Proceedings of Machine Learning Research vol 195:1–48, 2023 36th Annual Conference on Learning Theory

Oracle-Efficient Smoothed Online Learning for Piecewise Continuous Decision Making

Adam Block ABLOCK@MIT.EDU
Alexander Rakhlin RAKHLIN@MIT.EDU
Max Simchowitz MSIMCHOW@CSAIL.MIT.EDU
MIT

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Smoothed online learning has emerged as a popular framework to mitigate the substantial loss in statistical and computational complexity that arises when one moves from classical to adversarial learning. Unfortunately, for some spaces, it has been shown that efficient algorithms suffer an exponentially worse regret than that which is minimax optimal, even when the learner has access to an optimization oracle over the space. To mitigate that exponential dependence, this work introduces a new notion of complexity, the generalized bracketing numbers, which marries constraints on the adversary to the size of the space, and shows that an instantiation of Follow-the-Perturbed-Leader can attain low regret with the number of calls to the optimization oracle scaling optimally with respect to average regret. We then instantiate our bounds in several problems of interest, including online prediction and planning of piecewise continuous functions, which has many applications in fields as diverse as econometrics and robotics.

Keywords: Smoothed Online Learning, Oracle Efficiency, Bracketing Numbers

1. Introduction

Online learning has become the most popular regime for studying sequential decision making with dependent and potentially adversarial data. In this sequential setting, a learner makes predictions one at a time, with Nature providing the data sequentially and performance being measured via *regret* with respect to a best-in-hindsight comparator drawn from a fixed (and known) function class (Cesa-Bianchi and Lugosi, 2006). While this paradigm is attractive due to its great generality and minimal set of assumptions (Cesa-Bianchi and Lugosi, 2006), the worst-case nature can create severe statistical and computational challenges (Rakhlin et al., 2015; Littlestone, 1988; Hazan and Koren, 2016). In order to mitigate these difficulties, Rakhlin et al. (2011) proposed the *smoothed* setting, wherein the data are sampled from distributions whose likelihood ratios are bounded above by $1/\sigma$ with respect to a fixed dominating measure, which ensures that the data are not worst-case with high probability if $\sigma > 0$ is not too small. The smoothed setting interpolates between the classical regime, with $\sigma = 1$ and the data are sampled independently, and the worst-case regime as $\sigma \downarrow 0$.

Recent works have demonstrated strong computational-statistical tradeoffs in smoothed online learning: while there exist algorithms that enjoy regret *logarithmic* in $1/\sigma$, oracle-efficient algorithms necessarily suffer regret scaling *polynomially* in $1/\sigma$ (Haghtalab et al., 2022a,b; Block et al., 2022), where the learner is assumed access to an Empirical Risk Minimization (ERM) oracle that is able to efficiently optimize functionals on the parameter space. This gap is significant, because in many applications of interest, the natural scaling of σ is *exponential* in ambient problem dimension

(Block and Simchowitz, 2022). This is best illustrated in the two canonical examples that suggest smoothness is a reasonable assumption in practice: if the domain is finite, then the smoothness parameter σ is the inverse of the size of the domain, while if the domain is a subset of d-dimensional Euclidean space and Nature adds uniform or Gaussian noise to a worst-case datum, then σ is exponentially small in the dimension d. In either case, a polynomial dependence of regret on σ is unacceptable.

A natural question thus remains: under which types of smoothing is it possible to design oracle-efficient algorithms with regret that scales *polynomially* in problem dimension? A partial answer was provided by Block and Simchowitz (2022), who demonstrate an efficient algorithm based on the John Ellipsoid which attains $\log(T/\sigma) \cdot \mathrm{poly}(\text{dimension})$-regret for *noiseless* linear classification, and for a suitable generalization to classification with polynomial features. They also demonstrate that, under a different smoothness condition - σ_{dir}-directional smoothness - the perceptron algorithm automatically provides regret sublinear-in-T and polynomial in $1/\sigma_{\mathrm{dir}}$. Crucially, σ_{dir} is *dimension-free* for many distributions of interest, circumventing the curse-of-dimension encountered in previous $\mathrm{poly}(1/\sigma)$-regret bounds (Block et al., 2022; Haghtalab et al., 2022b).

In this work we take oracle-efficiency as a necessary precondition and expand the set of problems that efficient smoothed online learning can address. A central example to keep in mind is that of piecewise affine (PWA) functions, where a PWA function is defined by a finite set of regions in Euclidean space, within each of which the function is affine. Such classes naturally arise in segmented regression applications common in statistics and econometrics (Feder, 1975; Bai and Perron, 1998; Yamamoto and Perron, 2013), as well as in popular models for control systems (Borrelli, 2003; Henzinger and Sastry, 1998). Unfortunately, because of the discontinuities that arise when crossing regions, PWA regressors are *not* learnable in the adversarial setting even with unbounded computation time, due to the fact that they subsume (and vastly generalize) the class of linear thresholds, whose lack of online learnability is well-known (Littlestone, 1988); however, a smoothness assumption is natural in this setting, due to the injection of noise empiricists already incorporate (Posa et al., 2014; Suh et al., 2022a). Unfortunately, the nature of the injected noise is such that the smoothness parameter σ will be exponential in the dimension of the context space, as above, and thus previous guarantees do not suffice for applications. We are thus left with the question of designing practical algorithms that are provably (oracle-)efficient in the smoothed online learning setting.

Below, we will propose a measure of complexity based on classical bracketing numbers (Blum, 1955; Giné and Nickl, 2021) that, if bounded, leads to a practical algorithm that experiences provably small regret. In particular, we will consider instantiations of the well-known Follow-the-Perturbed-Leader (FTPL) algorithm (Kalai and Vempala, 2005) applied to a function class parameterized by some set Θ: at each time $1 \le t \le T$, we sample a random path $\omega_t(\theta)$ on Θ and select $\theta_t \in \arg\min_\theta L_{t-1}(\theta) + \omega_t(\theta)$, with $L_{t-1}(\theta)$ denoting the cumulative loss up to time $t-1$. Standard analyses of FTPL (Agarwal et al., 2019b; Suggala and Netrapalli, 2020; Haghtalab et al., 2022a; Block et al., 2022) require that the loss functions be Lipschitz in the parameter θ for any datum, which clearly does not hold for the central example of PWA functions. We show, however, that smoothness guarantees that many loss functions are Lipschitz *in expectation*, up to an additive constant depending on the complexity of the class as measured by our proposed generalization of bracketing numbers. Using this fact, we provide a template for proving regret guarantees for a lazy instantiation of FTPL.

While the theory described above may be of technical interest in its own right, we instantiate our results in several examples. We replace the standard notion of smoothness with the related concept of directional smoothness introduced above (Block and Simchowitz, 2022). We adapt results from Agarwal et al. (2019b); Suggala and Netrapalli (2020) on FTPL with an exponentially distributed perturbation and exhibit a practical and provably low-regret algorithm for piecewise continuous loss functions with generalized affine boundaries. We then generalize this result to loss functions with polynomial boundaries, assuming a more constrained adversary, and finally instantiate our results in a setting motivated by robotic planning. In more detail:

- In Section 3, we introduce a new measure of the size of a class, the generalized bracketing number, which combines assumptions on the adversary with the complexity of the space and thus can be small in many situations of interest. We use generalized bracketing numbers to prove Proposition 7, which says that if an adversary is suitably constrained and the generalized bracketing number with respect to a particular pseudo-metric is controlled, then a lazy version of FTPL experiences low regret. Along the way, we show in Proposition 5 that control of the generalized bracketing number leads to a concentration inequality that is uniform over both parameters and adversaries.

- In Theorem 9, we apply the general theory developed in Section 3 to the special case of finite dimensional Θ. In particular, by adapting arguments of Agarwal et al. (2019b); Suggala and Netrapalli (2020), we show that if the generalized bracketing numbers of Θ are controlled, then Algorithm 2 can achieve average regret at most ε with the optimal $\widetilde{\mathcal{O}}\left(\varepsilon^{-2}\right)$ number of calls to the ERM oracle.

- In Theorem 10 and Corollary 11, we consider an even more concrete setting, where the loss function is piecewise continuous with affine boundaries. In particular, we show that if the adversary is σ_{dir}-directionally smooth, then Algorithm 2 attains average regret ε with only $\widetilde{\mathcal{O}}\left(\sigma_{\mathrm{dir}}^{-1}\varepsilon^{-2}\right)$ calls to the ERM oracle, removing the exponential dependence on the dimension that would come from applying Block et al. (2022) and attaining optimal dependence on ε.

- In Theorem 13, we generalize the results of Corollary 11 and show that if the adversary is further constrained to be polynomially smooth (see Definition 12) and the loss function is piecewise continuous with boundaries defined by polynomials of degree at most r, then Algorithm 2 can achieve average regret ε with at most $\widetilde{\mathcal{O}}\left(\varepsilon^{-2r}\right)$ calls to the ERM oracle.

- In Section 5, we consider a setting of piecewise Lipschitz "hybrid" dynamical systems (Henzinger and Sastry, 1998), where the boundaries within regions are either linear or polynomial. These can model a number of dynamical systems popular in robotics, notably piecewise affine systems (Borrelli, 2003; Marcucci and Tedrake, 2019) and piecwise-polynomial systems (Posa et al., 2015). We demonstrate in Theorem 14 that, with smoothning in the inputs and dynamics, our proposed FTPL algorithm attains low-regret in an online planning setting. To our knowledge, this is the first low-regret algorithm for planning in hybrid systems that exhibit discontinuities.

We begin the paper by formally setting up the problem and introducing a number of prerequisite notions, before continuing to state and discuss our results. An extended discussion of related work is deferred to Appendix A for the sake of space.

2. Formal Setting and Notation

Formally, we consider the problem of online learning with constrained data. Given some decision space Θ and context space $\mathcal{Z}$, as well as a loss function $\ell : \Theta \times \mathcal{Z} \to [0, 1]$, online learning proceeds in rounds $1 \le t \le T$. At each time t, Nature selects some $z_t \in \mathcal{Z}$ and the learner selects some $\theta_t \in \Theta$ and suffers loss $\ell(\theta_t, z_t)$ with the goal of minimizing regret with respect to the best $\theta \in \Theta$ in hindsight, $\mathbb{E}\left[\mathrm{Reg}_T\right] = \mathbb{E}\left[\sum_{t=1}^T \ell(\theta_t, z_t) - \inf_{\theta\in\Theta} \sum_{t=1}^T \ell(\theta, z_t)\right]$. For the purposes of measuring oracle complexity, we will be particularly interested in the normalized regret $T^{-1}\,\mathrm{Reg}_T$. Frequently in applications, we will consider the special case of online supervised learning where $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ and $z = (x, y)$ consists of a context x and label y; in this case, we distinguish between *proper* learning, where the learner chooses θ_t before seeing x_t, and *improper* learning, where the learner is able to choose θ_t depending on the revealed x_t.

Due to the statistical and computational challenges of fully adversarial online learning (Rakhlin et al., 2015; Hazan and Koren, 2016), we will constrain the adversary to choose $z_t \sim p_t$, where $p_t \in \mathcal{M} \subset \Delta(\mathcal{Z})$ is a distribution on $\mathcal{Z}$ possibly depending on the history up to time t and $\mathcal{M}$ is some restricted class of distributions. In this work, we will mostly focus on the setting where $\mathcal{M}$ consists of smooth distributions in some sense:

Definition 1 *Given a space $\mathcal{X}$, a measure $\mu \in \Delta(\mathcal{X})$, and some $\sigma < 0$, we say that a measure p_t is σ-smooth with respect to μ if the likelihood ratio with respect to μ is uniformly bounded by σ^{-1}, i.e., $\left\|\frac{dp_t}{d\mu}\right\|_\infty \le \frac{1}{\sigma}$. If $\mathcal{Z} \subset \mathbb{R}^d$ for some d, we say that p_t is σ_{dir}-directionally smooth if, for any unit vector $\mathbf{w} \in \mathcal{S}^{d-1}$, the distribution of $\langle \mathbf{w}, \mathbf{x}\rangle$ is σ_{dir}-smooth with respect to the Lebesgue measure on the real line, where $\mathbf{x} \sim p_t$.*

As discussed further in the related work section, smoothness has recently become a popular assumption for smoothed online learning. Directional smoothness, introduced in Block and Simchowitz (2022) and used in Block et al. (2023), has provided a natural way to mitigate the dimensional dependence of standard smoothness in some commonly used systems.

Our algorithms will employ the computational primitive of an Empirical Risk Minimization (ERM) oracle:

Definition 2 *Given a space Θ, and functionals $\ell_i : \Theta \to \mathbb{R}$ for $1 \le i \le m$, define an Empirical Risk Minimization (ERM) oracle as any oracle that optimizes over Θ, i.e., $\widetilde{\theta} = \mathsf{ERMOracle}\left(\sum_{i=1}^m \ell_i(\theta)\right)$ if $\widetilde{\theta} \in \arg\min_{\theta\in\Theta} \sum_{i=1}^m \ell_i(\theta)$.*

We emphasize that some of the loss functions ℓ_i may be data independent; in this way, the ERM oracle is able to optimize the perturbed empirical losses required of it below. Definition 2 is a common assumption in the study of computationally efficient online learning (Hazan and Koren, 2016; Block et al., 2022; Haghtalab et al., 2022a), with many heuristics for popular function classes available for practical application (LeCun et al., 2015; Garulli et al., 2012). In the sequel, we will always suppose that ther learner has access to an ERM Oracle and measure the computational complexity of the algorithm by the number of calls to $\mathsf{ERMOracle}$. In particular, we are interested in the oracle complexity of achieving average regret ε, i.e., the number of oracle calls that suffice to ensure that $T^{-1} \cdot \mathbb{E}\left[\mathrm{Reg}_T\right] \le \varepsilon$. While in the main body we assume that $\mathsf{ERMOracle}$ is exact for the sake of clean presentation, in the appendix we provide statements and proofs requiring only an approximate oracle, with a possibly perturbation-dependent error contributing additively to our final regret guarantees.

In the following section, we will introduce a new notion of complexity, the generalized bracketing number of a space Θ. Here, we will recall the classical notion of bracketing entropy, both for the sake of comparison and for future reference with respect to one of our results:

Definition 3 (From Section 3.5.2 in Giné and Nickl (2021)) *For a function class $\mathcal{F} : \mathcal{Z} \to \mathbb{R}$ and a measure $\mu \in \Delta(\mathcal{Z})$, we say that a partition $\mathcal{N} = \{\mathcal{B}_i\}$ of $\mathcal{F}$ is an ε-bracket with respect to μ if for all $\mathcal{B}_i$, it holds that $\mathbb{E}_\nu \left[\sup_{f,g\in\mathcal{B}_i} |f(z) - g(z)|\right] \leq \varepsilon$. The bracketing number, $\mathcal{N}_{[]}(\mathcal{F}, \mu, \varepsilon)$ is the minimal size of such a partition.*

Control of the bracketing numbers of a function class classically lead to uniform laws of large numbers and uniform central limit theorems, with many common function classes having well-behaved such numbers; for more detail, see (Giné and Nickl, 2021).

Notation In the sequel, we will reserve z for contexts and θ for parameters. We will always denote the horizon by T, loss functions by ℓ, and will make vectors bold. We will use $\mathcal{O}(\cdot)$ notation to suppress universal constants and $\widetilde{\mathcal{O}}(\cdot)$ to suppress polylogarithmic factors. We will let $\|\cdot\|_1$ denote the ℓ_1 norm in Euclidean space and the unadorned $\|\cdot\|$ denote the Euclidean norm.

3. Follow the Perturbed Leader and Generalized Brackets

In this section, we propose our algorithm and define the complexity parameters that ensure we experience low expected regret. In the following section, we will provide examples. We will consider an instantiation of the Follow-the-Perturbed-Leader (FTPL) class of algorithms (Kalai and Vempala, 2005), where, at each time $1 \leq t \leq T$, we construct a sample path $\omega_t(\theta)$ drawn independently and identically across t from some stochastic process on Θ and select

$$\theta_t = \arg\min_{\theta\in\Theta} L_{t-1}(\theta) + \omega_t(\theta), \tag{3.1}$$

where $L_{t-1}(\theta) = \sum_{s=1}^{t-1} \ell(\theta, z_s)$. The classical analysis of FTPL uses the so-called 'Be-The-Leader' lemma (Kalai and Vempala, 2005, Lemma 3.1) to decompose regret into the size of the perturbation and the stability of the predictions, i.e., if the learner plays θ_t from (3.1), then regret is bounded as follows:

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq 2 \cdot \mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] + \sum_{t=1}^{T} \mathbb{E}\left[\ell(\theta_t, z_t) - \ell(\theta_{t+1}, z_t)\right]. \tag{3.2}$$

Typically, the challenge in analysing the regret incurred by FTPL is in bounding the second term in (3.2), the stability term. A common assumption involved in this analysis is that the loss ℓ is Lipschitz in θ (Agarwal et al., 2019b; Suggala and Netrapalli, 2020; Block et al., 2022); unfortunately, for many classes of interest, this assumption does not hold.

To motivate our approach, consider the simple setting of learning linear thresholds, where $\theta \in [0, 1]$ and $\ell(\theta, z) = \mathbb{I}[y \neq \mathrm{sign}(x - \theta)]$ for $z = (x, y) \in \mathcal{Z} = [0, 1] \times \{\pm 1\}$. In this case, it is clear that $\theta \mapsto \ell(\theta, z)$ is not Lipschitz (or even continuous) and so the results of Agarwal et al. (2019b); Suggala and Netrapalli (2020) do not apply; however, a simple computation tells us that if the adversary is σ-smooth with respect to the Lebesgue measure, then $\theta \mapsto \mathbb{E}_z[\ell(\theta, z)]$ *is* Lipschitz. Naïvely, we might then hope that the stability term $\mathbb{E}[\ell(\theta_t, z_t) - \ell(\theta_{t+1}, z_t)]$ can be controlled by $|\theta_t - \theta_{t+1}|$ and a similar argument as in Agarwal et al. (2019b); Suggala and Netrapalli (2020) could

be applied. This idea does not work because, while it is true that for any fixed $\theta \in \Theta$, smoothness of z_t conditioned on the history implies that $\mathbb{E}\left[\ell(\theta_t, z_t) - \ell(\theta, z_t)\right] \lesssim |\theta_t - \theta|$, in fact θ_{t+1} depends on z_t and so it is *not* true that the distribution of z_t conditioned on θ_{t+1} is necessarily smooth. We will not wholly discard the approach, however; instead, we will show that if the class of functions $\theta \mapsto \ell(\theta, z)$ is small with respect to a particular notion of complexity, then a similar argument holds. To make this precise, consider the following definition:

Definition 4 *Let $\mathcal{M}$ be a class of distributions on some space $\mathcal{Z}$ and suppose that $\rho : \Theta \times \Theta \times \mathcal{Z} \to \mathbb{R}$ is a pseudo-metric on the space Θ, parameterized by elements of $\mathcal{Z}$. We say that a set $\{(\theta_i, \mathcal{B}_i)\} \subset \Theta \times 2^{\Theta}$ is a generalized ε-bracket if $\Theta \subset \bigcup_i \mathcal{B}_i$ and for all i, it holds that*

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_{z \sim \nu}\left[\sup_{\theta \in \mathcal{B}_i} \rho(\theta, \theta_i, z)\right] \leq \varepsilon.$$

We denote by $\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon)$ the minimal size of a generalized ε-bracket.

Note the similarity of Definition 4 with the classical notion from Definition 3: generalized brackets require that the expected diameter of a given partition $\mathcal{B}_i$ is small *uniformly over measures* in some class $\mathcal{M}$; in fact, if $\mathcal{M}$ is a singleton and ρ is the standard metric on $\mathbb{R}$, we recover the classical notion. The utility of generalized ε-brackets over other notions of complexity, like standard covering numbers is as follows:

Proposition 5 *Let $\mathcal{M}$ and ρ be as in Definition 4 and suppose that $z_1, \dots, z_n$ are generated such that the law p_i of z_i conditioned on σ-algebra $\mathcal{F}_i$ generated by the z_j up to time i satisfies $p_i \in \mathcal{M}$ for all $1 \leq i \leq n$. Suppose further that for all $z \in \mathcal{Z}$, it holds that $\sup_{\theta,\theta' \in \Theta} \rho(\theta, \theta', z) \leq D$. Then for any $\varepsilon, \delta > 0$, with probability at least $1 - \delta$, it holds simultaneously for all $\theta, \theta' \in \Theta$ that:*

$$\left|\sum_{i=1}^{n} \rho(\theta, \theta', z_i)\right| \leq 4n \cdot \sup_{\nu \in \mathcal{M}} \mathbb{E}_{\nu}\left[\rho(\theta, \theta', z)\right] + 8\varepsilon n + 6D^2 \log\left(\frac{2\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon)}{\delta}\right). \tag{3.3}$$

The proof of Proposition 5 can be found in Appendix B and proceeds by applying Freedman's inequality and controlling the supremum of a sum by the sum of suprema. It is somewhat surprising that, despite this seemingly very loose bound, we are able to achieve below the expected $\widetilde{\mathcal{O}}\left(\varepsilon^{-2}\right)$ oracle complexity guarantees in a wide variety of settings.

Critically, because (3.3) holds uniformly over $\theta' \in \Theta$, we may apply Proposition 5 to $\theta' = \theta_{t+1}$ and escape the challenge presented by z_t not being smooth when conditioned on θ_{t+1}. There are two remaining problems before we can present our algorithm. First, due to the additive statistical error in (3.3), if n is too small, then Proposition 5 is vacuous. To mitigate this problem, we will run FTPL in epochs. For some fixed $n \in \mathbb{N}$, and for all $\tau \geq 1$, let $\widetilde{L}_\tau(\theta) = L_{\tau n}(\theta)$, and define $\mathcal{I}_\tau = \{i | (\tau - 1)n + 1 \leq i \leq \tau n\}$ as well as $\widetilde{\ell}_\tau(\theta) = \sum_{t \in \mathcal{I}_\tau} \ell(\theta, z_t)$. We will run a lazy version of FTPL, where we update $\theta_t = \widetilde{\theta}_\tau$ at the beginning of each $\mathcal{I}_\tau$ and let $\theta_t = \widetilde{\theta}_\tau$ until the next change of epoch. The laziness allows the first term in (3.3) to dominate when we apply Proposition 5. The full algorithm is summarized in Algorithm 1.

The second challenge is to relate the stability terms in (3.2) to the pseudo-metric ρ evaluated on successive $\widetilde{\theta}_\tau$. Thus, we will require that the losses satisfy the following structural condition:

Algorithm 1 Lazy FTPL

Initialize ERM Oracle ERMOracle, epoch length n, perturbation distribution Ω
for $\tau = 1, 2, \ldots, T/n$ **do**
  Sample $\omega_\tau : \Theta \to \mathbb{R}$ from Ω `(% Sample Perturbation)`
  $\widetilde{\theta}_\tau \leftarrow \mathsf{ERMOracle}\left(\widetilde{L}_\tau(\theta) + \omega_\tau(\theta)\right)$ `(% Call ERMOracle on perturbed losses)`
  for $t = (\tau - 1)n + 1, \ldots, \tau n$ **do**
    Observe z_t, **Predict** $\widetilde{\theta}_\tau$, **Receive** $\ell(\widetilde{\theta}_\tau, z_t)$

Definition 6 *Suppose that that Θ is a subset of some normed space equipped with norm $\|\cdot\|$. We say that the pseudo-metric $\rho : \Theta \times \Theta \times \mathcal{Z} \to \mathbb{R}$ satisfies the pseudo-isometry property with parameters (α, β) with respect to the class of distributions $\mathcal{M}$ and the norm $\|\cdot\|$ if for all $\theta, \theta' \in \Theta$, it holds that*

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_{z \sim \nu}\left[\rho(\theta, \theta', z)\right] \le \alpha \cdot \left\|\theta - \theta'\right\|^\beta.$$

We are now prepared to state our first result bounding the regret of an instance of Algorithm 1:

Proposition 7 *Suppose that we are in the constrained online learning setting, where the adversary is constrained to sample z_t from some distribution in the class $\mathcal{M}$. Suppose further that there is a pseudo-metric ρ on Θ parameterized by $\mathcal{Z}$ satisfying the pseudo-isometry property of Definition 6, and for all $\theta, \theta' \in \Theta$ it holds that $\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\ell(\theta, z) - \ell(\theta', z)\right] \le \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', z)\right]$. If the learner plays Algorithm 1 and $\sup_{\theta, \theta' \in \Theta} \rho(\theta, \theta', z) \le D$, then for any $\varepsilon > 0$, the expected regret is upper bounded by:*

$$\mathcal{O}\left(\mathbb{E}\left[\sup_{\theta \in \Theta} \omega_1(\theta)\right] + \varepsilon T + \frac{TD^2}{n} \cdot \log\left(T \cdot \mathcal{N}_{\mathcal{M},[]}\left(\Theta, \rho, \varepsilon\right)\right) + 2n\alpha \cdot \sum_{\tau=1}^{T/n} \mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|^\beta\right]\right).$$

While the upper bound above may look complicated, the intuition is relatively simple. The first term appears already in (3.2) and penalizes the size of the perturbation. The other terms provide an upper bound on the second part (the stability) of (3.2). Indeed, the last term is a measure of the stability of the predictions in the norm $\|\cdot\|$ and generally should decrease when the first term increases. The middle two terms, then, are the price we pay to convert the stability as measured in (3.2) to that measured by $\|\cdot\|$. We provide a complete proof in Appendix C by first proving a variant of the Be-the-Leader lemma from Kalai and Vempala (2005) that allows for lazy updates, before applying Proposition 5 along with the pseudo-isometry property to control the stability term of the lazy updates with respect to the evaluated loss functions by the stability of the learner's predictions with respect to the relevant norm. Putting everything together concludes the proof. We remark that, as presented, it might appear that there is no disadvantage to setting n as large as possible; indeed the n dependence in the final sum appears to cancel out and increasing n decreases the third term. Unsurprisingly, this is not the case as increasing n reduces the stability of the learner's predictions and thus implicitly increases the final term, as is clear in the applications of this result.

Proposition 7 provides a template for proving regret bounds for different instantiations of Algorithm 1. In particular, for a given loss function $\ell(\cdot, \cdot)$, it suffices to find a pseudo-metric ρ, norm $\|\cdot\|$, and noise distribution Ω such that (a) ρ is a pseudo-isometry with respect to the norm $\|\cdot\|$, (b) the generalized bracketing numbers of Θ are small with respect to ρ, and (c) the perturbation causes the

Algorithm 2 Lazy FTPL (Instantiated with Exponential Noise)

Initialize ERM Oracle ERMOracle, epoch length n, perturbation size η
for $\tau = 1, 2, \ldots, T/n$ **do**
  Sample $\xi = (\xi_1, \ldots, \xi_d) \overset{iid}{\sim} \mathrm{Exp}(1)$ `(% Sample Perturbation)`
  $\widetilde{\theta}_\tau \leftarrow \mathsf{ERMOracle}\left(\widetilde{L}_\tau(\theta) - \eta \langle \xi, \theta \rangle\right)$ `(% Call ERMOracle on perturbed losses)`
  for $t = (\tau - 1)n + 1, \ldots, \tau n$ **do**
    Observe z_t, **Predict** $\widetilde{\theta}_\tau$, **Receive** $\ell(\widetilde{\theta}_\tau, z_t)$

lazy updates to be stable in the sense that $\mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|\right]$ is small. As an easy warmup for the results in the next section, we show that we can recover a weak version of the oracle-complexity upper bound of proper, smoothed online learning with the Gaussian process perturbation from Block et al. (2022), using a substantially simpler proof when the relevant function class has small bracketing entropy in the classical sense.

In this motivating example, we suppose that $\Theta = \mathcal{F}$ denotes a function class and that we are in the online supervised learning setting, i.e., $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ with $\ell(\theta, z) = \widetilde{\ell}(f(x), y)$. We further suppose that the adversary is σ-smooth with respect to a known base measure μ (recall Definition 1). As in Block et al. (2022, Theorem 10), we consider a Gaussian process perturbation, where we draw $x_1, \ldots, x_m \sim \mu$ independently, $\gamma_1, \ldots, \gamma_m$ standard gaussians, and let $\omega(f) = \eta \cdot \sum_{i=1}^m \gamma_i f(x_i)$.

Corollary 8 *Suppose that we are in the smoothed online learning setting with a function class $\mathcal{F} : \mathcal{X} \to \{\pm 1\}$ and with $\widetilde{\ell}$ in the unit interval and Lipschitz with respect to the first argument for all choices of the second argument. If the learner plays Algorithm 1 with the Gaussian perturbation described above, then with the correct choice of hyperparameters, given in Appendix D, the learner can achieve average regret ε with $\widetilde{\mathcal{O}}\left(\frac{\varepsilon^{-4} L^{3/5}}{\sigma^{2/5}} \cdot \log^{3/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)\right)$ calls to the ERM oracle.*

Note that the oracle complexity guarantee is weaker than that of Block et al. (2022); we include Corollary 8, and its proof in Appendix D, merely as a simple demonstration of our techniques and how they relate to more classical notions of function class complexity. We now proceed to examples where our machinery provides novel regret bounds in fundamental settings.

4. Exponential Perturbations and Piecewise Continuous Functions

In the previous section, we observed that Proposition 7 provided a template for proving regret bounds for different instantiations of FTPL and applied this technique to recover earlier results from smoothed online learning. In this section, we provide new results for an important setting: piecewise continuous functions. Before we formally define piecewise continuous functions, we consider the more general case where the set $\Theta \subset \mathbb{R}^d$ for some dimension d. The template provided by Proposition 7 requires that we specify a perturbation distribution; whereas before we used a Gaussian process, here we adopt the approach of Agarwal et al. (2019b); Suggala and Netrapalli (2020) and use an exponential perturbation. Summarized in Algorithm 2, we keep the lazy updating from Algorithm 1 but specify $\omega(\theta) = -\eta \cdot \langle \xi, \theta \rangle$ for some scale parameter $\eta > 0$ and $\xi = (\xi_1, \ldots, \xi_d)$ for $\xi_i \sim \mathrm{Exp}(1)$ independently. With the exponential perturbation, we have the following regret bound:

Theorem 9 *Suppose that we are in the constrained online learning setting of Proposition 7 with* $\Theta \subset \mathbb{R}^d$ *such that* $\sup_{\theta,\theta' \in \Theta} \|\theta - \theta'\|_1 = D < \infty$. *Suppose further that the* $\mathcal{Z}$*-parameterized pseudo-metric* ρ *satisfies the pseudo-isometry property of Definition 6 with respect to* ℓ_1 *on* $\mathbb{R}^d$ *and that* $\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu [\ell(\theta, z) - \ell(\theta', z)] \leq \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu [\rho(\theta, \theta', z)]$. *If the learner plays Algorithm 2 and* $\eta = \Omega(n^2)$*, then the expected regret is bounded:*

$$\mathbb{E}[\mathrm{Reg}_T] \leq \mathcal{O}\left(\eta + \frac{T}{n} \cdot \log\left(\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, 1/T)\right) + T\alpha \left(\frac{\log \mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, 1/T)}{\eta}\right)^{\frac{\beta}{4-2\beta}}\right).$$

Tuning η and n, regret scales as $\widetilde{\mathcal{O}}\left(T^{\frac{4-2\beta}{4-\beta}}\right)$ with $\widetilde{\mathcal{O}}\left(T^{\frac{2}{4-\beta}}\right)$ calls to the optimization oracle and thus $\widetilde{\mathcal{O}}\left(\varepsilon^{-2/\beta}\right)$ calls to ERMOracle suffice to attain average regret ε. In particular, in the best case, when $\beta = 1$, we recover the optimal $\widetilde{\mathcal{O}}\left(\varepsilon^{-2}\right)$ oracle-complexity of attaining average regret bounded by ε that would arise if we called the oracle once per round and achieved regret $\widetilde{\mathcal{O}}\left(\sqrt{T}\right)$.

While a complete proof of Theorem 9 can be found in Appendix E, we provide a brief sketch here. Though we follow the general template of Proposition 7, we do not directly apply the result in order to get a slightly improved rate. As in the proof of the more general proposition, we appeal to the Be-the-Leader lemma to reduce the analysis to bounding the stability of the learner's predictions with respect to the losses. We then apply techniques from Agarwal et al. (2019b); Suggala and Netrapalli (2020) to show that if the stability term is small, then the learner's predictions are stable with respect to $\|\cdot\|_1$ in $\mathbb{R}^d$. Finally, we use pseudo-isometry and control of the generalized bracketing numbers along with Proposition 5 to conclude with a self-bounding argument.

4.1. Piecewise-Continuous Prediction

We now instantiate the previous result on several problems of interest. For the rest of this section, we show that piecewise continuous functions with well-behaved boundaries allow for both small bracketing numbers and the pseudo-isometry property for properly chosen ρ, assuming only directional smoothness of the adversary. Formally, we suppose that $\Theta = \Theta_{\mathrm{c}} \times \Theta_{\mathrm{d}}$ can be decomposed into continuous and discrete parts with $\Theta \subset \mathbb{R}^m$ for some dimension m. We construct a function g as follows. First, consider classes $g_k : \Theta_{\mathrm{c}} \times \mathcal{Z} \to \mathbb{R}$ for $1 \leq k \leq K$ such that for all $z \in \mathcal{Z}$, $g_k(\cdot, z)$ is Lipschitz as a function of θ_{c} with respect to the ℓ_1 norm on Θ. Now, for a fixed $\phi : \Theta_{\mathrm{d}} \times [K] \times \mathcal{Z} \to \mathbb{R}$, we define

$$k_\phi(\theta_{\mathrm{d}}, z) = \operatorname*{arg\,max}_{k \in [K]} \phi(\theta_{\mathrm{d}}, k, z), \qquad \ell(\theta, z) = g_{k_\phi(\theta_{\mathrm{d}}, z)}(\theta_{\mathrm{c}}, z). \tag{4.1}$$

While the above definition of ℓ may seem abstract, note that it is simply a way to formalize the class of functions that can be broken into K 'pieces' such that on each of which there exists a different parameterized function. For a concrete example, we show how the central PWA systems can be instantiated in this general framework below.

The formulation of (4.1) combines versatility and simplicity, but a related construction turns out to be easier to analyze: let $\overline{\phi} : \Theta_{\mathrm{d}} \times [K]^{\times 2} \times \mathcal{Z} \to \mathbb{R}$ such that $\overline{\phi}(\theta_{\mathrm{d}}, k, k', z) = -\overline{\phi}(\theta_{\mathrm{d}}, k', k, z)$ for all $\theta_{\mathrm{d}} \in \Theta_{\mathrm{d}}$, $k, k' \in [K]$, and $z \in \mathcal{Z}$. Further, let

$$\overline{k}_\phi(\theta_{\mathrm{d}}, z) = \operatorname*{arg\,max}_{k \in [K]} \sum_{k' \neq k} \mathbb{I}\left[\overline{\phi}(\theta_{\mathrm{d}}, k, k', z) \geq \phi(\theta_{\mathrm{d}}, k', k, z)\right],$$

with ties broken lexicagraphically, i.e., $\overline{k}_{\overline{\phi}}$ is the smallest index k that wins the most matches of a tournament, where victory is determined by the sign of $\overline{\phi}(\theta_{\mathrm{d}}, k, k', z)$. We then define

$$\overline{\ell}(\theta, z) = g_{\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, z)}(\theta_{\mathrm{c}}, z). \tag{4.2}$$

In this section, we will focus on the tournament formulation of (4.2) for the sake of simplicity. In Appendix F.1, we will extend our results to the case of (4.1) with an additional margin assumption. We further remark that (4.2) can be regarded as an improper relaxation of the natural function class in (4.1) and thus suffices for improper online learning[1]. Finally, we note that, while we have described a tournament-style aggregation system for the sake of simplicity, as can be seen from our proof, any aggregation of the $\binom{K}{2}$ events $\overline{\phi}(\theta_{\mathrm{d}}, k, k', z) \geq 0$ will result in a similar statement, resulting in much greater generality. This generalization allows, for example, to efficiently represent polytopic regions with K proportional to the number of faces.

4.2. Piecewise Continuous Prediction with Generalized Affine Boundaries

We begin our study with the important special case of affine decision boundaries. and note that the setting described by (4.1) encompasses the central example of PWA functions: by letting $\Theta_{\mathrm{c}} = \left(\mathbb{R}^{m\times d}\right)^{\times K}$, $\Theta_{\mathrm{d}} = (\mathbb{R}^{d+1})^{\times K}$, $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}^m$, and $\phi(\theta_{\mathrm{d}}, k, z) = \langle \mathbf{w}_k, (\mathbf{x}, 1)\rangle$, we may take

$$\ell(\theta, z) = \|\mathbf{y} - \mathbf{W}_{k^\star}\mathbf{x}\|^2, \qquad k^\star = \arg\max_{k\in[K]} \langle \mathbf{w}_k, (\mathbf{x}, 1)\rangle,$$

where we add an extra coordinate of 1 at the end to account for a possible affine constant. We show that if $\overline{\ell}$ is piecewise continuous as in (4.2) with affine boundaries, then the generalized bracketing numbers are small and pseudo-isometry holds with respect to the ℓ_1 norm as long as the adversary is σ_{dir}-directionally smooth.

Theorem 10 *Suppose that $\mathcal{Z} \subset \mathbb{R}^d$ and that Θ is a subset of Euclidean space of ℓ_1 diameter bounded by D, with $\Theta_{\mathrm{d}} \subset (\mathcal{S}^d)^{\times\binom{K}{2}}$; denote by $\mathbf{w}_{kk'}$ the coordinates of a given $\theta_{\mathrm{d}} \in \Theta_{\mathrm{d}}$. Suppose further that $\overline{\phi}(\theta_{\mathrm{d}}, k, k', \mathbf{z}) = \psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle)$ for some differentiable, odd, link function $\psi : \mathbb{R} \to \mathbb{R}$ satisfying $a \leq |\psi'(x)| \leq A$ for all x, and let $\mathcal{M}$ consists of the class of σ_{dir}-directionally smooth distributions such that $\|\mathbf{z}\|_\infty \leq B$. Let*

$$\rho(\theta, \theta', \mathbf{z}) = 2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right] + \max_{1\leq k\leq K} \left\|\theta_{\mathrm{c}}^{(k)} - \theta_{\mathrm{c}}^{\prime(k)}\right\|_1. \tag{4.3}$$

Then ρ is a pseudo-metric satisfying the pseudo-isometry property with $\alpha = \frac{2A(B\vee 1)}{a\sigma_{\mathrm{dir}}}$ and $\beta = 1$. Furthermore, for all $\varepsilon > 0$, $\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon) \leq \left(\frac{9AK^2BD}{a\sigma_{\mathrm{dir}}\varepsilon}\right)^{K^2(d+1)}$.

We prove Theorem 10 in full detail in Appendix F.1. The proofs of both statements rely on the same key step, given in Lemma 25, which demonstrates that for fixed θ_{d}^0, even though the event $\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}^0)\right]$ is not a continuous function of θ_{d}, its expectation is Lipschitz if $\mathbf{z}$ is σ_{dir}-directionally smooth. Thus, Lemma 25 is a vast generalization of the motivating argument involving one-dimensional thresholds in Section 3. This key lemma is proven by appealing to the

1. See Block et al. (2022) for a discussion on the difference between proper and improper online learning.

anti-concentration of affine functions applied to directionally smooth random variables and is the only place that the analysis of $\overline{\ell}$ is different from that of the function ℓ in (4.1). We then use this result to both imply pseudo-isometry and to show that a cover of Θ with respect to ℓ_1 gives rise to a generalized ε-bracket with respect to $\mathcal{M}$ and ρ.

Using Theorem 10, we are able to prove a concrete regret bound for Algorithm 2 on the class of piecewise continuous functions with affine boundaries:

Corollary 11 *Suppose that $\overline{\ell}$ is as in (4.2) with $\overline{\phi}$ and Θ as in Theorem 10 with $B \geq 1$ and $\overline{\ell}$ uniformly bounded in magnitude by 1. If we set $\eta = \widetilde{\Theta}\left(\left(TK^2dDBA(a\sigma_{\mathrm{dir}})^{-1}\right)^{2/3}\right)$ and $n = \sqrt{\eta}$, then Algorithm 2 experiences $\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\left(TAK^2dBD(a\sigma_{\mathrm{dir}})^{-1}\right)^{2/3}\right)$. In particular, to achieve average regret ε, it suffices to call* ERMOracle *only $\widetilde{\mathcal{O}}\left(\frac{AK^2dDB}{a\sigma_{\mathrm{dir}}\varepsilon^2}\right)$ times.*

The proof of Corollary 11 can be found in Appendix F.2 and follows almost immediately from Theorems 9 and 10. The simplest example of a link function is simply to let $\psi(x) = x$ the identity, in which case we obtain a regret bound for piecewise continuous functions with affine boundaries.

4.3. Piecewise Continuous Prediction with Polynomial Boundaries

In order to broaden the scope of applications, we now consider more general boundaries between regions. As mentioned above, the key to proving an analogue of Theorem 10 is the anti-concentration of affine functions applied to directionally smooth random variables. While anti-concentration properties of more general functions remain an active area of research, sub-classes of polynomials, such as multi-linear functions of independent variables, are known to anti-concentrate in great generality (Mossel et al., 2010) and suffice to extend our results loss functions with these decision boundaries using our techniques, we instead focus on general polynomial boundaries and further restrict $\mathcal{M}$:

Definition 12 *For a polynomial $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(x) = \sum_{\mathcal{I} \subset [n]} \alpha_{\mathcal{I}} x^{\mathcal{I}}$, let $r = \deg(f) = \max\left\{|\mathcal{I}| \, | \alpha_{\mathcal{I}} \neq 0\right\}$ denote the degree and let $\mathrm{coeff}_r(f) = \sqrt{\sum_{|\mathcal{I}|=r} \alpha_{\mathcal{I}}^2}$ be the Euclidean norm of the vector of coefficients on the top-degree terms of the polynomial f. We say that a distribution ν is $\sigma_{\mathrm{poly},r}$-polynomially smooth if for all $a \in \mathbb{R}$, and all degree r polynomials f such that $\mathrm{coeff}_r(f) = 1$, it holds that $\mathbb{P}_{x \sim \nu}\left(|f(x) - a| \leq \varepsilon\right) \leq \frac{\varepsilon^{\frac{1}{r}}}{\sigma_{\mathrm{poly},r}}$.*

Before proceeding, a few remarks are in order. First, we note that directional smoothness is *not* sufficient to ensure polynomial smoothness, as exhibited by Glazer and Mikulincer (2022, Example 3) and thus more constrained adversaries are indeed necessary to apply our methods. Second, we obseve that Definition 12 extends the notion of directional smoothness, with the latter corresponding to $\sigma_{\mathrm{poly},1}$-smoothness. Finally, we observe that several common families of distributions are easily seen to be polynomially smooth with dimension-independent $\sigma_{\mathrm{poly},r}$, such as Gaussians and, more generally, product measures of log-concave marginals (Glazer and Mikulincer, 2022, Corollary 4); we expand on this discussion in Appendix G.1. Assuming an adversary is polynomially smooth, we prove an analogue of Theorem 10, which then results in the following regret bound:

Theorem 13 *Suppose $\mathcal{Z} \subset \mathbb{R}^d$ and Θ is a subset of Euclidean space with ℓ_1 diameter bounded by D. Let Θ_{d} parameterize the set of tuples of $\binom{K}{2}$ degree r polynomials $(f_{\mathbf{w}_{kk'}})$ on $\mathbb{R}^d$ such that $\mathrm{coeff}_r(f_{\mathbf{w}_{kk'}}) = 1$ for all $k \in [K]$. Suppose that $\overline{\ell}$ is defined as in (4.2) with $\overline{\phi}(\theta_{\mathrm{d}}, k, k', \mathbf{z}) =$*

$f_{\mathbf{w}_{kk'}}(\mathbf{z})$ and $\overline{\ell}$ bounded in the unit interval. If $\mathcal{M}$ is the class of $\sigma_{\mathrm{poly},r}$-polynomially smooth distributions such that $\|\mathbf{z}\|_\infty \le B$ almost surely, then with the correct choices of η, n given in Appendix G.2, Algorithm 2 experiences $\mathbb{E}[\mathrm{Reg}_T] \le \widetilde{\mathcal{O}}\left(\left(TK^2r^2d^rDB\sigma_{\mathrm{poly},r}^{-1}\right)^{\frac{4r-2}{4r-1}}\right)$. Thus, the oracle complexity of achieving average regret at most ε is controlled by $\widetilde{\mathcal{O}}\left(\left(\frac{K^2r^2d^rDB}{\sigma_{\mathrm{poly},r}}\cdot\varepsilon^{-2r}\right)\right)$.

We prove Theorem 13 in similarly to how we prove Theorem 11, i.e., we show an analogue of Theorem 10 for polynomially smooth distributions to control the generalized bracketing numbers and pseudo-isometry constants with respect to ρ from (4.3) before applying Theorem 9. The full details are in Appendix G.2. We remark that the common thread between the proofs of Theorem 13 and Theorem 11 is that functions of random variables samples from distributions in $\mathcal{M}$ are sufficiently anti-concentrated as to smooth the non-continuous parts of the loss functions. Finally, note that we can replace $\overline{\ell}$ with ℓ from (4.1) with a similar margin assumption as in Appendix F.3.

5. Smoothed Multi-Step Planning

In previous sections, we were interested in online prediction; here we focus on the related problem of multi-step decision making. Specifically, we study the setting of multi-step planning, where the learner plays a sequence of dynamical inputs (in control parlance, an *open-loop plan*) to minimize a cumulative control loss over a finite planning horizon. We focus on "hybrid dynamics" (Borrelli, 2003; Henzinger and Sastry, 1998), where each state space is partitioned into regions (called modes) within which the dynamics are Lipschitz. We consider the case of affine decision boundaries between modes here and defer discussion of polynomial boundaries to Appendix H. We remark that this problem is challenging due to the introduction of possible discontinuities across modes, again limiting the applicability of previous techniques. This class is rich enough to model piecewise-affine dynamics frequently encountered in robotic-planning (Hogan and Rodriguez, 2016; Anitescu and Potra, 1997; Aydinoglu et al., 2021); in the appendix, we generalize further to polynomial decision boundaries (Posa et al., 2015). See also the related work in Appendix A.

Formally, we fix a planning horizon $H \in \mathbb{N}$ and consider a family of dynamical systems with states $\mathbf{x}_h \in \mathcal{X} \subset \mathbb{R}^m$ and inputs $\mathbf{u}_h \in \mathcal{U} \subset \mathbb{R}^d$. Our decision variables are *plans* $\theta = \bar{\mathbf{u}}_{1:H} \in \mathcal{K} \subset \mathcal{U}^{\times H}$ and our context are tuples $z_t = (\mathbf{x}_{t,1}, \boldsymbol{\eta}_{t,1:H}, \boldsymbol{\xi}_{t,1:H}, g_{t;1:H,1:K}, \ell_t^v, \mathbf{W}_{t,1:H})$ consisting of an initial state $\mathbf{x}_{t,1} \in \mathcal{X}$, noises $\boldsymbol{\eta}_{t,h} \in \mathcal{X}$ and $\boldsymbol{\xi}_{t,h} \in \mathcal{U}$, continuous functions $g_{t,h,k}$ defining the dynamics for mode-k at step h, time-dependent continuous losses ℓ_t^v, and matrices $\mathbf{W}_{t,h} \in \mathbb{R}^{K(m+d+1)}$ determining the boundaries between modes, where $\mathbf{W}_{t,h}$ has rows $\mathbf{w}_{t,h,k} \in \mathcal{S}^{m+d}$. We use $\mathbf{v} \in \mathcal{V} = \mathcal{X} \times \mathcal{U}$ to denote concatenations of state and input. We suppose piecewise-continuous dynamics, where

$$\begin{aligned} \mathbf{x}_{t,h+1}(\theta) &= g_{t,h,k_{t,h}(\mathbf{v}_{t,h}(\theta))}(\mathbf{v}_{t,h}(\theta)) + \boldsymbol{\eta}_{t,h}, \quad \text{and} \\ \mathbf{u}_{t,h}(\theta) &= \bar{\mathbf{u}}_{t,h} + \boldsymbol{\xi}_{t,h}, \quad \mathbf{v}_{t,h}(\theta) = (\mathbf{x}_{t,h}(\theta), \quad \mathbf{u}_{t,h}(\theta)), \\ k_{t,h}(\mathbf{v}) &= \arg\max_{k\in[K]} \phi_{t,h}(k,\mathbf{v}), \quad \text{and} \quad \phi_{t,h}(k,\mathbf{v}) = \langle \mathbf{w}_{t,h,k}, (\mathbf{v},1)\rangle . \end{aligned} \tag{5.1}$$

In words, for each time t, there are length H trajectories that evolve according to piecewise continuous dynamics, where each piece (mode) is determined by affine functions of both the previous state and an input. We aim to minimize regret with the loss $\ell(\theta, z_t) := \ell_t^v(\mathbf{v}_{t,1:H}(\theta))$, where $\ell_t^v : \mathcal{V}^H \to \mathbb{R}$ are 1- Lipschitz functions of both the state and input. We assume that, *for fixed*

mode sequences $k_{1:h} \in [K]^h$, the $h \in [H]$-fold compositions of the Lipschitz dynamic maps $g_{t,h,k_h} \circ g_{t,h-1,k_{h-1}} \circ \cdots \circ g_{t,1,k_1}$ are L-Lipschitz as functions of $\theta \in \mathcal{K}$ in an $\ell_1 \to \ell_1$ sense (see the appendix for a precise statement). Though L may be exponential in H in the worst-case, common stability conditions ensure that L is more reasonably bounded; for further elaboration, see Remark 38. Finally, in order to incorporate smoothness, let $\mathcal{F}_t$ denote the filtration generated by $(z_{1:t-1}, \ell_t^v, g_{t,1:H,1:K}, \mathbf{W}_{t,1:H})$, and for $h \geq 0$ let $\mathcal{F}_{t,h}$ denote the filtration generated by $\mathcal{F}_t$ and $\boldsymbol{\xi}_{t,1:h}, \boldsymbol{\eta}_{t,1:h}, \mathbf{x}_{t,1}$; we suppose that the tuple $(\boldsymbol{\xi}_{t,h}, \boldsymbol{\eta}_{t,h})$ of dynamics and input noise, conditioned on $\mathcal{F}_{t,h}$, is σ_{dir}-directionally smooth.

While the restriction to open-loop plans may seem limiting, we note that the flexibility in our definition of the $g_{t,h,k}$ allows us to incorporate a wide variety of state-dependent policies with minimal modification. For example, our framework includes the popular setting of linear controls, where the learner plays an affine function mapping the state to an input; by letting $g_{t,h,k}$ be multilinear in the input matrix and the state and letting the loss be quadratic, both of which remain Lipschitz due to our boundedness assumptions, we naturally recover a piecewise generalization of the well-known Linear Quadratic Regulator (LQR). Our main result is the following.

Theorem 14 *Suppose that we are in the situation described by (5.1), with* $(\boldsymbol{\eta}_{t,h}, \boldsymbol{\xi}_{t,h})|\mathcal{F}_{t,h-1}$ σ_{dir}*-directionally smooth,* $\sup_{\mathbf{v}\in\mathcal{V}} \|\mathbf{v}\|_1 \leq D$*, the* ℓ_t^v *are Lipschitz and bounded, and the* $g_{t,h,k}$ *satisfying technical continuity assumptions found in Theorem 37. If there is some margin parameter* $\gamma > 0$ *such that for all* $t \in [T]$ *and* $h \in [H]$ *it holds that* $\min_{k\neq k'\in[K]} \left\|\mathbf{w}_{t,h,k} - \mathbf{w}_{t,h,k'}\right\| \geq \gamma$ *and the planner plays* $\bar{\mathbf{u}}_{t,h}$ *according to Algorithm 2, then the oracle complexity of achieving average regret* ε *is at most* $\widetilde{\mathcal{O}}\left((dH^5K^4(DL/(\gamma\sigma_{\text{dir}}))^2)^{\frac{1}{3}}\varepsilon^{-2}\right)$.

The proof, elaboration of assumptions, and the extension to polynomial decision boundaries are given in Appendix H. The proof follows the template of the previous section; to handle the multi-step setup, we argue that smooth dynamical noise suffices to ensure that, when $\theta, \theta' \in \mathcal{K}$ are sufficiently close, smoothness ensures that the sequence of modes $k_{t,h}(\mathbf{v}_{t,h}(\theta)), k_{t,h}(\mathbf{v}_{t,h}(\theta))$ coincide for all $h \in [H]$ with high probability; this requires a telescoping argument similar in spirit to the performance-difference lemma in reinforcement learning (Kakade, 2003).

Acknowledgments

AB acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. We also acknowledge support from ONR under grant N00014-20-1-2336, DOE under grant DE-SC0022199, and NSF through award DMS-2031883. MS acknowledges support from Amazon.com Services LLC grant; PO 2D-06310236. We also acknowledge Russ Tedrake, Terry H.J. Suh, and Tao Pang for their helpful comments.

References

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26. JMLR Workshop and Conference Proceedings, 2011.

Jacob Abernethy, Chansoo Lee, Abhinav Sinha, and Ambuj Tewari. Online linear optimization via smoothing. In *Conference on Learning Theory*, pages 807–823. PMLR, 2014.

Jacob D Abernethy, Chansoo Lee, and Ambuj Tewari. Fighting bandits with a new kind of smoothness. *Advances in Neural Information Processing Systems*, 28, 2015.

Alekh Agarwal, Daniel Hsu, Satyen Kale, John Langford, Lihong Li, and Robert Schapire. Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning*, pages 1638–1646. PMLR, 2014.

Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119. PMLR, 2019a.

Naman Agarwal, Alon Gonen, and Elad Hazan. Learning in non-convex games with an optimization oracle. In *Conference on Learning Theory*, pages 18–29. PMLR, 2019b.

David Angeli. A lyapunov approach to incremental stability. In *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187)*, volume 3, pages 2947–2952. IEEE, 2000.

Mihai Anitescu and Florian A Potra. Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. *Nonlinear Dynamics*, 14:231–247, 1997.

Alp Aydinoglu, Philip Sieg, Victor M Preciado, and Michael Posa. Stabilization of complementarity systems via contact-aware controllers. *IEEE Transactions on Robotics*, 38(3):1735–1754, 2021.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

Adam Block and Max Simchowitz. Efficient and near-optimal smoothed online learning for generalized linear functions. *arXiv preprint arXiv:2205.13056*, 2022.

Adam Block, Yuval Dagan, Noah Golowich, and Alexander Rakhlin. Smoothed online learning is as easy as statistical learning. In *Conference on Learning Theory*, pages 1716–1786. PMLR, 2022.

Adam Block, Max Simchowitz, and Russ Tedrake. Smoothed online learning for prediction in piecewise affine systems. *arXiv preprint arXiv:2301.11187*, 2023.

JR Blum. On the convergence of empiric distribution functions. *The Annals of Mathematical Statistics*, pages 527–529, 1955.

Francesco Borrelli. *Constrained optimal control of linear and hybrid systems*, volume 290. Springer, 2003.

Anthony Carbery and James Wright. Distributional and lq norm inequalities for polynomials over convex bodies in rn. *Mathematical research letters*, 8(3):233–248, 2001.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Alon Cohen and Tamir Hazan. Following the perturbed leader for online structured learning. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1034–1042, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/cohena15.html.

Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1029–1038. PMLR, 2018.

Yan Dai, Haipeng Luo, and Liyu Chen. Follow-the-perturbed-leader for adversarial markov decision processes with bandit feedback. In *Advances in Neural Information Processing Systems*, 2022.

Sarah Dean and Benjamin Recht. Certainty equivalent perception-based control. In *Learning for Dynamics and Control*, pages 399–411. PMLR, 2021.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, 20(4): 633–679, 2020.

John DeHardt. Generalizations of the glivenko-cantelli theorem. *The Annals of Mathematical Statistics*, 42(6):2050–2055, 1971.

Manfred Deistler, K Peternell, and Wolfgang Scherrer. Consistency and relative efficiency of subspace methods. *Automatica*, 31(12):1865–1875, 1995.

Richard M Dudley. Central limit theorems for empirical measures. *The Annals of Probability*, pages 899–929, 1978.

Paul I Feder. On asymptotic distribution theory in segmented regression problems–identified case. *The Annals of Statistics*, 3(1):49–83, 1975.

Dylan Foster, Tuhin Sarkar, and Alexander Rakhlin. Learning nonlinear dynamical systems from a single trajectory. In *Learning for Dynamics and Control*, pages 851–861. PMLR, 2020.

Andrea Garulli, Simone Paoletti, and Antonio Vicino. A survey on switched and piecewise affine system identification. *IFAC Proceedings Volumes*, 45(16):344–355, 2012.

Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge university press, 2021.

Itay Glazer and Dan Mikulincer. Anti-concentration of polynomials: Dimension-free covariance bounds and decay of fourier coefficients. *Journal of Functional Analysis*, 283(9):109639, 2022.

Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis of online and differentially private learning. *Advances in Neural Information Processing Systems*, 33:9203–9215, 2020.

Nika Haghtalab, Yanjun Han, Abhishek Shetty, and Kunhe Yang. Oracle-efficient online learning for beyond worst-case adversaries. *arXiv preprint arXiv:2202.08549*, 2022a.

Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis with adaptive adversaries. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 942–953. IEEE, 2022b.

Elad Hazan and Tomer Koren. The computational power of optimization in online learning. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 128–141, 2016.

Elad Hazan and Karan Singh. Introduction to online nonstochastic control. *arXiv preprint arXiv:2211.09619*, 2022.

Elad Hazan, Karan Singh, and Cyril Zhang. Learning linear dynamical systems via spectral filtering. *Advances in Neural Information Processing Systems*, 30, 2017.

Thomas A Henzinger and Shankar Sastry. *Hybrid Systems: Computation and Control: First International Workshop, HSCC'98, Berkeley, California, USA, April 13-15, 1998: Proceedings*. Springer, 1998.

François Robert Hogan and Alberto Rodriguez. Feedback control of the pusher-slider system: A story of hybrid and underactuated contact dynamics. *arXiv preprint arXiv:1611.08268*, 2016.

Ali Jadbabaie and John Hauser. On the stability of unconstrained receding horizon control with a general terminal cost. In *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No. 01CH37228)*, volume 5, pages 4826–4831. IEEE, 2001.

Sham Kakade, Akshay Krishnamurthy, Kendall Lowrey, Motoya Ohnishi, and Wen Sun. Information theoretic regret bounds for online nonlinear control. *Advances in Neural Information Processing Systems*, 33:15312–15325, 2020.

Sham Machandranath Kakade. *On the sample complexity of reinforcement learning*. University of London, University College London (United Kingdom), 2003.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

Zifan Li and Ambuj Tewari. Beyond the hazard rate: More perturbation algorithms for adversarial multi-armed bandits. *J. Mach. Learn. Res.*, 18:183–1, 2017.

Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2:285–318, 1988.

Lennart Ljung. System identification: Theory for the users," prentice hall, new jersey, 1999.

Lennart Ljung and Bo Wahlberg. Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Advances in Applied Probability*, 24(2):412–440, 1992.

Horia Mania, Michael I Jordan, and Benjamin Recht. Active learning for nonlinear system identification with guarantees. *arXiv preprint arXiv:2006.10277*, 2020.

Tobia Marcucci and Russ Tedrake. Mixed-integer formulations for optimal control of piecewise-affine systems. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, pages 230–239, 2019.

Zakaria Mhammedi, Dylan J Foster, Max Simchowitz, Dipendra Misra, Wen Sun, Akshay Krishnamurthy, Alexander Rakhlin, and John Langford. Learning the linear quadratic regulator from nonlinear observations. *Advances in Neural Information Processing Systems*, 33:14532–14543, 2020.

Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, pages 295–341, 2010.

Richard Nickl and Benedikt M Pötscher. Bracketing metric entropy rates and empirical central limit theorems for function classes of besov-and sobolev-type. *Journal of Theoretical Probability*, 20: 177–199, 2007.

Samet Oymak and Necmiye Ozay. Non-asymptotic identification of lti systems from a single trajectory. In *2019 American control conference (ACC)*, pages 5655–5661. IEEE, 2019.

Daniel Pfrommer, Thomas TCK Zhang, Stephen Tu, and Nikolai Matni. Tasil: Taylor series imitation learning. *arXiv preprint arXiv:2205.14812*, 2022.

Michael Posa, Cecilia Cantu, and Russ Tedrake. A direct method for trajectory optimization of rigid bodies through contact. *The International Journal of Robotics Research*, 33(1):69–81, 2014.

Michael Posa, Mark Tobenkin, and Russ Tedrake. Stability analysis and control of rigid-body systems with impacts and friction. *IEEE Transactions on Automatic Control*, 61(6):1423–1437, 2015.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Stochastic, constrained, and smoothed adversaries. *Advances in neural information processing systems*, 24, 2011.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *J. Mach. Learn. Res.*, 16(1):155–186, 2015.

Yahya Sattar and Samet Oymak. Non-asymptotic and accurate learning of nonlinear dynamical systems. *Journal of Machine Learning Research*, 23(140):1–49, 2022.

Yahya Sattar, Zhe Du, Davoud Ataee Tarzanagh, Laura Balzano, Necmiye Ozay, and Samet Oymak. Identification and adaptive control of markov jump systems: Sample complexity and regret bounds. *arXiv preprint arXiv:2111.07018*, 2021.

Max Simchowitz and Dylan Foster. Naive exploration is optimal for online lqr. In *International Conference on Machine Learning*, pages 8937–8948. PMLR, 2020.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pages 439–473. PMLR, 2018.

Max Simchowitz, Karan Singh, and Elad Hazan. Improper learning for non-stochastic control. In *Conference on Learning Theory*, pages 3320–3436. PMLR, 2020.

Arun Sai Suggala and Praneeth Netrapalli. Online non-convex learning: Following the perturbed leader is optimal. In *Algorithmic Learning Theory*, pages 845–861. PMLR, 2020.

Hyung Ju Suh, Max Simchowitz, Kaiqing Zhang, and Russ Tedrake. Do differentiable simulators give better policy gradients? In *International Conference on Machine Learning*, pages 20668–20696. PMLR, 2022a.

Hyung Ju Terry Suh, Tao Pang, and Russ Tedrake. Bundled gradients through contact via randomized smoothing. *IEEE Robotics and Automation Letters*, 7(2):4000–4007, 2022b.

Anastasios Tsiamis and George J Pappas. Finite sample analysis of stochastic system identification. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3648–3654. IEEE, 2019.

Anastasios Tsiamis, Nikolai Matni, and George Pappas. Sample complexity of kalman filtering for unknown systems. In *Learning for Dynamics and Control*, pages 435–444. PMLR, 2020.

Anastasios Tsiamis, Ingvar Ziemann, Nikolai Matni, and George J Pappas. Statistical learning theory for control: A finite sample perspective. *arXiv preprint arXiv:2209.05423*, 2022.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Yohei Yamamoto and Pierre Perron. Estimating and testing multiple structural changes in linear models using band spectral regressions. *The Econometrics Journal*, 16(3):400–429, 2013.

Contents

Appendix A. Related Work

In this section, we continue our discussion of relevant related work from the introduction.

Smoothed Online Learning Smoothed online learning was originally proposed in Rakhlin et al. (2011), with more recent work including Haghtalab et al. (2020, 2022a,b); Block et al. (2022). In particular, Haghtalab et al. (2022b) characterized the statistical rates for smoothed online classification and Block et al. (2022) did the same for the more general setting of real-valued functions. The first analysis of oracle-efficient algorithms for smoothed online learning was conducted in Haghtalab et al. (2022a); Block et al. (2022), with both works providing both proper and improper algorithms. Both works also provided lower bounds, showing the exponential gap in dependence on the smoothing paramter σ of the regret incurred by inefficient and oracle-efficient algorithms.

In order to address the exponentially worse regret guarantees for oracle-efficient smoothed online learning, Block and Simchowitz (2022) examined a special case where the loss function is a linear threshold function, parameterized by elements in Θ. In the even more restricted, realizable setting, where there exists some $\theta \in \Theta$ achieving zero cumulative loss, that work was able to recover regret logarithmic in T/σ. Unfortunately, the noiseless assumption is unrealistic and the resulting algorithm is not robust to its removal; thus, Block and Simchowitz (2022) proposed a new notion, directional smoothness, that relaxed the smoothness assumption to one more specifically suited to linear structure. Building on this work, Block et al. (2023) demonstrated that oracle-efficient smoothed online learning was possible in the challenging Piecewise Affine (PWA) setting, with regret depending only polynomially on all relevant problem parameters, albeit with a somewhat impractical algorithm and a significantly worse dependence on the horizon in the regret. We note, however, that the results of Block et al. (2023) are not comparable with our results because our algorithm requires a stronger notion of ERM oracle than that of Block et al. (2023), a point on which we elaborate below.

Follow the Perturbed Leader and Oracle-Efficient Online Learning Follow the Perturbed Leader (FTPL) was first proposed and analyzed in Kalai and Vempala (2005) for the setting of linear losses. In that work, the authors introduced the Be-the-Leader lemma, decomposing regret

into a perturbation term and a stability term, which remains the most popular way to prove regret bounds for such algorithms. Since then, the algorithmic framework has seen much popularity, with applications to multi-armed bandits (Abernethy et al., 2015), Reinforcement Learning (Dai et al., 2022), and online structured learning (Cohen and Hazan, 2015), among others. Of greater relevance to this paper, are the works Agarwal et al. (2019b); Suggala and Netrapalli (2020), which demonstrate that in the adversarial online learning setting, if the loss functions are Lipschitz, then FTPL with an exponential perturbation can attain optimal regret. In our Theorem 9, we extend the approach of these two works beyond the Lipschitz case, using our new notion of complexity. Due to the memoryless property of the exponential distribution, it is one of the most popular perturbations used for analysis of FTPL instantiations (Kalai and Vempala, 2005; Suggala and Netrapalli, 2020; Agarwal et al., 2019b), although other distributions have been studied with different techniques (Abernethy et al., 2014, 2015; Li and Tewari, 2017; Block et al., 2022; Haghtalab et al., 2022a). Our analysis heavily relies on this memoryless property and thus we restrict our focus to this instantiation, leaving as an interesting question for future work whether similar results can hold with more general perturbation distributions.

Lower bounds for oracle-efficient online learning have proven substantially more difficult than upper bounds. In Hazan and Koren (2016), the authors demonstrated an exponential gap in the statistical and computational complexities of achieving average regret at most ε; similarly, the lower bounds of Block et al. (2022); Haghtalab et al. (2022a) are based on reductions to this result. On the other hand, this lower bound appears somewhat brittle, as it applies only to proper learning with a somewhat restricted notion of ERM oracle. While the oracle used in Block et al. (2023) fits into this model, the oracle we assume, as well as that used in Hazan and Koren (2016); Suggala and Netrapalli (2020), does not. For more discussion on this point, see Hazan and Koren (2016).

Prediction and Planning in Piecewise Affine Systems Our examples are motivated in part by the planning and prediction in piecewise affine systems, and more generally, systems with polynomial boundaries between Lipschitz regions. Piecewise affine dynamics are popular in the constrained MPC and hybrid systems literature, (Henzinger and Sastry, 1998; Borrelli, 2003), due in part to their ability to model contact dynamics in robotic systems (Marcucci and Tedrake, 2019; Anitescu and Potra, 1997; Suh et al., 2022b); polynomial boundaries are studied in (Posa et al., 2015). Suh et al. (2022b,b) have studied the advantages of randomized noise injection for trajectory planning through systems with discontinuities of these forms, demonstrating numerous advantages. For typical noise distributions (e.g. Gaussian), these randomized noise injections introduce the same sorts of smoothness properties leveraged in the present work.

Statistical and Online Learning for Control and Dynamical Prediction. Building on the decades-old literature for system-identification (Ljung, 1999; Ljung and Wahlberg, 1992; Deistler et al., 1995), recent work has provided finite-sample statistical guarantees for parameter recovery in linear dynamical systems for various regimes of interest (Simchowitz et al., 2018; Dean et al., 2020; Oymak and Ozay, 2019; Tsiamis and Pappas, 2019; Tsiamis et al., 2022). Further research has studied smooth nonlinear dynamics (Mania et al., 2020; Sattar and Oymak, 2022; Foster et al., 2020), and settings where only the observation model is nonlinear (Dean and Recht, 2021; Mhammedi et al., 2020). Relevant to this work, Sattar et al. (2021) study *Markov jump systems*, where the system dynamics alternate between one of a finite number of linear systems ("modes"), and switches between modes are governed by a (discrete) Markov chain. In constrast, the dynamics with piecewise affine boundaries studied in this work have modes which depend on state.

In addition to the recent advances in finite-sample system identification, a vast body of work has studied linear control tasks from the perspective of regret (Abbasi-Yadkori and Szepesvári, 2011; Agarwal et al., 2019a; Simchowitz et al., 2020; Simchowitz and Foster, 2020; Cohen et al., 2018). Kakade et al. (2020) studies nonlinear online control of *fixed* nonlinear systems with a certain linear-parametric structure; similarly to the present work, although in less generality, a crucial step is their use of Gaussian smoothing to guarantee low regret. In contrast, Section 5 allows for time-varying dynamics and does not rely on recovery of a low-dimensional parameter.

Our guarantees for prediction are similar in spirit to online prediction for linear control settings attained (Hazan et al., 2017; Tsiamis et al., 2020); though, of course, they pertain to a far broader class of dynamical systems.

Bracketing Entropy The notion of bracketing number originally dates back to Blum (1955); DeHardt (1971) and was used to prove uniform laws of large numbers. They were then used by Dudley (1978) to prove uniform central limit theorems. There exist many bounds on bracketing numbers for concrete function classes of interest, with the most notable likely being Besov and Sobolev classes (Nickl and Pötscher, 2007). Our Definition 4 generalizes this notion both by changing the absolute value to a general pseudo-metric and, more importantly, forcing the expectation to be uniform over a family of measures.

Appendix B. Proof of Proposition 5

In this section, we prove Proposition 5 by appealing to Freedman's inequality. We recall:

Lemma 15 (Freedman's Inequality, Agarwal et al. (2014)) *Let Z_t for $1 \le t \le T$ be a real-valued martingale difference sequence such that, conditional on $Z_{1:t-1}$, almost surely $|Z_t| \le R$. Then for any $0 < \eta < \frac{1}{R}$, with probability at least $1-\delta$, it holds that*

$$\sum_{t=1}^{T} Z_t \le \eta \cdot \sum_{t=1}^{T} \mathbb{E}_{t-1}\left[Z_t^2\right] + \frac{R \log\left(\frac{1}{\delta}\right)}{\eta}.$$

We are now ready to prove the key result:

Proof [Proof of Proposition 5] Let

$$Z_i(\theta, \theta') = \rho(\theta, \theta', x_i) - \mathbb{E}_i\left[\rho(\theta, \theta', x_i)\right],$$

where we use the convenient shorthand $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{i-1}]$, with $\mathcal{F}_i$ as in the statement of the proposition. We begin by observing that by assumption, $|Z_i| \le 2D$ and further, that

$$\mathbb{E}_i\left[Z_i^2\right] \le \mathbb{E}_i\left[\rho(\theta, \theta', x_i)^2\right] \le D \cdot \mathbb{E}_i\left[\rho(\theta, \theta', x_i)\right].$$

Applying Lemma 15, we see that for any fixed $0 < \eta < \frac{1}{R}$ and $\theta, \theta' \in \Theta$, with probability at least $1 - \frac{\delta}{2}$ it holds that

$$\begin{aligned}\sum_{i=1}^{n} \rho(\theta, \theta', x_i) &\le (1+\eta D) \sum_{i=1}^{n} \mathbb{E}_i\left[\rho(\theta, \theta', x_i)\right] + \frac{2D \log\left(\frac{2}{\delta}\right)}{\eta} \\ &\le (1+\eta D) n \cdot \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', x)\right] + \frac{2D \log\left(\frac{2}{\delta}\right)}{\eta}.\end{aligned}$$

Let $\mathcal{N} = \{(\theta_j, \mathcal{B}_j)\}$ denote a minimal generalized ε-bracket. By a union bound and setting $\eta = \frac{1}{3D}$, we see that with probability at least $1 - \frac{\delta}{2}$, it holds for all $\theta_j, \theta_k \in \mathcal{N}$ that

$$\sum_{i=1}^{n} \rho(\theta_j, \theta_k, x_i) \le 4n \cdot \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta_j, \theta_k, x)\right] + 2D^2 \cdot \log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right).$$

Similarly, we define

$$\widetilde{Z}_i^j = \sup_{\theta \in \mathcal{B}_j} \rho(\theta, \theta_j, x_i) - \mathbb{E}_i\left[\sup_{\theta \in \mathcal{B}_j} \rho(\theta, \theta_j, x_i)\right]$$

and note that by the definition of the generalized bracket,

$$\left|\widetilde{Z}_i^j\right| \le 2D \qquad\qquad \mathbb{E}_i\left[(\widetilde{Z}_i^j)^2\right] \le D\varepsilon.$$

Thus, again applying Lemma 15 and a union bound, it holds that with probability at least $1 - \frac{\delta}{2}$, for all $\theta_j \in \mathcal{N}$,

$$\sup_{\theta \in \mathcal{B}_j} \sum_{i=1}^{n} \rho(\theta, \theta_j, x_i) \le 4n\varepsilon + 2D^2 \log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right).$$

By the triangle inequality and a union bound, we then have that with probability at least $1 - \delta$, if for all $\theta, \theta' \in \Theta$, we let θ_j be the projection of θ to $\mathcal{N}$ and θ_k the projection of θ' to $\mathcal{N}$, it holds that

$$\begin{aligned}
\sum_{i=1}^{n} \rho(\theta, \theta', x_i) &\le \inf_{\theta_j \in \mathcal{N}}\left\{\sum_{i=1}^{n} \rho(\theta, \theta_j, x_i)\right\} + \sum_{i=1}^{n} \rho(\theta_j, \theta_k, x_i) + \inf_{\theta_k \in \mathcal{N}}\left\{\sum_{i=1}^{n} \rho(\theta_k, \theta', x_i)\right\} \\
&\le 4n \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta_j, \theta_k, x)\right] + 4n\varepsilon + 6D^2 \log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right) \\
&\le 4n \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta, \theta', x)\right] + \left(2n \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta, \theta_j, x)\right] + 2n \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta_k, \theta', x)\right]\right) \\
&\quad + 4n\varepsilon + 6D^2 \log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right) \\
&\le 4n \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta, \theta', x)\right] + 8n\varepsilon + 6D^2 \log\left(\frac{2\left|\mathcal{N}\right|^2}{\delta}\right).
\end{aligned}$$

The result follows. ■

Appendix C. Proof of Proposition 7

In this section, we prove a more general version of Proposition 7 by combining the classic Be-the-Leader Lemma from Kalai and Vempala (2005) with our Proposition 5. We begin by stating and proving a lazy version of the Be-the-Leader lemma. We follow the proof of Block et al. (2022, Lemma 31):

Lemma 16 (Be-The-Leader Lemma) *Let $n \in \mathbb{N}$ and suppose for each $1 \le \tau \le T/n$, the learner chooses some approximate minimizer of the perturbed cumulative loss. More precisely, for some real-valued function γ on the stochastic process, the learner chooses some $\widetilde{\theta}_\tau$ satisfying*

$$L_{(\tau-1)\cdot n}(\widetilde{\theta}_\tau) + \omega_{(\tau-1)\cdot n+1}(\widetilde{\theta}_\tau) \le \gamma(\omega_{(\tau-1)\cdot n+1}) + \inf_{\theta\in\Theta} L_{(\tau-1)\cdot n}(\theta) + \omega_{(\tau-1)\cdot n+1}(\theta)$$

and plays $\widetilde{\theta}_\tau$ for all $(\tau-1)\cdot n + 1 \le t \le \tau \cdot n$. Suppose that the ω_t are independent across t and identically distributed random processes on Θ satisfying $\mathbb{E}\left[\sup_{\theta\in\Theta} \omega_t(\theta)\right] \ge 0$. Then, the learner experiences the following regret:

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \frac{T}{n} \cdot \mathbb{E}\left[\gamma(\omega_1)\right] + \mathbb{E}\left[\sup_\theta \omega_1(\theta)\right] + \sum_{\tau=1}^{T/n} \mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \ell(\widetilde{\theta}_\tau, z_t) - \ell(\widetilde{\theta}_{\tau+1}, z_t)\right].$$

Proof We apply the proof of Block et al. (2022, Lemma 31) to the cumulative loss over n steps. Thus, for each $1 \le \tau \le T' = \lfloor T/n \rfloor$, recall that

$$\widetilde{\ell}_\tau(\theta) = \sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \ell(\theta, z_t).$$

We will apply induction on T' to the inequality

$$\mathbb{E}\left[\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta_{\tau+1})\right] \le \mathbb{E}\left[\sum_{\tau=1}^{T'} \ell_\tau(\theta_{T'+1}) + \omega_{T'\cdot n+1}(\theta_{T'+1})\right] + \mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] + \frac{T'}{n} \cdot \mathbb{E}\left[\gamma(\omega_1)\right]. \tag{C.1}$$

For the base case of $T' = 0$ the statement is trivial. Suppose that for some fixed $T' - 1$ that (C.1) holds. Then we see by construction that

$$\begin{aligned}\mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau(\theta_{T'}) + \omega_{T'\cdot n+1}(\theta_{T'})\right] &\le \mathbb{E}\left[\inf_{\theta\in\Theta} \sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau(\theta) + \omega_{T'\cdot n+1}(\theta)\right] + \mathbb{E}\left[\gamma(\omega_{T'\cdot n+1})\right] \\ &\le \mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}\left(\theta_{T'+1}\right) + \omega_{(T'+1)\cdot n+1}\left(\theta_{T'+1}\right)\right] + \mathbb{E}\left[\gamma(\omega_{(T'+1)\cdot n+1})\right],\end{aligned}$$

where the second inequality follows from the fact that the ω_t are independent and identically distributed as well as the construction of $\theta_{T'+1}$. Combining the induction hypothesis (C.1) with the above inequality tells us that

$$\mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau(\theta_{\tau+1})\right] \le \mathbb{E}\left[\sum_{\tau=1}^{T'-1} \widetilde{\ell}_\tau\left(\theta_{T'+1}\right) + \omega_{(T'+1)\cdot n+1}\left(\theta_{T'+1}\right)\right] + \mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] + T' \cdot \mathbb{E}\left[\gamma(\omega_1)\right].$$

Adding $\mathbb{E}\left[\widetilde{\ell}_{T'}(\theta_{T'+1})\right]$ to both sides finishes the induction proof.

To continue, we compute:

$$\mathbb{E}\left[\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta_{T'+1}) + \omega_{T'+1}(\theta_{T'+1})\right] \le \mathbb{E}\left[\inf_{\theta\in\Theta} \sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta) + \sup_{\theta'\in\Theta} \omega_1(\theta)\right] + \frac{T'}{n}\cdot \mathbb{E}\left[\gamma(\omega_1)\right]$$
$$\le \mathbb{E}\left[\inf_{\theta\in\Theta} \sum_{t=1}^{T} \ell(\theta, z_t)\right] + \mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] \frac{T'}{n}\cdot \mathbb{E}\left[\gamma(\omega_1)\right],$$

where we used the construction of $\theta_{T'+1}$ for the first inequality and the definition of $\widetilde{\ell}_\tau$ for the second inequality. To conclude, we apply (C.1) and observe:

$$\mathbb{E}\left[\mathrm{Reg}_T\right] = \mathbb{E}\left[\sum_{t=1}^{T} \ell(\theta_{\tau(t)}) - \inf_{\theta\in\Theta} \sum_{t=1}^{T} \ell(\theta, z_t)\right]$$
$$\le \mathbb{E}\left[\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta_\tau)\right] - \mathbb{E}\left[\sum_{\tau=1}^{T'} \widetilde{\ell}_\tau(\theta_{\tau+1})\right] + 2\mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] + \frac{T'}{n}\cdot \mathbb{E}\left[\gamma(\omega_1)\right],$$

where we denoted by $\tau(t) = \lfloor t/n \rfloor$. The result follows from the construction of $\widetilde{\ell}_\tau$. ■

We are now ready to prove the main result of the section.

Proposition 17 *Suppose that we are in the constrained online learning setting, where the adversary is constrained to sample z_t from some distribution in the class $\mathcal{M}$. Suppose further that there is a pseudo-metric ρ on Θ parameterized by $\mathcal{Z}$ satisfying the psuedo-isometry property of Definition 6 such that for all $\theta, \theta' \in \Theta$ it holds that $\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\ell(\theta, z) - \ell(\theta', z)\right] \le \sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', z)\right]$. If the learner plays $\widetilde{\theta}_\tau$ as in Lemma 16 and $\sup_{\theta,\theta'\in\Theta} \rho(\theta, \theta', z) \le D$, then for any $\varepsilon > 0$, the expected regret is bounded as:*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \frac{T}{n}\cdot \mathbb{E}\left[\gamma(\omega_1)\right] + \mathbb{E}\left[\sup_{\theta\in\Theta} \omega_1(\theta)\right] + 8\varepsilon T + 1$$
$$+ \frac{6TD^2}{n}\cdot \log\left(T\cdot \mathcal{N}_{\mathcal{M},[]}\left(\Theta, \rho, \varepsilon\right)\right) + 4n\alpha \cdot \sum_{\tau=1}^{T/n} \mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|^\beta\right].$$

Proof Applying Lemma 16, we have

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \frac{T}{n}\cdot \mathbb{E}\left[\gamma(\omega_1)\right] + \mathbb{E}\left[\sup_{\theta} \omega_1(\theta)\right] + \sum_{\tau=1}^{T/n} \mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \ell(\widetilde{\theta}_\tau, z_t) - \ell(\widetilde{\theta}_{\tau+1}, z_t)\right].$$

We thus only need to bound the final sum above. By the fact that the loss function is Lipschitz with respect to the pseudo-metric, we have:

$$\sum_{\tau=1}^{T/n} \mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \ell(\widetilde{\theta}_\tau, z_t) - \ell(\widetilde{\theta}_{\tau+1}, z_t)\right] \le \sum_{\tau=1}^{T/n} \mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, z_t)\right].$$

Thus for any fixed τ and for all $\varepsilon, \delta > 0$,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=(\tau-1)\cdot n+1}^{\tau\cdot n} \rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, z_t)\right] \\
&\leq \mathbb{E}\left[4n \sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, z)\right] + 8\varepsilon n + 6D^2 \log\left(\frac{2\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right) + \delta n\right] \\
&\leq 4n\alpha \cdot \mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|^\beta\right] + (8\varepsilon+\delta)n + 6D^2 \cdot \log\left(\frac{2\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right),
\end{aligned}$$

where the first inequality follows from Proposition 5 and the second follows from assuming that ρ satisfies the conditions of Definition 6. Summing over τ and setting $\delta = T^{-1}$ concludes the proof. ■

Finally, we observe that Proposition 7 is a special case of the preceding analysis:
Proof [Proof of Proposition 7] The result follows immediately by taking $\gamma = 0$ uniformly in Proposition 17. ■

Appendix D. Proof of Corollary 8

In this section, we prove Theorem 8 by first demonstrating that generalized brackets in this setting can simply be taken to be classical brackets and then by applying a stability bound from Block et al. (2022). In order to respect notational convention, we will replace Θ with a function class $\mathcal{F}$ and consider functions $f \in \mathcal{F}$ instead of parameters $\theta \in \Theta$. We will let

$$\rho(f, f', z) = L \cdot \left|f(x) - f'(x)\right|, \qquad z = (x, y) \in \mathcal{Z},$$

and show that with this ρ, classical brackets become generalized brackets after rescaling:

Lemma 18 *Let $\mathcal{M}$ denote the class of distributions that are σ-smooth with respect to some distribution μ on $\mathcal{X}$. If ρ is as above, then*

$$\mathcal{N}_{\mathcal{M},[]}(\mathcal{F}, \rho, \varepsilon) \leq \mathcal{N}_{[]}\left(\mathcal{F}, \mu, \sigma\varepsilon/L\right).$$

Moreover, ρ satisfies the pseudo-isometry for $\alpha = L \cdot \sigma^{-1}$ and $\beta = 1$ for the norm $L^1(\mu)$.

Proof Let $\mathcal{N} = \{\mathcal{B}_i\}$ denote an $\widetilde{\varepsilon}$-bracket, in the classical sense, of $\mathcal{F}$ with respect to μ, where $\widetilde{\varepsilon} = \frac{\sigma\varepsilon}{L}$, and let f_i denote an arbitrary member of $\mathcal{B}_i$. Then we see for all $\nu \in \mathcal{M}$,

$$\begin{aligned}
\mathbb{E}_\nu\left[\sup_{f\in\mathcal{B}_i} \rho(f, f_i, z)\right] &= \mathbb{E}_\nu\left[\sup_{f\in\mathcal{B}_i} |f(x) - f_i(x)|\right] \\
&= \mathbb{E}_\mu\left[\frac{d\nu}{d\mu} \sup_{f\in\mathcal{B}_i} |f(x) - f_i(x)|\right] \\
&\leq \frac{1}{\sigma} \cdot \mathbb{E}_\mu\left[\sup_{f\in\mathcal{B}_i} |f(x) - f_i(x)|\right] \\
&\leq \frac{\varepsilon}{\sigma},
\end{aligned}$$

by definition of a classical bracket. We conclude the proof of the first statement by observing that, again by definition, the $\mathcal{B}_i$ cover $\mathcal{F}$.

The second statement is trivial by definition of smooth distributions. ■

We now recall a stability result:

Lemma 19 (Lemma 34 from Block et al. (2022)) *Suppose that we are in the setting of Corollary 8 and let $\widehat{\mu}$ denote the empirical measure on the sampled x_i. If the function ℓ is L-Lipschitz in the first argument and*

$$\sup_{f,f' \in \mathcal{F}} \left| \|f - f'\|_{L^2(\mu)}^2 - \|f - f'\|_{L^2(\widehat{\mu})}^2 \right| \leq \Delta, \tag{D.1}$$

then for any fixed y,

$$\mathbb{E}\left[\|\ell(f_t(\cdot), y) - \ell(f_{t+1}(\cdot), y)\|_{L^1(p_t)} \right] \leq \frac{30L^3 \log(\eta)}{\sigma\eta} \cdot \mathbb{E}\left[1 + \sup_{f \in \mathcal{F}} \omega(f) \right] + \frac{2L\Delta}{\sigma}.$$

The assumption in Lemma 19 that the empirical and population norms are close to each other is a standard consequence of classical learning theory. We are thus ready to provide the main proof:

Proof [Proof of Corollary 8] By Proposition 7, it holds that Algorithm 1 experiences

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \mathcal{O}\left(\mathbb{E}\left[\sup_{f \in \mathcal{F}} \omega(f) \right] + \varepsilon T + \frac{T}{n} \cdot \log\left(T \cdot \mathcal{N}_{\mathcal{M},[]}(\mathcal{F}, \rho, \varepsilon) \right) + 2T\alpha \max_{\tau \leq T/n} \mathbb{E}\left[\|f_\tau - f_{\tau+1}\|^\beta \right] \right).$$

By the results of Lemma 18, we may take $\alpha = \frac{1}{\sigma}$ and $\beta = 1$ above to recover

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \mathcal{O}\left(\mathbb{E}\left[\sup_{f \in \mathcal{F}} \omega(f) \right] + \varepsilon T + \frac{T}{n} \cdot \log\left(T \cdot \mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma\varepsilon}{L} \right) \right) + \frac{T}{\sigma} \cdot \max_\tau \mathbb{E}\left[\|f_\tau - f_{\tau+1}\| \right] \right).$$

Observe now that if ℓ is L-Lipschitz, then $\widetilde{\ell}$, the cumulative loss over an epoch of length n, is Ln-Lipschitz by the triangle inequality. Thus, we see that

$$\max_\tau \mathbb{E}\left[\|f_\tau - f_{\tau+1}\| \right] \leq \mathcal{O}\left(\frac{L^3 n^3 \log(\eta)}{\sigma\eta} \cdot \mathbb{E}\left[1 + \sup_{f \in \mathcal{F}} \omega(f) \right] + \frac{2Ln\Delta}{\sigma} \right),$$

where we applied Lemma 19 on the event (D.1). Applying Block et al. (2022, Lemma 36), we see that with probability at least $1 - \delta$, it holds that we may take

$$\Delta \leq \mathcal{O}\left(\frac{\mathbb{E}\left[\sup_{f \in \mathcal{F}} \omega(f) \right]}{\eta m^{3/2}} + \frac{\sqrt{\log\left(\frac{1}{\delta}\right)}}{m} \right).$$

Applying Giné and Nickl (2021, Theorem 3.5.13), we see that

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} \omega(f) \right] \leq \mathcal{O}\left(\eta \sqrt{m \cdot \log\left(\mathcal{N}_{[]}(\mathcal{F}, \mu, \varepsilon) \right)} + \eta m \varepsilon \right)$$

for all $\varepsilon > 0$. Thus, setting

$$m = \sqrt{T}, \qquad \eta = \frac{1}{\sigma} \cdot \sqrt{\frac{TL^3n^3}{m}}, \qquad n = \frac{T^{1/5}\sigma^{2/5}}{L^{3/5}} \cdot \log^{2/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)$$

yields

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \widetilde{\mathcal{O}}\left(\frac{T^{4/5}L^{3/5}}{\sigma^{2/5}} \cdot \log^{3/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)\right)$$

with the same number of oracle calls. Thus, in particular, in order to achieve average regret at most ε, it suffices to call ERMOracle

$$\widetilde{\mathcal{O}}\left(\frac{\varepsilon^{-4}L^{3/5}}{\sigma^{2/5}} \cdot \log^{3/5}\left(\mathcal{N}_{[]}\left(\mathcal{F}, \mu, \frac{\sigma}{LT}\right)\right)\right)$$

times. ■

Appendix E. Proof of Theorem 9

In this section, we prove a more general version of Theorem 9. Recall that for fixed n and $1 \le \tau \le T/n$, we let

$$\mathcal{I}_\tau = \{t | (\tau - 1)n \le t \le \tau n\} \qquad \widetilde{\ell}_\tau(\theta) = \sum_{t \in \mathcal{I}_\tau} \ell(\theta, x_t)$$

$$L_t(\theta) = \sum_{s=1}^{t} \ell(\theta, x_s) \qquad \widetilde{L}_\tau(\theta) = \sum_{\tau'=1}^{\tau} \widetilde{\ell}_{\tau'}(\theta).$$

Further, we suppose that $\widetilde{\theta}_\tau$ is chosen such that for some real-valued function $\gamma : \mathbb{R}^d \to \mathbb{R}$, it holds that

$$\widetilde{L}_{\tau-1}(\widetilde{\theta}_\tau) - \eta \left\langle \xi, \widetilde{\theta}_\tau \right\rangle \le \gamma(\eta\xi) + \inf_{\theta \in \Theta} \widetilde{L}_{\tau-1}(\theta) - \eta \langle \xi, \theta \rangle . \tag{E.1}$$

We will assume that $\xi \sim \mathrm{Exp}(1)$ is a random vector in $\mathbb{R}^d$ whose coordinates are independently drawn according to a standard exponential distribution. For fixed ξ, let $\widetilde{\theta}_\tau(\xi)$ denote some $\widetilde{\theta}_\tau$ satisfying (E.1). We prove the following result:

Theorem 20 *Suppose that we are in the constrained online learning setting of Proposition 7 with $\Theta \subset \mathbb{R}^d$ such that $\sup_{\theta,\theta' \in \Theta} \|\theta - \theta'\|_1 = D < \infty$. Suppose further that the $\mathcal{Z}$-parameterized pseudo-metric ρ satisfies the pseudo-isometry property of Definition 6 with respect to ℓ^1 on $\mathbb{R}^d$ and that $\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\ell(\theta, z) - \ell(\theta', z)\right] \le \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', z)\right]$. If the learner plays Algorithm 2 and $\eta = \Omega(n^2)$ (with the exact relation given in (E.3)), then the expected regret is bounded:*

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \eta D d + 2 + + \frac{3DT}{n} \log\left(T\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, 1/T)\right)$$
$$+ 8T\alpha \cdot d^{\frac{\beta}{2-\beta}} \left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right] + 8\varepsilon n + \delta + 3D \log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}}, .$$

Note that Theorem 9 follows immediately by considering the special case $\gamma(\xi) = 0$.

The proof of Theorem 20 proceeds by first appealing to Lemma 16 and then bounding the stability term. The control of the stability term is broken into two parts: in the first, we apply the techniques of Agarwal et al. (2019b); Suggala and Netrapalli (2020) to show that if the stability term is small, then $\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1$ is small in expectation; in the second, we apply the pseudo-isometry assumption along with control of the generalized brackets to conclude the proof using a self-bounding approach.

E.1. Bounding the Stability Term

In this section, we apply the techniques of Suggala and Netrapalli (2020) to control the expected stability of $\widetilde{\theta}_\tau$ in $\|\cdot\|_1$. We have the following key lemma:

Lemma 21 *Suppose that $\theta_\tau \in \mathbb{R}^d$ is chosen according to (E.1). Suppose further that the ℓ^∞ diameter of Θ is bounded above by D. Then it holds that*

$$\mathbb{E}\left[\|\theta_\tau - \theta_{\tau+1}\|_1\right] \le d \cdot \sqrt{\frac{D}{\eta} \cdot \mathbb{E}\left[4\gamma(\xi) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|\right]}.$$

To prove the result, we require minor modifications of the key monotonicity lemmas from Suggala and Netrapalli (2020), where we apply their techniques without carrying a Lipschitz assumption on the losses. First, we have:

Lemma 22 *Suppose that $\xi, \xi' \in \mathbb{R}^d$ with $\theta = \widetilde{\theta}_\tau(\xi)$ and $\theta' = \widetilde{\theta}_\tau(\xi')$ for some fixed τ, as in (E.1). Then the following inequality holds:*

$$\eta \cdot \left\langle \xi' - \xi, \theta' - \theta \right\rangle \ge -\left(\gamma(\xi) + \gamma(\xi')\right).$$

Proof We compute:

$$\begin{aligned}
\widetilde{L}_\tau(\theta) - \eta\left\langle \xi, \theta \right\rangle &\le \widetilde{L}_\tau(\theta') - \eta\left\langle \xi, \theta' \right\rangle + \gamma(\xi) \\
&= \widetilde{L}_\tau(\theta') - \eta\left\langle \xi', \theta' \right\rangle + \eta\left\langle \xi' - \xi, \theta' \right\rangle + \gamma(\xi) \\
&\le \widetilde{L}_\tau(\theta) - \eta\left\langle \xi', \theta \right\rangle + \eta\left\langle \xi' - \xi, \theta' \right\rangle + \gamma(\xi) + \gamma(\xi').
\end{aligned}$$

The result follows. ■

The second necessary result is the analogue of Suggala and Netrapalli (2020, Lemma 6):

Lemma 23 *Suppose that $\xi, \xi' \in \mathbb{R}^d$ with $\widetilde{\theta}_\tau = \widetilde{\theta}_\tau(\xi)$, $\widetilde{\theta}_{\tau'} = \widetilde{\theta}_\tau(\xi')$ and $\widetilde{\theta}_{\tau+1}, \widetilde{\theta}'_{\tau+1}$ defined similarly for some fixed τ. Then the following inequality holds:*

$$\begin{aligned}
\min\left(\left\langle \widetilde{\theta}'_\tau, \xi' - \xi \right\rangle, \left\langle \widetilde{\theta}'_{\tau+1}, \xi' - \xi \right\rangle\right) \ge \max\left(\left\langle \widetilde{\theta}_\tau, \xi' - \xi \right\rangle, \left\langle \widetilde{\theta}_{\tau+1}, \xi' - \xi \right\rangle\right) \\
- \frac{2\left(\gamma(\xi) + \gamma(\xi')\right) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta}.
\end{aligned}$$

Proof By construction, we compute:

$$\begin{aligned}\widetilde{L}_\tau(\widetilde{\theta}_\tau) - \eta\left\langle \xi, \widetilde{\theta}_\tau\right\rangle &= \widetilde{L}_{\tau-1}(\widetilde{\theta}_\tau) - \eta\left\langle \xi, \widetilde{\theta}_\tau\right\rangle + \widetilde{\ell}_\tau(\widetilde{\theta}_\tau)\\ &\le \widetilde{L}_{\tau-1}(\widetilde{\theta}_{\tau+1}) - \eta\left\langle \xi, \widetilde{\theta}_{\tau+1}\right\rangle + \widetilde{\ell}_\tau(\widetilde{\theta}_\tau) + \gamma(\xi)\\ &= \widetilde{L}_\tau(\widetilde{\theta}_{\tau+1}) - \eta\left\langle \xi, \widetilde{\theta}_{\tau+1}\right\rangle + \gamma(\xi) + \widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1}).\end{aligned}$$

Again by construction, we have:

$$\begin{aligned}\widetilde{L}_\tau(\widetilde{\theta}_\tau) - \eta\left\langle \xi, \widetilde{\theta}_\tau\right\rangle &= \widetilde{L}_\tau(\widetilde{\theta}_\tau) - \eta\left\langle \xi', \widetilde{\theta}_\tau\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau\right\rangle\\ &\ge \widetilde{L}_\tau(\widetilde{\theta}'_{\tau+1}) - \eta\left\langle \xi', \widetilde{\theta}'_{\tau+1}\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau\right\rangle - \gamma(\xi')\\ &= \widetilde{L}_\tau(\widetilde{\theta}'_{\tau+1}) - \eta\left\langle \xi, \widetilde{\theta}'_{\tau+1}\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau - \widetilde{\theta}'_{\tau+1}\right\rangle - \gamma(\xi')\\ &\ge \widetilde{L}_\tau(\widetilde{\theta}_{\tau+1}) - \eta\left\langle \xi, \widetilde{\theta}_{\tau+1}\right\rangle + \eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau - \widetilde{\theta}'_{\tau+1}\right\rangle - \gamma(\xi').\end{aligned}$$

Combining the two preceding displays leads to the following inequality:

$$\eta\left\langle \xi' - \xi, \widetilde{\theta}_\tau - \widetilde{\theta}'_{\tau+1}\right\rangle \ge -2\left(\gamma(\xi) + \gamma(\xi')\right) - \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|.$$

An identical argument yields

$$\eta\left\langle \xi' - \xi, \widetilde{\theta}'_\tau - \widetilde{\theta}_{\tau+1}\right\rangle \ge -2\left(\gamma(\xi) + \gamma(\xi')\right) - \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|.$$

Applying Lemma 22 gives

$$\begin{aligned}\eta\left\langle \xi' - \xi, \widetilde{\theta}'_\tau - \widetilde{\theta}_\tau\right\rangle &\ge -\left(\gamma(\xi) + \gamma(\xi')\right)\\ \eta\left\langle \xi' - \xi, \widetilde{\theta}'_{\tau+1} - \widetilde{\theta}_{\tau+1}\right\rangle &\ge -\left(\gamma(\xi) + \gamma(\xi')\right).\end{aligned}$$

Combining the inequalities concludes the proof. ■

We are now ready to prove the stability bound:

Proof [Proof of Lemma 21] For some fixed τ and ξ, for all $1 \le i \le d$, let

$$\theta_{\max,i} = \max(\widetilde{\theta}_{\tau,i}, \widetilde{\theta}_{\tau+1,i}), \qquad \theta_{\min,i} = \min(\widetilde{\theta}_{\tau,i}, \widetilde{\theta}_{\tau+1,i}),$$

where $\widetilde{\theta}_{\tau,i}$ denotes the i^{th} coordinate of $\widetilde{\theta}_\tau$. Observe that $\left|\widetilde{\theta}_{\tau,i} - \widetilde{\theta}_{\tau+1,i}\right| = \theta_{\max,i} - \theta_{\min,i}$. Suppose that $\xi \sim \mathrm{Exp}(1)$ and let $\xi' = \xi + k\mathbf{e}_i$. Then, using the memoryless property of the exponential distribution and denoting

$$\mathbb{E}_{-i}[\cdot] = \mathbb{E}\left[\cdot|\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_d\right],$$

we have

$$\begin{aligned}\mathbb{E}_{-i}\left[\theta_{\min,i}\right] &= \mathbb{P}\left(\xi_i \le k\right)\cdot\mathbb{E}_{-i}\left[\theta_{\min,i}|\xi_i \le k\right] + \mathbb{P}\left(\xi_i > k\right)\cdot\mathbb{E}_{-i}\left[\theta_{\min,i}|\xi_i > k\right]\\ &\ge \left(1 - e^{-k}\right)\left(\mathbb{E}_{-i}\left[\theta_{\max,i}\right] - D\right) + e^{-k}\cdot\mathbb{E}_{-i}\left[\theta_{\min,i}|\xi_i > k\right]\\ &= \left(1 - e^{-k}\right)\left(\mathbb{E}_{-i}\left[\theta_{\max,i}\right] - D\right) + e^{-k}\cdot\mathbb{E}\left[\theta'_{\min,i}\right]\end{aligned}$$

where $\theta'_{\min,i} = \theta_{\min,i}(\xi')$. The inequality follows from the assumption on the diameter of Θ and the second equality follows from the memoryless property of the exponential distribution. Applying Lemma 23 and observing that $\left\langle \widetilde{\theta}_\tau, \xi' - \xi \right\rangle = k\widetilde{\theta}_{\tau,i}$, we see that

$$\mathbb{E}_{-i}\left[\theta'_{\min,i}\right] \geq \mathbb{E}_{-i}\left[\theta_{\max,i}\right] - \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right].$$

Thus, combining the previous displays tells us that

$$\mathbb{E}_{-i}\left[\theta_{\min,i}\right] \geq \mathbb{E}_{-i}[\theta_{\max,i}] - \left(1-e^{-k}\right)D - \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right].$$

Then,

$$\begin{aligned}
\mathbb{E}_{-i}\left[\left|\widetilde{\theta}_{\tau,i} - \widetilde{\theta}_{\tau+1,i}\right|\right] &= \mathbb{E}_{-i}\left[\theta_{\max,i} - \theta_{\min,i}\right] \\
&\leq \left(1-e^{-k}\right)D + \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right] \\
&\leq kD + \mathbb{E}_{-i}\left[\frac{2(\gamma(\xi)+\gamma(\xi')) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|}{\eta k}\right].
\end{aligned}$$

Summing over $1 \leq i \leq d$ and minimizing over k concludes the proof. ■

E.2. Concluding the Proof

We will apply the Be-the-Leader lemma; to do this, we need to bound the perturbation term and the stability terms. For the first, we have the following result:

Lemma 24 *Suppose that $\Theta \subset \mathbb{R}^d$ such that $\sup_\theta \|\theta\|_\infty \leq D$. Then*

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\langle\theta,\xi\rangle\right] \leq Dd,$$

where $\xi \sim \mathrm{Exp}(1)$.

Proof Observe that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\langle\theta,\xi\rangle\right] \leq \mathbb{E}\left[\sup_{\theta\in\Theta}\|\theta\|_\infty \|\xi\|_1\right] \leq D\mathbb{E}\left[\|\xi\|_1\right] = Dd.$$

■

We are now prepared to conclude the prove the main result:

Proof [Proof of Theorem 20] By Lemma 16, it suffices to bound the perturbation term and the stability terms independently. To bound the stability terms, note that by the assumption of Lipschitzness with respect to ρ, we have

$$\begin{aligned}\mathbb{E}\left[\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right] &= \mathbb{E}\left[\sum_{t\in\mathcal{I}_\tau} \ell(\widetilde{\theta}_\tau, x_t) - \ell(\widetilde{\theta}_{\tau+1}, x_t)\right] \\ &\le \mathbb{E}\left[\sum_{t\in\mathcal{I}_\tau} \rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, x_t)\right].\end{aligned}$$

We now apply Proposition 5 and observe that for all $\varepsilon, \delta > 0$,

$$\begin{aligned}\mathbb{E}\left[\sum_{t\in\mathcal{I}_\tau} \rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, x_t)\right] &\le \mathbb{E}\left[4n \sup_{\nu\in\mathcal{M}}\left[\rho(\widetilde{\theta}_\tau, \widetilde{\theta}_{\tau+1}, x)\right] + 8\varepsilon n + \delta + 3D\log\left(\frac{4\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)\right] \\ &\le 4n\left(\alpha\mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1^\gamma\right] + \beta\right) + 8\varepsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right) \\ &\le 4n\left(\alpha\left(\mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1\right]\right)^\beta\right) + 8\varepsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right),\end{aligned} \tag{E.2}$$

where the second inequality follows from the pseudo-isometry property and the last inequality follows by Jensen's and the fact that $\gamma \le 1$. By Lemma 21, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1\right] &\le d\sqrt{\frac{D}{\eta}\cdot\mathbb{E}\left[4\gamma(\xi) + \left|\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right|\right]} \\ &\le d\sqrt{\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right] + 8\varepsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)\right)} \\ &\quad + d\cdot\sqrt{\frac{2nD\alpha}{\eta}}\cdot\mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1\right]^{\frac{\beta}{2}},\end{aligned}$$

where the second inequality follows by plugging in the preceding display to Lemma 21 and applying subadditivity of the square root. Rearranging tells us that

$$\begin{aligned}&\mathbb{E}\left[\left\|\widetilde{\theta}_\tau - \widetilde{\theta}_{\tau+1}\right\|_1\right] \\ &\le \max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right] + 8\varepsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}}, \left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).\end{aligned}$$

Plugging this into (E.2) tells us that

$$\begin{aligned}&\mathbb{E}\left[\widetilde{\ell}_\tau(\widetilde{\theta}_\tau) - \widetilde{\ell}_\tau(\widetilde{\theta}_{\tau+1})\right] \le 8\varepsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right) \\ &+ 8n\alpha\cdot\max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right] + 8\varepsilon n + \delta + 3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}}, \left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).\end{aligned}$$

Summing over τ, we see that the stability term in Lemma 16 is bounded above by

$$8\varepsilon T + \delta\frac{T}{n} + \frac{3DT}{n}\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)$$
$$+ 8T\alpha\cdot\max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\varepsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}}, \left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).$$

Applying Lemma 16 and 24 tells us that the expected regret, then, is bounded by

$$\eta Dd + 8\varepsilon T + \delta\frac{T}{n} + \frac{3DT}{n}\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)$$
$$+ 8T\alpha\cdot\max\left(d^{\frac{\beta}{2-\beta}}\left(\frac{D}{\eta}\left(4\mathbb{E}\left[\gamma(\xi)\right]+8\varepsilon n+\delta+3D\log\left(\frac{\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,\varepsilon)}{\delta}\right)\right)\right)^{\frac{\beta}{4-2\beta}}, \left(\frac{2d^2Dn\alpha}{\eta}\right)^{\frac{1}{2-\beta}}\right).$$

If we set $\delta = \varepsilon = \frac{1}{T}$, note that as long as

$$\eta \geq \frac{d^{4-2\beta}\cdot D^{2-\beta}\cdot d^{4-2\beta}\cdot\alpha^2}{\left(4\mathbb{E}\left[\gamma(\xi)\right]+3D\log\left(T\cdot\mathcal{N}_{\mathcal{M},[]}(\Theta,\rho,1/T)\right)\right)^{2\beta}}\cdot n^2, \tag{E.3}$$

we have that the first argument to the maximum dominates the second, concluding the proof. ■

Appendix F. Proofs related to Piecewise Continuous Functions with Generalized Affine Boundaries

In this section, we provide a detailed proof of Theorem 10. We then state and prove a similar result, replacing $\overline{\ell}$ with the ℓ from (4.1), assuming an additional margin condition on the boundaries. The latter is included both for increased generality and for its application to the multi-step planning problem of Section 5.

F.1. Proof of Theorem 10

In this section we prove Theorem 10. We begin with the key step, showing that $\mathbb{P}\left(\overline{k}_\phi(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_\phi(\theta'_{\mathrm{d}},\mathbf{z})\right)\lesssim\left\|\theta_{\mathrm{d}}-\theta'_{\mathrm{d}}\right\|_1$ if $\mathbf{z}$ comes from a σ_{dir}-directionally smooth distribution. We then apply this result both to control the pseudo-isometry constant and to bound the generalized bracketing numbers.

We begin with the following lemma:

Lemma 25 *Suppose that Θ_{d}, $\overline{\phi}$, and $\overline{k}_{\overline{\phi}}$ are defined as in Theorem 10 and suppose $\mathbf{z}$ is chosen from a σ_{dir}-directionally smooth distribution such that $\|\mathbf{z}\|_\infty\leq B$ almost surely. Then,*

$$\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right)\leq\frac{AB}{a\sigma_{\mathrm{dir}}}\cdot\left\|\theta_{\mathrm{d}}-\theta'_{\mathrm{d}}\right\|_1.$$

Proof We begin by observing that

$$\begin{aligned}
\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right) &= \mathbb{P}\left(\arg\max_{k\in[K]} \sum_{k'\neq k} \mathbb{I}\left[\overline{\phi}(\theta_{\mathrm{d}}, k, k', z) \geq 0\right] \neq \arg\max_{k\in[K]} \sum_{k'\neq k} \mathbb{I}\left[\overline{\phi}(\theta'_{\mathrm{d}}, k, k', z) \geq 0\right]\right) \\
&\leq \mathbb{P}\left(\bigcup_{k,k'\in[K]} \left\{\overline{\phi}(\theta_{\mathrm{d}}, k, k', z) \geq 0 > \overline{\phi}(\theta'_{\mathrm{d}}, k, k', z)\right\}\right) \\
&= \mathbb{P}\left(\bigcup_{k,k'\in[K]} \left\{\psi\left(\langle \mathbf{w}_{kk'}, \mathbf{z}\rangle\right) \geq 0 > \psi\left(\left\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\right\rangle\right)\right\}\right) \\
&\leq \sum_{k,k'\in[K]} \mathbb{P}\left(\psi\left(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle\right) \geq 0 > \psi\left(\left\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\right\rangle\right)\right), \qquad \text{(F.1)}
\end{aligned}$$

where the first inequality follows from the fact that $\overline{\phi}$ is antisymmetric in (k, k'), the second equality follows from the construction of $\overline{\phi}$, and the last inequality follows from a union bound. We now observe that for fixed $k, k' \in [K]$,

$$\begin{aligned}
&\mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \geq 0 > \psi(\left\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\right\rangle)\right) \\
&\quad\leq \mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) \leq \left|\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle) - \psi(\left\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\right\rangle)\right|\right) \\
&\quad\leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle|) \leq A\left|\left\langle \mathbf{w}_{kk'} - \mathbf{w}'_{kk'}, (\mathbf{z}, 1)\right\rangle\right|\right) \\
&\quad\leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1)\rangle|) \leq AB\left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1\right) \\
&\quad\leq \frac{AB}{a\sigma_{\mathrm{dir}}} \cdot \left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1, \qquad \text{(F.2)}
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the assumption of ψ being A-Lipschitz, the third inequality follows from Hölder's inequality and the fact that $\|\mathbf{z}\|_\infty \leq B$ almost surely, and the final inequality follows from the fact that $\mathbf{w}_{kk'} \in \mathcal{S}^d$ and the directional smoothness of $\mathbf{z}$, along with Block and Simchowitz (2022, Lemma 36). Plugging in to the first display and summing tells us that

$$\begin{aligned}
\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right) &\leq \sum_{k,k'\in[K]} \frac{AB}{a\sigma_{\mathrm{dir}}} \cdot \left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1 \\
&= \frac{AB}{a\sigma_{\mathrm{dir}}} \left\|\theta_{\mathrm{d}} - \theta'_{\mathrm{d}}\right\|_1,
\end{aligned}$$

which concludes the proof. ■

We now use Lemma 25 to show that the pseudo-isometry property holds:

Lemma 26 *Suppose that we are in the situation of Theorem 10 and $\mathcal{M}$ is the class of σ_{dir}-directionally smooth distributions whose $\|\cdot\|_\infty$ is almost surely bounded by $B > 0$. Then*

$$\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta', \mathbf{z})\right] \leq \frac{2AB}{a\sigma_{\mathrm{dir}}} \cdot \left\|\theta - \theta'\right\|_1.$$

Proof We compute:

$$\begin{aligned}
\mathbb{E}_\nu\left[\rho(\theta,\theta',\mathbf{z})\right] &= \mathbb{E}_\nu\left[2\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right]+\max_{k\in[K]}\left\|\theta_{\mathrm{c}}^{(k)}-\theta_{\mathrm{c}}^{\prime(k)}\right\|_1\right]\\
&\le \frac{2AB}{a\sigma_{\mathrm{dir}}}\left\|\theta_{\mathrm{d}}-\theta'_{\mathrm{d}}\right\|_1+\max_{k\in[K]}\left\|\theta_{\mathrm{c}}^{(k)}-\theta_{\mathrm{c}}^{\prime(k)}\right\|_1\\
&\le \frac{2A(B\vee 1)}{a\sigma_{\mathrm{dir}}}\cdot\left(\left\|\theta_{\mathrm{d}}-\theta'_{\mathrm{d}}\right\|_1+\max_{k\in[K]}\left\|\theta_{\mathrm{c}}^{(k)}-\theta_{\mathrm{c}}^{\prime(k)}\right\|_1\right)\\
&\le \frac{2A(B\vee 1)}{a\sigma_{\mathrm{dir}}}\cdot\left\|\theta-\theta'\right\|_1,
\end{aligned}$$

where the first inequality follows from linearity of expectation and Lemma 25. The result follows. ■

We now control the generalized bracketing number of Θ:

Lemma 27 *If we are in the situation of Theorem 10 then for any $\varepsilon>0$,*

$$\mathcal{N}_{\mathcal{M},[]}\left(\Theta,\rho,\varepsilon\right)\le\left(\frac{9AK^2B}{a\sigma_{\mathrm{dir}}\varepsilon}\right)^{K^2d}.$$

Proof Let $\mathcal{N}=\left\{\theta_i=(\theta_{\mathrm{c}}^i,\theta_{\mathrm{d}}^i)\right\}$ denote an $\widetilde{\varepsilon}$-net of Θ with respect to ℓ_1, where $\widetilde{\varepsilon}=\frac{a\sigma_{\mathrm{dir}}}{3K^2AB}\cdot\varepsilon$, and let

$$\mathcal{B}_i=\left\{\theta\in\Theta|\left\|\theta-\theta_i\right\|_1\le\widetilde{\varepsilon}\right\}.$$

We claim that $\{(\theta_i,\mathcal{B}_i)\}$ forms a generalized ε-bracket with respect to $\mathcal{M}$, the class of σ_{dir}-directionally smooth distributions with ℓ_∞ norm bounded by B. To see this, observe first that by the definition of an $\widetilde{\varepsilon}$-net, it holds that the union of the $\mathcal{B}_i$ covers Θ. Now, fix $\theta_i,\mathbf{z}$ and observe that for $\theta\in\mathcal{B}_i$, we have

$$\begin{aligned}
\rho(\theta,\theta_i,\mathbf{z}) &= 2\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}^i,\mathbf{z})\right]+\max_{k\in[K]}\left\|\theta_{\mathrm{c}}-\theta_{\mathrm{c}}^i\right\|_1\\
&\le 2\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}^i,\mathbf{z})\right]+\left\|\theta_{\mathrm{c}}-\theta_{\mathrm{c}}^i\right\|_1.
\end{aligned}$$

Now we compute:

$$\begin{aligned}
&\mathbb{P}\left(\exists\theta\in\mathcal{B}_i\text{ s.t. }\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z})\neq\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}^i,\mathbf{z})\right)\\
&\le\mathbb{P}\left(\exists\theta\in\mathcal{B}_i,k,k'\in[K]\text{ s.t. }\psi(\langle\mathbf{w}_{kk'},(\mathbf{z},1)\rangle)\ge 0>\psi(\langle\mathbf{w}_{kk'}^i,(\mathbf{z},1)\rangle)\right)\\
&\le\mathbb{P}\left(\exists\theta\in\mathcal{B}_i,k,k'\in[K]\text{ s.t. }\left|\psi(\langle\mathbf{w}_{kk'}^i,(\mathbf{z},1)\rangle)\right|\le AB\left\|\mathbf{w}_{kk'}^i-\mathbf{w}_{kk'}\right\|_1\right)\\
&\le\mathbb{P}\left(\exists\theta\in\mathcal{B}_i,k,k'\in[K]\text{ s.t. }\left|\psi(\langle\mathbf{w}_{kk'}^i,(\mathbf{z},1)\rangle)\right|\le AB\widetilde{\varepsilon}\right)\\
&\le\sum_{k,k'\in[K]}\mathbb{P}\left(\left|\psi(\langle\mathbf{w}_{kk'}^i,(\mathbf{z},1)\rangle)\right|\le B\widetilde{\varepsilon}\right)\\
&\le\frac{K^2AB}{a\sigma_{\mathrm{dir}}}\widetilde{\varepsilon},
\end{aligned}$$

where the first inequality follows from the same reasoning as in (F.1), the second inequality follows from the same reasoning as in (F.2), the third inequality follows from the construction of $\mathcal{B}_i$, the fourth inequality follows from a union bound, and the final inequality follows from the assumption of directional smoothness. Thus, we note,

$$\begin{aligned}
\sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu \left[\sup_{\theta\in\mathcal{B}_i} \rho(\theta,\theta_i,\mathbf{z})\right] &= \sup_{\nu\in\mathcal{M}} \mathbb{E}_\nu \left[\sup_{\theta\in\mathcal{B}_i} 2\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta^i_{\mathrm{d}},\mathbf{z})\right] + \left\|\theta_{\mathrm{c}}-\theta^i_{\mathrm{c}}\right\|_1\right] \\
&\leq 2\cdot \sup_{\nu\in\mathcal{M}} \left\{\mathbb{E}_\nu \left[\sup_{\theta\in\mathcal{B}_i} \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta^i_{\mathrm{d}},\mathbf{z})\right]\right]\right\} + \sup_{\theta\in\mathcal{B}_i} \left\|\theta_{\mathrm{c}}-\theta^i_{\mathrm{c}}\right\|_1 \\
&\leq \left(\frac{2K^2AB}{a\sigma_{\mathrm{dir}}}+1\right)\widetilde{\varepsilon} \\
&\leq \varepsilon.
\end{aligned}$$

Thus we have shown that $\{(\theta_i,\mathcal{B}_i)\}$ is a generalized ε-bracket with respect to $\mathcal{M}$. It remains to bound the size. To do this, note that by construction, it suffices to bound the size of an $\widetilde{\varepsilon}$-cover with respect to ℓ_1 on Θ. But note that Θ is contained in an ℓ_1 ball of radius D and thus a simple volume argument (see Vershynin (2018, Section 4.2.1) for example) tells us that, because $\Theta \subset \mathbb{R}^{Kd+K^2(d-1)}$, we may take

$$|\mathcal{N}| \leq \left(\frac{3D}{\widetilde{\varepsilon}}\right)^{Kd+K^2(d-1)} \leq \left(\frac{9K^2DAB}{a\sigma_{\mathrm{dir}}\varepsilon}\right)^{K^2d}.$$

The result follows. ∎

The proof of Theorem 10 follows from combining Lemmas 27 and 26.

F.2. Proof of Corollary 11

By applying Theorems 9 and 10, it suffices to show that $\overline{\ell}$ is Lipschitz with respect to the ρ defined in (4.3). We observe, however, that

$$\begin{aligned}
&\overline{\ell}(\theta,\mathbf{z}) - \overline{\ell}(\theta',\mathbf{z}) \\
&= (\overline{\ell}(\theta,\mathbf{z}) - \overline{\ell}(\theta',\mathbf{z}))\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z}) = \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right] + (\overline{\ell}(\theta,\mathbf{z}) - \overline{\ell}(\theta',\mathbf{z}))\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right] \\
&\leq 2\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right] + \max_{k\in[K]} g_k(\theta_{\mathrm{c}},\mathbf{z}) - g_k(\theta'_{\mathrm{c}},\mathbf{z}) \\
&\leq 2\cdot\mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}},\mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}},\mathbf{z})\right] + \max_{k\in[K]} \left\|\theta^{(k)}_{\mathrm{c}} - \theta'^{(k)}_{\mathrm{c}}\right\|_1.
\end{aligned}$$

The result then follows by the definition of ρ.

F.3. Replacing $\overline{\ell}$ with ℓ

While the work in Appendix F.1 sufficed to prove Theorem 10, for the sake of planning, we may wish to replace the loss function $\overline{\ell}$ with the much simpler ℓ of (4.1). In order to apply our techniques, however, we will require that $\theta_{\mathrm{d}} = (\mathbf{w}_1,\ldots,\mathbf{w}_K)$ satisfies a certain margin condition. The analogue of Appendix F.1 is thus:

Theorem 28 *Suppose that $\mathcal{Z} \subset \mathbb{R}^d$ and Θ is a subset of a Euclidean space of ℓ_1 diameter bounded by D, and that*

$$\Theta_{\mathrm{d}} \subset \left\{ (\mathbf{w}_1, \dots, \mathbf{w}_K) \in (\mathcal{S}^d)^{\times K} \mid \min_{k \neq k' \in [K]} \left\| \mathbf{w}_{k,\widehat{d}} - \mathbf{w}_{k',\widehat{d}} \right\|_2 \geq \gamma \right\},$$

where we denote by $\mathbf{w}_k$ the coordinates of a given $\theta_{\mathrm{d}} \in \Theta_{\mathrm{d}}$ and let $\mathbf{w}_{k,\widehat{d}}$ denote the first d coordinates of $\mathbf{w}_k$. Suppose further that $\phi(\theta_{\mathrm{d}}, k, \mathbf{z}) = \psi(\langle \mathbf{w}_k, (\mathbf{z}, 1) \rangle)$ for some link function ψ, as in Theorem 10. If $\mathcal{M}$ consists of the class of σ_{dir}-directionally smooth distributions such that $\|\mathbf{z}\|_\infty \leq B$ almost surely for some $B \geq 1$, then with ρ as in (4.3), it holds that ρ is a pseudo-metric satisfying the pseudo-isometry property with $\alpha = \frac{4AB}{a\gamma\sigma_{\mathrm{dir}}}$ and $\beta = 1$. Furthermore, for all $\varepsilon > 0$,

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon) \leq \left(\frac{18AK^2BD}{a\gamma\sigma_{\mathrm{dir}}\varepsilon} \right)^{2K(d+1)}$$

Proof The proof is essentially the same as that of Theorem 10 given in Appendix F.1. In fact, we simply need to prove a version of Lemma 25 and the rest of the proof applies, *mutatis mutandis*. To see this, note that we may mimic the aforementioned proof by setting $\mathbf{w}_{kk'} = \mathbf{w}_k - \mathbf{w}_{k'}$; this is almost the same as the previous scenario with the exception that we may now take $\|\mathbf{w}_{kk'}\| \neq 1$. This causes a problem only in the application of directional smoothness; thus, suppose that $\mathbf{z}$ is σ_{dir} directionally smooth and observe that the chain of inequalities in (F.1) remains valid. Continuing, we see that for fixed $k, k' \in [K]$,

$$\begin{aligned}
&\mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1) \rangle) \geq 0 > \psi(\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1) \rangle)\right) \\
&\leq \mathbb{P}\left(\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1) \rangle) \leq \left|\psi(\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1) \rangle) - \psi(\langle \mathbf{w}'_{kk'}, (\mathbf{z}, 1) \rangle)\right|\right) \\
&\leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1) \rangle|) \leq A \left|\langle \mathbf{w}_{kk'} - \mathbf{w}'_{kk'}, (\mathbf{z}, 1) \rangle\right|\right) \\
&\leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_{kk'}, (\mathbf{z}, 1) \rangle|) \leq AB \left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1\right) \\
&\leq \mathbb{P}\left(\psi(|\langle \mathbf{w}_k - \mathbf{w}_{k'}, (\mathbf{z}, 1) \rangle|) \leq AB(\left\|\mathbf{w}_k - \mathbf{w}'_k\right\|_1 + \left\|\mathbf{w}_{k'} - \mathbf{w}'_{k'}\right\|_1)\right) \\
&\leq \frac{AB}{a\sigma_{\mathrm{dir}} \cdot \left\|\mathbf{w}_{k,\widehat{d}} - \mathbf{w}_{k',\widehat{d}}\right\|_2} \cdot \left(\left\|\mathbf{w}_k - \mathbf{w}'_k\right\|_1 + \left\|\mathbf{w}_{k'} - \mathbf{w}'_{k'}\right\|_1\right) \\
&\leq \frac{AB}{a\sigma_{\mathrm{dir}}\gamma} \cdot \left(\left\|\mathbf{w}_k - \mathbf{w}'_k\right\|_1 + \left\|\mathbf{w}_{k'} - \mathbf{w}'_{k'}\right\|_1\right),
\end{aligned}$$

where the first four inequalities follow as in (F.2), the fifth inequality follows from the definition of $\mathbf{w}_{kk'}$, the penultimate inequality follows as in the previous proof, and the last inequality follows from the margin assumption. We then may apply the identical logic as in the proof of Theorem 10 going forward and the result holds, after channging the dimension of Θ_{d} and adding a multiplicative factor of 2 to account for summing twice the differences $\|\mathbf{w}_k - \mathbf{w}'_k\|_1$ above. ■

Using the identical argument as in Corollary 11, we arrive at the following regret bound for Algorithm 2 in the situation of Theorem 28:

Corollary 29 *Suppose that ℓ is as in (4.1) with ϕ and Θ as in Theorem 28 with $B \geq 1$ and ℓ uniformly bounded in magnitude by 1. If we set $\eta = \widetilde{\mathcal{O}}\left(\left(TKAdDB(\gamma a \sigma_{\mathrm{dir}})^{-1}\right)^{2/3}\right)$ and $n = \sqrt{\eta}$,*

then Algorithm 2 expereinces

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\left(\frac{TAKdBD}{a\gamma\sigma_{\mathrm{dir}}}\right)^{2/3}\right).$$

In particular, the oracle complexity of achieving average regret ε is $\widetilde{\mathcal{O}}\left(\frac{AKdDB}{\gamma a\sigma_{\mathrm{dir}}\varepsilon^2}\right)$.

Appendix G. Proofs from Section 4.3

In this appendix, we discuss the polynomially smooth assumption and provide examples of common distributions satisfying this requirement. We also give a counter example that demonstrates that directional smoothness is *not* sufficient to ensure polynomial smoothness. We then prove Theorem 13.

G.1. Polynomial Smoothness

In this section, we discuss the notion of Polynomial smoothness found in Definition 12. We begin by recalling the simple Glazer and Mikulincer (2022, Example 3), which demonstrates that σ_{dir}-smoothness is strictly weaker than polynomial smoothness:

Example 1 *We show that there exists a class of distributions that is σ_{dir}-directionally smooth with σ_{dir} decaying polynomially with dimension, but is $\sigma_{\mathrm{poly},2}$-polynomially smooth only for $\sigma_{\mathrm{poly},2}$ decaying exponentially with dimension. Let ν_d denote the uniform measure on the unit Euclidean ball $\mathcal{B}^d \subset \mathbb{R}^d$. We observe that by Block and Simchowitz (2022, Example 1), ν_d is σ_{dir}-directionally smooth with $\sigma_{\mathrm{dir}} = \Omega\left(\frac{1}{d}\right)$. On the other hand, for the polynomial $f(x) = \frac{1}{d}\cdot\|x\|_2^2$, we see that $\mathrm{coeff}_2(f) = 1$, but concentration of measure (see Vershynin (2018) for example) tells us that $\mathbb{P}\left(|f(x)-1| \leq \varepsilon\right) \geq 1 - e^{-\Omega(d)\varepsilon}$.*

We now consider what kinds of distributions are polynomially smooth. The key tool in our arsenal is (a special case of) the famous inequality of Carbery-Wright, which says:

Theorem 30 (Theorem 8 from Carbery and Wright (2001)) *If ν is a log-concave measure on $\mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$ is a degree r polynomial, then for all $\varepsilon > 0$, if $X \sim \nu$,*

$$\mathbb{P}\left(|f(X)| \leq \varepsilon\right) \leq Cr\frac{\varepsilon^{\frac{1}{r}}}{\mathbb{E}\left[f(X)^2\right]^{\frac{1}{2r}}}.$$

Thus, Theorem 30 tells us that if ν is log-concave and we can be assured that all polynomials f with $\mathrm{coeff}_r(f) \geq 1$ have large second moment, then ν is $\sigma_{\mathrm{poly},r}$-polynomially smooth with $\sigma_{\mathrm{poly},r}$ depending nicely on the dimension. Proving that polynomials with large coefficients indeed have large second moment is still an active area of research, but we provide as an example the following result, rephrased into our language:

Theorem 31 (Corollary 4 from Glazer and Mikulincer (2022)) *Suppose that $\nu = \mu^{\otimes n}$ is a log-concave, isotropic product measure. Then ν is $\sigma_{\mathrm{poly},r}$-polynomially smooth with $\sigma_{\mathrm{poly},r} \geq \Omega\left(\frac{1}{r}\right)$.*

Note that the above result encompasses Gaussian measures and can be scaled as needed. Further results in the direction of Glazer and Mikulincer (2022) would translate directly into a wider class of measures known to be $\sigma_{\mathrm{poly},r}$-polynomially smooth.

G.2. Proof of Theorem 13

In this section, we provide a proof of Theorem 13 that follows the approach of Corollary 11. While we do not repeat the argument, we observe in passing that replacing the tournament-style $\overline{\ell}$ with the simpler function ℓ from (4.1) and including a margin assumption allows for an analogue of Theorem 28 in this setting. To prove Theorem 13, we begin by proving an analogue of Theorem 10:

Theorem 32 *Suppose that $\mathcal{Z} \subset \mathbb{R}^d$ and that Θ is a subset of Euclidean space with ℓ_1 diameter bounded by D. Let Θ_{d} parameterize the set of tuples of $\binom{K}{2}$ degree r polynomials $(f_{\mathbf{w}_{kk'}})$ on $\mathbb{R}^d$ such that $\mathrm{coeff}_r(f_{\mathbf{w}_{kk'}}) = 1$ for all $k \in [K]$. If $\overline{\phi}(\theta_{\mathrm{d}}, k, k', \mathbf{z}) = f_{\mathbf{w}_{kk'}}(\mathbf{z})$ and $\mathcal{M}$ is the class of $\sigma_{\mathrm{poly},r}$-polynomially smooth distributions such that $\|\mathbf{z}\|_\infty \le B$ almost surely, then the ρ defined in (4.3) is a pseudo-metric satisfying the pseudo-isometry property with $\alpha = \frac{2B^r D}{\sigma_{\mathrm{poly},r}}$ and $\beta = \frac{1}{r}$. Furthermore, for all $\varepsilon > 0$,*

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon) \le \left(\frac{9K^2 B}{\sigma_{\mathrm{poly},r}\varepsilon}\right)^{K^2 r^2 d^r}.$$

To prove the result, we need analogues of Lemmas 25,26, and 27. We begin with proving the Lipschitzness in expectation of the first term of ρ:

Lemma 33 *Suppose that Θ_{d}, $\overline{\phi}$, and $\overline{k}_{\overline{\phi}}$ are as in Theorem 32 and suppose that $\mathbf{z}$ is chosen from a $\sigma_{\mathrm{poly},r}$-polynomially smooth distribution such that $\|\mathbf{z}\|_\infty \le B$ almost surely for some $B \ge 1$. Then,*

$$\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right) \le \frac{B^r K^{2-\frac{4}{r}}}{\sigma_{\mathrm{poly},r}} \cdot \left\|\theta_{\mathrm{d}} - \theta'_{\mathrm{d}}\right\|_1^{\frac{1}{r}}.$$

Proof By the same argument as in (F.1), we have that

$$\mathbb{P}\left(\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right) \le \sum_{k,k' \in [K]} \mathbb{P}\left(f_{\mathbf{w}_{kk'}}(\mathbf{z}) \ge 0 > f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right).$$

Observe that by the triangle inequality,

$$\left|f_{\mathbf{w}_{kk'}}(\mathbf{z}) - f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right| \le B^r \cdot \left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1.$$

Thus, applying the argument in (F.2), we have

$$\begin{aligned}
\mathbb{P}\left(f_{\mathbf{w}_{kk'}}(\mathbf{z}) \ge 0 > f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right) &\le \mathbb{P}\left(\left|f_{\mathbf{w}_{kk'}}(\mathbf{z})\right| \le \left|\left|f_{\mathbf{w}_{kk'}}(\mathbf{z}) - f_{\mathbf{w}'_{kk'}}(\mathbf{z})\right|\right|\right) \\
&\le \mathbb{P}\left(\left|f_{\mathbf{w}_{kk'}}(\mathbf{z})\right| \le B^r \cdot \left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1\right) \\
&\le \frac{B}{\sigma_{\mathrm{poly},r}} \cdot \left\|\mathbf{w}_{kk'} - \mathbf{w}'_{kk'}\right\|_1^{\frac{1}{r}},
\end{aligned}$$

where the last inequality follows from the definition of polynomial smoothness. Applying Hölder's inequality and summing concludes the proof. ■

Using this result, we can prove an analogue of Lemma 26:

Lemma 34 *Suppose that we are in the situation of Theorem 32 and $\mathcal{M}$ is the class of $\sigma_{\mathrm{poly},r}$-polynomially smooth distributions such that the infinity norms of samples are uniformly bounded almost surely by some $B \geq 1$. If the ℓ_1 diameter of Θ is bounded by D, then*

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta, \theta', \mathbf{z})\right] \leq \frac{2BD}{\sigma_{\mathrm{poly},r}} \cdot \left\|\theta - \theta'\right\|^{\frac{1}{r}}.$$

Proof We compute:

$$\begin{aligned}
\mathbb{E}_\nu \left[\rho(\theta, \theta', \mathbf{z})\right] &= \mathbb{E}_\nu \left[2 \cdot \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta'_{\mathrm{d}}, \mathbf{z})\right] + \max_{k \in [K]} \left\|\theta_{\mathrm{c}}^{(k)} - \theta_{\mathrm{c}}^{\prime(k)}\right\|_1\right] \\
&\leq \frac{2B}{\sigma_{\mathrm{poly},r}} \cdot \left\|\theta_{\mathrm{d}} - \theta'_{\mathrm{d}}\right\|_1^{\frac{1}{r}} + \left\|\theta_{\mathrm{c}} - \theta'_{\mathrm{c}}\right\|_1 \\
&\leq \frac{2BD}{\sigma_{\mathrm{poly},r}} \cdot \left\|\theta - \theta'\right\|^{\frac{1}{r}},
\end{aligned}$$

where the second inequality follows from Lemma 33 and the last inequality follows by the assumption on the diameter. ■

Finally, we require an analogue of Lemma 27:

Lemma 35 *If we are in the situation of Theorem 32 then for any $\varepsilon > 0$, it holds that*

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \rho, \varepsilon) \leq \left(\frac{9K^2B}{\sigma_{\mathrm{poly},r}\varepsilon}\right)^{K^2r^2d^r}.$$

Proof We mimic the proof of Lemma 27 but apply Lemma 33 instead of Lemma 25. In particular, we suppose that $\mathcal{N} = \left\{\theta_i = \left(\theta_{\mathrm{c}}^i, \theta_{\mathrm{d}}^i\right)\right\}$ is an $\widetilde{\varepsilon}$-net of Θ with respect to ℓ_1, where $\widetilde{\varepsilon} = \left(\frac{\sigma_{\mathrm{poly},r}}{3K^2B} \cdot \varepsilon\right)^r$ and similarly let $\mathcal{B}_i \subset \Theta$ denote the set of parameters within $\widetilde{\varepsilon}$ of θ_i in ℓ_1 norm. We compute as in the proof of Lemma 27 that

$$\begin{aligned}
\mathbb{P}\left(\exists \theta \in \mathcal{B}_i \text{ s.t. } \overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}^i, \mathbf{z})\right) &\leq \mathbb{P}\left(\exists \theta \in \mathcal{B}_i \text{ s.t. } \left|f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right| \leq \left|f_{\mathbf{w}_{kk'}}(\mathbf{z}) - f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right|\right) \\
&\leq \mathbb{P}\left(\exists \theta \in \mathcal{B}_i \text{ s.t. } \left|f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right| \leq B^r \cdot \left\|\mathbf{w}_{kk'} - \mathbf{w}_{kk'}^i\right\|_1\right) \\
&\leq \mathbb{P}\left(\exists \theta \in \mathcal{B}_i \text{ s.t. } \left|f_{\mathbf{w}_{kk'}^i}(\mathbf{z})\right| \leq B^r \cdot \widetilde{\varepsilon}\right) \\
&\leq \frac{K^2B}{\sigma_{\mathrm{poly},r}} \cdot \widetilde{\varepsilon}^{\frac{1}{r}}.
\end{aligned}$$

Thus

$$\begin{aligned}
\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu \left[\rho(\theta, \theta_i, \mathbf{z})\right] &\leq 2 \cdot \sup_{\nu \in \mathcal{M}} \left\{\mathbb{E}_\nu \left[\sup_{\theta \in \mathcal{B}_i} \mathbb{I}\left[\overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}, \mathbf{z}) \neq \overline{k}_{\overline{\phi}}(\theta_{\mathrm{d}}^i, \mathbf{z})\right]\right]\right\} + \sup_{\theta \in \mathcal{B}_i} \left\|\theta_{\mathrm{c}} - \theta_{\mathrm{c}}^i\right\|_1 \\
&\leq \frac{2K^2B}{\sigma_{\mathrm{poly},r}} \cdot \widetilde{\varepsilon}^{\frac{1}{r}} + \widetilde{\varepsilon} \\
&\leq \frac{3K^2B}{\sigma_{\mathrm{poly},r}} \cdot \widetilde{\varepsilon}^{\frac{1}{r}} \\
&\leq \varepsilon.
\end{aligned}$$

Thus $\mathcal{N}$ is a generalized ε-bracket with respect to $\mathcal{M}$ and ρ. We may bound the size of $\mathcal{N}$ in the same way as in the proof of Lemma 27, after observing that $\mathbf{w}_{kk'}$ lives in a space of dimension

$$\sum_{i \le r} \binom{d}{i} \le \left(\frac{ed}{r}\right)^r.$$

■

Combining Lemmas 34 and 35 concludes the proof of Theorem 32. We are now ready to prove Theorem 13:

Proof [Proof of Theorem 13] We observe by the same logic as in the proof of Corollary 11 that $\overline{\ell}$ is Lipschitz with respect to ρ. Thus we may apply Theorem 9 and Theorem 32 to get that if Algorithm 2 is played, then

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \widetilde{\mathcal{O}}\left(\eta + \frac{T}{n} K^2 r^2 d^r \log\left(\frac{1}{\sigma_{\mathrm{poly},r}}\right) + \frac{TB^r D}{\sigma_{\mathrm{poly},r}} \cdot \left(\frac{K^2 r^2 d^r}{\eta}\right)^{\frac{1}{4r-2}}\right).$$

Setting

$$\eta = \widetilde{\Theta}\left(\left(\frac{TK^2 r^2 d^r DB}{\sigma_{\mathrm{poly},r}}\right)^{\frac{4r-2}{4r-1}}\right) \qquad n = \widetilde{\Theta}\left(\left(\frac{TK^2 r^2 d^r DB}{\sigma_{\mathrm{poly},r}}\right)^{\frac{2r-1}{4r-1}}\right)$$

concludes the proof. ■

Appendix H. Proof of Theorem 14

In this section, we state and prove a formal version of Theorem 14. We recall that we are in the situation of (5.1) and that our aim is to minimize the regret with respect to the best plan $\bar{\mathbf{u}}_{1:H}$. Throughout, we let $\|\cdot\|_1$ denote the ℓ_1 norm interpreted in the natural sense for concatenated vectors; e.g. $\|\bar{\mathbf{u}}_{1:H}\|_1 = \sum_{h=1}^H \|\mathbf{u}_h\|_1$. We begin by introducing a notation that will substantially simplify our presentation:

Definition 36 *For a given sequence of modes $k_{1:H} \in [K]^H$, recall from (5.1) that the states evolve as*

$$\tilde{\mathbf{x}}_{t,h+1}(\theta; k_{1:H}) = g_{t,h,k_h}(\tilde{\mathbf{x}}_{t,h}(\theta; k_{1:H}), \mathbf{u}_{t,h}(\theta)) + \boldsymbol{\eta}_{t,h}, \;\; \tilde{\mathbf{x}}_{t,1}(\theta; k_{1:H}) = \mathbf{x}_{t,1}, \;\; \mathbf{u}_{t,h}(\theta) = \bar{\mathbf{u}}_h + \boldsymbol{\xi}_{t,h},$$

where the difference between here and the situation in (5.1) is that here the mode sequence is given, whereas in (5.1) it was state and input dependent. We define the function

$$G_t(\theta; k_{1:H}) := (\tilde{\mathbf{x}}_{t,1}(\theta; k_{1:H}), \ldots, \tilde{\mathbf{x}}_{t,H}(\theta; k_{1:H})) \in \mathcal{X}^H \subset \mathbb{R}^{mH},$$

which maps a plan and given mode sequence to the associated trajectory.

We are now ready to state a formal version of Theorem 14:

Theorem 37 *For fixed planning horizon H, suppose that trajectories $\mathbf{x}_{t,1:H}$ evolve as in (5.1), where the learner chooses a plan $\theta \in \mathcal{K} \subset \mathcal{U}^{\times H}$ at each time t and the adversary presents the tuple $\mathbf{z}_t$ described in Section 5. Assume that for all $t \in [T]$ the following properties hold almost surely under the adversary's strategy p_t:*

1. *$\mathbf{x}_{t,1} \mid \mathcal{F}_t$ and $(\boldsymbol{\eta}_{t,h}, \boldsymbol{\xi}_{t,h}) \mid \mathcal{F}_{t,h-1}$ are σ_{dir}-directionally smooth.*
2. *For all mode sequences $k_{1:H} \in [K]^H$ and $\theta, \theta' \in \mathcal{U}^{\times H}$, $\|G_t(\theta, k_{1:H}) - G_t(\theta', k_{1:H})\| \le L \|\theta - \theta'\|_1$, i.e., the functions G_t are L-Lipschitz with respect to the ℓ_1 norm.*
3. *For all $h \in [H]$, $\sup_{\theta \in \mathcal{K}} \|\mathbf{u}_{t,h}(\theta)\|_1 \vee \|\mathbf{x}_{t,h}(\theta)\|_1 \le D$.*
4. *For some $\gamma > 0$, it holds for all $h \in [H]$ that $\min_{k \neq k'} \left\|\mathbf{w}_{t,h,k,\widehat{d}} - \mathbf{w}_{t,h,k',\widehat{d}}\right\|_2 \ge \gamma$, where we let $\mathbf{w}_{\widehat{d}}$ denote the first d coordinates of the vector $\mathbf{w} \in \mathbb{R}^{d+1}$.*
5. *For all $\mathbf{v}_{1:H}, \mathbf{v}'_{1:H} \in \mathcal{V}^H$ with $\|\mathbf{v}_{1:H}\|_1 \vee \|\mathbf{v}'_{1:H}\|_1 \le 2D$, we have that the loss functions ℓ_t^v are Lipschitz with respect to the ℓ_1 norm and bounded, i.e., $|\ell_t^v(\mathbf{v}_{1:H}) - \ell_t^v(\mathbf{v}'_{1:H})| \le \|\mathbf{v}_{1:H} - \mathbf{v}_{1:H}\|_1$ and $|\ell_t^v(\mathbf{v}_{1:H})| \le 1$.*

If the planner plays θ_t according to Algorithm 2 with $\eta = d^{1/3} H^{5/3} K^{4/3} \left(\frac{TLD}{\gamma \sigma_{\mathrm{dir}}}\right)^{2/3}$ and $n = \sqrt{\eta}$, then

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \widetilde{\mathcal{O}}\left(d^{1/3} H^{5/3} K^{4/3} \left(\frac{TLd}{\gamma \sigma_{\mathrm{dir}}}\right)^{2/3}\right).$$

Thus the oracle complexity of achieving average regret ε is $\widetilde{\mathcal{O}}\left(d^{1/3} D^{2/3} H^{5/3} K^{4/3} L^{2/3} (\gamma \sigma_{\mathrm{dir}})^{-2/3} \varepsilon^{-2}\right)$.

Remark 38 (Scaling of L and D) *Notice that the scaling of the parameters L and D depend on the H-fold compositions of the dynamic maps $g_{t,k,h}$. Thus in the second assumption in the above theorem, requiring that the maps G_t are L-Lipschitz, can naïvely allow L to scale like $\mathrm{poly}(H) \cdot (\Lambda)^H$ in the worst case when we only assume that the functions $g_{t,k,h}$ are Λ-Lipschitz. Similarly, if we only suppose that $g_{t,k,h}(\mathbf{v}) \le c_2 \|\mathbf{v}\| + c_1$, the bound D in the third and last assumption above could scale with $(c_2)^H$. While these bounds are tight in general, the exponential dependencies can be mitigated with common stability assumptions, often imposed in control settings. For example, under incremental stability of the composed dynamics (Pfrommer et al., 2022; Angeli, 2000), L would scale only polynomially in H. Further notions of input-to-state stability such as those found in Jadbabaie and Hauser (2001) result in polynomially-bounded D. Thus in many practical settings of interest, the parameters L, D scale only polynomially in all of the relevant problem parameters. Note that in the popular Linear Quadratic Regulator framework, of which our setting is a vast generalization, these stability assumptions are standard (Hazan and Singh, 2022).*

As in the previous applications of Theorem 9, we will prove an analogue of Theorem 10 where we introduce a pseudo-metric ρ and prove that it satisfies pseudo-isometry and provides control of the generalized bracketing numbers. We will then conclude by proving that our ℓ is Lipschitz with respect to ρ and appealing to Theorem 9. For the sake of simplicity, we will drop the index t temporarily and compare a given plan $\bar{\mathbf{u}}_{1:H}$ and its associated dynamics $\mathbf{x}_{1:H}$ with an alternative plan

$\bar{\mathbf{u}}'_{1:H}$ and its associated dynamics $\mathbf{x}'_{1:H}$, where these dynamics share the noise sequences $\eta_{1:H}, \xi_{1:H}$. We will abbreviate

$$\rho(\theta, \theta'') = \rho(\theta, \theta', \boldsymbol{\eta}_{1:H}, \boldsymbol{\xi}_{1:H}, \theta) = \left\|\theta - \theta'\right\|_1 + \sum_{h=1}^{H} \left\|\mathbf{x}_h - \mathbf{x}'_h\right\|. \tag{H.1}$$

We will also abbreviate $k_h = k_h(\mathbf{v}_h)$ and $k'_h = k_h(\mathbf{v}'_h)$. We begin by proving the following lemma:

Lemma 39 *Consider the event*

$$\mathcal{A}_h = \left\{k_{h'} = k'_{h'} \text{ for all } h' < h \text{ and } k_h \neq k'_h\right\}.$$

Then for all $h \in [H]$,

$$\left\|\mathbf{x}_h - \mathbf{x}'_h\right\|_1 \leq L \cdot \left\|\theta - \theta'\right\|_1 + 2D \cdot \sum_{h'=1}^{h-1} \mathbb{I}\left[\mathcal{A}_{h'}\right].$$

Proof By (5.1), we have

$$\begin{aligned}
\left\|\mathbf{x}_h - \mathbf{x}'_h\right\|_1 &= \left\|g_{k_{h-1}}(\mathbf{v}_{h-1}) - g_{k'_{h-1}}(\mathbf{v}'_{h-1})\right\|_1 \\
&\leq \left\|g_{k_{h-1}}(\mathbf{v}_{h-1}) - g_{k_{h-1}}(\mathbf{v}'_{h-1})\right\|_1 \cdot \mathbb{I}\left[\bigcup_{h' \geq h} \mathcal{A}_{h'}\right] + 2D \cdot \mathbb{I}\left[\bigcup_{h' < h} \mathcal{A}_{h'}\right] \\
&\leq \left\|G(\theta, k_{1:H}) - G\left(\theta', k'_{1:H}\right)\right\| + 2D \cdot \mathbb{I}\left[\bigcup_{h' < h} \mathcal{A}_{h'}\right] \\
&\leq L \cdot \left\|\bar{\mathbf{u}}_{1:H} - \bar{\mathbf{u}}'_{1:H}\right\|_1 + 2D \sum_{h'=1}^{h-1} \mathbb{I}\left[\mathcal{A}_{h'}\right],
\end{aligned}$$

where the equality is by construction, the first inequality follows from the boundedness of $\mathcal{Z}$, the second inequality follows from the fact that the definition of $\mathcal{A}_{h'}$ for $h' \geq h$ implies that the modes $k_{h''}$ for $h'' < h$ are the same as $k'_{h''}$, and the last inequality follows from a union bound and the second condition in Theorem 37. The result follows. ■

We are now ready to prove the pseudo-isometry property:

Lemma 40 *Let $\mathcal{M}$ denote the class of distributions induced by the setup in Theorem 37 and let ρ be as in (H.1). Then ρ satisfies the pseudo-isometry property with respect to $\|\cdot\|_1$ with $\alpha = \frac{6DH^2K^2L}{\gamma\sigma_{\mathrm{dir}}}$ and $\beta = 1$, i.e.,*

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta')\right] \leq \frac{6DH^2K^2L}{\gamma\sigma_{\mathrm{dir}}} \cdot \left\|\theta - \theta'\right\|_1.$$

Proof By (H.1) and Lemma 39, it holds that

$$\begin{aligned}
\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta')\right] &= \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\left\|\theta - \theta'\right\|_1 + \sum_{h=1}^{H} \left\|\mathbf{x}_h - \mathbf{x}'_h\right\|\right] \\
&\leq 2LH \cdot \left\|\theta - \theta'\right\|_1 + 2DH \cdot \sup_{\nu \in \mathcal{M}} \sum_{h=1}^{H} \mathbb{P}_\nu(\mathcal{A}_h).
\end{aligned}$$

We now compute,

$$\begin{aligned}\mathbb{P}(\mathcal{A}_h) &= \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } k_h \neq k'_h\right) \\ &= \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \arg\max_{k\in[K]} \left\langle \mathbf{w}^\star_{k,h}, (\mathbf{v}_h, 1)\right\rangle \neq \arg\max_{k'\in[K]} \left\langle \mathbf{w}^\star_{k',h}, (\mathbf{v}'_h, 1)\right\rangle\right) \\ &\leq \sum_{k\neq k'\in[K]} \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle \geq 0 > \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}'_h, 1)\right\rangle\right),\end{aligned} \tag{H.2}$$

where the argument is similar to that in (F.1). For fixed $k \neq k' \in [K]$, we then compute

$$\begin{aligned}&\mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle \geq 0 > \left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}'_h, 1)\right\rangle\right) \\ &\quad\leq \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle\right| \leq \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h - \mathbf{v}'_h, 0)\right\rangle\right|\right) \\ &\quad\leq \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle\right| \leq 2 \cdot \left\|\mathbf{v}_h - \mathbf{v}'_h\right\|_1\right),\end{aligned}$$

where the first inequality follows from the triangle inequality and the second from Hölder's inequality. We now observe that

$$\left\|\mathbf{v}_h - \mathbf{v}'_h\right\|_1 \leq \left\|\mathbf{u}_h - \mathbf{u}'_h\right\|_1 + \left\|\mathbf{x}_h - \mathbf{x}'_h\right\|_1,$$

and, furthermore, because the mode sequences $k_{1:h-1} = k'_{1:h-1}$, we have in this event that

$$\left\|\mathbf{x}_h - \mathbf{x}'_h\right\|_1 \leq \left\|G(\theta, k_{1:H}) - G(\theta', k_{1:H})\right\|_1 \leq L \cdot \left\|\theta - \theta'\right\|_1. \tag{H.3}$$

Thus, we have

$$\begin{aligned}&\mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, (\mathbf{v}_h, 1)\right\rangle\right| \leq 2 \cdot \left\|\mathbf{v}_h - \mathbf{v}'_h\right\|_1\right) \\ &\quad\leq \mathbb{P}\left(\left|\left\langle \mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}, \mathbf{v}_h\right\rangle\right| \leq 3L \cdot \left\|\theta - \theta'\right\|_1\right) \\ &\quad= \mathbb{P}\left(\left|\left\langle \frac{\mathbf{w}^\star_{k,h} - \mathbf{w}^\star_{k',h}}{\left\|\mathbf{w}^\star_{k,h,\widehat{d}} - \mathbf{w}^\star_{k',h,\widehat{d}}\right\|_2}, (\mathbf{v}_h, 1)\right\rangle\right| \leq \frac{3L}{\left\|\mathbf{w}^\star_{k,h,\widehat{d}} - \mathbf{w}^\star_{k',h,\widehat{d}}\right\|_2} \cdot \left\|\theta - \theta'\right\|_1\right) \\ &\quad\leq \frac{3L}{\sigma_{\mathrm{dir}} \cdot \left\|\mathbf{w}^\star_{k,h,\widehat{d}} - \mathbf{w}^\star_{k',h,\widehat{d}}\right\|_2} \cdot \left\|\theta - \theta'\right\|_1 \\ &\quad\leq \frac{3L}{\sigma_{\mathrm{dir}}\gamma} \cdot \left\|\theta - \theta'\right\|_1,\end{aligned} \tag{H.4}$$

where the first inequality follows from the preceding computation, the equality is trivial, the second inequality follows from the assmption of directional smoothness, and the last inequality follows from the margin assumption. Plugging back in, the result follows. ■

Finally, we prove that the bracketing numbers can be controlled:

Lemma 41 *Let $\mathcal{M}$ and ρ be as in Lemma 40 and $\mathcal{K}$ be as in Theorem 37. Then, for any $\varepsilon > 0$, it holds that*

$$\mathcal{N}_{\mathcal{M},[]}(\mathcal{K}, \rho, \varepsilon) \leq \left(\frac{36K^2DH^2L}{\gamma\sigma_{\mathrm{dir}}\varepsilon}\right)^{H(d+1)}.$$

Proof Let $\mathcal{N} = \{\theta^i\} \subset \mathcal{U}^{\times H}$ denote an $\widetilde{\varepsilon}$-net with respect to ℓ_1, where $\widetilde{\varepsilon} = \frac{\gamma\sigma_{\mathrm{dir}}}{12DH^2L} \cdot \varepsilon$. As in Lemmas 27 and 35, we will show that if $\mathcal{B}_i$ denotes the set of θ with distance at most $\widetilde{\varepsilon}$ to θ^i, then $\{(\theta^i, \mathcal{B}_i)\}$ forms a generalized ε-bracket with respect to $\mathcal{M}$. The argument is essentially identical after we replace Lemmas 25 and 33 with Lemma 39 and the argument in Lemma 40. In particular, for fixed i and any $\nu \in \mathcal{M}$, we see that

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\rho(\theta,\theta^i)\right] &= \mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\left\|\theta-\theta^i\right\|_1 + \sum_{h=1}^H\left\|\mathbf{x}_h-\mathbf{x}_h^i\right\|_1\right]\\ &\le \widetilde{\varepsilon} + \mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\sum_{h=1}^H\left\|\mathbf{x}_h-\mathbf{x}_h^i\right\|_1\right],\end{aligned}$$

where we let $\mathbf{x}^i_{1:H}$ denote the dynamics evolved with θ^i. For the second term, we invoke Lemma 39 and compute:

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\sum_{h=1}^H\left\|\mathbf{x}_h-\mathbf{x}_h^i\right\|_1\right] &\le \sum_{h=1}^H\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\left\|\mathbf{x}_h-\mathbf{x}_h^i\right\|_1\right]\\ &\le \sum_{h=1}^H\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}L\cdot\left\|\theta-\theta^i\right\| + 2D\cdot\sum_{h'=1}^{h-1}\mathbb{I}\left[\mathcal{A}_{h'}\right]\right]\\ &\le LH\widetilde{\varepsilon} + 2DH^2\cdot\max_{h\in[H]}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\mathbb{I}\left[\mathcal{A}_h\right]\right].\end{aligned}$$

Now we reason in a similar manner as in Lemma 40:

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta\in\mathcal{B}_i}\mathbb{I}\left[\mathcal{A}_h\right]\right] &= \mathbb{P}\left(\bigcup_{\bar{\mathbf{u}}_{1:H}\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'}\text{ for }h'<h\text{ and }k_h\neq k^i_h\right\}\right)\\ &\le \sum_{k\neq k'\in[K]}\mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'}\text{ for }h'<h\text{ and }\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 2\cdot\left\|\mathbf{v}_h-\mathbf{v}^i_h\right\|_1\right\}\right),\end{aligned}$$

where the inequality follows from the same chain of logic as in (H.2) and (H.2). As in (H.3), we observe that if the mode sequence $k_{h'} = k^i_{h'}$ for $h' < h$, then

$$\left\|\mathbf{v}_h-\mathbf{v}^i_h\right\|_1 \le 2L\cdot\left\|\theta-\theta^i\right\|_1.$$

Thus, by construction of $\mathcal{B}_i$, we see that

$$\begin{aligned}&\mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'}\text{ for }h'<h\text{ and }\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 2\cdot\left\|\mathbf{v}_h-\mathbf{v}^i_h\right\|_1\right\}\right)\\ &\le \mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 6L\cdot\left\|\theta-\theta^i\right\|_1\right\}\right)\\ &\le \mathbb{P}\left(\left|\left\langle\mathbf{w}^\star_{k,h}-\mathbf{w}^\star_{k',h},(\mathbf{v}^i_h,1)\right\rangle\right|\le 6L\widetilde{\varepsilon}\right)\\ &\le \frac{6L\widetilde{\varepsilon}}{\gamma\sigma_{\mathrm{dir}}},\end{aligned}$$

where the last inequality follows from directional smoothness and the margin assumption, as in (H.4). Plugging back in to the definition of $\widetilde{\varepsilon}$, we see that $\left\{(\theta^i, \mathcal{B}_i)\right\}$ forms a generalized ε-bracket as desired. Note that $\mathcal{K}$ lives in the ℓ_1 ball of radius D inside of $\mathbb{R}^{dH}$; thus, applying the same argument as in Theorem 10 concludes the proof. ■

Finally, we are ready to prove the main result:

Proof [Proof of Theorem 37] By Theorem 9, it suffices to show that the loss is Lipschitz with respect to ρ. To do this, note that

$$\begin{aligned}\ell(\mathbf{z}_t) - \ell(\mathbf{z}'_t) &\le \left\|\mathbf{v}_{t,1:H} - \mathbf{v}'_{t,1:H}\right\|_1 \\ &\le \left\|\theta - \theta'\right\|_1 + \sum_{h=1}^{H} \left\|\mathbf{x}_{t,h} - \mathbf{x}'_{t,h}\right\|_1 \\ &= \rho(\theta, \theta').\end{aligned}$$

Thus by Theorem 9, it holds that if the learner plays Algorithm 2 with $n = \sqrt{\eta}$, then he experiences

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \le \widetilde{\mathcal{O}}\left(\eta + \frac{TDH^2K^2L}{\gamma\sigma_{\mathrm{dir}}} \cdot \sqrt{\frac{Hd}{\eta}}\right)$$

by appealing to Lemmas 40 and 41. Setting η as in the statement of the theorem concludes the proof. ■

Finally, we prove an analogue of Theorem 37 where we now assume that the decision boundaries between modes are polynomials. The statement is almost equivalent to that of Theorem 37, with the exception that the boundaries between modes are now parameterized by polynomials, with the resulting increased oracle complexity along the lines of Theorem 13. The statement is as follows:

Theorem 42 *Suppose that we are in the situation of (5.1), with the exception that the regions are defined by polynomials of degree at most r. More precisely, we suppose that*

$$\begin{aligned}\mathbf{x}_{t,h+1}(\theta) &= g_{t,h,k_{t,h}(\mathbf{v}_{t,h}(\theta))}(\mathbf{v}_{t,h}(\theta)) + \boldsymbol{\eta}_{t,h}, \quad \textit{and} \\ \mathbf{u}_{t,h}(\theta) &= \bar{\mathbf{u}}_{t,h} + \boldsymbol{\xi}_{t,h}, \quad \mathbf{v}_{t,h}(\theta) = (\mathbf{x}_{t,h}(\theta), \quad \mathbf{u}_{t,h}(\theta)), \\ k_{t,h}(\mathbf{v}) &= \arg\max_{k\in[K]} \phi_{t,h}(k, \mathbf{v}), \quad \textit{and} \quad \phi_{t,h}(k, \mathbf{v}) = f_{\mathbf{w}_{t,k,h}}(\mathbf{v}),\end{aligned}$$

where the $f_{\mathbf{w}_{t,k,h}}$ are degree r polynomials with $\mathbf{w}_{t,k,h}$ parameterizing the coefficients. Suppose that for all $t \in [T]$, the following properties hold almost surely under the adversary's strategy p_t:

1. *$\mathbf{x}_{t,1} \mid \mathcal{F}_t$ and $(\boldsymbol{\eta}_{t,h}, \boldsymbol{\xi}_{t,h}) \mid \mathcal{F}_{t,h-1}$ are $\sigma_{\mathrm{poly},r}$-polynomially smooth.*
2. *For all mode sequences $k_{1:H} \in [K]^H$ and $\theta, \theta' \in \mathcal{U}^{\times H}$, $\|G_t(\theta, k_{1:H}) - G_t(\theta', k_{1:H})\| \le L\left\|\theta - \theta'\right\|_1$, i.e., the functions G_t are L-Lipschitz with respect to the ℓ_1 norm, where the G_t are the maps defined in Definition 36.*
3. *For all $h \in [H]$, $\sup_{\theta\in\mathcal{K}} \left\|\mathbf{u}_{t,h}(\theta)\right\|_1 \vee \left\|\mathbf{x}_{t,h}(\theta)\right\|_1 \le D$.*

4. *For some $\gamma > 0$, it holds for all $h \in [H]$ that $\min_{k \neq k' \in [K]} \mathrm{coeff}_r\left(f_{\mathbf{w}_{t,k,h}} - f_{\mathbf{w}_{t,k',h}}\right) \geq \gamma$, where $\mathrm{coeff}_r(\cdot)$ is as defined in Definition 12.*

5. *For all $\mathbf{v}_{1:H}, \mathbf{v}'_{1:H} \in \mathcal{V}^H$ with $\|\mathbf{v}_{1:H}\|_1 \vee \|\mathbf{v}'_{1:H}\|_1 \leq 2D$, we have that the loss functions ℓ_t^v are Lipschitz with respect to the ℓ_1 norm and bounded, i.e., $|\ell_t^v(\mathbf{v}_{1:H}) - \ell_t^v(\mathbf{v}'_{1:H})| \leq \|\mathbf{v}_{1:H} - \mathbf{v}_{1:H}\|_1$ and $|\ell_t^v(\mathbf{v}_{1:H})| \leq 1$.*

6. *The coefficients $\mathbf{w}_{t,k,h}$ of the polynomials $f_{\mathbf{w}_{t,k,h}}$ have unit Euclidean norm.*

If the planner plays θ_t according to Algorithm 2 with $\eta = \left(\frac{LTK^2r^2H^{2+r}d^rDB}{\gamma\sigma_{\mathrm{poly},r}}\right)^{\frac{4r-2}{4r-1}}$ and $n = \sqrt{\eta}$, then

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq \widetilde{\mathcal{O}}\left(\left(\frac{LTK^2r^2H^{2+r}d^rDB}{\gamma\sigma_{\mathrm{poly},r}}\right)^{\frac{4r-2}{4r-1}}\right).$$

Thus, the oracle complexity of achieving average regret ε is $\widetilde{\mathcal{O}}\left(\left(\frac{LK^2r^2H^{2+r}d^rDB}{\gamma\sigma_{\mathrm{poly},r}}\right)^{\frac{4r-2}{4r-1}}\right) \cdot \varepsilon^{-\frac{2}{r}}$.

Proof The proof will be similar to that of Theorem 37. In particular, we will still use the same ρ as in (H.1) and preserve notation from that proof. Applying Theorem 9, it suffices to control the pseudo-isometry and generalized bracketing numbers. We first claim that if $\mathcal{M}$ is the class of distributions for the adversary, induced by the setting at hand, then

$$\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta')\right] \leq \frac{4D^2H^2K^2L}{\gamma^{\frac{1}{r}} \cdot \sigma_{\mathrm{poly},r}} \cdot \left\|\theta - \theta'\right\|_1^{\frac{1}{r}}. \tag{H.5}$$

To see this, we observe that if $\mathcal{A}_h$ is as in Lemma 39, then by that same result,

$$\begin{aligned}\sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\rho(\theta, \theta')\right] &= \sup_{\nu \in \mathcal{M}} \mathbb{E}_\nu\left[\left\|\theta - \theta'\right\| + \sum_{h=1}^H \left\|\mathbf{x}_h - \mathbf{x}'_h\right\|\right] \\ &\leq 2LH \cdot \left\|\theta - \theta'\right\|_1 + 2DH \cdot \sup_{\nu \in \mathcal{M}} \sum_{h=1}^H \mathbb{P}_\nu\left(\mathcal{A}_h\right).\end{aligned}$$

We now compute, as in Lemma 40,

$$\begin{aligned}\mathbb{P}\left(\mathcal{A}_h\right) &= \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } k_h \neq k'_h\right) \\ &\leq \sum_{k \neq k' \in [K]} \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } f_{\mathbf{w}_{k,h}}(\mathbf{v}_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h) \geq 0 > f_{\mathbf{w}_{k,h}}(\mathbf{v}'_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}'_h)\right).\end{aligned}$$

For fixed $k \neq k'$, we then compute

$$\begin{aligned}&\mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } f_{\mathbf{w}_{k,h}}(\mathbf{v}_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h) \geq 0 > f_{\mathbf{w}_{k,h}}(\mathbf{v}'_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}'_h)\right) \\ &\leq \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right| \leq \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}'_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}'_h)\right| + \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right|\right) \\ &\leq \mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right| \leq 2D^r \cdot \left\|\mathbf{v}_h - \mathbf{v}'_h\right\|_1\right),\end{aligned}$$

where the second inequality follows from the assumption that $\mathbf{w}_{k,h}$ have unit norm, the fact that power functions area locally Lipschitz, and the fact that $\mathbf{v}_h$ has norm bounded by D. Applying the assumption of polynomial smoothness, we see that this last expression is bounded above:

$$\mathbb{P}\left(k_{h'} = k'_{h'} \text{ for } h' < h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}_h)\right| \leq 2D^r \cdot \left\|\mathbf{v}_h - \mathbf{v}'_h\right\|_1\right) \leq \frac{2D}{\gamma^{\frac{1}{r}} \cdot \sigma_{\text{poly},r}} \left\|\mathbf{v}_h - \mathbf{v}_{h'}\right\|_1^{\frac{1}{r}}.$$

We now conclude in a similar manner as in Lemma 40 and observe that

$$\left\|\mathbf{v}_h - \mathbf{v}'_h\right\|_1^{\frac{1}{r}} \leq 2\left\|\mathbf{u}_h - \mathbf{u}'_h\right\|_1^{\frac{1}{r}} + 2\left\|\mathbf{x}_h - \mathbf{x}'_h\right\|_1^{\frac{1}{r}}$$

and because the mode sequences $k_{1:h-1} = k'_{1:h-1}$ on this event,

$$\left\|\mathbf{x}_h - \mathbf{x}'_h\right\|_1^{\frac{1}{r}} \leq L^{\frac{1}{r}} \cdot \left\|\theta - \theta'\right\|_1^{\frac{1}{r}}.$$

Putting everything together suffices to prove that (H.5) holds.

We now claim that for $0 < \varepsilon < 1$,

$$\mathcal{N}_{\mathcal{M},[]}(\Theta, \mathcal{M}, \varepsilon) \leq \left(\frac{12D^2H^2K^2L}{\sigma_{\text{poly},r}\gamma^{\frac{1}{r}}}\varepsilon\right)^{K^2r^2(Hd)^r}. \tag{H.6}$$

To see this, we let $\mathcal{N} = \left\{\theta^i\right\} \subset \mathcal{U}^{\times H}$ denote an $\widetilde{\varepsilon}$-net with respect to ℓ_1, where $\widetilde{\varepsilon} = \gamma\left(\frac{\sigma_{\text{poly},r}}{4D^2H^2K^2L} \cdot \varepsilon\right)^r$. Our proof proceeds similarly to that in Lemma 41 and we demonstrate that if $\mathcal{B}_i$ denotes the set of $\widetilde{\varepsilon}$-balls around θ^i then the associated $(\theta^i, \mathcal{B}_i)$ forms a generalized ε-bracket with respect to $\mathcal{M}$. Indeed, for fixed i and $\nu \in \mathcal{M}$, we have:

$$\mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} \rho(\theta, \theta^i)\right] \leq \widetilde{\varepsilon} + \mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} \sum_{h=1}^H \left\|\mathbf{x}_h - \mathbf{x}^i_h\right\|_1\right].$$

For the second term, we again invoke Lemma 39 and observe that

$$\begin{aligned}\mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} \sum_{h=1}^H \left\|\mathbf{x}_h - \mathbf{x}^i_h\right\|_1\right] &\leq \sum_{h=1}^H \mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} L \cdot \left\|\theta - \theta^i\right\|_1 + 2D \cdot \sum_{h'=1}^{h-1} \mathbb{I}\left[\mathcal{A}_{h'}\right]\right] \\ &\leq LH\widetilde{\varepsilon} + 2DH^2 \cdot \max_{h \in [H]} \mathbb{E}_\nu\left[\sup_{\theta \in \mathcal{B}_i} \mathbb{I}\left[\mathcal{A}_h\right]\right].\end{aligned}$$

Using the identical logic combining as in Lemma 41, but using the polynomial smoothness assumptions in the same way as in Lemma 35, we see that

$$\mathbb{P}_\nu\left(\mathcal{A}_h\right) \leq \sum_{k \neq k' \in [K]} \mathbb{P}\left(\bigcup_{\theta \in \mathcal{B}_i}\left\{k_{h'} = k^i_{h'} \text{ for } h' < h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}^i_h) - f_{\mathbf{w}_{k',h}}(\mathbf{v}^i_h)\right| \leq 2D^r \cdot \left\|\mathbf{v}_h - \mathbf{v}^i_h\right\|_1\right\}\right).$$

Observing once again that

$$\left\|\mathbf{v}_h - \mathbf{v}^i_h\right\|_1 \leq 2L \cdot \left\|\theta - \theta^i\right\|$$

by the Lipschitzness of the G_t, we see that

$$\begin{aligned}\mathbb{P}_\nu(\mathcal{A}_h) &\le \sum_{k\neq k'\in[K]} \mathbb{P}\left(\bigcup_{\theta\in\mathcal{B}_i}\left\{k_{h'}=k^i_{h'} \text{ for } h'<h \text{ and } \left|f_{\mathbf{w}_{k,h}}(\mathbf{v}^i_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}^i_h)\right|\le 4D^rL\widetilde{\varepsilon}\right\}\right)\\ &\le \sum_{k\neq k'\in[K]} \mathbb{P}\left(\left|f_{\mathbf{w}_{k,h}}(\mathbf{v}^i_h)-f_{\mathbf{w}_{k',h}}(\mathbf{v}^i_h)\right|\le 4D^rL\widetilde{\varepsilon}\right)\\ &\le K^2\frac{4DL^{\frac{1}{r}}}{\gamma^{\frac{1}{r}}\cdot\sigma_{\mathrm{poly},r}}\cdot\widetilde{\varepsilon}^{\frac{1}{r}}.\end{aligned}$$

Plugging everything back in to the above work, we see that indeed $\mathcal{N}$ is a generalized ε-net. Using the same volume argument as in the proofs of Lemmas 27, 35, and 41, we see that (H.6) holds.

Finally, we note that ℓ^v is clearly lipschitz with respect to ρ and thus we may apply Theorem 9, which tells us that if we run Algorithm 2, then for $n=\sqrt{\eta}$,

$$\mathbb{E}\left[\mathrm{Reg}_T\right]\le\widetilde{\mathcal{O}}\left(\eta+\frac{T}{n}K^2r^2H^rd^r\log\left(\frac{1}{\sigma_{\mathrm{poly},r}}\right)+\frac{TD}{\gamma^{\frac{1}{r}}\sigma_{\mathrm{poly},r}}\cdot\left(\frac{K^2r^2H^rd^r}{\eta}\right)^{\frac{1}{4r-2}}\right).$$

Setting η as in the statement then concludes the proof. ■