-1}}\left\|\nabla\bar{\ell}_{s+1}(\bar{\theta}_s)\right\|_{\mathbf{W}_{s+1}^{-1}} \\
&= (1+D)|\mu(a_s^\mathsf{T}\bar{\theta}_s) - r_{s+1}||\mu(a_s^\mathsf{T}\bar{\theta}_s) - 1 + r_{s+1}|\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} \\
&= (1+D)|\mu(a_s^\mathsf{T}\bar{\theta}_s)||\mu(a_s^\mathsf{T}\bar{\theta}_s) - 1|\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} && (r_{s+1}\in\{0,1\}) \\
&= (1+D)\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} && (\mu(1-\mu)=\dot{\mu})
\end{aligned}$$

Summing yields the announced result. □

C.3.6 Proof of Lemma 7

Lemma 7. *The following result holds for any $t \geq 1$:*

$$\sum_{s=1}^{t}\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} \leq \exp(D)d\log\left((1+t/(4d)\right) .$$

Proof. By Equation (22), for all $s \geq 1$:

$$\begin{aligned}
\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s) &\leq \exp\left(|a_s^\mathsf{T}(\theta'_{s+1} - \bar{\theta}_s)|\right)\dot{\mu}(a_s^\mathsf{T}\theta_{s+1}) \\
&\leq \exp(D)\dot{\mu}(a_s^\mathsf{T}\theta'_{s+1}) . && (\theta_{s+1},\bar{\theta}_s\in\Theta_s,\ D\geq\mathrm{diam}_{\mathcal{A}}(\Theta_s))
\end{aligned}$$

Denoting $x_s = \sqrt{\mu(a_s^\mathsf{T}\theta'_{s+1})}a_s$ and $\mathbf{M}_{t+1} = \sum_{s=1}^t x_s x_s^\mathsf{T}$, we have:

$$\begin{aligned}
\sum_{s=1}^{t}\dot{\mu}(a_s^\mathsf{T}\bar{\theta}_s)\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} &\leq \exp(D)\dot{\sum_{s=1}^{t}}\mu(a_s^\mathsf{T}\theta'_{s+1})\left\|a_s\right\|^2_{\mathbf{W}_{s+1}^{-1}} \\
&\leq \exp(D)\sum_{s=1}^{t}\left\|x_s\right\|^2_{\mathbf{M}_{s+1}^{-1}} \\
&= \exp(D)\sum_{s=1}^{t}\mathrm{Tr}(\mathbf{M}_{s+1}^{-1}x_s x_s^\mathsf{T}) \\
&= \exp(D)\sum_{s=1}^{t}\mathrm{Tr}(\mathbf{M}_{s+1}^{-1}(\mathbf{M}_{s+1} - \mathbf{M}_s)) \\
&\leq \exp(D)\log\left(|\mathbf{M}_{t+1}|\,/\,|\mathbf{M}_1|\right) && \text{(Lemma 4.6 of Hazan (2016))} \\
&\leq \exp(D)d\log(1+t/(4d)) ,
\end{aligned}$$

where we last used Lemma 10 along with $\left\|x_s\right\|^2 \leq \dot{\mu}(a_s^\mathsf{T}\theta'_{s+1}) \leq 1/4$. □

C.4 Proof of Proposition 3

We prove below Proposition 3 from the main paper. It justifies the confidence sets used in `OFU-ECOLog`.

In this context, we have $\Theta_t \equiv \Theta$, the set returned by the warm-up procedure run with the conditions of Proposition 5 and $\varepsilon_s = 1/s$.

Proposition 3 (Confidence Set)**.** *Let $\delta \in (0,1]$ and $\{(\theta_t, \mathbf{W}_t)\}_t$ the parameters maintained by Algorithm 1 with τ set according to Proposition 5. Then:*

$$\mathbb{P}\left(\forall t\geq 1,\ \left\|\theta_\star - \theta'_{t+1}\right\|^2_{\mathbf{W}_{t+1}} \leq \sigma_t(\delta)\ \text{and}\ \theta_\star\in\Theta\right) \geq 1-2\delta .$$

The function $\sigma_t(\delta)$ is defined in Equation (14) and checks $\sigma_t(\delta) \leq \mathfrak{C}S^2 d\log(t/\delta)$.

Proof. By Proposition 5 we know that $\mathrm{diam}_{\mathcal{A}}(\Theta) \leq 1$ so we can set $D = 1$. For the rest of the proof we assume that the event $\{\theta_\star \in \Theta\}$ holds - this happens with probability at least $1-\delta$ according to Proposition 5. Theorem 3 therefore applies since Θ is convex and compact. This yields:

$$\mathbb{P}\left(\forall t \geq 1,\ \|\theta_\star - \theta_{t+1}\|^2_{\mathbf{W}_t} \leq 8S^2 + 4\sum_{s=1}^{t} s\varepsilon_s^2 + 2 + 9\nu_t(\delta) + 18\exp(1)d\log(1+t/(4d))\right) \geq 1-\delta\ .$$

After a classic bound on the harmonic function; for $t \geq 1$:

$$\sum_{s=1}^{t} s\varepsilon_s^2 = \sum_{s=1}^{t} 1/s \leq 1 + \log(t)\ ,$$

we are left to apply a naive union bound with the event $\{\theta_\star \in \Theta\}$ to finish the proof. □

C.5 A Data-Dependent Version

The following result justifies the confidence regions used in `ada-OFU-ECOLog`.

Proposition 7. *Let $\delta \in (0,1]$ and $\{(\theta_t, \mathbf{W}_t, \Theta_t)\}_t$ maintained by Algorithm 2. Then:*

$$\mathbb{P}\left(\forall t \geq 1\ , \theta_\star \in \Theta_t \text{ and } \left\|\theta_\star - \theta'_{t+1}\right\|^2_{\mathbf{W}_{t+1}} \leq \eta_t(\delta)\right) \geq 1 - 2\delta\ .$$

The function $\eta_t(\delta)$ is defined in Equation (15) and checks $\eta_t(\delta) \leq \mathfrak{C}S^2 d\log(t/\delta)$.

Proof. This result can be easily be retrieved from the proof of Theorem 3. The sets:

$$\Theta_{t+1} = \left\{\theta, \left\|\theta - \hat{\theta}^{\mathcal{H}}_{t+1}\right\|^2_{\mathbf{V}^{\mathcal{H}}_t} \leq \beta_t(\delta)\right\}\ ,$$

maintained in Algorithm 2 are indeed compact and convex. Further, they contain $\theta_\star$ with high probability:

$$\begin{aligned}
1-\delta &\leq \mathbb{P}\left(\forall t \geq 1,\ \left\|\theta_\star - \hat{\theta}^{\mathcal{H}}_{t+1}\right\|^2_{\mathbf{H}^{\mathcal{H}}_t(\theta_\star)} \leq \beta_t(\delta)\right) && \text{(Lemma 1)}\\
&\leq \mathbb{P}\left(\forall t \geq 1,\ \left\|\theta_\star - \hat{\theta}^{\mathcal{H}}_{t+1}\right\|^2_{\mathbf{V}^{\mathcal{H}}_t} \leq \beta_t(\delta)\right) && \text{(Equation (9))}\\
&= \mathbb{P}\left(\forall t \geq 1, \theta_\star \in \Theta_t\right)\ .
\end{aligned}$$

Further, recall that the inequality:

$$\dot{\mu}(a_s^\intercal \bar{\theta}_s) \leq 2\dot{\mu}(a_s^\intercal \theta_{s+1})\ ,$$

holds *by construction* in Algorithm 2. When it is not satisfied, the couple (a_s, r_{s+1}) is not fed to the `ECOLog` procedure. This essentially allows to replace $\exp(D)$ in Lemma 7 by a constant factor (here, 2). From there, following the demonstration of Theorem 3 up to straight-forward adaptations (*e.g* to deal with the fact that some rounds are ignored from the learning when the above inequality is not satisfied) yields that under the event $\{\forall t \geq 1, \theta_\star \in \Theta_t\}$:

$$\mathbb{P}\left(\forall t \geq 1,\ \left\|\theta_\star - \theta'_{t+1}\right\|^2_{\mathbf{W}_{t+1}} \leq \eta_t(\delta)\right) \geq 1-\delta\ .$$

A union bound finishes the proof. □

D REGRET BOUNDS

To reduce clutter and fit with the notations adopted in the main text, we go back in this section to identifying θ_t and its ε-approximation θ'_t. This does not impact the validity of the regret bounds - the effects of optimization errors are fully dealt with in the radius of the confidence sets we designed in Appendix C.2.

D.1 Proof of Theorem 1

Theorem 1 (Regret Bound). *Let $\delta \in (0,1]$. Setting $\tau = \mathfrak{C}\kappa S^6 d^2 \log(T/\delta)^2$ ensures the regret of* `OFU-ECOLog`(δ, τ) *satisfies with probability at least $1-2\delta$:*

$$\mathrm{Regret}(T) \leq \mathfrak{C} S d \log(T/\delta)\sqrt{T\dot{\mu}(a_\star^\mathsf{T}\theta_\star)} + \mathfrak{C} S^6 \kappa d^2 \log(T/\delta)^2 \ .$$

Algorithm 1 `OFU-ECOLog`

input: failure level δ, warm-up length τ.
  Set $\Theta \leftarrow$ `WarmUp`(τ) (see Procedure 1). ▷ *forced-exploration*
  Initialize $\theta_{\tau+1} \in \Theta$, $\mathbf{W}_{\tau+1} \leftarrow \mathbf{I}_d$ and $\mathcal{C}_{\tau+1}(\delta) \leftarrow \Theta$.
  for $t \geq \tau+1$ **do**
    Play $a_t \in \arg\max_{a\in\mathcal{A}} \max_{\theta\in\mathcal{C}_t(\delta)} a^\mathsf{T}\theta$. ▷ *planning*
    Observe reward r_{t+1}, construct loss $\ell_{t+1}(\theta) = \ell(a_t^\mathsf{T}\theta, r_{t+1})$.
    Compute $(\theta_{t+1}, \mathbf{W}_{t+1}) \leftarrow$ `ECOLog`$(1/t, \Theta, \ell_{t+1}, \mathbf{W}_t, \theta_t)$ (see Procedure 2). ▷ *learning*
    Compute $\mathcal{C}_{t+1}(\delta) \leftarrow \left\{\|\theta - \theta_{t+1}\|^2_{\mathbf{W}_{t+1}} \leq \sigma_t(\delta)\right\}$. ▷ $\sigma_t(\delta)$ *is defined in Equation (14)*
  end for

Proof. According to Proposition 3 (its detailed version in Appendix C.4) setting $\tau = \mathfrak{C}\kappa S^6 d^2 \log(T/\delta)$ ensures:

$$\mathbb{P}\left(\theta_\star \in \Theta \text{ and } \ \|\theta_t - \theta_\star\|^2_{\mathbf{W}_t} \leq \sigma_t(\delta) \text{ for all } t \geq \tau+1\right) \geq 1-2\delta \ ,$$

In the rest of the proof we assume that the above event, denoted E_δ, holds.

Since $\mu(\cdot) \in (0,1)$ the regret incurred during warm-up can be directly bounded by τ. Therefore for $T \geq \tau+1$;

$$\begin{aligned}\mathrm{Regret}(T) &\leq \mathfrak{C}\kappa S^6 d^2 \log(T/\delta) + \sum_{t=\tau+1}^{T} \mu(a_\star^\mathsf{T}\theta_\star) - \mu(a_t^\mathsf{T}\theta_\star) \\ &\leq \mathfrak{C}\kappa S^6 d^2 \log(T/\delta) + R(T) \ ,\end{aligned}$$

where we defined $R(T) = \sum_{t=\tau+1}^{T} \mu(a_\star^\mathsf{T}\theta_\star) - \mu(a_t^\mathsf{T}\theta_\star)$. To control this term we follow the usual strategy for bounding the regret of optimistic algorithms, and re-use tools introduced by Faury et al. (2020); Abeille et al. (2021) - adapted to our confidence set. In the following, we denote for $t \geq \tau+1$:

$$(a_t, \tilde{\theta}_t) \in \underset{\mathcal{A}, \mathcal{C}_t(\delta)}{\arg\max}\, a^\mathsf{T}\theta \ .$$

where $\mathcal{C}_t(\delta) = \{\theta, \ \|\theta_t - \theta\|^2_{\mathbf{W}_t} \leq \sigma_t(\delta)\}$. Because E_δ holds this implies that the couple $(a_t, \tilde{\theta}_t)$ is optimistic. Formally: $a_t^\mathsf{T}\tilde{\theta}_t \geq a_\star^\mathsf{T}\theta_\star$. We start by tying the regret to the prediction error of $\tilde{\theta}_{t+1}$ and continue with a second-order Taylor expansion.

$$\begin{aligned}R(T) &= \sum_{t=\tau+1}^{T} \mu(a_\star^\mathsf{T}\theta_\star) - \mu(a_t^\mathsf{T}\theta_\star) \\ &\leq \sum_{t=\tau+1}^{T} \mu(a_t^\mathsf{T}\tilde{\theta}_t) - \mu(a_t^\mathsf{T}\theta_\star) && \text{(optimism, } \mu \nearrow\text{)} \\ &\leq \sum_{t=\tau+1}^{T} \dot{\mu}(a_t^\mathsf{T}\theta_\star) a_t^\mathsf{T}(\tilde{\theta}_t - \theta_\star) + \tilde{\alpha}(a_t^\mathsf{T}\theta_\star, a_t^\mathsf{T}\tilde{\theta}_t)(a_t^\mathsf{T}(\tilde{\theta}_t - \theta_\star))^2 && \text{(Taylor, } |\ddot{\mu}| \leq \dot{\mu}\text{)} \\ &=: R_1(T) + R_2(T) \ .\end{aligned}$$

Above, we defined $R_1(T) = \sum_{t=\tau+1}^{T} \dot{\mu}(a_t^\intercal \theta_\star) a_t^\intercal (\tilde{\theta}_t - \theta_\star)$ and $R_2(T) = \sum_{t=\tau+1}^{T} \tilde{\alpha}(a_t^\intercal \theta_\star, a_t^\intercal \tilde{\theta}_t)(a_t^\intercal (\tilde{\theta}_t - \theta_\star))^2$. We start by bounding $R_2(T)$;

$$
\begin{aligned}
R_2(T) &\leq \sum_{t=\tau+1}^{T} (a_t^\intercal (\tilde{\theta}_t - \theta_\star))^2/2 && (|\dot{\mu}| \leq 1) \\
&\leq \sum_{t=\tau+1}^{T} \|a_t\|^2_{\mathbf{W}_t^{-1}} \|\tilde{\theta}_t - \theta_\star\|^2_{\mathbf{W}_t}/2 && \text{(Cauchy-Schwarz)} \\
&\leq 2\sigma_t(\delta) \sum_{t=\tau+1}^{T} \|a_t\|^2_{\mathbf{W}_t^{-1}} && (\tilde{\theta}_t, \theta_\star \in \mathcal{C}_t(\delta)) \\
&\leq \mathfrak{C} d S^2 \log(T/\delta) \sum_{t=\tau+1}^{T} \|a_t\|^2_{\mathbf{W}_t^{-1}} && \text{(Equation (14))} \\
&\leq \mathfrak{C} d S^2 \log(T/\delta) \sum_{t=\tau+1}^{T} \|a_t\|^2_{\mathbf{V}_t^{-1}} \\
&\leq \mathfrak{C} d^2 \kappa S^2 \log(T/\delta)^2 && \text{(Lemma 9)}
\end{aligned}
$$

We last applied Lemma 9 with $x_t = a_t/\sqrt{\kappa}$, and proceeded with some simple upper-bounding operations. The second to last inequality is a consequence of Abeille et al. (2021, Lemma 9) which ensures:

$$
\begin{aligned}
\dot{\mu}(a_s^\intercal \theta'_{s+1}) &\geq \dot{\mu}(a_s^\intercal \theta_\star) \exp(-|a_s^\intercal (\theta'_{s+1} - \theta_\star)|) \\
&\geq \dot{\mu}(a_s^\intercal \theta_\star) \exp(-1) && (\theta_\star, \theta'_{s+1} \in \Theta,\, \text{diam}_{\mathcal{A}}(\Theta) \leq 1) \\
&\geq \exp(-1)\kappa\,.
\end{aligned}
$$

We now turn our attention to $R_1(T)$.

$$
\begin{aligned}
R_1(T) &= \sum_{t=\tau+1}^{T} \dot{\mu}(a_t^\intercal \theta_\star) a_t^\intercal (\tilde{\theta}_t - \theta_\star) \\
&\leq \sum_{t=\tau+1}^{T} \sqrt{\dot{\mu}(a_t^\intercal \theta_\star)} \sqrt{\exp(|a_t^\intercal (\theta_{t+1} - \theta_\star)|) \dot{\mu}(a_t^\intercal \theta_{t+1})} a_t^\intercal (\tilde{\theta}_t - \theta_\star) && \text{(Abeille et al., 2021, Lemma 9))} \\
&\leq \sqrt{e} \sum_{t=\tau+1}^{T} \sqrt{\dot{\mu}(a_t^\intercal \theta_\star)} \sqrt{\dot{\mu}(a_t^\intercal \theta_{t+1})} a_t^\intercal (\tilde{\theta}_t - \theta_\star) && (\text{diam}_{\mathcal{A}}(\Theta) \leq 1) \\
&\leq \sqrt{e} \sum_{t=\tau+1}^{T} \sqrt{\dot{\mu}(a_t^\intercal \theta_\star)} \sqrt{\dot{\mu}(a_t^\intercal \theta_{t+1})} \|a_t\|_{\mathbf{W}_t^{-1}} \|\tilde{\theta}_t - \theta_\star\|_{\mathbf{W}_t} && \text{(Cauchy-Schwarz)} \\
&\leq e \sum_{t=\tau+1}^{T} \sqrt{\dot{\mu}(a_t^\intercal \theta_\star)} \sqrt{\dot{\mu}(a_t^\intercal \theta_{t+1})} \|a_t\|_{\mathbf{W}_t^{-1}} \left(\|\theta_t - \theta_\star\|_{\mathbf{W}_t} + \|\tilde{\theta}_t - \theta_t\|_{\mathbf{W}_t} \right) && \text{(Triangle ineq.)} \\
&\leq 2e\sqrt{\sigma_T(\delta)} \sum_{t=\tau+1}^{T} \sqrt{\dot{\mu}(a_t^\intercal \theta_\star)} \sqrt{\dot{\mu}(a_t^\intercal \theta_{t+1})} \|a_t\|_{\mathbf{W}_t^{-1}} && (\tilde{\theta}_t, \theta_\star \in \mathcal{C}_t(\delta)) \\
&\leq \mathfrak{C} S \sqrt{d \log(T/\delta)} \sqrt{\sum_{t=\tau+1}^{T} \dot{\mu}(a_t^\intercal \theta_\star)} \sqrt{\sum_{t=\tau+1}^{T} \dot{\mu}(a_t^\intercal \theta_{t+1}) \|a_t\|^2_{\mathbf{W}_t^{-1}}} && \text{(Cauchy-Schwarz)} \\
&\leq \mathfrak{C} S \sqrt{d \log(T/\delta)} \sqrt{d \log(1 + T/d)} \sqrt{\sum_{t=\tau+1}^{T} \dot{\mu}(a_t^\intercal \theta_\star)} && \text{(Lemma 9)}
\end{aligned}
$$

where Lemma 9 was used with $x_t = \sqrt{\dot\mu(a_t^\intercal\theta_{t+1})}a_t$ (and $\dot\mu \le 1$). From then, we can directly follow the proof of Theorem 1 from Abeille et al. (2021) (more precisely, follow the reasoning employed in their Section C.1 page 18) for which we extract that:

$$\sum_{t=\tau+1}^{T} \dot\mu(a_t^\intercal\theta_\star) \le R_T + T\dot\mu(a_\star^\intercal\theta_\star) \,.$$

Assembling the bounds on $R_2(T)$, $R_1(T)$ and $\sum_{t=\tau+1}^{T}\dot\mu(a_t^\intercal\theta_\star)$ we obtain that:

$$R(T) \le \mathfrak{C}\kappa S^2 d^2 \log(T/\delta)^2 + \mathfrak{C}Sd\log(T/\delta)\sqrt{R_T + T\dot\mu(a_\star\theta_\star)} \,.$$

Because $x^2 - bx - c \le 0 \Rightarrow x^2 \le 2b^2 + 2c$ we have:

$$R(T) \le \mathfrak{C}Sd\log(T/\delta)\sqrt{T\dot\mu(a_\star\theta_\star)} + \mathfrak{C}\kappa S^2 d^2\log(T/\delta)^2 \,,$$

which concludes the proof. □

Remark 2. *The scaling w.r.t S of the regret's second-order term is driven by the length τ of the warm-up phase. As anticipated in Remark 1 this scaling is reduced when $\|\hat\theta_\tau\| \le S$ which often happens in practice. In this case, we obtain a second-order term which exactly matches the one of Abeille et al. (2021).*

D.2 The `TS-ECOLog` algorithm

In this section we introduce the TS version of `OFU-ECOLog` whose pseudo-code is provided in Algorithm 3.

Algorithm 3 `TS-ECOLog`

input: failure level δ, warm-up length τ, distribution $\mathcal{D}^{\mathrm{TS}}$
  Set $\Theta \leftarrow$ `WarmUp`(τ) (see Procedure 1). ▷ *forced-exploration*
  Initialize $\theta_{\tau+1} \in \Theta$, $\mathbf{W}_{\tau+1} \leftarrow \mathbf{I}_d$.
  for $t \ge \tau + 1$ **do**
    Set `reject` $\leftarrow$ `true` ▷ *sampling*
    while `reject` **do**
      Sample $\eta \sim \mathcal{D}^{\mathrm{TS}}$, let $\tilde\theta_t = \theta_t + \sigma_t(\delta)\mathbf{W}_t^{-1/2}\eta$.
      If $\tilde\theta_t \in \Theta$ set `reject` $\leftarrow$ `false`
    end while
    Play $a_t \in \arg\max_{a\in\mathcal{A}} a^\intercal\tilde\theta_t$.
    Observe reward r_{t+1}, construct loss $\ell_{t+1}(\theta) = \ell(a_t^\intercal\theta, r_{t+1})$.
    Compute $(\theta_{t+1}, \mathbf{W}_{t+1}) \leftarrow$ `ECOLog`$(1/t, \Theta, \ell_{t+1}, \mathbf{W}_t, \theta_t)$ (see Procedure 2). ▷ *learning*
  end for

The algorithm display little novelty compared to the linear case studied by Agrawal and Goyal (2013); Abeille and Lazaric (2017). The only difference is a rejection sampling step on Θ. The analysis is also similar, up to minor modifications. The following statement provides a regret guarantee for `TS-ECOLog`.

Theorem 5. *Let $\delta \in (0,1]$ and $\mathcal{D}^{\mathrm{TS}}$ a distribution satisfying Definition 1 of Abeille and Lazaric (2017). Setting $\tau = \mathfrak{C}\kappa S^6 d^2\log(T/\delta)^2$ ensures that the regret of* `TS-ECOLog`$(\delta, \tau, \mathcal{D}^{\mathrm{TS}})$ *satisfies with probability at least $1-\delta$:*

$$\mathrm{Regret}(T) \le \mathfrak{C}Sd^{3/2}\log(T/\delta)\sqrt{T\dot\mu(a_\star^\intercal\theta_\star)} + \mathfrak{C}S^6\kappa d^3\log(T/\delta)^2 \,.$$

Proof. According to Proposition 3 (its detailed version in Appendix C.4) setting $\tau = \mathfrak{C}\kappa S^6 d^2\log(T/\delta)$ ensures:

$$\mathbb{P}\left(\theta_\star \in \Theta \text{ and } \|\theta_t - \theta_\star\|_{\mathbf{W}_t}^2 \le \sigma_t(\delta) \text{ for all } t \ge \tau+1\right) \ge 1 - 2\delta \,.$$

As in the proof of Theorem 1 we assume that the above event, denoted E_δ, holds.

We decompose the regret as:

$$\begin{aligned}\text{Regret}(T) &\leq \tau + \sum_{t=\tau+1}^{T} \mu(a_\star^\intercal\theta_\star) - \mu(a_t^\intercal\theta_\star) \\ &= \tau + \sum_{t=\tau+1}^{T} \mu(a_\star^\intercal\theta_\star) - \mu(a_t^\intercal\tilde{\theta}_t) + \sum_{t=\tau+1}^{T} \mu(a_t^\intercal\tilde{\theta}_t) - \mu(a_t^\intercal\theta_\star) \\ &\leq \mathfrak{C}\kappa S^6 d^2 \log(T/\delta)^2 + R^{\text{TS}}(T) + R^{\text{PRED}}(T)\ .\end{aligned}$$

Above, we defined $R^{\text{TS}}(T) = \sum_{t=\tau+1}^{T} \mu(a_\star^\intercal\theta_\star) - \mu(a_t^\intercal\tilde{\theta}_t)$ and $R^{\text{PRED}}(T) = \sum_{t=\tau+1}^{T} \mu(a_t^\intercal\tilde{\theta}_t) - \mu(a_t^\intercal\theta_\star)$. To bound $R^{\text{PRED}}(T)$ one can directly follow the strategy employed in Appendix D.1. The only difference comes from the radius of the "effective" confidence set that is used - inflated by $\sqrt{d}$ because of the concentration properties of $\mathcal{D}^{\text{TS}}$. This leads to:

$$R^{\text{PRED}}(T) \leq \mathfrak{C}Sd^{3/2}\log(T/\delta)\sqrt{T\dot{\mu}(a_\star\theta_\star)} + \mathfrak{C}\kappa S^2 d^3 \log(T/\delta)^2$$

We now turn to $R^{\text{TS}}(T)$. Following Abeille and Lazaric (2017) we denote $J(\theta) = \max_{a\in\mathcal{A}} a^\intercal\theta$. We have:

$$\begin{aligned}R^{\text{TS}}(T) &= \sum_{t=\tau+1}^{T} \mu(a_\star^\intercal\theta_\star) - \mu(a_t^\intercal\tilde{\theta}_t) \\ &= \sum_{t=\tau+1}^{T} \alpha(a_\star^\intercal\theta_\star, a_t^\intercal\tilde{\theta}_t)(a_\star^\intercal\theta_\star - a_t^\intercal\tilde{\theta}_t) && \text{(exact first-order Taylor)} \\ &= \sum_{t=\tau+1}^{T} \alpha(J(\theta_\star), J(\tilde{\theta}_t))(J(\theta_\star) - J(\tilde{\theta}_t)) && \text{(def. of } J\text{)} \qquad (34)\end{aligned}$$

By convexity of J along with the computations of its sub-gradients (see Section C of Abeille and Lazaric (2017));

$$\begin{aligned}|J(\theta_\star) - J(\tilde{\theta}_t)| &\leq \max\left\{|\nabla J(\theta_\star)^\intercal(\theta_\star - \tilde{\theta}_t)|, |\nabla J(\tilde{\theta}_t)^\intercal(\theta_\star - \tilde{\theta}_t)|\right\} && \text{(convexity of } J\text{)} \\ &\leq \max\left\{|a_\star^\intercal(\theta_\star - \tilde{\theta}_t)|, |a_t^\intercal(\theta_\star - \tilde{\theta}_t)|\right\} && (\nabla J(\theta) = \arg\max_{a\in\mathcal{A}} a^\intercal\theta) \\ &\leq \text{diam}_{\mathcal{A}}(\Theta) && (\tilde{\theta}_t, \theta_\star \in \Theta) \\ &\leq 1\ . && \text{(Proposition 5)}\end{aligned}$$

Therefore:

$$\begin{aligned}\alpha(J(\theta_\star), J(\tilde{\theta}_t)) &= \int_{v=0}^{1} \dot{\mu}(J(\theta_\star) + v(J(\tilde{\theta}_t) - J(\theta_\star))dv \\ &\leq \dot{\mu}(J(\theta_\star)) \int_{v=0}^{1} \exp(v|J(\tilde{\theta}_t) - J(\theta_\star)|)dv && \text{(Lemma 9 of Abeille et al. (2021))} \\ &\leq \dot{\mu}(J(\theta_\star)) \int_{v=0}^{1} \exp(v)dv && (|J(\theta_\star) - J(\tilde{\theta}_t) \leq 1) \\ &\leq 2\dot{\mu}(J(\theta_\star))\ .\end{aligned}$$

Plugging the above inequality in Equation (34) yields:

$$\begin{aligned}R^{\text{TS}}(T) &\leq 2\dot{\mu}(J(\theta_\star)) \sum_{t=\tau+1}^{T} J(\theta_\star) - J(\tilde{\theta}_t) \\ &= 2\dot{\mu}(a_\star^\intercal\theta_\star) \sum_{t=\tau+1}^{T} J(\theta_\star) - J(\tilde{\theta}_t)\ .\end{aligned}$$

From then on we can follow the proof of Abeille and Lazaric (2017) which in the linear case studies exactly $\sum_t J(\theta_\star) - J(\tilde{\theta}_t)$. Directly following their line of proof yields $\sum_t J(\theta_\star) - J(\tilde{\theta}_t) \lesssim \mathfrak{C}\sqrt{d}\sqrt{\sigma_T(\delta)}\sum_t \|a_t\|_{\mathbf{W}_t^{-1}} + \sqrt{T}$. This concludes the proof since $\sigma_T(\delta) \leq \mathfrak{C}S^2 d\log(T/\delta)$ and:

$$\begin{aligned}
\dot{\mu}(a_\star^\mathsf{T}\theta_\star)\sum_{t=\tau+1}^{T}\|a_t\|_{\mathbf{W}_t^{-1}} &= \sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_\star)}\sum_{t=\tau+1}^{T}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_\star)}\|a_t\|_{\mathbf{W}_t^{-1}} \\
&\leq \mathfrak{C}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_\star)}\sum_{t=\tau+1}^{T}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_{t+1})}\sqrt{\exp(|a_\star^\mathsf{T}(\theta_\star-\theta_{t+1})|)}\|a_t\|_{\mathbf{W}_t^{-1}} \\
&\leq \mathfrak{C}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_\star)}\sum_{t=\tau+1}^{T}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_{t+1})}\sqrt{\exp(2\mathrm{diam}_{\mathcal{A}}(\Theta))}\|a_t\|_{\mathbf{W}_t^{-1}} && (\theta_{t+1},\theta_\star\in\Theta) \\
&\leq \mathfrak{C}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_\star)}\sum_{t=\tau+1}^{T}\sqrt{\dot{\mu}(a_\star^\mathsf{T}\theta_{t+1})}\|a_t\|_{\mathbf{W}_t^{-1}} && (\mathrm{diam}_{\mathcal{A}}(\Theta)\leq 1) \\
&\leq \mathfrak{C}\sqrt{T\dot{\mu}(a_\star^\mathsf{T}\theta_\star)}\sqrt{d\log(T)}\,.
\end{aligned}$$

where we last used Cauchy-Schwarz inequality and Lemma 9. □

D.3 Proof of Theorem 2

Theorem 2. *Let $\delta\in[0,1)$. With probability at least $1-\delta$ the regret of* `ada-OFU-ECOLog`(δ) *satisfies:*

$$\mathrm{Regret}(T) \leq \mathfrak{C}Sd\sqrt{\sum_{t=1}^{T}\dot{\mu}(a_{\star,t}^\mathsf{T}\theta_\star)\log(T/\delta)} + \mathfrak{C}S^6\kappa d^2\log(T/\delta)^2\,,$$

where $a_{\star,t} = \arg\max_{a\in\mathcal{A}_t} a^\mathsf{T}\theta_\star$.

Algorithm 2 `ada-OFU-ECOLog`

input: failure level δ.
Initialize $\Theta_1 = \{\|\theta\|\leq S\}$, $\mathcal{C}_1(\delta)\leftarrow\Theta_1$, $\theta_1\in\Theta$, $\mathbf{W}_1\leftarrow\mathbf{I}_d$ and $\mathcal{H}_1\leftarrow\emptyset$.
for $t\geq 1$ **do**
  Play $a_t\in\arg\max_{a\in\mathcal{A}}\max_{\theta\in\mathcal{C}_t(\delta)} a^\mathsf{T}\theta$, observe reward r_{t+1}.
  Compute the estimators θ_t^0, θ_t^1 (see Equation (8)) and $\bar{\theta}_t$.
  if $\dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t)\leq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^0)$ and $\dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t)\leq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^1)$ **then**
    Form the loss ℓ_{t+1} and compute $(\theta_{t+1}, \mathbf{W}_{t+1})\leftarrow$ `ECOLog`$(1/t,\Theta_t,\ell_{t+1},\mathbf{W}_t,\theta_t)$.
    Compute $\mathcal{C}_{t+1}(\delta)\leftarrow\left\{\|\theta-\theta_{t+1}\|^2_{\mathbf{W}_{t+1}}\leq\eta_t(\delta)\right\}$, set $\mathcal{H}_{t+1}\leftarrow\mathcal{H}_t$. ▷ *$\eta_t(\delta)$ is defined in Equation (15)*
  else
    Set $\mathcal{H}_{t+1}\leftarrow\mathcal{H}_t\cup\{a_t,r_{t+1}\}$ and compute $\hat{\theta}^{\mathcal{H}}_{t+1} = \arg\min\sum_{(a,r)\in\mathcal{H}_{t+1}}\ell(a^\mathsf{T}\theta,r)+\gamma_t(\delta)\|\theta\|^2$.
    Update $\mathbf{V}_t^{\mathcal{H}}\leftarrow\sum_{a\in\mathcal{H}_{t+1}}aa^\mathsf{T}/\kappa+\gamma_t(\delta)\mathbf{I}_d$, $\theta_{t+1}\leftarrow\theta_t$ and $\mathbf{W}_{t+1}\leftarrow\mathbf{W}_t$.
    Compute $\Theta_{t+1} = \left\{\|\theta-\hat{\theta}^{\mathcal{H}}_{t+1}\|^2_{\mathbf{V}_t^{\mathcal{H}}}\leq\beta_t(\delta)\right\}\cap\Theta_1$. ▷ *$\beta_t(\delta)$ is defined in Equation (12)*
  end if
end for

Proof. We denote $\mathcal{T}$ the set of rounds at which condition (C_1) breaks. Formally;

$$\mathcal{T} := \left\{t\in[T],\ \dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t)\geq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^1) \text{ or } \dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t)\geq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^0)\right\}\,.$$

We claim the following result bounding the cardinality of $\mathcal{T}$. The proof is deferred to Appendix D.3.1.

Lemma 8. *The following inequality holds:*

$$|\mathcal{T}| \leq \mathfrak{C}S^6\kappa d^2\log(T/\delta)^2\,.$$

We follow a naive (but sufficient) bounding strategy. For rounds $t \in \mathcal{T}$ we crudely bound the instantaneous regret by its maximal value (*e.g.* 1);

$$\begin{aligned}\text{Regret}(T) &\leq |\mathcal{T}| + \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_{\star,t}^\intercal\theta_\star) - \dot{\mu}(a_t^\intercal\theta_\star) && (\mu \in (0,1)) \\ &\leq \mathfrak{C}S^6\kappa d^2\log(T/\delta)^2 + R_T && \text{(Lemma 8)}\end{aligned}$$

where $R_T := \sum_{t\notin\mathcal{T}} \dot{\mu}(a_{\star,t}^\intercal\theta_\star) - \dot{\mu}(a_t^\intercal\theta_\star)$. In the following, we denote for $t \notin \tau$:

$$(a_t, \tilde{\theta}_t) \in \underset{\mathcal{A}_t, \mathcal{C}_t(\delta)}{\arg\max}\, a^\intercal\theta \;.$$

where $\mathcal{C}_t(\delta) = \{\theta,\ \|\theta_t - \theta\|^2_{\mathbf{W}_t} \leq \eta_t(\delta)\}$ and $\eta_t(\delta)$ is defined in Equation (15). In the following, we assume that the following event holds:

$$E_\delta = \{t \in [T]\setminus\mathcal{T},\ \theta_\star \in \mathcal{C}_t(\delta)\cap\Theta_t\}\ ,$$

This happens with probability at least $1-2\delta$ according to Proposition 7. This implies that the couple $(a_t, \tilde{\theta}_t)$ is optimistic. Formally: $a_t^\intercal\tilde{\theta}_t \geq a_{\star,t}^\intercal\theta_\star$. Therefore:

$$\begin{aligned}R(T) &= \sum_{t\in[T]\setminus\mathcal{T}} \mu(a_{\star,t}^\intercal\theta_\star) - \mu(a_t^\intercal\theta_\star) \\ &\leq \sum_{t\in[T]\setminus\mathcal{T}} \mu(a_t^\intercal\tilde{\theta}_t) - \mu(a_t^\intercal\theta_\star) && \text{(optimism, } \mu\nearrow\text{)} \\ &\leq \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_\star)a_t^\intercal(\tilde{\theta}_t - \theta_\star) + \tilde{\alpha}(a_t^\intercal\theta_\star, a_t^\intercal\tilde{\theta}_t)(a_t^\intercal(\tilde{\theta}_t-\theta_\star))^2 && \text{(Taylor, } |\ddot{\mu}|\leq\dot{\mu}\text{)} \\ &=: R_1(T) + R_2(T)\;.\end{aligned}$$

The bound on $R_2(T)$ is directly extracted from the proof of Theorem 1 presented in Appendix D.1.

$$R_2(T) \leq \mathfrak{C}d^2\kappa S^2\log(T/\delta)^2\;.$$

The story is slightly different for $R_1(T)$ and the proof laid out in Appendix D.1 needs to be slightly adapted. We need to differentiate the rounds where $\dot{\mu}(a_t^\intercal\theta_\star) \leq \dot{\mu}(a_t^\intercal\theta_{t+1})$ and the rounds where $\dot{\mu}(a_t^\intercal\theta_\star) \geq \dot{\mu}(a_t^\intercal\theta_{t+1})$. In what follows we focus only on the latter (for the former we can directly adapt the approach laid out in Appendix D.1).

$$\begin{aligned}R_1(T) &= \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_\star)a_t^\intercal(\tilde{\theta}_t - \theta_\star) \\ &\leq \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_{t+1})a_t^\intercal(\tilde{\theta}_t-\theta_\star) + a_t^\intercal(\tilde{\theta}_t-\theta_\star)|a_t^\intercal(\theta_\star-\theta_{t+1})| && (|\ddot{\mu}|\leq 1) \\ &\leq \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_{t+1})\|a_t\|_{\mathbf{W}_t^{-1}}\|\tilde{\theta}_t-\theta_\star\|_{\mathbf{W}_t} + \|a_t\|^2_{\mathbf{W}_t^{-1}}\|\tilde{\theta}_t-\theta_\star\|_{\mathbf{W}_t}\|\theta_{t+1}-\theta_\star\|_{\mathbf{W}_t} \\ &\leq \mathfrak{C}\sqrt{\eta_T(\delta)}\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_{t+1})\|a_t\|_{\mathbf{W}_t^{-1}} + \mathfrak{C}\sqrt{\eta_T(\delta)}\sum_{t\in[T]\setminus\mathcal{T}}\|a_t\|^2_{\mathbf{W}_t^{-1}}\|\theta_{t+1}-\theta_\star\|_{\mathbf{W}_t} \\ &\leq \mathfrak{C}\sqrt{\eta_T(\delta)}\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_{t+1})\|a_t\|_{\mathbf{W}_t^{-1}} + \mathfrak{C}\sqrt{\eta_T(\delta)}\sum_{t\in[T]\setminus\mathcal{T}}\|a_t\|^2_{\mathbf{W}_t^{-1}}\|\theta_{t+1}-\theta_\star\|_{\mathbf{W}_{t+1}} && (\mathbf{W}_t \succeq \mathbf{W}_{t+1}) \\ &\leq \mathfrak{C}\sqrt{\eta_T(\delta)}\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\intercal\theta_{t+1})\|a_t\|_{\mathbf{W}_t^{-1}} + \mathfrak{C}\eta_T(\delta)\sum_{t\in[T]\setminus\mathcal{T}}\|a_t\|^2_{\mathbf{W}_t^{-1}}\end{aligned}$$

The second term in the above inequality is bounded exactly as in $R_2(T)$; this yields:

$$\begin{aligned}
R_1(T) &\leq \mathfrak{C}\sqrt{\eta_T(\delta)} \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\mathsf{T}\theta_{t+1})\|a_t\|_{\mathbf{W}_t^{-1}} + \mathfrak{C}\kappa S^2 d^2 \log(T)^2 && \text{(cf. bound on } R_2(T)\text{)} \\
&\leq \mathfrak{C}\sqrt{\eta_T(\delta)} \sqrt{\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\mathsf{T}\theta_{t+1})} \sqrt{\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\mathsf{T}\theta_{t+1}\|a_t\|^2_{\mathbf{W}_t^{-1}}} + \mathfrak{C}\kappa S^2 d^2 \log(T/\delta)^2 && \text{(Cauchy-Schwarz)} \\
&\leq \mathfrak{C}Sd\log(T/\delta) \sqrt{\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\mathsf{T}\theta_{t+1})} + \mathfrak{C}\kappa S^2 d^2 \log(T/\delta)^2 && \text{(Lemma 9)} \\
&\leq \mathfrak{C}Sd\log(T/\delta) \sqrt{\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\mathsf{T}\theta_\star)} + \mathfrak{C}\kappa S^2 d^2 \log(T/\delta)^2 && \text{(by hyp.)}
\end{aligned}$$

Again, by following the proof of Theorem 1 from Abeille et al. (2021) we get that:

$$\sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_t^\mathsf{T}\theta_\star) \leq R_T + \sum_{t\in[T]\setminus\mathcal{T}} \dot{\mu}(a_{\star,t}^\mathsf{T}\theta_\star) \,.$$

Assembling the different bounds and solving the implicit inequation on R_T yields the announced result. □

D.3.1 Proof of Lemma 8

Lemma 8. *The following inequality holds:*

$$|\mathcal{T}| \leq \mathfrak{C}S^6\kappa d^2 \log(T/\delta)^2 \,.$$

Proof. Denote for $u \in \{0,1\}$:

$$\mathcal{T}_u := \left\{t \in [T],\ \dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t) \geq 2\dot{\mu}(a_t^\mathsf{T}\theta_t^u)\right\} \,,$$

so that $\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1$. By Abeille et al. (2021, Lemma 9) we know that:

$$\dot{\mu}(a_t^\mathsf{T}\bar{\theta}_t) \leq \dot{\mu}(a_t^\mathsf{T}\theta_t^u) \exp\left(|a_t^\mathsf{T}(\bar{\theta}_t - \theta_t^u)|\right)$$

Therefore by straight-forward manipulations:

$$t \in \mathcal{T} \Longrightarrow \exists u \in \{0,1\} \text{ s.t } |a_t^\mathsf{T}(\bar{\theta}_t - \theta_t^u)| \geq \log(2) \,. \tag{35}$$

We can now bound $|\mathcal{T}|$ thanks to the form of Θ_t (which contains $\bar{\theta}_t$ and θ_t^u by construction) and the Elliptical Potential lemma.

$$\begin{aligned}
|\mathcal{T}|\log(2)^2 &\leq \sum_{t\in\mathcal{T}} \left|a_t^\mathsf{T}(\bar{\theta}_t - \theta_t^u)\right|^2 \\
&\leq \sum_{t\in\mathcal{T}_0} \|a_t\|^2_{(\mathbf{V}_t^{\mathcal{H}})^{-1}} \left\|\bar{\theta}_t - \theta_t^u\right\|^2_{\mathbf{V}_t^{\mathcal{H}}} && \text{(Cauchy-Schwarz)} \\
&\leq 4\beta_T(\delta) \sum_{t\in\mathcal{T}} \|a_t\|^2_{(\mathbf{V}_t^{\mathcal{H}})^{-1}} && (\bar{\theta}_t,\, \theta_t^u \in \Theta_t) \\
&\leq 8\kappa\beta_T(\delta) d\log(T) && \text{(Lemma 9)}
\end{aligned}$$

We applied Lemma 9 with $x_s = a_s/\sqrt{\kappa}$, and after checking that in Algorithm 2 the matrix $\mathbf{V}_t^{\mathcal{H}}$ is indeed updated in rounds $t \in \mathcal{T}$. This conclude the proof since $\beta_t(\delta) \leq \mathfrak{C}S^6 d\log(T/\delta)$.

□

E Computational Costs

E.1 Proof of Propositions 1 and 4

The goal of this section is to examine the per-round complexity of the algorithms laid out in the main paper.

E.1.1 Per-Round Cost of ECOLog

We start by the main computational bottleneck of our approach, which is the ECOLog procedure (see its sequential form in Algorithm 4). It involves computing θ'_{t+1} - an ε_t-approximation (in ℓ_2-norm) of θ_{t+1}, and updating the matrix $\mathbf{W}_{t+1}$ (along with its inverse which will be used for the planning mechanism). We claim the following result, a slightly more detailed version of Proposition 2 in the main text.

Proposition 8. *Fix $t \in \mathbb{N}^+$. Assume that Θ_t is a bounded and closed ellipsoid and $\varepsilon_t > 0$. Completing round t of* ECOLog *can be done within $\mathcal{O}(d^2 \log(\mathrm{diam}(\Theta_t))/\varepsilon_t)^2)$ operations.*

Proof. Given θ'_{t+1} the matrix $\mathbf{W}_{t+1}$ can be updated at cost $\mathcal{O}(d^2)$ since:

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \dot{\mu}(a_t^\mathsf{T}\theta'_{t+1})a_t a_t \ .$$

The cost of maintaining $\mathbf{W}_{t+1}^{-1}$ is the same thanks to the Sherman-Morrison formula. The main computational complexity therefore stems from the computation of θ'_{t+1}. Recall:

$$\theta_{t+1} = \operatorname*{arg\,min}_{\theta\in\Theta_t} \frac{1}{2+D}\|\theta - \theta'_t\|^2_{\mathbf{W}_t} + \ell_{t+1}(\theta) \ ,$$

where $D \geq \mathrm{diam}_{\mathcal{A}}(\Theta_t)$. Let $\mathbf{W}_t = \mathbf{L}_t\mathbf{L}_t^\mathsf{T}$ the Cholesky decomposition of $\mathbf{W}_t$ (it exists since $\mathbf{W}_t$ is p.s.d). By denoting $z_t = \mathbf{L}_t^\mathsf{T}\theta'_t$, performing the change of variable $z \leftarrow \mathbf{L}_t^\mathsf{T}\theta$ and removing constants we obtain:

$$\theta_{t+1} = \mathbf{L}_t^{-\mathsf{T}} \operatorname*{arg\,min}_{\mathbf{L}_t^{-\mathsf{T}}z\in\Theta} \left(\bar{L}_{t+1}(z) := \frac{1}{2+D}\|z\|^2 + \frac{2}{2+D}z^\mathsf{T}z_t + \ell_{t+1}(\mathbf{L}_t^{-\mathsf{T}}z) \right) \ .$$

By direct computations:

$$\nabla^2\bar{L}_{t+1}(z) = (1+D/2)^{-1}\mathbf{I}_d + \dot{\mu}(a_t^\mathsf{T}\mathbf{L}_t^{-\mathsf{T}}z)\mathbf{L}_t^{-1}a_t a_t^\mathsf{T}\mathbf{L}_t^{-\mathsf{T}} \ .$$

proving that for all $z \in \mathbb{R}^d$ (using the fact that $\dot{\mu} \in [0, 1/4]$ and $\mathbf{W}_t \succeq \mathbf{I}_d$) :

$$0 \prec (1+D/2)^{-1} \preceq \nabla^2\bar{L}_{t+1}(z) \preceq (1+D/2)^{-1} + 1/4 \ .$$

The function $\bar{L}_{t+1}(z)$ is therefore strongly convex and $(5/4 + D/8)^{-1}$ well-conditioned. Furthermore, note the convexity of the constraint $\{z,\ \mathbf{L}_t^{-\mathsf{T}}z \in \Theta\}$ since Θ itself is convex.

Let θ'_{t+1} be returned by the Projected Gradient Descent algorithm (see (Hazan, 2016, Algorithm 2) for instance) ran for T steps, where:

$$T = (9/4 + D/8)\log(\mathrm{diam}(\Theta_t)/\varepsilon_t) \ .$$

By Lemma 12 this ensures that:

$$\|\theta_{t+1} - \theta'_{t+1}\| \leq \varepsilon_t \ ,$$

which is enough to complete round t of the ECOLog procedure. Because the gradients of $\bar{L}_{t+1}(\theta)$ only take $\mathcal{O}(d^2)$ operations to compute, the cost of running the Projected Gradient Descent algorithm for T rounds is $\mathcal{O}(T(d^2 + \mathrm{proj}(\Theta_t)))$. The quantity $\mathrm{proj}(\Theta_t)$ is the cost of projection the estimate on the set $\{\mathbf{L}_t^{-\mathsf{T}}z \in \Theta_t\}$. This constraint set is ellipsoidal since Θ_t is an ellipsoid (by assumption). Projecting on this set therefore boils down to solving a one-dimensional convex problem (see Lemma 13). Similarly, this program is solved to accuracy ε in $\mathcal{O}(d^2\log(1/\varepsilon))$ (it involves some matrix-vector multiplications and triangular inverse solving, hence the d^2 dependency). To finish the proof we are therefore left with evaluating the cost of computing the Cholesky factor $\mathbf{L}_t$. This quantity can be maintained online and updated at cost $\mathcal{O}(d^2)$ thanks to the rank-one nature of $\mathbf{W}_t$'s update (see for instance Golub and van Loan (2013, Section 6.5.4)). □

E.1.2 Proof of Proposition 1

Proposition 1 (Computational Cost). *Let $|\mathcal{A}| = K < \infty$. Each round t of* `OFU-ECOLog` *can be completed within $\mathcal{O}(Kd^2 + d^2\log(t)^2)$ operations.*

Proof. Recall that in `OFU-ECOLog` we have $\Theta_t \equiv \Theta$ where $D = 1$ satisfies $D \geq \mathrm{diam}_{\mathcal{A}}(\Theta)$ (see Proposition 5) and $\varepsilon_t = 1/t$. Easy computations further show that $\mathrm{diam}(\Theta) \leq \mathrm{poly}(S)$; for instance, a crude bound yields that for all $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned}
\|\theta_1 - \theta_2\|^2 &\leq \lambda_\tau(\delta)^{-1} \|\theta_1 - \theta_2\|_{\mathbf{V}_\tau}^2 && (\mathbf{V}_\tau \geq \lambda_\tau(\delta)\mathbf{I}_d) \\
&\leq 4\lambda_\tau(\delta)^{-1}\beta_\tau(\delta) && (\theta_1, \theta_2 \in \Theta) \\
&= \mathrm{poly}(S) && \text{(see Equations (10) and (12))}
\end{aligned}$$

Proposition 8 hence ensures that the cost running the `ECOLog` routine at round t of `OFU-ECOLog` is at most $\mathfrak{C}d^2\log(\mathrm{poly}(S)t)^2$. The optimistic planning mechanism requires performing K matrix-vector products, which cost is $\mathcal{O}(Kd^2)$. This finishes the proof. □

Remark. *The proof discards the cost of the warm-up; its only computational bottleneck is the computation of $\hat{\theta}_\tau$. This happens only once and boils down to the minimization of a well-conditioned (after preconditioning by $\mathbf{V}_\tau$) convex function - which is therefore cheap, typically $\mathcal{O}(\tau\log(T))$ where τ is the length of the warm-up.*

E.1.3 Proof of Proposition 4

Proposition 4. *The per-round computational cost of Algorithm 2 is bounded by $\mathcal{O}\left(\kappa + Kd^2 + d^2\log(T)^2\right)$.*

Proof. The proof is essentially the same as for Proposition 1. The main difference is the value of $\mathrm{diam}_{\mathcal{A}}(\Theta)$; it is now bounded by $\mathrm{poly}(S)$ (by using a similar argument that in Appendix E.1.2 when we bounded $\mathrm{diam}(\Theta)$). As discussed in the main text there is however an additional cost inherited from the computations of $\hat{\theta}_t^{\mathcal{H}}$. This requires minimizing a well-conditioned (after preconditioning by $\mathbf{V}_t^{\mathcal{H}}$) convex function which gradients are computed at $\mathcal{O}(d|\mathcal{H}_t|)$ cost. Lemma 8 proves that $|\mathcal{H}_t| \leq \kappa$; therefore the computational overhead is $\mathcal{O}(\kappa d\log(T))$. Note that precisely because of Lemma 8, it turns out that this extra-cost only needs to be paid at most $\approx \kappa$ times (and not at every round as suggested by Proposition 4). □

E.2 Computational Costs of Other Approaches

We briefly discuss the computational cost we announced in Table 1 for `GLM-UCB` Filippi et al. (2010) and `OFULog-r` Abeille et al. (2021). Both require the computation of the MLE estimator:

$$\hat{\theta}_{t+1} = \arg\min_\theta \left\{ L_{t+1}(\theta) := \sum_{s=1}^{t} \ell_{s+1}(\theta) + \lambda\|\theta\|^2 \right\} .$$

An efficient way to solve $\hat{\theta}_{t+1}$ to ε accuracy (typically with $\varepsilon = 1/T$ to preserve regret guarantees) is to run a gradient descent (GD) algorithm with $\mathbf{V}_t$-preconditioning. This step is important as in all generality L_{t+1} can be $1/t$ well-conditioned; running GD directly on L_{t+1} will therefore require $\mathcal{O}(t\log(1/\varepsilon)$ to reach ε-accuracy. Preconditioning allows to reduce this cost to $\mathcal{O}(\log(1/\varepsilon))$. The cost of computing the gradient of L_{t+1} (and its pre-conditioned version) is still high, typically $\Omega(t)$ (more precisely, $\Omega(d^2t)$ for the preconditioned version which involves matrix-vector multiplication). Overall, the cost of computing $\hat{\theta}_{t+1}$ to ε accuracy is therefore $\mathcal{O}(d^2t\log(1/\varepsilon))$.

For `GLM-UCB` a $\mathcal{O}(d^2K)$ additional cost is to be added to account for the optimistic planning. Things are worse for `OFULog-r` as at round t and *for every arm* $a \in \mathcal{A}$ it needs to solve a convex program of the form:

$$\max_\theta \left\{ a^\mathsf{T}\theta \text{ s.t } L_{t+1}(\theta) \leq \gamma_t(\delta) \right\} .$$

Projecting on the set $\{L_{t+1}(\theta) \leq \gamma_t(\delta)\}$ is as costly as computating of $\hat{\theta}_{t+1}$, hence the additional $\widetilde{\mathcal{O}}(Kd^2T)$ computational cost.

F AUXILIARY RESULTS

The following version of the Elliptical Potential lemma (see, *e.g*, (Abbasi-Yadkori et al., 2011, Lemma 11)) is a direct consequence of (Faury et al., 2020, Lemma 15) along with the determinant-trace inequality (see Lemma 10).

Lemma 9 (Elliptical potential)**.** *Let $\lambda \geq 1$ and $\{x_s\}_{s=1}^{\infty}$ a sequence in $\mathbb{R}^d$ such that $\|x_s\| \leq X$ for all $s \in \mathbb{N}$. Further, define $\mathbf{V}_t := \sum_{s=1}^{t} x_s x_s^\mathsf{T} + \lambda \mathbf{I}_d$. Then:*

$$\sum_{t=1}^{T} \|x_t\|^2_{\mathbf{V}_{t-1}^{-1}} \leq 2d(1+X^2)\log\left(1+\frac{TX^2}{d\lambda}\right)$$

The following is extracted from (Abbasi-Yadkori et al., 2011, Lemma 10).

Lemma 10 (Determinant-Trace inequality)**.** *Let $\{x_s\}_{s=1}^{\infty}$ a sequence in $\mathbb{R}^d$ such that $\|x_s\| \leq X$ for all $s \in \mathbb{N}$, and let λ be a non-negative scalar. For $t \geq 1$ define $\mathbf{V}_t := \sum_{s=1}^{t} x_s x_s^\mathsf{T} + \lambda \mathbf{I}_d$. The following inequality holds:*

$$\det(\mathbf{V}_t) \leq \left(\lambda + tX^2/d\right)^d$$

The following statements are standard results from the convex optimization literature.

Lemma 11 (Section 4.2.3 of Boyd and Vandenberghe (2004))**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ a differentiable and convex function and $\mathcal{C} \subset \mathbb{R}^d$ a convex set. Further, denote:*

$$x_0 := \arg\min_{x \in \mathcal{C}} f(x) \ .$$

Then for any *$y \in \mathcal{C}$:*

$$\nabla f(x_0)^\mathsf{T}(y - x) \geq 0 \ .$$

Lemma 12. *Let $f : \mathbb{R}^d \to \mathbb{R}$ a twice differentiable and strongly convex function such that for all $x \in \mathbb{R}^d$:*

$$0 \preceq \alpha \mathbf{I}_d \preceq \nabla^2 f(x) \preceq \beta \mathbf{I}_d \ .$$

Let $\mathcal{C} \subset \mathbb{R}^d$ a convex set, $x_0 = \arg\min_{\mathcal{C}} f(x)$ and $\gamma = \alpha/\beta$. Let x_{T+1} be the estimator returned by the projected gradient descent algorithm (Algorithm 2 in Hazan (2016)) with step-size $1/\beta$ run for T rounds. For $\varepsilon > 0$ setting:

$$T = (1+\gamma^{-1})\log\left(\mathrm{diam}(C)/\varepsilon\right) \ ,$$

ensures that $\|x_{T+1} - x_0\| \leq \varepsilon$.

Proof. The proof is standard in the convex optimization literature. We remind it briefly for completeness.

$$\begin{aligned} f(x_{T+1}) &\geq f(x_0) + \nabla f(x_0)^\mathsf{T}(x_{T+1} - x_0) + \frac{\alpha}{2}\|x_{T+1} - x_0\|^2 \\ &\geq f(x_0) + \frac{\alpha}{2}\|x_{T+1} - x_0\|^2 \qquad \text{(Lemma 11)} \end{aligned}$$

Furthermore by convexity:

$$\begin{aligned} f(x_{T+1}) &\leq f(x_T) + \nabla f(x_T)^\mathsf{T}(x_{T+1} - x_T) + \frac{\beta}{2}\|x_{T+1} - x_T\|^2 \\ &\leq f(x_T) + \nabla f(x_T)^\mathsf{T}(x_0 - x_T) - \frac{\beta}{2}\|x_{T+1} - x_0\|^2 + \frac{\beta}{2}\|x_T - x_0\|^2 \\ &\leq f(x_0) - \frac{\beta}{2}\|x_0 - x_{T+1}\|^2 + \frac{\beta - \alpha}{2}\|x_T - x_0\|^2 \ . \end{aligned}$$

The second to last inequality uses the definition of x_{T+1} (given by the projected gradient descent algorithm). Plugging everything together yields:

$$\begin{aligned}\|x_{T+1} - x_0\|^2 &\leq \frac{\beta-\alpha}{\beta+\alpha}\|x_T - x_0\|^2 \\ &\leq \left(\frac{\beta-\alpha}{\beta+\alpha}\right)^T \|x_1 - x_0\|^2 \\ &\leq \left(1 - \frac{2\alpha}{\beta+\alpha}\right)^T \mathrm{diam}(\mathcal{C})^2 \\ &\leq \exp(-2T\alpha/(\beta+\alpha))\mathrm{diam}(\mathcal{C})^2 \ .\end{aligned}$$

Solving for $\|x_{T+1} - x_0\|^2 \leq \varepsilon^2$ yields the announced result. □

Lemma 13 (Ellipsoidal Projection)**.** *Let $x \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{d\times d}$ a p.s.d matrix. Let y be the projection of x onto the set $\{z, \|z\|^2_{\mathbf{A}}/2 \leq 1\}$. Then $y = (\mathbf{I}_d + \lambda_\star \mathbf{A}^{-1})^{-1}x$ where $\lambda_\star$ is the solution of the following one-dimensional strongly concave program:*

$$\lambda_\star = \operatorname*{arg\,max}_{\lambda\geq 0} -2\lambda - x^\mathsf{T}\mathbf{A}^{1/2}(\lambda\mathbf{I}_d + \mathbf{A})^{-1}\mathbf{A}^{1/2}x \ .$$

Proof. By definition of the projection onto a convex set:

$$y := \operatorname*{arg\,min}_{\frac{1}{2}\|z\|^2_{\mathbf{A}^{-1}}\leq 1} \left\{ f(z) := \frac{1}{2}\|x - z\|^2 \right\}$$

Introducing the Lagrangian $L(z,\lambda) := \|x-z\|^2/2 + \lambda(\|z\|^2_{\mathbf{A}^{-1}}/2 - 1)$ and by strong duality:

$$\begin{aligned} f(y) &= \min_z \max_{\lambda\geq 0} L(z,\lambda) \\ &= \max_{\lambda\geq 0} \min_z L(z,\lambda) \ .\end{aligned}$$

Denoting $z(\lambda) = \arg\min_z L(z,\lambda)$, direct computation yields that:

$$z(\lambda) = (\mathbf{I}_d + \lambda\mathbf{A}^{-1})^{-1}x \ .$$

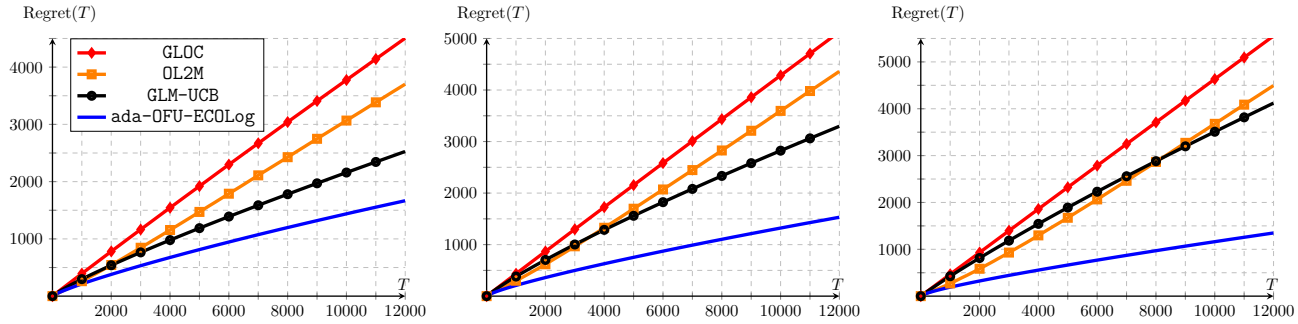
Replacing into the dual problem, one obtains $y = z(\lambda_\star)$ where $\lambda_\star$ solves the program:

$$\lambda_\star = \operatorname*{arg\,max}_{\lambda\geq 0} -\lambda - x^\mathsf{T}\mathbf{A}^{1/2}(\lambda\mathbf{I}_d + \mathbf{A})^{-1}\mathbf{A}^{1/2}x/2 \ .$$

□

G ADDITIONAL EXPERIMENTS

The results reported in Figure 2 complements Figure 1 from the main text, for **LogB** instances of higher dimension and varying values of κ. As promised by the regret bounds, the improvement brought by `ada-OFU-ECOLog` over its statistically sub-optimal predecessors increases as κ grows (*i.e* as the reward signal gets more non-linear). We did not evaluate the performances of `OFULog-r` in this setting - it is unfortunately too computationally demanding to complete in reasonable time.



$d = 10$, $\kappa = 50$ $d = 10$, $\kappa = 150$ $d = 10$, $\kappa = 400$.

Figure 2: Numerical simulations on **LogB** problems of dimensions $d = 10$ and varying value of κ. Regret curves are averaged over 100 independent trajectories, for fixed arm-sets of cardinality 200.