On the Complexity of Representation Learning in Contextual Linear Bandits

Andrea Tirinzoni
Meta

Matteo Pirotta
Meta

Alessandro Lazaric
Meta

Abstract

In contextual linear bandits, the reward function is assumed to be a linear combination of an unknown reward vector and a given embedding of context-arm pairs. In practice, the embedding is often learned at the same time as the reward vector, thus leading to an *online representation learning* problem. Existing approaches to representation learning in contextual bandits are either very generic (e.g., model-selection techniques or algorithms for learning with arbitrary function classes) or specialized to particular structures (e.g., nested features or representations with certain spectral properties). As a result, the understanding of the cost of representation learning in contextual linear bandit is still limited. In this paper, we take a systematic approach to the problem and provide a comprehensive study through an instance-dependent perspective. We show that representation learning is fundamentally more complex than linear bandits (i.e., learning with a given representation). In particular, learning with a given set of representations is never simpler than learning with the *worst* realizable representation in the set, while we show cases where it can be arbitrarily harder. We complement this result with an extensive discussion of how it relates to existing literature and we illustrate positive instances where representation learning is as complex as learning with a fixed representation and where sub-logarithmic regret is achievable.

1 INTRODUCTION

Stochastic contextual linear bandits (CLBs) focus on the interplay between exploration and exploitation when the reward $f^\star(x, a)$ of each context-arm pair $(x, a) \in \mathcal{X} \times \mathcal{A}$ is a

linear function of a known feature map $\phi^\star : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^{d_{\phi^\star}}$ and an unknown parameter $\theta^\star$. CLBs have been widely studied due to their broad applicability and strong theoretical guarantees (e.g., Lattimore and Szepesvári, 2020, and references therein). Unfortunately, the assumption that a realizable linear representation is known is often violated in real applications, where one only observes raw context-arm data and a suitable representation has to be learned online. Representation learning in CLBs relaxes this assumption by providing the learner with a set of representations $\Phi = \{\phi : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^{d_\phi}\}$ (e.g., a neural network) among which a realizable one exists (i.e., $\phi^\star \in \Phi$).

Representation learning can be viewed as a special case of learning with a general realizable function class (i.e., $\mathcal{F}_\Phi := \{f(\cdot, \cdot) = \phi(\cdot, \cdot)^\mathsf{T}\theta \mid \phi \in \Phi, \theta \in \mathbb{R}^{d_\phi}\}$), which has been extensively studied in the literature (e.g., Agarwal et al., 2014; Foster and Rakhlin, 2020; Simchi-Levi and Xu, 2020) with algorithms achieving $O(\sqrt{AT \log(|\mathcal{F}_\Phi|)})$ worst-case regret, where $|\mathcal{F}_\Phi|$ is the covering number of $\mathcal{F}_\Phi$. However, these algorithms do not explicitly leverage the bi-linear structure of the function class $\mathcal{F}_\Phi$. Another direction is to leverage model-selection techniques. While generic model-selection approaches (e.g. Abbasi-Yadkori et al., 2020; Pacchiano et al., 2020; Cutkosky et al., 2021) can be directly applied when Φ is finite, more specialized techniques can be used when Φ has additional structure (e.g., nested features (Foster et al., 2019)). Interestingly, some of these algorithms (e.g., Foster et al., 2019; Cutkosky et al., 2021; Ghosh et al., 2021) achieve regret guarantees matching the performance of the best representation in the set, up to a representation learning cost that depends on the number of representations $|\Phi|$, the problem horizon T, or other quantities specific to the structure of Φ. Nonetheless, these results are worst-case in nature and general model-selection algorithms are limited by an unavoidable $\Omega(\sqrt{T})$ regret (Pacchiano et al., 2020), which may hinder them from fully exploiting the structure of Φ and achieve instance-optimal performance (e.g., logarithmic regret). Alternatively, Papini et al. (2021) and Tirinzoni et al. (2022) proposed specialized representation learning algorithms that exploit the bi-linear structure of $\mathcal{F}_\Phi$ to obtain the *instance-dependent* regret bound of the best unknown realizable representation up to a logarithmic factor in $|\Phi|$. Furthermore, they showed that constant regret is achievable

(i.e., after a finite time τ the algorithm only plays optimal arms) when a realizable representation satisfies a certain spectral property. However, these results rely on the strong assumption that either all the representations in Φ are realizable or any misspecified representation can be identified by playing any sequence of arms.

In this paper, we focus on the following question:

What is the cost of representation learning compared to a CLB with a given representation?

In order to address this question, we first provide a systematic analysis of representation learning in CLBs through an instance-dependent lens. By specializing existing results, we derive an instance-dependent lower bound on the regret of any "good" representation learning algorithm which shows that the asymptotic regret must be at least $\mathcal{C}(f^\star, \mathcal{F}_\Phi)\log(T)$, where $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ is a complexity measure depending both on the reward function $f^\star$ and the given set of representations Φ. Moreover, this complexity is tight, as there exist algorithms attaining $\mathcal{C}(f^\star, \mathcal{F}_\Phi)\log(T)$ regret in the large T regime. This instance-dependent view allows us to have a more fine grained comparison to CLBs with a given representation, thus providing insights on the complexity of representation learning that may remain "hidden" in worst-case studies.

Leveraging this lower bound we are then able to derive the following results: **(1)** We show that the regret of representation learning is never smaller than the regret of learning with the *worst* realizable representation in the set, i.e., $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \geq \sup_{\phi \in \Phi, \text{realizable}} \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}})$. This reveals a fundamental limit to representation learning, showing that it is impossible to adapt to representations with better complexity. Surprisingly, this result holds even for instances $f^\star$ where all representations $\phi \in \Phi$ are realizable. Indeed, this is due to a subtle but crucial effect of representation learning: as in general all representations $\phi \in \Phi$ may be misspecified for some of the reward functions $f' \in \mathcal{F}_\Phi$, an algorithm needs to be robust to such misspecification and it cannnot fully adapt to cases that are favorable for some representations. **(2)** We further strengthen this result by showing examples where the inequality is strict and the gap arbitrarily large. In particular, we construct instances where all representations are realizable and have small dimensionality and yet the regret can be as large as learning with "tabular" features assigning a distinct dimension to each context-arm pair. **(3)** We characterize favorable instances where misspecified representations in Φ can be discarded without increasing the regret so that $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = \sup_{\phi \in \Phi, \text{realizable}} \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}})$.

Finally, we instantiate our analysis in widely studied representation structures (e.g., tabular, nested features, features with spectral properties, and the special case where all representations are realizable) and provide novel insights on the complexity of representation learning in these settings.

2 PRELIMINARIES

We consider a *stochastic contextual bandit* problem with a finite set of contexts $\mathcal{X}$ and a finite set of arms $\mathcal{A}$. Let $X := |\mathcal{X}|$ and $A := |\mathcal{A}|$. At each time step $t \in \mathbb{N}$, the learner first observes a context $x_t \in \mathcal{X}$ drawn i.i.d. from a distribution ρ[1], it selects an arm $a_t \in \mathcal{A}$, and it receives a scalar reward drawn from a Gaussian distribution with mean $f^\star(x_t, a_t)$ and unit variance.

Let Φ be a set of *representations*, where each $\phi \in \Phi$ is a d_ϕ-dimensional *feature map* $\phi : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^{d_\phi}$. We define the associated function class $\mathcal{F}_\Phi := \{f(\cdot, \cdot) = \phi(\cdot, \cdot)^\mathsf{T}\theta \mid \phi \in \Phi, \theta \in \mathbb{R}^{d_\phi}\}$. The set Φ and function class $\mathcal{F}_\Phi$ are realizable when:

Assumption 1 (Realizability). *There exist $\phi^\star \in \Phi$ and $\theta^\star \in \mathbb{R}^{d^\star}$, where $d^\star := d_{\phi^\star}$, such that*

$$f^\star(x, a) = \phi^\star(x, a)^\mathsf{T}\theta^\star \quad \forall x \in \mathcal{X}, a \in \mathcal{A}.$$

This assumption is required only for a representation $\phi^\star \in \Phi$ (which is said to be *realizable*), while, for any $\phi \neq \phi^\star$, the approximation error $\max_{x,a} |f^\star(x, a) - \phi(x, a)^\mathsf{T}\theta|$ may be non-zero for any θ, meaning that $f^\star$ cannot be approximated as a linear function of ϕ. In this case, we shall say that representation ϕ is *misspecified*.

Learning problem. We consider the problem of (bilinear) *representation learning* for regret minimization.

Definition 1 (Representation learning problem $(f^\star, \mathcal{F}_\Phi)$). *Consider an unknown stochastic contextual bandit problem with reward function $f^\star$. The learner is provided only with a set of representations Φ (equiv. function class $\mathcal{F}_\Phi$) satisfying Assumption 1 ($\phi^\star$ unknown) and it aims at minimizing the cumulative regret over T steps,*

$$R_T(f^\star) := \sum_{t=1}^{T} \left(\max_{a \in \mathcal{A}} f^\star(x_t, a) - f^\star(x_t, a_t) \right). \quad (1)$$

When $\Phi = \{\phi^\star\}$, the learning problem $(f^\star, \mathcal{F}_{\{\phi^\star\}})$ is known as stochastic *contextual linear bandit* (CLB), where the learner knows the realizable representation $\phi^\star$, while in representation learning the learner needs to learn within the realizable *non-linear* function class $\mathcal{F}_\Phi$. Note also that Φ may be an *infinite uncountable* set.

Notation We use $M^\dagger$ to denote the pseudo-inverse of a matrix $M \in \mathbb{R}^{n \times m}$, while $\mathrm{Im}(M)$ and $\mathrm{Ker}(M)$ denote its column and null spaces, respectively. For a vector $v \in \mathbb{R}^d$ and a matrix $M \in \mathbb{R}^{d \times d}$, we let $\|v\|_M^2 := v^\mathsf{T} M v$. We use $\pi^\star_{f^\star}(x) := \arg\max_{a \in \mathcal{A}} f^\star(x, a)$ to denote the optimal arm for context x when facing a problem with reward $f^\star$. We assume $\pi^\star_{f^\star}(x)$ to be unique for all x. We define the

---

[1]We assume ρ to be full-support over $\mathcal{X}$ w.l.o.g.

sub-optimality gap of arm $a \in \mathcal{A}$ for context $x \in \mathcal{X}$ as $\Delta_{f^\star}(x, a) := f^\star(x, \pi^\star_{f^\star}(x)) - f^\star(x, a)$. Note that, under Assumption 1, we have $\Delta_{f^\star}(x, a) = z^\star_{\phi^\star}(x, a)^\mathsf{T}\theta^\star$, where we call $z^\star_\phi(x, a) := \phi(x, \pi^\star_{f^\star}(x)) - \phi(x, a)$ the *feature gap*.

We will often use a matrix notation for all quantities. We denote by $f^\star \in \mathbb{R}^{XA}$ a vectorized reward function and by $D_\eta := \text{diag}(\{\eta(x, a)\}_{x \in \mathcal{X}, a \in \mathcal{A}})$ the $XA \times XA$ matrix containing a function $\eta : \mathcal{X} \times \mathcal{A} \to [0, \infty)$. For any $\phi \in \Phi$, let $F_\phi \in \mathbb{R}^{XA \times d_\phi}$ be the matrix containing the feature vectors $\{\phi(x, a)\}_{x \in \mathcal{X}, a \in \mathcal{A}}$ as rows and $V_\eta(\phi) := F_\phi^\mathsf{T} D_\eta F_\phi = \sum_{x,a} \eta(x, a)\phi(x, a)\phi(x, a)^\mathsf{T}$. Note that $f^\star = F_{\phi^\star}\theta^\star$.

Using this notation, $\|f^\star - F_\phi\theta\|^2_{D_\eta}$ is exactly the mean square error of the function $\phi(\cdot, \cdot)^\mathsf{T}\theta$ in predicting $f^\star$ when the learner has $\eta(x, a)$ samples from each (x, a). We define $\theta^\star_\eta(\phi) := \arg\min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi\theta\|^2_{D_\eta}$ as the best fit for the reward parameter using representation ϕ. By standard regression theory, it is easy to show that $\theta^\star_\eta(\phi) = V_\eta(\phi)^\dagger \sum_{x,a} \eta(x, a)\phi(x, a)f^\star(x, a)$. Similarly, the quantity $\|f^\star - F_\phi\theta^\star_\eta(\phi)\|^2_{D_\eta}$ is related to the *misspecification* of representation ϕ: it is zero for all η if ϕ is realizable, while it is positive for at least one η if ϕ is misspecified.

3 INSTANCE-DEPENDENT REGRET LOWER BOUND

We start by stating a novel asymptotic regret lower bound for the representation learning problem $(f^\star, \mathcal{F}_\Phi)$ (see Definition 1). Let $\mathfrak{A}$ be any bandit strategy, i.e., a sequence $\{\mathfrak{A}_t\}_{t \geq 1}$ where each $\mathfrak{A}_t : (\mathcal{X} \times \mathcal{A} \times \mathbb{R})^{t-1} \times \mathcal{X} \to \mathcal{A}$ is a measurable mapping w.r.t. the history up to time step $t-1$. We say that a $\mathfrak{A}$ is *uniformly good* on a function class $\mathcal{F}$ if $\mathbb{E}^{\mathfrak{A}}_f[R_T(f)] = o(T^\alpha)$ for any $\alpha > 0$ and any $f \in \mathcal{F}$[2], where $\mathbb{E}^{\mathfrak{A}}_f$ denotes the expectation under algorithm $\mathfrak{A}$ in a contextual bandit problem with reward function $f \in \mathcal{F}$.

Theorem 1. *Let $\mathfrak{A}$ be a uniformly good strategy on the class $\mathcal{F}_\Phi$ and suppose that $\pi^\star_{f^\star}$ is unique. Then,*

$$\liminf_{T \to \infty} \frac{\mathbb{E}^{\mathfrak{A}}_{f^\star}[R_T(f^\star)]}{\log(T)} \geq \mathcal{C}(f^\star, \mathcal{F}_\Phi),$$

where $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ is the value of the optimization problem

$$\inf_{\{\eta(x,a)\} \geq 0} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x, a)\Delta_{f^\star}(x, a) \quad \text{s.t.}$$

$$\inf_{\phi \in \Phi} \min_{x, a \neq \pi^\star_{f^\star}(x)} \left(\|f^\star - F_\phi\theta^\star_\eta(\phi)\|^2_{D_\eta} + c^\eta_{x,a}(f^\star, \phi) \right) \geq 2,$$

---

[2]Our analysis easily extends to the weaker notion of uniformly good algorithm requiring $O(T^\alpha)$ regret on all $f \in \mathcal{F}$ only for some $\alpha \in (0, 1)$. In this case, the stated lower bound remains the same as in Theorem 1 up to a factor $1 - \alpha$ (Tirinzoni et al., 2021).

with

$$c^\eta_{x,a}(f^\star, \phi) = \begin{cases} 0 & \text{if } z^\star_\phi(x, a)^\mathsf{T}\theta^\star_\eta(\phi) \leq 0, \\ 0 & \text{if } z^\star_\phi(x, a) \notin \text{Im}(V_\eta(\phi)), \\ \frac{(z^\star_\phi(x,a)^\mathsf{T}\theta^\star_\eta(\phi))^2}{\|z^\star_\phi(x,a)\|^2_{V_\eta(\phi)^\dagger}} & \text{otherwise.} \end{cases}$$

The proof (see Appendix B) builds on the asymptotic regret lower bound for contextual bandits with general function classes (a.k.a. structured bandits), which can be extracted as a special case of the one for Markov decision processes (Ok et al., 2018). While Ok et al. (2018) provide an implicit complexity measure $\mathcal{C}(f, \mathcal{F})$ for learning any instance f when knowing that it belongs to a given class $\mathcal{F}$, we derive a more explicit complexity measure $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ for representation learning. The general lower bound follows from a fundamental result stating that any uniformly good algorithm must guarantee $\sum_{t=1}^T \mathbb{E}^{\mathfrak{A}}_{f^\star}[(f^\star(x_t, a_t) - f(x_t, a_t))^2] \geq 2\log(T)$ as $T \to \infty$ for any *alternative* reward $f \in \mathcal{F}$ that induces a different optimal policy than $\pi^\star_{f^\star}$. Our explicit complexity follows by leveraging a novel reformulation of the set of such *alternative* rewards for representation learning which allows us to derive a closed-form expression of the above general condition.

As common in existing instance-dependent lower bounds (e.g., Combes et al., 2017; Ok et al., 2018), the complexity $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ is the value of an optimization problem which seeks an allocation of samples η minimizing the regret while collecting sufficient information about the instance $f^\star$. Such an information constraint is the peculiar component in our setting as it formally establishes the minimal level of exploration that any uniformly good representation learning algorithm must guarantee. In particular, for any representation $\phi \in \Phi$, context $x \in \mathcal{X}$, and sub-optimal action $a \neq \pi^\star_{f^\star}(x)$, any feasible allocation η must guarantee

$$\underbrace{\|f^\star - F_\phi\theta^\star_\eta(\phi)\|^2_{D_\eta}}_{\text{misspecification}} + \underbrace{c^\eta_{x,a}(f^\star, \phi)}_{\text{sub-optimality}} \geq 2. \tag{2}$$

Here we recognize the contribution of two terms. The first one is related to the misspecification error of representation ϕ induced by η (i.e., the minimum achievable mean square error when linearly estimating $f^\star$ with ϕ using samples collected according to η). It is trivially zero for any η if ϕ is realizable. The second term is related to the complexity for learning that a is a sub-optimal action for context x when using representation ϕ to estimate the reward. Interestingly, $c^\eta_{x,a}(f^\star, \phi)$ resembles the complexity term appearing in the existing lower bound for a CLB problem with given representation ϕ (e.g., Hao et al., 2020; Tirinzoni et al., 2020).

The constraint requires the sum of these two terms to be large. This means that any feasible allocation η, and thus *any uniformly good representation learning algorithm, must either learn that ϕ is misspecified or that a is sub-optimal in context x under the best fit of $f^\star$ with represen-*

tation ϕ. We now discuss relevant possible cases to better undestand the complexity for achieving so.

Case 1. ϕ **is realizable and** $z_\phi^\star(x, a) \in \text{Im}(V_\eta(\phi))$**.** In this case, the misspecification term in (2) is zero for any η and $z_\phi^\star(x, a)^\mathsf{T} \theta_\eta^\star(\phi) = \Delta_{f^\star}(x, a) > 0$ by realizability, definition of $z_\phi^\star$, and sub-optimality of a. From (2), this implies that η must guarantee that $\|z_\phi^\star(x, a)\|_{V_\eta(\phi)^\dagger}^2 \leq \Delta_{f^\star}(x, a)^2/2$. It turns out that this is exactly the same complexity measure we have for learning that (x, a) is sub-optimal in the CLB $(f^\star, \mathcal{F}_{\{\phi\}})$. Since $\|z_\phi^\star(x, a)\|_{V_\eta(\phi)^\dagger}$ represents the uncertainty that allocation η has on the rewards of (x, a) and $(x, \pi_{f^\star}^\star(x))$, this condition simply requires any uniformly good algorithm to reduce such uncertainty below a factor of the gap of (x, a).

Case 2. ϕ **is realizable and** $z_\phi^\star(x, a) \notin \text{Im}(V_\eta(\phi))$**.** In this case, both the misspecification term and $c_{x,a}^\eta(f^\star, \phi)$ are zero. From (2), this means that η is infeasible. This is intuitive since, when the feature gap $z_\phi^\star(x, a)$ is not in the column space of the design matrix $V_\eta(\phi)$, the allocation η does not provide any information about arm a in the representation space of ϕ, and thus it cannot learn whether a is sub-optimal or not. Therefore, any feasible η must guarantee $z_\phi^\star(x, a) \in \text{Im}(V_\eta(\phi))$ for all (x, a) when ϕ is realizable, i.e., any good algorithm must explore all feature directions. This has an interesting implication: when $\text{span}(\{\phi(x, a)\}_{x,a}) = d_\phi$, any feasible design matrix must be invertible. This result was already proved by Lattimore and Szepesvári (2017) in the linear bandit setting using an ad-hoc derivation, while here we establish it in greater generality as a consequence of our lower bound.

Case 3. ϕ **is misspecified and** $c_{x,a}^\eta(f^\star, \phi) = 0$**.** This can happen in two cases: either $z_\phi^\star(x, a)^\mathsf{T} \theta_\eta^\star(\phi) \leq 0$, which means that the sub-optimality gap of (x, a) cannot be accurately estimated using representation ϕ, or $z_\phi^\star(x, a) \notin \text{Im}(V_\eta(\phi))$. In both cases, a feasible η must make the first term in (2) large, i.e., it must learn that ϕ is misspecified. Interestingly, this implies that, differently from the realizable case, a feasible allocation does not need to explore the whole feature space for ϕ (e.g., it does not have to make the design matrix $V_\eta(\phi)$ invertible). This is particularly relevant when ϕ is high-dimensional, as identifying the misspecification may be easier than covering all dimensions.

Case 4. ϕ **is misspecified and** $c_{x,a}^\eta(f^\star, \phi) > 0$**.** This is the case with most freedom: a feasible allocation can either learn that ϕ is misspecified or that (x, a) is sub-optimal. As we shall see in Section 4.3, this flexibility may be exploited to find allocations that manage to "discard" representations without significantly affecting the regret.

3.1 Known-Representation Case

As expected, when instantiating Theorem 1 in the standard CLB $(f^\star, \mathcal{F}_{\{\phi^\star\}})$, we recover the existing lower bound for

such a setting (Hao et al., 2020; Tirinzoni et al., 2020).[3]

Corollary 1. *Let* $\text{span}(\{\phi^\star(x, a)\}_{x,a}) = d^\star$*. In the CLB* $(f^\star, \mathcal{F}_{\{\phi^\star\}})$*, the complexity* $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$ *of Theorem 1 is*

$$\inf_{\eta: V_\eta(\phi^\star)^{-1} \text{exists}} \sum_{x,a} \eta(x, a) \Delta_{f^\star}(x, a)$$

$$\text{s.t.} \quad \min_{x, a \neq \pi_{f^\star}^\star(x)} \frac{\Delta_{f^\star}(x, a)^2}{\|z_{\phi^\star}(x, a)\|_{V_\eta(\phi^\star)^{-1}}^2} \geq 2.$$

Comparing this result with Theorem 1, we notice that adding one representation to the set Φ implies adding one constraint to the optimization problem, hence making the problem harder. On the positive side, Theorem 1 does not impose the strong constraint of Corollary 1 for every $\phi \in \Phi$, which would require any good algorithm to learn an optimal action at every context for all representations. In fact, it may be possible to leverage the misspecification of a representation ϕ to lower the additional complexity w.r.t. the one imposed in the realizable case (see Equation 2 and, e.g., Case 4 above). In Section 4, we further elaborate on how the complexity of representation learning is impacted by these elements and how it compares with the complexity of CLBs when given a realizable representation.

3.2 The Lower Bound Is Attainable

It is known that instance-dependent lower bounds in the general form of Ok et al. (2018) can be attained. Since Theorem 1 is an instantiation of such a result, this implies that $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ is a *tight* complexity measure for representation learning as there exist algorithms matching it.

Proposition 1. *There exists an algorithm* $\mathfrak{A}$ *(e.g., Dong and Ma, 2022) such that, for any representation learning problem* $(f^\star, \mathcal{F}_\Phi)$*,*

$$\limsup_{T \to \infty} \frac{\mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)]}{\log(T)} \leq \mathcal{C}(f^\star, \mathcal{F}_\Phi).$$

While, to the best of our knowledge, the algorithm of Dong and Ma (2022) is the only one attaining instance-optimal complexity in contextual bandits with general function classes, it is actually easy to adapt existing strategies for non-contextual bandits to our setting (Combes et al., 2017; Degenne et al., 2020; Jun and Zhang, 2020). In particular, the algorithm of Jun and Zhang (2020) would obtain an *anytime* regret of order $O(\mathcal{C}(f^\star, \mathcal{F}_\Phi) \log(T) + \log\log(T))$. This shows that $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ is also a relevant finite-time complexity measure (and not only asymptotic), up to a $O(\log\log(T))$ term depending on other instance-dependent factors.

---

[3]Existing lower bounds are derived under the assumption that the full set of features $\{\phi^\star(x, a)\}_{x,a}$ span $\mathbb{R}^{d^\star}$. This is without loss of generality since one can always remove redundant features by computing the low-rank SVD of $F_{\phi^\star}$.

4 COMPLEXITY OF REPRESENTATION LEARNING

We now provide a series of results to better characterize the instance-dependenet complexity $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ of representation learning in comparison with the complexity $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$ of the single-representation CLB problem.

4.1 Representation Learning Cannot Be Easier Than Learning With A Given Representation

We first prove that the complexity of learning with a single representation is a lower bound for representation learning.

Proposition 2. *For any Φ such that $f^\star \in \mathcal{F}_\Phi$, $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \geq \sup_{\phi \in \Phi : f^\star \in \mathcal{F}_{\{\phi\}}} \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}})$.*

This result leverages the instance-dependent nature of the complexity derived in Theorem 1 to compare representation learning with a single-representation CLB for *every reward function $f^\star$*. This is in contrast with a worst-case analysis, where we would compare the two approaches w.r.t. their respective worst-case reward functions.

Whenever there is only one realizable representation $\phi^\star$ in Φ, the result is intuitive, since adding misspecified representations to Φ cannot make the problem any easier. Nonetheless, Proposition 2 has another, less obvious, implication: representation learning is at least as hard as the *hardest* CLB ($f^\star, \mathcal{F}_{\{\phi\}}$) among all realizable representations. More surprisingly, this result holds even when all the representations in Φ are realizable for $f^\star$. In fact, this is the unavoidable price for an algorithm to be *robust* (i.e., uniformly good) to any other reward function $f' \in \mathcal{F}_\Phi$ for which some representation ϕ may not be realizable and it defines an intrinsic limit to the level of adaptivity to $f^\star$ that we can expect in representation learning (see Section 5 for a discussion on how this result relates to existing literature).

4.2 There Exist Instances Where Representation Learning Is Strictly Harder Than Learning With A Given Representation

After establishing that representation learning cannot be easier than CLBs, a natural question is: how much harder can it be? Here we show that, for any reward function $f^\star$, there exists a set of representations Φ with $f^\star \in \mathcal{F}_\Phi$ such that any uniformly good representation learning algorithm must suffer regret scaling linearly with the number of contexts and actions, whereas the regret of learning with any realizable representation in the set only scales with the feature dimensionality $d \ll XA$.

Proposition 3. *Let $X, A \geq 1$ and $2 \leq d \leq XA$. Fix an arbitrary instance $f^\star : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ and denote by $\Delta_{\min}$ its minimum positive gap. Then, there exists a set of d-dimensional representations Φ of cardinality $|\Phi| =$*

$\lceil \frac{X(A-1)}{d-1} \rceil$ *such that $f^\star \in \cap_{\phi \in \Phi} \mathcal{F}_{\{\phi\}}$ and*

$$\mathcal{C}(f^\star, \mathcal{F}_\Phi) = \sum_{x \in \mathcal{X}} \sum_{a \neq \pi^\star_{f^\star}(x)} \frac{2}{\Delta_{f^\star}(x, a)}.$$

Moreover, for any $\phi \in \Phi$,

$$\mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}}) \leq \frac{2(d-1)}{\Delta_{\min}}.$$

Note that the complexity $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ of the representation learning problem built in Proposition 3 is exactly the complexity for learning the contextual bandit problem $f^\star$ when ignoring the set of representations Φ, i.e., the (unstructured) tabular setting[4]. Therefore, Proposition 3 proves that there exist "hard" representation learning problems whose complexity is the same as learning without any prior knowledge about $f^\star$. While this may be expected as the set Φ is constructed to be worst-case for $f^\star$, the second statement of Proposition 3 is more surprising. In fact, Φ is constructed using only realizable representations for $f^\star$ with dimension $d \ll XA$. As such, the complexity $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}})$ for learning with any $\phi \in \Phi$ only scales (in the worst case) with d and it can be arbitrarily smaller than $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$.

Remark 1. *Rather than constructing a single hard instance $(f^\star, \mathcal{F}_\Phi)$ where representation learning is difficult, we prove that for any reward function $f^\star$ we can find a set Φ such that $(f^\star, \mathcal{F}_\Phi)$ is difficult. Hence, representation learning can be difficult regardless of the reward function.*

4.3 There Exist Instances Where Representation Learning Is Not Harder Than Learning With A Given Representation

Unlike the previous hardness results, here we show that there exist favorable instances $(f^\star, \mathcal{F}_\Phi)$ where the complexity of representation learning is the same as the one of a CLB with a realizable representation in Φ. This means that representation learning comes "for free" on such instances.

Proposition 4. *Let $\eta^\star(x, a) = \mathbb{1}\left(a = \pi^\star_{f^\star}(x)\right)$. Let Φ contain a unique realizable representation $\phi^\star$ and suppose that there exists $\varepsilon > 0$ such that, for all $\phi \in \Phi$ with $f^\star \notin \mathcal{F}_{\{\phi\}}$, $\|f^\star - F_\phi \theta^\star_{\eta^\star}(\phi)\|^2_{D_{\eta^\star}} \geq \varepsilon$. Then, $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = \mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$.*

Intuitively, the condition on Φ in Proposition 4 requires every misspecified representation to have a minimum positive mean square error in fitting $f^\star$ when samples are collected by an optimal policy. This means that a learner is able to

---

[4]Since the context-action space is finite, we can always run a trivial variant of UCB (Auer et al., 2002a) that estimates the reward of each (x, a) independently and achieve regret $\mathbb{E}_{f^\star}[R_T(f^\star)] \lesssim \sum_{x \in \mathcal{X}} \sum_{a \neq \pi^\star_{f^\star}(x)} \frac{\log(T)}{\Delta_{f^\star}(x,a)}$. This is also the instance-optimal rate of the unstructured setting (Ok et al., 2018).

detect all misspecified representation by playing optimal actions, i.e., while suffering zero regret, hence making representation learning costless in the long run.

Consider the following scheme as an example of how a simple strategy can leverage the condition Proposition 4. Assuming finite Φ, take any algorithm with sub-linear regret on the class $\mathcal{F}_\Phi$, e.g., any of the algorithms for general function classes (e.g., Foster and Rakhlin, 2020; Simchi-Levi and Xu, 2020). Run the algorithm in combination with an elimination rule for misspecified representations (e.g., the one proposed by Tirinzoni et al. (2022)) until only one representation remains active. When this happens, switch to playing an instance-optimal algorithm for CLBs on the remaining representation. It is easy to see that, since the starting algorithm has sub-linear regret, it plays optimal actions linearly often and, thus, thanks to the assumption in Proposition 4, it collects sufficient information to eliminate all misspecified representations in a finite time. This means that the algorithm suffers only constant regret for eliminating misspecified representations, while it never discards $\phi^\star$ with high probability. After that, playing an instance-optimal strategy on $\phi^\star$ implies that the total regret is roughly $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}}) \log(T)$ in the long run, which is exactly the same regret we would have by running the instance optimal algorithm on $\phi^\star$ from the very beginning.

5 SPECIFIC REPRESENTATION STRUCTURES

We now discuss some of the specific representation learning problems studied in the literature, while providing additional insights on their instance-dependent complexity.

5.1 Trivial Representations

It is well known that the realizability of $\mathcal{F}_\Phi$ (Assumption 1) is crucial for efficient learning, as sub-linear regret may be impossible otherwise (Lattimore et al., 2020). In practice, when little prior knowledge about the reward function $f^\star$ is available to design a suitable class $\mathcal{F}_\Phi$, a common technique is to reduce the approximation error by expanding Φ, in the hope of ensuring realizability.

When the context-arm pairs are finite, a trivial realizable representation can always be constructed as the canonical basis of $\mathbb{R}^{XA}$. Let $\{(x_i, a_i)\}_{i=1}^N$ be an enumeration of all $N = XA$ context-arm pairs. Then, we can define the XA-dimensional features $\bar{\phi}$ as $\bar{\phi}_i(x, a) := \mathbb{1}(x = x_i, a = a_i)$. It is easy to see that $f^\star = F_{\bar{\phi}}\theta$ for $\theta_i = f^\star(x_i, a_i)$.

A natural idea to build a class for representation learning is thus to start from a set Φ of "good" features (e.g., low dimensional or with nice spectral properties) and then add the trivial representation $\bar{\phi}$ so as to ensure realizability. The hope is that a good algorithm would still be able to leverage

the "good" representations to achieve better results whenever possible. The following result shows that this is *impossible*: every uniformly good algorithm must pay the complexity of learning without any prior knowledge on $f^\star$ as far as $\bar{\phi}$ is in the set of candidate representations.

Proposition 5. *Let Φ be any set of representations (not necessarily realizable for $f^\star$). Then,*

$$\mathcal{C}(f^\star, \mathcal{F}_{\Phi \cup \{\bar{\phi}\}}) = \sum_{x \in \mathcal{X}} \sum_{a \neq \pi_{f^\star}^\star(x)} \frac{2}{\Delta_{f^\star}(x, a)}.$$

As already noted in the discussion of Proposition 3, the complexity $\mathcal{C}(f^\star, \mathcal{F}_{\Phi \cup \{\bar{\phi}\}})$ of Proposition 5 is equivalent to the complexity of learning $f^\star$ without any prior knowledge. Hence, no uniformly good algorithm can leverage the representations in Φ when $\bar{\phi}$ is also considered, no matter how good they are. For instance, the set Φ could even be a singleton $\{\phi^\star\}$ containing a realizable representation of dimension $d \ll XA$, and still representation learning over the set $\{\phi^\star, \bar{\phi}\}$ so as to achieve regret scaling with the properties of $\phi^\star$ is impossible. A similar result was derived by Réda et al. (2021), who showed that learning an instance $f^\star$ which is known to be approximately linear in given features ϕ without knowing the amount of misspecification is as complex as learning $f^\star$ without any prior knowledge.

5.2 Nested Features

A popular design choice for representation learning is to be build a set of *nested features* (Foster et al., 2019; Pacchiano et al., 2020; Cutkosky et al., 2021; Ghosh et al., 2021) $\Phi = \{\phi_1, \ldots, \phi_N\}$ of increasing dimension (i.e., such that $d_i := d_{\phi_i} < d_{i+1} := d_{\phi_{i+1}}$ for all $i \in [N-1]$) that satisfy the following property: for all $i \in [N-1]$ and (x, a), the first d_i components of $\phi_{i+1}(x, a)$ are equal to $\phi_i(x, a)$. Let $i^\star \in [N]$ be such that $\phi_{i^\star}$ is the realizable representation of smallest dimension (which exists by Assumption 1). The nestedness implies that ϕ_i is realizable for all $i \geq i^\star$.

Several approaches have been proposed for this setting. While Foster et al. (2019) designed a strategy with regret $\widetilde{O}(\sqrt{d_{i^\star} T} + T^{3/4})$, model-selection algorithms (e.g., Pacchiano et al., 2020; Cutkosky et al., 2021) achieve regret of order $\widetilde{O}(poly(N)\sqrt{d_{i^\star} T})$. Interestingly, Ghosh et al. (2021) obtained $\widetilde{O}(\sqrt{d_{i^\star} T})$ regret, that is of the same order as the worst-case regret achievable by, e.g., LinUCB on the (unknown) smallest realizable representation $\phi_{i^\star}$.

We show that things are considerably more complex from an instance-dependent perspective.

Proposition 6. *Let Φ be a set of N nested features and $f^\star \in \mathcal{F}_\Phi$. Then,*

$$\mathcal{C}(f^\star, \mathcal{F}_\Phi) = \mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}}). \tag{3}$$

This result claims that representation learning on a set of nested features Φ is as difficult as a CLB problem with the

realizable representation of largest dimension (ϕ_N). This is somehow surprising since it essentially states that representation learning over nested features is useless, and one may simply learn with the highest dimensional representation (which is known to be realizable) from the very beginning. The intuition is that, while the learner knows ϕ_N to be realizable for any reward function (by assumption), it does not know whether this is true for ϕ_{N-1}, ϕ_{N-2}, etc. Even if, say, ϕ_{N-1} is realizable for $f^\star$, there might be another reward function f' where this is not true. Any good algorithm must explore sufficiently to eventually discriminate between $f^\star$ and f' in the long run, and it turns out that the complexity for doing so is exactly $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}})$, hence making any finer level of adaptivity impossible. Moreover, we prove in Appendix D that there exist problems with $d_{i^\star} \ll d_N$ where $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}}) \gtrsim d_N$ but $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_{i^\star}\}}) \lesssim d_{i^\star}$. This implies that, in the worst-case, any uniformly good algorithm must suffer a dependence on the dimensionality of the largest representation, regardless of the fact that a smaller realizable representation is nested into it.

Note that this does not contradict existing results for model-selection (e.g., Foster et al., 2019; Ghosh et al., 2021; Cutkosky et al., 2021). In fact, while they achieve a dependence on the worst-case regret of the best representation $\phi_{i^\star}$, they also feature some representation learning cost which dominates in the long run. For instance, the bound of Ghosh et al. (2021) has a $O(d_N^2 \log(T))$ additive term. Therefore, while some gains are possible in the small T regime (e.g., scaling with $\sqrt{d_{i^\star}T}$ instead of $\sqrt{d_N T}$), in the long run the logarithmic term dominates and $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}}) \log(T)$ becomes the optimal complexity.

5.3 Hls Representations And Sub-Logarithmic Regret

Hao et al. (2020) and Papini et al. (2021) recently showed that in a CLB $(f^\star, \mathcal{F}_{\{\phi\}})$ it is possible to achieve constant regret when the given realizable representation ϕ satisfies a certain spectral condition.

Definition 2 (HLS representation). *A representation ϕ is HLS for an instance $f^\star$ if, for all $x \in \mathcal{X}$ and $a \neq \pi^\star_{f^\star}(x)$, $\phi(x, a) \in \mathrm{span}(\{\phi(x, \pi^\star_{f^\star}(x))\}_{x \in \mathcal{X}})$.*[5]

Intuitively, a representation satisfying this property allows exploring the full feature space by playing an optimal policy. That is, playing optimal actions allows refining the reward estimates at all (x, a), even those that are not played. Interestingly, Papini et al. (2021) showed that this condition is both necessary and sufficient to achieve constant regret.

Theorem 2 (Papini et al. (2021)). *Constant regret is*

---

[5]The original definition (Hao et al., 2020) requires the stronger condition $\mathrm{span}(\{\phi(x, \pi^\star_{f^\star}(x))\}_{x \in \mathcal{X}}) = \mathbb{R}^{d_\phi}$. This is because the authors assumed that $\mathrm{span}(\{\phi(x, a)\}_{x,a}) = \mathbb{R}^{d_\phi}$. Here we state a generalization that works even without such an assumption.

achievable on an instance $f^\star$ if, and only if, the learner is provided with a HLS realizable representation $\phi^\star$.

When a realizable HLS representation $\phi^\star$ is not known a-priori and one must perform representation learning, it is natural to ask whether such a strong result can still be achieved. Papini et al. (2021); Tirinzoni et al. (2022) showed that this is indeed the case under strong conditions on Φ: either **1)** all the representation are realizable or **2)** misspecified representations are detectable by any policy (i.e., such that $\min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi \theta\|^2_{D_\eta} > 0$ for any η).

We now provide a necessary and sufficient condition on the representations Φ to allow $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$. This implies that, whenever such a condition is not met, any form of sub-logarithmic regret (e.g., constant) is *impossible* for *any* uniformly good algorithm. On the other hand, when the condition is met, sub-logarithmic regret is achievable (and it is achieved by the algorithm mentioned in Section 3.2).[6]

Proposition 7. *A necessary and sufficient condition for $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$ is that the following two properties hold for any $\phi \in \Phi$ such that $\min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi \theta\|^2_{D_{\eta^\star}} = 0$ and for all $x \in \mathcal{X}, a \neq \pi^\star_{f^\star}(x)$:*

1. $z^\star_\phi(x, a)^\mathsf{T} \theta^\star_{\eta^\star}(\phi) > 0$;
2. $\phi(x, a) \in \mathrm{Im}(V_{\eta^\star}(\phi))$.

Proposition 7 can be read as follows. Any representation ϕ whose misspecification is detectable by an optimal policy (i.e., such that $\min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi \theta\|^2_{D_{\eta^\star}} > 0$) does not bring any contribution to the regret lower bound, as already noted in Section 4.3. For any other representation ϕ, the optimal policy must be able to detect that all sub-optimal pairs (x, a) of $f^\star$ are indeed sub-optimal. This, in turns, requires $\phi(x, a)$ to be in the span of $V_{\eta^\star}(\phi)$ (i.e., the optimal policy explores the direction $\phi(x, a)$) and that $z^\star_\phi(x, a)^\mathsf{T} \theta^\star_{\eta^\star}(\phi) > 0$ (i.e., the best approximation to the gap of (x, a) using ϕ remains strictly positive). On the other hand, suppose that, for some $\phi \in \Phi$ with zero misspecification under an optimal policy, one of the two properties in Proposition 7 does not hold. Then, if $z^\star_\phi(x, a)^\mathsf{T} \theta^\star_{\eta^\star}(\phi) \leq 0$, (x, a) has higher reward than $(x, \pi^\star_{f^\star}(x))$ in the linear instance $(\phi, \theta^\star_{\eta^\star}(\phi))$, which means that it is impossible to learn its sub-optimality. Similarly, if $\phi(x, a) \notin \mathrm{Im}(V_{\eta^\star}(\phi))$, an optimal policy does not explore the direction $\phi(x, a)$ at all, which means that it cannot estimate the corresponding reward. In both cases, it is necessary to repeatedly play at least some sub-optimal action, which implies that $\mathcal{C}(f^\star, \mathcal{F}_\Phi) > 0$ and sub-logarithmic is thus impossible.

Perhaps surprisingly, an immediate consequence of Proposition 7 is that sub-logarithmic regret is *impossible* if Φ contains at least one *realizable* non-HLS representation.

---

[6]The best algorithm mentioned in Section 3.2 achieves $O(\log \log(T))$ regret when Φ satisfies Proposition 7. How to achieve constant regret in this setting remains an open question.

Corollary 2. *If there exists a realizable representation* $\phi \in \Phi$ *which does not satisfy the HLS condition (Definition 2),* $\mathcal{C}(f^\star, \mathcal{F}_\Phi) > 0$ *(i.e., sub-logarithmic regret is impossible).*

This implies that, even when Φ contains *only* realizable representations and all but one are HLS, constant regret cannot be attained by any uniformly good algorithm.

5.4 Fully-Realizable Representations

Another specific structure is when *all* representantions in Φ are realizable for all reward functions of interest. Papini et al. (2021) proved that, in this case, a LinUCB-based algorithm can adapt to the *best* instance-dependent regret bound of a representation in Φ (e.g., it achieves constant regret when at least one representation is HLS). While our Proposition 2 and Corollary 2 seem to contradict their result, it turns out that Papini et al. (2021) consider a simpler problem: they assume the learner to be aware of Φ containing only realizable representations, while we consider the more general setting where it only knows that one of them is realizable. Intuitively, when the learner has such a strong prior knowledge, specialized strategies can be designed to achieve better results. However, such strategies would not be uniformly good on all problems in our class $\mathcal{F}_\Phi$ as they do not account for misspecified representations. Therefore, the price to pay for being robust to misspecification is in general very large, no matter how "good" Φ is.

As a complement to the results of Papini et al. (2021), in Appendix F we show that the instance-optimal complexity of representation learning with prior knowledge about full realizability is indeed never larger than the instance-optimal complexity of a CLB with *any* representation in Φ, while there even exist cases where the former complexity is significantly smaller. This makes the problem of fully-realizable representation learning statistically "easier" than CLBs, as opposed to our general setting (Definition 1).

6 GENERAL FUNCTIONS AND WORST-CASE REGRET

While the instance-dependent viewpoint we considered so far allowed us to provide sharp insights on the complexity of representation learning, it is still asymptotic in nature and may "hide" other phenomena happening in the finite-time regime. Existing algorithms for general function classes (e.g., Foster and Rakhlin, 2020; Simchi-Levi and Xu, 2020) achieve $O(\sqrt{AT \log(|\mathcal{F}|)})$ regret when given an arbitrary class $\mathcal{F}$. This is known to be optimal in the worst possible choice of $\mathcal{F}$ (Agarwal et al., 2012). When applied to representation learning with finite $|\Phi|$, i.e., to learn any instance $f^\star \in \mathcal{F}_\Phi$ given $\mathcal{F}_\Phi$, their regret bound reduces to $O(\sqrt{AT(\log(|\Phi|) + d)})$ and it is an open question whether this is optimal in the worst possible set Φ. A similar result was obtained by Moradipari et al. (2022).

In particular, one might be wondering whether a polynomial dependence on the number of actions A is really unavoidable even when the learner is provided with a set of d-dimensional representations with $d \ll A$. The question arises mostly because some model-selection algorithms (Cutkosky et al., 2021) achieve $O(\sqrt{dT \log(A)})$ regret on this problem, with some extra dependences on other problem-independent variables, like $|\Phi|$ or T. Such bounds give hope that adapting to the worst-case complexity of a CLB with one of the realizable representations in Φ may be possible at least in the small T regime. Once again, we show that this is impossible. We state a worst-case lower bound for representation learning proving that a polynomial depedence on the number of actions is unavoidable.

Theorem 3. *Let $N \geq 1$, $A \geq 4$ and $d \geq 12 \log_2(A)$. There exists a context distribution, a set of d-dimensional representations Φ of size $|\Phi| = N$ over A arms, and a universal constant $c > 0$ such that, for any learning algorithm $\mathfrak{A}$ and $T \geq \max\{\lfloor \log(dN)/\log(A) \rfloor, d/\log_2(A)\}$,*

$$\sup_{f \in \mathcal{F}_\Phi} \mathbb{E}_f^{\mathfrak{A}}[R_T(f)] \geq c\sqrt{T\left(d\log_2(A) + A\left\lfloor \frac{\log(dN)}{\log(A)} \right\rfloor\right)}.$$

The proof of this result combines techniques used to derive two existing lower bounds: the $\Omega(\sqrt{dT \log(A)})$ lower bound for CLB problems of He et al. (2022) and the $\Omega(\sqrt{AT \log(|\mathcal{F}|)})$ lower bound for general function classes of Agarwal et al. (2012). Differently from existing upper bounds that scale as $\Omega(\sqrt{Ad})$, our lower bound decouples the polynomial dependencies on A and d. Whether this is matchable by a specialized algorithm for representation learning, or whether existing algorithm for general function classes are already worst-case optimal in our setting, remains an intriguing open question.

7 DISCUSSION

Our main contributions can be summarized in two fundamental hardness results. **1)** Through an instance-dependent lens, representation learning is never easier than a CLB with a given realizable representation, while the former problem can be strictly harder, up to the point that knowing that one of some given low-dimensional representations is realizable is useless. **2)** Adaptivity to the best representation is impossible in general, both in the instance-dependent long-horizon and in the worst-case small-horizon regimes. In particular, as opposed to worst-case results, instance-dependent adaptivity is impossible for representation learning on nested features, and the same holds when all representations are realizable if the learner does not know it a-priori. On the positive side, we characterized "simple" instances where representation learning is not harder than a CLB and where sub-logarithmic regret can be achieved.

Following literature on linear bandits (Tirinzoni et al., 2020; Kirschner et al., 2021), an interesting open question is how to design computationally-efficient representation learning strategies with good (e.g., worst-case optimal) finite-time regret and asymptotically instance-optimal performance.

References

Yasin Abbasi-Yadkori, Aldo Pacchiano, and My Phan. Regret balancing for bandit and RL model selection. *CoRR*, abs/2006.05491, 2020.

Alekh Agarwal, Miroslav Dudík, Satyen Kale, John Langford, and Robert Schapire. Contextual bandit learning with predictable rewards. In *Artificial Intelligence and Statistics*, pages 19–26. PMLR, 2012.

Alekh Agarwal, Daniel J. Hsu, Satyen Kale, John Langford, Lihong Li, and Robert E. Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. In *ICML*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 1638–1646. JMLR.org, 2014.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002a.

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002b.

Richard Combes, Stefan Magureanu, and Alexandre Proutière. Minimal exploration in structured stochastic bandits. In *NIPS*, pages 1763–1771, 2017.

Ashok Cutkosky, Christoph Dann, Abhimanyu Das, Claudio Gentile, Aldo Pacchiano, and Manish Purohit. Dynamic balancing for model selection in bandits and RL. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 2276–2285. PMLR, 2021.

Rémy Degenne, Han Shao, and Wouter Koolen. Structure adaptive algorithms for stochastic bandits. In *International Conference on Machine Learning*, Vienna, Austria, 2020. Virtual conference.

Kefan Dong and Tengyu Ma. Asymptotic instance-optimal algorithms for interactive decision making. *arXiv preprint arXiv:2206.02326*, 2022.

Dylan J. Foster and Alexander Rakhlin. Beyond UCB: optimal and efficient contextual bandits with regression oracles. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 3199–3210. PMLR, 2020.

Dylan J. Foster, Akshay Krishnamurthy, and Haipeng Luo. Model selection for contextual bandits. In *NeurIPS*, pages 14714–14725, 2019.

Aurélien Garivier, Pierre Ménard, and Gilles Stoltz. Explore first, exploit next: The true shape of regret in bandit problems. *Mathematics of Operations Research*, 44 (2):377–399, 2019.

Avishek Ghosh, Abishek Sankararaman, and Kannan Ramchandran. Problem-complexity adaptive model selection for stochastic linear bandits. In *AISTATS*, volume 130 of *Proceedings of Machine Learning Research*, pages 1396–1404. PMLR, 2021.

Botao Hao, Tor Lattimore, and Csaba Szepesvári. Adaptive exploration in linear contextual bandit. In *AISTATS*, volume 108 of *Proceedings of Machine Learning Research*, pages 3536–3545. PMLR, 2020.

Jiahao He, Jiheng Zhang, and Rachel Q. Zhang. A reduction from linear contextual bandit lower bounds to estimation lower bounds. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8660–8677. PMLR, 17–23 Jul 2022.

Kwang-Sung Jun and Chicheng Zhang. Crush optimism with pessimism: Structured bandits beyond asymptotic optimality. *Advances in Neural Information Processing Systems*, 33:6366–6376, 2020.

Johannes Kirschner, Tor Lattimore, Claire Vernade, and Csaba Szepesvári. Asymptotically optimal information-directed sampling. In *Conference on Learning Theory*, pages 2777–2821. PMLR, 2021.

Tor Lattimore and Csaba Szepesvári. The end of optimism? an asymptotic analysis of finite-armed linear bandits. In *AISTATS*, volume 54 of *Proceedings of Machine Learning Research*, pages 728–737. PMLR, 2017.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Tor Lattimore, Csaba Szepesvári, and Gellért Weisz. Learning with good feature representations in bandits and in RL with a generative model. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 5662–5670. PMLR, 2020.

Ahmadreza Moradipari, Berkay Turan, Yasin Abbasi-Yadkori, Mahnoosh Alizadeh, and Mohammad Ghavamzadeh. Feature and parameter selection in stochastic linear bandits. In *ICML*, volume 162 of *Proceedings of Machine Learning Research*, pages 15927–15958. PMLR, 2022.

Jungseul Ok, Alexandre Proutière, and Damianos Tranos. Exploration in structured reinforcement learning. In *NeurIPS*, pages 8874–8882, 2018.

Aldo Pacchiano, My Phan, Yasin Abbasi-Yadkori, Anup Rao, Julian Zimmert, Tor Lattimore, and Csaba Szepesvári. Model selection in contextual stochastic bandit problems. In *NeurIPS*, 2020.

Matteo Papini, Andrea Tirinzoni, Marcello Restelli, Alessandro Lazaric, and Matteo Pirotta. Leveraging good representations in linear contextual bandits. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 8371–8380. PMLR, 2021.

Clémence Réda, Andrea Tirinzoni, and Rémy Degenne. Dealing with misspecification in fixed-confidence linear top-m identification. *Advances in Neural Information Processing Systems*, 34:25489–25501, 2021.

David Simchi-Levi and Yunzong Xu. Bypassing the monster: A faster and simpler optimal algorithm for contextual bandits under realizability. *CoRR*, abs/2003.12699, 2020.

Andrea Tirinzoni, Matteo Pirotta, Marcello Restelli, and Alessandro Lazaric. An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits. *Advances in Neural Information Processing Systems*, 33: 1417–1427, 2020.

Andrea Tirinzoni, Matteo Pirotta, and Alessandro Lazaric. A fully problem-dependent regret lower bound for finite-horizon mdps. *arXiv preprint arXiv:2106.13013*, 2021.

Andrea Tirinzoni, Matteo Papini, Ahmed Touati, Alessandro Lazaric, and Matteo Pirotta. Scalable representation learning in linear contextual bandits with constant regret guarantees. *EWRL*, 2022.

A NOTATION

Symbol	Meaning
	Learning problem
$\mathcal{X}$	Finite set of $X = |\mathcal{X}|$ contexts
$\mathcal{A}$	Finite set of $A = |\mathcal{A}|$ arms
ρ	Context distribution (full-support)
$f^\star : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$	Reward function
$R_T(f^\star)$	Cumulative regret when learning $f^\star$ (see Equation 1)
$\phi : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^{d_\phi}$	A d_ϕ-dimensional representation
Φ	Set of representations known to the agent
$\phi^\star$	The true realizable representation for $f^\star$
$\theta^\star$	The true parameter in $\mathbb{R}^{d_{\phi^\star}}$ such that $f^\star(x,a) = \phi^\star(x,a)^\mathsf{T}\theta^\star$
$\mathcal{F}_\Phi := \{f(\cdot,\cdot) = \phi(\cdot,\cdot)^\mathsf{T}\theta \mid \phi \in \Phi, \theta \in \mathbb{R}^{d_\phi}\}$	Function class known to the agent (s.t. $f^\star \in \mathcal{F}_\Phi$)
$\pi_f^\star(x) = \arg\max_{a \in \mathcal{A}} f(x,a)$	Optimal policy for reward function f
$\Delta_f(x,a) := \max_{a' \in \mathcal{A}} f(x,a') - f(x,a)$	Sub-optimality gap of (x,a) with reward f
$\mathrm{KL}_{x,a}(f,f') := \frac{1}{2}(f(x,a) - f'(x,a))^2$	KL divergence between the (Gaussian) observations in (x,a) under rewards f and f'
$\eta : \mathcal{X} \times \mathcal{A} \to [0,\infty)$	An allocation of samples
$z_\phi^\star(x,a) := \phi(x, \pi_{f^\star}^\star(x)) - \phi(x,a)$	Different between optimal and sub-optimal features
	Matrix notation
$F_\phi \in \mathbb{R}^{XA \times d_\phi}$	Feature matrix for representation ϕ (containing $\{\phi(x,a)\}_{x \in \mathcal{X}, a \in \mathcal{A}}$ as rows)
$f \in \mathbb{R}^{XA}$	Vectorized reward function (s.t. $f^\star = F_{\phi^\star}\theta^\star$)
$D_\eta := \mathrm{diag}(\{\eta(x,a)\}_{x \in \mathcal{X}, a \in \mathcal{A}})$	Diagonal matrix containing allocation η
$V_\eta(\phi) := F_\phi^\mathsf{T} D_\eta F_\phi$	Design matrix for representation ϕ (equiv. $V_\eta(\phi) := \sum_{x,a} \eta(x,a)\phi(x,a)\phi(x,a)^\mathsf{T}$)
$\|f^\star - F_\phi\theta\|_{D_\eta}^2$	MSE of model $\phi(\cdot,\cdot)^\mathsf{T}\theta$ in predicting $f^\star$ under allocation η
$\theta_\eta^\star(\phi) := \arg\min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi\theta\|_{D_\eta}^2$	Best fit for $f^\star$ using representation ϕ and allocation η
$\theta_\eta^\star(\phi) = V_\eta(\phi)^\dagger \sum_{x,a} \eta(x,a)\phi(x,a)f^\star(x,a)$	Closed-form expression for $\theta_\eta^\star(\phi)$ (equiv. $\theta_\eta^\star(\phi) = V_\eta(\phi)^\dagger F_\phi^\mathsf{T} D_\eta f^\star$)
	Linear algebra
$M^\dagger$	Pseudo-inverse of a matrix $M \in \mathbb{R}^{n \times m}$
$\mathrm{Im}(M)$	Column space of a matrix $M \in \mathbb{R}^{n \times m}$
$\mathrm{Row}(M)$	Row space of a matrix $M \in \mathbb{R}^{n \times m}$
$\mathrm{Ker}(M)$	Null space of a matrix $M \in \mathbb{R}^{n \times m}$
$\|v\|_M^2 := v^\mathsf{T} M v$	Weighted norm for a vector $v \in \mathbb{R}^d$ and a matrix $M \in \mathbb{R}^{d \times d}$

Table 1: The notation adopted in this paper.

Additional notation We introduce some additional terms w.r.t. those considered in the main paper.

Given two mean-reward functions $f, f' : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$, we define the Kullback-Leibler (KL) divergence between the corresponding distributions in a context-arm pair (x,a) as $\mathrm{KL}_{x,a}(f,f') := \frac{1}{2}(f(x,a) - f'(x,a))^2$.

B INSTANCE-DEPENDENT LOWER BOUNDS

B.1 General Lower Bound

We state the lower bound of Ok et al. (2018), which defines a complexity measure $\mathcal{C}(f^\star, \mathcal{F})$ for learning any reward $f^\star$ in any given function class $\mathcal{F}$. Our lower bound for representation learning (Theorem 1) will be derived by instantiating this result for the specific class $\mathcal{F}_\Phi$.

Theorem 4 (Ok et al. (2018)). *Let $\mathfrak{A}$ be a uniformly good strategy on a class $\mathcal{F}$. Then, for any $f^\star \in \mathcal{F}$ such that $\pi_{f^\star}^\star$ is unique,*

$$\liminf_{T \to \infty} \frac{\mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)]}{\log(T)} \geq \mathcal{C}(f^\star, \mathcal{F}),$$

where $\mathcal{C}(f^\star, \mathcal{F})$ is the value of the optimization problem

$$\inf_{\{\eta(x,a)\} \geq 0} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a)\Delta_{f^\star}(x,a) \quad \text{s.t.} \quad \inf_{f \in \Lambda(f^\star, \mathcal{F})} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a)\mathrm{KL}_{x,a}(f^\star, f) \geq 1,$$

where $\Lambda(f^\star, \mathcal{F}) := \{f \in \mathcal{F} \mid \exists x \in \mathcal{X}, a \neq \pi_{f^\star}^\star(x) : f(x,a) > f(x, \pi_{f^\star}^\star(x))\}$ *is the set of alternatives for* f.

Proposition 8 (Monotonicity of $\mathcal{C}(f^\star, \mathcal{F})$)**.** *For any two function classes* $\underline{\mathcal{F}}, \overline{\mathcal{F}}$ *such that* $\underline{\mathcal{F}} \subseteq \overline{\mathcal{F}}$ *and* $f^\star \in \underline{\mathcal{F}} \cap \overline{\mathcal{F}}$, $\mathcal{C}(f^\star, \underline{\mathcal{F}}) \leq \mathcal{C}(f^\star, \overline{\mathcal{F}})$.

Proof. This result is immediate from Theorem 4: $\underline{\mathcal{F}} \subseteq \overline{\mathcal{F}}$ implies $\Lambda(f^\star, \underline{\mathcal{F}}) \subseteq \Lambda(f^\star, \overline{\mathcal{F}})$ by definition of alternative set, which in turns implies that the feasibility set of the optimization problem $\Lambda(f^\star, \overline{\mathcal{F}})$ includes the one of the optimization problem $\Lambda(f^\star, \underline{\mathcal{F}})$. The objective functions are the same, hence the second value must be smaller. $\square$

B.2 The Unstructured Case

Let $\mathcal{F}_{\mathrm{uns}}$ denote the set of all possible functions mapping $\mathcal{X} \times \mathcal{A}$ into $\mathbb{R}$. Note that an algorithm learning an instance $f^\star$ with $\mathcal{F}_{\mathrm{uns}}$ as input has no prior knowledge about $f^\star$ itself. We call this setting "unstructured", as opposed to the "structured" setting where we are given $\mathcal{F} \subset \mathcal{F}_{\mathrm{uns}}$ (e.g., $\mathcal{F} = \mathcal{F}_\Phi$). The following result formally shows that the complexity measure from Theorem 4 for such an unstructured setting reduces exactly to the sum of inverse gaps appearing both in Proposition 3 and 5.

Theorem 5. *For any* $f^\star \in \mathcal{F}$ *such that* $\pi_{f^\star}^\star$ *is unique,*

$$\mathcal{C}(f, \mathcal{F}_{\mathrm{uns}}) = \sum_{x \in \mathcal{X}} \sum_{a \neq \pi_{f^\star}^\star(x)} \frac{2}{\Delta_{f^\star}(x,a)}.$$

Proof. Note that the alternative set $\Lambda(f^\star, \mathcal{F}_{\mathrm{uns}})$ can be decomposed into a union of $X(A-1)$ half-spaces, each associated to a sub-optimal context-arm pair (x,a) containing those instances such that $f(x,a) > f(x, \pi_{f^\star}^\star(x))$. It is easy to see that the closest alternative for the half-space associated with (x,a) (i.e., the one minimizing the KL divergence in the constraint of Theorem 4) is $f(x,a) = f^\star(x,a) + \Delta_{f^\star}(x,a)$ and $f(x',a') = f^\star(x',a')$ for all $(x',a') \neq (x,a)$. Therefore, the constraint associated with (x,a) yields $\eta(x,a) \geq 2/\Delta_{f^\star}(x,a)^2$. There is one such constraint for each suboptimal (x,a), hence yielding the stated lower bound. $\square$

B.3 Proof of Theorem 1

The result follows by instantiating the general lower bound of Theorem 4 to our specific function class $\mathcal{F}_\Phi := \{f(\cdot,\cdot) = \phi(\cdot,\cdot)^\mathsf{T}\theta \mid \phi \in \Phi, \theta \in \mathbb{R}^{d_\phi}\}$. First note that the resulting set of alternatives can be decomposed into a union of half-spaces,

$$\Lambda(f^\star, \mathcal{F}_\Phi) := \bigcup_{\phi \in \Phi} \bigcup_{x \in \mathcal{X}} \bigcup_{a \neq \pi_{f^\star}^\star(x)} \left\{ \theta \in \mathbb{R}^{d_\phi} \mid \phi(x,a)^\mathsf{T}\theta > \phi(x, \pi_{f^\star}^\star(x))^\mathsf{T}\theta \right\}. \tag{4}$$

Moreover, for any $f \in \mathcal{F}_\Phi$ such that $f(x,a) = \phi(x,a)^\mathsf{T}\theta$,

$$\sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a) \mathrm{KL}_{x,a}(f^\star, f) = \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a) \left(f^\star(x,a) - \phi(x,a)^\mathsf{T}\theta \right)^2.$$

Therefore, the infimum over alternatives in the constraint of Theorem 4 can be computed by performing one minimization of a quadratic function over (the closure of) each half-space in (4). Formally, for each $\phi \in \Phi, \bar{x} \in \mathcal{X}, \bar{a} \neq \pi_{f^\star}^\star(\bar{x})$, we compute

$$\mathcal{I}_\eta(f^\star, \phi, \bar{x}, \bar{a}) := \min_{\theta \in \mathbb{R}^{d_\phi}} \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a) \left(f^\star(x,a) - \phi(x,a)^\mathsf{T}\theta \right)^2 \quad \text{s.t.} \quad \phi(\bar{x},\bar{a})^\mathsf{T}\theta \geq \phi(\bar{x}, \pi_{f^\star}^\star(\bar{x}))^\mathsf{T}\theta. \tag{5}$$

This can be re-written in matrix notation as

$$\mathcal{I}_\eta(f^\star, \phi, \bar{x}, \bar{a}) = \min_{\theta \in \mathbb{R}^{d_\phi}} \frac{1}{2} \|f^\star - F_\phi \theta\|_{D_\eta}^2 \quad \text{s.t.} \quad (\phi(\bar{x},\bar{a}) - \phi(\bar{x}, \pi_{f^\star}^\star(\bar{x})))^\mathsf{T}\theta \geq 0. \tag{6}$$

The optimal value can be found in closed-form by using Lemma 1 (proved below) with $z = \phi(\bar{x},\bar{a}) - \phi(\bar{x}, \pi_{f^\star}^\star(\bar{x}))$. Theorem 1 is then proved by plugging this result into Theorem 4, while noting that the constraint in the latter lower bound can be written as

$$\mathcal{I}_\eta(f^\star, \phi, x, a) \geq 1 \quad \forall \phi \in \Phi, x \in \mathcal{X}, a \neq \pi_{f^\star}^\star(x). \tag{7}$$

$\square$

B.4 Implicit Formulation of Theorem 1

An immediate corollary from the proof of Theorem 1 (see in particular Equation 7) is the following implicit version of our lower bound for representation learning.

Corollary 3. *For any Φ and $f^\star \in \mathcal{F}_\Phi$ such that $\pi_{f^\star}^\star$ is unique,*

$$\mathcal{C}(f^\star, \mathcal{F}_\Phi) = \inf_{\{\eta(x,a)\} \geq 0} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a)\Delta_{f^\star}(x,a) \quad \text{s.t.} \quad \mathcal{I}_\eta(f^\star, \phi, x, a) \geq 1 \ \forall \phi \in \Phi, x \in \mathcal{X}, a \neq \pi_{f^\star}^\star(x).$$

B.5 Minimizing Over Half-Spaces

We derive a general result which gives a closed-form expression for the minimization of the mean-square error for predicting a function $f^\star$ using the linear space $\mathcal{F}_{\{\phi\}}$ subject to the constraint that the parameter θ lies into some half-space.

Lemma 1. *Let $\phi \in \Phi$, $z \in \mathbb{R}^{d_\phi}$, and D_η be any $XA \times XA$ diagonal matrix with non-negative entries. Consider the optimization problem*

$$\mathcal{I}_\eta(f^\star, \phi, z) := \min_{\theta \in \mathbb{R}^{d_\phi}} \left\{ \frac{1}{2}\|f^\star - F_\phi\theta\|_{D_\eta}^2 \quad \text{s.t.} \quad z^\mathsf{T}\theta \geq 0 \right\}. \tag{8}$$

Let $\theta_\eta(f^\star, \phi, z)$ be a parameter attaining the minimum. Then,

$$\mathcal{I}_\eta(f^\star, \phi, z) = \frac{1}{2}\|f^\star - F_\phi\theta_\eta^\star(\phi)\|_{D_\eta}^2 + \mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0, z \in \operatorname{Im}(V_\eta(\phi))\right) \frac{(z^\mathsf{T}\theta_\eta^\star(\phi))^2}{2\|z\|_{V_\eta(\phi)^\dagger}^2},$$

$$\theta_\eta(f^\star, \phi, z) = \theta_\eta^\star(\phi) - \mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0, z \in \operatorname{Im}(V_\eta(\phi))\right) \frac{z^\mathsf{T}\theta_\eta^\star(\phi)}{\|z\|_{V_\eta(\phi)^\dagger}^2} V_\eta(\phi)^\dagger z.$$

Proof. Note that, for any vector $y \in \mathbb{R}^{XA}$, $\|y\|_{D_\eta}^2 = \|D_\eta^{1/2}y\|_2^2$. The Lagrange dual problem corresponding to (8) is

$$\max_{\lambda \in \mathbb{R}_{\geq 0}} \min_{\theta \in \mathbb{R}^{d_\phi}} \underbrace{\left\{ \frac{1}{2}\|D_\eta^{1/2}f^\star - D_\eta^{1/2}F_\phi\theta\|_2^2 - \lambda z^\mathsf{T}\theta \right\}}_{:=g(\theta,\lambda)}.$$

Let us fix λ and optimize $g(\theta, \lambda)$ over θ. Let (U, Σ, V) be an SVD decomposition of $D_\eta^{1/2}F_\phi$, i.e., such that $D_\eta^{1/2}F_\phi = U\Sigma V^\mathsf{T}$ with $U \in \mathbb{R}^{XA \times XA}$, $\Sigma \in \mathbb{R}^{XA \times d_\phi}$ (diagonal), and $V \in \mathbb{R}^{d_\phi \times d_\phi}$. Since both U and V are orthogonal matrices (i.e., $U^\mathsf{T}U = UU^\mathsf{T} = I_{XA}$ and $V^\mathsf{T}V = VV^\mathsf{T} = I_{d_\phi}$),

$$g(\theta, \lambda) = \frac{1}{2}\|U^\mathsf{T}D_\eta^{1/2}f^\star - \Sigma V^\mathsf{T}\theta\|_2^2 - \lambda z^\mathsf{T}VV^\mathsf{T}\theta.$$

Note that this follows since, for any two vectors $x, y \in \mathbb{R}^{XA}$, by orthogonality of U,

$$\|x - Uy\|_2^2 = \|UU^\mathsf{T}x - Uy\|_2^2 = (U^\mathsf{T}x - y)^\mathsf{T}U^\mathsf{T}U(U^\mathsf{T}x - y) = \|U^T x - y\|_2^2.$$

We can now perform a change of variables $y = V^\mathsf{T}\theta$ and define the function

$$\tilde{g}(y, \lambda) = \frac{1}{2}\|U^\mathsf{T}D_\eta^{1/2}f^\star - \Sigma y\|_2^2 - \lambda z^\mathsf{T}Vy.$$

Since V is invertible and $V^\mathsf{T} = V^{-1}$, if y_λ is a minimizer of $\tilde{g}(y, \lambda)$, then $\theta_\lambda = Vy_\lambda$ is a minimizer of $g(\theta, \lambda)$.

Let us thus minimize $\tilde{g}(y, \lambda)$ as a function of y. Its gradient w.r.t. y is

$$\nabla_y\tilde{g}(y, \lambda) = -\Sigma^\mathsf{T}(U^\mathsf{T}D_\eta^{1/2}f^\star - \Sigma y) - \lambda V^\mathsf{T}z.$$

Equating it to zero, we obtain the inequality

$$\Sigma^\mathsf{T}\Sigma y = \Sigma^\mathsf{T}U^\mathsf{T}D_\eta^{1/2}f^\star + \lambda V^\mathsf{T}z.$$

We now distinguish three cases.

Case 1: $z = 0_{d_\phi}$ Note that $\Sigma^\mathsf{T}\Sigma$ is a $d_\phi \times d_\phi$ diagonal matrix with $\{\sigma_i^2\}_{i \in [d_\phi]}$ on its diagonal, i.e., the squared singular values of $D_\eta^{1/2}F_\phi$ or, equivalently, the eigenvalues of $(D_\eta^{1/2}F_\phi)^\mathsf{T}D_\eta^{1/2}F_\phi = F_\phi^\mathsf{T}D_\eta F_\phi = V_\eta(\phi)$. Note that some of these might be zero as $V_\eta(\phi)$ might not be full rank. In this case, it is easy to see that a solution is

$$y_\lambda = (\Sigma^\mathsf{T}\Sigma)^\dagger \Sigma^\mathsf{T}U^\mathsf{T}D_\eta^{1/2}f^\star = \Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star,$$

where we used Property 2 of the pseudo-inverse (see Appendix G). Note that λ has no impact on the optimization problem (in other words, there is no constraint to be satisfied). That is, the optimal θ solving the original optimization problem (8) is

$$\theta = Vy_\lambda = V\Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star \overset{(a)}{=} (D_\eta^{1/2}F_\phi)^\dagger D_\eta^{1/2}f^\star \overset{(b)}{=} (F_\phi^\mathsf{T}D_\eta F_\phi)^\dagger F_\phi^\mathsf{T}D_\eta f^\star = V_\eta(\phi)^\dagger F_\phi^\mathsf{T}D_\eta f^\star = \theta_\eta^\star(\phi),$$

where (a) uses the definition of pseudo-inverse of $D_\eta^{1/2}F_\phi$, (b) uses Property 2 of the pseudo-inverse, and the last two equalities use respectively the definition of $V_\eta(\phi)$ and of $\theta_\eta^\star(\phi)$ (see Appendix A).

Case 2: $z \in \mathrm{Ker}(V_\eta(\phi))$ **and** $z \neq 0_{d_\phi}$ Suppose that $V_\eta(\phi)$ has rank $d' < d_\phi$ and that the singular values in Σ are sorted in non-increasing order. Then, since the columns of V from index $d'+1$ to index d_ϕ span $\mathrm{Ker}(V_\eta(\phi))$, the vector $V^\mathsf{T}z$ has at least one non-zero element in a coordinate $i \in \{d'+1, \ldots, d_\phi\}$. However, the vectors $\Sigma^\mathsf{T}\Sigma y$ and $\Sigma^\mathsf{T}U^\mathsf{T}D_\eta^{1/2}f^\star$ have clearly all zero components in those coordinates. This means that the gradient cannot be equated to zero. In particular, this implies that, for any $\lambda \neq 0$, we can find y such that $\tilde{g}(y, \lambda) = -\infty$. That is, the optimal solution must be at $\lambda = 0$, in which case we reduce to Case 1 (the constraint has no impact) and we get the same optimal parameter/value.

Case 3: $z \notin \mathrm{Ker}(V_\eta(\phi))$ In this case, $[V^\mathsf{T}z]_i = 0$ for all $i \in \{d'+1, \ldots, d_\phi\}$. That is, the gradient can now be equated to zero yielding, by Property 2 of the pseudo-inverse,

$$y_\lambda = (\Sigma^\mathsf{T}\Sigma)^\dagger \Sigma^\mathsf{T}U^\mathsf{T}D_\eta^{1/2}f^\star + \lambda(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T}z = \Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star + \lambda(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T}z.$$

Then,

$$\tilde{g}(y_\lambda, \lambda) = \frac{1}{2}\|U^\mathsf{T}D_\eta^{1/2}f^\star - \Sigma\Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star - \lambda\Sigma(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T}z\|_2^2 - \lambda z^\mathsf{T}V\Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star - \lambda^2 z^\mathsf{T}V(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T}z.$$

Note that $\Sigma(\Sigma^\mathsf{T}\Sigma)^\dagger = (\Sigma^\mathsf{T})^\dagger$, $\Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star = V^\mathsf{T}\theta_\eta^\star(\phi)$ (proved in Case 1), and, by Property 4 of pseudo-inverses,

$$V(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T} = V\Sigma^\dagger(\Sigma^\mathsf{T})^\dagger V^\mathsf{T} = V\Sigma^\dagger U^\mathsf{T}U(\Sigma^\mathsf{T})^\dagger V^\mathsf{T} = (D_\eta^{1/2}F_\phi)^\dagger((D_\eta^{1/2}F_\phi)^\mathsf{T})^\dagger = ((D_\eta^{1/2}F_\phi)^\mathsf{T}D_\eta^{1/2}F_\phi)^\dagger = V_\eta(\phi)^\dagger.$$

Hence,

$$\tilde{g}(y_\lambda, \lambda) = \frac{1}{2}\|\underbrace{U^\mathsf{T}D_\eta^{1/2}f^\star - \Sigma V^\mathsf{T}\theta_\eta^\star(\phi)}_{:=w} - \lambda(\Sigma^\mathsf{T})^\dagger V^\mathsf{T}z\|_2^2 - \lambda z^\mathsf{T}\theta_\eta^\star(\phi) - \lambda^2\|z\|_{V_\eta(\phi)^\dagger}^2$$

$$= \frac{1}{2}\|w\|_2^2 + \frac{\lambda^2}{2}\underbrace{\|(\Sigma^\mathsf{T})^\dagger V^\mathsf{T}z\|_2^2}_{=\|z\|_{V_\eta(\phi)^\dagger}^2} - \lambda z^\mathsf{T}V\Sigma^\dagger w - \lambda z^\mathsf{T}\theta_\eta^\star(\phi) - \lambda^2\|z\|_{V_\eta(\phi)^\dagger}^2$$

$$= \frac{1}{2}\|w\|_2^2 - \lambda z^\mathsf{T}\underbrace{V\Sigma^\dagger U^\mathsf{T}D_\eta^{1/2}f^\star}_{=\theta_\eta^\star(\phi)} + \lambda z^\mathsf{T}\underbrace{V\Sigma^\dagger\Sigma V^\mathsf{T}\theta_\eta^\star(\phi)}_{=\theta_\eta^\star(\phi)} - \lambda z^\mathsf{T}\theta_\eta^\star(\phi) - \frac{\lambda^2}{2}\|z\|_{V_\eta(\phi)^\dagger}^2$$

$$= \frac{1}{2}\|w\|_2^2 - \lambda z^\mathsf{T}\theta_\eta^\star(\phi) - \frac{\lambda^2}{2}\|z\|_{V_\eta(\phi)^\dagger}^2.$$

Differentiating w.r.t. λ,

$$\frac{\mathrm{d}\tilde{g}(y_\lambda, \lambda)}{\mathrm{d}\lambda} = -z^\mathsf{T}\theta_\eta^\star(\phi) - \lambda\|z\|_{V_\eta(\phi)^\dagger}^2.$$

If $z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0$, then the optimum is obtained with $\lambda = -z^\mathsf{T}\theta_\eta^\star(\phi)/\|z\|_{V_\eta(\phi)^\dagger}^2$, otherwise the optimum is at $\lambda = 0$. Therefore, the optimal λ is

$$\lambda = -\frac{z^\mathsf{T}\theta_\eta^\star(\phi)\mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0\right)}{\|z\|_{V_\eta(\phi)^\dagger}^2}.$$

Plugging this into the expression of y_λ and then into $\theta_\lambda = V y_\lambda$, we find that

$$\theta_\lambda = \theta_\eta^\star(\phi) - \frac{z^\mathsf{T}\theta_\eta^\star(\phi)\mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0\right)}{\|z\|_{V_\eta(\phi)^\dagger}^2}V(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T}z = \theta_\eta^\star(\phi) - \frac{z^\mathsf{T}\theta_\eta^\star(\phi)\mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0\right)}{\|z\|_{V_\eta(\phi)^\dagger}^2}V_\eta(\phi)^\dagger z.$$

Final result Combining the three cases, it is easy to see that the optimal parameter is

$$\theta_\eta(f^\star, \phi, z) = \theta_\eta^\star(\phi) - \mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0, z \in \mathrm{Im}(V_\eta(\phi))\right)\frac{z^\mathsf{T}\theta_\eta^\star(\phi)}{\|z\|_{V_\eta(\phi)^\dagger}^2}V_\eta(\phi)^\dagger z.$$

The corresponding optimal value is

$$\mathcal{I}_\eta(f^\star, \phi, z) := \frac{1}{2}\|f^\star - F_\phi\theta_\eta(f^\star, \phi, z)\|_{D_\eta}^2$$

$$= \frac{1}{2}\|f^\star - F_\phi\theta_\eta^\star(\phi)\|_{D_\eta}^2 + \mathbb{1}\left(z^\mathsf{T}\theta_\eta^\star(\phi) \leq 0, z \in \mathrm{Im}(V_\eta(\phi))\right)\left(\frac{(z^\mathsf{T}\theta_\eta^\star(\phi))^2}{2\|z\|_{V_\eta(\phi)^\dagger}^2} + \frac{z^\mathsf{T}\theta_\eta^\star(\phi)}{\|z\|_{V_\eta(\phi)^\dagger}^2}\alpha\right),$$

where

$$\alpha = z^\mathsf{T}V_\eta(\phi)^\dagger F_\phi^\mathsf{T}D_\eta(f^\star - F_\phi\theta_\eta^\star(\phi)) = z^\mathsf{T}\theta_\eta^\star(\phi) - z^\mathsf{T}V_\eta(\phi)^\dagger V_\eta(\phi)\theta_\eta^\star(\phi)$$

$$= z^\mathsf{T}\theta_\eta^\star(\phi) - z^\mathsf{T}V_\eta(\phi)^\dagger V_\eta(\phi)V_\eta(\phi)^\dagger F_\phi^\mathsf{T}D_\eta f^\star$$

$$= z^\mathsf{T}\theta_\eta^\star(\phi) - z^\mathsf{T}V_\eta(\phi)^\dagger F_\phi^\mathsf{T}D_\eta f^\star = z^\mathsf{T}\theta_\eta^\star(\phi) - z^\mathsf{T}\theta_\eta^\star(\phi) = 0.$$

This concludes the proof.

$\square$

C PROOF OF PROPOSITIONS 2-7

C.1 Proof of Proposition 2

This result follows easily from the definition of $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ in Theorem 1: for any realizable $\phi \in \Phi$ (i.e., such that $f^\star \in \mathcal{F}_{\{\phi\}}$), it is sufficient to drop the constraints associated with all representations except those for ϕ itself. The resulting optimization problem yields exactly the lower bound for the CLB $(f^\star, \mathcal{F}_{\{\phi\}})$ (Corollary 1). Then it must be that $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \geq \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}})$ since we enlarged the feasibility set.

Alternatively, one can see this as a consequence of Proposition 8 by noting that $\mathcal{F}_\Phi = \cup_{\phi \in \Phi}\mathcal{F}_{\{\phi\}}$. $\square$

C.2 Proof of Proposition 3

Suppose, for simplicity, that $X(A-1)/(d-1)$ is an integer value.[7] Let $\mathcal{Z} = \{(x, a) \in \mathcal{X} \times \mathcal{A} : a \neq \pi_{f^\star}^\star(x)\}$ be the set of sub-optimal context-arm pairs for $f^\star$. Note that it has cardinality $X(A-1)$. Let us partition it in $n = X(A-1)/(d-1)$ subsets $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$, each of size $d-1$. Let us enumerate the sub-optimal context-arm pairs as follows: for $i \in [n], j \in [d-1]$, (x_{ij}, a_{ij}) denotes the j-th context-arm pair contained in $\mathcal{Z}_i$.

Let us define a set $\Phi = \{\phi_1, \ldots, \phi_n\}$ of n d-dimensional representations as follows. For each $i \in [n]$, we choose

$$\phi_i(x, a) = \begin{bmatrix} f^\star(x, a) \\ \Delta_{f^\star}(x, a)\mathbb{1}\left((x, a) = (x_{i1}, a_{i1})\right) \\ \vdots \\ \Delta_{f^\star}(x, a)\mathbb{1}\left((x, a) = (x_{i(d-1)}, a_{i(d-1)})\right) \end{bmatrix}$$

Moreover, for each sub-optimal pair (x_{ij}, a_{ij}), let us define the function f_{ij} as

$$f_{ij}(x, a) = \begin{cases} f^\star(x, a) + \Delta_{f^\star}(x, a) & \text{if } (x, a) = (x_{ij}, a_{ij}) \\ f^\star(x, a) & \text{otherwise.} \end{cases}$$

---

[7]The proof trivially extends to the case where $X(A-1)/(d-1)$ is not an integer by consider $\lceil X(A-1)/(d-1) \rceil$ instead.

Then, it is easy to see that $f \in \mathcal{F}_{\{\phi_i\}}$ for each $f \in \{f^\star, f_{i1}, \ldots, f_{i(d-1)}\}$. In particular, $f^\star$ is realized by choosing $\theta = (1, 0, \ldots, 0)^\top$, while f_{ij} with $\theta = (1, 0, \ldots, 1, 0, \ldots, 0)^\top$ with the second 1 in position $j+1$. This implies that *all* $\phi \in \Phi$ are realizable for $f^\star$.

Lower bounding $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$ Note that, by our choice of Φ, $f_{ij} \in \mathcal{F}_\Phi$ for all $i \in [n], j \in [d-1]$. Moreover, f_{ij} corresponds to the closest alternative for the half-space associated with the suboptimal pair (x_{ij}, a_{ij}) in the unstructured lower bound (see the proof of Theorem 5). This implies that, if we evaluate the constraint at (x_{ij}, a_{ij}) and ϕ_i in the implicit form of Theorem 1 (see Corollary 3), we obtain that

$$
\mathcal{I}_\eta(f^\star, \phi_i, x_{ij}, a_{ij}) = \min_{\theta : \phi_i(x_{ij}, a_{ij})^\top \theta \geq \phi_i(x_{ij}, \pi_{f^\star}^\star(x_{ij}))^\top \theta} \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x, a) \left(f^\star(x, a) - \phi_i(x, a)^\top \theta \right)^2
$$

$$
= \frac{1}{2} \eta(x_{ij}, a_{ij}) \Delta_{f^\star}(x_{ij}, a_{ij}),
$$

since the optimum is attained by θ such that $\phi_i(x, a)^\top \theta = f_{ij}(x, a)$. The constraint imposes that the right-hand side is larger than 1. This holds for all sub-optimal pairs (x_{ij}, a_{ij}). Therefore, if we call $\bar{\eta}$ an optimal feasible allocation,

$$
\mathcal{C}(f^\star, \mathcal{F}_\Phi) = \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \bar{\eta}(x, a) \Delta_{f^\star}(x, a) = \sum_{x \in \mathcal{X}} \sum_{a \neq \pi_{f^\star}^\star(x)} \frac{2}{\Delta_{f^\star}(x, a)}.
$$

Upper bounding $\sup_{\phi \in \Phi} \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}})$ Now take some $\phi_i \in \Phi$ for $i \in [n]$. From Corollary 3, we know that any feasible η for the optimization problem in $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_i\}})$ must satisfy

$$
\inf_{\theta : \phi_i(\bar{x}, \bar{a})^\top \theta \geq \phi_i(\bar{x}, \pi_{f^\star}^\star(\bar{x}))^\top \theta} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x, a) \left(f^\star(x, a) - \phi_i(x, a)^\top \theta \right)^2 \geq 2 \quad \forall \bar{x} \in \mathcal{X}, \bar{a} \neq \pi_{f^\star}^\star(\bar{x}).
$$

For any θ, from the definition of ϕ_i,

$$
\sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x, a) \left(f^\star(x, a) - \phi_i(x, a)^\top \theta \right)^2
$$

$$
= \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x, a) \left(f^\star(x, a)(1 - \theta_1) - \Delta_{f^\star}(x, a) \sum_{j : (x, a) = (x_{ij}, a_{ij})} \theta_{j+1} \right)^2
$$

$$
= \sum_{(x, a) \notin \{(x_{ij}, a_{ij})\}_{j \in [d-1]}} \eta(x, a) \left(f^\star(x, a)(1 - \theta_1) \right)^2 + \sum_{j=1}^{d-1} \eta(x_{ij}, a_{ij}) \left(f^\star(x_{ij}, a_{ij})(1 - \theta_1) - \Delta_{f^\star}(x_{ij}, a_{ij}) \theta_{j+1} \right)^2.
$$

Let us now minimize this quantity for θ such that $\phi_i(\bar{x}, \bar{a})^\top \theta \geq \phi_i(\bar{x}, \pi(\bar{x}))^\top \theta$ for some fixed $(\bar{x}, \bar{a})$.

Clearly, for any $(\bar{x}, \bar{a}) \notin \{(x_{ij}, a_{ij})\}_{j \in [d-1]}$, we have $\phi_i(\bar{x}, \bar{a})^\top \theta = f^\star(\bar{x}, \bar{a})$. This implies that there is no constraint associated with such a pair $(\bar{x}, \bar{a})$ (the set of which we take the infimum is actually empty). Thus, we can set $\eta(\bar{x}, \bar{a}) = 0$ for such pairs.

For $(\bar{x}, \bar{a}) = (x_{ij}, a_{ij})$ for some $j \in [d-1]$, the condition on θ becomes $\theta_{j+1} \Delta_{f^\star}(x_{ij}, a_{ij}) \geq \theta_1 \Delta_{f^\star}(x_{ij}, a_{ij})$ which is equivalent to $\theta_{j+1} \geq \theta_1$ since $\Delta_{f^\star}(x_{ij}, a_{ij}) > 0$. Let us suppose that $\eta(x, \pi_{f^\star}^\star(x)) = \infty$ (or a very large value) for all x. Note that this has no impact on the final objective value since we pay zero regret for playing optimal actions. Then, it is easy to see that the solution to the optimization problem above is attained with $\theta_1 = 1$ (otherwise the first term would be extremely large), $\theta_{j+1} = 1$, and $\theta_{l+1} = 0$ for $l \neq j$. The corresponding constraint is thus $\eta(x_{ij}, a_{ij}) \geq 2/\Delta_{f^\star}(x_{ij}, a_{ij})^2$ for all $j \in [d-1]$. Therefore, the resulting optimal allocation $\bar{\eta}$ requires playing (1) exactly $2/\Delta_{f^\star}(x_{ij}, a_{ij})^2$ times all pairs in $\{(x_{ij}, a_{ij})\}_{j \in [d-1]}$, (2) a large number of times the optimal pairs, and (3) zero times all the other pairs. Its regret is trivially

$$
\sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \bar{\eta}(x, a) \Delta_{f^\star}(x, a) = \sum_{j=1}^{d-1} \frac{2}{\Delta_{f^\star}(x_{ij}, a_{ij})} \leq \frac{2(d-1)}{\Delta_{\min}}.
$$

This holds for all $i \in [n]$, which concludes the proof.

$\square$

C.3 Proof of Proposition 4

We already proved in Proposition 2 that $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \geq \sup_{\phi \in \Phi: f^\star \in \mathcal{F}_{\{\phi\}}} \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}}) = \mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$, where the equality holds since $\phi^\star$ is the unique realizable representation in Φ. We now prove that $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \leq \mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$. Combining these two results clearly proves the statement.

Consider any optimal feasible allocation $\bar{\eta}$ for the optimization problem associated to $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$ (i.e., the CLB with known $\phi^\star$). Let $\tilde{\eta} = \bar{\eta} + \frac{2}{\varepsilon}\eta^\star$. Clearly, since $\eta^\star$ plays only optimal actions (i.e., with zero regret),

$$\sum_{x,a} \tilde{\eta}(x,a)\Delta_{f^\star}(x,a) = \sum_{x,a} \bar{\eta}(x,a)\Delta_{f^\star}(x,a) = \mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}}).$$

Moreover, for any $\phi \in \Phi$ such that $\phi \neq \phi^\star$,

$$\|f^\star - F_\phi\theta_{\tilde{\eta}}^\star(\phi)\|_{D_{\tilde{\eta}}}^2 = \min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi\theta\|_{D_{\tilde{\eta}}}^2 \geq \frac{2}{\varepsilon} \min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi\theta\|_{D_{\eta^\star}}^2 = \frac{2}{\varepsilon}\|f^\star - F_\phi\theta_{\eta^\star}^\star(\phi)\|_{D_{\eta^\star}}^2 \geq 2,$$

where the last inequality follows by the assumption on all misspecified representations. This implies that $\tilde{\eta}$ is feasible in the optimization problem of $\mathcal{C}(f^\star, \mathcal{F}_\Phi)$, which, together with the equality above, proves that $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \leq \mathcal{C}(f^\star, \mathcal{F}_{\{\phi^\star\}})$.

$\square$

C.4 Proof of Proposition 5

First note that $\mathcal{F}_{\{\bar{\phi}\}}$ is the set of all possible reward functions. In fact, for any mapping $f : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$, $f(x,a) = \bar{\phi}(x,a)^\mathsf{T}\theta$ for θ the vectorization of f, which implies that $f \in \mathcal{F}_{\{\bar{\phi}\}}$. Then, since $\mathcal{F}_{\{\bar{\phi}\}} \subseteq \mathcal{F}_{\Phi \cup \{\bar{\phi}\}}$ by definition, $\mathcal{F}_{\Phi \cup \{\bar{\phi}\}}$ is also the set of all possible reward functions, which implies that $\mathcal{C}(f^\star, \mathcal{F}_{\Phi \cup \{\bar{\phi}\}})$ is exactly the complexity of Theorem 5 for learning $f^\star$ without any prior knowledge. This is exactly the one stated in Proposition 5. $\square$

C.5 Proof of Proposition 6

The inequality $\mathcal{C}(f^\star, \mathcal{F}_\Phi) \geq \mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}})$ is an immediate consequence of Proposition 2 since ϕ_N is realizable. The equality follows since one could simply run an instance optimal algorithm for CLB problems (e.g., Tirinzoni et al., 2020) on representation ϕ_N and obtain regret asymptotically approaching $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}}) \log(T)$. $\square$

C.6 Proof of Proposition 7

Note that a necessary and sufficient condition for $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$ is that there exists a feasible allocation η with $\eta(x,a) = 0$ for all $x \in \mathcal{X}$ and $a \neq \pi_{f^\star}^\star(x)$. That is, a feasible allocation which plays only optimal arms.

Let us start by proving that the condition is necessary for $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$. Let us proceed by contradiction. Suppose that there exists ϕ with $\min_\theta \|f^\star - F_\phi\theta\|_{D_{\eta^\star}}^2 = 0$ and $x \in \mathcal{X}, a \neq \pi_{f^\star}^\star(x)$ such that $z_\phi(x,a)^\mathsf{T}\theta_{\eta^\star}^\star(\phi) \leq 0$ or $z_\phi(x,a) \notin \mathrm{Im}(V_{\eta^\star}(\phi))$, while $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$. Clearly, if $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$, there exists a value $M > 0$ (possibly very large) such that the allocation $\eta^\star$ scaled by M is feasible . Note that scaling by a constant does not affect the column space of the resulting design matrix (i.e., $\mathrm{Im}(V_{\eta^\star}(\phi)) = \mathrm{Im}(V_{M\eta^\star}(\phi))$), nor does it affect the best fit in $\mathcal{F}_{\{\phi\}}$ (i.e., $\theta_{\eta^\star}^\star(\phi) = \theta_{M\eta^\star}^\star(\phi)$). However, the negation of the stated condition implies that there exists (ϕ, x, a) such that $\mathcal{I}_{M\eta^\star}(f^\star, \phi, x, a) = \mathcal{I}_{\eta^\star}(f^\star, \phi, x, a) = 0$. Here $\mathcal{I}_{\eta^\star}(f^\star, \phi, x, a) = 0$ is the constraint function associated to (ϕ, x, a) in Corollary 3. That is, $M\eta^\star$ is infeasible. This is clearly a contradiction, and thus the stated condition is necessary.

Let us now prove that the condition is sufficient for $\mathcal{C}(f^\star, \mathcal{F}_\Phi) = 0$. Take any ϕ. We consider two cases.

Case 1: $\min_\theta \|f^\star - F_\phi\theta\|_{D_{\eta^\star}}^2 > 0$ Simply rescaling $\eta^\star$ by $M > 0$ yields $\min_\theta \|f^\star - F_\phi\theta\|_{D_{M\eta^\star}}^2 = M \min_\theta \|f^\star - F_\phi\theta\|_{D_{\eta^\star}}^2$. Since the latter term is strictly positive, we can set $M = 1/\min_\theta \|f^\star - F_\phi\theta\|_{D_{\eta^\star}}^2$ to guarantee that $\mathcal{I}_{M\eta^\star}(f^\star, \phi, x, a) \geq 1$. That is $M\eta^\star$ is feasible while maintaining the objective value to zero.

Case 2: $\min_\theta \|f^\star - F_\phi \theta\|^2_{D_{\eta^\star}} = 0$ Take any sub-optimal (x, a). We shall prove that $z_\phi(x, a)^\mathsf{T} \theta^\star_{\eta^\star}(\phi) > 0$ and $z_\phi(x, a) \in \mathrm{Im}(V_{\eta^\star}(\phi))$ imply that $M\eta^\star$ is feasible for some sufficiently large M. We have

$$\mathcal{I}_{M\eta^\star}(f^\star, \phi, x, a) = \frac{(z_\phi(x, a)^\mathsf{T} \theta^\star_{\eta^\star}(\phi))^2}{2\|z_\phi(x, a)\|^2_{V_{M\eta^\star}(\phi)^\dagger}}.$$

Let (U, Σ, V) be the SVD of $D^{1/2}_{M\eta^\star} F_\phi$. Note that

$$\|z_\phi(x, a)\|^2_{V_{M\eta^\star}(\phi)^\dagger} = \|z_\phi(x, a)\|^2_{(F_\phi^\mathsf{T} D_{M\eta^\star} F_\phi)^\dagger} = \|z_\phi(x, a)\|^2_{((D^{1/2}_{M\eta^\star} F_\phi)^\mathsf{T} D^{1/2}_{M\eta^\star} F_\phi)^\dagger}$$

$$= \|z_\phi(x, a)\|^2_{V(\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T}}$$

$$= z_\phi(x, a)^\mathsf{T} V (\Sigma^\mathsf{T}\Sigma)^\dagger V^\mathsf{T} z_\phi(x, a).$$

Suppose that $V_{M\eta^\star}(\phi)$ has rank $d' \le d_\phi$ and that the singular values in Σ are sorted in non-increasing order. It is easy to see that the first d' singular values of Σ are $\sqrt{M}\sigma_1, \dots, \sqrt{M}\sigma_{d'}$, where $\{\sigma_i\}_{i\in[d']}$ are the first d' singular values of $D^{1/2}_{\eta^\star} F_\phi$ and are all strictly positive (this is because $\mathrm{rank}(V_{M\eta^\star}(\phi)) = \mathrm{rank}(V_{\eta^\star}(\phi))$). All other singular values are zero. Thus, we have

$$\|z_\phi(x, a)\|^2_{V_\eta(\phi)^\dagger} = \sum_{i=1}^{d'} \frac{[V^\mathsf{T} z_\phi(x, a)]^2_i}{M\sigma_i^2}.$$

Recall that $\mathrm{Im}(V_{M\eta^\star}(\phi)) = \mathrm{Im}(V_{\eta^\star}(\phi)) = \mathrm{span}(v_1, \dots, v'_d)$, where $\{v_i\}_{i\in[d']}$ denote the columns of V associated with a non-zero singular value in Σ. Since $z_\phi(x, a) \in \mathrm{Im}(V_{\eta^\star}(\phi))$, the vector $V^\mathsf{T} z_\phi(x, a)$ has some non-zero element among the first d' components and all zeros in the remaing $d_\phi - d'$. This implies that we can set M to

$$M = \frac{2 \sum_{i=1}^{d'} \frac{[V^\mathsf{T} z_\phi(x,a)]^2_i}{\sigma_i^2}}{(z_\phi(x, a)^\mathsf{T} \theta^\star_{\eta^\star}(\phi))^2},$$

which implies that $\mathcal{I}_{M\eta^\star}(f^\star, \phi, x, a) \ge 1$ and thus $M\eta^\star$ is feasible. Overall we proved that there exists a sufficiently large M such that $M\eta^\star$ satisfy all the constraints while achieving an objective value of zero. This concludes the proof. $\square$

C.7   Proof of Corollary 2

The proof is trivial from Proposition 7: if there exists a realizable non-HLS representation ϕ, we have by definition that $\min_{\theta \in \mathbb{R}^{d_\phi}} \|f^\star - F_\phi \theta\|^2_{D_\eta} = 0$ for any η, while there exist $x \in \mathcal{X}, a \neq \pi^\star_{f^\star}(x)$ such that $\phi(x, a) \in \mathrm{Im}(V_{\eta^\star}(\phi))$. This violates the conditions of Proposition 7, hence $\mathcal{C}(f^\star, \mathcal{F}_\Phi) > 0$. $\square$

D   ADDITIONAL RESULTS ON NESTED FEATURES

While learning with nested features Φ cannot reduce the instance-dependent complexity with respect to learning with ϕ_N alone, when $d_{i^\star} \ll d_N$ one may still be wondering whether the complexity $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}})$ can really be much larger than $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_{i^\star}\}})$. After all, if $f^\star$ can be described by a low-dimensional representation ($\phi_{i^\star}$) that is nested into ϕ_N, it essentially means that ϕ_N is highly redundant and it might easily be compressed. We show that even in the single representation case, this is not possible in the worst case: there are instances where the regret of any uniformly good algorithm must scale with d_N even though $d_{i^\star} \ll d_N$.

Proposition 9. *Let $2 \le d_{\min} < d_{\max} \le X(A-1)$ and $2 \le N \le d_{\max} - d_{\min} + 1$. There exists an instance $f^\star : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ with minimum positive gap $\Delta_{\min}$ and a set $\Phi = \{\phi_1, \dots, \phi_N\}$ of N nested features with ϕ_N of dimension $d_{\max}$, $\phi_{i^\star} = \phi_1$ of dimension $d_{\min}$, and*

$$\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_{i^\star}\}}) \le \frac{2d_{\min}}{\Delta_{\min}}, \qquad \mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}}) \ge \frac{d_{\max}}{\Delta_{\min}}.$$

Proof. Let $\pi : \mathcal{X} \to \mathcal{A}$ be any policy and define an instance $f^\star$ as $f^\star(x, \pi(x)) = \Delta_{\min}$ for all $x \in \mathcal{X}$ and $f^\star(x, a) = 0$ for all $x \in \mathcal{X}, a \neq \pi(x)$. Clearly, $\pi^\star_{f^\star} = \pi$ and all sub-optimal context-arm pairs have gap $\Delta_{\min}$. Let $(x_1, a_1), \dots, (x_m, a_m)$ be an arbitrary enumeration of all $m = X(A - 1)$ sub-optimal context-arm pairs.

We define the set of N nested features $\Phi = \{\phi_1, \ldots, \phi_N\}$ as follows. For each $i \in [N]$, representation ϕ_i has dimension $d_i \in \{d_{\min}, \ldots, d_{\max}\}$ with $d_{\min} \geq 2$ and $d_{\max} \leq X(A-1) + 1$. In particular, $d_1 < d_2 < \cdots < d_N$, $d_1 = d_{\min}$, and $d_N = d_{\max}$. Representation ϕ_i is defined as

$$
\phi_i(x,a) = \begin{bmatrix} f^\star(x,a) \\ \Delta_{\min} \mathbb{1}\left((x,a) = (x_1, a_1)\right) \\ \vdots \\ \Delta_{\min} \mathbb{1}\left((x,a) = (x_{d_i-1}, a_{d_i-1})\right) \end{bmatrix}
$$

That is, all representations contain the true reward function in their first component, while the i-th representation contains indicators over the first $d_i - 1$ sub-optimal context-arm pairs in the remaining $d_i - 1$ components. Note that these representations are nested and all realizable (i.e., $i^\star = 1$). To prove the proposition, we only need to compute the complexities $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_1\}})$ and $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}})$. For all $i \in [N]$, is it easy to see from Case 2 of the proof of Proposition 3 that,

$$
\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_i\}}) = \sum_{j=1}^{d_i - 1} \frac{2}{\Delta_{f^\star}(x_j, a_j)} = \frac{2(d_i - 1)}{\Delta_{\min}}.
$$

Then, $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_1\}}) \leq \frac{2d_{\min}}{\Delta_{\min}}$ is trivial since $d_1 = d_{\min}$, while $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_N\}}) \geq \frac{d_{\max}}{\Delta_{\min}}$ follows since $d_n = d_{\max}$ and $d_{\max} \geq 3$. $\qquad\square$

Proposition 9 implies that achieving regret scaling with the dimensionality $d_{i^\star}$ of the smallest realizable representation is impossible in general. In the worst-case, any uniformly good algorithm must suffer a dependence on the dimensionality of the largest representation, regardless of the fact that a smaller realizable representation is nested into it.

E   WORST-CASE LOWER BOUND (PROOF OF THEOREM 3)

We start by proving two important lemmas. Then, we use them to prove a $\Omega(\sqrt{AT \log(|\Phi|)})$ lower bound (Theorem 6) and a $\Omega(\sqrt{dT \log(A)})$ lower bound (Theorem 7). Theorem 3 (formally stated in Theorem 8 below with precise constants) will then follow by combining these two.

Lemma 2. *Let $T \in \mathbb{N}_{>0}$ and denote by $\mathbb{P}_{f|x_{1:T}}, \mathbb{E}_{f|x_{1:T}}$ the probability and expectation operators over the full T-step history when learning problem f with some fixed algorithm conditioned on the sequence of contexts $x_1, \ldots, x_T$. For any couple of bandit instances $f_1, f_2 : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ and any $(\bar{x}, \bar{a}) \in \mathcal{X} \times \mathcal{A}$,*

$$
\mathbb{E}_{f_1|x_{1:T}}[N_T(\bar{x}, \bar{a})] \leq \mathbb{E}_{f_2|x_{1:T}}[N_T(\bar{x}, \bar{a})] + \frac{N_T(\bar{x})}{2} \sqrt{\sum_{x,a} \mathbb{E}_{f_2|x_{1:T}}[N_T(x,a)](f_1(x,a) - f_2(x,a))^2}.
$$

Proof. By Lemma 1 of Garivier et al. (2019), we have that, for any random variable Z_T with values in $[0,1]$ and that is measurable w.r.t. the T-step history conditioned on the context sequence $x_{1:T}$,

$$
\text{KL}(\mathbb{P}_{f_2|x_{1:T}}, \mathbb{P}_{f_1|x_{1:T}}) \geq \text{kl}(\mathbb{E}_{f_2|x_{1:T}}[Z_T], \mathbb{E}_{f_1|x_{1:T}}[Z_T]),
$$

where kl denotes the KL divergence between two bernoulli distributions with parameter $\mathbb{E}_{f_2|x_{1:T}}[Z_T]$ and $\mathbb{E}_{f_1|x_{1:T}}[Z_T]$, respectively. The left-hand side can be simplified as in Equation 8 of Garivier et al. (2019) by using the chain rule of KL divergences together with the fact that rewards are Gaussian with unit variance:

$$
\text{KL}(\mathbb{P}_{f_2|x_{1:T}}, \mathbb{P}_{f_1|x_{1:T}}) = \frac{1}{2} \sum_{x,a} \mathbb{E}_{f_2|x_{1:T}}[N_T(x,a)](f_1(x,a) - f_2(x,a))^2.
$$

On the other hand, by Pinsker's inequality,

$$
\text{kl}(\mathbb{E}_{f_2|x_{1:T}}[Z_T], \mathbb{E}_{f_1|x_{1:T}}[Z_T]) \geq 2(\mathbb{E}_{f_1|x_{1:T}}[Z_T] - \mathbb{E}_{f_2|x_{1:T}}[Z_T])^2.
$$

Choosing $Z_T = N_T(\bar{x}, \bar{a})/N_T(\bar{x})$ and rearranging concludes the proof. $\qquad\square$

Lemma 3 (Worst-case lower bound for unstructured contextual bandits). *Take any $X \geq 1, A \geq 2$. Let Π be the set of all deterministic policies mapping $[X]$ to $[A]$ and, for $\varepsilon > 0$, define $\mathcal{F}_\Pi := \{f_\pi : \pi \in \Pi\}$ where $f_\pi(x, a) := \varepsilon \mathbb{1}\left(a = \pi(x)\right)$ for all x, a. Then, choosing $\varepsilon = \sqrt{\frac{XA}{20T}}$, for any learning algorithm $\mathfrak{A}$ and $T \geq X$,*

$$\max_{f^\star \in \mathcal{F}_\Pi} \mathbb{E}^{\mathfrak{A}}_{f^\star}[R_T(f^\star)] \geq \frac{\sqrt{XAT}}{8\sqrt{5}}.$$

Proof. The proof follows a construction of Auer et al. (2002b). Let us consider a uniform context distribution. Take any $\pi \in \Pi$ and consider the corresponding instance $f_\pi \in \mathcal{F}_\Pi$. Note that, on such an instance, each action not prescribed by π is sub-optimal with a gap of ε. Then, the expected regret of any algorithm $\mathfrak{A}$ is

$$\mathbb{E}^{\mathfrak{A}}_{f_\pi}[R_T(f^\star)] = \varepsilon \sum_{x \in \mathcal{X}} \sum_{a \neq \pi(x)} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[N_T(x, a)] = \varepsilon \left(T - \sum_{x \in \mathcal{X}} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[N_T(x, \pi(x))] \right).$$

We can lower bound the maximum regret over $\mathcal{F}_\Pi$ as

$$\max_{f^\star \in \mathcal{F}_\Pi} \mathbb{E}^{\mathfrak{A}}_{f^\star}[R_T(f^\star)] = \max_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[R_T(f_\pi)] \geq \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[R_T(f_\pi)] = \varepsilon \left(T - \sum_{x \in \mathcal{X}} \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[N_T(x, \pi(x))] \right).$$

We shall thus focus on upper bounding the second term within brackets. Fix any context $\bar{x}$. Since the context distribution is independent of the specific instance,

$$\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[N_T(\bar{x}, \pi(\bar{x}))] = \mathbb{E}_{x_{1:T}} \left[\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi | x_{1:T}}[N_T(\bar{x}, \pi(\bar{x}))] \right].$$

Take any π. Let $\bar{f}_\pi(\bar{x}, a) = 0 \ \forall a$ and $\bar{f}_\pi(x, a) = f_\pi(x, a) \ \forall x \neq \bar{x}, a$. Applying Lemma 2 on the couple $(\bar{x}, \pi(\bar{x}))$ with $f_1 = f_\pi$ and $f_2 = \bar{f}_\pi$,

$$\mathbb{E}^{\mathfrak{A}}_{f_\pi | x_{1:T}}[N_T(\bar{x}, \pi(\bar{x}))] \leq \mathbb{E}_{\bar{f}_\pi | x_{1:T}}[N_T(\bar{x}, \pi(\bar{x}))] + \frac{N_T(\bar{x})\varepsilon}{2} \sqrt{\mathbb{E}_{\bar{f}_\pi | x_{1:T}}[N_T(\bar{x}, \pi(\bar{x}))]}.$$

Now take A policies $\pi_1, \ldots, \pi_A$ which are equal to π in all contexts except $\bar{x}$, where they play each a different action. Note that one of these policies must be π itself. Averaging both sides over these policies, and using Jensen's inequality,

$$\frac{1}{A} \sum_{a \in \mathcal{A}} \mathbb{E}^{\mathfrak{A}}_{f_{\pi_a} | x_{1:T}}[N_T(\bar{x}, \pi_a(\bar{x}))] \leq \frac{1}{A} \sum_{a \in \mathcal{A}} \mathbb{E}_{\bar{f}_\pi | x_{1:T}}[N_T(\bar{x}, \pi_a(\bar{x}))] + \frac{N_T(\bar{x})\varepsilon}{2} \sqrt{\frac{1}{A} \sum_{a \in \mathcal{A}} \mathbb{E}_{\bar{f}_\pi | x_{1:T}}[N_T(\bar{x}, \pi_a(\bar{x}))]}$$

$$= \mathbb{E}_{\bar{f}_\pi | x_{1:T}} \left[\frac{1}{A} \sum_{a \in \mathcal{A}} N_T(\bar{x}, a) \right] + \frac{N_T(\bar{x})\varepsilon}{2} \sqrt{\mathbb{E}_{\bar{f}_\pi | x_{1:T}} \left[\frac{1}{A} \sum_{a \in \mathcal{A}} N_T(\bar{x}, a) \right]}$$

$$= \frac{N_T(\bar{x})}{A} + \frac{N_T(\bar{x})^{3/2}\varepsilon}{2\sqrt{A}},$$

where the first equality holds since $\bar{f}_\pi$ does not depend on the choice of π_a.

Let $\Pi_{\mathcal{X} \setminus \{\bar{x}\}}$ be the set of all policies defined on the context space $\mathcal{X} \setminus \{\bar{x}\}$ and, for $\pi \in \Pi_{\mathcal{X} \setminus \{\bar{x}\}}$, let π_a be the corresponding policy extended to $\bar{x}$, where $\pi_a(\bar{x}) = a$. We have

$$\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi | x_{1:T}}[N_T(\bar{x}, \pi(\bar{x}))] = \frac{1}{|\Pi_{\mathcal{X} \setminus \{\bar{x}\}}|} \sum_{\pi \in \Pi_{\mathcal{X} \setminus \{\bar{x}\}}} \frac{1}{A} \sum_{a \in \mathcal{A}} \mathbb{E}^{\mathfrak{A}}_{f_{\pi_a} | x_{1:T}}[N_T(\bar{x}, \pi_a(\bar{x}))] \leq \frac{N_T(\bar{x})}{A} + \frac{N_T(\bar{x})^{3/2}\varepsilon}{2\sqrt{A}},$$

where in the last step we used the inequality derived above. Thus, by Jensen's inequality,

$$\frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{E}^{\mathfrak{A}}_{f_\pi}[N_T(\bar{x}, \pi(\bar{x}))] \leq \frac{\mathbb{E}_{x_{1:T}}[N_T(\bar{x})]}{A} + \frac{\mathbb{E}_{x_{1:T}}[N_T(\bar{x})^{3/2}]\varepsilon}{2\sqrt{A}} \leq \frac{T}{XA} + \frac{\sqrt{\mathbb{E}_{x_{1:T}}[N_T(\bar{x})^3]}\varepsilon}{2\sqrt{A}}.$$

With a uniform context distribution, $N_T(\bar{x}) \sim \text{Bin}(T, 1/X)$ and its third moment has the closed-form expression

$$\mathbb{E}_{x_{1:T}}[N_T(\bar{x})^3] = \frac{T}{X} + \frac{3T(T-1)}{X^2} + \frac{T(T-1)(T-2)}{X^3} \leq 5\frac{T^3}{X^3}$$

for $T \geq X$. Plugging everything back into our regret bound,

$$\max_{f^\star \in \mathcal{F}_\Pi} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)] \geq \varepsilon \left(T - \frac{T}{A} - \varepsilon\sqrt{\frac{5T^3}{4XA}} \right) \geq \varepsilon \left(\frac{T}{2} - \varepsilon\sqrt{\frac{5T^3}{4XA}} \right),$$

where we used $A \geq 2$. The proof is concluded by optimizing over ε. $\qquad\square$

Theorem 6. *Let $N, d \geq 1, A \geq 2$. There exist a context distribution and a set of d-dimensional representations Φ of size $|\Phi| = N$ over A arms such that, for any learning algorithm $\mathfrak{A}$ and $T \geq \lfloor \log(dN)/\log(A) \rfloor$,*

$$\max_{f^\star \in \mathcal{F}_\Phi} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)] \geq \frac{\sqrt{AT\lfloor \log(dN)/\log(A) \rfloor}}{8\sqrt{5}}.$$

Proof. Let $X = \lfloor \log(dN)/\log(A) \rfloor$. We shall build a set of d-dimensional representations Φ over X contexts and A arms such that $\mathcal{F}_\Phi \supseteq \mathcal{F}_\Pi$, where $\mathcal{F}_\Pi$ is the set of functions from $[X]$ to $[A]$ defined in Lemma 3. Then, from Lemma 3 we directly have

$$\max_{f^\star \in \mathcal{F}_\Phi} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)] \geq \max_{f^\star \in \mathcal{F}_\Pi} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star, \mathcal{F}_\Phi)] \geq \frac{\sqrt{XAT}}{8\sqrt{5}}.$$

Recall that $\mathcal{F}_\Pi := \{f_\pi : \pi \in \Pi\}$ where $f_\pi(x, a) := \varepsilon\mathbb{1}(a = \pi(x))$ for all x, a. Suppose we want to represent the A^X functions $\mathcal{F}_\Pi$ in $\mathcal{F}_\Phi$. Clearly, with a single representation ϕ we can represent at least d policies $\pi_1, \ldots, \pi_d$ by setting

$$\phi(x, a) = \begin{bmatrix} \mathbb{1}(a = \pi_1(x)) \\ \vdots \\ \mathbb{1}(a = \pi_d(x)) \end{bmatrix}.$$

Then, the corresponding functions $f_{\pi_1}, \ldots, f_{\pi_d}$ are realized by choosing parameters with value ε on a single component and zero on all the others. In total we have N feature maps, so that we can represent at least Nd functions. Thus, it is enough to have $A^X \leq Nd$ to guarantee $\mathcal{F}_\Phi \supseteq \mathcal{F}_\Pi$. Rearranging this condition, we find that $X = \lfloor \log(dN)/\log(A) \rfloor$ contexts are enough. This concludes the proof. $\qquad\square$

Theorem 7. *Let $N \geq 1, A \geq 4$ and $d \geq 12\log_2(A)$. There exist a context distribution and a set of d-dimensional representations Φ of size $|\Phi| = N$ over A arms such that, for any learning algorithm $\mathfrak{A}$ and $T \geq d/\log_2(A)$,*

$$\max_{f^\star \in \mathcal{F}_\Phi} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)] \geq \frac{\sqrt{Td\log_2(A)}}{16\sqrt{5}}.$$

Proof. To gain intuition, let us start from $A = 2$. Suppose d is even and consider a set $\mathcal{X} = \{x_1, \ldots, x_{d/2}\}$ of $X = d/2$ contexts. Consider a problem with 2 arms a_1, a_2 and a d-dimensional representation

$$\phi(x, a) = \begin{bmatrix} \mathbb{1}(x = x_1, a = a_1) \\ \mathbb{1}(x = x_1, a = a_2) \\ \vdots \\ \mathbb{1}(x = x_{d/2}, a = a_1) \\ \mathbb{1}(x = x_{d/2}, a = a_2) \end{bmatrix}.$$

It is clear that with ϕ we can represent any function from $\mathcal{X}$ to $\{a_1, a_2\}$. Let Π be the set of all 2^X policies mapping $\mathcal{X}$ to $\{a_1, a_2\}$. Then, $\mathcal{F}_{\{\phi\}} \supseteq \mathcal{F}_\Pi$, where $\mathcal{F}_\Pi$ is defined in Lemma 3. Thus, from Lemma 3, as far as $T \geq d/2$,

$$\max_{f^\star \in \mathcal{F}_{\{\phi\}}} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star, \mathcal{F}_{\{\phi\}})] \geq \frac{\sqrt{dT}}{8\sqrt{5}}.$$

Let us now extend this reasoning to $A \geq 2$ arms.

We use a construction inspired by He et al. (2022). Let us define a representation $\bar{\phi}$ of dimension $\lfloor d/\lfloor \log_2(A) \rfloor \rfloor$ for a 2-armed problem defined over $X = \lfloor \lfloor d/\lfloor \log_2(A) \rfloor \rfloor/2 \rfloor$ contexts and arms $\{\bar{a}_1, \bar{a}_2\}$ as

$$\forall x \in [X], a \in \{\bar{a}_1, \bar{a}_2\} : \bar{\phi}(x, a) = \begin{bmatrix} \mathbb{1}\,(x = x_1, a = \bar{a}_1) \\ \mathbb{1}\,(x = x_1, a = \bar{a}_2) \\ \vdots \\ \mathbb{1}\,(x = x_X, a = \bar{a}_1) \\ \mathbb{1}\,(x = x_X, a = \bar{a}_2) \end{bmatrix}.$$

For $a \in [2^{\lfloor \log_2(A) \rfloor}]$, let $b(a)$ denote the binary vector of size $\lfloor \log_2(A) \rfloor$ encoding arm a, and let $b_i(a)$ denote its i-th component (such that $b_1(a)$ is the least significant digit and viceversa for $b_{\lfloor \log_2(A) \rfloor}(a)$) Then, we define the feature map for our A-armed problem as

$$\forall x \in [X], a \in [2^{\lfloor \log_2(A) \rfloor}] : \phi(x, a) = \begin{bmatrix} \bar{\phi}(x, \bar{a}_{1+b_1(a)}) \\ \vdots \\ \bar{\phi}(x, a_{1+b_{\lfloor \log_2(A) \rfloor}(a)}) \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where the number of zeros is $d - \lfloor d/\lfloor \log_2(A) \rfloor \rfloor \lfloor \log_2(A) \rfloor$. If A is not a power of 2, for all remaining arms, we set $\phi(x, a) = 0$. Intuitively, ϕ encodes $\lfloor \log_2(A) \rfloor$ copies of the linear bandit problem defined by $\bar{\phi}$. Moreover, selecting an action $a \in [2^{\lfloor \log_2(A) \rfloor}]$ in the problem represented by ϕ is equivalent to selecting actions $\{\bar{a}_{1+b_i(a)}\}_{i \in [\lfloor \log_2(A) \rfloor]}$ in the $\lfloor \log_2(A) \rfloor$ copies of the problem represented by $\bar{\phi}$.

Now fix some parameter $\theta \in \mathbb{R}^d$ and split it into consecutive vectors $\{\theta_i\}_{i \in [\lfloor \log_2(A) \rfloor]}$ each of size $\lfloor d/\lfloor \log_2(A) \rfloor \rfloor$ (and disregard the remaining components). It is easy to see that, for any $x \in [X]$ and $a \in [2^{\lfloor \log_2(A) \rfloor}]$,

$$\max_{a' \in \mathcal{A}} \phi(x, a')^\mathsf{T} \theta - \phi(x, a)^\mathsf{T} \theta = \sum_{i=1}^{\lfloor \log_2(A) \rfloor} \left(\max_{a' \in \{\bar{a}_1, \bar{a}_2\}} \bar{\phi}(x, a')^\mathsf{T} \theta_i - \bar{\phi}(x, \bar{a}_{1+b_i(a)})^\mathsf{T} \theta_i \right).$$

That is, the sub-optimality gap of (x, a) in the A-armed instance (ϕ, θ) is equal to the sum of gaps of the "binarized arms" over the instances $\{(\bar{\phi}, \theta_i)\}_{i \in [\lfloor \log_2(A) \rfloor]}$. Note also that we can always choose θ such that the reward of optimal arms is strictly positive, so that the remaining $A - 2^{\lfloor \log_2(A) \rfloor}$ arms are sub-optimal.

For an instance $f^\star$ that is linear in ϕ and θ, let us rewrite the regret $\mathbb{E}^{\mathfrak{A}}_{f^\star}[R_T(f^\star)]$ in the more explicit form $\mathbb{E}^{\mathfrak{A}}_{\phi, \theta}[R_T(\phi, \theta)]$. Then, the derivation above implies that

$$\max_{\theta \in \mathbb{R}^d} \mathbb{E}^{\mathfrak{A}}_{\phi, \theta}[R_T(\phi, \theta)] \geq \max_{\theta \in \mathbb{R}^d} \sum_{i=1}^{\lfloor \log_2(A) \rfloor} \mathbb{E}^{\mathfrak{A}}_{\bar{\phi}, \theta_i}[R_T(\bar{\phi}, \theta_i)] = \sum_{i=1}^{\lfloor \log_2(A) \rfloor} \max_{\theta \in \mathbb{R}^{\lfloor d/\lfloor \log_2(A) \rfloor \rfloor}} \mathbb{E}^{\mathfrak{A}}_{\bar{\phi}, \theta}[R_T(\bar{\phi}, \theta)]$$

$$\geq \frac{\lfloor \log_2(A) \rfloor \sqrt{2T \lfloor \lfloor d/\lfloor \log_2(A) \rfloor \rfloor/2 \rfloor}}{8\sqrt{5}}$$

$$\geq \frac{\sqrt{T(d\lfloor \log_2(A) \rfloor - 3\lfloor \log_2(A) \rfloor^2)}}{8\sqrt{5}}$$

$$\geq \frac{\sqrt{Td \log_2(A)}}{16\sqrt{5}},$$

where in the second inequality we used Lemma 3 to lower bound the regret in each of the 2-armed instances using that $T \geq d/\log_2(A)$, exactly as we did in the initial example. In the third and fourth inequalities we simplified the expression using the conditions $A \geq 4$ and $d \geq 12\log_2(A)$.

Therefore, we proved that there exists a linear bandit problem (with a single representation) that satisfies the stated result. Clearly, the same applies to representation learning with $N > 1$ by simply ignoring the extra representations. $\qquad \square$

Theorem 8. *[Restatement of Theorem 3] Let $N \geq 1, A \geq 4$ and $d \geq 12 \log_2(A)$. There exist a context distribution and a set of d-dimensional representations Φ of size $|\Phi| = N$ over A arms such that, for any learning algorithm $\mathfrak{A}$ and $T \geq \max\{\lfloor \log(dN)/\log(A) \rfloor, d/\log_2(A)\}$,*

$$\max_{f^\star \in \mathcal{F}_\Phi} \mathbb{E}_{f^\star}^{\mathfrak{A}}[R_T(f^\star)] \geq \frac{\sqrt{T \left(d \log_2(A) + A \lfloor \log(dN)/\log(A) \rfloor\right)}}{32\sqrt{5}}.$$

Proof. This is easy by contradiction. Suppose the statement does not hold. Then, for any set of representations and context distribution, there exists an algorithm such that the maximum regret over the family is at most the stated quantity. Using the sub-additivity of the square root followed by upper bounding the sum of the resulting two terms by twice the maximum among them, we find that such an algorithm must violate either the lower bound of Theorem 6 or the one of Theorem 7. This is a contradiction. □

F THE FULLY-REALIZABLE CASE

We provide novel insights on the complexity of representation learning in the setting studied by Papini et al. (2021), where the agent knows that $f^\star$ is a linear function of *all* representations $\phi \in \Phi$. That is, we consider the set of instances

$$\mathcal{F}_\Phi^{\mathrm{FR}} := \left\{ f : \mathcal{X} \times \mathcal{A} \to \mathbb{R} \mid \forall \phi \in \Phi, \exists \theta \in \mathbb{R}^{d_\phi} : f(x,a) = \phi(x,a)^\mathsf{T}\theta \; \forall x, a \right\}. \tag{9}$$

Clearly, $\mathcal{F}_\Phi^{\mathrm{FR}} \subseteq \mathcal{F}_\Phi$ and thus learning with $\mathcal{F}_\Phi^{\mathrm{FR}}$ is not harder than learning with $\mathcal{F}_\Phi$. This is intuitive since the agent is given more prior knowledge about $f^\star$ itself.

F.1 Instance-dependent Lower Bound

The following result formally establishes the complexity of a representation learning problem $(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}})$.

Theorem 9. *Let $f^\star \in \mathcal{F}_\Phi^{\mathrm{FR}}$ be an instance with unique optimal policy. Then, the complexity $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}})$ of Theorem 4 is*

$$\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}}) = \inf_{\eta(x,a) \geq 0} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a) \Delta_{f^\star}(x,a) \quad \text{s.t.} \quad \mathcal{I}_\eta(f^\star, x, a) \geq 1 \quad x \in \mathcal{X}, a \neq \pi_{f^\star}^\star(x),$$

where

$$\mathcal{I}_\eta(f^\star, x, a) := \max_{\phi \in \Phi} \frac{\Delta_{f^\star}(x,a)^2 \mathbb{1}\left(z_\phi(x,a) \in \mathrm{Im}(V_\eta(\phi))\right)}{2\|\phi(x,a) - \phi(x, \pi_{f^\star}^\star(x))\|_{V_\eta(\phi)^\dagger}^2}.$$

Proof. It is easy to see that the constraint in Theorem 4 decomposes into $X(A-1)$ constraints, one for each sub-optimal context-arm pair. In particular, the constraint associated with $\bar{x} \in \mathcal{X}, \bar{a} \neq \pi_{f^\star}^\star(\bar{x})$ is that the following quantity is larger than one:

$$\inf_{f \in \mathcal{F}_\Phi^{\mathrm{FR}}} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a) \mathrm{KL}_{x,a}(f^\star, f) \quad \text{s.t.} \quad f(\bar{x}, \bar{a}) > f(\bar{x}, \pi_{f^\star}^\star(\bar{x})). \tag{10}$$

Clearly, $\mathcal{F}_\Phi^{\mathrm{FR}} = \cap_{\phi \in \Phi} \mathcal{F}_{\{\phi\}}$. This implies that the infimum over the former set is equal to the maximum of the infima over the latter sets. This implies that the quantity above is

$$\mathcal{I}_\eta(f^\star, \bar{x}, \bar{a}) := \max_{\phi \in \Phi} \left\{ \min_{\theta \in \mathbb{R}^{d_\phi}} \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} \eta(x,a) \left(f^\star(x,a) - \phi(x,a)^\mathsf{T}\theta\right)^2 \quad \text{s.t.} \quad \phi(\bar{x}, \bar{a})^\mathsf{T}\theta \geq \phi(\bar{x}, \pi_{f^\star}^\star(\bar{x}))^\mathsf{T}\theta \right\}.$$

The inner problem is a minimization over a single half-space (the same as the one we compute in the realizable single-representation setting). Using Lemma 1 while noting that $f^\star(x,a) \in \mathcal{F}_{\{\phi\}}$ for all $\phi \in \Phi$,

$$\mathcal{I}_\eta(f^\star, \bar{x}, \bar{a}) := \max_{\phi \in \Phi} \frac{\Delta_{f^\star}(x,a)^2 \mathbb{1}\left(z_\phi(x,a) \in \mathrm{Im}(V_\eta(\phi))\right)}{2\|\phi(x,a) - \phi(x, \pi_{f^\star}^\star(x))\|_{V_\eta(\phi)^\dagger}^2}.$$

□

F.2 Complexity of Representation Learning

We provide a series of results to characterize the complexity of representation learning in the fully-realizable setting. In particular, we show that the problem is significantly easier than in our general setting (Assumption 1).

Fully-realizable representation learning is never harder than learning with a given representation

Proposition 10. *For any Φ such that $f^\star \in \mathcal{F}_\Phi^{\mathrm{FR}}$, $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}}) \leq \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}}) \leq \mathcal{C}(f^\star, \mathcal{F}_\Phi)$.*

Proof. The second inequality is proved by Proposition 2 while noting that $f^\star \in \mathcal{F}_{\{\phi\}}$ for all $\phi \in \Phi$. The first one is an immediate consequence of Theorem 9: it is sufficient to lower bound the maximum over ϕ in each constraints using a single representation. $\square$

Remark 2. *An immediate consequence of this result is that representation learning in the fully-realizable setting is never harder than a CLB with any of the representations in Φ. This is in striking contrast with the general setting of Assumption 1, where representation learning is never easier than learning with any representation in Φ. The intuition from Theorem 9 is that, in the fully-realizable setting, we are allowed to choose a different representation for each x, a in order to facilitate satisfying the exploration constraints, while in Theorem 1 we have one independent constraint for each representation.*

Fully-realizable representation learning can be much easier than learning with a given representation We present an example inspired by Lattimore and Szepesvári (2017). In our context, learning with each single representation yields a dependence on the minimum gap, while representation learning in the fully-realizable setting does not.

Proposition 11. *For any $\varepsilon > 0$, there exist an instance $f^\star$, a universal constant c, and a set of representations Φ such that $f^\star \in \cap_{\phi \in \Phi} \mathcal{F}_{\{\phi\}}$ and $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}}) \leq c$, while $\min_{\phi \in \Phi} \mathcal{C}(f^\star, \mathcal{F}_{\{\phi\}}) \geq c/\varepsilon$.*

Proof. Let us consider a finite-armed (non-contextual) bandit problem with 4 arms. The mean-reward vector is $f^\star = (1, 1 - \varepsilon, 1 - \varepsilon, 0)^\mathsf{T}$. We have two realizable representations ϕ_1, ϕ_2 of dimension $d = 3$ defined as

$$\phi_1(a_1) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \phi_1(a_2) = \begin{bmatrix} 1 - \varepsilon \\ \varepsilon \\ 0 \end{bmatrix} \quad \phi_1(a_3) = \begin{bmatrix} 0 \\ 0 \\ 1 - \varepsilon \end{bmatrix} \quad \phi_1(a_4) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\phi_2(a_1) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \phi_2(a_2) = \begin{bmatrix} 1 - \varepsilon \\ 0 \\ 0 \end{bmatrix} \quad \phi_2(a_3) = \begin{bmatrix} 0 \\ \varepsilon \\ 1 - \varepsilon \end{bmatrix} \quad \phi_2(a_4) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The parameter realizing $f^\star$ is in both cases $\theta = (1, 0, 1)^\mathsf{T}$.

We start by computing the lower bound in the FR representation learning setting. We shall look for an upper bound to the optimal value $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}})$ which does not scale by $1/\varepsilon$. Let us choose an allocation η for which $\eta(a_1) = M$ (some very large quantity), $\eta(a_2) = \eta(a_3) = 0$. We need to find the required number of pulls to a_4. Since we are looking for an upper bound to the optimal value, it is enough to find $\eta(a_4)$ such that η satisfies the constraints in Theorem 9 for some specific representations (possibly different for different sub-optimal arms). We choose ϕ_1 for a_2 and a_4, and ϕ_2 for a_3. These yield the constraints

$$\frac{\Delta_{f^\star}(a_2)^2}{2\|\phi_1(a_2) - \phi_1(a_1)\|^2_{V_\eta(\phi_1)^\dagger}} = \frac{\varepsilon^2}{2\|(-\varepsilon, \varepsilon, 0)^\mathsf{T}\|^2_{V_\eta(\phi_1)^\dagger}} \geq 1.$$

$$\frac{\Delta_{f^\star}(a_3)^2}{2\|\phi_2(a_3) - \phi_2(a_1)\|^2_{V_\eta(\phi_2)^\dagger}} = \frac{\varepsilon^2}{2\|(0, \varepsilon, -\varepsilon)^\mathsf{T}\|^2_{V_\eta(\phi_2)^\dagger}} \geq 1.$$

$$\frac{\Delta_{f^\star}(a_4)^2}{2\|\phi_1(a_4) - \phi_1(a_1)\|^2_{V_\eta(\phi_1)^\dagger}} = \frac{1}{2\|(-1, 1, 0)^\mathsf{T}\|^2_{V_\eta(\phi_1)^\dagger}} \geq 1.$$

Note that, by our choice of η, $V_\eta(\phi_1)$ and $V_\eta(\phi_2)$ are diagonal matrices with diagonal elements $(M, \eta(a_4), 0)$ and $(0, \eta(a_4), M)$, respectively. Therefore, the constraints above reduce to

$$\frac{\varepsilon^2}{\varepsilon^2/M + \varepsilon^2/\eta(a_4)} \geq 2, \quad \frac{\varepsilon^2}{\varepsilon^2/\eta(a_4) + \varepsilon^2/M} \geq 2, \quad \frac{1}{1/M + 1/\eta(a_4)} \geq 2.$$

Letting M go to infinity (which does not alter the objective value), we find that $\eta(a_4) \geq 2$ suffices. Therefore, we proved that the allocation $\eta = (\infty, 0, 0, 2)^\mathsf{T}$ is feasible, and thus the optimal value is bounded by $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}}) \leq 2$.

We now show that the regret when learning with each of the single representations scales at least by $1/\varepsilon$. Let us do it for ϕ_1. For ϕ_2 the argument will be the same since the two representations are equal up to a permutation of the first and third component.

Clearly, since we want to lower bound the optimal value $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_1\}})$, we can drop all constraints but the one associated with a_3, i.e.,

$$\frac{\Delta_{f^\star}(a_3)^2}{2\|\phi_1(a_3) - \phi_1(a_1)\|_{V_\eta(\phi_1)^\dagger}^2} = \frac{\varepsilon^2}{2\|(-1, 0, 1 - \varepsilon)^\mathsf{T}\|_{V_\eta(\phi_1)^\dagger}^2} \geq 1.$$

Let us set once again $\eta(a_1) = M$ (some very large value). Since the constraint associated with a_3 requires to make the feature norm of $\phi_1(a_3) - \phi_1(a_1) = (-1, 0, 1 - \varepsilon)^\mathsf{T}$ small, clearly both a_2 and a_4 do not serve to this purpose (they do not cover the third dimension). So the optimal strategy must have $\eta(a_2) = \eta(a_4) = 0$. The matrix $V_\eta(\phi_1)^\dagger$ is then diagonal with elements $(M, 0, (1 - \varepsilon)^2 \eta(a_3))$. Thus, the constraint reduces to

$$\frac{\varepsilon^2}{1/M + 1/\eta(a_3)} \geq 2.$$

This implies that $\eta(a_3) \geq 2/\varepsilon^2$. Plugging this into the regret, recalling that action a_3 has gap ε, we obtain that $\mathcal{C}(f^\star, \mathcal{F}_{\{\phi_1\}}) \geq 2/\varepsilon$.

$\square$

F.3 Necessary and Sufficient Condition for Constant Regret

Proposition 12. *A necessary and sufficient condition for $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}}) = 0$ is that, for all $x \in \mathcal{X}, a \neq \pi_{f^\star}^\star(x)$, there exists $\phi \in \Phi$ such that $z_\phi(x, a) \in \mathrm{Im}(V_{\eta^\star}(\phi))$ (equiv. $\phi(x, a) \in \mathrm{Im}(V_{\eta^\star}(\phi)))$.*

Remark 3. *This result shows that the mixing HLS condition assumed by Papini et al. (2021) (see their Definition 1) is actually necessary for constant regret.*

Proof. Proving that the condition is necessary can be easily done by contradiction. If $\mathcal{C}(f^\star, \mathcal{F}_\Phi^{\mathrm{FR}}) = 0$, then a rescaling of $\eta^\star$ must be feasible. However, if for some sub-optimal (x, a) we have $z_\phi(x, a) \notin \mathrm{Im}(V_{\eta^\star}(\phi))$ for all ϕ, that would imply that any rescaling of $\eta^\star$ is actually infeasible, hence yielding a contradiction.

The proof that the condition is sufficient can be done as a simple extension of the one of Proposition 7. Simply take any sub-optimal (x, a) and show that a re-scaling of $\eta^\star$ is feasible by following exactly the same steps as in Proposition 7. $\square$

G USEFUL LINEAR ALGEBRA RESULTS

G.1 Singular value decomposition

We recall that the SVD of a real matrix $A \in \mathbb{R}^{n \times m}$ is a factorization of the form $A = U\Sigma V^\mathsf{T}$, with $U \in \mathbb{R}^{n \times n}$ orthogonal (i.e., such that $U^\mathsf{T}U = UU^\mathsf{T} = I$), $\Sigma \in \mathbb{R}^{n \times m}$ diagonal, and $V \in \mathbb{R}^{m \times m}$ orthogonal. Suppose A has rank $d \leq \min\{n, m\}$ and that the diagonal entries of Σ (i.e., the singular values of A) are in decreasing order ($\sigma_1 \geq \cdots \geq \sigma_{\min\{n,m\}} \geq 0$). We list some well-known properties of the SVD decomposition.

Properties

- The number of non-zero entries in Σ correspond to the rank of A.

- Let $u_1, \ldots, u_d$ be the columns of U (i.e., the left singular vectors) corresponding to non-zero singular values. Then $\mathrm{span}(u_1, \ldots, u_d) = \mathrm{Im}(A)$.

- Let $v_1, \ldots, v_d$ be the columns of V (i.e., the right singular vectors) corresponding to non-zero singular values. Then $\mathrm{span}(v_1, \ldots, v_d) = \mathrm{Row}(A)$.

- Let $v_{d+1}, \ldots, v_{\min\{n,m\}}$ be the columns of V (i.e., the right singular vectors) corresponding to zero singular values. Then $\mathrm{span}(v_{d+1}, \ldots, v_{\min\{n,m\}}) = \mathrm{Ker}(A)$.

Let $A \in \mathbb{R}^{n \times m}$ with $n \geq m$ and $\mathrm{rank}(A) = d$. Then,

- $\mathrm{rank}(A^T A) = \mathrm{rank}(A)$.

- $\mathrm{Im}(A^T A) = \mathrm{Row}(A) = \mathrm{span}(v_1, \ldots, v_d)$.

- $\mathrm{Ker}(A^T A) = \mathrm{Ker}(A) = \mathrm{span}(v_{d+1}, \ldots, v_{\min\{n,m\}})$.

G.2 Pseudo-inverse

We recall that the pseudo-inverse of a matrix $A \in \mathbb{R}^{n \times m}$ is defined as $A^\dagger = V \Sigma^\dagger U^\mathsf{T}$, where (U, Σ, V) is the SVD of A and $\Sigma^\dagger$ is a diagonal matrix with the inverse of the non-zero elements of Σ.

Properties

1. $A A^\dagger A = A$ and $A^\dagger A A^\dagger = A^\dagger$.

2. $(A^\mathsf{T} A)^\dagger A^T = A^\dagger$.

3. $(A^\dagger)^\mathsf{T} = (A^\mathsf{T})^\dagger$.

4. $(A A^\mathsf{T})^\dagger = (A^T)^\dagger A^\dagger$.

5. If either $A^\mathsf{T} A = I$ or $B B^\mathsf{T} = I$ or $A = B^\mathsf{T}$: $(AB)^\dagger = B^\dagger A^\dagger$.