Sparse Nonparametric Contextual Bandits

Hamish Flynn
Universitat Pompeu Fabra
hamishflynn.gm@gmail.com

Julia Olkhovskaya
Delft University of Technology
julia.olkhovskaya@gmail.com

Paul Rognon-Vael
Universitat Pompeu Fabra
paul.rognon@gmail.com

Abstract

We study the benefits of sparsity in nonparametric contextual bandit problems, in which the set of candidate features is countably or uncountably infinite. Our contribution is two-fold. First, using a novel reduction to sequences of multi-armed bandit problems, we provide lower bounds on the minimax regret, which show that polynomial dependence on the number of actions is generally unavoidable in this setting. Second, we show that a variant of the Feel-Good Thompson Sampling algorithm enjoys regret bounds that match our lower bounds up to logarithmic factors of the horizon, and have logarithmic dependence on the effective number of candidate features. When we apply our results to kernelised and neural contextual bandits, we find that sparsity enables better regret bounds whenever the horizon is large enough relative to the sparsity and the number of actions.

1 Introduction

The contextual bandit problem is a general model for sequential decision-making problems, in which at each step, a learner observes a context, plays an action in response to the context and then receives a reward. The goal of the learner is to maximise the reward accumulated over n rounds, which is usually measured by the regret with respect to playing the best action for each context. The contextual bandit problem has attracted a great deal of attention because it is a faithful model of many real-world problems, such as personalised advertising (Abe & Nakamura, 1999), personalised news recommendation (Li et al., 2010) and medical treatment (Durand et al., 2018). In many practical situations, the set of possible contexts is very large, and the learner must use some sort of function approximation to learn general patterns that apply to new contexts. Previous works have considered finite-dimensional linear function approximation (Abe & Long, 1999; Dani et al., 2008; Abbasi-Yadkori et al., 2011), nonparametric function approximation (Bubeck et al., 2008; Kleinberg et al., 2008; Srinivas et al., 2010; Valko et al., 2013; Slivkins, 2014) and wide neural networks (Zhou et al., 2020; Zhang et al., 2021; Kassraie & Krause, 2022). However, none of these approaches is entirely satisfactory. While linear methods lead to efficient estimation of the reward function, they typically only work well when one has considerable prior knowledge about the relationship between contexts, actions and rewards. In particular, the user is required to specify a (small) set of features such that the reward function is a linear combination of these features. Nonparametric methods are much more flexible, but they suffer from a curse of dimensionality. If the contexts are vectors, then the regret of a nonparametric contextual bandit algorithm can grow exponentially with the dimension of the contexts. Neural contextual bandit algorithms typically operate in the lazy regime (Jacot et al., 2018; Chizat et al., 2019), in which neural networks behave like kernel methods. As a result, these algorithms suffer from the same drawbacks as nonparametric methods. Can we achieve the best of

18th European Workshop on Reinforcement Learning (EWRL 2025).

both worlds: a contextual bandit algorithm that selects a small set of useful features from an infinite set of candidate features to achieve both flexibility and sample efficiency?

In this work, we study a class of contextual bandit problems that we call *sparse nonparametric contextual bandits*. Briefly, we assume the expected reward for each context-action pair is a linear combination of features selected from a set of infinitely many candidate features, for a flexible function modelling. We further assume a sparse structure opening the way to efficient estimation. We consider both the case of *countable sparsity*, where the set of candidate features is countable, and the case of *uncountable sparsity*, where the set is uncountable. We fully describe the class under study in Section 2.2.

Contributions. We consider sparse nonparametric contextual bandits with n rounds, K actions per round and sparsity s. In the uncountable sparsity model, we consider candidate features parameterised by a d-dimensional vector. Our contribution is as follows:

- First, using a novel reduction to sequences of multi-armed bandit problems, we establish lower bounds of order $\sqrt{Ksn}$ for countable sparsity and $\sqrt{Ksdn}$ for uncountable sparsity. These lower bounds show that it is not possible to achieve low worst-case regret when the number of actions is large relative to the number of rounds.
- Second, we propose an algorithm based on the Feel-Good Thompson Sampling (FGTS) algorithm (Zhang, 2022), which uses novel sparsity priors for nonparametric models. We prove that this algorithm enjoys regret bounds that nearly match our lower bounds, even when the sparsity s is unknown. When the sparsity is known, our results demonstrate that FGTS is minimax optimal, up to logarithmic factors of n (cf. Section 4).
- In addition, we identify regimes in which sparsity enables improved regret bounds in kernelised and neural contextual bandit problems.

Outline. The remainder of the paper is structured as follows. In Section 2, we formally describe the setting of sparse nonparametric contextual bandits and introduce some mild regularity conditions. We also show that our framework includes interesting settings that require feature learning, such as sparse kernelised contextual bandits and neural contextual bandits. In Section 3, we state our lower bounds on the minimax regret and outline their proofs. In Section 4, we describe FGTS with our sparsity priors, and we show that it enjoys regret bounds that closely match our lower bounds. Finally, in Section 5, we summarise our findings. The proofs and the discussions of related work, limitations and directions for future work are gathered in the Appendix. Additionally, in Appendix B.2, we use our upper and lower bounds on the minimax regret to identify regimes in which sparsity is helpful for regret minimisation in kernelised and neural contextual bandits.

Notation. For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the greatest integer that is less than or equal to x and $\lceil x \rceil$ is the least integer that is greater than or equal to x. For any positive integer d, $[d]$ is the set $\{1, \dots, d\}$. For positive integers a and m, $a \bmod m$ is defined to be the unique integer $r \in \{0, 1, \dots, m-1\}$, such that $a = qm + r$, for some non-negative integer q. For any $q \in [1, \infty]$, $d \in \mathbb{N}$ and $R > 0$, we let $\mathbb{B}_q^d(R) = \{\theta \in \mathbb{R}^d : \|\theta\|_q \leq R\}$ denote the d-dimensional ℓ_q-ball with radius R. For any set $\mathbb{T}$ and any $\epsilon > 0$, we use $\mathcal{M}(\mathbb{T}, \|\cdot\|, \epsilon)$ to denote the ϵ-packing number of $\mathbb{T}$ (w.r.t. the norm $\|\cdot\|$) and $\mathcal{N}(\mathbb{T}, \|\cdot\|, \epsilon)$ to denote the ϵ-covering number of $\mathbb{T}$.

2 Problem Setting

2.1 Contextual Bandits

We consider the following contextual bandit protocol, in which a learner interacts with an environment over a sequence of n rounds. At the start of each round $t \in [n]$, the environment reveals a context $X_t \in \mathcal{X}$. In response, the learner selects an action $A_t \in \mathcal{A}$ and observes a real-valued reward Y_t. We let $\mathcal{F}_t = \sigma(X_1, A_1, Y_1, \dots, X_t, A_t, Y_t)$ denote the σ-field generated by the interaction history between the learner and the environment up to the end of round t, and we introduce the shorthand $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t-1}, X_t, A_t]$.

We assume that the action set is $\mathcal{A} = [K]$, and that each context X_t is of the form $X_t = (X_{t,a})_{a \in \mathcal{A}}$, where each $X_{t,a}$ lies in some set $\mathcal{Z}$. For an arbitrary $x \in \mathcal{X}$ and $a \in [K]$, we use x_a to denote the element in $\mathcal{Z}$ corresponding to the context x and the action a. We allow the contexts to be selected by an adaptive adversary, which means that the environment can take the history $\mathcal{F}_{t-1}$ into account before selecting X_t. The rewards are assumed to be stochastic, and of the form $Y_t = f^*(X_t, A_t) + \epsilon_t$, where $f^* : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ is a fixed reward function and ϵ_t is zero-mean, conditionally sub-Gaussian noise, meaning $\mathbb{E}_t[\exp(\lambda\epsilon_t)] \leq \exp(\lambda^2/8)$ for all $\lambda \in \mathbb{R}$.[1] The goal of the learner is to minimise the expected cumulative regret, which is defined as

$$R_n(f^*) = \mathbb{E}\left[\sum_{t=1}^{n} \max_{a \in \mathcal{A}}\{f^*(X_t, a)\} - f^*(X_t, A_t)\right] . \tag{1}$$

The conditional distribution of the action A_t, conditioned on $\mathcal{F}_{t-1}$ and X_t is denoted by $\pi_t(\cdot|\mathcal{F}_{t-1}, X_t)$. We call the sequence $(\pi_t)_{t=1}^n$ the policy of a contextual bandit algorithm.

2.2 Sparse Nonparametric Contextual Bandits

In sparse nonparametric contextual bandits, the reward function f^* is assumed to be an unknown linear combination of s features that belong to a known set of infinitely many candidate features. The sparsity s may or may not be known in advance. We consider two notions of sparsity that we refer to as *countable sparsity* and *uncountable sparsity*. To describe them, we start by recalling the standard, parametric, sparse linear contextual bandit problem, and then we show how our framework extends it.

In sparse linear contextual bandits, the reward function can be expressed as a weighted sum $f^*(x, a) = \sum_{i=1}^{p} w_i^* \phi_i(x_a)$ of finitely many features $\phi_1, \dots, \phi_p$. It is assumed that the weight vector $w^* \in \mathbb{R}^p$ contains only s non-zero elements. A natural way to make this model more flexible is to express the reward function as the infinite weighted sum $f^*(x, a) = \sum_{i=1}^{\infty} w_i^* \phi_i(x_a)$, where w^* is now a parameter sequence, as opposed to a parameter vector. We say that such a reward function is s-sparse when $\|w^*\|_0 = s$, where s is finite and ideally small. That is, there exists a finite subset $S \subset \mathbb{N}$ of size s, such that for all $i \notin S$, $w_i^* = 0$. We refer to this as *countable sparsity*. We assume that $\|w^*\|_1 \leq 1$ and $\|\phi_i\|_\infty \leq 1$ for all $i \in \mathbb{N}$, which implies that $\|f^*\|_\infty \leq 1$.

One can view the features $\phi_1, \phi_2, \dots$ as a sequence of functions, each mapping $\mathcal{Z}$ to $\mathbb{R}$, or as a single function $\phi : \mathcal{Z} \times \mathbb{N} \to \mathbb{R}$ that maps any $x_a \in \mathcal{Z}$ and $i \in \mathbb{N}$ to the value $\phi_i(x_a)$. Adopting the latter view, we can further generalise this model by replacing the index $i \in \mathbb{N}$ with a continuous parameter $\theta \in \Theta \subset \mathbb{R}^d$. The reward function can now be written as $f^*(x, a) = \int_\Theta \phi(x_a, \theta) \mathrm{d}w^*(\theta)$, where w^* is a signed measure on Θ. We say that such a reward function is s-sparse if w^* is a discrete measure that can be written as a sum of s Dirac measures, where the i^{th} Dirac measure is weighted by w_i^* and centred at θ_i^*. That is, there exist $s \in \mathbb{N}$, $\theta_1^*, \dots, \theta_s^* \in \Theta$ and $w^* \in \mathbb{R}^s$, such that $f^*(x, a) = \sum_{i=1}^{s} w_i^* \phi(x_a, \theta_i^*)$. Since the set of candidate features is uncountable, we refer to this as *uncountable sparsity*. The set of all functions of the form $f(x, a) = \sum_{i=1}^{\infty} w_i \phi(x_a, \theta_i)$ contains every s-sparse reward function, so we will restrict our attention to functions of this form, where $w = (w_1, w_2, \dots)$ is a parameter sequence, and $\theta_1, \theta_2, \dots$ is a sequence of elements in Θ. We assume that $\|w^*\|_1 \leq 1$, $\|\theta_i^*\|_2 \leq 1$ for all $i \in [s]$ and $\|\phi\|_\infty \leq 1$, which implies that $\|f^*\|_\infty \leq 1$.

We introduce the following shared notation to describe the classes of reward functions that we consider. We use s and S to denote the sparsity and the support of f^*. For countable sparsity, the support of f^* is $S = \{i \in \mathbb{N} : w_i^* \neq 0\} \subseteq \mathbb{N}$. For uncountable sparsity, the support of f^* is $S = \{\theta_1^*, \dots, \theta_s^*\}$. We use m and M to denote the sparsity and the support of an arbitrary model. We use ν to denote the parameter(s) of interest in both types of sparsity. That is, for countable sparsity $\nu = w$, and for uncountable sparsity $\nu = (w, \theta_1, \dots, \theta_m)$. Depending on the type of sparsity, we define f_ν to be the function $f_\nu(x, a) = \sum_{i=1}^{\infty} w_i \phi_i(x_a)$ or $f_\nu(x, a) = \sum_{i=1}^{m} w_i \phi(x_a, \theta_i)$. For countable sparsity, we use the notations f_ν and f_w interchangeably.

2.3 Regularity Conditions

In our analysis, we make the following regularity assumptions. For countable sparsity, we assume that the feature map satisfies one of following uniform decay conditions.

[1] More precisely, this means that ϵ_t is conditionally $\frac{1}{2}$-sub-Gaussian. Our regret analysis applies to any sub-Gaussian parameter. We assume that it is $\frac{1}{2}$ so that Theorem E.1 is consistent with Theorem 1 in Zhang (2022).

Definition 2.1 (Uniform decay.)**.** We say that $(\phi_i)_{i=1}^{\infty}$ satisfies the polynomial decay condition if, for some $\beta > 1$, $\|\phi_i\|_\infty \leq i^{-\beta/2}$. We say that $(\phi_i)_{i=1}^{\infty}$ satisfies the exponential decay condition if, for some $\beta > 0$, $\|\phi_i\|_\infty \leq \exp(-i^\beta/2)$.[2]

As discussed in Section 2.4, this is a natural assumption for kernelised bandits with countable sparsity. More generally, smoothness assumptions of this form appear frequently in nonparametric statistics (Wasserman, 2006). When the features satisfy one of these decay conditions, we can define a notion of *effective dimension*. In particular, we define the effective dimension for sample size n as

$$d_{\text{eff}} := \min\{i \in \mathbb{N} : \forall j > i, \|\phi_j\|_\infty^2 \leq \tfrac{1}{n}\}. \tag{2}$$

This definition ensures that if we approximate $f_\nu(x, a) = \sum_{i=1}^{\infty} w_i \phi_i(x_a)$ by the function $\tilde{f}_\nu(x, a) = \sum_{i=1}^{d_{\text{eff}}} w_i \phi_i(x_a)$, then the squared approximation error is bounded by $1/n$. In kernelised bandits, which will be one of our main points of reference (cf. Section B.2), d_{eff} is usually defined to be the effective degrees of freedom of the kernel ridge estimate, which is a data-dependent quantity (see e.g. Valko et al. (2013); Calandriello et al. (2019)). However, both definitions of d_{eff} are equivalent in the sense that, in the worst case, the scaling in n of the effective degrees of freedom of the ridge estimate with the features $\phi_1, \phi_2, \ldots$ matches that of our d_{eff} quantity in (2). One can verify that for polynomial decay, $d_{\text{eff}} \leq n^{1/\beta}$, and for exponential decay, $d_{\text{eff}} \leq \log^{1/\beta}(n)$. For uncountable sparsity, we require that the feature map is uniformly Lipschitz continuous. We assume that the Lipschitz constant is 1 for the sake of simplicity.

Definition 2.2 (Uniform Lipschitz continuity)**.** We say that ϕ is uniformly Lipschitz if

$$\forall x \in \mathcal{X}, a \in [K], \theta, \theta' \in \Theta,\ |\phi(x_a, \theta) - \phi(x_a, \theta')| \leq \|\theta - \theta'\|_2.$$

As discussed in Section 2.4, this is a natural assumption for kernelised bandits with uncountable sparsity and neural bandits. Moreover, similar assumptions have been used in previous works on contextual bandits (Zhang, 2022; Neu et al., 2022, 2024).

2.4 Examples

We conclude this section by highlighting some specific instantiations of our framework.

Kernelised Bandits with Countable Sparsity. In kernelised contextual bandits, each context-action pair corresponds to a vector $x_a \in \mathcal{Z} \subset \mathbb{R}^p$, and the reward function is $f^*(x, a) = h^*(x_a)$, where $h^* : \mathcal{Z} \to \mathbb{R}$ is a function in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, with reproducing kernel $k : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$. If k is continuous and $\mathcal{Z}$ is a compact metric space, then k is called a Mercer kernel, and due to Mercer's theorem, the kernel function can be written as

$$k(z, z') = \sum_{i=1}^{\infty} \xi_i \varphi_i(z) \varphi_i(z'),$$

where $(\xi_i)_{i=1}^{\infty}$ and $(\varphi_i)_{i=1}^{\infty}$ are the (non-negative) eigenvalues and eigenfunctions of the kernel (cf. Section 12.3 in Wainwright (2019)). Moreover, $\mathcal{H}$ can be represented as

$$\mathcal{H} = \{h(z) = \textstyle\sum_{i=1}^{\infty} w_i \sqrt{\xi_i} \varphi_i(z) : \sum_{i=1}^{\infty} w_i^2 < \infty\},$$

and the squared RKHS norm is $\|h\|_{\mathcal{H}}^2 = \sum_{i=1}^{\infty} w_i^2$. We notice that if we define $\phi_i := \sqrt{\xi_i}\varphi_i$, and consider functions in $\mathcal{H}$ corresponding to s-sparse sequences w, then we find ourselves in a sparse nonparametric contextual bandit problem, with countable sparsity. Note that for the commonly-used Matérn and RBF kernels, the eigenfunctions can be uniformly bounded and the eigenvalues decay to 0 as i tends to ∞, which means the features ϕ_i will typically satisfy one of the uniform decay conditions in Definition 2.1.

[2] The factors of $\frac{1}{2}$ are to make these definitions consistent with the usual eigenvalue decay conditions for Mercer kernels (see for instance Vakili et al. (2021)). For positive constants C, C_1 and C_2, one can easily replace these conditions with $\|\phi_i\|_\infty \leq Ci^{-\beta/2}$ and $\|\phi_i\|_\infty \leq C_1 \exp(-C_2 i^\beta/2)$.

Kernelised Bandits with Uncountable Sparsity. The kernelised contextual bandit problem can also be modelled as a sparse nonparametric contextual bandit problem with uncountable sparsity. One can alternatively express the RKHS $\mathcal{H}$ as

$$\mathcal{H} = \overline{\{h(z) = \sum_{i=1}^{\infty} w_i k(z, z_i) : \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} w_i w_j k(z_i, z_j) < \infty\}},$$

where for any set A, $\overline{A}$ denotes the closure of A. In addition, the squared RKHS norm can be expressed as $\|h\|_{\mathcal{H}}^2 = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} w_i w_j k(z_i, z_j)$. If we set $\Theta = \mathcal{Z}$, $\phi(z, \theta) = k(z, z')$ and assume that the function h^* is s-sparse, meaning $h^*(z) = \sum_{i=1}^{s} w_i^* k(z, z_i^*)$, then we find ourselves in a sparse nonparametric contextual bandit problem with uncountable sparsity (assuming $\mathcal{Z}$ is uncountable). The Matérn and RBF kernel functions are both Lipschitz, which means that the uniform Lipschitz continuity property is satisfied for these kernels.

Neural Bandits. For our final example, we consider a class of contextual bandit problems in which the reward function is a single-layer neural network, of possibly unknown width. Each context-action pair corresponds to a vector $x_a \in \mathcal{Z} \subset \mathbb{R}^p$. The reward function can be written as the infinite-width neural network

$$f^*(x, a) = \sum_{i=1}^{\infty} w_i^* \sigma(\langle \theta_i^*, x_a \rangle)\,,$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is an activation function and $\theta_i^* \in \Theta$. If we assume that f^* is s-sparse, and define $\phi(z, \theta) = \sigma(\langle \theta, z \rangle)$, then this is a sparse nonparametric contextual bandit problem with uncountable sparsity (assuming Θ is uncountable). If the norm of x_a is bounded for every $x \in \mathcal{X}$ and $a \in [K]$, and σ is a Lipschitz activation function, then the uniform Lipschitz continuity property is satisfied.

3 Minimax Lower Bounds

In this section, we establish lower bounds on the minimax regret for each type of sparsity. Both of our lower bounds are inspired by the lower bound for sparse linear bandits in Theorem 24.3 of Lattimore & Szepesvári (2020a), which exploits the fact that a sparse linear bandit problem can mimic a multi-task bandit problem. The main idea behind our lower bounds is that a sparse nonparametric contextual bandit problem, with either type of sparsity, can mimic a sequence of K-armed (non-contextual) bandit problems. In each of these sub-problems, there is a single good action with expected reward Δ and $K - 1$ bad actions with expected reward 0. The regret suffered in each sub-problem can be quantified using the lower bound for K-armed bandits in Exercise 24.1 of Lattimore & Szepesvári (2020a) (see also Lemma D.1). Due to space limitations, we only describe the reduction to sequences of K-armed bandits in detail in Appendix D.

The lower bound obtained through this reasoning is typically larger when the original problem is split into a larger number of K-armed bandits, each with a smaller horizon. At one extreme, if we impose no restrictions on the features and the reward function other than boundedness, then the sparse nonparametric contextual bandit problem can be reduced a sequence of n K-armed bandits, each with horizon 1, and with $\Delta = 1$. In this case, any algorithm is reduced to guessing in each round, and so the minimax regret is linear in n. If the features satisfy one of the regularity conditions in Section 2.3, then this constrains the largest possible value of Δ. In particular, the maximum value of Δ decreases as the number of K-armed bandits increases. One can still reduce the problem to a sequence of n K-armed bandit problems, but Δ would be so small that the regret of any algorithm is negligible. The worst-case reduction is to a sequence of fewer than n K-armed bandits, where some learning is possible, and so the minimax regret is sublinear. In the following subsections, we state our lower bounds for each type of sparsity.

3.1 Countable Sparsity

In the countable sparsity setting, we obtain the following lower bound.

Theorem 3.1. *Consider the sparse nonparametric contextual bandit problem with countable sparsity described in Section 2. Let $\mathcal{A} = [K]$ for some $K \geq 2$ and assume that the noise variables are standard Gaussian, i.e. $\epsilon_t \sim \mathcal{N}(0, 1)$. Suppose that for some $\beta > 1$ and some integer $m \geq s^{\beta+2}K^{\beta+1}$, $n = sm$. Then for any policy, there exists a sequence of contexts $x_1, \ldots, x_n$, a parameter* *sequence $w = (w_1, w_2, \dots)$ with $\|w\|_0 = s$, $\|w\|_1 = 1$ and a sequence of functions $(\phi_i)_{i=1}^{\infty}$ with $\|\phi_i\|_\infty \leq i^{-\beta/2}$, such that*

$$R_n(f_w) \geq \frac{1}{8}\sqrt{Ksn}.$$

Instead, suppose that for some $\beta > 0$ and some integer $m \geq \lceil 1/\beta \rceil s^2 K \exp(s^\beta K^{\beta \lceil 1/\beta \rceil})$, $n = sm$. Then for any policy, there exists a sequence of contexts $x_1, \dots, x_n$, a parameter sequence $w = (w_1, w_2, \dots)$ with $\|w\|_0 = s$, $\|w\|_1 = 1$ and a sequence of functions $(\phi_i)_{i=1}^{\infty}$ with $\|\phi_i\|_\infty \leq \exp(-i^\beta/2)$, such that

$$R_n(f_w) \geq \frac{1}{8}\sqrt{\max(1, 1/\beta)Ksn}.$$

Note that $R(f_w)$ denotes the expected regret when the reward function is f_w, where w is the difficult parameter sequence. The proof of this result can be found in Appendix D.2. Here, we provide some intuition about the manner in which the decay conditions constrain the number of K-armed bandits in the reduction. To reduce the problem to a sequence of ms K-armed bandits, for any $m \geq 1$, we require that the features satisfy $\|\phi_i\|_\infty = \Delta$ for all $i \in [sK^m]$. Combined with either the polynomial or exponential decay condition, this means that Δ must satisfy either $\Delta \leq (sK^m)^{-\beta/2}$ or $\Delta \leq \exp((sK^m)^{-\beta}/2)$. Thus, there is a trade-off between choosing a large number of switches and a large gap Δ. On the one hand, this lower bound does not rule out the possibility that there exist algorithms with regret bounds that depend polynomially on $s\log(d_{\text{eff}})$. In the absence of sparsity, one would expect that the minimax regret depends polynomially on d_{eff}. On the other hand, this lower bound has polynomial dependence on K, whereas in non-sparse linear or kernelised contextual bandits, the minimax regret depends only logarithmically on K. This suggests that sparsity is only helpful when the number of actions is sufficiently small (cf. Section B.2).

3.2 Uncountable Sparsity

In the uncountable sparsity setting, we obtain the following lower bound.

Theorem 3.2. *Consider the sparse nonparametric contextual bandit problem with uncountable sparsity described in Section 2. Let $\mathcal{A} = [K]$ for some $K \geq 2$ and let $n = sdm$ for some integer $m \geq s^{2+2/d}K^3$. Assume that the noise variables are standard Gaussian. For any policy, there exists a sequence of contexts $x_1, \dots, x_n \in \mathcal{X}$, parameters $w \in \mathbb{B}_1^s(1)$, $\theta_1, \dots, \theta_s \in \Theta \subset \mathbb{B}_2^d(1)$ and a uniformly Lipschitz continuous function ϕ, with $\|\phi\|_\infty \leq 1$, such that*

$$R_n(f_\nu) \geq \frac{1}{8}\sqrt{Ksdn}.$$

This lower bound has similar implications to the previous one. In particular, it does not rule out the possibility of an algorithm with a regret bound that is polynomial in sd, but it does suggest that sparsity is only helpful when the number of actions is sufficiently small. For the special case of $s = 1$, Theorem 3.2 provides a lower bound of order $\sqrt{Kdn}$ for the setting studied in Neu et al. (2022, 2024), where the reward function $f^*(x, a) = \phi(x_a, \theta^*)$ is a Lipschitz function of a d-dimensional parameter vector. The proof of Theorem 3.2 can be found in Appendix D.3. It uses a similar reduction to a sequence of K-armed bandits, though the number of switches is constrained in a different manner. For any given number of switches, the Lipschitz property is used to relate the maximum value of Δ to the packing number of the ball $\mathbb{B}_2^d(1)$.

4 Matching Upper Bounds

In this section, we present an algorithm with regret bounds that match the lower bounds in the previous section up to logarithmic factors of n. We first describe the algorithm, and then discuss the regret bounds for countable and uncountable sparsity separately.

4.1 Method

Our algorithm is a particular instance of the Feel-Good Thompson Sampling (FGTS) algorithm proposed by Zhang (2022). FGTS is a modification of the popular Thompson Sampling algorithm (Thompson, 1933) that achieves improved bounds on (frequentist) regret by more aggressively exploring promising actions. Our algorithm uses novel priors inspired by the PAC-Bayes literature on sparse regression to extend FGTS to sparse nonparametric contextual bandits.

To describe FGTS in detail, we must first introduce some notation. We let $a(\nu, x) = \arg\max_{a\in[K]} f_\nu(x, a)$ denote the optimal action for context x and reward function f_ν. We let $f_\nu(x) = \max_{a\in[K]}\{f_\nu(x, a)\} = f_\nu(x, a(\nu, x))$ denote the maximum of $f_\nu(x, a)$ with respect to the action. Similarly, we let $a^*(x) = \arg\max_{a\in[K]} f^*(x, a)$ and $f^*(x) = f^*(x, a^*(x))$. We use $p_1(\nu)$ to denote the prior distribution, which will be specified in the subsequent subsections. We consider the following negative log-likelihood for the reward, conditioned on ν, x and a:

$$L(\nu, x, a, y) = \eta(f_\nu(x, a) - y)^2 - \lambda f_\nu(x)\,. \tag{3}$$

The quadratic term $\eta(f_\nu(x, a) - y)^2$ corresponds to a Gaussian log-likelihood with mean $f_\nu(x, a)$ and variance $1/(2\eta)$. The Feel-Good exploration term, $\lambda f_\nu(x)$, favours parameters ν where the maximum of $f_\nu(x, a)$ w.r.t. the action is large. Finally, $\eta > 0$ and $\lambda > 0$ are tuning parameters, which we will set later. With the negative log-likelihood L and the prior p_1, the posterior p_t after $t-1$ rounds have been completed is

$$p_t(\nu) \propto \exp\Big(-\sum_{l=1}^{t-1} L(\nu, X_l, A_l, Y_l)\Big)p_1(\nu). \tag{4}$$

In each round t, FGTS draws a sample ν_t from the posterior p_t, and then plays the action $A_t = a(\nu_t, X_t)$. We give the pseudocode for FGTS in Algorithm 1.

Algorithm 1 Feel-Good Thompson Sampling

```
Input: prior $p_1$, parameters $\eta$, $\lambda$
for $t = 1, \dots, n$ do
    Observe context $X_t$,
    Draw $\nu_t \sim p_t$ according to (4),
    Select action $A_t = a(\nu_t, X_t)$,
    Observe reward $Y_t$.
end for
```

To bound the regret of FGTS, we use the decoupling technique developed by Zhang (2022). As long as the prior satisfies $\mathbb{P}_{\nu\sim p_1}[\max_{x,a} |f_\nu(x, a)| \le 1] = 1$, Theorem 1 of Zhang (2022) (see also Theorem E.1) states that

$$R_n(f^*) \le \frac{\lambda K n}{\eta} + 6\lambda n - \frac{1}{\lambda} Z_n\,,$$

where $Z_n := \mathbb{E}\log\mathbb{E}_{\nu\sim p_1}\exp(-\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t))$ is a log-partition function, and $\Delta L(\nu, x, a, y) := \eta[(f_\nu(x, a) - y)^2 - (f^*(x, a) - y)^2] - \lambda[f_\nu(x) - f^*(x)]$ is the logarithm of a likelihood ratio. The main technical contribution of our regret analysis is to derive new bounds on Z_n, which depend on the true (possibly unknown) sparsity $\|w^*\|_0$. These new bounds on Z_n may be of independent interest, since they can be used to prove regret bounds for sparse nonparametric regression (cf. chapters 2 and 3 in Cesa-Bianchi & Lugosi (2006)). Due to space limitations, we state and prove our bounds on Z_n in Appendix E.2 and Appendix F.2. In the following subsections, we describe the priors that we use and the regret bounds that we obtain for each type of sparsity.

4.2 Countable Sparsity

For countable sparsity, we use a subset selection prior that can be factorised into a discrete distribution $p_1(M)$ over subsets $M \subseteq \mathbb{N}$, and a conditional distribution $p_1(w|M)$ over parameter sequences $w \in \mathbb{R}^\infty$ with support M. In particular, we adapt the prior used in Section 3 of Alquier & Lounici

(2011), which is designed for finite-dimensional sparse linear models with a p-dimensional parameter vector. Alquier & Lounici (2011) use a prior $p_1(M)$ over subsets $M \subseteq [p]$ that can be factorised into a distribution $p_1(m)$ over subset sizes $m \in [p]$ and a conditional distribution $p_1(M \mid m)$ over subsets $M \subseteq [p]$ of size m. To penalise large subsets and express no preference between any two subsets of the same size, $p_1(m)$ is chosen to be $p_1(m) \propto 2^{-m}$ (with $p_1(0) = 0$) and $p_1(M \mid m)$ is chosen to be the uniform distribution over subsets of size m. The resulting distribution over subsets of $[p]$ is

$$p_1(M) = \sum_{m=1}^{p} p_1(m)p_1(M \mid m) = p_1(|M|)p_1(M \mid |M|) = \frac{2^{-|M|}}{\binom{p}{|M|}\sum_{m=1}^{p} 2^{-m}}\,,$$

for any non-empty $M \subseteq [p]$. The conditional distribution $p_1(w \mid M)$ over parameter vectors with support M is the uniform distribution on the set $\mathcal{W}_M := \{w : \|w\|_1 \leq 1, \text{ and } w_i = 0\,\forall i \notin M\}$. The resulting prior over parameter vectors w is $p_1(w) = \sum_{M\subseteq[p]} p_1(M)p_1(w \mid M)\,.$

A naïve way of extending this prior to the infinite-dimensional case would be to replace the sums over m from 1 to p with sums over m from 1 to infinity. However, this fails for two reasons. First, the power set of $\mathbb{N}$ is uncountable, so it is not clear that a sum over all subsets of $\mathbb{N}$ is well-defined. Second, for a fixed subset size m, the set of all subsets of $\mathbb{N}$ of size m is countably infinite, so a uniform distribution on this set cannot be defined. Fortunately, these problems can be fixed by exploiting the uniform decay condition in Definition 2.1. In particular, we restrict the support of the distribution $p_1(M)$ to subsets $M \subseteq [d_{\text{eff}}]$. Intuitively, if we include at least d_{eff} features, then the uniform decay condition ensures that the regret suffered by ignoring some features will be negligible compared to the regret suffered while estimating the best d_{eff}-dimensional approximation of w^*. This intuition is made rigorous in the proof of Theorem 4.1. We set

$$p_1(M) = \mathbb{I}\{M \subseteq [d_{\text{eff}}]\}\frac{2^{-|M|}}{\binom{d_{\text{eff}}}{|M|}\sum_{m=1}^{d_{\text{eff}}} 2^{-m}}\,,$$

for non-empty $M \subseteq [d_{\text{eff}}]$ and $p_1(\emptyset) = 0$, which penalises large subsets and assigns probability zero to any subset not contained within $[d_{\text{eff}}]$. For any subset $M \subseteq [d_{\text{eff}}]$, we choose the conditional distribution over parameter sequences with support M to be the uniform distribution $\mathcal{U}(\mathcal{W}_M)$ over $\mathcal{W}_M$, which has the density function

$$p_1(w \mid M) = \tfrac{|M|!}{2^{-|M|}}\mathbb{I}\{\|w\|_1 \leq 1\}\mathbb{I}\{w_i = 0 \ \ \forall i \notin M\}\,.$$

This reflects our assumption that $\|w^*\|_1 \leq 1$. The resulting prior over parameter sequences is

$$p_1(w) = \sum_{M\subseteq[d_{\text{eff}}]} p_1(M)p_1(w \mid M)\,. \tag{5}$$

For every w in the support of this prior, we have $\|w\|_1 \leq 1$. This means that, as required by Theorem E.1, we have $\mathbb{P}_{w\sim p_1}[\max_{x,a} |f_w(x, a)| \leq 1] = 1$. Our regret bound for FGTS in the countable sparsity setting is stated in the following theorem.

Theorem 4.1. *Consider FGTS with $\eta = 1/4$ and the prior p_1 defined in* (5). *The expected regret of FGTS with $\lambda \propto \sqrt{\log(d_{\text{eff}}n)/(Kn)}$ is at most*

$$R_n(f^*) = \mathcal{O}(\|w^*\|_0\sqrt{Kn\log(d_{\text{eff}}n)})\,.$$

Suppose that $s \geq \|w^*\|_0$ is a known upper bound on the sparsity. The expected regret of FGTS with $\lambda \propto \sqrt{s\log(d_{\text{eff}}n)/(Kn)}$ is at most

$$R_n(f^*) = \mathcal{O}(\sqrt{Ksn\log(d_{\text{eff}}n)})\,. \tag{6}$$

This regret bound shows that when either of the uniform decay conditions in Definition 2.1 is satisfied, FGTS with the prior p_1 is nearly minimax optimal. When the polynomial decay condition is satisfied, the effective dimension satisfies $d_{\text{eff}} \leq n^{1/\beta}$. Therefore, the regret bound for FGTS is of order at most $\sqrt{(1 + 1/\beta)Ksn\log(n)}$. Since $\beta > 1$, $1 + 1/\beta$ less than 2, so this matches the lower bound in Theorem 3.1 up to a factor of $\sqrt{\log(n)}$. When the exponential decay condition is satisfied, $d_{\text{eff}} \leq \log^{1/\beta}(n)$, and so the regret bound for FGTS is of order at most $\sqrt{\max(1, 1/\beta)Ksn\log(n)}$. Once again, this matches the lower bound in Theorem 3.1 up to a factor of $\sqrt{\log(n)}$.

4.3 Uncountable Sparsity

For uncountable sparsity, we use a prior that can be factorised into a discrete distribution $p_1(m)$ over the number of features m, and conditional distributions $p_1(w \mid m)$ and $p_1(\theta_1, \dots, \theta_m \mid m)$ on the parameters $w \in \mathbb{R}^m$ and $\theta_1, \dots, \theta_m \in \mathbb{R}^d$ given m. To penalise large numbers of features, we choose $p_1(m) \propto 2^{-m}$ for $m \in \mathbb{N}$. We set both $p_1(w \mid m)$ and $p_1(\theta_1, \dots, \theta_m \mid m)$ to be uniform distributions. In particular, we set $p_1(w \mid m)$ to be the uniform distribution $\mathcal{U}(\mathbb{B}_1^m(1))$ on the m-dimensional unit ℓ_1 ball $\mathbb{B}_1^m(1)$ to reflect our assumption that $\|w^*\|_1 \leq 1$. For each $m \in \mathbb{N}$, we factorise the conditional distribution over $\theta_1, \dots, \theta_m$ as $p_1(\theta_1, \dots, \theta_m \mid m) = \prod_{i=1}^m p_1(\theta_i)$. We choose $p_1(\theta)$ to be the uniform distribution $\mathcal{U}(\mathbb{B}_2^d(1))$ on the d-dimensional unit ℓ_2 ball $\mathbb{B}_2^d(1)$ to reflect our assumption that for each i, $\|\theta_i^*\|_2 \leq 1$. We then have

$$p_1(w \mid m) = \tfrac{m!}{2^{-m}} \mathbb{I}\{\|w\|_1 \leq 1\} \quad \text{and} \quad p_1(\theta) = \tfrac{\Gamma(d/2+1)}{\pi^{d/2}} \mathbb{I}\{\|\theta\|_2 \leq 1\} \,.$$

The resulting prior over parameters ν is

$$p_1(\nu) = \sum_{m=1}^{\infty} p_1(m) p_1(w \mid m) p_1(\theta_1, \dots, \theta_m \mid m) \,. \tag{7}$$

For every ν in the support of our prior, we have $\|w\|_1 \leq 1$, which means that the prior satisfies $\mathbb{P}_{\nu \sim p_1}[\max_{x,a} |f_\nu(x,a)| \leq 1] = 1$. Our regret bound for FGTS in the uncountable sparsity setting is stated in the following theorem.

Theorem 4.2. *Consider FGTS with $\eta = 1/4$ and the prior p_1 defined in (7). The expected regret of FGTS with $\lambda \propto \sqrt{d \log(n)/(Kn)}$ is at most*

$$R_n(f^*) = \mathcal{O}(\|w^*\|_0 \sqrt{Kdn \log(n)}) \,.$$

Suppose that $s \geq \|w^*\|_0$ is a known upper bound on the sparsity. The expected regret of FGTS with $\lambda \propto \sqrt{sd \log(n)/(Kn)}$ is at most

$$R_n(f^*) = \mathcal{O}(\sqrt{Ksdn \log(n)}) \,.$$

This regret bound shows that FGTS with the prior p_1 in (7) is nearly minimax optimal for the uncountable sparsity setting. In particular, the regret bound for a known sparsity matches the lower bound in Theorem 3.2 up to a factor of $\sqrt{\log(n)}$. The proof of Theorem can be found in Appendix F.

5 Discussion

In this work, we studied a new class of contextual bandit problems, called sparse nonparametric contextual bandits, which captures the challenge of simultaneously learning relevant features and minimising regret. Our main goal was to establish whether it is possible to exploit sparsity to obtain a flexible and sample-efficient contextual bandit algorithm. To this end, we proved upper and lower bounds on the minimax regret for this class of problems, which match up to logarithmic factors of n. Our findings are mixed. On the one hand, our lower bounds have polynomial dependence on the number of actions, which suggests that it is difficult to exploit sparsity when the number of actions is large. On the other hand, we showed that the Feel-Good Thompson Sampling algorithm, with suitable sparsity priors, enjoys regret bounds with mild dependence on the number of candidate features. When applied to sparse kernelised contextual bandits and neural contextual bandits (cf. Appendix B.2), we found that sparsity always enables better regret bounds, as long as the horizon and/or the dimension of the contexts is large enough relative to the sparsity and the number of actions. We discuss some limitations of our work, as well as some questions raised by our work, in Appendix C.

Acknowledgments

We would like to thank Gergely Neu for highlighting reductions from contextual bandits to online regression as a promising algorithmic framework for this setting and also for pointing out the multi-task lower bound in Theorem 24.3 of Lattimore & Szepesvári (2020a) as a starting point for deriving our lower bounds. We would like to thank Eugenio Clerico for discussions about the Beurling LASSO and Ludovic Schwartz for discussions about tuning the λ parameter in FGTS and related algorithms.

Hamish was funded by the European Research Council (ERC), under the European Union's Horizon 2020 research and innovation programme (grant agreement 950180). Julia was supported by the Netherlands Organization for Scientific Research (NWO) under grant number VI.Veni.232.068. Paul was supported by the Department of Research and Universities of the Government of Catalunya and the European Social Fund.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Abbasi-Yadkori, Y., Pal, D., and Szepesvari, C. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pp. 1–9. PMLR, 2012.

Abe, N. and Long, P. M. Associative reinforcement learning using linear probabilistic concepts. In *ICML*, pp. 3–11. Citeseer, 1999.

Abe, N. and Nakamura, A. Learning to optimally schedule internet banner advertisements. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 12–21, 1999.

Agarwal, A., Luo, H., Neyshabur, B., and Schapire, R. E. Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pp. 12–38. PMLR, 2017.

Alquier, P. and Lounici, K. PAC-Bayesian theorems for sparse regression estimation with exponential weights. *Electronic Journal of Statistics*, 5:127–145, 2011.

Ariu, K., Abe, K., and Proutière, A. Thresholded lasso bandit. In *International Conference on Machine Learning*, pp. 878–928. PMLR, 2022.

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

Azais, J.-M., De Castro, Y., and Gamboa, F. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.

Bach, F. Breaking the curse of dimensionality with convex neural networks. *The Journal of Machine Learning Research*, 18(1):629–681, 2017.

Bastani, H. and Bayati, M. Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294, 2020.

Bastani, H., Bayati, M., and Khosravi, K. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 67(3):1329–1349, 2021.

Belkin, M. Approximation beats concentration? an approximation view on inference with smooth radial kernels. In *Conference On Learning Theory*, pp. 1348–1361. PMLR, 2018.

Bengio, Y., Roux, N., Vincent, P., Delalleau, O., and Marcotte, P. Convex neural networks. *Advances in neural information processing systems*, 18, 2005.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. Simultaneous analysis of lasso and Dantzig selector. *Annals of Statistics*, 37(4):1705–1732, 2009.

Bodmann, B. G., Flinth, A., and Kutyniok, G. Compressed sensing for analog signals. *arXiv preprint arXiv:1803.04218*, 2018.

Bredies, K. and Pikkarainen, H. K. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

Bubeck, S., Stoltz, G., Szepesvári, C., and Munos, R. Online optimization in x-armed bandits. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L. (eds.), *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc., 2008. URL `https://proceedings.neurips.cc/paper_files/paper/2008/file/f387624df552cea2f369918c5e1e12bc-Paper.pdf`.

Bühlmann, P. and Van De Geer, S. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Calandriello, D., Carratino, L., Lazaric, A., Valko, M., and Rosasco, L. Gaussian process optimization with adaptive sketching: Scalable and no regret. In *Conference on Learning Theory*, pp. 533–557. PMLR, 2019.

Candès, E. and Fernandez-Granda, C. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

Candès, E. and Tao, T. Decoding by linear programming. *IEEE transactions on information theory*, 51(12):4203–4215, 2005.

Candès, E. and Tao, T. The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35:2313–2351, 2007.

Cesa-Bianchi, N. and Lugosi, G. *Prediction, learning, and games*. Cambridge university press, 2006.

Chakraborty, S., Roy, S., and Tewari, A. Thompson sampling for high-dimensional sparse linear contextual bandits. In *International Conference on Machine Learning*, pp. 3979–4008. PMLR, 2023.

Chen, B., Castro, R., and Krause, A. Joint optimization and variable selection of high-dimensional Gaussian processes. In *Proc. International Conference on Machine Learning (ICML)*, 2012.

Chen, S. S., Donoho, D. L., and Saunders, M. A. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

Chizat, L., Oyallon, E., and Bach, F. On lazy training in differentiable programming. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper_files/paper/2019/file/ae614c557843b1df326cb29c57225459-Paper.pdf`.

Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Cortes, C. Support-vector networks. *Machine Learning*, 1995.

Dalalyan, A. and Tsybakov, A. Aggregation by exponential weighting, sharp oracle inequalities and sparsity. *Machine Learning*, 72(1-2):39–61, 2008.

Dalalyan, A. and Tsybakov, A. Sparse regression learning by aggregation and Langevin monte-carlo. *Journal of Computer and System Sciences*, 78(5):1423–1443, 2012a. ISSN 0022-0000. doi: https://doi.org/10.1016/j.jcss.2011.12.023. URL `https://www.sciencedirect.com/science/article/pii/S0022000012000220`. JCSS Special Issue: Cloud Computing 2011.

Dalalyan, A. S. and Tsybakov, A. Mirror averaging with sparsity priors. *Bernoulli*, 18(3):914–944, 2012b.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *COLT*, volume 2, pp. 3, 2008.

De Castro, Y. and Gamboa, F. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and applications*, 395(1):336–354, 2012.

Djolonga, J., Krause, A., and Cevher, V. High-dimensional Gaussian process bandits. *Advances in neural information processing systems*, 26, 2013.

Donoho, D. L. and Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3): 425–455, 1994.

Donoho, D. L. and Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432):1200–1224, 1995.

Donoho, D. L. and Johnstone, I. M. Minimax estimation via wavelet shrinkage. *The annals of Statistics*, 26(3):879–921, 1998.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(2):301–337, 1995.

Durand, A., Achilleos, C., Iacovides, D., Strati, K., Mitsis, G. D., and Pineau, J. Contextual bandits for adapting treatment in a mouse model of de novo carcinogenesis. In *Machine learning for healthcare conference*, pp. 67–82. PMLR, 2018.

Duval, V. and Peyré, G. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

Foster, D. and Rakhlin, A. Beyond UCB: Optimal and efficient contextual bandits with regression oracles. In *International Conference on Machine Learning*, pp. 3199–3210. PMLR, 2020.

Foster, D., Agarwal, A., Dudík, M., Luo, H., and Schapire, R. Practical contextual bandits with regression oracles. In *International Conference on Machine Learning*, pp. 1539–1548. PMLR, 2018.

Foster, D. J. and Krishnamurthy, A. Efficient first-order contextual bandits: Prediction, allocation, and triangular discrimination. *Advances in Neural Information Processing Systems*, 34:18907–18919, 2021.

Foster, D. J., Gentile, C., Mohri, M., and Zimmert, J. Adapting to misspecification in contextual bandits. *Advances in Neural Information Processing Systems*, 33:11478–11489, 2020.

Foucart, S. and Rauhut, H. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.

Gardner, J., Guo, C., Weinberger, K., Garnett, R., and Grosse, R. Discovering and Exploiting Additive Structure for Bayesian Optimization. In Singh, A. and Zhu, J. (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1311–1319. PMLR, 20–22 Apr 2017.

Gerchinovitz, S. Sparsity regret bounds for individual sequences in online linear regression. In *Proceedings of the 24th Annual Conference on Learning Theory*, pp. 377–396. JMLR Workshop and Conference Proceedings, 2011.

Guedj, B. and Alquier, P. PAC-Bayesian estimation and prediction in sparse additive models. *Electronic Journal of Statistics*, 7:264–291, 2013. doi: 10.1214/13-EJS771. URL `https://doi.org/10.1214/13-EJS771`.

Hao, B., Lattimore, T., and Szepesvari, C. Adaptive exploration in linear contextual bandit. In *International Conference on Artificial Intelligence and Statistics*, pp. 3536–3545. PMLR, 2020a.

Hao, B., Lattimore, T., and Wang, M. High-dimensional sparse linear bandits. *Advances in Neural Information Processing Systems*, 33:10753–10763, 2020b.

Jacot, A., Gabriel, F., and Hongler, C. Neural tangent kernel: Convergence and generalization in neural networks. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper_files/paper/2018/file/5a4be1fa34e62bb8a6ec6b91d2462f5a-Paper.pdf`.

Jang, K., Zhang, C., and Jun, K.-S. Popart: Efficient sparse regression and experimental design for optimal sparse linear bandits. *Advances in Neural Information Processing Systems*, 35:2102–2114, 2022.

Kandasamy, K., Schneider, J., and Poczos, B. High dimensional Bayesian optimisation and bandits via additive models. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 295–304, Lille, France, 07–09 Jul 2015. PMLR.

Kassraie, P. and Krause, A. Neural contextual bandits without regret. In Camps-Valls, G., Ruiz, F. J. R., and Valera, I. (eds.), *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 240–278. PMLR, 28–30 Mar 2022. URL `https://proceedings.mlr.press/v151/kassraie22a.html`.

Kim, G.-S. and Paik, M. C. Doubly-robust lasso bandit. *Advances in Neural Information Processing Systems*, 32, 2019.

Kim, S.-J. and Oh, M.-h. Local anti-concentration class: Logarithmic regret for greedy linear contextual bandit. *Advances in Neural Information Processing Systems*, 37, 2024.

Kleinberg, R., Slivkins, A., and Upfal, E. Multi-armed bandits in metric spaces. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pp. 681–690, 2008.

Koltchinskii, V. and Yuan, M. Sparse recovery in large ensembles of kernel machines. In *Proceedings of COLT*, volume 69, 2008.

Kumar, S., Sarkar, P., Tian, K., and Zhu, Y. Spike-and-slab posterior sampling in high dimensions. *arXiv preprint arXiv:2503.02798*, 2025.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020a.

Lattimore, T. and Szepesvári, C. Solutions to selected exercises in bandit algorithms, September 2020b. URL `https://tor-lattimore.com/downloads/book/solutions.pdf`.

Leung, G. and Barron, A. R. Information theory and mixing least-squares regressions. *IEEE Transactions on information theory*, 52(8):3396–3410, 2006.

Levin, A., Weiss, Y., Durand, F., and Freeman, W. T. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 1964–1971. IEEE, 2009.

Li, K., Yang, Y., and Narisetty, N. N. Regret lower bound and optimal algorithm for high-dimensional contextual linear bandit. *Electronic Journal of Statistics*, 15(2):5652–5695, 2021.

Li, L., Chu, W., Langford, J., and Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

Li, W., Chen, N., and Hong, L. J. Dimension reduction in contextual online learning via nonparametric variable selection. *Journal of Machine Learning Research*, 24(136):1–84, 2023.

Lin, Y. and Zhang, H. H. Component selection and smoothing in multivariate nonparametric regression. *The Annals of Statistics*, 34(5):2272–2297, 2006.

Mutny, M. and Krause, A. Efficient high dimensional Bayesian optimization with additivity and quadrature fourier features. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Neu, G., Olkhovskaya, J., Papini, M., and Schwartz, L. Lifting the information ratio: An information-theoretic analysis of Thompson sampling for contextual bandits. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 9486–9498. Curran Associates, Inc., 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/file/3d84d9b523e6e82916d496e58761002e-Paper-Conference.pdf`.

Neu, G., Papini, M., and Schwartz, L. Optimistic information directed sampling. In Agrawal, S. and Roth, A. (eds.), *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 3970–4006. PMLR, 30 Jun–03 Jul 2024. URL `https://proceedings.mlr.press/v247/neu24a.html`.

Oh, M.-h., Iyengar, G., and Zeevi, A. Sparsity-agnostic lasso bandit. In *International Conference on Machine Learning*, pp. 8271–8280. PMLR, 2021.

Papini, M., Tirinzoni, A., Restelli, M., Lazaric, A., and Pirotta, M. Leveraging good representations in linear contextual bandits. In *International Conference on Machine Learning*, pp. 8371–8380. PMLR, 2021.

Poon, C., Keriven, N., and Peyré, G. The geometry of off-the-grid compressed sensing. *Foundations of Computational Mathematics*, 23(1):241–327, 2023.

Raskutti, G., J Wainwright, M., and Yu, B. Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *Journal of machine learning research*, 13(2), 2012.

Ravikumar, P., Lafferty, J., Liu, H., and Wasserman, L. Sparse additive models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(5):1009–1030, 2009.

Rigollet, P. and Tsybakov, A. Exponential Screening and optimal rates of sparse estimation. *The Annals of Statistics*, 39(2):731 – 771, 2011. doi: 10.1214/10-AOS854. URL `https://doi.org/10.1214/10-AOS854`.

Rolland, P., Scarlett, J., Bogunovic, I., and Cevher, V. High-dimensional Bayesian optimization via additive models with overlapping groups. In Storkey, A. and Perez-Cruz, F. (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 298–307. PMLR, 09–11 Apr 2018.

Rosset, S., Swirszcz, G., Srebro, N., and Zhu, J. ℓ_1 regularization in infinite dimensional feature spaces. In *Learning Theory: 20th Annual Conference on Learning Theory, COLT 2007, San Diego, CA, USA; June 13-15, 2007. Proceedings 20*, pp. 544–558. Springer, 2007.

Santin, G. and Schaback, R. Approximation of eigenfunctions in kernel-based spaces. *Advances in Computational Mathematics*, 42(4):973–993, 2016.

Shi, L., Huang, X., Feng, Y., and Suykens, J. Sparse kernel regression with coefficient-based lq-regularization. *Journal of Machine Learning Research*, 20, 2019.

Slivkins, A. Contextual bandits with similarity information. *Journal of Machine Learning Research*, 15:2533–2568, 2014.

Smola, A. and Bartlett, P. Sparse greedy Gaussian process regression. In Leen, T., Dietterich, T., and Tresp, V. (eds.), *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2000. URL `https://proceedings.neurips.cc/paper_files/paper/2000/file/3214a6d842cc69597f9edf26df552e43-Paper.pdf`.

Srinivas, N., Krause, A., Kakade, S., and Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proc. International Conference on Machine Learning (ICML)*, 2010.

Tang, G., Bhaskar, B. N., Shah, P., and Recht, B. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.

Thompson, W. R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. ISSN 00063444. URL `http://www.jstor.org/stable/2332286`.

Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Tipping, M. The relevance vector machine. *Advances in neural information processing systems*, 12, 1999.

Tirinzoni, A., Papini, M., Touati, A., Lazaric, A., and Pirotta, M. Scalable representation learning in linear contextual bandits with constant regret guarantees. *Advances in Neural Information Processing Systems*, 35:2307–2319, 2022.

Vakili, S., Khezeli, K., and Picheny, V. On information gain and regret bounds in Gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 82–90. PMLR, 2021.

Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. Finite-Time Analysis of Kernelised Contextual Bandits. In *Uncertainty in Artificial Intelligence*, 2013.

Vapnik, V. N. The support vector method. In *International conference on artificial neural networks*, pp. 261–271. Springer, 1997.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

Wang, X., Wei, M., and Yao, T. Minimax concave penalized multi-armed bandit model with high-dimensional covariates. In *International Conference on Machine Learning*, pp. 5200–5208. PMLR, 2018.

Wang, Z., Hutter, F., Zoghi, M., Matheson, D., and De Feitas, N. Bayesian optimization in a billion dimensions via random embeddings. *Journal of Artificial Intelligence Research*, 55:361–387, 2016.

Wasserman, L. *All of nonparametric statistics*. Springer Science & Business Media, 2006.

Williams, C. and Seeger, M. Using the Nyström method to speed up kernel machines. In Leen, T., Dietterich, T., and Tresp, V. (eds.), *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2000.

Wu, W., Yang, J., and Shen, C. Stochastic linear contextual bandits with diverse contexts. In *International Conference on Artificial Intelligence and Statistics*, pp. 2392–2401. PMLR, 2020.

Zhang, T. Feel-good Thompson sampling for contextual bandits and reinforcement learning. *SIAM Journal on Mathematics of Data Science*, 4(2):834–857, 2022.

Zhang, W., Zhou, D., Li, L., and Gu, Q. Neural Thompson sampling. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=tkAtoZkcUnm`.

Zhang, Y., Wainwright, M. J., and Jordan, M. I. Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In Balcan, M. F., Feldman, V., and Szepesvári, C. (eds.), *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 921–948, Barcelona, Spain, 13–15 Jun 2014. PMLR. URL `https://proceedings.mlr.press/v35/zhang14.html`.

Zhou, D., Li, L., and Gu, Q. Neural contextual bandits with UCB-based exploration. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11492–11502. PMLR, 13–18 Jul 2020. URL `https://proceedings.mlr.press/v119/zhou20a.html`.

A Related Work

A.1 Sparse Models

The sparsity priors that we use take inspiration from a line of work on exponentially weighted aggregates, which have previously been used to derive oracle inequalities for sparse linear regression (Leung & Barron, 2006; Dalalyan & Tsybakov, 2008, 2012a,b; Alquier & Lounici, 2011) and regret bounds for online sparse linear regression (Gerchinovitz, 2011). Several works have applied traditional approaches for sparse linear regression, such as the LASSO (Tibshirani, 1996), Basis Pursuit (Chen et al., 2001) and thresholding, to nonparametric regression settings that resemble the countable sparsity model that we consider. Examples include wavelet regression (Donoho & Johnstone, 1994, 1995, 1998; Donoho et al., 1995) and sparse additive models (Lin & Zhang, 2006; Koltchinskii & Yuan, 2008; Ravikumar et al., 2009; Raskutti et al., 2012). The uncountable sparsity model has been studied in supervised learning settings, under the name of convex neural networks (Bengio et al., 2005; Rosset et al., 2007; Bach, 2017). Several previous works have developed sparse kernel methods, such as the Support Vector Machine (Cortes, 1995; Vapnik, 1997), the Relevance Vector Machine (Tipping, 1999) and sparse Gaussian processes (Smola & Bartlett, 2000; Williams & Seeger, 2000). Many of these approaches use sparse estimators for the purpose of improved computational efficiency, though some sparse kernel methods exploit sparsity for improved sample efficiency (Shi et al., 2019).

A.2 Sparse Linear Bandits

The challenge of simultaneously learning relevant features and minimising regret in contextual bandits has been partially addressed by some previous works on sparse linear contextual bandits. Several works consider sparse linear contextual bandits, with p-dimensional feature vectors, which can be thought of as a finite-dimensional version of the problem that we consider. Abbasi-Yadkori et al. (2012) used an online-to-confidence-set conversion to develop an algorithm with a regret bound of order $\sqrt{spn}$. When the contexts are chosen by an adversary and the number of actions K is allowed to be large or infinite, there is a matching lower bound (Lattimore & Szepesvári, 2020a). Because of this negative result, it has become popular to study sparse linear contextual bandits with i.i.d. contexts drawn from a well-conditioned distribution. Here, well-conditioned means that, with high probability, the empirical context covariance matrix satisfies a compatibility condition (Bühlmann & Van De Geer, 2011), or one of a number of similar conditions studied in high-dimensional statistics. Under these conditions, various methods enjoy regret bounds that depend only logarithmically on the dimension p (Foster et al., 2018; Bastani & Bayati, 2020; Wang et al., 2018; Kim & Paik, 2019; Oh et al., 2021; Ariu et al., 2022; Chakraborty et al., 2023). In non-contextual sparse linear bandits, Hao et al. (2020b) and Jang et al. (2022) showed that the existence of a policy that collects well-conditioned data enables problem-dependent regret bounds with logarithmic dependence on p. However, these approaches cannot easily be extended to sparse nonparametric bandits, because the compatibility condition (as well as other similar conditions) can fail to hold at all in both the countable and uncountable sparsity models.

A.3 Contextual Bandits and Feature Selection

Several works discovered conditions on the context distribution and the feature representation under which optimistic (Hao et al., 2020a; Wu et al., 2020) or greedy (Bastani et al., 2021; Kim & Oh, 2024) contextual bandit algorithms are guaranteed to have constant or logarithmic (in n) regret. Subsequently, contextual bandit algorithms have been developed to identify these good feature representations while simultaneously minimising regret (Papini et al., 2021; Tirinzoni et al., 2022). These results are complementary to our own, which hold when the contexts are selected by an adversary.

A.4 Bandits With Low-Dimensional Structure

Various methods have been developed to exploit other forms of low-dimensional structure in bandit and Bayesian optimisation problems. Chen et al. (2012), Djolonga et al. (2013) and Wang et al. (2016) designed Bayesian optimisation algorithms for the setting where the reward function is a composition of a low-dimensional linear embedding and a function drawn from a Gaussian process. Li et al.

(2023) designed a contextual bandit algorithm for a similar setting, in which the reward function is a composition of a linear embedding and a Lipschitz function. Kandasamy et al. (2015), Gardner et al. (2017), Rolland et al. (2018) and Mutny & Krause (2018) developed Bayesian optimisation algorithms for (generalised) additive reward functions.

A.5 Compressed Sensing Off the Grid

The nonparametric contextual bandit problem with uncountable sparsity is related to a problem known as compressed sensing off the grid (Tang et al., 2013). This is also called an inverse problem in the space of measures (Bredies & Pikkarainen, 2013), or (blind) deconvolution when applied to signals (Levin et al., 2009). In this problem, the aim is to recover the underlying discrete measure in a sparse infinite-dimensional linear model, using as few measurements as possible. In the finite-dimensional compressed sensing problem, conditions on the measurements similar to the compatibility condition used in sparse linear bandits are sufficient to guarantee recovery via, for instance, the LASSO (see e.g. chapters 4-6 in Foucart & Rauhut (2013)). To guarantee recovery in the off-the-grid compressed sensing problem, several works have developed weaker and more refined conditions on the measurements (Duval & Peyré, 2015; Bodmann et al., 2018; Poon et al., 2023). These conditions ensure that the underlying sparse measure can be recovered using an infinite-dimensional formulation of the LASSO, known as the Beurling LASSO (De Castro & Gamboa, 2012).

B Application to Linear, Kernelised and Neural Bandits

B.1 Implications for Sparse Linear Contextual Bandits

Our results have interesting implications for particular cases of sparse linear contextual bandits. With a small modification, Theorem 3.1 also provides a lower bound of $\sqrt{Ksn}/8$ for sparse linear contextual bandits with p-dimensional feature vectors and $K \leq p/s$. This complements the existing lower bound of order $\sqrt{spn}$ (Lattimore & Szepesvári, 2020a), in which the difficult instance has $K \geq p$. When $K \leq p/s$, our Theorem 4.1 gives a matching upper bound up to a factor of $\sqrt{\log(n)}$.

In linear contextual bandits with adversarial contexts, p-dimensional feature vectors and K actions, SupLinRel (Auer, 2002) and SupLinUCB (Chu et al., 2011) both have regret bounds of order approximately $\sqrt{pn\log(K)}$. If the parameter vector is s-sparse (and s is known), then as discussed in Section A, there is an algorithm that achieves an upper bound of $\sqrt{spn}$, and a matching lower bound. This suggests that sparsity is helpful when the number of actions is large, relative to the sparsity, but not particularly helpful when p is large. When applied to this setting, where $d_{\text{eff}} = p$, Theorem 4.1 would give a regret bound of order $\sqrt{Ks\log(p)n}$. Roughly speaking, this is an improvement whenever the number of actions is less than the ratio of the total number of features and the number of useful features, so when $K \leq p/s$.

B.2 When Does Sparsity Help In Kernelised and Neural Bandits?

In this section, we use our upper and lower bounds from sections 3 and 4 to identify regimes in which sparsity is helpful for regret minimisation in kernelised and neural contextual bandits. For simplicity, we ignore polylogarithmic factors of n in this section. Our findings are summarised in Table B.2, for kernelised bandits with the Matérn kernel, Table B.2, for kernelised bandits with the RBF kernel, and Table B.2, for neural bandits.

Upper Bound	Regret	Sparsity Assumption
Valko et al. (2013)	$\mathcal{O}\big(n^{\frac{\nu+p}{2\nu+p}}\sqrt{\log(K)}\big)$	none
This paper (Theorem 4.1)	$\mathcal{O}(\sqrt{Ksn\log(n)})$	countable
This paper (Theorem 4.2)	$\mathcal{O}(\sqrt{Kspn})$	uncountable
Lower Bound		
This paper (Theorem 3.1)	$\Omega(\sqrt{Ksn})$	countable
This paper (Theorem 3.2)	$\Omega(\sqrt{Kspn})$	uncountable

Table 1: Regret upper and lower bounds for kernelised contextual bandits with the Matérn kernel.

Upper Bound	Regret	Sparsity Assumption
Valko et al. (2013)	$\mathcal{O}\big(\sqrt{n\log^p(n)\log(K)}\big)$	none
This paper (Theorem 4.1)	$\mathcal{O}(\sqrt{Kspn\log\log(n)})$	countable
This paper (Theorem 4.2)	$\mathcal{O}(\sqrt{Kspn})$	uncountable
Lower Bound		
This paper (Theorem 3.1)	$\Omega(\sqrt{Kspn})$	countable
This paper (Theorem 3.2)	$\Omega(\sqrt{Kspn})$	uncountable

Table 2: Regret upper and lower bounds for kernelised contextual bandits with the RBF kernel.

In kernelised contextual bandits with adversarial contexts (of dimension p) and K actions, SupKernelUCB (Valko et al., 2013) has a regret bound of order $\sqrt{d_{\text{eff}} n \log(K)}$, where d_{eff} depends on the choice of the kernel. If the Matérn kernel with smoothness parameter $\nu > 0$ is used, then the polynomial decay condition is satisfied with $\beta = (2\nu + p)/p$ (Santin & Schaback, 2016), and so $d_{\text{eff}} \leq n^{p/(2\nu+p)}$. If the RBF kernel is used, then the exponential decay condition is satisfied with $\beta = 1/p$ (Belkin, 2018), and so $d_{\text{eff}} \leq \log^p(n)$. In the countable sparsity model, Theorem 4.1 gives a regret bound of order $\sqrt{Ksn\log(d_{\text{eff}})}$. For the Matérn kernel, this translates to a bound of order $\sqrt{(p/(2\nu+p))Ksn\log(n)}$, which can be compared to the bound of $n^{\frac{\nu+p}{2\nu+p}}\sqrt{\log(K)}$ for SupKernelUCB. For the RBF kernel, the bound from Theorem 4.1 becomes $\sqrt{Kspn\log\log(n)}$, whereas the bound for SupKernelUCB is $\sqrt{n\log^p(n)\log(K)}$. With both of these kernels, we observe that when n is sufficiently large, approximately $n \geq (sK)^{(2\nu+p)/p}$ for the Matérn kernel and $n \geq \exp((Ksp)^{1/p})$ for the RBF kernel, sparsity enables better regret bounds. In the uncountable sparsity model, the dimension of θ is $d = p$. Theorem 4.2 gives a regret bound of order $\sqrt{Kspn}$ with both the Matérn and RBF kernels. Again, we observe that for sufficiently large n, approximately $n \geq (Ksp)^{(2\nu+p)/p}$ for the Matérn kernel and $n \geq \exp((Ksp)^{1/p})$ for the RBF kernel, sparsity enables better regret bounds.

In neural contextual bandits, it is often assumed that the reward function lies in the RKHS associated with an infinite-width neural tangent kernel (NTK). In this setting, with adversarial contexts and K actions, the SupNN-UCB algorithm (Kassraie & Krause, 2022) has a regret bound of order approximately $\sqrt{d_{\text{eff}} n \log(K)}$, where d_{eff} is the effective dimension of the NTK. If the contexts are p-dimensional vectors on the unit sphere, then for the NTK of a single-layer ReLU neural network, $d_{\text{eff}} \leq n^{1-1/p}$ (cf. Theorem 3.1 in Kassraie & Krause (2022)). In this setting, the regret bound of SupNN-UCB is of order $n^{1-1/(2p)}\sqrt{\log(K)}$. In the neural bandits example in Section 2.4, the reward function is assumed to be a single-layer neural network of possibly unknown width. If we use a ReLU activation function, then Theorem 4.2 gives a regret bound of order $\sqrt{Kspn}$. For sufficiently large n, approximately $n \geq (Ksp)^{1-1/p}$, sparsity enables a better regret bound.

Upper Bound	Regret	Sparsity Assumption
Kassraie & Krause (2022)	$\mathcal{O}\big(n^{1-\frac{1}{2p}}\sqrt{\log(K)}\big)$	none
This paper (Theorem 4.2)	$\mathcal{O}(\sqrt{Kspn})$	uncountable
Lower Bound		
This paper (Theorem 3.2)	$\Omega(\sqrt{Kspn})$	uncountable

Table 3: Regret upper and lower bounds for neural contextual bandits.

C Further Discussion

C.1 Computationally Efficient Algorithms

An important question is whether or not there exists a computationally efficient algorithm that has a (nearly) matching regret bound. For FGTS, this question boils down to whether or not one can sample from the posterior efficiently. In particular, if there is access to an oracle that samples from the posterior, then FGTS is oracle-efficient. We are required to sample from the posterior only once per round and the rest of the algorithm has time complexity which polynomial in K, s, n, d and the dimension of the contexts. Several works have demonstrated that Markov Chain Monte Carlo methods can be used to approximately sample from the sparse posteriors that appear in our analysis (Alquier & Lounici, 2011; Rigollet & Tsybakov, 2011; Guedj & Alquier, 2013). However, we are not aware of any way to sample from the posteriors in Section 4 with polynomial time complexity.

Several oracle-based contextual bandit algorithms, such as SquareCB (Foster & Rakhlin, 2020), are known to lead to sample-efficient and computationally efficient algorithms whenever the oracle is both sample-efficient and computationally efficient. For SquareCB in particular, this reduces the problem to that of finding a sample-efficient and computationally efficient algorithm for sparse nonparametric regression. However, it is known that even sparse linear regression exhibits an unfortunate trade-off between sample and computational efficiency (Zhang et al., 2014). If we require a polynomial-time algorithm, we could resort to using "slow-rate" bounds for the LASSO (see e.g. Section 7.4 in Wainwright (2019)) for finite-dimensional or countable sparse regression, or the Beurling LASSO (Bach, 2017) for uncountable sparse regression. However, the regret bound of SquareCB would have a sub-optimal $n^{3/4}$ rate. Moreover, it is not clear that this actually would lead to a polynomial-time algorithm for sparse nonparametric regression. The time complexity of the LASSO is polynomial in the number of candidate features, but in the countable sparsity model, the (effective) number of features is d_{eff}, which can be exponential in the dimension of the contexts. Similarly, it is not known whether the Beurling LASSO can be computed with polynomial time complexity.

C.2 Large Action Sets

Our lower bounds show that under our mild set of assumptions it is not possible to have low worst-case regret when the set of actions is large. It would be interesting to investigate which additional assumptions can lead to regret bounds with improved dependence on the number of actions. In sparse linear contextual bandits with contexts drawn i.i.d. from a well-conditioned distribution, there are several algorithms that achieve logarithmic (or better) dependence on both the dimension and the number of actions (Kim & Paik, 2019; Oh et al., 2021). Loosely speaking, the context distribution is well-conditioned if there is low correlation between the features. This notion of a well-conditioned distribution can be made precise using conditions studied in high-dimensional statistics, such as the compatibility condition (Bühlmann & Van De Geer, 2011), restricted eigenvalue conditions (Bickel et al., 2009) and the restricted isometry property (Candès & Tao, 2005, 2007) to name a few. Unfortunately, when the number of features is infinite, these conditions can fail to hold at all. In the field of compressed sensing, a number of weaker conditions, known as individual or non-uniform recovery conditions (cf. Section 4.4 in Foucart & Rauhut (2013)), have been studied and applied to sparse infinite-dimensional linear models. In the uncountable sparsity model, these conditions roughly state that recovery is possible if $\theta_1^*, \dots, \theta_s^*$ are well-separated (Candès & Fernandez-Granda, 2014; Duval & Peyré, 2015; Azais et al., 2015; Poon et al., 2023). One could investigate whether these individual recovery conditions enable regret bounds for sparse nonparametric contextual bandits with logarithmic (or better) dependence on the number of actions.

These conditions on the context distribution also have implications for designing computationally efficient algorithms. If the restricted isometry property is satisfied, then it is possible to sample from spike-and-slab posteriors for sparse (finite-dimensional) linear models with polynomial time complexity (Kumar et al., 2025). If one could prove a similar result for sparse nonparametric models, then (under suitable conditions on the context distribution) it would be possible to run FGTS with polynomial time complexity. Similarly, if a restricted eigenvalue condition is satisfied, then the LASSO enjoys fast-rate bounds for sparse regression (cf. Theorem 7.20 in Wainwright (2019)), and so SquareCB would have a $\sqrt{n}$ regret bound. However, as previously mentioned, it is not clear that the LASSO (or BLASSO) has polynomial time complexity for sparse nonparametric models.

C.3 Instance-Dependent Analysis

We could also consider conditions on the context and/or reward distributions that allow for improved dependence of the regret on n. Several sparse linear bandit algorithms enjoy improved regret bounds when a margin condition (cf. Assumption 1 in Li et al. (2021)) is satisfied. The margin condition roughly states that, with high probability over the random draw of the context, the gap between the reward of the best action and that of the second best action is large. In the extreme case where, with probability 1, there is a strictly positive gap, regret bounds of order $\log(n)$ can be achieved (Wang et al., 2018; Bastani & Bayati, 2020; Li et al., 2021). It would be interesting to investigate whether the margin condition enables logarithmic regret bounds for sparse nonparametric contextual bandits.

Similarly, one could investigate first-order bounds, where the dependence of the regret on n is replaced with the cumulative loss (negation of the reward) of the best policy, which will be much smaller when, for instance, a margin condition is satisfied and/or the rewards are (almost) noiseless. The first efficient algorithm with first-order guarantees for contextual bandits was introduced by Foster & Krishnamurthy (2021), and is based on a reduction from contextual bandits to online regression with cross-entropy loss. More recently, the Optimistic Information Directed Sampling (OIDS) algorithm was shown to achieve a first-order regret bound for a general class of contextual bandit problems (Neu, Papini, and Schwartz, 2024). As noted by Neu, Papini, and Schwartz (2024), their information-theoretic regret analysis is closely related to the decoupling technique proposed by Zhang (2022), which we used in this paper. In particular, both approaches reduce the problem of bounding the regret to that of bounding the log partition function Z_n in (11). We therefore expect that our bounds on Z_n could be used to show that OIDS satisfies a first-order regret bound for sparse nonparametric contextual bandits.

C.4 Misspecification

For simplicity, and to keep our attention focused solely on the challenge of exploiting sparsity, we assumed that the model is well-specified. To make the model more realistic, we could assume that the reward function is approximated well by a sparse function. For FGTS, achieving a good balance between the regret suffered while estimating the best s-sparse approximation of the reward function and the regret suffered due to approximation error likely requires the tuning parameter λ to be set according to the misspecification level, which will decrease as s increases, but the rate at which it does so may be unknown. One could investigate whether an aggregation (or "corralling") procedure (Agarwal et al., 2017; Foster et al., 2020) could be used to tune λ adaptively.

D Proofs of the Lower Bounds

The proofs of Theorem 3.1 and Theorem 3.2 both work by first reducing the sparse nonparametric contextual bandit problem to a sequence of K-armed bandits, and then lower bounding the total regret in the sequence of K-armed bandit problems. Each type of sparsity requires a different reduction. However, both proofs use the same auxiliary lemmas to lower bound the regret in the sequence of K-armed bandit problems. In Section D.1, we state and prove these auxiliary lemmas. In the subsequent subsections, we use these lemmas to prove Theorem 3.1 and then Theorem 3.2.

D.1 Auxiliary Lemmas

The first auxiliary lemma is a lower bound on the regret for K-armed bandit problems in which one action has mean reward Δ and the remaining $K-1$ actions have mean reward 0. The rewards of all actions are subject to standard Gaussian noise. For some $b \in [K]$, let $R_m(b)$ denote the expected regret over m steps in a K-armed bandit with optimal action b. Therefore,

$$R_m(b) = \mathbb{E}\left[\sum_{t=1}^{m} \Delta \cdot \mathbb{I}\{A_t \neq b\}\right].$$

Exercise 24.1 in Lattimore & Szepesvári (2020a) asks the reader to prove the following lower bound on the averaged (over b) expected regret $\frac{1}{K}\sum_{b\in[K]} R_m(b)$.

Lemma D.1 (Exercise 24.1 in Lattimore & Szepesvári (2020a))**.** *Let $K \geq 2$ and set $\Delta = \sqrt{K/m}/2$. For any policy,*

$$\frac{1}{K}\sum_{b\in[K]} R_m(b) \geq \frac{1}{8}\sqrt{Km}.$$

A proof can be found in Lattimore & Szepesvári (2020b). To use this result, we need to reduce the sparse nonparametric contextual bandit problem to a sequence of K-armed bandit problems. In the proofs of each lower bound, we always factorise the horizon as $n = m_1 m_2$ for some $m_1, m_2 \in \mathbb{N}$ and fix the sequence of contexts $x_1, \dots, x_n$ to be $z_1, \dots, z_1, z_2, \dots, z_2, \dots, z_{m_1}, \dots, z_{m_1}$, where each z_i is repeated m_2 times. For each lower bound, we choose a set of parameters N of size $|N| = K^{m_1}$ and a feature map ϕ, such that for every $i \in [m_1]$, $a \in [K]$ and $\nu \in N$, we can express the expected reward as

$$f_\nu(z_i, a) = \tfrac{\Delta}{s}\mathbb{I}\{a = b_i(\nu)\}\,. \tag{8}$$

The function $b : N \to [K]^{m_1}$ is any bijection between N and the set of sequences of actions of length m_1. This reward function mimics a sequence of m_1 K-armed bandit problems, in which the index of the optimal action switches every m_2 rounds. For each context z_i and each $\nu \in N$, there is a single good action $b_i(\nu)$ with expected reward Δ/s and $K-1$ bad actions with expected reward 0. Since each $\nu \in N$ corresponds to a unique sequence $(b_1, \dots, b_{m_1}) \in [K]^{m_1}$ of optimal actions (and vice versa), we can equivalently parameterise the reward function by $b_{1:m_1} := (b_1, \dots, b_{m_1}) \in [K]^{m_1}$. In particular, for each $b_{1:m_1} \in [K]^{m_1}$, we can write the expected regret $R_n(b_{1:m_1})$ as

$$R_n(b_{1:m_1}) := \mathbb{E}\left[\sum_{t=1}^{n}\sum_{i=1}^{m_1} \tfrac{\Delta}{s}\mathbb{I}\{x_t = z_i\}\mathbb{I}\{A_t \neq b_i\}\right].$$

The following lemma shows that, for any $b_{1:m_1} \in [K]^{m_1}$, $R_n(b_{1:m_1})$ can be re-written as the total expected regret suffered in a sequence of K-armed bandit problems.

Lemma D.2 (Regret decomposition)**.** *For any* $b_{1:m_1} \in [K]^{m_1}$*,*

$$R_n(b_{1:m_1}) = \sum_{i=1}^{m_1} R_{m_2,i}(b_{1:i})\,,$$

where

$$R_{m_2,i}(b_{1:i}) = \mathbb{E}\left[\sum_{t=m_2(i-1)+1}^{m_2 i} \tfrac{\Delta}{s}\mathbb{I}\{A_t \neq b_i\}\right].$$

Proof. The proof only requires us to swap a sum and an expectation. In particular,

$$\begin{aligned}R_n(b_{1:m_1}) &= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{i=1}^{m_1} \tfrac{\Delta}{s}\mathbb{I}\{x_t = z_i\}\mathbb{I}\{A_t \neq b_i\}\right]\\ &= \sum_{i=1}^{m_1}\mathbb{E}\left[\sum_{t=1}^{n} \tfrac{\Delta}{s}\mathbb{I}\{x_t = z_i\}\mathbb{I}\{A_t \neq b_i\}\right]\\ &= \sum_{i=1}^{m_1}\mathbb{E}\left[\sum_{t=m_2(i-1)+1}^{m_2 i} \tfrac{\Delta}{s}\mathbb{I}\{A_t \neq b_i\}\right] = \sum_{i=1}^{m_1} R_{m_2,i}(b_{1:i}).\end{aligned}$$

□

We notice that $R_{m_2,i}(b_{1:i})$ is the regret suffered in m_2 steps of a K-armed bandit problem, in which the optimal action is b_i and all other actions result in an expected regret of Δ/s. We write $R_{m_2,i}(b_{1:i})$ (as opposed to $R_{m_2,i}(b_{1:m_1})$) because the expected regret for the i^{th} sub-problem is independent of $b_{i+1:m_1}$. $R_{m_2,i}(b_{1:i})$ can depend on the first $i-1$ elements $b_{1:i-1}$ of the sequence $b_{1:m_1}$, since the rewards obtained in the first $(i-1)m_2$ rounds can influence the policy played in rounds $t = (i-1)m_2 + 1$ to $t = im_2$. However, the lower bound in Lemma D.1 applies to any policy, and hence any $b_{1:i-1}$. The final auxiliary lemma combines the previous two, and provides a lower bound on the averaged regret $\frac{1}{K^{m_1}}\sum_{b_{1:m_1}\in[K]^{m_1}} R_n(b_{1:m_1})$.

Lemma D.3 (Lower bound for sequences of K-armed bandits)**.** *Let* $K \geq 2$ *and set* $\Delta = s\sqrt{K}/\sqrt{4m_2}$*. For any* $b_{1:m_1} \in [K]^{m_1}$ *and any policy,*

$$\frac{1}{K^{m_1}}\sum_{b_{1:m_1}\in[K]^{m_1}} R_n(b_{1:m_1}) \geq \frac{1}{8} m_1\sqrt{Km_2}\,.$$

Proof. Choosing $\Delta = s\sqrt{K}/\sqrt{Km_2}$ ensures that $\Delta/s = \sqrt{K}/\sqrt{4m_2}$, which is required for the lower bound in Lemma D.1. Using Lemma D.1, and the fact that $R_{m_2,i}(b_{1:i})$ is independent of $b_{i+1:m_1}$, we obtain,

$$\begin{aligned}
\frac{1}{K^{m_1}} \sum_{b_{1:m_1} \in [K]^{m_1}} R_n(b_{1:m_1}) &= \sum_{i=1}^{m_1} \frac{1}{K^{m_1}} \sum_{b_{1:m_1} \in [K]^{m_1}} R_{m_2,i}(b_{1:m_1}) \\
&= \sum_{i=1}^{m_1} \frac{1}{K^{i-1}} \sum_{b_{1:i-1} \in [K]^{i-1}} \frac{1}{K} \sum_{b_i \in [K]} \frac{1}{K^{m_1-i}} \sum_{b_{i+1:m_1} \in [K]^{m_1-i}} R_{m_2,i}(b_{1:i}) \\
&= \sum_{i=1}^{m_1} \frac{1}{K^{i-1}} \sum_{b_{1:i-1} \in [K]^{i-1}} \frac{1}{K} \sum_{b_i \in [K]} R_{m_2,i}(b_{1:i}) \\
&\geq \frac{1}{8} \sum_{i=1}^{m_1} \frac{1}{K^{i-1}} \sum_{b_{1:i-1} \in [K]^{i-1}} \sqrt{Km_2} \\
&= \frac{1}{8} m_1 \sqrt{Km_2}\,.
\end{aligned}$$

□

D.2 Proof of Theorem 3.1

For the convenience of the reader, we repeat the statement of Theorem 3.1 here.

Statement of Theorem 3.1. Consider the sparse nonparametric contextual bandit problem with countable sparsity described in Section 2. Let $\mathcal{A} = [K]$ for some $K \geq 2$ and assume that the noise variables are standard Gaussian. Suppose that for some $\beta > 1$ and some integer $m \geq s^{\beta+2}K^{\beta+1}$, $n = sm$. Then for any policy, there exists a sequence of contexts $x_1, \dots, x_n$, a parameter sequence $w = (w_1, w_2, \dots)$ with $\|w\|_0 = s$, $\|w\|_1 = 1$ and a sequence of functions $(\phi_i)_{i=1}^{\infty}$ with $\|\phi_i\|_\infty \leq i^{-\beta/2}$, such that

$$R_n(f_w) \geq \frac{1}{8}\sqrt{Ksn}.$$

Instead, suppose that for some $\beta > 0$ and some integer $m \geq \lceil 1/\beta \rceil s^2 K \exp(s^\beta K^{\beta \lceil 1/\beta \rceil})$, $n = sm$. Then for any policy, there exists a sequence of contexts $x_1, \dots, x_n$, a parameter sequence $w = (w_1, w_2, \dots)$ with $\|w\|_0 = s$, $\|w\|_1 = 1$ and a sequence of functions $(\phi_i)_{i=1}^{\infty}$ with $\|\phi_i\|_\infty \leq \exp(-i^\beta/2)$, such that

$$R_n(f_w) \geq \frac{1}{8}\sqrt{\max(1, 1/\beta)Ksn}.$$

Proof. First, we prove the lower bound for the scenario with polynomial decay, in which $\|\phi_i\|_\infty \leq i^{-\beta/2}$ for some $\beta > 1$. The sequence of contexts $x_1, \dots, x_n$ is fixed in advance, and selected from the set $\{z_1, \dots, z_s\}$, where each z_i is of the form $z_i = (z_{i,a})_{a \in [K]}$. For each $i \in [s]$ and $a \in [K]$, $z_{i,a}$ can be arbitrary, as long as each $z_{i,a}$ is distinct. We choose the sequence of contexts $x_1, \dots, x_n$ to be

$$z_1, \dots, z_1, z_2, \dots, z_2, \dots, z_s, \dots, z_s,$$

where each z_i is repeated m times. Next, we define the sequence of functions $(\phi_j)_{j=1}^{\infty}$. In fact, we will consider a finite sequence of functions $(\phi_j)_{j=1}^{sK}$. We could extend this to an infinite sequence by choosing $\phi_j \equiv 0$ for $j > sK$ and nothing would change. We define $e_1, \dots, e_{sK}$ to be the standard basis vectors of $\mathbb{R}^{sK}$. For each $i \in [s]$ and $a \in [K]$, and some $\Delta \in (0, 1]$ to be chosen later, we define

$$\phi(z_{i,a}) = (\phi_1(z_{i,a}), \dots, \phi_{sK}(z_{i,a})) = \Delta \cdot e_{(i-1)K+a}\,.$$

We could also write this as $\phi_j(z_{i,a}) = \Delta \cdot \mathbb{I}\{j = (i-1)K + a\}$. Note that if $\Delta \leq (sK)^{-\beta/2}$, then $\|\phi_j\|_\infty \leq j^{-\beta/2}$ is satisfied for all $j \in [sK]$. Next, we define a set of s-sparse parameter sequences. In fact, we will consider sK-dimensional parameter vectors, but if we wished to use an infinite sequence, we could just append zeros to the end. We let $u_1, \dots, u_K$ denote the standard basis vectors of $\mathbb{R}^K$, and we define the set $\mathcal{W} = \{(1/s)u_i : i \in [K]\} \subset \mathbb{R}^K$. We consider parameter vectors in the set $\mathcal{W}^s \subset \mathbb{R}^{sK}$, meaning each w is of the form

$$w = [(1/s)u_{i_1}^\top, \dots, (1/s)u_{i_s}^\top]^\top ,$$

for some collection of indices $(i_1, \dots, i_s) \in [K]^s$. We define the bijection $b : \mathcal{W}^s \to [K]^s$ to be the function that maps each w to its corresponding sequence of indices $(i_1, \dots, i_s)$. Thus for any $i \in [s]$, $b_i(w)$ is the position of the non-zero element within the i^{th} block of w. We notice that each $w \in \mathcal{W}^s$ satisfies $\|w\|_0 = s$ and $\|w\|_1 = 1$. Moreover, for any $i \in [s]$, $a \in [K]$ and $w \in \mathcal{W}^s$ we can write down the expected reward as

$$f_w(z_i, a) = \sum_{j=1}^{sK} w_j \phi_j(z_{i,a}) = \tfrac{\Delta}{s}\mathbb{I}\{a = b_i(w)\} .$$

This reward function (and this sequence of contexts) mimics a sequence of K-armed bandit problems, in which the index of the optimal action switches every m rounds. In particular, it is of the same form as the reward function in (8). Since each $w \in \mathcal{W}^s$ corresponds to a unique sequence $(b_1, \dots, b_s) \in [K]^s$ of optimal actions (and vice versa), we can equivalently parameterise the reward function by the sequence $b_{1:s} := (b_1, \dots, b_s) \in [K]^s$. Now, using Lemma D.2, we can re-write the regret $R_n(b_{1:s})$ as

$$R_n(b_{1:s}) = \sum_{i=1}^{s} R_{m,i}(b_{1:i}) ,$$

where $R_{m,i}(b_{1:i}) = \mathbb{E}[\sum_{t=m(i-1)+1}^{mi} \frac{\Delta}{s}\mathbb{I}\{A_t \neq b_i\}]$ is the expected regret for the i^{th} sub-problem. Using this regret decomposition and Lemma D.3 (and the fact that $n = sm$), we can lower bound the averaged expected regret as

$$\frac{1}{K^s} \sum_{b_{1:s} \in [K]^s} R_n(b_{1:s}) \geq \frac{1}{8} s\sqrt{Km} = \frac{1}{8}\sqrt{Ksn} .$$

Lemma D.3 requires us to set Δ such that $\Delta/s = \sqrt{K/m}/2$. The only thing left to do is to find the values of m such that $\Delta = s\sqrt{\frac{K}{4m}} \leq (sK)^{-\beta/2}$, which ensures that $\phi_1, \dots, \phi_{sK}$ satisfies the polynomial decay condition. This constraint for m can be rearranged into

$$m \geq \tfrac{1}{4} s^{\beta+2} K^{\beta+1} .$$

Thus we can choose any m satisfying $m \geq s^{\beta+2}K^{\beta+1}$. Next, we prove the lower bound for the scenario with exponential decay, in which $\|\phi_i\|_\infty \leq \exp(-i^\beta/2)$, for some $\beta > 0$. First, we consider the case where $\beta \geq 1$. In this case, $\max(1, 1/\beta) = 1$, so we need to lower bound the expected regret by $\sqrt{Ksn}/8$. To do so, we can mostly repeat the proof for polynomial decay. We set $(\phi_j)_{j=1}^\infty$ to be the same sequence of functions, which satisfies $\|\phi_j\|_\infty \leq \Delta$ for $j \leq sK$ and $\|\phi_j\|_\infty = 0$ for $j \geq sK$. If $\Delta \leq \exp(-s^\beta K^\beta/2)$, then $(\phi_j)_{j=1}^\infty$ satisfies the exponential decay condition. We can factorise n as $n = sm$ and repeat the rest of the proof to obtain the desired lower bound of $\sqrt{Ksn}/8$. In order to use Lemma D.3, we again require $\Delta = s\sqrt{K}/\sqrt{4m}$. This means that the exponential decay condition is satisfied when m satisfies the inequality $s\sqrt{K}/\sqrt{4m} \leq \exp(-s^\beta K^\beta/2)$. This is equivalent to

$$m \geq \tfrac{1}{4} s^2 K \exp(s^\beta K^\beta) .$$

Thus we can choose any $m \geq s^2 K \exp(s^\beta K^\beta)$. Due to the condition on n in the statement of the theorem, when $\beta \geq 1$, we can indeed factorise n as $n = sm$, for some integer $m \geq s^2 K \exp(s^\beta K^\beta)$.

For the case where $\beta \in (0, 1)$, we require a more elaborate reduction to sequences of K-armed bandit problems. The proof for this case generalises the previous one in the sense that, for some positive integer $l \neq 1$, we reduce the problem to a sequence of sl K-armed bandit problems, each with horizon m. We factorise the horizon as $n = s\lceil 1/\beta \rceil m$. The sequence of contexts $x_1, \dots, x_n$ is fixed in advance, and selected from the set $\{z_1, \dots, z_{s\lceil 1/\beta \rceil}\}$. For each $i \in [s\lceil 1/\beta \rceil]$ and $a \in [K]$,

$z_{i,a}$ can be arbitrary, as long as each $z_{i,a}$ is distinct. We choose the sequence of contexts $x_1, \ldots, x_n$ to be

$$z_1, \ldots, z_1, z_2, \ldots, z_2, \ldots, z_{s\lceil 1/\beta\rceil}, \ldots, z_{s\lceil 1/\beta\rceil},$$

where each z_i is repeated m times. Next, we define the sequence of functions $(\phi_j)_{j=1}^{\infty}$. For each $i \in [s\lceil 1/\beta\rceil]$, we define the bijection $\rho : [s\lceil 1/\beta\rceil] \to [s] \times [\lceil 1/\beta\rceil]$ by

$$\rho(i) = (\rho_1(i), \rho_2(i)) = (\lceil \tfrac{i}{\lceil 1/\beta\rceil} \rceil, (i-1) \bmod \lceil 1/\beta\rceil + 1).$$

Some values of ρ are $\rho(1) = (1,1)$, $\rho(2) = (1,2)$, $\rho(\lceil 1/\beta\rceil) = (1, \lceil 1/\beta\rceil)$, $\rho(\lceil 1/\beta\rceil + 1) = (2,1)$, $\rho(s\lceil 1/\beta\rceil) = (s, \lceil 1/\beta\rceil)$. We let $\zeta : [K^{\lceil 1/\beta\rceil}] \to [K]^{\lceil 1/\beta\rceil}$ be any bijection between $[K^{\lceil 1/\beta\rceil}]$, the set of all positive integers up to $K^{\lceil 1/\beta\rceil}$, and $[K]^{\lceil 1/\beta\rceil}$, the set of sequences of integers in $[K]$ of length $\lceil 1/\beta\rceil$. For instance, we could choose

$$(\zeta_1(i), \zeta_2(i), \ldots, \zeta_{\lceil 1/\beta\rceil}(i)) = (\lfloor \tfrac{i-1}{K^{\lceil 1/\beta\rceil - 1}} \rfloor + 1, \lfloor \tfrac{i-1}{K^{\lceil 1/\beta\rceil - 2}} \rfloor \bmod K + 1, \ldots, (i-1) \bmod K + 1).$$

We can now define the sequence of functions $(\phi_j)_{j=1}^{\infty}$. Similarly to before, we will consider a finite sequence of $sK^{\lceil 1/\beta\rceil}$ functions, but we could extend it to an infinite sequence by choosing $\phi_j \equiv 0$ for all $j > sK^{\lceil 1/\beta\rceil}$. For each $i \in [s\lceil 1/\beta\rceil]$, $a \in [K]$ and $j \in [sK^{\lceil 1/\beta\rceil}]$, we define the j^{th} function in the sequence to be

$$\phi_j(z_{i,a}) = \Delta \cdot \mathbb{I}\{\lfloor \tfrac{j-1}{K^{\lceil 1/\beta\rceil}} \rfloor + 1 = \rho_1(i)\} \cdot \mathbb{I}\{a = \zeta_{\rho_2(i)}((j-1) \bmod K^{\lceil 1/\beta\rceil} + 1)\}\,, \tag{9}$$

where $\Delta \in (0, 1]$ is some positive constant to be chosen later. If we let $\phi(z_{i,a})$ denote the vector in $\mathbb{R}^{sK^{\lceil 1/\beta\rceil}}$ whose j^{th} element is $\phi_j(z_{i,a})$, we can describe the functions $\phi_1, \ldots, \phi_{sK^{\lceil 1/\beta\rceil}}$ in a simpler way. We split the vector $\phi(z_{i,a})$ into s blocks of $K^{\lceil 1/\beta\rceil}$ elements and write

$$\phi(z_{i,a}) = [0, \ldots, 0, \Delta \cdot \mathbb{I}\{a = \zeta_{\rho_2(i)}(1)\}, \ldots, \Delta \cdot \mathbb{I}\{a = \zeta_{\rho_2(i)}(K^{\lceil 1/\beta\rceil})\}, 0, \ldots, 0]\,, \tag{10}$$

where the block of $K^{\lceil 1/\beta\rceil}$ elements that are not identically equal to zero is the $\rho_1(i)^{\text{th}}$ block. To see that this is the same as the previous definition, we can observe that the first indicator in (9) sets $\phi_j(z_{i,a})$ to 0 whenever j does not correspond to an index in the $\rho_1(i)^{\text{th}}$ block of $\phi(z_{i,a})$. For j in the $\rho_1(i)^{\text{th}}$ block of $\phi(z_{i,a})$, i.e. j satisfying

$$(\rho_1(i) - 1)K^{\lceil 1/\beta\rceil} + 1 \le j \le \rho_1(i)K^{\lceil 1/\beta\rceil}\,,$$

$(j-1) \bmod K^{\lceil 1/\beta\rceil} + 1$ is equal to the position of j within the $\rho_1(i)^{\text{th}}$ block, which means the second indicator in (9) is equal to the indicator in (10) for these j. To satisfy the exponential decay condition, Δ must be set small enough. In particular, since $\|\phi_j\|_\infty \le \Delta$ for all $j \in [sK^{\lceil 1/\beta\rceil}]$, then as long as Δ satisfies $\Delta \le \exp(-s^\beta K^{\beta\lceil 1/\beta\rceil}/2)$, each ϕ_j satisfies $\|\phi_j\|_\infty \le \exp(-j^\beta/2)$.

Next, we define a set of s-sparse parameter sequences. We will consider s-sparse vectors in $\mathbb{R}^{sK^{\lceil 1/\beta\rceil}}$, but we could extend this to infinite sequences by appending zeros. We let $u_1, \ldots, u_{K^{\lceil 1/\beta\rceil}}$ denote the standard basis vectors of $\mathbb{R}^{K^{\lceil 1/\beta\rceil}}$, and we define the set $\mathcal{W} = \{(1/s)u_i : i \in [K^{\lceil 1/\beta\rceil}]\}$. We consider parameter vectors in the set $\mathcal{W}^s$, so each w is of the form

$$w = [(1/s)u_{i_1}^\top, \ldots, (1/s)u_{i_s}^\top]^\top\,,$$

for some collection of indices $(i_1, \ldots, i_s) \in [K^{\lceil 1/\beta\rceil}]^s$. We define the bijection $\omega : \mathcal{W}^s \to [K^{\lceil 1/\beta\rceil}]^s$ to be the function that maps each w to it's corresponding sequence of indices $(i_1, \ldots, i_s)$. Thus, for any $i \in [s]$, $\omega_i(w)$ is the position of the non-zero element within the i^{th} block of w. We notice that each $w \in \mathcal{W}^s$ satisfies $\|w\|_0 = s$ and $\|w\|_1 = 1$. Now, we can write down the expression for the expected reward function. In particular,

$$f_w(z_i, a) = \sum_{j=1}^{sK^{\lceil 1/\beta\rceil}} w_j \phi_j(z_{i,a}) = \tfrac{\Delta}{s}\mathbb{I}\{a = \zeta_{\rho_2(i)}(\omega_{\rho_1(i)}(w))\}.$$

To simplify this expression, we define $b : \mathcal{W}^s \to [K]^{s\lceil 1/\beta \rceil}$ to be the bijection that maps each w to the corresponding sequence of good actions. Thus $b(w) = (b_1(w), \dots, b_{s\lceil 1/\beta \rceil}(w))$, where for each $i \in [s\lceil 1/\beta \rceil]$,

$$b_i(w) = \zeta_{\rho_2(i)}(\omega_{\rho_1(i)}(w)).$$

Thus, for any $i \in [s\lceil 1/\beta \rceil]$, $a \in [K]$ and $w \in \mathcal{W}^s$ we can express the expected reward as

$$f_w(z_i, a) = \tfrac{\Delta}{s}\mathbb{I}\{a = b_i(w))\}.$$

We notice that we have another reward function that is in the same form as the reward function in (8). Since b is a bijection between $\mathcal{W}^s$ and $[K]^{s\lceil 1/\beta \rceil}$, each $w \in \mathcal{W}^s$ corresponds to a unique sequence $(b_1, \dots, b_{s\lceil 1/\beta \rceil}) \in [K]^{s\lceil 1/\beta \rceil}$ of optimal actions (and vice versa). Therefore, we can equivalently parameterise the reward function by the sequence $b_{1:s\lceil 1/\beta \rceil} := (b_1, \dots, b_{s\lceil 1/\beta \rceil}) \in [K]^{s\lceil 1/\beta \rceil}$. Using Lemma D.2, we can re-write the regret $R_n(b_{1:s\lceil 1/\beta \rceil})$ as

$$R_n(b_{1:s\lceil 1/\beta \rceil}) = \sum_{i=1}^{s\lceil 1/\beta \rceil} R_{m,i}(b_{1:i})\,,$$

where $R_{m,i}(b_{1:i}) = \mathbb{E}[\sum_{t=m(i-1)+1}^{mi} \frac{\Delta}{s}\mathbb{I}\{A_t \neq b_i\}]$ is the expected regret for the i^{th} sub-problem. Using this regret decomposition and Lemma D.3 (and the fact that $n = s\lceil 1/\beta \rceil m$), we can lower bound the averaged expected regret as

$$\frac{1}{K^{s\lceil 1/\beta \rceil}} \sum_{b_{1:s\lceil 1/\beta \rceil} \in [K]^{s\lceil 1/\beta \rceil}} R_n(b_{1:s\lceil 1/\beta \rceil}) \geq \frac{1}{8} s\lceil 1/\beta \rceil \sqrt{Km} = \frac{1}{8}\sqrt{\lceil 1/\beta \rceil Ksn} \geq \frac{1}{8}\sqrt{(1/\beta)Ksn}\,.$$

Lemma D.3 requires us to set $\Delta/s = \sqrt{K/m}/2$. The only thing left to do is to find the values of m such that $\Delta = s\sqrt{\frac{K}{4m}} \leq \exp(-s^\beta K^{\beta\lceil 1/\beta \rceil}/2)$, which ensures that the exponential decay condition is satisfied. This constraint for m can be rearranged into

$$m \geq \tfrac{1}{4}s^2 K \exp(s^\beta K^{\beta\lceil 1/\beta \rceil})\,.$$

Thus, we can choose any m satisfying $m \geq s^2 K \exp(s^\beta K^{\beta\lceil 1/\beta \rceil})$. Since the condition on n was $n = sm$ for some integer $m \geq \lceil 1/\beta \rceil s^2 K \exp(s^\beta K^{\beta\lceil 1/\beta \rceil})$, we can indeed factorise n into $n = s\lceil 1/\beta \rceil m$ for some $m \geq s^2 K \exp(s^\beta K^{\beta\lceil 1/\beta \rceil})$. □

D.3 Proof of Theorem 3.2

For the convenience of the reader, we repeat the statement of Theorem 3.2 here.

Statement of Theorem 3.2. Consider the sparse nonparametric contextual bandit problem with uncountable sparsity described in Section 2. Let $\mathcal{A} = [K]$ for some $K \geq 2$ and let $n = sdm$ for some integer $m \geq s^{2+2/d}K^3$. Assume that the noise variables are standard Gaussian. For any policy, there exists a sequence of contexts $x_1, \dots, x_n \in \mathcal{X}$, parameters $w \in \mathbb{B}_1^s(1)$, $\theta_1, \dots, \theta_s \in \Theta \subset \mathbb{B}_2^d(1)$ and a uniformly Lipschitz continuous function ϕ, with $\|\phi\|_\infty \leq 1$, such that

$$R_n(f_\nu) \geq \frac{1}{8}\sqrt{Ksdn}.$$

Proof. The sequence of contexts $x_1, \dots, x_n$ is fixed in advance, and selected from the set $\{z_1, \dots, z_{sd}\}$. For each $i \in [sd]$ and $a \in [K]$, $z_{i,a}$ can be arbitrary, as long as each $z_{i,a}$ is distinct. In particular, we choose the sequence of contexts $x_1, \dots, x_n$ to be

$$z_1, \dots, z_1, z_2, \dots, z_2, \dots, z_{sd}, \dots, z_{sd},$$

where each z_i is repeated m times. Next, we choose the parameters w and $\theta_1, \dots, \theta_s$. First, we set $w_j = 1/s$ for every $j \in [s]$, which means that $\|w\|_1 = 1$. The parameters $\theta_1, \dots, \theta_s$ lie in a set Θ, which we specify now. We choose Θ such that: a) Θ is a Δ-packing of $\mathbb{B}_2^d(1)$ w.r.t. the ℓ_2

norm, for some $\Delta \in (0, 1]$ to be chosen later; b) Θ is a union of s disjoint sets $\Theta_1, \ldots, \Theta_s$, each with cardinality $|\Theta_j| = K^d$. To guarantee that there is a set Θ that satisfies these requirements, we must choose Δ such that the Δ-packing number $\mathcal{M}(\mathbb{B}_2^d(1), \|\cdot\|_2, \Delta)$ is at least sK^d. Using Lemma 5.5 and Lemma 5.7 in Wainwright (2019), $\mathcal{M}(\mathbb{B}_2^d(1), \|\cdot\|_2, \Delta)$ can be lower bounded as

$$\mathcal{M}(\mathbb{B}_2^d(1), \|\cdot\|_2, \Delta) \geq \mathcal{N}(\mathbb{B}_2^d(1), \|\cdot\|_2, \Delta) \geq (\tfrac{1}{\Delta})^d,$$

which means that if $sK^d \leq (\frac{1}{\Delta})^d$, then we can construct a Δ-packing of $\mathbb{B}_2^d(1)$ that has at least sK^d elements. This inequality can be rearranged into $\Delta \leq 1/(s^{1/d}K)$. We will consider sequences of parameters $\theta_1, \ldots, \theta_s$ where each θ_j is in Θ_j. For each $j \in [s]$, since Θ_j has cardinality K^d, we can define a bijection $\omega^j : \Theta_j \to [K]^d$ between Θ_j and the set $[K]^d$ of sequences of actions of length d. We define the bijection $\rho : [sd] \to [s] \times [d]$ by

$$\rho(i) = (\rho_1(i), \rho_2(i)) = (\lceil \tfrac{i}{d} \rceil, (i-1) \bmod d + 1).$$

Now, using ρ and $\omega^1, \ldots \omega^s$, we can define a bijection $b : \Theta_1 \times \cdots \times \Theta_s \to [K]^{sd}$ between $\Theta_1 \times \cdots \times \Theta_s$ and the set $[K]^{sd}$ of sequences of actions of length d. In particular, for any $(\theta_1, \ldots \theta_s) \in \Theta_1 \times \cdots \times \Theta_s$ and $i \in [sd]$, we define

$$b_i(\theta_1, \ldots, \theta_s) = \omega^{\rho_1(i)}_{\rho_2(i)}(\theta_{\rho_1(i)}).$$

To define ϕ, we introduce one more function. We define $\kappa : \Theta \to [s]$ to be the function that maps each $\theta \in \Theta$ to the unique integer $j \in [s]$ such that $\theta \in \Theta_j$. For each $i \in [sd]$, $a \in [K]$ and $\theta \in \Theta$, we define (with the same $\Delta \in (0, 1]$ as before)

$$\phi(z_{i,a}, \theta) = \Delta \mathbb{I}\{\rho_1(i) = \kappa(\theta)\} \mathbb{I}\{a = \omega^{\kappa(\theta)}_{\rho_2(i)}(\theta)\}.$$

Since $\phi(z_{i,a}, \theta) \in \{0, \Delta\}$ and Θ is a Δ-packing, we have

$$\forall i \in [sd], a \in [K], \theta, \theta' \in \Theta,\ |\phi(z_{i,a}, \theta) - \phi(z_{i,a}, \theta')| \leq \Delta < \|\theta - \theta'\|_2,$$

which means that ϕ satisfies the uniform Lipschitz continuity property. Using the definition of ϕ, for any $\nu = (\theta_1, \ldots, \theta_s) \in \Theta_1 \times \cdots \times \Theta_s$, the expected reward function is

$$f_\nu(z_i, a) = \sum_{j=1}^{s} \tfrac{1}{s} \phi(z_{i,a}, \theta_j) = \tfrac{\Delta}{s} \mathbb{I}\{a = \omega^{\rho_1(i)}_{\rho_2(i)}(\theta_{\rho_1(i)})\} = \tfrac{\Delta}{s} \mathbb{I}\{a = b_i(\theta_1, \ldots, \theta_s)\}.$$

We have another reward function that is in the same form as the reward function in (8). Since b is a bijection between $\Theta_1 \times \cdots \times \Theta_s$ and $[K]^{s\lceil 1/\beta \rceil}$, each $\nu \in \Theta_1 \times \cdots \times \Theta_s$ corresponds to a unique sequence $(b_1, \ldots, b_{sd}) \in [K]^{sd}$ of optimal actions (and vice versa). Therefore, we can equivalently parameterise the reward function by the sequence $b_{1:sd} := (b_1, \ldots, b_{sd}) \in [K]^{sd}$. Using Lemma D.2, we can re-write the regret $R_n(b_{1:sd})$ as

$$R_n(b_{1:sd}) = \sum_{i=1}^{sd} R_{m,i}(b_{1:i}),$$

where $R_{m,i}(b_{1:i}) = \mathbb{E}[\sum_{t=m(i-1)+1}^{mi} \frac{\Delta}{s} \mathbb{I}\{A_t \neq b_i\}]$ is the expected regret for the i^{th} sub-problem. Using this regret decomposition and Lemma D.3 (and the fact that $n = sdm$), we can lower bound the averaged expected regret as

$$\frac{1}{K^{sd}} \sum_{b_{1:sd} \in [K]^{sd}} R_n(b_{1:sd}) \geq \frac{1}{8} sd\sqrt{Km} = \frac{1}{8}\sqrt{Ksdn}\,.$$

Lemma D.3 requires us to set $\Delta/s = \sqrt{K/m}/2$. The only thing left to do is to find the values of m such that $\Delta = s\sqrt{\frac{K}{4m}} \leq 1/(s^{1/d}K)$, which ensures that $|\Theta|$ is smaller than $\mathcal{M}(\mathbb{B}_2^d(1), \|\cdot\|_2, \Delta)$. This constraint for m can be rearranged into

$$m \geq \tfrac{1}{4} s^{2+2/d} K^3\,.$$

Thus we can choose any m satisfying $m \geq s^{2+2/d}K^3$. □

E Proof of Theorem 4.1

To bound the expected regret of FGTS, we use the decoupling technique developed by Zhang (2022). Theorem 1 in Zhang (2022) provides the following upper bound on the expected regret of FGTS.

Theorem E.1 (Theorem 1 in Zhang (2022))**.** *Consider FGTS with the posterior defined in* (4) *and the likelihood defined in* (3)*. Suppose that the prior p_1 is chosen such that $\mathbb{P}_{\nu\sim p_1}[\max_{x,a}|f_\nu(x,a)| \leq 1] = 1$. For any $\eta \leq 1/4$ and any $\lambda > 0$,*

$$R_n(f^*) \leq \frac{\lambda K n}{\eta} + 6\lambda n - \frac{1}{\lambda} Z_n\,, \quad \textit{where} \;\; Z_n := \mathbb{E}\log\mathbb{E}_{\nu\sim p_1}\exp\bigg(-\sum_{t=1}^{n}\Delta L(\nu, X_t, A_t, Y_t)\bigg)\,, \tag{11}$$

and

$$\Delta L(\nu, x, a, y) := \eta\left[(f_\nu(x,a) - y)^2 - (f^*(x,a) - y)^2\right] - \lambda\left[f_\nu(x) - f^*(x)\right]\,.$$

Note that $\Delta L(\nu, x, a, y)$ is the logarithm of the likelihood ratio with parameter ν^* in the denominator and ν in the numerator. Hence, we call ΔL the log-likelihood ratio. We use the shorthand $\Delta L_t(\nu) := \Delta L_t(\nu, X_t, A_t, Y_t)$. For countable sparsity, we use $\Delta L_t(\nu)$ and $\Delta L_t(w)$ interchangeably.

The proof of Theorem 4.1 uses Theorem E.1 and a bound on the log partition function Z_n (defined in (11)). To bound Z_n, we use some auxiliary lemmas. In Section E.1, we state and prove these auxiliary lemmas. In Section E.2, we then state and prove a bound on Z_n. Finally, in Section E.3, we prove Theorem 4.1.

We recall here some notation introduced in Sections 2.2 and 4.2. We use s and S to denote the sparsity and support of w^*. For any non-empty subset $M \subseteq \mathbb{N}$, we define $\mathcal{W}_M := \{w : \|w\|_1 \leq 1,\ w_i = 0\ \forall i \notin M\}$. We let $\bar{w}$ denote the projection of the sequence w^* onto the set of parameter sequences with support contained in $[d_{\text{eff}}]$. Thus $\bar{w}$ is the sequence such that for all $i \in S \cap [d_{\text{eff}}]$, $\bar{w}_i = w_i^*$, and for all $i \notin S \cap [d_{\text{eff}}]$, $\bar{w}_i = 0$. We let $\bar{S} = \text{supp}(\bar{w}) = \text{supp}(w^*) \cap [d_{\text{eff}}]$. Note that if $\bar{S} = \emptyset$, then the regret of any algorithm is $\mathcal{O}(\sqrt{n})$. In particular,

$$\begin{aligned} R_n(f^*) &= \mathbb{E}\left[\textstyle\sum_{t=1}^{n}\max_{a\in[K]}\left\{\sum_{i=d_{\text{eff}}+1}^{\infty} w_i^*\phi_i(X_{t,a})\right\} - \sum_{i=d_{\text{eff}}+1}^{\infty} w_i^*\phi_i(X_{t,A_t})\right] \\ &\leq 2n\|w^*\|_1\|\phi_{d_{\text{eff}}+1}\|_\infty \leq 2\sqrt{n}\,. \end{aligned}$$

Therefore, we continue under the assumption that $\bar{S} \neq \emptyset$. For each $c \in (0,1]$, we define the set $\mathcal{W}_c = \{(1-c)\bar{w} + cw : w \in \mathcal{W}_{\bar{S}}\} \subseteq \mathcal{W}_{\bar{S}}$. We notice that for every $w \in \mathcal{W}_c$,

$$\|\bar{w} - w\|_1 = \|\bar{w} - (1-c)\bar{w} - cw'\|_1 = c\|\bar{w} - w'\|_1 \leq 2c\,. \tag{12}$$

E.1 Auxiliary Lemmas

The first auxiliary lemma provides an alternative expression for the log-likelihood ratio ΔL.

Lemma E.2.

$$\begin{aligned} \Delta L(\nu, X_t, A_t, Y_t) = \eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))^2 - 2\eta\epsilon_t(f_\nu(X_t, A_t) - f^*(X_t, A_t)) \\ + \lambda(f_\nu(X_t) - f^*(X_t))\,. \end{aligned}$$

Proof. By definition of ΔL, we have

$$\Delta L(\nu, X_t, A_t, Y_t) = \eta\left[(f_\nu(X_t, A_t) - Y_t)^2 - (f^*(X_t, A_t) - Y_t)^2\right] - \lambda(f_\nu(X_t)) - f^*(X_t))\,.$$

By expanding and rearranging the squared terms, we obtain

$$\begin{aligned} (f_\nu(X_t, A_t) - Y_t)^2 - (f^*(X_t, A_t) - Y_t)^2 &= (f_\nu(X_t, A_t) - f^*(X_t, A_t) - \epsilon_t)^2 - \epsilon_t^2 \\ &= (f_\nu(X_t, A_t) - f^*(X_t, A_t))^2 \\ &\quad - 2\epsilon_t(f_\nu(X_t, A_t) - f^*(X_t, A_t)). \end{aligned}$$

□

The next lemma utilises the uniform decay condition. It shows that for any parameter sequence w with support contained in $[d_{\text{eff}}]$, the difference between f_w and f^* can be bounded in terms of the ℓ_1 distance between w and $\bar{w}$ and a small approximation error caused by ignoring any components of w^* with indices greater than d_{eff}.

Lemma E.3. *For any w with support contained in $[d_{\text{eff}}]$, and all $x \in \mathcal{X}$ and $a \in [K]$,*

$$|f_w(x,a) - f^*(x,a)| \leq \|w - \bar{w}\|_1 + 1/\sqrt{n}\,.$$

Proof. Using the triangle inequality, we obtain

$$\begin{aligned}|f_w(x,a) - f^*(x,a)| &= \left|\sum_{i=1}^{d_{\text{eff}}}(w_i - \bar{w}_i)\phi_i(x_a) - \sum_{i=d_{\text{eff}}+1}^{\infty} w_i^*\phi_i(x_a)\right| \\ &\leq \|w - \bar{w}\|_1\|\phi_1\|_\infty + \|w^*\|_1\|\phi_{d_{\text{eff}}+1}\|_\infty \\ &\leq \|w - \bar{w}\|_1 + 1/\sqrt{n}\,.\end{aligned}$$

The fact that $\|\phi_{d_{\text{eff}}+1}\|_\infty \leq 1/\sqrt{n}$ follows from the definition of d_{eff}. □

As a consequence of this lemma, we also have

$$(f_w(x,a) - f^*(x,a))^2 \leq \|w - \bar{w}\|_1^2 + 2\|w - \bar{w}\|_1/\sqrt{n} + 1/n\,.$$

In addition, since $|f_w(x) - f^*(x)| \leq \max_{a\in[K]}|f_w(x,a) - f^*(x,a)|$, we also have

$$|f_w(x) - f^*(x)| \leq \|w - \bar{w}\|_1 + 1/\sqrt{n}\,.$$

Using Lemma E.3, we obtain the following exponential moment bound, which we will use later to control the terms depending on $\epsilon_1, \dots, \epsilon_n$ that appear in Lemma E.2.

Lemma E.4. *For any fixed $w \in \mathcal{W}_c$,*

$$\mathbb{E}\left[\exp(\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t)\right] \leq \exp\left(\frac{\eta^2(4c^2n + 4c\sqrt{n} + 1)}{2}\right)\,.$$

Proof. We recall that $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_{t-1}, X_t, A_t]$ and that each ϵ_t is sub-Gaussian, meaning $\mathbb{E}_t[\exp(\lambda\epsilon_t)] \leq \exp(\lambda^2/8)$ for any $\lambda \in \mathbb{R}$. Using this, Lemma E.3 and the fact that $w \in \mathcal{W}_c$, we have

$$\begin{aligned}&\mathbb{E}\left[\exp(\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t)\right] \\ &= \mathbb{E}\left[\exp(\textstyle\sum_{t=1}^{n-1} - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t)\mathbb{E}_n\left[\exp(-2\eta(f_w(X_n,A_n) - f^*(X_n,A_n))\epsilon_n)\right]\right] \\ &\leq \mathbb{E}\left[\exp(\textstyle\sum_{t=1}^{n-1} - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t)\exp(\frac{\eta^2(f_w(X_n,A_n) - f^*(X_n,A_n))^2}{2})\right] \\ &\leq \mathbb{E}\left[\exp(\textstyle\sum_{t=1}^{n-1} - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t)\right]\exp\left(\frac{\eta^2(4c^2 + 4c/\sqrt{n} + 1/n)}{2}\right)\,,\end{aligned}$$

where the last inequality follows from Lemma E.3 and (12). By iterating this argument, we obtain the inequality in the statement of the Lemma. □

The next lemma provides a bound on the covering number of $\mathcal{W}_c$.

Lemma E.5 (Lemma 7 in Wainwright (2019))**.** *For any $p \geq 1$, $d \geq 1$, $c > 0$ and $\Delta > 0$,*

$$\mathcal{N}(\mathbb{B}_p^d(c), \|\cdot\|_p, \Delta) \leq (1 + \tfrac{2c}{\Delta})^d\,.$$

It is easy to see that there is a surjective isometric embedding from the set $\mathcal{W}_c$ (with the ℓ_1 norm) to the ball $\mathbb{B}_1^{\bar{s}}(c)$ (also with the ℓ_1 norm). In particular, to embed $w \in \mathcal{W}_c$ into $\mathbb{B}_1^{\bar{s}}(c)$, one can subtract $(1-c)\bar{w}$ and then remove all the components corresponding to indices not in $\bar{S}$. Therefore, a consequence of Lemma E.5 is that $\mathcal{N}(\mathcal{W}_c, \|\cdot\|_1, \Delta) \le (1 + \frac{2c}{\Delta})^{\bar{s}}$. The final auxiliary lemma controls the expected value of the maximum of the noise process originating from Lemma E.2.

Lemma E.6. *For any $c \in (0,1]$ and $\Delta > 0$,*

$$\mathbb{E}\left[\max_{w\in\mathcal{W}_c} \{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\}\right] \le \bar{s}\log(1+\frac{2c}{\Delta}) + \frac{\eta^2(4c^2n+4c\sqrt{n}+1)}{2} + \Delta\eta n\,.$$

Proof. We set $\mathcal{W}_{c,\Delta}$ to be any minimal ℓ_1-norm Δ-covering of $\mathcal{W}_c$. We define $[w] := \arg\min_{w'\in\mathcal{W}_{c,\Delta}} \|w - w'\|_1$ to be the ℓ_1-norm projection of $w \in \mathcal{W}_c$ into $\mathcal{W}_{c,\Delta}$. The first step is to replace the maximum over the infinite set $\mathcal{W}_c$ by a maximum over the finite set $\mathcal{W}_{c,\Delta}$ and a discretisation error. We have

$$\begin{aligned}\max_{w\in\mathcal{W}_c} \{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\} &\le \max_{w\in\mathcal{W}_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_{[w]}(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\right\}\\ &\quad+ \max_{w\in\mathcal{W}_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f_{[w]}(X_t,A_t))\epsilon_t\right\}\\ &= \max_{w\in\mathcal{W}_{c,\Delta}} \{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\}\\ &\quad+ \max_{w\in\mathcal{W}_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f_{[w]}(X_t,A_t))\epsilon_t\right\}\,.\end{aligned}$$

To bound the expectation of the maximum over $\mathcal{W}_{c,\Delta}$, we use Jensen's inequality, Lemma E.4 and Lemma E.5 to obtain

$$\begin{aligned}&\mathbb{E}\left[\max_{w\in\mathcal{W}_{c,\Delta}} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\right\}\right]\\ &\le \log\mathbb{E}\left[\max_{w\in\mathcal{W}_{c,\Delta}} \left\{\exp\left(\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\right)\right\}\right]\\ &\le \log\mathbb{E}\left[\textstyle\sum_{w\in\mathcal{W}_{c,\Delta}}\left\{\exp\left(\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f^*(X_t,A_t))\epsilon_t\right)\right\}\right]\\ &\le \log\left(|\mathcal{W}_{c,\Delta}|\exp\left(\tfrac{\eta^2(4c^2n+4c\sqrt{n}+1)}{2}\right)\right)\\ &= \log(\mathcal{N}(\mathcal{W}_c, \|\cdot\|_1, \Delta)) + \tfrac{\eta^2(4c^2n+4c\sqrt{n}+1)}{2}\\ &\le \bar{s}\log(1+\tfrac{2c}{\Delta}) + \tfrac{\eta^2(4c^2n+4c\sqrt{n}+1)}{2}\,.\end{aligned}$$

Since, $\mathcal{W}_{c,\Delta}$ is a Δ-covering and $\|\phi_i\|_\infty \le 1$, we have $|f_w(x,a) - f_{[w]}(x,a)| \le \Delta$ for all $x \in \mathcal{X}$ and $a \in [K]$. Also, since each ϵ_t is 1/2-sub-Gaussian, $\mathbb{E}[\epsilon_t^2] \le 1/4$. Using these facts and the Cauchy-Schwartz inequality, we can bound the discretisation error as

$$\begin{aligned}&\mathbb{E}\left[\max_{w\in\mathcal{W}_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t,A_t) - f_{[w]}(X_t,A_t))\epsilon_t\right\}\right]\\ &\le \mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\sqrt{\textstyle\sum_{t=1}^n 4\eta^2(f_w(X_t,A_t) - f_{[w]}(X_t,A_t))^2}\right\}\sqrt{\textstyle\sum_{t=1}^n\epsilon_t^2}\right]\\ &\le 2\eta\Delta\sqrt{n}\mathbb{E}\left[\sqrt{\textstyle\sum_{t=1}^n\epsilon_t^2}\right]\\ &\le 2\eta\Delta\sqrt{n}\sqrt{\textstyle\sum_{t=1}^n\mathbb{E}[\epsilon_t^2]}\\ &\le \eta\Delta n\,.\end{aligned}$$

□

E.2 Bounding the Log Partition Function

Using the auxiliary lemmas established in the previous subsection, we can now prove a bound on the log partition function Z_n.

Lemma E.7. *If we use the prior p_1 in* (5)*, then for every n,*

$$-Z_n \leq \|w^*\|_0 \log(8ed_{\text{eff}}n) + 2\eta^2 + 5\eta + 2\lambda\sqrt{n}\,.$$

Proof. For each $c \in (0,1]$, we define the set $\mathcal{W}_c = \{(1-c)\bar{w} + cw : w \in \mathcal{W}_{\bar{S}}\} \subseteq \mathcal{W}_{\bar{S}}$ and define the event

$$\mathcal{E}_c := \{w \in \mathcal{W}_c\}\,.$$

When $w \sim p_1(w \mid M = \bar{S}) = \mathcal{U}(\mathcal{W}_{\bar{S}})$, we have

$$\mathbb{P}_{w\sim p_1(\cdot|M=\bar{S})}[\mathcal{E}_c] = \tfrac{\text{Vol}(\mathcal{W}_c)}{\text{Vol}(\mathcal{W}_{\bar{S}})} = c^{\bar{s}}\,.$$

Since we know that $p_1(M) = 2^{-|M|}\binom{d_{\text{eff}}}{|M|}^{-1}(\sum_{i=1}^{d_{\text{eff}}} 2^{-i})$, we have

$$\begin{aligned}
Z_n &= \mathbb{E}\left[\log \mathbb{E}_{w\sim p_1}\left[\exp(-\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t))\right]\right] \\
&\geq \mathbb{E}\left[\log p_1(\bar{S})\mathbb{E}_{w\sim p_1|\bar{S}}\left[\exp(\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t))\right]\right] \\
&\geq \mathbb{E}\left[\log p_1(\bar{S})\mathbb{P}_{w\sim p_1|M=\bar{S}}[\mathcal{E}_c] \min_{\nu\in\mathcal{W}_c}\left\{\exp(-\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t))\right\}\right] \\
&= \log(p_1(\bar{S})) + \log(\mathbb{P}_{w\sim p_1|M=\bar{S}}[\mathcal{E}_c]) - \mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t)\right\}\right] \\
&= -\bar{s}\log(2) - \log(\tbinom{d_{\text{eff}}}{\bar{s}}) - \log(\textstyle\sum_{i=1}^{d_{\text{eff}}} 2^{-i}) - \bar{s}\log(1/c) - \mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t)\right\}\right]\,.
\end{aligned}$$

Next, using Lemma E.2, Lemma E.3 and (12), for any $w \in \mathcal{W}_c$, we have

$$\begin{aligned}
\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t) &= \textstyle\sum_{t=1}^n \eta(f_w(X_t, A_t) - f^*(X_t, A_t))^2 + \lambda(f_w(X_t) - f^*(X_t)) \\
&\quad + \textstyle\sum_{t=1}^n - 2\eta(f_w(X_t, A_t) - f^*(X_t, A_t))\epsilon_t \\
&\leq 4c^2\eta n + 4c\eta\sqrt{n} + \eta + 2c\lambda n + \lambda\sqrt{n} \\
&\quad + \textstyle\sum_{t=1}^n - 2\eta(f_w(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t)\right\}\right] &\leq 4c^2\eta n + 4c\eta\sqrt{n} + \eta + 2c\lambda n + \lambda\sqrt{n} \\
&\quad + \mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\}\right]\,.
\end{aligned}$$

Using Lemma E.6, for any $c \in (0,1]$ and $\Delta > 0$, we also have

$$\mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\}\right] \leq \bar{s}\log(1 + \tfrac{2c}{\Delta}) + \tfrac{\eta^2(4c^2n + 4c\sqrt{n} + 1)}{2} + \Delta\eta n\,.$$

If we choose $c = 1/(2\sqrt{n})$ and $\Delta = 1/n$, then this bound becomes

$$\mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n - 2\eta(f_w(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\}\right] \leq \bar{s}\log(1 + \sqrt{n}) + 2\eta^2 + \eta\,.$$

Thus, with the choice $c = 1/(2\sqrt{n})$, we obtain the bound

$$\mathbb{E}\left[\max_{w\in\mathcal{W}_c}\left\{\textstyle\sum_{t=1}^n \Delta L(w, X_t, A_t, Y_t)\right\}\right] \leq \bar{s}\log(1 + \sqrt{n}) + 2\eta^2 + 5\eta + 2\lambda\sqrt{n}\,.$$

If we combine everything, and use the inequality $\binom{d_{\mathrm{eff}}}{\bar{s}} \le (ed_{\mathrm{eff}}/\bar{s})^{\bar{s}}$, then we obtain

$$\begin{aligned}
-Z_n &\le \bar{s}\log(2) + \log(\tbinom{d_{\mathrm{eff}}}{\bar{s}}) + \log(\textstyle\sum_{i=1}^{d_{\mathrm{eff}}} 2^{-i}) + \bar{s}\log(2\sqrt{n}) + \bar{s}\log(1+\sqrt{n}) + 2\eta^2 + 5\eta + 2\lambda\sqrt{n} \\
&\le \|w^*\|_0\big(\log(2) + \log(ed_{\mathrm{eff}}) + 2\log(2\sqrt{n})\big) + 2\eta^2 + 5\eta + 2\lambda\sqrt{n} \\
&\le \|w^*\|_0 \log(8ed_{\mathrm{eff}}n) + 2\eta^2 + 5\eta + 2\lambda\sqrt{n}\,.
\end{aligned}$$

□

E.3 Proof of Theorem 4.1

Proof. Using Theorem E.1, and then Lemma E.7 (with $\eta = 1/4$), we have

$$\begin{aligned}
R_n(f^*) &\le \lambda(4K+6)n - \frac{1}{\lambda}Z_n \\
&\le \lambda(4K+6)n + \frac{1}{\lambda}\big(\|w^*\|_0\log(8ed_{\mathrm{eff}}n) + 2\big) + 2\sqrt{n}\,.
\end{aligned}$$

If we choose

$$\lambda = \sqrt{\frac{\log(8ed_{\mathrm{eff}}n)}{(4K+6)n}}\,,$$

then we obtain the regret bound

$$\begin{aligned}
R_n(f^*) &\le (\|w^*\|_0 + 1)\sqrt{(4K+6)n\log(8ed_{\mathrm{eff}}n)} + 2\sqrt{\frac{(4K+6)n}{\log(8ed_{\mathrm{eff}}n)}} + 2\sqrt{n} \\
&= \mathcal{O}\Big(\|w^*\|_0\sqrt{Kn\log(d_{\mathrm{eff}}n)}\Big)\,.
\end{aligned}$$

If s is a known upper bound on $\|w^*\|_0$ and we choose

$$\lambda = \sqrt{\frac{s\log(8ed_{\mathrm{eff}}n) + 2}{(4K+6)n}}\,,$$

then we obtain the regret bound

$$\begin{aligned}
R_n(f^*) &\le 2\sqrt{(4K+6)(s\log(8ed_{\mathrm{eff}}n) + 2)n} + 2\sqrt{n} \\
&= \mathcal{O}(\sqrt{Ksn\log(d_{\mathrm{eff}}n)})\,.
\end{aligned}$$

□

F Proof of Theorem 4.2

This section follows a similar structure to the previous one. In Section F.1, we state and prove some auxiliary lemmas. In Section F.2, we state and prove a bound on Z_n. Finally, in Section F.3, we prove Theorem 4.2.

In this section, unless stated otherwise, we let $s = \|w^*\|_0$ denote the true sparsity. We recall here some notation introduced in Section 4.3. For each $c \in (0,1]$ we define the sets

$$\mathcal{W}_c := \{(1-c)w^* + cw : w \in \mathbb{B}_1^s(1)\}, \qquad \Theta_{i,c} := \{(1-c)\theta_i^* + c\theta : \theta \in \mathbb{B}_2^d(1)\}\,.$$

We notice that for each $w \in \mathcal{W}_c$,

$$\|w^* - w\|_1 = c\|w^* - w'\|_1 \le c\|w^*\|_1 + c\|w'\|_1 \le 2c\,,$$

where w' is some element in $\mathbb{B}_1^s(1)$. Similarly, for each $i \in [s]$ and $\theta_i \in \Theta_{i,c}$, we have $\|\theta_i^* - \theta_i\|_2 \le 2c$. For each $c \in (0,1]$, let $N_c = \mathcal{W}_c \times \Theta_{1,c} \times \cdots \times \Theta_{s,c}$

F.1 Auxiliary Lemmas

First, we show that if ϕ satisfies the Lipschitz property in Definition 2.2, then for any $x \in \mathcal{X}$ and $a \in [K]$, the function value $f_\nu(x, a)$ changes smoothly as ν is varied.

Lemma F.1. *For any $m \in \mathbb{N}$ and any $\nu, \nu' \in \mathbb{R}^m \times \Theta^m$,*

$$\forall x \in \mathcal{X}, a \in [K],\ |f_\nu(x,a) - f_{\nu'}(x,a)| \le \max_{i \in [m]}\{\|\theta_i - \theta_i'\|_2\} + \|w - w'\|_1.$$

Proof. Using the triangle inequality and the uniform Lipschitz property, we have

$$\begin{aligned}
|f_\nu(x,a) - f_{\nu'}(x,a)| &= \left|\sum_{i=1}^m (w_i\phi(x_a, \theta_i) - w_i'\phi(x_a, \theta_i'))\right| \\
&= \left|\sum_{i=1}^m (w_i\phi(x_a, \theta_i) - w_i\phi(x_a, \theta_i') + w_i\phi(x_a, \theta_i') - w_i'\phi(x_a, \theta_i'))\right| \\
&\le \sum_{i=1}^m |w_i||\phi(x_a, \theta_i) - \phi(x_a, \theta_i')| + \sum_{i=1}^m |w_i - w_i'||\phi(x_a, \theta_i')| \\
&\le \max_{i \in [m]}\{\|\theta_i - \theta_i'\|_2\} + \|w - w'\|_1 \,.
\end{aligned}$$

□

Note that this upper bound also applies to $f_\nu(x)$ and $f_{\nu'}(x)$. In particular,

$$\begin{aligned}
|f_\nu(x) - f_{\nu'}(x)| &= |\max_{a \in [K]}\{f_\nu(x,a)\} - \max_{a \in [K]}\{f_{\nu'}(x,a)\}| \\
&\le |\max_{a \in [K]}\{f_\nu(x,a) - f_{\nu'}(x,a)\}| \\
&\le \max_{i \in [m]}\{\|\theta_i - \theta_i'\|_2\} + \|w - w'\|_1 \,.
\end{aligned}$$

Using Lemma F.1, we obtain the following exponential moment bound.

Lemma F.2. *For any fixed $\nu \in N_c$,*

$$\mathbb{E}\left[\exp\left(\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t)\epsilon_t\right)\right] \le \exp\left(8c^2\eta^2 n\right).$$

Proof. Using the sub-Gaussian property of $\epsilon_1, \dots, \epsilon_n$, Lemma F.1 and the fact that $\nu \in N_c$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right)\right] \\
&= \mathbb{E}\Big[\exp\left(\textstyle\sum_{t=1}^{n-1} - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right)\mathbb{E}_n\left[\exp\left(- 2\eta(f_\nu(X_n, A_n) - f^*(X_n, A_n))\epsilon_n\right)\right]\Big] \\
&\le \mathbb{E}\Big[\exp\left(\textstyle\sum_{t=1}^{n-1} - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right)\exp\left(\tfrac{\eta^2(f_\nu(X_n,A_n) - f^*(X_n,A_n))^2}{2}\right)\Big] \\
&\le \mathbb{E}\Big[\exp\left(\textstyle\sum_{t=1}^{n-1} - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right)\Big]\exp\left(8c^2\eta^2\right).
\end{aligned}$$

By iterating this argument, we obtain the inequality in the statement of the lemma. □

We set $\mathcal{W}_{c,\Delta}$ to be any minimal ℓ_1-norm Δ-covering of $\mathcal{W}_c$ and (for each $i \in s$) $\Theta_{i,c,\Delta}$ to be any minimal ℓ_2-norm Δ-covering of $\Theta_{i,c}$. We let $N_{c,\Delta} = \mathcal{W}_{c,\Delta} \times \Theta_{1,c,\Delta} \times \cdots \times \Theta_{s,c,\Delta}$. Using Lemma E.5, we obtain a bound on the cardinality of $N_{c,\Delta}$.

Corollary F.3. *For any $c \in (0,1]$ and $\Delta > 0$,*

$$|N_{c,\Delta}| \leq \left(1 + \tfrac{2c}{\Delta}\right)^{s(d+1)} .$$

Proof. Using Lemma E.5, we have

$$|\mathcal{W}_{c,\Delta}| \leq \mathcal{N}(\mathbb{B}_1^s(c), \|\cdot\|_1, \Delta) \leq \left(1 + \tfrac{2c}{\Delta}\right)^{s} ,$$

and, for any $i \in [s]$,

$$|\Theta_{i,c,\Delta}| \leq \mathcal{N}(\mathbb{B}_2^d(c), \|\cdot\|_2, \Delta) \leq \left(1 + \tfrac{2c}{\Delta}\right)^{d} .$$

Therefore,

$$|N_{c,\Delta}| = |\mathcal{W}_{c,\Delta}| \times |\Theta_{1,c,\Delta}| \times \cdots \times |\Theta_{s,c,\Delta}| \leq \left(1 + \tfrac{2c}{\Delta}\right)^{s(d+1)} .$$

□

The final auxiliary lemma is analogous to Lemma E.6. It controls the expectation of the maximum of the noise process in from Lemma E.2 for the case of uncountable sparsity.

Lemma F.4. *For any $c \in (0,1]$ and $\Delta > 0$,*

$$\mathbb{E}\left[\max_{\nu \in N_c} \{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\}\right] \leq s(d+1)\log(1 + \tfrac{2c}{\Delta}) + 8c^2\eta^2 n + 2\eta\Delta n .$$

Proof. We define $[w] := \arg\min_{w' \in \mathcal{W}_{c,\Delta}} \|w - w'\|_1$ to be the ℓ_1-norm projection of $w \in \mathcal{W}_c$ into $\mathcal{W}_{c,\Delta}$. For each $i \in [s]$, we define $[\theta_i] := \arg\min_{\theta' \in \Theta_{i,c,\Delta}} \|\theta_i - \theta'\|_2$ to be the ℓ_2-norm projection of $\theta_i \in \Theta_{i,c}$ into $\Theta_{i,c,\Delta}$. For any $\nu \in N_c$, we define $[\nu] := ([w], [\theta_1], \ldots, [\theta_s])$. Using Lemma F.1, and the fact that $\mathcal{W}_{c,\Delta}, \Theta_{1,c,\Delta}, \ldots, \Theta_{s,c,\Delta}$ are Δ-coverings, we have

$$\forall x \in \mathcal{X}, a \in [K], \quad |f_\nu(x,a) - f_{[\nu]}(x,a)| \leq \max_{i \in [s]}\{\|\theta_i - [\theta_i]\|_2\} + \|w - [w]\|_1 \leq 2\Delta .$$

The first step is to replace the maximum over the infinite set N_c by a maximum over the finite set $N_{c,\Delta}$ and a discretisation error. We have

$$\begin{aligned}
\max_{\nu \in N_c} \{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\} &\leq \max_{\nu \in N_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_{[\nu]}(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\} \\
&\quad + \max_{\nu \in N_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f_{[\nu]}(X_t, A_t))\epsilon_t\right\} \\
&= \max_{\nu \in N_{c,\Delta}} \{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\} \\
&\quad + \max_{\nu \in N_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f_{[\nu]}(X_t, A_t))\epsilon_t\right\} .
\end{aligned}$$

To bound the expectation of the maximum over $N_{c,\Delta}$, we use Jensen's inequality, Lemma F.2 and Corollary F.3 to obtain

$$\begin{aligned}
&\mathbb{E}\left[\max_{\nu \in N_{c,\Delta}} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\}\right] \\
&\leq \log \mathbb{E}\left[\max_{\nu \in N_{c,\Delta}} \left\{\exp\left(\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right)\right\}\right] \\
&\leq \log \mathbb{E}\left[\textstyle\sum_{\nu \in N_{c,\Delta}} \left\{\exp\left(\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right)\right\}\right] \\
&\leq \log\left(|N_{c,\Delta}| \exp\left(8c^2\eta^2 n\right)\right) \\
&= s(d+1)\log(1 + \tfrac{2c}{\Delta}) + 8c^2\eta^2 n .
\end{aligned}$$

Using the sub-Gaussian property of $\epsilon_1, \dots, \epsilon_n$ and the Cauchy-Schwartz inequality, we can bound the discretisation error as

$$\begin{aligned}
&\mathbb{E}\left[\max_{\nu \in N_c} \left\{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f_{[\nu]}(X_t, A_t))\epsilon_t\right\}\right] \\
&\le \mathbb{E}\left[\max_{\nu \in N_c} \left\{\sqrt{\textstyle\sum_{t=1}^n 4\eta^2(f_\nu(X_t, A_t) - f_{[\nu]}(X_t, A_t))^2}\right\} \sqrt{\textstyle\sum_{t=1}^n \epsilon_t^2}\right] \\
&\le 4\eta\Delta\sqrt{n}\,\mathbb{E}\left[\sqrt{\textstyle\sum_{t=1}^n \epsilon_t^2}\right] \\
&\le 4\eta\Delta\sqrt{n}\sqrt{\textstyle\sum_{t=1}^n \mathbb{E}[\epsilon_t^2]} \\
&\le 2\eta\Delta n\,.
\end{aligned}$$

□

F.2 Bounding the Log Partition Function

Lemma F.5. *If we use the prior p_1 in* (7)*, then for every n,*

$$-Z_n \le \|w^*\|_0(d+1)\log(4\sqrt{n}) + 4\eta + 2\lambda\sqrt{n}\,.$$

Proof. For each $c \in (0, 1]$, we define the event

$$\mathcal{E}_c := \{\nu \in N_c\}\,.$$

When $w \sim p_1(w \mid m = s) = \mathcal{U}(\mathbb{B}_1^s(1))$ and $\theta_i \sim p_1(\theta) = \mathcal{U}(\mathbb{B}_2^d(1))$, we have

$$\mathbb{P}_{\nu \sim p_1 \mid m = s}[\mathcal{E}_c] = \tfrac{\mathrm{Vol}(\mathcal{W}_c)}{\mathrm{Vol}(\mathbb{B}_1^s(1))} \times \tfrac{\mathrm{Vol}(\Theta_{1,c})}{\mathrm{Vol}(\mathbb{B}_2^d(1))} \times \cdots \times \tfrac{\mathrm{Vol}(\Theta_{s,c})}{\mathrm{Vol}(\mathbb{B}_2^d(1))} = c^{s(d+1)}\,.$$

Using this, and the fact that $p_1(m) = 2^{-m}$, we have

$$\begin{aligned}
Z_n &= \mathbb{E}\left[\log \mathbb{E}_{\nu \sim p_1}\left[\exp(-\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t))\right]\right] \\
&\ge \mathbb{E}\left[\log p_1(s)\mathbb{E}_{\nu \sim p_1 \mid m = s}\left[\exp(-\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t))\right]\right] \\
&\ge \mathbb{E}\left[\log p_1(s)\mathbb{P}_{\nu \sim p_1 \mid m = s}[\mathcal{E}_c] \min_{\nu \in N_c}\left\{\exp(-\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t))\right\}\right] \\
&= \log(p_1(s)) + \log(\mathbb{P}_{\nu \sim p_1 \mid m = s}[\mathcal{E}_c]) - \mathbb{E}\left[\max_{\nu \in N_c}\left\{\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t)\right\}\right] \\
&= -s\log(2) - s(d+1)\log(1/c) - \mathbb{E}\left[\max_{\nu \in N_c}\left\{\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t)\right\}\right]
\end{aligned}$$

Next, using Lemma E.2 and then Lemma F.1, for any $\nu \in N_c$, we have

$$\begin{aligned}
\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t) &= \textstyle\sum_{t=1}^n \eta(f_\nu(X_t, A_t) - f_{\nu^*}(X_t, A_t))^2 + \lambda(f_\nu(X_t) - f_{\nu^*}(X_t)) \\
&\quad + \textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t \\
&\le 16c^2\eta n + 4c\lambda n + \textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\max_{\nu \in N_c}\left\{\textstyle\sum_{t=1}^n \Delta L(\nu, X_t, A_t, Y_t)\right\}\right] &\le 16c^2\eta n + 4c\lambda n \\
&\quad + \mathbb{E}\left[\max_{\nu \in N_c}\left\{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\}\right]\,.
\end{aligned}$$

Using Lemma F.4, for any $c \in (0, 1]$ and $\Delta > 0$, we also have

$$\mathbb{E}\left[\max_{\nu \in N_c}\left\{\textstyle\sum_{t=1}^n - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\right\}\right] \le s(d+1)\log(1 + \tfrac{2c}{\Delta}) + 8c^2\eta^2 n + 2\eta\Delta n\,.$$

If we choose $c = 1/(2\sqrt{n})$ and $\Delta = 1/n$, then this bound becomes

$$\mathbb{E}\left[\max_{\nu \in N_c} \{\textstyle\sum_{t=1}^{n} - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\}\right] \leq s(d+1)\log(1+\sqrt{n}) + 2\eta^2 + 2\eta\,.$$

Thus with the choice $c = 1/(2\sqrt{n})$, we obtain the bound

$$\mathbb{E}\left[\max_{\nu \in N_c} \{\textstyle\sum_{t=1}^{n} - 2\eta(f_\nu(X_t, A_t) - f^*(X_t, A_t))\epsilon_t\}\right] \leq s(d+1)\log(1+\sqrt{n}) + 2\eta^2 + 6\eta + 2\lambda\sqrt{n}\,.$$

If we combine everything, then we obtain

$$\begin{aligned} -Z_n &\leq s\log(2) + s(d+1)\log(2\sqrt{n}) + s(d+1)\log(1+\sqrt{n}) + 2\eta^2 + 6\eta + 2\lambda\sqrt{n} \\ &\leq s(d+1)(\log(2) + 2\log(2\sqrt{n})) + 2\eta^2 + 6\eta + 2\lambda\sqrt{n} \\ &= s(d+1)\log(8n) + 2\eta^2 + 6\eta + 2\lambda\sqrt{n}\,. \end{aligned}$$

□

F.3 Proof of Theorem 4.2

Proof. Using Theorem E.1, and then Lemma F.5 (with $\eta = 1/4$), we have

$$\begin{aligned} R_n(f^*) &\leq \lambda(4K+6)n - \frac{1}{\lambda}Z_n \\ &\leq \lambda(4K+6)n + \frac{1}{\lambda}\big(\|w^*\|_0(d+1)\log(8n) + 2\big) + 2\sqrt{n}\,. \end{aligned}$$

If we choose

$$\lambda = \sqrt{\frac{(d+1)\log(8n)}{(4K+6)n}}\,,$$

then we obtain the regret bound

$$\begin{aligned} R_n(f^*) &\leq (\|w^*\|_0 + 1)\sqrt{(4K+6)(d+1)n\log(8n)} + 2\sqrt{\frac{(4K+6)n}{\log(8n)}} + 2\sqrt{n} \\ &= \mathcal{O}\Big(\|w^*\|_0\sqrt{Kdn\log(n)}\Big)\,. \end{aligned}$$

If s is a known upper bound on $\|w^*\|_0$ and we choose

$$\lambda = \sqrt{\frac{s(d+1)\log(8n) + 2}{(4K+6)n}}\,,$$

then we obtain the regret bound

$$\begin{aligned} R_n(f^*) &\leq 2\sqrt{(4K+6)(s(d+1)\log(8n) + 2)n} + 2\sqrt{n} \\ &= \mathcal{O}(\sqrt{Ksdn\log(n)})\,. \end{aligned}$$

□