Oracle-Efficient Hybrid Online Learning with Unknown Distribution

Changlong Wu WUCHANGL@HAWAII.EDU
CSoI, Purdue University

Jin Sima JSIMA@ILLINOIS.EDU
ECE Department, UIUC

Wojciech Szpankowski SZPAN@PURDUE.EDU
CSoI, Purdue University

Editors: Shipra Agrawal and Aaron Roth

Abstract

We study the problem of *oracle-efficient* hybrid online learning when the features are generated by an *unknown* i.i.d. process and the labels are generated adversarially. Assuming access to an (offline) ERM oracle, we show that there exists a computationally efficient online predictor that achieves a regret upper bounded by $\tilde{O}(T^{\frac{3}{4}})$ for a finite-VC class, and upper bounded by $\tilde{O}(T^{\frac{p+1}{p+2}})$ for a class with α fat-shattering dimension α^{-p}. This provides the first known *oracle-efficient* sublinear regret bounds for hybrid online learning with an *unknown* feature generation process. In particular, it confirms a conjecture of Lazaric and Munos (2012). We then extend our result to the scenario of shifting distributions with K changes, yielding a regret of order $\tilde{O}(T^{\frac{4}{5}}K^{\frac{1}{5}})$. Finally, we establish a regret of $\tilde{O}((K^{\frac{2}{3}}(\log|\mathcal{H}|)^{\frac{1}{3}}+K)\cdot T^{\frac{4}{5}})$ for the contextual K-armed bandits with a finite policy set $\mathcal{H}$, i.i.d. generated contexts from an *unknown* distribution, and adversarially generated costs.
Keywords: Hybrid online learning, ERM oracle, oracle-efficiency, relaxation, random playout

1. Introduction

We study the problem of *hybrid* stochastic-adversary online learning, where the features are assumed to be sampled from an *unknown* stochastic source while the labels are selected adversarially. Recent advancements Lazaric and Munos (2009); Rakhlin et al. (2012); Haghtalab et al. (2020, 2022a,b); Block et al. (2022); Wu et al. (2023a,b) have demonstrated that such hybrid settings provide a fundamental paradigm shift beyond the classical *worst-case* adversarial online setting to accommodate broader stochastic scenarios, while still preserving the capacity to handle various adversarial situations and maintain minimal assumptions on the expert class.

We are interested in *oracle-efficient* regret minimization methods as in Kakade and Kalai (2005). Here, we assume that the learner has access to an Empirical Risk Minimization (ERM) optimization oracle which, given any sequence of feature-label pairs, identifies the expert within the class that achieves the minimal cumulative loss. This effectively reduces the online learning problem to a batch learning problem, for which algorithms such as gradient descent have been highly successful in computing ERM optimization, even in complex classes like neural networks. Previous studies have applied this methodology in various online learning scenarios, including transductive online learning as in Kakade and Kalai (2005), online learning with *smooth* adversary samples as in Rakhlin et al. (2012); Haghtalab et al. (2022a); Block et al. (2022), and contextual bandits with a *known* i.i.d. feature generation distribution as in Rakhlin and Sridharan (2016); Syrgkanis et al. (2016); Banihashem et al. (2023). However, all of these works have assumed some form of access

to a sampling oracle for the feature generation process. This may not be realistic when feature generation is costly, such as in medical data, or when the underlying probability law is unknown a priori. Other studies, like Lazaric and Munos (2009); Wu et al. (2023a,b), do address scenarios with *unknown* distributions, but provide only computationally inefficient prediction rules.

This paper initiates the study of *oracle-efficient* hybrid online learning without assuming any access to the underlying probability law of the feature generation process. For the clarity of presentation, we will mainly focus on scenarios of online learning where features are generated by an *unknown* i.i.d. process. However, we will also consider extensions to other scenarios, such as shifting distributions and contextual multi-armed bandits. Our approach also provides a general methodology that concentrates on the *feature efficiency* in online learning.

Problem formulation. Let $\mathcal{X}$ be an instance (feature) space and $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ be a function class mapping $\mathcal{X} \to [0,1]$. We consider the following *hybrid* online learning scenario. Nature selects an (unknown) distribution μ over $\mathcal{X}$ at the start of the game. At each time step t, Nature independently samples $\mathbf{x}_t \sim \mu$ and selects *adversarially* a $y_t \in [0,1]$, but reveals only $\mathbf{x}_t$. A predictor then (randomly) generates $\hat{y}_t \in [0,1]$ based on $\mathbf{x}^t, y^{t-1}$, where $\mathbf{x}^t = \{\mathbf{x}_1, \cdots, \mathbf{x}_t\}$ and $y^{t-1} = \{y_1, \cdots, y_{t-1}\}$. Nature then reveals y_t, and the predictor incurs a loss $\ell(\hat{y}_t, y_t)$, for a predefined loss function $\ell : [0,1]^2 \to \mathbb{R}^+$. Here, we assume that the loss ℓ is *convex* in its first argument and L-Lipschitz in *both* arguments, e.g., the *absolute loss* $\ell(\hat{y}, y) = |\hat{y} - y|$. A prediction rule is a function Φ that takes inputs from $(\mathcal{X} \times [0,1])^* \times \mathcal{X}$ and generates a *distribution* over $[0,1]$. For any prediction rule Φ and function class $\mathcal{H}$, we define the hybrid minimax regret as:

$$\tilde{r}_T(\mathcal{H}, \Phi) = \sup_{\mu} \mathbb{E}_{\mathbf{x}_1} \sup_{y_1 \in [0,1]} \mathbb{E}_{\hat{y}_1} \cdots \mathbb{E}_{\mathbf{x}_T} \sup_{y_T \in [0,1]} \mathbb{E}_{\hat{y}_T} \left[\sum_{t=1}^{T} \ell(\hat{y}_t, y_t) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{T} \ell(h(\mathbf{x}_t), y_t) \right], \quad (1)$$

where $\mathbf{x}_t \sim \mu$ and $\hat{y}_t \sim \Phi(\mathbf{x}^t, y^{t-1})$ for $t \in [T]$. Our goal is to find an *oracle-efficient* prediction rule Φ that minimizes $\tilde{r}_T(\mathcal{H}, \Phi)$.

1.1. Results and Techniques

In this work, we provide the first known oracle-efficient sub-linear regret bounds for the hybrid online learning with *unknown* feature generation distributions.

Theorem 1 (Informal) *Let $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ be a class with Rademacher complexity $O(T^q)$. Then, there exists an oracle-efficient predictor with at most $O(\sqrt{T}\log T)$ calls to the ERM oracle that achieves the hybrid minimax regret of order $\tilde{O}(T^{\frac{2-q}{3-2q}})$. In particular, for a VC-class, this implies a regret of $\tilde{O}(T^{\frac{3}{4}})$, and for classes with α-fat shattering dimension α^{-p}, it results in a regret of $\tilde{O}(T^{\max\{\frac{3}{4}, \frac{p+1}{p+2}\}})$, where $\tilde{O}$ hides poly-log factors.*

To the best of our knowledge, the regret bounds presented in Theorem 1 are the first known *oracle-efficient* sub-linear regrets for hybrid online learning with *unknown* feature generation distributions and generic (non-parametric) hypothesis classes. In particular, our $\tilde{O}(T^{\frac{3}{4}})$ bound confirms a conjecture of Lazaric and Munos (2012) regarding the oracle-efficient regret bounds for finite-VC classes under absolute loss [1]. Note that, it was demonstrated by (Block et al., 2022, Theorem 7) that for *known* feature generation distributions, an $\tilde{O}(T^{\max\{\frac{1}{2}, \frac{p-1}{p}\}})$ regret bound is achievable for a class

1. They also obtained an $O(\sqrt{T\log T})$ bound using a computationally *inefficient* covering-based approach.

with an α-fat shattering dimension of order α^{-p}. However, to the best of our knowledge, such a chaining-based bound was not known for the *unknown* distribution case, even in the information-theoretical sense. The closest comparison is the regret bound obtained in Wu et al. (2023b) that *matches* our $\tilde{O}(T^{\frac{p+1}{p+2}})$ bound, but established via an (inefficient) one-step covering approach.

At a high level, our oracle-efficient prediction rule is based on the *relaxation* and *random play-out* techniques, as introduced in Rakhlin et al. (2012). However, a distinguishing feature of our setup is that we are *not* able to access the sampling oracle of the underlying feature generating process. Our main idea is to employ an *epoch-based* approach. We partition the time horizon into a set of carefully *designed* epochs. At each epoch, we *estimate* the underlying distribution μ by $\hat{\mu}$ using samples observed in prior epochs. We then use the estimated distribution $\hat{\mu}$ to generate the *hallucinated samples* as needed in the relaxation framework for the current epoch. Observe, however, that the estimation $\hat{\mu}$ can arbitrarily deviate from μ under total variation, as we do not make any structural assumption on μ, and the adversary *knows* $\hat{\mu}$ when generating adversary samples. Therefore, the *randomness matching* argument, as in Rakhlin et al. (2012); Block et al. (2022), will not work. To overcome this issue, we introduce a *surrogate* relaxation based on $\hat{\mu}$ and relate it to the regret via a novel concept of *approx-admissibility*, which is further controlled by a novel *symmetrization* argument. The regret will then follow by carefully designing the epochs to balance the error introduced by the *approx-admissibility* and the Rademacher complexity of the class restricted to each epoch.

Tighter bounds for special cases. Going beyond the general result in Theorem 1, we show in Theorem 23 (Appendix F) that an oracle-efficient $\tilde{O}(T^{\max\{\frac{1}{2},\frac{p-1}{p}\}})$ regret is achievable for finite fat-shattering classes of order p if we assume a weaker *oblivious* adversary. This is tight upto poly-logarithmic factors. Furthermore, for the class $\mathcal{H}$ of all Lipschitz functions $[0,1]^d \to [0,1]$ we establish in Theorem 15 the (optimal) regret $\tilde{O}(T^{\max\{\frac{1}{2},\frac{d-1}{d}\}})$ against the *adaptive* adversary.

Shifting distributions. Our next result drops the i.i.d. assumption and allows distributions to change over time. We assume that the number of changes is upper bounded by K and that the possible distributions and change points are completely unconstrained and unknown a priori. We show that the *oracle-efficient* regret is upper bounded by $\tilde{O}(T^{\frac{4}{5}}K^{\frac{1}{5}})$ for a finite VC class. Note that an $O(\sqrt{KT\log T})$ bound was demonstrated by Wu et al. (2023a). However, their algorithm relies on constructing an exponentially large covering, thus being computationally inefficient.

Contextual K-arm Bandits. Finally, we establish an $O((K^{\frac{2}{3}}(\log|\mathcal{H}|)^{\frac{1}{3}} + K\sqrt{\log K}) \cdot T^{\frac{4}{5}})$ *oracle-efficient* regret bound for the contextual K-armed bandits with a finite policy set $\mathcal{H}$, where the contexts are generated by an *unknown* i.i.d. process and the costs are selected adversarially. The closest comparison to this result is the $O(T^{\frac{2}{3}}(K\log|\mathcal{H}|)^{\frac{1}{3}})$ bound established in Banihashem et al. (2023), only for the *known* i.i.d. distribution case. Notably, our result answers positively a question of Banihashem et al. (2023) regarding relaxing the sampling access to the context distribution.

2. Notation and Preliminaries

Oracle-Efficient Predictors. We adopt the following *mixed*-ERM oracle from Block et al. (2022). Let $(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_m, y_m) \in \mathcal{X} \times [0,1]$, $C \in \mathbb{R}^+$, $\epsilon^n \in \{-1,+1\}^n$, and $\tilde{\mathbf{x}}^n \in \mathcal{X}^n$. The mixed ERM oracle is the task of finding $\inf_{h\in\mathcal{H}} \left\{\sum_{i=1}^m \ell(h(\mathbf{x}_i), y_i) + C\sum_{j=1}^n \epsilon_j h(\tilde{\mathbf{x}}_j)\right\}$. Note that the

unintuitive parts $\epsilon h(\tilde{\mathbf{x}})$ can be interpreted as an *absolute* loss since $\epsilon h(\tilde{\mathbf{x}}) = |h(\tilde{\mathbf{x}}) - \frac{(1-\epsilon)}{2}| - \frac{1-\epsilon}{2}$ for $\epsilon \in \{+1, -1\}$. Therefore, the mixed ERM oracle is reduced to a regular (weighted) ERM oracle if ℓ is the absolute loss. Moreover, the weight C can be understood as repeating the same sample C times (rounding to an integer if necessary). We say a predictor Φ is *oracle-efficient* if the running time of computing $\hat{y}_t \sim \Phi(\mathbf{x}^t, y^{t-1})$ is polynomial with respect to t by accessing a mixed ERM oracle (with each oracle call treated as unit time) for any $\mathbf{x}^t, y^{t-1}$.

Adaptive v.s. Oblivious. Note that our formulation in (1) assumes that the generation of y_ts is *adaptive*, since the selection of y_t at each time step depends on all prior information $\mathbf{x}^t$, y^{t-1}, and $\hat{y}^{t-1}$. For comparison, we also introduce a weaker notion of adversary, namely, the *oblivious* adversary, which selects the y_ts based only on the current instance $\mathbf{x}_t$ (see Theorem 23 in Appendix F). It turns out that the adaptive nature of the adversary constitutes the main obstacle in our analysis.

Hybrid Contextual Bandits. We now formulate the contextual K-arm bandits within our framework. Let $\mathcal{D}_K$ be the set of all probability distributions over $[K]$. A policy set $\mathcal{H}$ is a class of functions $\mathcal{X} \to [K]$. We consider the following bandit setup: Nature selects some μ at the start of the game. At each time step t, Nature samples $\mathbf{x}_t \sim \mu$ and selects *adversarially* a cost vector $c_t \in [0,1]^K$, but reveals only $\mathbf{x}_t$. A predictor then selects a distribution $q_t \in \mathcal{D}_K$ based on the history observed thus far and samples $\hat{y}_t \sim q_t$. Nature reveals only $c_t[\hat{y}_t]$ which is also the predictor incurred loss. The goal is to find an *oracle-efficient* prediction rule $\Phi : (\mathcal{X} \times [0,1])^* \times \mathcal{X} \to \mathcal{D}_K$ that minimizes: $\tilde{r}_T^{\mathsf{bandit}}(\mathcal{H}, \Phi) = \sup_{\mu} \mathbb{E}_{\mathbf{x}_1} \sup_{c_1} \cdots \mathbb{E}_{\mathbf{x}_T} \sup_{c_T} \mathbb{E}_{\hat{y}^T} \left[\sum_{t=1}^{T} \langle q_t, c_t \rangle - \inf_{h \in \mathcal{H}} \sum_{t=1}^{T} c_t[h(\mathbf{x}_t)] \right]$, where $q_t = \Phi(\mathbf{x}^t, c_1[\hat{y}_1], \cdots, c_{t-1}[\hat{y}_{t-1}])$. Here, the ERM oracle is to find $\inf_{h \in \mathcal{H}} \sum_{i=1}^{m} \hat{c}_i[h(\mathbf{x}_i)]$ for any $(\mathbf{x}_1, \hat{c}_1), \cdots, (\mathbf{x}_m, \hat{c}_m) \in \mathcal{X} \times \mathbb{R}^K$, as Rakhlin and Sridharan (2016).

3. Oracle-Efficient Regret Bounds for Online Learning

In this section, we focus on bounding the hybrid minimax regret for online learning as in (1) with a generic hypothesis class $\mathcal{H}$, using *oracle-efficient* predictors by accessing to an mixed ERM oracle. We first recall the following standard notion of Rademacher complexity:

Definition 2 *Let $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ be a function class and $T \in \mathbb{N}^+$. The Rademacher complexity of $\mathcal{H}$ at horizon T is defined to be $\mathsf{Rad}_T(\mathcal{H}) = \sup_{\mathbf{x}^T \in \mathcal{X}^T} \mathbb{E}_{\epsilon^T} \left[\sup_{h \in \mathcal{H}} \sum_{t=1}^{T} \epsilon_t h(\mathbf{x}_t) \right]$, where ϵ_t is i.i.d. sampled from the uniform distribution over $\{\pm 1\}$.*

We are now ready to state our first main result:

Theorem 3 *Let $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ be a class with $\mathsf{Rad}_T(\mathcal{H}) \le O(T^q)$ for some $q \in [\frac{1}{2}, 1]$, and let ℓ be a L-Lipschitz loss that is convex in its first argument. Then there exists an* oracle-efficient *prediction rule Φ with at most $O(L\sqrt{T}\log T)$ calls to the ERM oracle per round, such that*

$$\tilde{r}_T(\mathcal{H}, \Phi) \le O\left(L\sqrt{\log(LT)} \cdot T^{\frac{2-q}{3-2q}} \right).$$

In particular, for a binary-valued class with finite VC-dimension, we have $\tilde{r}_T(\mathcal{H}, \Phi) \le O(L\sqrt{\mathsf{VC}(\mathcal{H})\log(LT)} \cdot T^{\frac{3}{4}})$, and for a real-valued class $\mathcal{H}$ with an α-fat shattering dimension of order α^{-p} for $p > 0$ (Alon et al., 1997), we have $\tilde{r}_T(\mathcal{H}, \Phi) \le \tilde{O}(LT^{\max\{\frac{3}{4}, \frac{p+1}{p+2}\}})$.

The rest of this section is devoted to establishing Theorem 3. At a high level, we will partition the time horizon into a set of epochs and make predictions at each epoch using features in prior epochs as *side-information*. Our main proof technique is an epoch-based predictor introduced in Section 3.2, together with a novel *approx-admissibility* framework for handling prediction with side-information, as developed in Section 3.1.

3.1. Regret Analysis with Side-Information

We first consider a *hypothetical* scenario where we assume the predictor has access to some *side-information* $\mathbf{x}^0_{-N+1}$ sampled $i.i.d.$ from the same distribution μ. It is crucial to note that this information is *known* to the adversarial as well, i.e., the adversary's strategy could also depend on $\mathbf{x}^0_{-N+1}$, which turns out to be the main obstacle in our analysis.

Formally, we consider the following learning game proceeds over a horizon of length M: At the start of the game, Nature selects an unknown distribution μ over $\mathcal{X}$, samples an $i.i.d.$ sample $\mathbf{x}^0_{-N+1}$ of size N from μ and reveals $\mathbf{x}^0_{-N+1}$ to a predictor; At each time step $j \in [M]$, Nature samples $\mathbf{x}_j \sim \mu$ and selects *adversarially* $y_j \in [0,1]$ (depends on $\mathbf{x}^j_{-N+1}$ and $\hat{y}^{t-1}$) but reveals only $\mathbf{x}_j$; The predictor then (randomly) generates $\hat{y}_j \in [0,1]$ based on $\mathbf{x}^j_{-N+1}$ and y^{j-1}; Nature reveals y_j and the predictor incurs loss $\ell(\hat{y}_j, y_j)$, for some predefined convex and L-Lipschitz loss.

Predictor via *surrogate* relaxation. Let $\hat{\mu}_N$ be the *empirical* distribution $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\mathbf{x}_{-N+i}}$, where $\delta_{\mathbf{x}}$ is the Dirac measure on $\mathbf{x}$. For any time step $j \in [M]$ and horizon M satisfying $M \le N/2$, we construct the following *randomized* prediction rule:

1. Sample (internally) the *hallucinated samples* $\tilde{\mathbf{x}}_{j+1}, \cdots, \tilde{\mathbf{x}}_M$ from $\hat{\mu}_N$ *without replacement* [2] and $\epsilon_{j+1}, \cdots, \epsilon_M$ $i.i.d.$ from the uniform distribution over $\{-1, +1\}$;

2. Make prediction

$$\hat{y}_j = \operatorname*{arg\,min}_{\hat{y}\in[0,1]} \sup_{y\in[0,1]} \left\{ \ell(\hat{y}, y) + \sup_{h\in\mathcal{H}} \left[2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - \ell(h(\mathbf{x}_j), y) - \sum_{i=1}^{j-1} \ell(h(\mathbf{x}_i), y_i) \right] \right\}. \tag{2}$$

Note that the main difference from the classical *random play-out* techniques such as in Rakhlin et al. (2012); Block et al. (2022) is that the hallucinated samples are generated from $\hat{\mu}_N$ instead of μ. Crucially, our sampling is performed *without replacement* (not $i.i.d.$), which is essential for our following analysis (Lemma 9). More generally, one may also replace the estimation $\hat{\mu}_N$ with other estimation rules instead of the empirical distribution we used here. This could provide tighter bounds if the distribution μ is well structured, see Section 3.3. The following lemma shows that the predictor $\hat{y}_j$ can be computed *efficiently* by accessing to a mixed-ERM oracle.

Lemma 4 *The predictor $\hat{y}_j$ can be computed upto error $\pm\frac{1}{L\sqrt{M}}$ by making at most $O(L\sqrt{M}\log M)$ mixed-ERM oracle calls. Moreover, for binary valued class $\mathcal{H}$ with $y \in \{0,1\}$ and* absolute *loss, we need only 2 (regular) ERM orcale calls to compute $\hat{y}_j$ exactly.*

2. For technical reasons, we assume here that $\tilde{\mathbf{x}}^M_{j+1}$ is sampled from $\hat{\mu}_N$ *without replacement*. Equivalently, $\tilde{\mathbf{x}}^M_{j+1}$ is sampled uniformly from all (permuted) *subseqeunces* of $\mathbf{x}^0_{-N+1}$ of length $M-j$.

Proof Clearly, a naive approach for discretizing both $\hat{y}$ and y with scale $\frac{1}{L\sqrt{M}}$ yields an algorithm with L^2M oracle calls. The $O(L\sqrt{M}\log M)$ bound follows from (Block et al., 2022, Thm 7) leveraging the convexity on $\hat{y}$. The second part follows from the relation $\epsilon h(\tilde{\mathbf{x}}) = |h(\tilde{\mathbf{x}}) - \frac{(1-\epsilon)}{2}| - \frac{1-\epsilon}{2}$ for $\epsilon \in \{+1, -1\}$ and the second assertion of (Block et al., 2022, Thm 7). ■

Analysis of the regret. Denote by Φ the prediction rule derived from (2). We consider the following analogous hybrid minimax regret as in (1) with the additional *side-information*:

$$\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H}, \Phi) = \sup_{\mu} \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}_1} \sup_{y_1} \mathbb{E}_{\hat{y}_1} \cdots \mathbb{E}_{\mathbf{x}_M} \sup_{y_M} \mathbb{E}_{\hat{y}_M} \left[\sum_{j=1}^{M} \ell(\hat{y}_j, y_j) - \inf_{h\in\mathcal{H}} \sum_{j=1}^{M} \ell(h(\mathbf{x}_j), y_j) \right], \tag{3}$$

where the randomness of $\hat{y}_j$s is over the $\tilde{\mathbf{x}}$'s and ϵ's as in (2), while $\mathbf{x}_j$s are sampled $i.i.d.$ from μ.

To proceed, we first introduce the following key concept. Let $(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_M, y_M) \in \mathcal{X} \times [0, 1]$ be any realization of the feature-label pairs. We write $L^h_j = \sum_{i=1}^{j} \ell(h(\mathbf{x}_i), y_i)$ to simplify our discussion. The *surrogate* relaxation is defined as [3]

$$R_j = \mathbb{E}_{\tilde{\mathbf{x}},\epsilon} \left[\sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L^h_j \right], \tag{4}$$

where $\tilde{\mathbf{x}}_i$s and ϵ_is are generated the same way as in (2). We also define the following variation that replaces the single $\tilde{\mathbf{x}}_{j+1}$ with a sample $\mathbf{x} \sim \mu$:

$$\tilde{R}_j = \mathbb{E}_{\mathbf{x}\sim\mu} \mathbb{E}_{\tilde{\mathbf{x}},\epsilon} \left[\sup_{h\in\mathcal{H}} 2L\epsilon_{j+1} h(\mathbf{x}) + 2L \sum_{i=j+2}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L^h_j \right]. \tag{5}$$

Note that the main technique for proving the relaxation based regret bounds, such as Rakhlin et al. (2012), is through the concept of *admissibility*, which essentially asserts that $\mathbb{E}_{\mathbf{x}_j} \sup_{y_j} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}} [\ell(\hat{y}_j, y_j) + R_j] \le R_{j-1}$. However, a major technical step for establishing such an result is based on the so-called *randomness matching* argument by leveraging the fact that the *hallucinated samples* used to define the relaxation are the same as the actual feature generating process. This, unfortunately, is not true in our case since the empirical distribution $\hat{\mu}_N$ can deviate arbitrarily from μ under total variation, regardless of how large the sample size N is. We instead establish the following *approx-admissibility* of our surrogate relaxation, with the proof deferred to Appendix A.

Lemma 5 (Approx-Admissibility) *Let $\hat{y}_j$ be as in (2), then for all $j \in [M]$ we have:*

$$\mathbb{E}_{\mathbf{x}_j} \sup_{y_j} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}} [\ell(\hat{y}_j, y_j) + R_j] \le \tilde{R}_{j-1}. \tag{6}$$

We are now ready to state our first main technical lemma of this section, which follows from Lemma 5 by a "backward tracing" argument. The detailed proof is deferred to Appendix B.

3. Throughout the paper, we use the convention $\mathbb{E}_{\tilde{\mathbf{x}},\epsilon} \equiv \mathbb{E}_{\tilde{\mathbf{x}}^M_{j+1},\epsilon^M_{j+1}}$ to simplify notation.

Lemma 6 (Regret Bound via Approx-Admissibility) *Let Φ be the predictor as in (2). Then for any class $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ with a convex and L-Lipschitz loss ℓ, we have*

$$\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H}, \Phi) \leq \mathbb{E}_{\mathbf{x}^0_{-N+1}} \left[\tilde{R}_0 + \sum_{j=1}^{M-1} \mathbb{E}_{\mathbf{x}^j} \sup_{y^j} (\tilde{R}_j - R_j) \right], \tag{7}$$

where $\mathbf{x}^M_{-N+1}$ are sampled $i.i.d.$ from μ and R_j, $\tilde{R}_j$ are defined as in (4) and (5).

Remark 7 *Note that the decomposition presented in Lemma 6 holds whenever the approx-admissibility condition of Lemma 5 is satisfied. We believe this could be applicable to a broader set of problems and is of independent interest.*

Bounding the relaxations. By Lemma 6, we know that the regret $\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H}, \Phi)$ can be upper bounded by $\tilde{R}_0$ and the discrepancies between R_j and $\tilde{R}_j$. Clearly, by the definition of $\tilde{R}_j$, we have $\tilde{R}_0 \leq 2L\mathsf{Rad}_M(\mathcal{H})$, where $\mathsf{Rad}_M(\mathcal{H})$ is the Rademacher complexity of $\mathcal{H}$ as in Definition 2. To bound the discrepancies, for any $j \in [M-1]$, $\mathbf{x}^j, \tilde{\mathbf{x}}^M_{j+2} \in \mathcal{X}^*$, $\epsilon^M_{j+1} \in \{\pm 1\}^*$ and $y^j \in [0,1]^j$, we define the following function:

$$f_{\mathbf{x}^j, \tilde{\mathbf{x}}^M_{j+2}, \epsilon^M_{j+1}, y^j}(\mathbf{x}) = \sup_{h \in \mathcal{H}} \left\{ 2L\epsilon_{j+1} h(\mathbf{x}) + 2L \sum_{i=j+2}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L^h_j \right\}. \tag{8}$$

The following fact is a consequence of our definitions.

Fact 1 *We have $R_j = \mathbb{E}_{\tilde{\mathbf{x}}, \epsilon} \left[f_{\mathbf{x}^j, \tilde{\mathbf{x}}^M_{j+2}, \epsilon^M_{j+1}, y^j}(\tilde{\mathbf{x}}_{j+1}) \right]$ and $\tilde{R}_j = \mathbb{E}_{\mathbf{x} \sim \mu} \mathbb{E}_{\tilde{\mathbf{x}}, \epsilon} \left[f_{\mathbf{x}^j, \tilde{\mathbf{x}}^M_{j+2}, \epsilon^M_{j+1}, y^j}(\mathbf{x}) \right]$.*

Let $\mathbf{z}_j = (\mathbf{x}^j, \tilde{\mathbf{x}}^M_{j+2}, \epsilon^M_{j+1})$. We now observe the following key properties of the functions $f_{\mathbf{z}_j, y^j}(\mathbf{x})$, which demonstrates that $f_{\mathbf{z}_j, y^j}(\mathbf{x})$ has *sensitivity* upper bounded by $4L$ and is Lipschitz on y^j. We refer to Appendix D for the detailed proof.

Proposition 8 *For any $\mathbf{z}_j$ and y^j, we have $\sup_{\mathbf{x}, \mathbf{x}'} |f_{\mathbf{z}_j, y^j}(\mathbf{x}) - f_{\mathbf{z}_j, y^j}(\mathbf{x}')| \leq 4L$. Moreover, for all $\mathbf{z}_j$, $\mathbf{x}$ and $y^j, y'^j \in [0,1]^j$, we have $|f_{\mathbf{z}_j, y^j}(\mathbf{x}) - f_{\mathbf{z}_j, y'^j}(\mathbf{x})| \leq jL||y^j - y'^j||_\infty$.*

Note that Proposition 8 and Fact 1 immediately imply that $\tilde{R}_j - R_j \leq 4L||\mu - \hat{\mu}_N||_{\mathsf{TV}}$ [4]. Unfortunately, we are unable to bound the total variation distance $||\mu - \hat{\mu}_N||_{\mathsf{TV}}$ due to the lack of any structure we impose on μ. We instead establish the following key technical result, which bounds the discrepancies via a Rademacher sum of the functions $f_{\mathbf{z}_j, y^j}$. This result constitutes the main technical ingredient in our following analysis.

Lemma 9 *For all $j \in [M-1]$, $M \leq N/2$ and $B = N - M + j + 1$, we find*

$$\mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}^j} \sup_{y^j} (\tilde{R}_j - R_j) \leq \sup_{\mathbf{x}^{-N+B}_{-N+1}, \mathbf{x}'^B, \mathbf{z}_j} \mathbb{E}_{\epsilon'^B} \left[\sup_{y^j} \frac{1}{B} \sum_{i=1}^{B} \epsilon'_i (f_{\mathbf{z}_j, y^j}(\mathbf{x}'_i) - f_{\mathbf{z}_j, y^j}(\mathbf{x}_{-N+i})) \right], \tag{9}$$

where $\mathbf{x}^{-N+B}_{-N+1}, \mathbf{x}'^B, \mathbf{z}_j$ run over all possible values and ϵ'^B is distributed uniformly over $\{\pm 1\}^B$.

4. Using the fact that $\mathbb{E}_{\mathbf{x} \sim \mu}[f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim \nu}[f(\mathbf{x})] \leq \sup_{\mathbf{x}, \mathbf{x}'} |f(\mathbf{x}) - f(\mathbf{x}')| \cdot ||\mu - \nu||_{\mathsf{TV}}$.

Proof [Sketch] We highlight only the main idea here and refer to Appendix C for the complete proof. By Fact 1, we can upper bound the discrepancies by $\mathbb{E}_{\mathbf{x}^0_{-N+1}}\mathbb{E}_{\mathbf{z}_j}\sup_{y^j}[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}}[f_{\mathbf{z}_j,y^j}(\tilde{\mathbf{x}}_{j+1})]]$, where $\mathbf{z}_j = (\mathbf{x}^j, \tilde{\mathbf{x}}^M_{j+2}, \epsilon^M_{j+1})$. Note that $\tilde{\mathbf{x}}^M_{j+1}$ is sampled uniformly from $\mathbf{x}^0_{-N+1}$ *without replacement* as in (2). Therefore, the randomness of $\tilde{\mathbf{x}}^M_{j+1}$ can be described as follows: we first sample $\tilde{\mathbf{x}}^M_{j+2}$ from $\mathbf{x}^0_{-N+1}$ and then sample $\tilde{\mathbf{x}}_{j+1}$ *uniformly* from the remaining samples in $\mathbf{x}^0_{-N+1}$. Now, the key observation is that, by symmetries of $\mathbf{x}^0_{-N+1}$ (which are $i.i.d.$), we can *fix* $\tilde{\mathbf{x}}^M_{j+2}$ being the last $M-j-1$ samples in $\mathbf{x}^0_{-N+1}$. Therefore, we have $\mathbb{E}_{\tilde{\mathbf{x}}_{j+1}}[f_{\mathbf{z}_j,y^j}(\tilde{\mathbf{x}}_{j+1})] = \frac{1}{B}\sum_{i=1}^{B} f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})$, where $B = N-M+j+1$. Since $\mathbf{z}_j$ is *decoupled* from $\mathbf{x}^{-N+B}_{-N+1}$ by our construction, we obtain the upper bound $\mathbb{E}_{\mathbf{z}_j}\mathbb{E}_{\mathbf{x}^{-N+B}_{-N+1}}\sup_{y^j}[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \frac{1}{B}\sum_{i=1}^{B} f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})]$. The lemma then follows by *symmetrization* with $\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})]$ (see Appendix C). ■

For any $j \in [M-1]$ and $\mathbf{z}_j$ as above, we define the following function class [5]:

$$\mathcal{G}_{\mathbf{z}_j} = \{g_{\mathbf{z}_j,y^j}(\mathbf{x},\mathbf{x}') \stackrel{\text{def}}{=} f_{\mathbf{z}_j,y^j}(\mathbf{x}') - f_{\mathbf{z}_j,y^j}(\mathbf{x}) : y^j \in [0,1]^j, (\mathbf{x},\mathbf{x}') \in \mathcal{X}^2\}. \tag{10}$$

Lemma 9 essentially states that the discrepancy between R_j and $\tilde{R}_j$ is upper bounded by the Rademacher complexity of the class $\mathcal{G}_{\mathbf{z}_j}$ as $\mathbb{E}_{\mathbf{x}^0_{-N+1}}\mathbb{E}_{\mathbf{x}^j}\sup_{y^j}(\tilde{R}_j - R_j) \le \sup_{\mathbf{z}_j}\frac{1}{B}\mathsf{Rad}_B(\mathcal{G}_{\mathbf{z}_j})$.

The following lemma provides a useful bound on such Rademacher complexities.

Lemma 10 *Let $\mathcal{G}_{\mathbf{z}_j}$ be as in (10), $M \le N/2$ and $B = N-M+j+1$. Then*

$$\sup_{\mathbf{z}_j}\frac{1}{B}\mathsf{Rad}_B(\mathcal{G}_{\mathbf{z}_j}) \le O\left(\sqrt{\frac{jL^2\log(jLB)}{B}}\right) \le O\left(\sqrt{\frac{2jL^2\log(jLN/2)}{N}}\right). \tag{11}$$

Proof Let $\mathcal{C} \subset [0,1]^j$ be a covering of $[0,1]^j$ with norm L_∞ radius $\frac{1}{jLB}$. We have $|\mathcal{C}| \le (jLB)^j$. By the second part of Proposition 8, we know that the class $\mathcal{G}'_{\mathbf{z}_j} \stackrel{\text{def}}{=} \{g_{\mathbf{z}_j,y^j} : y^j \in \mathcal{C}\}$ forms a uniform L_∞-covering of $\mathcal{G}_{\mathbf{z}_j}$ with radius $\frac{2}{B}$. Therefore, $\frac{1}{B}\mathsf{Rad}_B(\mathcal{G}_{\mathbf{z}_j}) \le \frac{1}{B}\mathsf{Rad}_B(\mathcal{G}'_{\mathbf{z}_j}) + \frac{2}{B}$. The first inequality then follows by a simple application of Massart's lemma (Shalev-Shwartz and Ben-David, 2014, Lemma 26.8) over $\mathcal{G}'_{\mathbf{z}_j}$, since $|\mathcal{G}'_{\mathbf{z}_j}| \le |\mathcal{C}| \le (jLB)^j$ and $\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{X}^2}\{g_{\mathbf{z}_j,y^j}(\mathbf{x},\mathbf{x}')\} \le 4L$ for all $g_{\mathbf{z}_j,y^j} \in \mathcal{G}_{\mathbf{z}_j}$ due to the first part of Proposition 8. The second inequality is implied by that $B \ge N/2$ and the fact that the function $\frac{\log B}{B}$ is monotone decreasing. ■

Putting everything together, we arrive at:

Theorem 11 *Let Φ be the predictor as in (2) and $M \le N/2$. Then for any class $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ with a convex and L-Lipschitz loss ℓ, the predictor Φ can be computed efficiently with access to at most $O(L\sqrt{M}\log M)$ mixed-ERM oracle calls per round such that*

$$\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H},\Phi) \le 2L\mathsf{Rad}_M(\mathcal{H}) + \sqrt{M} + O\left(\sqrt{\frac{M^3L^2\log(MLN)}{N}}\right). \tag{12}$$

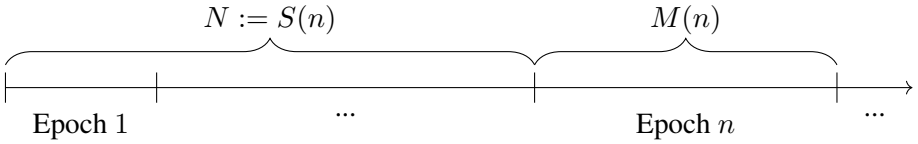
Proof The regret bound follows directly from Lemma 4, Lemma 6 and Lemma 9. We then invoke Lemma 10 to bound the discrepancies by noticing that $j \le M$. ■

5. Note that the "complexity" of $\mathcal{G}_{\mathbf{z}_j}$ arises from the $y^j \in [0,1]^j$.

Remark 12 *Note that Theorem 11 shows that if $N \gg M^2 \log M$ then the regret with side-information is reduced to the Rademacher complexities of $\mathcal{H}$, and thus matches the case when the distribution is known in advance. However, in reality such side-information is not available for the unknown distribution case, which can only be obtained from prior samples.*

3.2. Proof of Theorem 3: the Epoch Approach

We are now equipped with all the technical tools to prove Theorem 3, with the only missing ingredient of constructing the *side-information*. For this purpose, we employ an *epoch-based* approach, resembling those used in Lazaric and Munos (2009); Wu et al. (2023a), but in a completely different context. We partition the time horizon into epochs, with epoch n of length $M(n)$. Let $S(n) = \sum_{i=1}^{n-1} M(i)$ be the total time steps after $n-1$ epochs. We will use the features observed upto time $S(n)$ as the side-information introduces in Section 3.1 and apply the predictor constructed in (2) to make the prediction during the nth epoch.



To this end, our main technical part is to *optimize* the epoch length $M(n)$ that balances the trade-off in (12) and achieving the minimal total regret. Let Φ be the predictor derived from (2), which we write as $\Phi(\mathbf{x}^0_{-N+1}, \mathbf{x}^j, y^{j-1})$ for the side-information $\mathbf{x}^0_{-N+1}$, features $\mathbf{x}^j$ and labels y^{j-1} observed thus far. We define the following *epoch* predictor Ψ: for any epoch n and time step j during such epoch, we set

$$\Psi(\mathbf{x}^{S(n)+j}, y^{S(n)+j-1}) = \Phi\left(\mathbf{x}^{S(n)}, \mathbf{x}^{S(n)+j}_{S(n)+1}, y^{S(n)+j-1}_{S(n)+1}\right). \tag{13}$$

Let $S^{-1}(T)$ be the largest number n such that $S(n) < T$. The following lemma upper bounds the hybrid minimax regret (1) of Ψ using the regrets with side information (3) incurred by Φ. Note that this is *not* immediately obvious since we have *reused* the side-information among different epochs.

Lemma 13 *For any $\mathcal{H}$ and convex L-Lipschitz loss ℓ, we have*

$$\tilde{r}_T(\mathcal{H}, \Psi) \le \sum_{n=1}^{S^{-1}(T)} \tilde{r}^{\mathsf{side}}_{M(n),S(n)}(\mathcal{H}, \Phi).$$

Proof Define the operator $\mathbb{Q}_i^j \equiv \mathbb{E}_{\mathbf{x}_i} \sup_{y_i} \mathbb{E}_{\hat{y}_i} \cdots \mathbb{E}_{\mathbf{x}_j} \sup_{y_j} \mathbb{E}_{\hat{y}_j}$, where $\hat{y}_t \sim \Psi(\mathbf{x}^t, y^{t-1})$ for all $t \in [T]$. We have (truncate the last $S(n+1)$ above T if necessary):

$$\begin{aligned}
\tilde{r}_T(\mathcal{H}, \Psi) &= \mathbb{Q}_1^T \sup_{h\in\mathcal{H}} \left[\sum_{n=1}^{S^{-1}(T)} \sum_{j=S(n)+1}^{S(n+1)} \ell(\hat{y}_j, y_j) - \ell(h(\mathbf{x}_j), y_j) \right] \\
&\overset{(a)}{\le} \sum_{n=1}^{S^{-1}(T)} \mathbb{Q}_1^T \sup_{h\in\mathcal{H}} \left[\sum_{j=S(n)+1}^{S(n+1)} \ell(\hat{y}_j, y_j) - \ell(h(\mathbf{x}_j), y_j) \right]
\end{aligned}$$

$$\overset{(b)}{=} \sum_{n=1}^{S^{-1}(T)} \mathbb{E}_{\mathbf{x}^{S(n)}} \mathbb{Q}_{S(n)+1}^{S(n+1)} \sup_{h\in\mathcal{H}} \left[\sum_{j=S(n)+1}^{S(n+1)} \ell(\hat{y}_j, y_j) - \ell(h(\mathbf{x}_j), y_j) \right]$$
$$\overset{(c)}{=} \sum_{n=1}^{S^{-1}(T)} \tilde{r}^{\mathsf{side}}_{M(n),S(n)}(\mathcal{H}, \Phi),$$

where (a) follows by $\sup(A+B) \le \sup A + \sup B$ and linearity of expectation; (b) follows since $\hat{y}_j$ depends only on $\mathbf{x}^j$ and $y^j_{S(n)}$ for $j \in (S(n), S(n+1)]$; (c) follows by definition. ■

Proof [Proof of Theorem 3] Assume $\mathsf{Rad}_T(\mathcal{H}) \le O(T^q)$ for some $q \in [\frac{1}{2}, 1]$. By Theorem 11 and $M(n), S(n) \le T$ we have

$$\tilde{r}^{\mathsf{side}}_{M(n),S(n)}(\mathcal{H}, \Phi) \le O\left(LM(n)^q + \sqrt{\frac{M(n)^3 L^2 \log(LT^2)}{S(n)}} \right).$$

Let $M(n) = n^\alpha$ for some $\alpha > 0$ to be determined later. We have $S(n) = \sum_{i=1}^{n-1} i^\alpha = \Theta(n^{\alpha+1})$ by integration approximation, and $S^{-1}(T) \le O(T^{1/(\alpha+1)})$. This implies that $\tilde{r}^{\mathsf{side}}_{M(n),S(n)}(\mathcal{H}, \Phi) \le O(Ln^{\alpha q} + L\sqrt{\log(LT^2)} n^{\alpha - \frac{1}{2}})$. By Lemma 13 and integration approximation again, we conclude

$$\tilde{r}_T(\mathcal{H}, \Psi) \le O\left(LT^{\frac{\alpha q+1}{\alpha+1}} + L\sqrt{\log(LT^2)} T^{\frac{\alpha+\frac{1}{2}}{\alpha+1}} \right). \tag{14}$$

Optimizing $\arg\min_{\alpha>0} \max\{\frac{\alpha q+1}{\alpha+1}, \frac{\alpha+\frac{1}{2}}{\alpha+1}\}$, we find (14) is minimized when taking $\alpha = \frac{1}{2(1-q)}$. Plugging back to (14), we find $\tilde{r}_T(\mathcal{H}, \Psi) \le O\left(L\sqrt{\log(LT)} T^{\frac{2-q}{3-2q}} \right)$. This completes the proof of the first part. The second and third parts follow by the facts that $\mathsf{Rad}_T(\mathcal{H}) \le O(\sqrt{\mathsf{VC}(\mathcal{H})T})$ for finite-VC class (Wainwright, 2019, Example 5.24), and $\mathsf{Rad}_T(\mathcal{H}) \le \tilde{O}(T^{\max\{\frac{1}{2}, \frac{p-1}{p}\}})$ for classes with α-fat shattering dimension of order α^{-p} (Block et al., 2022). This completes the proof and the big-O notations and $M(n) \le S(n)/2$ are justified by noting that $\alpha \ge 1$ since $q \ge \frac{1}{2}$. ■

3.3. Tighter Bounds for Special Classes

As demonstrated in Section 3.1, the main technical obstacle for analyzing the regret is to upper bound the discrepancies between $\tilde{R}_j$ and R_j as in Lemma 6. It was shown in Lemma 9 that such discrepancies can be upper bounded by the Rademacher complexity of the class $\mathcal{G}_{\mathbf{z}_j}$ in (10). We demonstrate in this section how to leverage the *structural* information of $\mathcal{G}_{\mathbf{z}_j}$ leading to tighter regret bounds for certain special classes, when compared to the general bounds from Theorem 3.

Binary valued classes. Let $\mathcal{H} \subset \{0,1\}^{\mathcal{X}}$ be a binary valued class and $\ell(\hat{y}, y) = |\hat{y} - y|$. For any given $\mathbf{z}_j$ (assume, w.l.o.g., $\epsilon_{j+1} = 1$) and $y^j \in \{0,1\}^j$, the function $f_{\mathbf{z}_j, y^j}$ can be expressed as $f_{\mathbf{z}_j, y^j}(\mathbf{x}) = \sup_h \{2h(\mathbf{x}) + F(h)\}$ (see definition in (8)), where $F(h)$ is a discrete valued function taking values in $[-2M, 2M]$. Define $\mathcal{H}^0 = \{h \in \mathcal{H} : F(h) = \sup_{h' \in \mathcal{H}} F(h')\}$ and $\mathcal{H}^1 = \{h \in \mathcal{H} : F(h) = \sup_{h' \in \mathcal{H}} F(h') - 1\}$. Let $h^0(\mathbf{x}) = \sup_{h \in \mathcal{H}^0}\{h(\mathbf{x})\}$, $h^1(\mathbf{x}) = \sup_{h \in \mathcal{H}^1}\{h(\mathbf{x})\}$ and $\hat{h} = \arg\max_{h \in \mathcal{H}} F(h)$. The following *structural* characterization of $f_{\mathbf{z}_j, y^j}$ holds:

Fact 2 *For any* $\mathbf{x}$*, if* $h^0(\mathbf{x}) = 1$ *then* $f_{\mathbf{z}_j,y^j}(\mathbf{x}) = F(\hat{h}) + 2$*; if* $h^0(\mathbf{x}) = 0$ *and* $h^1(\mathbf{x}) = 1$ *then* $f_{\mathbf{z}_j,y^j}(\mathbf{x}) = F(\hat{h}) + 1$ *; else* $f_{\mathbf{z}_j,y^j}(\mathbf{x}) = F(\hat{h})$.

Theorem 14 *Let* $\mathcal{H} \subset \{0,1\}^{\mathcal{X}}$, $\mathcal{F}^{\mathsf{u}} = \{f_{\mathcal{H}'}(\mathbf{x}) = \sup_{h\in\mathcal{H}'}\{h(\mathbf{x})\} : \mathcal{H}' \subset \mathcal{H}\}$, $\mathcal{F}^{\mathsf{i}} = \{f_{\mathcal{H}'}(\mathbf{x}) = \inf_{h\in\mathcal{H}'}\{h(\mathbf{x})\} : \mathcal{H}' \subset \mathcal{H}\}$ *be two classes of functions and* ℓ *be the absolute loss. Then there exists an oracle-efficient predictor* Φ *satisfying* $\tilde{r}_T(\mathcal{H},\Phi) \le O(\sqrt{\max\{\mathsf{VC}(\mathcal{F}^{\mathsf{u}}),\mathsf{VC}(\mathcal{F}^{\mathsf{i}})\}T})$.

Proof [Sketch] By Fact 2, we know $f_{\mathbf{z}_j,y^j}$ is completely determined by the functions h^0, h^1. Therefore, the "complexity" of $\mathcal{G}_{\mathbf{z}_j}$ is reduced to that of $\{h^0, h^1\}$s. The proof then follows by the key observation that $h^0, h^1 \in \mathcal{F}^{\mathsf{u}}$ and techniques in Section 3.1 & 3.2. See Appendix D for details. ■

Note that for the threshold functions $\mathcal{H} = \{1\{x \ge a\} : a, x \in [0,1]\}$ we have $\mathcal{F}^{\mathsf{u}} = \mathcal{F}^{\mathsf{i}} = \mathcal{H}$. Theorem 14 implies an oracle efficient $O(\sqrt{T})$ regret, which matches the information-theoretical lower bound and is tighter than the covering-based $O(\sqrt{T\log T})$ bound implied by Lazaric and Munos (2009). Another example is the class of indicators of intervals with bounded length $\{1\{x \in [a,b]\} : b - a \ge \gamma, [a,b] \subset [0,1]\}$, for which we have $\mathsf{VC}(\mathcal{F}^{\mathsf{i}}) = 2$ and $\mathsf{VC}(\mathcal{F}^{\mathsf{u}}) \le O(\frac{1}{\gamma})$.

Lipschitz functions. Let $\mathcal{X} = [0,1]^d$ and $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ be the class of *all* 1-Lipschitz functions under L_∞ norm. Assume $\ell(\hat{y}, y) = |\hat{y} - y|$ is the absolute loss. Let μ and $\hat{\mu}_N$ be the true and empirical distributions, respectively, as in Section 3.1. By Fact 1 and assuming that $\tilde{\mathbf{x}}_{j+1}^M$ is sampled $i.i.d.$ from $\hat{\mu}_N$, we have $\mathbb{E}_{\mathbf{x}_{-N+1}^j} \sup_{y^j}(\tilde{R}_j - R_j) \le \mathbb{E}_{\mathbf{x}_{-N+1}^0} \sup_{y^j,\mathbf{z}_j}(\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \mathbb{E}_{\mathbf{x}\sim\hat{\mu}_N}[f_{\mathbf{z}_j,y^j}(\mathbf{x})])$. By the same argument as Proposition 8 (second part) and Lipschitz property of $h \in \mathcal{H}$, we have:

Fact 3 *For all* $\mathbf{z}_j, y^j$ *and* $\mathbf{x}, \mathbf{x}'$, $|f_{\mathbf{z}_j,y^j}(\mathbf{x}) - f_{\mathbf{z}_j,y^j}(\mathbf{x}')| \le 2||\mathbf{x} - \mathbf{x}'||_\infty$.

Theorem 15 *Let* $\mathcal{H}$ *and* ℓ *be as above. Then, there exists an oracle-efficient predictor* Φ *such that* $\tilde{r}_T(\mathcal{H},\Phi) \le \tilde{O}(T^{\max\{\frac{1}{2},\frac{d-1}{d}\}})$*, and this bound is tight upto poly-logarithmic factors.*

Proof By Fact 3, we know that for all $\mathbf{z}_j, y^j$ the function $f_{\mathbf{z}_j,y^j}(\mathbf{x})$ is 2-Lipschitz. Therefore, by Kantorovich-Rubinstein duality (Villani, 2021) we have $\sup_{y^j,\mathbf{z}_j}(\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \mathbb{E}_{\mathbf{x}\sim\hat{\mu}_N}[f_{\mathbf{z}_j,y^j}(\mathbf{x})]) \le 2W_1(\mu,\hat{\mu}_N)$, where $W_1(\mu,\hat{\mu}_N) = \inf_{\gamma\in\Gamma(\mu,\hat{\mu}_N)} \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\gamma}[||\mathbf{x} - \mathbf{x}'||_\infty]$ is the Wasserstein 1-distance with $\Gamma(\mu,\hat{\mu}_N)$ being the class of all coupling between $\mu, \hat{\mu}_N$. Therefore, we have $\mathbb{E}_{\mathbf{x}_{-N+1}^j} \sup_{y^j}(\tilde{R}_j - R_j) \le 2\mathbb{E}_{\mathbf{x}_{-N+1}^0}[W_1(\mu,\hat{\mu}_N)]$, i.e., the discrepancy is upper bounded by the convergence rate of empirical distribution under Wasserstein 1-distance. Invoking (Fournier and Guillin, 2015, Thm 1) and boundedness of $\mathcal{X}$, we have $\mathbb{E}_{\mathbf{x}_{-N+1}^0}[W_1(\mu,\hat{\mu}_N)] \le \tilde{O}(N^{-1/d})$. Let Φ be the predictor in (2). By Lemma 6 and $\mathsf{Rad}_M(\mathcal{H}) \le \tilde{O}(M^{\max\{\frac{1}{2},\frac{d-1}{d}\}})$ (Wainwright, 2019), we have $\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H}) \le \tilde{O}(M^{\max\{\frac{1}{2},\frac{d-1}{d}\}} + MN^{-1/d})$. The result then follows by Lemma 13 with $M(n) = 2^n$ (which ensures $N = S(n) = M(n) - 1$). The last part follows by that the ϵ-metric entropy of $\mathcal{H}$ is $\Theta(\frac{1}{\epsilon^d})$ (Wainwright, 2019). ■

Remark 16 *Note that, if we assume certain structure on* μ *that admits a computationally efficient estimator* $\hat{\mu}_N$ *that satisfies* $||\mu - \hat{\mu}_N||_{\mathsf{TV}} \le O(\frac{1}{\sqrt{N}})$ *(such as for Gaussian distributions (Ashtiani et al., 2018)), then the (optimal)* $O(\mathsf{Rad}_T(\mathcal{H}) + \sqrt{T})$ *bound is achievable for* any *class* $\mathcal{H} \subset [0,1]^{\mathcal{X}}$.

4. Shifting Distributions

In this section, we consider a scenario where the underlying feature-generating distribution is allowed to *change* over time. We assume that the total number of changes is upper bounded by K, while the selection of distributions and change points are completely arbitrary. It was demostrated by Wu et al. (2023a) that for finite-VC classes and absolute loss, the regret grows as $O(\sqrt{KT\log T})$ under such feature generation processes. However, their algorithm depends on the construction of an exponentially sized cover, making it computationally inefficient. We will demonstrate in this section an *oracle-efficient* algorithm that achieves a slightly worse regret.

We first observe that if we *know* the positions of the change points, then we can simply run our oracle-efficient predictor from Theorem 3 on each of the segments independently, leading to an $\tilde{O}(T^{\frac{3}{4}}K^{\frac{1}{4}})$ regret. Since there are only T^K possible configurations of the change points, we can therefore run an expert algorithm to aggregate each of such configurations, leading to an $\tilde{O}(\sqrt{KT\log T} + T^{\frac{3}{4}}K^{\frac{1}{4}})$ regret. Unfortunately, this approach has computational cost dominated by $\Omega(T^K)$ and therefore not efficient for large K. To address this issue, we instead partition the time horizon into epochs with *fixed* length B and run our oracle-efficient algorithm on each of the epochs independently. The rationale behind this approach is that, if we select B small enough, there will be at least $\frac{T}{B} - K$ epochs with $i.i.d.$ sampling. By tuning the epoch length B, we arrive at:

Proposition 17 *Let $\mathcal{H} \subset \{0,1\}^{\mathcal{X}}$ be a binary valued class with finite VC-dimension under a convex and L-Lipschitz loss. Then there exists an oracle-efficient predictor Φ such that if the features are generated by the process with change cost K and the labels are generated adversarially, then the hybrid minimax regret as (1) is upper bounded by $O\left(L\sqrt{\mathsf{VC}(\mathcal{H})\log(LT)}K^{\frac{1}{5}}T^{\frac{4}{5}}\right)$.*

Proof We partition the time horizon into $\frac{T}{B}$ epochs each of length B. Let Φ be the predictor from Theorem 3 that we run on each epochs independently. By independence, we know that the total regret is upper bounded by the sum of regrets incurred on each of the epochs. Note that, if an epoch does not contain a change point, then the regret is upper bounded by $O(L\sqrt{\mathsf{VC}(\mathcal{H})\log(LT/B)}B^{\frac{3}{4}})$ by Theorem 3, else we naively upper bounded by B. Since there can be at most K epochs containing a change point, the total regret is upper bounded by $O\left(\left(\frac{T}{B} - K\right)L\sqrt{\mathsf{VC}(\mathcal{H})\log(LT/B)}B^{\frac{3}{4}} + KB\right)$ by optimizing over B: We find that $B = T^{\frac{4}{5}}K^{-\frac{4}{5}}$ attains the minimum and the result follows. ■

5. Contextual K-arm Bandits

We now briefly discuss the extension to the contextual K-arm bandits as introduced in Section 2, leaving details to Appendix E. The basic idea follows the same path as in the online learning case, where we partition the time steps into epochs, at each epoch we use the sample observed thus far to estimate the underlying distribution and use the estimated distribution to generate the *hallucinated samples* as needed in the relaxation based algorithms.

Bandit predictor with side-information. Let $(\mathbf{x}_1, c_1), \cdots, (\mathbf{x}_M, c_M)$ be any realization of the context-cost pairs and $\mathbf{x}_{-N+1}^0$ be the side-information with $\mathbf{x}_{-N+1}^M$ sampled $i.i.d.$ from an (unknown) distribution μ. At each time step $j \in [M]$, we construct the following prediction rule adapted from Syrgkanis et al. (2016) by generating the hallucinated samples from $\hat{\mu}_N$ instead of μ:

1. Sample $\tilde{\mathbf{x}}_{j+1}^M$ from $\hat{\mu}_N$ *without replacement*; ϵ_{j+1}^M *i.i.d.* from uniform distribution over $\{\pm 1\}^K$; Z_{j+1}^M *i.i.d.* from distribution over $\{0, \frac{1}{\gamma}\}$ such that $\Pr[Z_i = \frac{1}{\gamma}] = \gamma K$, where γ is a parameter to be tuned; Let e_k be the standard base of $\mathbb{R}^K$ with coordinate k being 1;

2. Let $\mathcal{D}_K$ be the class of distributions over $[K]$. Find (using ERM oracle):

$$\hat{q}_j = \operatorname*{arg\,min}_{q \in \mathcal{D}_K} \sup_{p_j \in \mathcal{D}'} \mathbb{E}_{\hat{c}_j \sim p_j} \left\{ \langle q, \hat{c}_j \rangle - \inf_{h \in \mathcal{H}} \left(\sum_{i=1}^{j} \hat{c}_i[h(\mathbf{x}_i)] + \sum_{i=j+1}^{M} 2\epsilon_i[h(\tilde{\mathbf{x}}_i)] Z_i \right) + \gamma(M-j)K \right\},$$

   where $\mathcal{D}'$ is the class of distributions over $\{\frac{1}{\gamma} e_k : k \in [K]\} \cup \{\mathbf{0}\}$ such that the probability equals $\frac{1}{\gamma} e_k$ is upper bounded by γ for all $k \in [K]$; and $\mathbf{0} \in \mathbb{R}^K$ is the all 0s vector;

3. Define $q_j = (1 - \gamma K)\hat{q}_j + \gamma \mathbf{1}$ and make prediction $\hat{y}_j \sim q_j$, where $\mathbf{1}$ is the all 1s vector.

Here, the $\hat{c}_i$ at step 2 is an unbiased estimation of c_i as $\hat{c}_i = \frac{1}{\gamma} I_i e_{\hat{y}_i}$, where I_i is the indicator that takes value 1 w.p. $\frac{\gamma c_i[\hat{y}_i]}{q_i[\hat{y}_i]}$ with $\hat{y}_i$ and q_i being the predictions at step $i \le j-1$.

Analysis of regret. Our key idea is to define a *surrogate* relaxation R_j and $\tilde{R}_j$ for the predictor $\hat{q}_j$ and establish a bandit version of decomposition for the regret with side-information as in Lemma 6. This is achieved via the concept of *approx-admissibility* as in Lemma 5 and a careful adaption of the admissibility proof from Syrgkanis et al. (2016). The technical challenge then boils down to bounding the discrepancies between R_j and $\tilde{R}_j$, as in Lemma 6. To this end, we employ a technique similar to Lemma 9 that relates the discrepancies to a Rademacher sum and show that the *sensitivity* of the functions in the sum is upper bounded by $O(K)$. Crucially, the sensitivity is *independent* of γ, and therefore the regret with side information is optimized at $\gamma = (\log|\mathcal{H}|/KM)^{\frac{1}{3}}$. By leveraging a similar epoch approach as in the proof of Theorem 3, and setting the epoch length to $n^{\frac{3}{2}}$, we arrive at the main result of this section (see Appendix E for a detailed proof):

Theorem 18 *Let $\mathcal{H} \subset [K]^{\mathcal{X}}$ be a finite policy set. Then there exists an oracle-efficient predictor Φ such that the hybrid bandit minimax regret defined in Section 2 is upper bounded by*

$$\tilde{r}_T^{\mathsf{bandit}}(\mathcal{H}, \Phi) \le O\left((K^{\frac{2}{3}} (\log|\mathcal{H}|)^{\frac{1}{3}} + K\sqrt{\log K}) \cdot T^{\frac{4}{5}} \right).$$

Remark 19 *Note that, our primary focus here is on the dependency on T. We believe a more careful selection of the epoch length and employing techniques from Banihashem et al. (2023) could result in a better dependency on K. We leave it as an open problem to improve the $\frac{4}{5}$ exponent of T.*

6. Additional Related Work

The relaxation-based approach was first introduced by Rakhlin et al. (2012), providing a generic method for constructing sequential prediction algorithms (albeit potentially inefficient) for a wide range of online learning scenarios. Rakhlin et al. (2012) demonstrated that an *oracle-efficient* online learning algorithm is feasible via the so-called *random play-out* approach, provided one can access a sampling oracle for *future* features. This includes applications such as transduction online learning (Kakade and Kalai, 2005; Cesa-Bianchi and Shamir, 2013) and *known* i.i.d. feature generation distributions (Rakhlin and Sridharan, 2016; Syrgkanis et al., 2016; Banihashem et al., 2023). A more

sophisticated scenario, the *smooth adversarial* setting, was investigated by Rakhlin et al. (2011); Haghtalab et al. (2022b); Block et al. (2022); Haghtalab et al. (2022a). In this setting, the future sampling distribution is not directly accessible but can be stochastically controlled via a *coupling* argument introduced by Haghtalab et al. (2022b). However, this approach still requires access to a sampling oracle of the underlying reference measure.

An alternative technical approach, the *follow the perturbed leader* (FTPL) algorithm, has been widely used in the literature for obtaining oracle-efficient algorithms (Kakade and Kalai, 2005; Haghtalab et al., 2022a; Block et al., 2022; Bhatt et al., 2023). A key technical benefit of this approach, compared to the relaxation-based approach, is that the prediction rule is *proper* (i.e., the prediction can be generated by a function within the hypothesis class) and involves fewer ERM oracle calls. However, it also suffers from higher regret guarantees for general fat-shattering classes Block et al. (2022). Interestingly, the FTPL-based approaches share some technical similarities with our work as well. For instance, our Lemma 6 is similar to the decomposition of regret as presented in (Cesa-Bianchi and Lugosi, 2006, Chapter 4.3). Additionally, the way we control the discrepancies of our *surrogate* relaxation in Lemma 9 is similar in spirit to methods employed by Haghtalab et al. (2022a); Block et al. (2022) for bounding their "generalization errors". However, it is unclear whether this FTPL-based approach can be adapted to our unknown distribution setting.

Concurrent work. We note also a recent concurrent work by Block et al. (2024), which specifically studied the ERM rule for smoothed adversaries with an *unknown* reference measure under the assumption of *realizable* labels and with L_2 loss. However, this work does not provide sublinear regret bounds for *adversarially* generated labels as investigated in our paper.

Acknowledgments

This work was partially supported by the NSF Center for Science of Information (CSoI) Grant CCF-0939370, and also by NSF Grants CCF-2006440, CCF-2007238, and CCF-2211423.

References

Noga Alon, Shai Ben-David, Nicolo Cesa-Bianchi, and David Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *Journal of the ACM (JACM)*, 44(4):615–631, 1997.

Hassan Ashtiani, Shai Ben-David, Nicholas Harvey, Christopher Liaw, Abbas Mehrabian, and Yaniv Plan. Nearly tight sample complexity bounds for learning mixtures of gaussians via sample compression schemes. *Advances in Neural Information Processing Systems*, 31, 2018.

Kiarash Banihashem, MohammadTaghi Hajiaghayi, Suho Shin, and Max Springer. An improved relaxation for oracle-efficient adversarial contextual bandits. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Alankrita Bhatt, Nika Haghtalab, and Abhishek Shetty. Smoothed analysis of sequential probability assignment. *arXiv preprint arXiv:2303.04845*, 2023.

Adam Block, Yuval Dagan, Noah Golowich, and Alexander Rakhlin. Smoothed online learning is as easy as statistical learning. *arXiv preprint arXiv:2202.04690*, 2022.

Adam Block, Alexander Rakhlin, and Abhishek Shetty. On the performance of empirical risk minimization with smoothed data. *arXiv preprint arXiv:2402.14987*, 2024.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.

Nicolò Cesa-Bianchi and Ohad Shamir. Efficient transductive online learning via randomized rounding. *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pages 177–194, 2013.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, 162(3-4):707–738, 2015.

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 smoothed adversaries. In *Advances in Neural Information Processing Systems*, 2022a.

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

David Haussler. Sphere packing numbers for subsets of the boolean n-cube with bounded vapnik-chervonenkis dimension. *Journal of Combinatorial Theory, Series A*, 69(2):217–232, 1995.

Sham Kakade and Adam T Kalai. From batch to transductive online learning. *Advances in Neural Information Processing Systems*, 18, 2005.

Alessandro Lazaric and Rémi Munos. Hybrid stochastic-adversarial on-line learning. In *Conference on Learning Theory*, 2009.

Alessandro Lazaric and Rémi Munos. Learning with stochastic inputs and adversarial outputs. *Journal of Computer and System Sciences*, 78(5):1516–1537, 2012.

Alexander Rakhlin and Karthik Sridharan. Bistro: An efficient relaxation-based method for contextual bandits. In *International Conference on Machine Learning*, volume 48 of *PMLR*, pages 1977–1985. PMLR, 20–22 Jun 2016.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Stochastic and constrained adversaries. *arXiv preprint arXiv:1104.5070*, 2011.

Sasha Rakhlin, Ohad Shamir, and Karthik Sridharan. Relax and randomize: From value to algorithms. *Advances in Neural Information Processing Systems*, 25, 2012.

Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

Vasilis Syrgkanis, Haipeng Luo, Akshay Krishnamurthy, and Robert E Schapire. Improved regret bounds for oracle-based adversarial contextual bandits. *Advances in Neural Information Processing Systems*, 29, 2016.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Changlong Wu, Ananth Grama, and Wojciech Szpankowski. Online learning in dynamically changing environments. In *Conference on Learning Theory (COLT)*. PMLR, 2023a.

Changlong Wu, Mohsen Heidari, Ananth Grama, and Wojciech Szpankowski. Expected worst case regret via stochastic sequential covering. *Transactions on Machine Learning Research*, 2023b.

Appendix A. Proof of Lemma 5

In this section, we establish the *approx-admissibility* of our predictor in (2). The reasoning follows closely to the arguments as in (Rakhlin et al., 2012, Lemma 11&12) but needs careful adaption for handling the hallucinated samples $\tilde{\mathbf{x}}$s generated from $\hat{\mu}_N$. We have

$$
\begin{aligned}
\mathbb{E}_{\mathbf{x}_j} \sup_{y_j} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}} \left[\ell(\hat{y}_j, y_j) + R_j\right] &\overset{(a)}{=} \mathbb{E}_{\mathbf{x}_j} \sup_{y_j} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}} \left[\ell(\hat{y}_j, y_j) + \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_j^h\right] \\
&\le \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\sup_{y_j} \ell(\hat{y}_j, y_j) + \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - \ell(h(\mathbf{x}_j), y_j) - L_{j-1}^h\right] \\
&\overset{(b)}{\le} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{y_j} \ell(\hat{y}, y_j) + \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - \ell(h(\mathbf{x}_j), y_j) - L_{j-1}^h\right] \\
&= \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{y_j} \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + \ell(\hat{y}, y_j) - \ell(h(\mathbf{x}_j), y_j)\right] \\
&\overset{(c)}{\le} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{y_j} \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + \partial\ell(\hat{y}, y_j)(\hat{y} - h(\mathbf{x}_j))\right] \\
&\overset{(d)}{\le} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{y_j} \sup_{g_j\in[-L,L]} \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + g_j(\hat{y} - h(\mathbf{x}_j))\right] \\
&\overset{(e)}{\le} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{g_j\in\{-L,L\}} \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + g_j(\hat{y} - h(\mathbf{x}_j))\right]
\end{aligned}
$$

where (a) follows by the definition of R_j and that $\hat{y}_j$ has the same randomness as R_j (i.e, the $\tilde{\mathbf{x}}$s and ϵs); (b) is due to definition of $\hat{y}_j$; (c) is due to convexity of ℓ; (d) is due to L-Lipschitz property of ℓ; (e) follows by that the inner function is convex w.r.t. g_j and thus the $\sup_{g_j\in[-L,L]}$ is attained on the boundary $\{-L, L\}$. We have

$$
\begin{aligned}
&\mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{g_j\in\{-L,L\}} \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + g_j(\hat{y} - h(\mathbf{x}_j))\right] \\
&\quad\overset{(a)}{=} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\inf_{\hat{y}} \sup_{d_j\in\Delta(\{-L,L\})} \mathbb{E}_{g_j\sim d_j} \left[\sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + g_j(\hat{y} - h(\mathbf{x}_j))\right]\right] \\
&\quad\overset{(b)}{=} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\sup_{d_j\in\Delta(\{-L,L\})} \inf_{\hat{y}} \mathbb{E}_{g_j\sim d_j} \left[\sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + g_j(\hat{y} - h(\mathbf{x}_j))\right]\right] \\
&\quad\overset{(c)}{=} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j} \left[\sup_{d_j} \inf_{\hat{y}} \mathbb{E}_{g_j\sim d_j} \left[g_j\hat{y} + \sup_{h\in\mathcal{H}} 2L \sum_{i=j+1}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h - g_j h(\mathbf{x}_j)\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\inf_{\hat{y}}\left(\mathbb{E}_{g_j\sim d_j}[g_j\hat{y}] + \mathbb{E}_{g_j\sim d_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h - g_j h(\mathbf{x}_j)\right]\right)\right] \\
&\overset{(d)}{=} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\left(\inf_{\hat{y}}\mathbb{E}_{g_j'\sim d_j}[g_j'\hat{y}]\right) + \mathbb{E}_{g_j\sim d_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h - g_j h(\mathbf{x}_j)\right]\right] \\
&= \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j\sim d_j}\left[\inf_{\hat{y}}\mathbb{E}_{g_j'\sim d_j}[g_j'\hat{y}] + \sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h - g_j h(\mathbf{x}_j)\right]\right] \\
&= \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j\sim d_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + \inf_{\hat{y}}\mathbb{E}_{g_j'\sim d_j}[g_j'\hat{y}] - g_j h(\mathbf{x}_j)\right]\right] \\
&\overset{(e)}{\leq} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j\sim d_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + \mathbb{E}_{g_j'\sim d_j}[g_j' h(\mathbf{x}_j)] - g_j h(\mathbf{x}_j)\right]\right] \\
&\overset{(f)}{\leq} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j,g_j'\sim d_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + (g_j' - g_j)h(\mathbf{x}_j)\right]\right] \\
&\overset{(g)}{=} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j,g_j'\sim d_j}\mathbb{E}_{\epsilon_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + \epsilon_j(g_j' - g_j)h(\mathbf{x}_j)\right]\right] \\
&= \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j,g_j'\sim d_j}\mathbb{E}_{\epsilon_j}\left[\sup_{h\in\mathcal{H}}\underbrace{\left(2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h\right)}_{A} + \underbrace{\epsilon_j g_j' h(\mathbf{x}_j)}_{B} + \underbrace{(-\epsilon_j g_j h(\mathbf{x}_j))}_{C}\right]\right] \\
&\overset{(h)}{\leq} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\sup_{d_j}\mathbb{E}_{g_j\sim d_j}\mathbb{E}_{\epsilon_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + 2\epsilon_j g_j h(\mathbf{x}_j)\right]\right] \\
&\overset{(i)}{=} \mathbb{E}_{\epsilon,\tilde{\mathbf{x}}}\mathbb{E}_{\mathbf{x}_j}\left[\mathbb{E}_{\epsilon_j}\left[\sup_{h\in\mathcal{H}} 2L\sum_{i=j+1}^{M}\epsilon_i h(\tilde{\mathbf{x}}_i) - L_{j-1}^h + 2\epsilon_j L h(\mathbf{x}_j)\right]\right] \\
&= \tilde{R}_{j-1},
\end{aligned}
$$

where (a) follows by $\sup_{g_j\in\{-L,L\}} \equiv \sup_{d_j\in\Delta(\{-L,L\})}\mathbb{E}_{g_j\sim d_j}$ where $\Delta(\{-L,L\})$ is the set of all probability distributions over $\{-L,L\}$; (b) follows by the minimax theorem and noticing that the inner expectation is bi-linear w.r.t. $\hat{y}$ and d_j; (c) follows by the fact that $g_j\hat{y}$ is independent of $\sup_h$; (d) follows by that the $\sup_h$ term is independent of $\hat{y}$ and introducing an $i.i.d.$ copy g_j' of g_j; (e) follows by the fact that replacing $\hat{y}$ with $h(\mathbf{x}_j)$ does not decrease the $\inf$ term; (f) is due to $\sup\mathbb{E} \leq \mathbb{E}\sup$; (g) is due to symmetries of g_j, g_j' and ϵ_j is uniform over $\{-1,1\}$; (h) follows by $\sup(A+B+C) \leq \sup(A/2+B) + \sup(A/2+C) = (\sup(A+2B) + \sup(A+2C))/2$, the linearity of expectation and symmetries of B, C; (i) follows by that the inner expectation takes the

same value for all $g_j \in \{-L, L\}$ and therefore the $\sup_{d_j} \mathbb{E}_{g_j \sim d_j}$ can be eliminated. This completes the proof.

Appendix B. Proof of Lemma 6

Denote $\mathbb{Q}_j \equiv \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}_1} \sup_{y_1} \mathbb{E}_{\hat{y}_1} \cdots \mathbb{E}_{\mathbf{x}_j} \sup_{y_j} \mathbb{E}_{\hat{y}_j}$ for notation convenience. We have

$$
\begin{aligned}
\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H}, \Phi) &\overset{(a)}{=} \mathbb{Q}_M \left[\sum_{j=1}^{M} \ell(\hat{y}_j, y_j) + R_M \right] \\
&\overset{(b)}{=} \mathbb{Q}_{M-1} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, y_j) + \mathbb{E}_{\mathbf{x}_M} \sup_{y_M} \mathbb{E}_{\hat{y}_M} \left[\ell(\hat{y}_M, y_M) + R_M \right] \right] \\
&\overset{(c)}{\leq} \mathbb{Q}_{M-1} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, y_j) + \tilde{R}_{M-1} \right] \\
&= \mathbb{Q}_{M-1} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, y_j) + R_{M-1} + (\tilde{R}_{M-1} - R_{M-1}) \right] \\
&\overset{(d)}{\leq} \mathbb{Q}_{M-1} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, y_j) + R_{M-1} \right] + \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}^{M-1}} \sup_{y^{M-1}} (\tilde{R}_{M-1} - R_{M-1}) \\
&\overset{(e)}{\leq} \mathbb{E}_{\mathbf{x}^0_{-N+1}}[\tilde{R}_0] + \sum_{j=1}^{M-1} \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}^j} \sup_{y^j} (\tilde{R}_j - R_j),
\end{aligned}
$$

where (a) follows by definition of R_M; (b) follows by extracting the last layer of $\mathbb{Q}_M$; (c) follows by Lemma 5 and noticing that $\hat{y}_j$ has the same randomness as R_j; (d) follows by the the facts that $\sup(A + B) \leq \sup A + \sup B$, $\sup \mathbb{E} \leq \mathbb{E} \sup$, the linearity of expectation and $\tilde{R}_{M-1} - R_{M-1}$ is independent of $\hat{y}_j$ for all $j \leq M - 1$; (e) follows by repeating the same arguments for another $M - 1$ steps. This completes the proof.

Appendix C. Proof of Lemma 9

We have

$$
\begin{aligned}
\mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}^j} \sup_{y^j} (\tilde{R}_j - R_j) &\overset{(a)}{=} \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}^j} \sup_{y^j} \mathbb{E}_{\tilde{\mathbf{x}}^M_{j+1}, \epsilon^M_{j+1}} \mathbb{E}_{\mathbf{x} \sim \mu} [f_{\mathbf{z}_j, y^j}(\mathbf{x}) - f_{\mathbf{z}_j, y^j}(\tilde{\mathbf{x}}_{j+1})] \\
&\leq \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}^j} \mathbb{E}_{\tilde{\mathbf{x}}^M_{j+2}, \epsilon^M_{j+1}} \sup_{y^j} \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}, \mathbf{x}} [f_{\mathbf{z}_j, y^j}(\mathbf{x}) - f_{\mathbf{z}_j, y^j}(\tilde{\mathbf{x}}_{j+1})] \\
&\overset{(b)}{=} \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{z}_j} \sup_{y^j} \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}, \mathbf{x}} [f_{\mathbf{z}_j, y^j}(\mathbf{x}) - f_{\mathbf{z}_j, y^j}(\tilde{\mathbf{x}}_{j+1})] \\
&= \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{z}_j} \sup_{y^j} [\mathbb{E}_{\mathbf{x} \sim \mu} [f_{\mathbf{z}_j, y^j}(\mathbf{x})] - \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}} [f_{\mathbf{z}_j, y^j}(\tilde{\mathbf{x}}_{j+1})]] \\
&\overset{(c)}{=} \mathbb{E}_{\mathbf{z}_j} \mathbb{E}_{\mathbf{x}^{-N+B}_{-N+1}} \sup_{y^j} \left[\mathbb{E}_{\mathbf{x} \sim \mu} [f_{\mathbf{z}_j, y^j}(\mathbf{x})] - \frac{1}{B} \sum_{i=1}^{B} f_{\mathbf{z}_j, y^j}(\mathbf{x}_{-N+i}) \right]
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(d)}{=} \mathbb{E}_{\mathbf{z}_j}\mathbb{E}_{\mathbf{x}_{-N+1}^{-N+B}} \sup_{y^j}\left[\frac{1}{B}\sum_{i=1}^{B}\mathbb{E}_{\mathbf{x}_i'\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x}_i')] - f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})\right]\\
&= \mathbb{E}_{\mathbf{z}_j}\mathbb{E}_{\mathbf{x}_{-N+1}^{-N+B}} \sup_{y^j}\mathbb{E}_{\mathbf{x}'^B\sim\mu^{\otimes B}}\left[\frac{1}{B}\sum_{i=1}^{B} f_{\mathbf{z}_j,y^j}(\mathbf{x}_i') - f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})\right]\\
&\le \mathbb{E}_{\mathbf{z}_j}\mathbb{E}_{\mathbf{x}_{-N+1}^{-N+B}}\mathbb{E}_{\mathbf{x}'^B} \sup_{y^j}\left[\frac{1}{B}\sum_{i=1}^{B} f_{\mathbf{z}_j,y^j}(\mathbf{x}_i') - f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})\right]\\
&\overset{(e)}{=} \mathbb{E}_{\mathbf{z}_j}\mathbb{E}_{\mathbf{x}_{-N+1}^{-N+B}}\mathbb{E}_{\mathbf{x}'^B}\mathbb{E}_{\epsilon'^B} \sup_{y^j}\left[\frac{1}{B}\sum_{i=1}^{B} \epsilon_j'(f_{\mathbf{z}_j,y^j}(\mathbf{x}_i') - f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i}))\right]\\
&\le \sup_{\mathbf{x}_{-N+1}^{-N+B},\mathbf{x}'^B,\mathbf{z}_j}\mathbb{E}_{\epsilon'^B} \sup_{y^j}\left[\frac{1}{B}\sum_{i=1}^{B} \epsilon_j'(f_{\mathbf{z}_j,y^j}(\mathbf{x}_i') - f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i}))\right]
\end{aligned}
$$

where (a) follows by Fact 1 (in Section 3.1); (b) follows by definition of $\mathbf{z}_j$; (c) follows by Fact 4 below and taking $B = N-M+j+1$; (d) follows by introducing B fresh $i.i.d.$ samples $\mathbf{x}'^B \sim \mu^{\otimes B}$; (e) follows by symmetries of $\mathbf{x}'^B$ and $\mathbf{x}_{-N+1}^{-N+B}$ (which are independent of $\mathbf{z}_j$) and introducing the $i.i.d.$ random variables ϵ'^B uniform over $\{-1,1\}^B$;

Fact 4 *Let $B = N - M + j + 1$, then*

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{x}_{-N+1}^0}\mathbb{E}_{\mathbf{z}_j} \sup_{y^j}[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}}[f_{\mathbf{z}_j,y^j}(\tilde{\mathbf{x}}_{j+1})]]\\
&\qquad\qquad = \mathbb{E}_{\mathbf{z}_j}\mathbb{E}_{\mathbf{x}_{-N+1}^{-N+B}} \sup_{y^j}\left[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \frac{1}{B}\sum_{i=1}^{B} f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})\right].
\end{aligned}
$$

Proof Note that $\mathbf{z}_j = (\mathbf{x}^j, \tilde{\mathbf{x}}_{j+2}^M, \epsilon_{j+1}^M)$, where $\tilde{\mathbf{x}}_{j+1}^M$ are sampled uniformly from $\mathbf{x}_{-N+1}^0$ *without replacement*, and $\mathbf{x}^j$, ϵ_{j+1}^M are independent of $\mathbf{x}_{-N+1}^0$. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{x}_{-N+1}^0}\mathbb{E}_{\mathbf{z}_j} \sup_{y^j}[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}}[f_{\mathbf{z}_j,y^j}(\tilde{\mathbf{x}}_{j+1})]]\\
&\qquad = \mathbb{E}_{\mathbf{x}^j,\epsilon_{j+1}^M}\mathbb{E}_{\mathbf{x}_{-N+1}^0}\mathbb{E}_{\tilde{\mathbf{x}}_{j+2}^M} \sup_{y^j}[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \mathbb{E}_{\tilde{\mathbf{x}}_{j+1}}[f_{\mathbf{z}_j,y^j}(\tilde{\mathbf{x}}_{j+1})]]\\
&\qquad \overset{(\star)}{=} \mathbb{E}_{\mathbf{x}^j,\epsilon_{j+1}^M}\mathbb{E}_{\mathbf{x}_{-N+1}^0}\mathbb{E}_I \sup_{y^j}\left[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j,y^j}(\mathbf{x})] - \frac{1}{B}\sum_{i\in[N]\backslash I} f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i})\right]
\end{aligned}
$$

where the key step $(\star)$ follows by noticing that the randomness of $\tilde{\mathbf{x}}_{j+2}^M$ is equivalent to selecting a *random* index set $I \subset [N]$ uniformly with size $|I| = M - j - 1$ and the index of $\tilde{\mathbf{x}}_{j+1}$ (in $\mathbf{x}_{-N+1}^0$) is then uniform over $[N]\backslash I$ [6], where the size of $[N]\backslash I$ is $B = N - M + j + 1$; Therefore,

$$
\mathbb{E}_{\tilde{\mathbf{x}}_{j+1}}[f_{\mathbf{z}_j,y^j}(\tilde{\mathbf{x}}_{j+1})] = \frac{1}{B}\sum_{i\in[N]\backslash I} f_{\mathbf{z}_j,y^j}(\mathbf{x}_{-N+i}).
$$

6. By the definition of sampling *without replacement*.

Note that $\mathbf{x}^0_{-N+1}$ is an $i.i.d.$ sample, by *symmetries*, we can *fix* $I = \{B+1, \cdots, N\}$ (i.e., we take $\tilde{\mathbf{x}}^M_{j+2}$ being $\mathbf{x}^0_{-N+B+1}$) and therefore $\tilde{\mathbf{x}}^M_{j+2}$ can be decoupled from $\mathbf{x}^{-N+B}_{-N+1}$, leading to

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}^j, \epsilon^M_{j+1}} \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_I \sup_{y^j} \left[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j, y^j}(\mathbf{x})] - \frac{1}{B} \sum_{i\in[N]\setminus I} f_{\mathbf{z}_j, y^j}(\mathbf{x}_{-N+i}) \right] \\
&\qquad = \mathbb{E}_{\mathbf{x}^j, \epsilon^M_{j+1}} \mathbb{E}_{\tilde{\mathbf{x}}^M_{j+2}} \mathbb{E}_{\mathbf{x}^{-N+B}_{-N+1}} \sup_{y^j} \left[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j, y^j}(\mathbf{x})] - \frac{1}{B} \sum_{i=1}^{B} f_{\mathbf{z}_j, y^j}(\mathbf{x}_{-N+i}) \right] \\
&\qquad = \mathbb{E}_{\mathbf{z}_j} \mathbb{E}_{\mathbf{x}^{-N+B}_{-N+1}} \sup_{y^j} \left[\mathbb{E}_{\mathbf{x}\sim\mu}[f_{\mathbf{z}_j, y^j}(\mathbf{x})] - \frac{1}{B} \sum_{i=1}^{B} f_{\mathbf{z}_j, y^j}(\mathbf{x}_{-N+i}) \right].
\end{aligned}$$

This completes the proof of the Fact. ■

Appendix D. Omitted Proofs

In this section, we collect all other proofs that are omitted from the main text.

Proof [Proof of Proposition 8] Denote

$$F(h) = 2L \sum_{i=j+2}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L^h_j.$$

Let $\hat{h} = \arg\max_{h\in\mathcal{H}} F(h)$ (find an approximation if necessary). We claim that for any $\mathbf{x} \in \mathcal{X}$,

$$F(\hat{h}) - 2L \le \sup_{h\in\mathcal{H}} \{2\epsilon_{j+1} L h(\mathbf{x}) + F(h)\} \le F(\hat{h}) + 2L.$$

This will complete the proof of the first part. To see the upper bound, we have

$$\sup_{h\in\mathcal{H}} \{2\epsilon_{j+1} L h(\mathbf{x}) + F(h)\} \le \sup_h \{2\epsilon_{j+1} L h(\mathbf{x})\} + \sup_h F(h) \le 2L + F(\hat{h}),$$

since $h(\mathbf{x}) \in [0,1]$. For the lower bound, we have

$$\sup_{h\in\mathcal{H}} \{2\epsilon_{j+1} L h(\mathbf{x}) + F(h)\} \ge 2\epsilon_{j+1} L \hat{h}(\mathbf{x}) + F(\hat{h}) \ge F(\hat{h}) - 2L,$$

since $\sup$ do not increase by replacing h with any specific $\hat{h}$ and $\hat{h}(\mathbf{x}) \in [0,1]$.

To prove the second part, for any given $h \in \mathcal{H}$, we denote

$$g_h(y^j) = 2L\epsilon_{j+1} h(\mathbf{x}) + 2L \sum_{i=j+2}^{M} \epsilon_i h(\tilde{\mathbf{x}}_i) - L^h_j.$$

Note that, y^j only appears in the L^h_j term. By definition of L^h_j and L-Lipschitz property of the loss ℓ, we have

$$\forall h \in \mathcal{H},\ |g_h(y^j) - g_h(y'^j)| \le jL||y^j - y'^j||_\infty.$$

Let $\hat{h} = \arg\max_h g_h(y^j)$, we have

$$\sup_h g_h(y^j) - \sup_h g_h(y'^j) \le g_{\hat{h}}(y^j) - g_{\hat{h}}(y'^j) \le jL||y^j - y'^j||_\infty.$$

Let $\hat{h}' = \arg\max_h g_h(y'^j)$, we have

$$\sup_h g_h(y^j) - \sup_h g_h(y'^j) \ge g_{\hat{h}'}(y^j) - g_{\hat{h}'}(y'^j) \ge -jL||y^j - y'^j||_\infty.$$

The proposition follows by noticing that $f_{\mathbf{z}_j,y^j}(\mathbf{x}) = \sup_h g_h(y^j)$. ■

Proof [Proof of Fact 2] Note that $f_{\mathbf{z}_j,y^j}(\mathbf{x}) = \sup_h\{2h(\mathbf{x}) + F(h)\}$. If $h^0(\mathbf{x}) = 1$, then $\exists h \in \mathcal{H}^0$ such that $h(\mathbf{x}) = 1$ and $F(h) = F(\hat{h})$, thus $f_{\mathbf{z}_j,y^j}(\mathbf{x}) \ge 2+F(\hat{h})$. Clearly, we also have $f_{\mathbf{z}_j,y^j}(\mathbf{x}) \le 2\sup_h h(\mathbf{x})+\sup_h F(h) \le 2+F(\hat{h})$, the first case follows. If $h^0(\mathbf{x}) = 0$ and $h^1(\mathbf{x}) = 1$, then there exists $h \in \mathcal{H}^1$ such that $h(\mathbf{x}) = 1$ and $F(h) = F(\hat{h})-1$, thus $f_{\mathbf{z}_j,y^j}(\mathbf{x}) \ge F(\hat{h})-1+2 = F(\hat{h})+1$. On the other-hand, since $h^0(\mathbf{x}) = 0$, we have for all $h \in \mathcal{H}^0$, $2h(\mathbf{x}) + F(h) = F(\hat{h})$. For any other $h \notin \mathcal{H}^0 \cup \mathcal{H}^1$, we have $2h(\mathbf{x}) + F(h) \le 2 + F(\hat{h}) - 2 = F(\hat{h})$. Therefore, $f_{\mathbf{z}_j,y^j}(\mathbf{x}) \le F(\hat{h}) + 1$, this completes the second case. Finally, if both $h^0(\mathbf{x}) = h^1(\mathbf{x}) = 0$, we have for any $h \in \mathcal{H}^0$, $2h(\mathbf{x}) + F(h) = F(\hat{h})$, i.e., $f_{\mathbf{z}_j,y^j}(\mathbf{x}) \ge F(\hat{h})$. Moreover, for any $h \notin \mathcal{H}^0$, it is easy to verify that $2h(\mathbf{x}) + F(h) \le F(\hat{h})$. This completes the proof. ■

Proof [Proof of Theorem 14] Assume, w.o.l.g., $\epsilon_{j+1} = 1$. The functions h^0, h^1 as in Fact 2 are within $\mathcal{F}^{\mathsf{u}}$. For *any* $\mathbf{x}^{2N} \in \mathcal{X}^{2N}$ and $\hat{\mu}$ uniform over $\mathbf{x}^{2N}$, there exists a γ-cover $\mathcal{C}_\gamma$ of $\mathcal{F}^{\mathsf{u}}$ under distance $d_{\hat{\mu}}(f_1, f_2) \stackrel{\text{def}}{=} \Pr_{\mathbf{x}\sim\hat{\mu}}[f_1(\mathbf{x}) \ne f_2(\mathbf{x})]$ such that $|\mathcal{C}_\gamma| \le O(\frac{1}{\gamma^{\mathsf{VC}(\mathcal{F}^{\mathsf{u}})}})$ (Haussler, 1995). By Fact 2, there exists a function $\mathcal{T} : (\mathcal{F}^{\mathsf{u}})^2 \to \{0, 1, 2\}^{\mathcal{X}}$ such that for *any* $f_{\mathbf{z}_j,y^j}$, there exist $h^0, h^1 \in \mathcal{F}^{\mathsf{u}}$ such that $f_{\mathbf{z}_j,y^j}(\mathbf{x}) = \mathcal{T}(h^0(\mathbf{x}), h^1(\mathbf{x})) + c_{\mathbf{z}_j,y^j}$, where $c_{\mathbf{z}_j,y^j} = F(\hat{h})$ as in Fact 2. Therefore, the function class $\mathcal{C}' \stackrel{\text{def}}{=} \{\mathcal{T}(h^0, h^1) : h^0, h^1 \in \mathcal{C}_\gamma\}$ forms a 2γ-cover of $\{(f_{\mathbf{z}_j,y^j}(\mathbf{x}) - c_{\mathbf{z}_j,y^j}) : y^j \in [0, 1]^j\}$ under distance $d_{\hat{\mu}}(f_1, f_2)$ and $|\mathcal{C}'| \le O(\frac{1}{\gamma^{2\mathsf{VC}(\mathcal{F}^{\mathsf{u}})}})$. This implies that the function class $\mathcal{C}'' \stackrel{\text{def}}{=} \{g(\mathbf{x}', \mathbf{x}) = f(\mathbf{x}') - f(\mathbf{x}) : f \in \mathcal{C}',\ (\mathbf{x}', \mathbf{x}) \in \mathcal{X}^2\}$ forms a 4γ-cover of

$$\mathcal{G}_{\mathbf{z}_j} = \{g_{\mathbf{z}_j,y^j}(\mathbf{x}', \mathbf{x}) = f_{\mathbf{z}_j,y^j}(\mathbf{x}') - f_{\mathbf{z}_j,y^j}(\mathbf{x}) : y^j \in [0, 1]^j,\ (\mathbf{x}', \mathbf{x}) \in \mathcal{X}^2\}$$

under distance $d_{\hat{\nu}}(g_1, g_2) = \Pr_{(\mathbf{x}',\mathbf{x})\sim\hat{\nu}}[g_1(\mathbf{x}', \mathbf{x}) \ne g_2(\mathbf{x}', \mathbf{x})]$ for any distribution $\hat{\nu}$ uniform over a fixed *pairing* of $\mathbf{x}^{2N}$ and $|\mathcal{C}''| \le O(\frac{1}{\gamma^{2\mathsf{VC}(\mathcal{F}^{\mathsf{u}})}})$. We have by the chaining argument (Wainwright, 2019, Example 5.24) that $\mathsf{Rad}_N(\mathcal{G}_{\mathbf{z}_j}) \le O(\sqrt{\mathsf{VC}(\mathcal{F}^{\mathsf{u}})N})$. This implies by Lemma 6 & 9 that

$$\tilde{r}^{\mathsf{side}}_{M,N}(\mathcal{H}, \Phi) \le O\left(\sqrt{\mathsf{VC}(\mathcal{H})M} + \frac{M\sqrt{\mathsf{VC}(\mathcal{F}^{\mathsf{u}})}}{\sqrt{N}}\right). \tag{15}$$

Taking $M(n) = 1.5^n$ in (13), we have $N = S(n) = 2 \cdot 1.5^n - 3$, which ensures $M(n) \le S(n)/2 + O(1)$ (as required for (15) to hold). Invoking Lemma 13, we conclude

$$\tilde{r}_T(\mathcal{H}, \Psi) \le O(\sqrt{\mathsf{VC}(\mathcal{H})} + \sqrt{\mathsf{VC}(\mathcal{F}^{\mathsf{u}})}) \sum_{n=1}^{\lceil \log_{1.5}(T)\rceil} 1.5^{n/2} \le O(\sqrt{\mathsf{VC}(\mathcal{F}^{\mathsf{u}})T}),$$

where the last inequality follows by $\mathcal{H} \subset \mathcal{F}^{\mathsf{u}}$. This completes the proof and the case for $\epsilon_{j+1} = -1$ is symmetric with $\mathcal{F}^{\mathsf{i}}$. ■

Appendix E. Analysis of Contextual K-arm Bandits

In this appendix, we provide a detailed analysis of the contextual K-arm bandit problem with the contexts generated by an *unknown* $i.i.d.$ process and costs generated adversarially, as defined in Section 2. Let $\mathcal{H} \subset [K]^{\mathcal{X}}$ be a finite policy set. Following the same steps as in the online learning case, we first consider the scenario with side-information as in Section 3.1. Let $(\mathbf{x}_1, c_1), \cdots, (\mathbf{x}_M, c_M)$ be any realization of the feature-loss pairs and $\mathbf{x}^0_{-N+1}$ be the side-information with $\mathbf{x}^M_{-N+1}$ sampled $i.i.d.$ from an (unknown) distribution μ. We consider the following *surrogate* relaxation:

$$R_j = \mathbb{E}_{\tilde{\mathbf{x}},\epsilon,Z}\left[-\inf_{h\in\mathcal{H}}\left(2\epsilon_{j+1}[h(\tilde{\mathbf{x}}_{j+1})]Z_{j+1} + \sum_{i=1}^{j}\hat{c}_i[h(\mathbf{x}_i)] + \sum_{i=j+2}^{M} 2\epsilon_i[h(\tilde{\mathbf{x}}_i)]Z_i\right) + \gamma(M-j)K\right],$$

where ϵ_is are (vectors) $i.i.d.$ uniform over $\{\pm 1\}^K$, $\tilde{\mathbf{x}}_i$s are sampled from $\hat{\mu}_N$ (i.e., the empirical distribution over $\mathbf{x}^0_{-N+1}$) *without replacement* and Z_is are $i.i.d.$ with $Z_i \in \{0, \frac{1}{\gamma}\}$ such that $\Pr[Z_i = \frac{1}{\gamma}] = \gamma K$ and γ is a parameter that needs to be tuned. Moreover, $\hat{c}_i$ is a random vector constructed from the prediction $\hat{y}_i \in [K]$ and distribution q_i (see construction below) as $\hat{c}_i = \frac{1}{\gamma} I_i e_{\hat{y}_i}$, where e_k is the standard base of $\mathbb{R}^K$ with coordinate k being 1 and I_i is the indicator that takes value 1 w.p. $\frac{\gamma c_i[\hat{y}_i]}{q_i[\hat{y}_i]}$. Similarly, we define the variational relaxation as

$$\tilde{R}_j = \mathbb{E}_{\mathbf{x}\sim\mu}\mathbb{E}_{\tilde{\mathbf{x}},\epsilon,Z}\left[-\inf_{h\in\mathcal{H}}\left(2\epsilon_{j+1}[h(\mathbf{x})]Z_{j+1} + \sum_{i=1}^{j}\hat{c}_i[h(\mathbf{x}_i)] + \sum_{i=j+2}^{M} 2\epsilon_i[h(\tilde{\mathbf{x}}_i)]Z_i\right) + \gamma(M-j)K\right].$$

Prediction rule. The prediction rule at step j is given as follows:

1. Sample the data $\tilde{\mathbf{x}}_i$s, ϵ_is and Z_is as in the definition of R_j;
2. Let $\mathcal{D}_K$ be the class of distribution over $[K]$. Find

$$\hat{q}_j = \arg\min_{q\in\mathcal{D}_K}\sup_{p_j\in\mathcal{D}'}\mathbb{E}_{\hat{c}_j\sim p_j}\left\{\langle q, \hat{c}_j\rangle - \inf_{h\in\mathcal{H}}\left(\sum_{i=1}^{j}\hat{c}_i[h(\mathbf{x}_i)] + \sum_{i=j+1}^{M} 2\epsilon_i[h(\tilde{\mathbf{x}}_i)]Z_i\right) + \gamma(M-j)K\right\},$$

   where $\mathcal{D}'$ is the class of distributions over $\{\frac{1}{\gamma}e_k : k \in [K]\} \cup \{\mathbf{0}\}$ s.t. $\forall k \in [K], p[k] \le \gamma$ and $\langle q, \hat{c}_j\rangle$ is the scalar product;
3. Define $q_j = (1-\gamma K)\hat{q}_j + \gamma\mathbf{1}$ and make prediction $\hat{y}_j \sim q_j$.

Note that, the prediction rule is iterative in the sense that the predictions q_js and $\hat{y}_j$s are used to construct $\hat{c}_j$s for the predictors in the following steps.

Let Φ be the prediction rule as define above. The bandit minimax regret with side-information is defined as follows:

$$\tilde{r}^{\mathsf{bandit}}_{M,N}(\mathcal{H},\Phi) = \sup_{\mu}\mathbb{E}_{\mathbf{x}^0_{-N+1}}\mathbb{E}_{\mathbf{x}_1}\sup_{c_1}\cdots\mathbb{E}_{\mathbf{x}_M}\sup_{c_M}\mathbb{E}_{\hat{y}^M}\left[\sum_{j=1}^{M}\langle q_j, c_j\rangle - \inf_{h\in\mathcal{H}}\sum_{j=1}^{M} c_j[h(\mathbf{x}_j)]\right], \quad (16)$$

where $\mathbb{E}_{\hat{y}_j}$ is over all the internal randomness used to construct the predictor, including $\tilde{\mathbf{x}}$, ϵ, Z, I_j and the randomness of $\hat{y}_j \sim q_j$. It is important to note that the selection of costs c_js is oblivious to

the predictions $\hat{y}_j$s [7], although it still (adaptively) depends on the contexts $\mathbf{x}_j$s, which differs from the online learning case as in (1).

We have the following bandit version of *approx-admissibility*.

Lemma 20 *For the predictors q_js and $\hat{y}_j$s, we have*

$$\mathbb{E}_{\mathbf{x}_j} \sup_{c_j} \mathbb{E}_{\hat{y}_j}[c_j[\hat{y}_j] + R_j] \le \tilde{R}_{j-1}.$$

Proof [Sketch] We sketch only the high-level ideas. The proof essentially follows the same arguments as in Lemma 5 and the admissibility proof of Syrgkanis et al. (2016, Theorem 3), by noticing that the primary distinction between our R_j and the relaxation described in Syrgkanis et al. (2016) is the randomness of $\tilde{\mathbf{x}}_i$s. Their entire proof of admissibility is conducted by *conditioning* on such randomness, which only enters in the final step. Therefore, it can be applied in parallel to our case, as in the proof of Lemma 5. ■

Similarly, we have the following decomposition as in Lemma 6.

Lemma 21 *For the predictor Φ and any given μ, the bandit minimax regret with side-information is upper bounded by*

$$\tilde{r}^{\mathsf{bandit}}_{M,N}(\mathcal{H}, \Phi) \le \mathbb{E}_{\mathbf{x}^0_{-N+1}} \left[\tilde{R}_0 + \sum_{j=1}^{M-1} \mathbb{E}_{\mathbf{x}^j, \hat{y}^j} \sup_{c^j} (\tilde{R}_j - R_j) \right].$$

Proof We first observe that $\hat{c}_j$ is a unbiased estimation of c_j for all $j \in [M]$, i.e., $\mathbb{E}_{\hat{y}_j}[\hat{c}_j[k]] = c_j[k]$ for all $k \in [K]$. Therefore,

$$\begin{aligned}
\mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^M \langle q_j, c_j \rangle - \inf_{h \in \mathcal{H}} \sum_{j=1}^M c_j[h(\mathbf{x}_j)] \right] &= \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^M \langle q_j, c_j \rangle \right] - \inf_{h \in \mathcal{H}} \sum_{j=1}^M c_j[h(\mathbf{x}_j)] \\
&= \sup_{h \in \mathcal{H}} \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^M \langle q_j, c_j \rangle \right] - \sum_{j=1}^M c_j[h(\mathbf{x}_j)] \\
&\overset{(a)}{=} \sup_{h \in \mathcal{H}} \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^M \langle q_j, c_j \rangle - \hat{c}_j[h(\mathbf{x}_j)] \right] \\
&\le \mathbb{E}_{\hat{y}^M} \sup_{h \in \mathcal{H}} \left[\sum_{j=1}^M \langle q_j, c_j \rangle - \hat{c}_j[h(\mathbf{x}_j)] \right] \\
&= \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^M \langle q_j, c_j \rangle - \inf_{h \in \mathcal{H}} \sum_{j=1}^M \hat{c}_j[h(\mathbf{x}_j)] \right].
\end{aligned}$$

7. This assumption was also made implicitly in prior literature for contextual bandits, such as Rakhlin and Sridharan (2016); Syrgkanis et al. (2016); Banihashem et al. (2023).

where (a) follows by that $\hat{c}_j$s are unbiased estimations of c_js. This implies that

$$\tilde{r}^{\mathsf{bandit}}_{M,N}(\mathcal{H},\Phi) \le \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}_1} \sup_{c_1} \cdots \mathbb{E}_{\mathbf{x}_M} \sup_{c_M} \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^{M} \langle q_j, c_j \rangle + R_M \right]$$
$$\le \mathbb{E}_{\mathbf{x}^0_{-N+1}} \mathbb{E}_{\mathbf{x}_1} \sup_{c_1} \mathbb{E}_{\hat{y}_1} \cdots \mathbb{E}_{\mathbf{x}_M} \sup_{c_M} \mathbb{E}_{\hat{y}_M} \left[\sum_{j=1}^{M} c_j[\hat{y}_j] + R_M \right],$$

where the second inequality follows by $\sup \mathbb{E} \le \mathbb{E} \sup$ and $\mathbb{E}_{\hat{y}_j}[c_j[\hat{y}_j]] = \langle q_j, c_j \rangle$. The lemma then follows from Lemma 20 and the same argument as in the proof of Lemma 6. ■

In order to analyze the discrepancy between R_j and $\tilde{R}_j$, we define the following function:

$$f_{\mathbf{z}_j,\hat{c}^j}(\mathbf{x}) = \mathbb{E}_{Z_{j+1}} \left[\inf_{h\in\mathcal{H}} \left(2\epsilon_{j+1}[h(\mathbf{x})]Z_{j+1} + \sum_{i=1}^{j} \hat{c}_i[h(\mathbf{x}_i)] + \sum_{i=j+2}^{M} 2\epsilon_i[h(\tilde{\mathbf{x}}_i)]Z_i \right) \right],$$

where $\mathbf{z}_j$ is composed of all other variables that define R_j except Z_{j+1} and $\hat{c}^j$.

The following key property bounds the sensitivity of $f_{\mathbf{z}_j,\hat{c}^j}$:

Lemma 22 *We have for any* $\mathbf{z}_j, \hat{c}^j$ *and* $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$

$$|f_{\mathbf{z}_j,\hat{c}^j}(\mathbf{x}) - f_{\mathbf{z}_j,\hat{c}^j}(\mathbf{x}')| \le 4K.$$

Proof Clearly, if $Z_{j+1} = 0$ then the sensitivity is 0, else, $Z_{j+1} = \frac{1}{\gamma}$ and the sensitivity is upper bounded by $\frac{4}{\gamma}$ using the same argument in the proof of Proposition 8. Since $Z_{j+1} \neq 0$ happens w.p. $\le \gamma K$, the expected function has sensitivity upper bounded by $\frac{4}{\gamma} \times \gamma K = 4K$. ■

We now observe that

$$\mathbb{E}_{\mathbf{x}^j,\hat{y}^j} \sup_{c^j}(\tilde{R}_j - R_j) \le \mathbb{E}_{\mathbf{x}^j} \sup_{\hat{c}^j}(\tilde{R}_j - R_j),$$

since R_j and $\tilde{R}_j$ depend only on $\hat{c}_j$. Notably, the support set size of $\hat{c}_j$s is *finite* and upper bounded by $K+1$. Invoking Lemma 9, Lemma 22 and a simple application of Massart's lemma gives

$$\mathbb{E}_{\mathbf{x}^j_{-N+1},\hat{y}^j} \sup_{c_j}(\tilde{R}_j - R_j) \le O\left(\sqrt{\frac{K^2 j \log K}{N}}\right). \tag{17}$$

Note that this upper bound is *independent* of the parameter γ. Invoking (Syrgkanis et al., 2016, Theorem 3) and taking $\gamma = (\log|\mathcal{H}|/KM)^{\frac{1}{3}}$, we find $\tilde{R}_0 \le O((KM)^{\frac{2}{3}}(\log|\mathcal{H}|)^{\frac{1}{3}})$. This, together with Lemma 21 and (17), implies that

$$\tilde{r}^{\mathsf{bandit}}_{M,N}(\mathcal{H},\Phi) \le O\left((KM)^{\frac{2}{3}}(\log|\mathcal{H}|)^{\frac{1}{3}} + K\sqrt{\frac{M^3 \log K}{N}}\right).$$

Setting $M(n) = n^{\frac{3}{2}}$ and using a similar argument as Lemma 13, we arrive at our main result of this section

$$\tilde{r}^{\mathsf{bandit}}_{T}(\mathcal{H},\Phi) \le O\left((K^{\frac{2}{3}}(\log|\mathcal{H}|)^{\frac{1}{3}} + K\sqrt{\log K}) \cdot T^{\frac{4}{5}}\right),$$

where Φ can be computed *efficiently* by accessing to an ERM oracle (with computational cost same as Syrgkanis et al. (2016)).

Appendix F. Oblivious Adversaries

In this section, we provide the regret analysis for online learning against an *oblivious* adversary as introduced in Section 2. We follow the same online learning game as in (1) with the exception that the adversary fixes functions $f_1, \cdots, f_T : \mathcal{X} \to [0,1]$ before the game and sets the adversary labels $y_t = f_t(\mathbf{x}_t)$ for each time step $t \in [T]$. Formally, for any expert class $\mathcal{H}$ and prediction rule Φ, we are interested in the following *oblivious* minimax regret:

$$\tilde{r}_T^{\mathsf{ob}}(\mathcal{H}, \Phi) = \sup_{f_1, \cdots, f_T \in [0,1]^{\mathcal{X}}} \sup_{\mu} \mathbb{E}_{\mathbf{x}^T} \mathbb{E}_{\hat{y}^T} \left[\sum_{t=1}^{T} \ell(\hat{y}_t, f_t(\mathbf{x}_t)) - \inf_{h \in \mathcal{H}} \sum_{t=1}^{T} \ell(h(\mathbf{x}_t), f_t(\mathbf{x}_t)) \right],$$

where $\mathbf{x}^T$ are sampled $i.i.d.$ from μ and $\hat{y}_t \sim \Phi(\mathbf{x}^t, y^{t-1})$ for $t \in [T]$. For the clarity of presentation, we assume that $\ell(\hat{y}, y) = |\hat{y} - y|$ is the absolute loss. We now ready to state the main result of this appendix:

Theorem 23 *Let $\mathcal{H} \subset [0,1]^{\mathcal{X}}$ be a class of Rademacher complexity $\mathsf{Rad}_T(\mathcal{H}) = O(T^q)$ for some $q \in [\frac{1}{2}, 1]$ and ℓ be the absolute loss. Then there exists an oracle-efficient prediction rule Φ with at most $O(\sqrt{T}\log T)$ calls to the ERM oracle per round, such that $\tilde{r}_T^{\mathsf{ob}}(\mathcal{H}, \Phi) \le O(T^q)$. In particular, for finite-VC class $\mathcal{H}$, we have $\tilde{r}_T^{\mathsf{ob}}(\mathcal{H}, \Phi) \le O(\sqrt{\mathsf{VC}(\mathcal{H})T})$. For a class $\mathcal{H}$ with α-fat shattering dimension $O(\alpha^{-p})$ for some $p > 0$, we have $\tilde{r}_T^{\mathsf{ob}}(\mathcal{H}, \Phi) \le \tilde{O}(T^{\max\{\frac{1}{2}, \frac{p-1}{p}\}})$.*

Proof We will follow the same path as the regret analysis for the *non-oblivious* adversaries as established in Section 3. We first consider the scenario with side-information $\mathbf{x}_{-N+1}^0$, and define for any predictor Φ the following oblivious minimax regret with side-information:

$$\tilde{r}_{M,N}^{\mathsf{ob,side}}(\mathcal{H}, \Phi) = \sup_{f_1, \cdots, f_M \in [0,1]^{\mathcal{X}}} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^M} \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^{M} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) - \inf_{h \in \mathcal{H}} \sum_{j=1}^{M} \ell(h(\mathbf{x}_j), f_j(\mathbf{x}_j)) \right],$$

where $\mathbf{x}_{-N+1}^M$ are sampled $i.i.d.$ from μ. Let Φ be the predictor as in (2) and R_j and $\tilde{R}_j$ be the same *surrogate* relaxations as in (4) and (5). We claim that:

$$\tilde{r}_{M,N}^{\mathsf{ob,side}}(\mathcal{H}, \Phi) \le \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^0} \left[\tilde{R}_0 + \sum_{j=1}^{M-1} \mathbb{E}_{\mathbf{x}^j} [\tilde{R}_j - R_j] \right]. \tag{18}$$

To see this, we find

$$\begin{aligned}
\tilde{r}_{M,N}^{\mathsf{ob,side}}(\mathcal{H}, \Phi) &= \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^M} \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^{M} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) - \inf_{h \in \mathcal{H}} \sum_{j=1}^{M} \ell(h(\mathbf{x}_j), f_j(\mathbf{x}_j)) \right] \\
&= \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^M} \mathbb{E}_{\hat{y}^M} \left[\sum_{j=1}^{M} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) + R_M \right] \\
&= \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^{M-1}} \mathbb{E}_{\hat{y}^{M-1}} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) + \mathbb{E}_{\mathbf{x}_M} \mathbb{E}_{\hat{y}_M} [\ell(\hat{y}_M, f_M(\mathbf{x}_M)) + R_M] \right]
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^{M-1}} \mathbb{E}_{\hat{y}^{M-1}} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) + \mathbb{E}_{\mathbf{x}_M} \sup_{y_M} \mathbb{E}_{\hat{y}_M} [\ell(\hat{y}_M, y_M) + R_M] \right] \\
&\overset{(b)}{\leq} \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^{M-1}} \mathbb{E}_{\hat{y}^{M-1}} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) + \tilde{R}_{M-1} \right] \\
&= \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^{M-1}} \mathbb{E}_{\hat{y}^{M-1}} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) + R_{M-1} + \tilde{R}_{M-1} - R_{M-1} \right] \\
&= \sup_{f^M} \sup_{\mu} \left(\mathbb{E}_{\mathbf{x}_{-N+1}^{M-1}} \mathbb{E}_{\hat{y}^{M-1}} \left[\sum_{j=1}^{M-1} \ell(\hat{y}_j, f_j(\mathbf{x}_j)) + R_{M-1} \right] + \mathbb{E}_{\mathbf{x}_{-N+1}^{M-1}} (\tilde{R}_{M-1} - R_{M-1}) \right) \\
&\overset{(c)}{\leq} \sup_{f^M} \sup_{\mu} \mathbb{E}_{\mathbf{x}_{-N+1}^{0}} \left[\tilde{R}_0 + \sum_{j=1}^{M-1} \mathbb{E}_{\mathbf{x}^j} [\tilde{R}_j - R_j] \right]
\end{aligned}
$$

where (a) follows by that replacing $f_M(\mathbf{x}_M)$ with $\sup_{y_M}$ do not decrease the value; (b) follows by Lemma 5; (c) follows by repeating the same argument for another $M-1$ steps.

Now, the key observation is that $\mathbb{E}_{\mathbf{x}_{-N+1}^{j}}[\tilde{R}_j - R_j] = 0$ for all $j \in [M-1]$ whenever $N \geq M-1$. This follows by the same argument as in the proof of Lemma 9 by noticing that the $\sup_{y^j}$ is outside the expectation $\mathbb{E}_{\epsilon' B}$ for oblivious adversaries. Moreover, this argument holds for all $B = N - M + j + 1 \geq 1$, i.e., $N \geq M - j$ (since by our assumption $N \geq M-1$ and $j \geq 1$). Therefore, we have

$$
\tilde{r}_{M,N}^{\mathsf{ob,side}}(\mathcal{H}, \Phi) \leq \mathbb{E}_{\mathbf{x}_{-N+1}^{0}}[\tilde{R}_0] \leq \mathsf{Rad}_M(\mathcal{H}) \leq O(M^q),
$$

whenever $N \geq M-1$. By the epoch approach as in Section 3.2 and taking the epoch length $M(n) = 2^n$ (which ensures $S(n) \geq M(n) - 1$) we conclude

$$
\tilde{r}_T^{\mathsf{ob}}(\mathcal{H}, \Psi) \leq \sum_{n=1}^{\lceil \log T \rceil} 2^{nq} \leq O(T^q),
$$

where Ψ is the epoch predictor derived from Φ as (13). The theorem now follows by Lemma 4 and noticing that the computational error only contributes $O(\sqrt{T})$ to the regret. ■

Remark 24 *Theorem 23 demonstrates that the* oblivious *minimax regret with* unknown $i.i.d.$ *feature generation process is equivalent to the regret achievable with* known *feature generation distribution and non-oblivious adversaries (Block et al., 2022, Thm 7), which also matches the information-theoretical lower bound (upto poly-logarithmic factors).*

Remark 25 *It is interesting to note that the proof of Theorem 23 can also be applied to the* semi-adaptive *adversaries that selects the adversary label y_j depending on $\mathbf{x}_{j-B}^{j}$ for some $B \geq 0$. This provides a way to interpolate all the ranges of regret from $\tilde{O}(T^{\max\{\frac{1}{2}, \frac{p-1}{p}\}})$ to $\tilde{O}(T^{\max\{\frac{3}{4}, \frac{p+1}{p+2}\}})$.*