Online Learning: Stochastic, Constrained, and Smoothed Adversaries

Alexander Rakhlin
Department of Statistics
University of Pennsylvania
rakhlin@wharton.upenn.edu

Karthik Sridharan
Toyota Technological Institute at Chicago
karthik@ttic.edu

Ambuj Tewari
Computer Science Department
University of Texas at Austin
ambuj@cs.utexas.edu

Abstract

Learning theory has largely focused on two main learning scenarios: the classical statistical setting where instances are drawn i.i.d. from a fixed distribution, and the adversarial scenario wherein, at every time step, an adversarially chosen instance is revealed to the player. It can be argued that in the real world neither of these assumptions is reasonable. We define the minimax value of a game where the adversary is restricted in his moves, capturing stochastic and non-stochastic assumptions on data. Building on the sequential symmetrization approach, we define a notion of distribution-dependent Rademacher complexity for the spectrum of problems ranging from i.i.d. to worst-case. The bounds let us immediately deduce variation-type bounds. We study a smoothed online learning scenario and show that exponentially small amount of noise can make function classes with infinite Littlestone dimension learnable.

1 Introduction

In the papers [1, 10, 11], an array of tools has been developed to study the minimax value of diverse sequential problems under the *worst-case* assumption on Nature. In [10], many analogues of the classical notions from statistical learning theory have been developed, and these have been extended in [11] for performance measures well beyond the additive regret. The process of *sequential symmetrization* emerged as a key technique for dealing with complicated nested minimax expressions. In the worst-case model, the developed tools give a unified treatment to such sequential problems as regret minimization, calibration of forecasters, Blackwell's approachability, Φ-regret, and more.

Learning theory has been so far focused predominantly on the i.i.d. and the worst-case learning scenarios. Much less is known about learnability in-between these two extremes. In the present paper, we make progress towards filling this gap by proposing a framework in which it is possible to variously restrict the behavior of Nature. By restricting Nature to play i.i.d. sequences, the results boil down to the classical notions of statistical learning in the supervised learning scenario. By not placing any restrictions on Nature, we recover the worst-case results of [10]. Between these two endpoints of the spectrum, particular assumptions on the adversary yield interesting bounds on the minimax value of the associated problem. Once again, the sequential symmetrization technique arises as the main tool for dealing with the minimax value, but the proofs require more care than in the i.i.d. or completely adversarial settings.

Adapting the game-theoretic language, we will think of the learner and the adversary as the two players of a zero-sum repeated game. Adversary's moves will be associated with "data", while the moves of the learner – with a function or a parameter. This point of view is not new: game-theoretic minimax analysis has been at the heart of statistical decision theory for more than half a century (see [3]). In fact, there is a well-developed theory of minimax estimation when restrictions are put on either the choice of the adversary or the allowed estimators by the player. *We are not aware of a similar theory for sequential problems with non-i.i.d. data.*

The main contribution of this paper is the development of tools for the analysis of online scenarios where the adversary's moves are restricted in various ways. In additional to general theory, we consider several interesting scenarios which can be captured by our framework. All proofs are deferred to the appendix.

2 Value of the Game

Let $\mathcal{F}$ be a closed subset of a complete separable metric space, denoting the set of moves of the learner. Suppose the adversary chooses from the set $\mathcal{X}$. Consider the `Online Learning Model`, defined as a T-round interaction between the learner and the adversary: On round $t = 1, \ldots, T$, the learner chooses $f_t \in \mathcal{F}$, the adversary simultaneously picks $x_t \in \mathcal{X}$, and the learner suffers loss $f_t(x_t)$. The goal of the learner is to minimize regret, defined as $\sum_{t=1}^T f_t(x_t) - \inf_{f\in\mathcal{F}} \sum_{t=1}^T f(x_t)$. It is a standard fact that simultaneity of the choices can be formalized by the first player choosing a mixed strategy; the second player then picks an action based on this mixed strategy, but not on its realization. We therefore consider randomized learners who predict a distribution $q_t \in \mathcal{Q}$ on every round, where $\mathcal{Q}$ is the set of probability distributions on $\mathcal{F}$, assumed to be weakly compact. The set of probability distributions on $\mathcal{X}$ (mixed strategies of the adversary) is denoted by $\mathcal{P}$.

We would like to capture the fact that sequences $(x_1, \ldots, x_T)$ cannot be arbitrary. This is achieved by defining restrictions on the adversary, that is, subsets of "allowed" distributions for each round. These restrictions limit the scope of available mixed strategies for the adversary.

Definition 1. A *restriction* $\mathcal{P}_{1:T}$ on the adversary is a sequence $\mathcal{P}_1, \ldots, \mathcal{P}_T$ of mappings $\mathcal{P}_t : \mathcal{X}^{t-1} \mapsto 2^{\mathcal{P}}$ such that $\mathcal{P}_t(x_{1:t-1})$ is a *convex* subset of $\mathcal{P}$ for any $x_{1:t-1} \in \mathcal{X}^{t-1}$.

Note that the restrictions depend on the past moves of the adversary, but not on those of the player. We will write $\mathcal{P}_t$ instead of $\mathcal{P}_t(x_{1:t-1})$ when $x_{1:t-1}$ is clearly defined. Using the notion of restrictions, we can give names to several types of adversaries that we will study in this paper.

(1) A *worst-case adversary* is defined by vacuous restrictions $\mathcal{P}_t(x_{1:t-1}) = \mathcal{P}$. That is, any mixed strategy is available to the adversary, including any deterministic point distribution.
(2) A *constrained adversary* is defined by $\mathcal{P}_t(x_{1:x_{t-1}})$ being the set of all distributions supported on the set $\{x \in \mathcal{X} : C_t(x_1, \ldots, x_{t-1}, x) = 1\}$ for some deterministic binary-valued constraint C_t. The deterministic constraint can, for instance, ensure that the length of the path determined by the moves $x_1, \ldots, x_t$ stays below the allowed budget.
(3) A *smoothed adversary* picks the worst-case sequence which gets corrupted by i.i.d. noise. Equivalently, we can view this as restrictions on the adversary who chooses the "center" (or a parameter) of the noise distribution.

Using techniques developed in this paper, we can also study the following adversaries (omitted due to lack of space):

(4) A *hybrid adversary* in the supervised learning game picks the worst-case label y_t, but is forced to draw the x_t-variable from a fixed distribution [6].
(5) An *i.i.d. adversary* is defined by a time-invariant restriction $\mathcal{P}_t(x_{1:t-1}) = \{p\}$ for every t and some $p \in \mathcal{P}$.

For the given restrictions $\mathcal{P}_{1:T}$, we define the value of the game as

$$\mathcal{V}_T(\mathcal{P}_{1:T}) \triangleq \inf_{q_1\in\mathcal{Q}} \sup_{p_1\in\mathcal{P}_1} \mathop{\mathbb{E}}_{f_1,x_1} \inf_{q_2\in\mathcal{Q}} \sup_{p_2\in\mathcal{P}_2} \mathop{\mathbb{E}}_{f_2,x_2} \cdots \inf_{q_T\in\mathcal{Q}} \sup_{p_T\in\mathcal{P}_T} \mathop{\mathbb{E}}_{f_T,x_T} \left[\sum_{t=1}^T f_t(x_t) - \inf_{f\in\mathcal{F}} \sum_{t=1}^T f(x_t)\right] \quad (1)$$

where f_t has distribution q_t and x_t has distribution p_t. As in [10], the adversary is *adaptive*, that is, chooses p_t based on the history of moves $f_{1:t-1}$ and $x_{1:t-1}$. At this point, the only difference from the setup of [10] is in the restrictions $\mathcal{P}_t$ on the adversary. Because these restrictions might not allow point distributions, suprema over p_t's in (1) cannot be equivalently written as the suprema over x_t's.

A word about the notation. In [10], the value of the game is written as $\mathcal{V}_T(\mathcal{F})$, signifying that the main object of study is $\mathcal{F}$. In [11], it is written as $\mathcal{V}_T(\ell, \Phi_T)$ since the focus is on the complexity of the set of transformations Φ_T and the payoff mapping ℓ. In the present paper, the main focus is indeed on the restrictions on the adversary, justifying our choice $\mathcal{V}_T(\mathcal{P}_{1:T})$ for the notation.

The first step is to apply the minimax theorem. To this end, we verify the necessary conditions. Our assumption that $\mathcal{F}$ is a closed subset of a complete separable metric space implies that $\mathcal{Q}$ is tight and Prokhorov's theorem states that compactness of $\mathcal{Q}$ under weak topology is equivalent to tightness [14]. Compactness under weak topology allows us to proceed as in [10]. Additionally, we require that the restriction sets are compact and convex.

Theorem 1. *Let $\mathcal{F}$ and $\mathcal{X}$ be the sets of moves for the two players, satisfying the necessary conditions for the minimax theorem to hold. Let $\mathcal{P}_{1:T}$ be the restrictions, and assume that for any $x_{1:t-1}$, $\mathcal{P}_t(x_{1:t-1})$ satisfies the necessary conditions for the minimax theorem to hold. Then*

$$\mathcal{V}_T(\mathcal{P}_{1:T}) = \sup_{p_1 \in \mathcal{P}_1} \mathbb{E}_{x_1 \sim p_1} \dots \sup_{p_T \in \mathcal{P}_T} \mathbb{E}_{x_T \sim p_T} \left[\sum_{t=1}^{T} \inf_{f_t \in \mathcal{F}} \mathbb{E}_{x_t \sim p_t} \left[f_t(x_t) \right] - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} f(x_t) \right]. \quad (2)$$

The nested sequence of suprema and expected values in Theorem 1 can be re-written succinctly as

$$\begin{aligned} \mathcal{V}_T(\mathcal{P}_{1:T}) &= \sup_{\mathbf{p} \in \mathfrak{P}} \mathbb{E}_{x_1 \sim p_1} \mathbb{E}_{x_2 \sim p_2(\cdot|x_1)} \dots \mathbb{E}_{x_T \sim p_T(\cdot|x_{1:T-1})} \left[\sum_{t=1}^{T} \inf_{f_t \in \mathcal{F}} \mathbb{E}_{x_t \sim p_t} \left[f_t(x_t) \right] - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} f(x_t) \right] \\ &= \sup_{\mathbf{p} \in \mathfrak{P}} \mathbb{E} \left[\sum_{t=1}^{T} \inf_{f_t \in \mathcal{F}} \mathbb{E}_{x_t \sim p_t} \left[f_t(x_t) \right] - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} f(x_t) \right] \end{aligned} \quad (3)$$

where the supremum is over all joint distributions $\mathbf{p}$ over sequences, such that $\mathbf{p}$ satisfies the restrictions as described below. Given a joint distribution $\mathbf{p}$ on sequences $(x_1, \dots, x_T) \in \mathcal{X}^T$, we denote the associated conditional distributions by $p_t(\cdot|x_{1:t-1})$. We can think of the choice $\mathbf{p}$ as a sequence of oblivious strategies $\{p_t : \mathcal{X}^{t-1} \mapsto \mathcal{P}\}_{t=1}^T$, mapping the prefix $x_{1:t-1}$ to a conditional distribution $p_t(\cdot|x_{1:t-1}) \in \mathcal{P}_t(x_{1:t-1})$. We will indeed call $\mathbf{p}$ a "joint distribution" or an "oblivious strategy" interchangeably. We say that a joint distribution $\mathbf{p}$ *satisfies restrictions* if for any t and any $x_{1:t-1} \in \mathcal{X}^{t-1}$, $p_t(\cdot|x_{1:t-1}) \in \mathcal{P}_t(x_{1:t-1})$. The set of all joint distributions satisfying the restrictions is denoted by $\mathfrak{P}$. We note that Theorem 1 cannot be deduced immediately from the analogous result in [10], as it is not clear how the restrictions on the adversary per each round come into play after applying the minimax theorem. Nevertheless, it is comforting that the restrictions directly translate into the set $\mathfrak{P}$ of oblivious strategies satisfying the restrictions.

Before continuing with our goal of upper-bounding the value of the game, we state the following interesting facts.

Proposition 2. *There is an oblivious minimax optimal strategy for the adversary, and there is a corresponding minimax optimal strategy for the player that does not depend on its own moves.*

The latter statement of the proposition is folklore for worst-case learning, yet we have not seen a proof of it in the literature. The proposition holds for all online learning settings with legal restrictions $\mathcal{P}_{1:T}$, encompassing also the no-restrictions setting of worst-case online learning [10]. The result crucially relies on the fact that the objective is external regret.

3 Symmetrization and Random Averages

Theorem 1 is a useful representation of the value of the game. As the next step, we upper bound it with an expression which is easier to study. Such an expression is obtained by introducing Rademacher random variables. This process can be termed *sequential symmetrization* and has been exploited in [1, 10, 11]. The restrictions $\mathcal{P}_t$, however, make sequential symmetrization considerably more involved than in the papers cited above. The main difficulty arises from the fact that the set $\mathcal{P}_t(x_{1:t-1})$ depends on the sequence $x_{1:t-1}$, and symmetrization (that is, replacement of x_s with x'_s) has to be done with care as it affects this dependence. Roughly speaking, in the process of symmetrization, a tangent sequence $x'_1, x'_2, \dots$ is introduced such that x_t and x'_t are independent and identically distributed given "the past". However, "the past" is itself an interleaving choice of the original sequence and the tangent sequence.

Define the "selector function" $\chi : \mathcal{X} \times \mathcal{X} \times \{\pm 1\} \mapsto \mathcal{X}$ by $\chi(x, x', \epsilon) = x'$ if $\epsilon = 1$ and $\chi(x, x', \epsilon) = x$ if $\epsilon = -1$. When x_t and x'_t are understood from the context, we will use the shorthand $\chi_t(\epsilon) := \chi(x_t, x'_t, \epsilon)$. In other words, χ_t selects between x_t and x'_t depending on the sign of ϵ. Throughout the paper, we deal with binary trees, which arise from symmetrization [10]. Given some set $\mathcal{Z}$, an *$\mathcal{Z}$-valued tree of depth T* is a sequence $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_T)$ of T mappings $\mathbf{z}_i : \{\pm 1\}^{i-1} \mapsto \mathcal{Z}$. The T-tuple $\epsilon = (\epsilon_1, \dots, \epsilon_T) \in \{\pm 1\}^T$ defines a path. For brevity, we write $\mathbf{z}_t(\epsilon)$ instead of $\mathbf{z}_t(\epsilon_{1:t-1})$.

Given a joint distribution $\mathbf{p}$, consider the "$(\mathcal{X} \times \mathcal{X})^{T-1} \mapsto \mathcal{P}(\mathcal{X} \times \mathcal{X})$"- valued probability tree $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_T)$ defined by

$$\boldsymbol{\rho}_t(\epsilon_{1:t-1})\left((x_1, x'_1), \dots, (x_{T-1}, x'_{T-1})\right) = (p_t(\cdot|\chi_1(\epsilon_1), \dots, \chi_{t-1}(\epsilon_{t-1})), p_t(\cdot|\chi_1(\epsilon_1), \dots, \chi_{t-1}(\epsilon_{t-1}))).$$

In other words, the values of the mappings $\boldsymbol{\rho}_t(\epsilon)$ are products of conditional distributions, where conditioning is done with respect to a sequence made from x_s and x'_s depending on the sign of ϵ_s. We note that the difficulty in intermixing the x and x' sequences *does not arise in i.i.d. or worst-case symmetrization*. However, in-between these extremes the notational complexity seems to be unavoidable if we are to employ symmetrization and obtain a version of Rademacher complexity.

As an example, consider the "left-most" path $\epsilon = -\mathbf{1}$ in a binary tree of depth T, where $\mathbf{1} = (1, \dots, 1)$ is a T-dimensional vector of ones. Then all the selectors $\chi(x_t, x'_t, \epsilon_t)$ choose the sequence $x_1, \dots, x_T$. The probability tree $\boldsymbol{\rho}$ on the "left-most" path is, therefore, defined by the conditional distributions $p_t(\cdot|x_{1:t-1})$; on the path $\epsilon = \mathbf{1}$, the conditional distributions are $p_t(\cdot|x'_{1:t-1})$.

Slightly abusing the notation, we will write $\boldsymbol{\rho}_t(\epsilon)\left((x_1, x'_1), \dots, (x_{t-1}, x'_{t-1})\right)$ for the probability tree since $\boldsymbol{\rho}_t$ clearly depends only on the prefix up to time $t-1$. Throughout the paper, it will be understood that the tree $\boldsymbol{\rho}$ is obtained from $\mathbf{p}$ as described above. Since all the conditional distributions of $\mathbf{p}$ satisfy the restrictions, so do the corresponding distributions of the probability tree $\boldsymbol{\rho}$. By saying that $\boldsymbol{\rho}$ satisfies restrictions we then mean that $\mathbf{p} \in \mathfrak{P}$.

Sampling of a pair of $\mathcal{X}$-valued trees from $\boldsymbol{\rho}$, written as $(\mathbf{x}, \mathbf{x}') \sim \boldsymbol{\rho}$, is defined as the following recursive process: for any $\epsilon \in \{\pm 1\}^T$, $(\mathbf{x}_1(\epsilon), \mathbf{x}'_1(\epsilon)) \sim \boldsymbol{\rho}_1(\epsilon)$ and

$$(\mathbf{x}_t(\epsilon), \mathbf{x}'_t(\epsilon)) \sim \boldsymbol{\rho}_t(\epsilon)((\mathbf{x}_1(\epsilon), \mathbf{x}'_1(\epsilon)), \dots, (\mathbf{x}_{t-1}(\epsilon), \mathbf{x}'_{t-1}(\epsilon))) \quad \text{for} \quad 2 \le t \le T \tag{4}$$

To gain a better understanding of the sampling process, consider the first few levels of the tree. The roots $\mathbf{x}_1, \mathbf{x}'_1$ of the trees $\mathbf{x}, \mathbf{x}'$ are sampled from p_1, the conditional distribution for $t = 1$ given by $\mathbf{p}$. Next, say, $\epsilon_1 = +1$. Then the "right" children of $\mathbf{x}_1$ and $\mathbf{x}'_1$ are sampled via $\mathbf{x}_2(+1), \mathbf{x}'_2(+1) \sim p_2(\cdot|\mathbf{x}'_1)$ since $\chi_1(+1)$ selects $\mathbf{x}'_1$. On the other hand, the "left" children $\mathbf{x}_2(-1), \mathbf{x}'_2(-1)$ are both distributed according to $p_2(\cdot|\mathbf{x}_1)$. Now, suppose $\epsilon_1 = +1$ and $\epsilon_2 = -1$. Then, $\mathbf{x}_3(+1, -1), \mathbf{x}'_3(+1, -1)$ are both sampled from $p_3(\cdot|\mathbf{x}'_1, \mathbf{x}_2(+1))$.

The proof of Theorem 3 reveals why such intricate conditional structure arises, and Proposition 5 below shows that this structure greatly simplifies for i.i.d. and worst-case situations. Nevertheless, the process described above allows us to define a *unified notion of Rademacher complexity* for the spectrum of assumptions between the two extremes.

Definition 2. The *distribution-dependent sequential Rademacher complexity* of a function class $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ is defined as

$$\mathfrak{R}_T(\mathcal{F}, \mathbf{p}) \triangleq \mathbb{E}_{(\mathbf{x}, \mathbf{x}') \sim \boldsymbol{\rho}} \mathbb{E}_\epsilon \left[\sup_{f \in \mathcal{F}} \sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon)) \right]$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_T)$ is a sequence of i.i.d. Rademacher random variables and $\boldsymbol{\rho}$ is the probability tree associated with $\mathbf{p}$.

We now prove an upper bound on the value $\mathcal{V}_T(\mathcal{P}_{1:T})$ of the game in terms of this distribution-dependent sequential Rademacher complexity. The result cannot be deduced directly from [10], and it greatly increases the scope of problems whose learnability can now be studied in a unified manner.

Theorem 3. *The minimax value is bounded as*

$$\mathcal{V}_T(\mathcal{P}_{1:T}) \le 2 \sup_{\mathbf{p} \in \mathfrak{P}} \mathfrak{R}_T(\mathcal{F}, \mathbf{p}). \tag{5}$$

More generally, for any measurable function M_t such that $M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon) = M_t(\mathbf{p}, f, \mathbf{x}', \mathbf{x}, -\epsilon)$,

$$\mathcal{V}_T(\mathcal{P}_{1:T}) \le 2 \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}} \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t (f(\mathbf{x}_t(\epsilon)) - M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon)) \right]$$

The following corollary provides a natural "centered" version of the distribution-dependent Rademacher complexity. That is, the complexity can be measured by relative shifts in the adversarial moves.

Corollary 4. *For the game with restrictions $\mathcal{P}_{1:T}$,*

$$\mathcal{V}_T(\mathcal{P}_{1:T}) \le 2 \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}} \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \Big(f(\mathbf{x}_t(\epsilon)) - \mathbb{E}_{t-1} f(\mathbf{x}_t(\epsilon)) \Big) \right]$$

where $\mathbb{E}_{t-1}$ denotes the conditional expectation of $\mathbf{x}_t(\epsilon)$.

Example 1. *Suppose $\mathcal{F}$ is a unit ball in a Banach space and $f(x) = \langle f, x\rangle$. Then*

$$\mathcal{V}_T(\mathcal{P}_{1:T}) \le 2 \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}} \mathbb{E}_\epsilon \left\| \sum_{t=1}^T \epsilon_t \Big(\mathbf{x}_t(\epsilon) - \mathbb{E}_{t-1} \mathbf{x}_t(\epsilon) \Big) \right\|$$

Suppose the adversary plays a simple random walk (e.g., $p_t(x|x_1, \ldots, x_{t-1}) = p_t(x|x_{t-1})$ is uniform on a unit sphere). For simplicity, suppose this is the only strategy allowed by the set $\mathfrak{P}$. Then $\mathbf{x}_t(\epsilon) - \mathbb{E}_{t-1}\mathbf{x}_t(\epsilon)$ are independent increments when conditioned on the history. Further, the increments do not depend on ϵ_t. Thus, $\mathcal{V}_T(\mathcal{P}_{1:T}) \le 2\mathbb{E} \left\| \sum_{t=1}^T Y_t \right\|$ where $\{Y_t\}$ is the corresponding random walk.

We now show that the distribution-dependent sequential Rademacher complexity *for i.i.d. data* is precisely the classical Rademacher complexity, and further show that the distribution-dependent sequential Rademacher complexity is always upper bounded by the worst-case sequential Rademacher complexity defined in [10].

Proposition 5. *First, consider the i.i.d. restrictions $\mathcal{P}_t = \{p\}$ for all t, where p is some fixed distribution on $\mathcal{X}$, and let $\boldsymbol{\rho}$ be the process associated with the joint distribution $\mathbf{p} = p^T$. Then*

$$\mathfrak{R}_T(\mathcal{F}, \mathbf{p}) = \mathfrak{R}_T(\mathcal{F}, p), \qquad \textit{where} \qquad \mathfrak{R}_T(\mathcal{F}, p) \triangleq \mathbb{E}_{x_1,\ldots,x_T\sim p} \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(x_t) \right] \tag{6}$$

is the classical Rademacher complexity. Second, for any joint distribution $\mathbf{p}$,

$$\mathfrak{R}_T(\mathcal{F}, \mathbf{p}) \le \mathfrak{R}_T(\mathcal{F}), \qquad \textit{where} \qquad \mathfrak{R}_T(\mathcal{F}) \triangleq \sup_{\mathbf{x}} \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon)) \right] \tag{7}$$

is the sequential Rademacher complexity defined in [10].

In the case of hybrid learning, adversary chooses a sequence of pairs (x_t, y_t) where the instance x_t's are i.i.d. but the labels y_i's are fully adversarial. The distribution-dependent Rademacher complexity in such a hybrid case can be upper bounded by a vary natural quantity: a random average where expectation is taken over x_t's and a supremum over $\mathcal{Y}$-valued trees. So, the distribution dependent Rademacher complexity itself becomes a hybrid between the classical Rademacher complexity and the worst case sequential Rademacher complexity. For more details, see Lemma 17 in the Appendix as another example of an analysis of the distribution-dependent sequential Rademacher complexity.

Distribution-dependent sequential Rademacher complexity enjoys many of the nice properties satisfied by both classical and worst-case Rademacher complexities. As shown in [10], these properties are handy tools for proving upper bounds on the value in various examples. We have: (a) If $\mathcal{F} \subset \mathcal{G}$, then $\mathfrak{R}(\mathcal{F}, \mathbf{p}) \le \mathfrak{R}(\mathcal{G}, \mathbf{p})$; (b) $\mathfrak{R}(\mathcal{F}, \mathbf{p}) = \mathfrak{R}(\mathrm{conv}(\mathcal{F}), \mathbf{p})$; (c) $\mathfrak{R}(c\mathcal{F}, \mathbf{p}) = |c|\mathfrak{R}(\mathcal{F}, \mathbf{p})$ for all $c \in \mathbb{R}$; (d) For any h, $\mathfrak{R}(\mathcal{F} + h, \mathbf{p}) = \mathfrak{R}(\mathcal{F}, \mathbf{p})$ where $\mathcal{F} + h = \{f + h : f \in \mathcal{F}\}$.

In addition to the above properties, upper bounds on $\mathfrak{R}(\mathcal{F}, \mathbf{p})$ can be derived via sequential covering numbers defined in [10]. This notion of a cover captures the sequential complexity of a function class on a given $\mathcal{X}$-valued tree $\mathbf{x}$. One can then show an analogue of the Dudley integral bound, where the complexity is averaged with respect to the underlying process $(\mathbf{x}, \mathbf{x}') \sim \boldsymbol{\rho}$.

4 Application: Constrained Adversaries

In this section, we consider adversaries who are deterministically constrained in the sequences of actions they can play. It is often useful to consider scenarios where the adversary is worst case, yet has some *budget* or *constraint* to satisfy while picking the actions. Examples of such scenarios include, for instance, games where the adversary is constrained to make moves that are close in some fashion to the previous move, linear games with bounded variance, and so on. Below we formulate such games quite generally through arbitrary constraints that the adversary has to satisfy on each round. We easily derive several results to illustrate the versatility of the developed framework.

For a T round game consider an adversary who is only allowed to play sequences $x_1, \ldots, x_T$ such that at round t the constraint $C_t(x_1, \ldots, x_t) = 1$ is satisfied, where $C_t : \mathcal{X}^t \mapsto \{0, 1\}$ represents the constraint on the sequence played so far. The constrained adversary can be viewed as a stochastic adversary with restrictions on the conditional distribution at time t given by the set of all Borel distributions on the set $\mathcal{X}_t(x_{1:t-1}) \triangleq \{x \in \mathcal{X} : C_t(x_1, \ldots, x_{t-1}, x) = 1\}$. Since this set includes all point distributions on each $x \in \mathcal{X}_t$, the sequential complexity simplifies in a way similar to worst-case adversaries. We write $\mathcal{V}_T(C_{1:T})$ for the value of the game with the given constraints. Now, assume that for any $x_{1:t-1}$, the set of all distributions on $\mathcal{X}_t(x_{1:t-1})$ is weakly compact in a way similar to compactness of $\mathcal{P}$. That is, $\mathcal{P}_t(x_{1:t-1})$ satisfy the necessary conditions for the minimax theorem to hold. We have the following corollaries of Theorems 1 and 3.

Corollary 6. *Let $\mathcal{F}$ and $\mathcal{X}$ be the sets of moves for the two players, satisfying the necessary conditions for the minimax theorem to hold. Let $\{C_t : \mathcal{X}^{t-1} \mapsto \{0, 1\}\}_{t=1}^T$ be the* constraints*. Then*

$$\mathcal{V}_T(C_{1:T}) \;=\; \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sum_{t=1}^T \inf_{f_t\in\mathcal{F}} \mathbb{E}_{x_t\sim p_t}\left[f_t(x_t)\right] - \inf_{f\in\mathcal{F}} \sum_{t=1}^T f(x_t)\right] \tag{8}$$

where $\mathbf{p}$ ranges over all distributions over sequences $(x_1, \ldots, x_T)$ such that $\forall t, C_t(x_{1:t-1}) = 1$.

Corollary 7. *Let the set $\mathcal{T}$ be a set of pairs $(\mathbf{x}, \mathbf{x}')$ of $\mathcal{X}$-valued trees with the property that for any $\epsilon \in \{\pm1\}^T$ and any $t \in [T]$, $C(\chi_1(\epsilon_1), \ldots, \chi_{t-1}(\epsilon_{t-1}), \mathbf{x}_t(\epsilon)) = C(\chi_1(\epsilon_1), \ldots, \chi_{t-1}(\epsilon_{t-1}), \mathbf{x}'_t(\epsilon)) = 1$. The minimax value is bounded as*

$$\mathcal{V}_T(C_{1:T}) \le 2 \sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}} \mathfrak{R}_T(\mathcal{F}, \mathbf{p}).$$

More generally, for any measurable function M_t such that $M_t(f, \mathbf{x}, \mathbf{x}', \epsilon) = M_t(f, \mathbf{x}', \mathbf{x}, -\epsilon)$,

$$\mathcal{V}_T(C_{1:T}) \le 2 \sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}} \mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t(f(\mathbf{x}_t(\epsilon)) - M_t(f, \mathbf{x}, \mathbf{x}', \epsilon))\right] .$$

Armed with these results, we can recover and extend some known results on online learning against budgeted adversaries. The first result says that if the adversary is not allowed to move by more than σ_t away from its previous average of decisions, the player has a strategy to exploit this fact and obtain lower regret. For the ℓ_2-norm, such "total variation" bounds have been achieved in [4] up to a $\log T$ factor. Our analysis seamlessly incorporates variance measured in arbitrary norms, not just ℓ_2. We emphasize that such certificates of learnability are not possible with the analysis of [10].

Proposition 8 (Variance Bound)**.** *Consider the online linear optimization setting with $\mathcal{F} = \{f : \Psi(f) \le R^2\}$ for a λ-strongly function $\Psi : \mathcal{F} \mapsto \mathbb{R}_+$ on $\mathcal{F}$, and $\mathcal{X} = \{x : \|x\|_* \le 1\}$. Let $f(x) = \langle f, x\rangle$ for any $f \in \mathcal{F}$ and $x \in \mathcal{X}$. Consider the sequence of constraints $\{C_t\}_{t=1}^T$ given by $C_t(x_1, \ldots, x_{t-1}, x) = 1$ if $\|x - \frac{1}{t-1}\sum_{\tau=1}^{t-1} x_\tau\|_* \le \sigma_t$ and 0 otherwise. Then*

$$\mathcal{V}_T(C_{1:T}) \le 2\sqrt{2}R\sqrt{\lambda^{-1}\sum_{t=1}^T \sigma_t^2}$$

In particular, we obtain the following ℓ_2 variance bound. Consider the case when $\Psi : \mathcal{F} \mapsto \mathbb{R}_+$ is given by $\Psi(f) = \frac{1}{2}\|f\|^2$, $\mathcal{F} = \{f : \|f\|_2 \le 1\}$ and $\mathcal{X} = \{x : \|x\|_2 \le 1\}$. Consider the constrained game where the move x_t played by adversary at time t satisfies $\left\|x_t - \frac{1}{t-1}\sum_{\tau=1}^{t-1} x_\tau\right\|_2 \le \sigma_t$. In this case we can conclude that $\mathcal{V}_T(C_{1:T}) \le 2\sqrt{2}\sqrt{\sum_{t=1}^T \sigma_t^2}$. We can also derive a variance bound over the simplex. Let $\Psi(f) = \sum_{i=1}^d f_i \log(df_i)$ is defined over the d-simplex $\mathcal{F}$, and $\mathcal{X} = \{x : \|x\|_\infty \le 1\}$. Consider the constrained game where the move x_t played by adversary at time t satisfies $\max_{j\in[d]} \left| x_t[j] - \frac{1}{t-1}\sum_{\tau=1}^{t-1} x_\tau[j] \right| \le \sigma_t$. For any $f \in \mathcal{F}$, $\Psi(f) \le \log(d)$ and so we conclude that $\mathcal{V}_T(C_{1:T}) \le 2\sqrt{2}\sqrt{\log(d)\sum_{t=1}^T \sigma_t^2}$.

The next Proposition gives a bound whenever the adversary is constrained to choose his decision from a small ball around the previous decision.

Proposition 9 (Slowly-Changing Decisions)**.** *Consider the online linear optimization setting where adversary's move at any time is close to the move during the previous time step. Let $\mathcal{F} = \{f : \Psi(f) \le R^2\}$ where $\Psi : \mathcal{F} \mapsto \mathbb{R}_+$ is a λ-strongly function on $\mathcal{F}$ and $\mathcal{X} = \{x : \|x\|_* \le B\}$. Let $f(x) = \langle f, x\rangle$ for any $f \in \mathcal{F}$ and $x \in \mathcal{X}$. Consider the sequence of constraints $\{C_t\}_{t=1}^T$ given by $C_t(x_1, \ldots, x_{t-1}, x) = 1$ if $\|x - x_{t-1}\|_* \le \delta$ and 0 otherwise. Then,*

$$\mathcal{V}_T(C_{1:T}) \le 2R\delta\sqrt{2T/\lambda} .$$

In particular, consider the case of a Euclidean-norm restriction on the moves. Let $\Psi : \mathcal{F} \mapsto \mathbb{R}_+$ is given by $\Psi(f) = \frac{1}{2}\|f\|^2$, $\mathcal{F} = \{f : \|f\|_2 \le 1\}$ and $\mathcal{X} = \{x : \|x\|_2 \le 1\}$. Consider the constrained game where the move x_t played by adversary at time t satisfies $\|x_t - x_{t-1}\|_2 \le \delta$. In this case we can conclude that $\mathcal{V}_T(C_{1:T}) \le 2\delta\sqrt{2T}$. For the case of decision-making on the simplex, we obtain the following result. Let $\Psi(f) = \sum_{i=1}^d f_i \log(df_i)$ is defined over the d-simplex $\mathcal{F}$, and $\mathcal{X} = \{x : \|x\|_\infty \le 1\}$. Consider the constrained game where the move x_t played by adversary at time t satisfies $\|x_t - x_{t-1}\|_\infty \le \delta$. In this case note that for any $f \in \mathcal{F}$, $\Psi(f) \le \log(d)$ and so we can conclude that $\mathcal{V}_T(C_{1:T}) \le 2\delta\sqrt{2T\log(d)}$.

5 Application: Smoothed Adversaries

The development of *smoothed analysis* over the past decade is arguably one of the landmarks in the study of complexity of algorithms. In contrast to the overly optimistic *average complexity* and the overly pessimistic *worst-case complexity*, smoothed complexity can be seen as a more realistic measure of algorithm's performance. In their groundbreaking work, Spielman and Teng [13] showed that the smoothed running time complexity of the simplex method is polynomial. This result explains good performance of the method in practice despite its exponential-time worst-case complexity. In this section, we consider the effect of smoothing on *learnability*.

It is well-known that there is a gap between the i.i.d. and the worst-case scenarios. In fact, we do not need to go far for an example: a simple class of threshold functions on a unit interval is learnable in the i.i.d. supervised learning scenario, yet difficult in the online worst-case model [8, 2, 9]. This fact is reflected in the corresponding combinatorial dimensions: the Vapnik-Chervonenkis dimension is one, whereas the Littlestone dimension is infinite. The proof of the latter fact, however, reveals that the infinite number of mistakes on the part of the player is due to the infinite resolution of the carefully chosen adversarial sequence. We can argue that this infinite precision is an unreasonable assumption on the power of a real-world opponent. The idea of limiting the power of the malicious adversary through perturbing the sequence can be traced back to Posner and Kulkarni [9]. The authors considered on-line learning of functions of bounded variation, but in the so-called realizable setting (that is, when labels are given by some function in the given class).

We define the *smoothed online learning model* as the following T-round interaction between the learner and the adversary. On round t, the learner chooses $f_t \in \mathcal{F}$; the adversary simultaneously chooses $x_t \in \mathcal{X}$, which is then perturbed by some noise $s_t \sim \sigma$, yielding a value $\tilde{x}_t = \omega(x_t, s_t)$; and the player suffers $f_t(\tilde{x}_t)$. Regret is defined with respect to the perturbed sequence. Here $\omega : \mathcal{X} \times S \mapsto \mathcal{X}$ is some measurable mapping; for instance, additive disturbances can be written as $\tilde{x} = \omega(x, s) = x + s$. If ω keeps x_t unchanged, that is $\omega(x_t, s_t) = x_t$, the setting is precisely the standard online learning model. In the full information version, we assume that the choice $\tilde{x}_t$ is revealed to the player at the end of round t. We now recognize that the setting is nothing but a particular way to restrict the adversary. That is, the choice $x_t \in \mathcal{X}$ defines a parameter of a mixed strategy from which a actual move $\omega(x_t, s_t)$ is drawn; for instance, for additive zero-mean Gaussian noise, x_t defines the center of the distribution from which $x_t + s_t$ is drawn. In other words, noise does not allow the adversary to play any desired mixed strategy.

The value of the smoothed online learning game (as defined in (1)) can be equivalently written as

$$\mathcal{V}_T = \inf_{q_1} \sup_{x_1} \mathop{\mathbb{E}}_{\substack{f_1 \sim q_1 \\ s_1 \sim \sigma}} \inf_{q_2} \sup_{x_2} \mathop{\mathbb{E}}_{\substack{f_2 \sim q_2 \\ s_2 \sim \sigma}} \cdots \inf_{q_T} \sup_{x_T} \mathop{\mathbb{E}}_{\substack{f_T \sim q_T \\ s_T \sim \sigma}} \left[\sum_{t=1}^{T} f_t(\omega(x_t, s_t)) - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} f(\omega(x_t, s_t)) \right]$$

where the infima are over $q_t \in \mathcal{Q}$ and the suprema are over $x_t \in \mathcal{X}$. Using sequential symmetrization, we deduce the following upper bound on the value of the smoothed online learning game.

Theorem 10. *The value of the smoothed online learning game is bounded above as*

$$\mathcal{V}_T \le 2 \sup_{x_1 \in \mathcal{X}} \mathop{\mathbb{E}}_{s_1 \sim \sigma} \mathbb{E}_{\epsilon_1} \ldots \sup_{x_T \in \mathcal{X}} \mathop{\mathbb{E}}_{s_T \sim \sigma} \mathbb{E}_{\epsilon_T} \left[\sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \epsilon_t f(\omega(x_t, s_t)) \right]$$

We now demonstrate how Theorem 10 can be used to show learnability for smoothed learning of threshold functions. First, consider the supervised game with threshold functions on a unit interval (that is, non-homogenous hyperplanes). The moves of the adversary are pairs $x = (z, y)$ with $z \in [0,1]$ and $y \in \{0,1\}$, and the binary-valued function class $\mathcal{F}$ is defined by

$$\mathcal{F} = \{ f_\theta(z,y) = |y - \mathbf{1}\{z < \theta\}| : \theta \in [0,1] \}, \tag{9}$$

that is, every function is associated with a threshold $\theta \in [0,1]$. The class $\mathcal{F}$ has infinite Littlestone's dimension and is not learnable in the worst-case online framework. Consider a smoothed scenario, with the z-variable of the adversarial move (z,y) perturbed by an additive uniform noise $\sigma = \text{Unif}[-\gamma/2, \gamma/2]$ for some $\gamma \ge 0$. That is, the actual move revealed to the player at time t is $(z_t + s_t, y_t)$, with $s_t \sim \sigma$. Any non-trivial upper bound on regret has to depend on particular noise assumptions, as $\gamma = 0$ corresponds to the case with infinite Littlestone dimension. For the uniform disturbance, the intuition tells us that noise implies a margin, and we should expect a $1/\gamma$ complexity parameter appearing in the bounds. The next lemma quantifies the intuition that additive noise limits precision of the adversary.

Lemma 11. *Let $\theta_1, \ldots, \theta_N$ be obtained by discretizing the interval $[0,1]$ into $N = T^a$ bins $[\theta_i, \theta_{i+1})$ of length T^{-a}, for some $a \ge 3$. Then, for any sequence $z_1, \ldots, z_T \in [0,1]$, with probability at least $1 - \frac{1}{\gamma T^{a-2}}$, no two elements of the sequence $z_1 + s_1, \ldots, z_T + s_T$ belong to the same interval $[\theta_i, \theta_{i+1})$, where $s_1, \ldots, s_T$ are i.i.d.* Unif$[-\gamma/2, \gamma/2]$.

We now observe that, conditioned on the event in Lemma 11, the upper bound on the value in Theorem 10 is a supremum of N martingale difference sequences! We then arrive at:

Proposition 12. *For the problem of smoothed online learning of thresholds in 1-D, the value is*

$$\mathcal{V}_T \le 2 + \sqrt{2T\left(4 \log T + \log(1/\gamma)\right)}$$

What we found is somewhat surprising: for a problem which is not learnable in the online worst-case scenario, an exponentially small noise added to the moves of the adversary yields a learnable problem. This shows, at least in the given example, that the worst-case analysis and Littlestone's dimension are brittle notions which might be too restrictive in the real world, where some noise is unavoidable. It is comforting that small additive noise makes the problem learnable!

The proof for smoothed learning of half-spaces in higher dimension follows the same route as the one-dimensional exposition. For simplicity, assume the hyperplanes are homogenous and $\mathcal{Z} = S_{d-1} \subset \mathbb{R}^d$, $\mathcal{Y} = \{-1, 1\}$, $\mathcal{X} = \mathcal{Z} \times \mathcal{Y}$. Define $\mathcal{F} = \{ f_\theta(z,y) = \mathbf{1}\{y \langle z, \theta \rangle > 0\} : \theta \in S_{d-1} \}$, and assume that the noise is distributed uniformly on a square patch with side-length γ on the surface of the sphere S_{d-1}. We can also consider other distributions, possibly with support on a d-dimensional ball instead.

Proposition 13. *For the problem of smoothed online learning of half-spaces,*

$$\mathcal{V}_T = O\left(\sqrt{dT \left(\log\left(\frac{1}{\gamma}\right) + \frac{3}{d-1} \log T \right)} + v_{d-2} \cdot \left(\frac{1}{\gamma}\right)^{\frac{3}{d-1}} \right)$$

where v_{d-2} is constant depending only on the dimension d.

We conclude that half spaces are online learnable in the smoothed model, since the upper bound of Proposition 13 guarantees existence of an algorithm which achieves this regret. In fact, for the two examples considered in this section, the Exponential Weights Algorithm on the discretization given by Lemma 11 is a (computationally infeasible) algorithm achieving the bound.

References

[1] J. Abernethy, A. Agarwal, P. Bartlett, and A. Rakhlin. A stochastic view of optimal regret through minimax duality. In *COLT*, 2009.

[2] S. Ben-David, D. Pal, and S. Shalev-Shwartz. Agnostic online learning. In *Proceedings of the 22th Annual Conference on Learning Theory*, 2009.

[3] J.O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, 1985.

[4] E. Hazan and S. Kale. Better algorithms for benign bandits. In *SODA*, 2009.

[5] S.M. Kakade, K. Sridharan, and A. Tewari. On the complexity of linear prediction: Risk bounds, margin bounds, and regularization. *NIPS*, 22, 2008.

[6] A. Lazaric and R. Munos. Hybrid Stochastic-Adversarial On-line Learning. In *COLT*, 2009.

[7] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer-Verlag, New York, 1991.

[8] N. Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2(4):285–318, 04 1988.

[9] S. Posner and S. Kulkarni. On-line learning of functions of bounded variation under various sampling schemes. In *Proceedings of the sixth annual conference on Computational learning theory*, pages 439–445. ACM, 1993.

[10] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Random averages, combinatorial parameters, and learnability. In *NIPS*, 2010. Full version available at arXiv:1006.1138.

[11] A. Rakhlin, K. Sridharan, and A. Tewari. Online learning: Beyond regret. In *COLT*, 2011. Full version available at arXiv:1011.3168.

[12] S. Shalev-Shwartz, O. Shamir, N. Srebro, and K. Sridharan. Learnability, stability and uniform convergence. *JMLR*, 11:2635–2670, Oct 2010.

[13] D. A. Spielman and S. H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.

[14] A. W. Van Der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes : With Applications to Statistics*. Springer Series, March 1996.

A Proofs

Proof of Theorem 1. The proof is identical to that in [10]. For simplicity, denote $\psi(x_{1:T}) = \inf_{f\in\mathcal{F}} \sum_{t=1}^T f(x_t)$. The first step in the proof is to appeal to the minimax theorem for every couple of inf and sup:

$$\begin{aligned}
&\inf_{q_1\in\mathcal{Q}} \sup_{p_1\in\mathcal{P}_1} \mathop{\mathbb{E}}_{\substack{f_1\sim q_1\\ x_1\sim p_1}} \cdots \inf_{q_T\in\mathcal{Q}} \sup_{p_T\in\mathcal{P}_T} \mathop{\mathbb{E}}_{\substack{f_T\sim q_T\\ x_T\sim p_T}} \left[\sum_{t=1}^T f_t(x_t) - \psi(x_{1:T})\right] \\
&= \sup_{p_1\in\mathcal{P}_1} \inf_{q_1\in\mathcal{Q}} \mathop{\mathbb{E}}_{\substack{f_1\sim q_1\\ x_1\sim p_1}} \dots \sup_{p_T\in\mathcal{P}_T} \inf_{q_T\in\mathcal{Q}} \mathop{\mathbb{E}}_{\substack{f_T\sim q_T\\ x_T\sim p_T}} \left[\sum_{t=1}^T f_t(x_t) - \psi(x_{1:T})\right] \\
&= \sup_{p_1\in\mathcal{P}_1} \inf_{f_1\in\mathcal{F}} \mathbb{E}_{x_1\sim p_1} \dots \sup_{p_T\in\mathcal{P}_T} \inf_{f_T\in\mathcal{F}} \mathbb{E}_{x_T\sim p_T} \left[\sum_{t=1}^T f_t(x_t) - \psi(x_{1:T})\right]
\end{aligned}$$

From now on, it will be understood that x_t has distribution p_t and that the suprema over p_t are in fact over $p_t \in \mathcal{P}_t(x_{1:t-1})$. By moving the expectation with respect to x_T and then the infimum with respect to f_T inside the expression, we arrive at

$$\begin{aligned}
&\sup_{p_1}\inf_{f_1}\mathbb{E}_{x_1}\dots\sup_{p_{T-1}}\inf_{f_{T-1}}\mathbb{E}_{x_{T-1}}\sup_{p_T}\left[\sum_{t=1}^{T-1} f_t(x_t) + \left[\inf_{f_T}\mathbb{E}_{x_T} f_T(x_T)\right] - \mathbb{E}_{x_T}\psi(x_{1:T})\right] \\
&= \sup_{p_1}\inf_{f_1}\mathbb{E}_{x_1}\dots\sup_{p_{T-1}}\inf_{f_{T-1}}\mathbb{E}_{x_{T-1}}\sup_{p_T}\mathbb{E}_{x_T}\left[\sum_{t=1}^{T-1} f_t(x_t) + \left[\inf_{f_T}\mathbb{E}_{x_T} f_T(x_T)\right] - \psi(x_{1:T})\right]
\end{aligned}$$

Let us now repeat the procedure for step $T-1$. The above expression is equal to

$$\begin{aligned}
&\sup_{p_1}\inf_{f_1}\mathbb{E}_{x_1}\dots\sup_{p_{T-1}}\inf_{f_{T-1}}\mathbb{E}_{x_{T-1}}\left[\sum_{t=1}^{T-1} f_t(x_t) + \sup_{p_T}\mathbb{E}_{x_T}\left[\inf_{f_T}\mathbb{E}_{x_T} f_T(x_T) - \psi(x_{1:T})\right]\right] \\
&= \sup_{p_1}\inf_{f_1}\mathbb{E}_{x_1}\dots\sup_{p_{T-1}}\left[\sum_{t=1}^{T-2} f_t(x_t) + \left[\inf_{f_{T-1}}\mathbb{E}_{x_{T-1}} f_{T-1}(x_{T-1})\right] + \mathbb{E}_{x_{T-1}}\sup_{p_T}\mathbb{E}_{x_T}\left[\inf_{f_T}\mathbb{E}_{x_T} f_T(x_T) - \psi(x_{1:T})\right]\right] \\
&= \sup_{p_1}\inf_{f_1}\mathbb{E}_{x_1}\dots\sup_{p_{T-1}}\mathbb{E}_{x_{T-1}}\sup_{p_T}\mathbb{E}_{x_T}\left[\sum_{t=1}^{T-2} f_t(x_t) + \left[\inf_{f_{T-1}}\mathbb{E}_{x_{T-1}} f_{T-1}(x_{T-1})\right] + \left[\inf_{f_T}\mathbb{E}_{x_T} f_T(x_T)\right] - \psi(x_{1:T})\right]
\end{aligned}$$

Continuing in this fashion for $T-2$ and all the way down to $t=1$ proves the theorem. □

Proof of Proposition 2. Even though Theorem 1 shows equality to some quantity with a supremum over oblivious strategies $\mathbf{p}$, it is not immediate that there exists an oblivious minimax strategy for the adversary, and a proof is required. To this end, for any oblivious strategy $\mathbf{p}$, define the regret the player would get playing optimally against $\mathbf{p}$:

$$\mathcal{V}_T^{\mathbf{p}} \triangleq \inf_{f_1\in\mathcal{F}}\mathbb{E}_{x_1\sim p_1}\inf_{f_2\in\mathcal{F}}\mathbb{E}_{x_2\sim p_2(\cdot|x_1)}\cdots\inf_{f_T\in\mathcal{F}}\mathbb{E}_{x_T\sim p_T(\cdot|x_{1:T-1})}\left[\sum_{t=1}^T f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]. \tag{10}$$

We will prove that for any oblivious strategy $\mathbf{p}$,

$$\mathcal{V}_T(\mathcal{P}_{1:T}) \geq \mathcal{V}_T^{\mathbf{p}} = \inf_{\boldsymbol{\pi}}\mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_{f_t\sim\pi_t(\cdot|x_{1:t-1})}\mathbb{E}_{x_t\sim p_t} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right] \tag{11}$$

with equality holding for $\mathbf{p}^*$ which achieves the supremum in (3). Importantly, the infimum is over strategies $\boldsymbol{\pi} = \{\pi_t\}_{t=1}^T$ of the player that *do not depend* on player's previous moves, that is $\pi_t : \mathcal{X}^{t-1} \mapsto \mathcal{Q}$.

Fix an oblivious strategy $\mathbf{p}$ and note that $\mathcal{V}_T(\mathcal{P}_{1:T}) \geq \mathcal{V}_T^{\mathbf{p}}$. From now on, it will be understood that x_t has distribution $p_t(\cdot|x_{1:t-1})$. Let $\boldsymbol{\pi} = \{\pi_t\}_{t=1}^T$ be a strategy of the player, that is, a sequence of mappings $\pi_t : (\mathcal{F}\times\mathcal{X})^{t-1} \mapsto \mathcal{Q}$.

By moving to a functional representation in Eq. (10),

$$\mathcal{V}_T^{\mathbf{P}} = \inf_{\boldsymbol{\pi}} \mathbb{E}_{f_1\sim\pi_1}\mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{f_T\sim\pi_T(\cdot|f_{1:T-1},x_{1:T-1})}\mathbb{E}_{x_T\sim p_T(\cdot|x_{1:T-1})}\left[\sum_{t=1}^T f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]$$

Note that the last term does not depend on $f_1,\dots,f_T$, and so the expression above is equal to

$$\inf_{\boldsymbol{\pi}}\Bigg\{\mathbb{E}_{f_1\sim\pi_1}\mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{f_T\sim\pi_T(\cdot|f_{1:T-1},x_{1:T-1})}\mathbb{E}_{x_T\sim p_T(\cdot|x_{1:T-1})}\left[\sum_{t=1}^T f_t(x_t)\right]$$
$$-\mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{x_T\sim p_T(\cdot|x_{1:T-1})}\left[\inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]\Bigg\}$$
$$= \inf_{\boldsymbol{\pi}}\left\{\mathbb{E}_{f_1\sim\pi_1}\mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{f_T\sim\pi_T(\cdot|f_{1:T-1},x_{1:T-1})}\mathbb{E}_{x_T\sim p_T(\cdot|x_{1:T-1})}\left[\sum_{t=1}^T f_t(x_t)\right]\right\} - \left\{\mathbb{E}\left[\inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]\right\}$$

Now, by linearity of expectation, the first term can be written as

$$\inf_{\boldsymbol{\pi}}\left\{\sum_{t=1}^T \mathbb{E}_{f_1\sim\pi_1}\mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{f_T\sim\pi_T(\cdot|f_{1:T-1},x_{1:T-1})}\mathbb{E}_{x_T\sim p_T(\cdot|x_{1:T-1})} f_t(x_t)\right\}$$
$$= \inf_{\boldsymbol{\pi}}\left\{\sum_{t=1}^T \mathbb{E}_{f_1\sim\pi_1}\mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{f_t\sim\pi_t(\cdot|f_{1:t-1},x_{1:t-1})}\mathbb{E}_{x_t\sim p_t(\cdot|x_{1:t-1})} f_t(x_t)\right\}$$
$$= \inf_{\boldsymbol{\pi}}\left\{\sum_{t=1}^T \mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{x_t\sim p_t(\cdot|x_{1:t-1})}\left[\mathbb{E}_{f_1\sim\pi_1}\dots\mathbb{E}_{f_t\sim\pi_t(\cdot|f_{1:t-1},x_{1:t-1})} f_t(x_t)\right]\right\} \tag{12}$$

Now notice that for any strategy $\boldsymbol{\pi} = \{\pi_t\}_{t=1}^T$, there is an equivalent strategy $\boldsymbol{\pi}' = \{\pi'_t\}_{t=1}^T$ that (a) gives the same value to the above expression as $\boldsymbol{\pi}$ and (b) does not depend on the past decisions of the player, that is $\pi'_t : \mathcal{X}^{t-1} \mapsto \mathcal{Q}$. To see why this is the case, fix any strategy $\boldsymbol{\pi}$ and for any t define

$$\pi'_t(\cdot|x_{1:t-1}) = \mathbb{E}_{f_1\sim\pi_1}\dots\mathbb{E}_{f_{t-1}\sim\pi_t(\cdot|f_{1:t-2},x_{1:t-2})}\pi_t(\cdot|f_{1:t-1},x_{1:t-1})$$

where we integrated out the sequence $f_1,\dots,f_{t-1}$. Then

$$\mathbb{E}_{f_1\sim\pi_1}\dots\mathbb{E}_{f_t\sim\pi_t(\cdot|f_{1:t-1},x_{1:t-1})} f_t(x_t) = \mathbb{E}_{f_t\sim\pi'_t(\cdot|x_{1:t-1})} f_t(x_t)$$

and so $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ give the same value in (12).

We conclude that the infimum in (12) can be restricted to those strategies $\boldsymbol{\pi}$ that do not depend on past randomizations of the player. In this case,

$$\mathcal{V}_T^{\mathbf{P}} = \inf_{\boldsymbol{\pi}}\left\{\sum_{t=1}^T \mathbb{E}_{x_1\sim p_1}\dots\mathbb{E}_{x_t\sim p_t(\cdot|x_{1:t-1})}\mathbb{E}_{f_t\sim\pi_t(\cdot|x_{1:t-1})} f_t(x_t)\right]\right\} - \left\{\mathbb{E}\left[\inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]\right\}$$
$$= \inf_{\boldsymbol{\pi}}\left\{\sum_{t=1}^T \mathbb{E}_{x_1,\dots,x_{t-1}}\mathbb{E}_{f_t\sim\pi_t(\cdot|x_{1:t-1})}\mathbb{E}_{x_t} f_t(x_t)\right]\right\} - \left\{\mathbb{E}\left[\inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]\right\}$$
$$= \inf_{\boldsymbol{\pi}}\mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_{f_t\sim\pi_t(\cdot|x_{1:t-1})}\mathbb{E}_{x_t\sim p_t} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right].$$

Now, notice that we can choose the Bayes optimal response f_t in each term:

$$\mathcal{V}_T^{\mathbf{P}} = \inf_{\boldsymbol{\pi}}\mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_{f_t\sim\pi_t(\cdot|x_{1:t-1})}\mathbb{E}_{x_t\sim p_t} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]$$
$$\geq \inf_{\boldsymbol{\pi}}\mathbb{E}\left[\sum_{t=1}^T \inf_{f_t\in\mathcal{F}}\mathbb{E}_{x_t\sim p_t} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]$$
$$= \mathbb{E}\left[\sum_{t=1}^T \inf_{f_t\in\mathcal{F}}\mathbb{E}_{x_t\sim p_t} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right].$$

Together with Theorem 1, this implies that

$$\mathcal{V}_T^{\mathbf{p}^*} = \mathcal{V}_T(\mathcal{P}_{1:T}) = \inf_{\boldsymbol{\pi}} \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_{f_t\sim\pi_t(\cdot|x_{1:t-1})}\mathbb{E}_{x_t\sim p_t^*} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right]$$

for any $\mathbf{p}^*$ achieving supremum in (3). Further, the infimum is over strategies that do not depend on the moves of the player.

We conclude that there is an oblivious minimax optimal strategy of the adversary, and there is a corresponding minimax optimal strategy for the player that does not depend on its own moves.

□

Proof of Theorem 3. From Eq. (3),

$$\begin{aligned}
\mathcal{V}_T &= \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sum_{t=1}^T \inf_{f_t\in\mathcal{F}} \mathbb{E}_{t-1}\left[f_t(x_t)\right] - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right] \\
&= \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \inf_{f_t\in\mathcal{F}} \mathbb{E}_{t-1}\left[f_t(x_t)\right] - f(x_t)\right\}\right] \\
&\leq \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \mathbb{E}_{t-1}\left[f(x_t)\right] - f(x_t)\right\}\right] \qquad (13)
\end{aligned}$$

The upper bound is obtained by replacing each infimum by a particular choice f. Note that $\mathbb{E}_{t-1}\left[f(x_t)\right] - f(x_t)$ is a martingale difference sequence. We now employ a symmetrization technique. For this purpose, we introduce a *tangent sequence* $\{x'_t\}_{t=1}^T$ that is constructed as follows. Let x'_1 be an independent copy of x_1. For $t \geq 2$, let x'_t be both identically distributed as x_t as well as independent of it conditioned on $x_{1:t-1}$. Then, we have, for any $t \in [T]$ and $f \in \mathcal{F}$,

$$\mathbb{E}_{t-1}\left[f(x_t)\right] = \mathbb{E}_{t-1}\left[f(x'_t)\right] = \mathbb{E}_T\left[f(x'_t)\right] \ . \qquad (14)$$

The first equality is true by construction. The second holds because x'_t is independent of $x_{t:T}$ conditioned on $x_{1:t-1}$. We also have, for any $t \in [T]$ and $f \in \mathcal{F}$,

$$f(x_t) = \mathbb{E}_T\left[f(x_t)\right] \ . \qquad (15)$$

Plugging in (14) and (15) into (13), we get,

$$\begin{aligned}
\mathcal{V}_T &\leq \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \mathbb{E}_T\left[f(x'_t)\right] - \mathbb{E}_T\left[f(x_t)\right]\right\}\right] \\
&= \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\mathbb{E}_T\left[\sum_{t=1}^T f(x'_t) - f(x_t)\right]\right\}\right] \\
&\leq \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t) - f(x_t)\right\}\right] \ .
\end{aligned}$$

For any $\mathbf{p}$, the expectation in the above supremum can be written as

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t) - f(x_t)\right\}\right] = \mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|x_1)}\ldots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|x_1,\ldots,x_{T-1})}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t) - f(x_t)\right\}\right].$$

Now, let's see what happens when we rename x_1 and x'_1 in the right-hand side of the above inequality. The equivalent expression we then obtain is

$$\mathbb{E}_{x'_1,x_1\sim p_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|x'_1)}\mathbb{E}_{x_3,x'_3\sim p_3(\cdot|x'_1,x_2)}\ldots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|x'_1,x_{2:T-1})}\left[\sup_{f\in\mathcal{F}}\left\{-(f(x'_1)-f(x_1)) + \sum_{t=2}^T f(x'_t) - f(x_t)\right\}\right].$$

Now fix any $\epsilon \in \{\pm 1\}^T$. Informally, $\epsilon_t = 1$ indicates whether we rename x_t and x'_t. It is not hard to verify that

$$\mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|x_1)}\dots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|x_1,\dots,x_{T-1})}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t)-f(x_t)\right\}\right]$$

$$= \mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(-1))}\dots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(-1),\dots,\chi_{T-1}(-1))}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t)-f(x_t)\right\}\right] \tag{16}$$

$$= \mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(\epsilon_1))}\dots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x'_t)-f(x_t))\right\}\right] \tag{17}$$

Since Eq. (16) holds for any $\epsilon \in \{\pm 1\}^T$, we conclude that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t)-f(x_t)\right\}\right] \tag{18}$$

$$= \mathbb{E}_\epsilon\mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(\epsilon_1))}\dots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x'_t)-f(x_t))\right\}\right]$$

$$= \mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x'_t)-f(x_t))\right\}\right].$$

The process above can be thought of as taking a path in a binary tree. At each step t, a coin is flipped and this determines whether x_t or x'_t is to be used in conditional distributions in the following steps. This is precisely the process outlined in (4). Using the definition of $\boldsymbol{\rho}$, we can rewrite the last expression in Eq. (18) as

$$\mathbb{E}_{(x_1,x'_1)\sim\boldsymbol{\rho}_1(\epsilon)}\mathbb{E}_{\epsilon_1}\mathbb{E}_{(x_2,x'_2)\sim\boldsymbol{\rho}_2(\epsilon)(x_1,x'_1)}\dots\mathbb{E}_{\epsilon_{T-1}}\mathbb{E}_{(x_T,x'_T)\sim\boldsymbol{\rho}_T(\epsilon)\left((x_1,x'_1),\dots,(x_{T-1},x'_{T-1})\right)}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t(f(x_t)-f(x'_t))\right\}\right].$$

More succinctly, Eq. (18) can be written as

$$\mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(\mathbf{x}'_t(-\mathbf{1}))-f(\mathbf{x}_t(-\mathbf{1}))\right\}\right] = \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t(f(\mathbf{x}_t(\epsilon))-f(\mathbf{x}'_t(\epsilon)))\right\}\right]. \tag{19}$$

It is worth emphasizing that the values of the mappings $\mathbf{x},\mathbf{x}'$ are drawn conditionally-independently, however the distribution depends on the ancestors in *both* trees. In some sense, the path ϵ defines "who is tangent to whom".

We now split the supremum into two:

$$\mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t(f(\mathbf{x}_t(\epsilon))-f(\mathbf{x}'_t(\epsilon)))\right\}\right]$$

$$\le \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon))\right] + \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T -\epsilon_t f(\mathbf{x}'_t(\epsilon))\right] \tag{20}$$

$$= 2\mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon))\right]$$

The last equality is not difficult to verify but requires understanding the symmetry between the paths in the $\mathbf{x}$ and $\mathbf{x}'$ trees. This symmetry implies that the two terms in Eq. (20) are equal. Each $\epsilon \in$

$\{\pm 1\}^T$ in the first term defines time steps t when values in $\mathbf{x}$ are used in conditional distributions. To any such ϵ, there corresponds a $-\epsilon$ in the second term which defines times when values in $\mathbf{x}'$ are used in conditional distributions. This implies the required result. As a more concrete example, consider the path $\epsilon = -\mathbf{1}$ in the first term. The contribution to the overall expectation is the supremum over $f \in \mathcal{F}$ of evaluation of $-f$ on the left-most path of the $\mathbf{x}$ tree which is defined as successive draws from distributions p_t conditioned on the values on the left-most path, irrespective of the $\mathbf{x}'$ tree. Now consider the corresponding path $\epsilon = \mathbf{1}$ in the second term. Its contribution to the overall expectation is a supremum over $f \in \mathcal{F}$ of evaluation of $-f$ on the right-most path of the $\mathbf{x}'$ tree, defined as successive draws from distributions p_t conditioned on the values on the right-most path, irrespective of the $\mathbf{x}$ tree. Clearly, the contributions are the same, and the same argument can be done for any path ϵ.

Alternatively, we can see that the two terms in Eq. (20) are equal by expanding the notation. We thus claim that

$$\mathbb{E}_{x_1,x_1'\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x_2'\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\mathbb{E}_{x_T,x_T'\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t f(x_t')\right\}\right]$$
$$= \mathbb{E}_{x_1,x_1'\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x_2'\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\mathbb{E}_{x_T,x_T'\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t f(x_t)\right\}\right]$$

The identity can be verified by simultaneously renaming $\mathbf{x}$ with $\mathbf{x}'$ and ϵ with $-\epsilon$. Since $\chi(x, x', \epsilon) = \chi(x', x, -\epsilon)$, the distributions in the two expressions are the same while the sum of the first term becomes the sum of the second term.

More generally, the split of Eq. (20) can be performed via an additional "centering" term. For any t, let M_t be a function with the property $M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon) = M_t(\mathbf{p}, f, \mathbf{x}', \mathbf{x}, -\epsilon)$

We then have

$$\begin{aligned}
&\mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t(f(\mathbf{x}_t(\epsilon)) - f(\mathbf{x}_t'(\epsilon)))\right\}\right]\\
&\le \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T \epsilon_t(f(\mathbf{x}_t(\epsilon)) - M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon))\right]\\
&+ \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T -\epsilon_t(f(\mathbf{x}_t'(\epsilon)) - M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon))\right]\\
&= 2\mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T \epsilon_t(f(\mathbf{x}_t(\epsilon)) - M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon))\right]
\end{aligned} \tag{21}$$

To verify equality of the two terms in (21) we can expand the notation.

$$\mathbb{E}_{x_1,x_1'\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x_2'\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\mathbb{E}_{x_T,x_T'\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}$$
$$\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x_t') - M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon))\right\}\right]$$
$$= \mathbb{E}_{x_1,x_1'\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x_2'\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\mathbb{E}_{x_T,x_T'\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}$$
$$\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t(f(x_t) - M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon))\right\}\right]$$

□

Proof of Corollary 4. Define a function M_t as the conditional expectation

$$M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon) = \mathbb{E}_{x\sim p_t(\cdot|\chi_1(\epsilon_1),\dots,\chi_{t-1}(\epsilon_{t-1}))} f(x).$$

The property $M_t(\mathbf{p}, f, \mathbf{x}, \mathbf{x}', \epsilon) = M_t(\mathbf{p}, f, \mathbf{x}', \mathbf{x}, -\epsilon)$ holds because $\chi(x, x', \epsilon) = \chi(x', x, -\epsilon)$. □

Proof of Proposition 5. By definition, we have,

$$\mathfrak{R}_T(\mathcal{F}, \mathbf{p}) = \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon))\right] \tag{22}$$

In the i.i.d. case, however, the tree generation according to the $\boldsymbol{\rho}$ process simplifies: for any $\epsilon \in \{\pm 1\}^T, t \in [T]$,

$$(\mathbf{x}_t(\epsilon), \mathbf{x}'_t(\epsilon)) \sim p \times p\ .$$

Thus, the $2 \cdot (2^T - 1)$ random variables $\mathbf{x}_t(\epsilon), \mathbf{x}'_t(\epsilon)$ are all i.i.d. drawn from p. Writing the expectation (22) explicitly as an average over paths, we get

$$\begin{aligned}\mathfrak{R}_T(\mathcal{F}, \mathbf{p}) &= \frac{1}{2^T}\sum_{\epsilon\in\{\pm1\}^T} \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon))\right] \\ &= \frac{1}{2^T}\sum_{\epsilon\in\{\pm1\}^T} \mathbb{E}_{x_1,\ldots,x_T\sim p} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(x_t)\right] \\ &= \mathbb{E}_\epsilon \mathbb{E}_{x_1,\ldots,x_T\sim p} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(x_t)\right]\ .\end{aligned}$$

The second equality holds because, for any fixed path ϵ, the T random variables $\{\mathbf{x}_t(\epsilon)\}_{t\in[T]}$ have joint distribution p^T. This proves the first claim.

We now prove the second claim. To make the $\boldsymbol{\rho}$ process associated with $\mathbf{p}$ more explicit, we use the expanded definition:

$$\begin{aligned}&\mathfrak{R}_T(\mathcal{F}, \mathbf{p}) \\ &= \mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\ \ldots\ \mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(\epsilon_1),\ldots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(x_t)\right] \\ &\le \sup_{x_1,x'_1}\mathbb{E}_{\epsilon_1}\sup_{x_2,x'_2}\mathbb{E}_{\epsilon_2}\ \ldots\ \sup_{x_T,x'_T}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(x_t)\right] \qquad (23) \\ &= \sup_{x_1}\mathbb{E}_{\epsilon_1}\sup_{x_2}\mathbb{E}_{\epsilon_2}\ \ldots\ \sup_{x_T}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t f(x_t)\right] \\ &= \mathfrak{R}_T(\mathcal{F})\ .\end{aligned}$$

The inequality holds by replacing expectation over x_t, x'_t by a supremum over the same. We then get rid of x'_t's since they do not appear anywhere. □

Proof of Corollary 7. The first steps follow the proof of Theorem 3:

$$\mathcal{V}_T \le \sup_{\mathbf{p}\in\mathfrak{P}} \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t) - f(x_t)\right\}\right]$$

and for a fixed $\mathbf{p} \in \mathfrak{P}$,

$$\begin{aligned}&\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(x'_t) - f(x_t)\right\}\right] \qquad (24) \\ &= \mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\ldots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(\epsilon_1),\ldots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x'_t) - f(x_t))\right\}\right].\end{aligned}$$

At this point we pass to an upper bound, unlike the proof of Theorem 3. Notice that $p_t(\cdot|\chi_1(\epsilon_1), \ldots, \chi_{t-1}(\epsilon_{t-1}))$ is a distribution with support in $\mathcal{X}_t(\chi_1(\epsilon_1), \ldots, \chi_{t-1}(\epsilon_{t-1}))$. That is, the sequence $\chi_1(\epsilon_1), \ldots, \chi_{t-1}(\epsilon_{t-1})$ defines the constraint at time t. Passing from $t = T$ down to $t = 1$, we can replace all the expectations over p_t by the suprema over the set $\mathcal{X}_t$, only increasing the value:

$$\mathbb{E}_{x_1,x'_1\sim p_1}\mathbb{E}_{\epsilon_1}\mathbb{E}_{x_2,x'_2\sim p_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\mathbb{E}_{x_T,x'_T\sim p_T(\cdot|\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x'_t)-f(x_t))\right\}\right]$$
$$\le \sup_{x_1,x'_1\in\mathcal{X}_1}\mathbb{E}_{\epsilon_1}\sup_{x_2,x'_2\in\mathcal{X}_2(\cdot|\chi_1(\epsilon_1))}\mathbb{E}_{\epsilon_2}\dots\sup_{x_T,x'_T\in\mathcal{X}_T(\chi_1(\epsilon_1),\dots,\chi_{T-1}(\epsilon_{T-1}))}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(x'_t)-f(x_t))\right\}\right]$$
$$= \sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T -\epsilon_t(f(\mathbf{x}'_t(\epsilon))-f(\mathbf{x}_t(\epsilon)))\right\}\right]$$

In the last equality, we passed to the tree representation. Indeed, at each step, we are choosing x_t, x'_t from the appropriate set and then flipping a coin ϵ_t which decides which of x_t, x'_t will be used to define the constraint set through $\chi_t(\epsilon_t)$. This once again defines a tree structure and we may pass to the supremum over trees $(\mathbf{x},\mathbf{x}')\in\mathcal{T}$. However, $\mathcal{T}$ is not a set of all possible $\mathcal{X}$-valued trees: for each t, $\mathbf{x}_t(\epsilon),\mathbf{x}'_t(\epsilon)\in\mathcal{X}_t(\chi_1(\mathbf{x}_1,\mathbf{x}'_1,\epsilon_1),\dots,\chi_{t-1}(\mathbf{x}_{t-1}(\epsilon_{t-1}),\mathbf{x}'_{t-1}(\epsilon_{t-1}),\epsilon_{t-1}))$. That is, the choice at each node of the tree is constrained by the values of both trees according to the path. As before, the left-most path of the $\mathbf{x}$ tree (as well as the right-most path of the $\mathbf{x}'$ tree) is defined by constraints applied to the values on the path only disregarding the other tree.

The rest of the proof exactly follows the proof of Theorem 3. □

Proof of Proposition 8. Let $M_t(f,\mathbf{x},\mathbf{x}',\epsilon)=\frac{1}{t-1}\sum_{\tau=1}^{t-1}f(\chi_\tau(\epsilon_\tau))$. Note that since $\chi(x,x',\epsilon)=\chi(x',x,-\epsilon)$, we have that $M_t(f,\mathbf{x},\mathbf{x}',\epsilon)=M_t(f,\mathbf{x}',\mathbf{x},-\epsilon)$. Using 7 we conclude that

$$\mathcal{V}_T \le 2\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T\epsilon_t\left(\langle f,\mathbf{x}_t(\epsilon)\rangle-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\langle f,\chi_\tau(\epsilon_\tau)\rangle\right)\right]$$
$$= 2\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\sup_{f\in\mathcal{F}}\left\langle f,\sum_{t=1}^T\epsilon_t\left(\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right)\right\rangle\right]$$

By linearity and Fenchel's inequality, the last expression is upper bounded by

$$\frac{2}{\alpha}\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\sup_{f\in\mathcal{F}}\left\langle f,\alpha\sum_{t=1}^T\epsilon_t\left(\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right)\right\rangle\right]$$
$$\le \frac{2}{\alpha}\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\sup_{f\in\mathcal{F}}\Psi(f)+\Psi^*\left(\alpha\sum_{t=1}^T\epsilon_t\left(\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right)\right)\right]$$
$$\le \frac{2}{\alpha}\left(\sup_{f\in\mathcal{F}}\Psi(f)+\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\Psi^*\left(\alpha\sum_{t=1}^T\epsilon_t\left(\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right)\right)\right]\right)$$
$$\le \frac{2R^2}{\alpha}+\frac{2}{\alpha}\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_{\epsilon}\left[\Psi^*\left(\alpha\sum_{t=1}^T\epsilon_t\left(\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right)\right)\right]$$
$$\le \frac{2R^2}{\alpha}+\frac{\alpha}{\lambda}\sum_{t=1}^T\mathbb{E}_{\epsilon}\left[\left\|\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right\|_*^2\right] \tag{25}$$

Where the last step follows from Lemma 2 of [5] (with a slight modification). However since $(\mathbf{x},\mathbf{x}')\in\mathcal{T}$ are pairs of tree such that for any $\epsilon\in\{\pm1\}^T$ and any $t\in[T]$.

$$C(\chi_1(\epsilon_1),\dots,\chi_{t-1}(\epsilon_{t-1}),\mathbf{x}_t(\epsilon))=1$$

we can conclude that for any $\epsilon\in\{\pm1\}^T$ and any $t\in[T]$,

$$\left\|\mathbf{x}_t(\epsilon)-\frac{1}{t-1}\sum_{\tau=1}^{t-1}\chi_\tau(\epsilon_\tau)\right\|_*\le\sigma_t$$

Using this with Equation 25 and the fact that α is arbitrary, we can conclude that

$$\mathcal{V}_T \le \inf_{\alpha>0}\left\{\frac{2R^2}{\alpha} + \frac{\alpha}{\lambda}\sum_{t=1}^{T}\sigma_t^2\right\} \le 2\sqrt{2}R\sqrt{\sum_{t=1}^{T}\sigma_t^2}$$

□

Proof of Proposition 9. Let $M_t(f,\mathbf{x},\mathbf{x}',\epsilon) = f(\chi_{t-1}(\epsilon_{t-1}))$. Note that since $\chi(x,x',\epsilon) = \chi(x',x,-\epsilon)$ we have that $M_t(f,\mathbf{x},\mathbf{x}',\epsilon) = M_t(f,\mathbf{x}',\mathbf{x},-\epsilon)$. Using 7 we conclude that

$$\begin{aligned}\mathcal{V}_T &\le 2\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^{T}\epsilon_t\left(\langle f,\mathbf{x}_t(\epsilon)\rangle - \langle f,\chi_{t-1}(\epsilon_{t-1})\rangle\right)\right]\\ &= 2\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left\langle f,\sum_{t=1}^{T}\epsilon_t\left(\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right)\right\rangle\right]\end{aligned}$$

As before, using linearity and Fenchel's inequality we pass to the upper bound

$$\begin{aligned}&\frac{2}{\alpha}\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\left\langle f,\alpha\sum_{t=1}^{T}\epsilon_t\left(\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right)\right\rangle\right]\\ &\le \frac{2}{\alpha}\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_\epsilon\left[\sup_{f\in\mathcal{F}}\Psi(f) + \Psi^*\left(\alpha\sum_{t=1}^{T}\epsilon_t\left(\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right)\right)\right]\\ &\le \frac{2}{\alpha}\left(\sup_{f\in\mathcal{F}}\Psi(f) + \sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_\epsilon\left[\Psi^*\left(\alpha\sum_{t=1}^{T}\epsilon_t\left(\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right)\right)\right]\right)\\ &\le \frac{2R^2}{\alpha} + \frac{2}{\alpha}\sup_{(\mathbf{x},\mathbf{x}')\in\mathcal{T}}\mathbb{E}_\epsilon\left[\Psi^*\left(\alpha\sum_{t=1}^{T}\epsilon_t\left(\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right)\right)\right]\\ &\le \frac{2R^2}{\alpha} + \frac{\alpha}{\lambda}\sum_{t=1}^{T}\mathbb{E}_\epsilon\left[\left\|\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right\|_*^2\right] \qquad (26)\end{aligned}$$

Where the last step follows from Lemma 2 of [5] (with slight modification). However since $(\mathbf{x},\mathbf{x}')\in\mathcal{T}$ are pairs of tree such that for any $\epsilon\in\{\pm1\}^T$ and any $t\in[T]$.

$$C(\chi_1(\epsilon_1),\ldots,\chi_{t-1}(\epsilon_{t-1}),\mathbf{x}_t(\epsilon)) = 1$$

we can conclude that for any $\epsilon\in\{\pm1\}^T$ and any $t\in[T]$,

$$\left\|\mathbf{x}_t(\epsilon)-\chi_{t-1}(\epsilon_{t-1})\right\|_* \le \delta$$

Using this with Equation 26 and the fact that α is arbitrary, we can conclude that

$$\mathcal{V}_T \le \inf_{\alpha>0}\left\{\frac{2R^2}{\alpha} + \frac{\alpha\delta^2 T}{\lambda}\right\} \le 2R\delta\sqrt{2T}$$

□

Proof of Theorem 10. First, using the fact that the maximum of a linear functional over a simplex is achieved at the corners,

$$\begin{aligned}\mathcal{V}_T &= \inf_{q_1}\sup_{x_1}\mathop{\mathbb{E}}_{\substack{f_1\sim q_1\\ s_1\sim\sigma}}\ldots\inf_{q_T}\sup_{x_T}\mathop{\mathbb{E}}_{\substack{f_T\sim q_T\\ s_T\sim\sigma}}\left[\sum_{t=1}^{T}f_t(\omega(x_t,s_t)) - \inf_{f\in\mathcal{F}}\sum_{t=1}^{T}f(\omega(x_t,s_t))\right]\\ &= \inf_{q_1}\sup_{p_1}\mathop{\mathbb{E}}_{\substack{f_1\sim q_1,x_1\sim p_1\\ s_1\sim\sigma}}\ldots\inf_{q_T}\sup_{p_T}\mathop{\mathbb{E}}_{\substack{f_T\sim q_T,x_T\sim p_T\\ s_T\sim\sigma}}\left[\sum_{t=1}^{T}f_t(\omega(x_t,s_t)) - \inf_{f\in\mathcal{F}}\sum_{t=1}^{T}f(\omega(x_t,s_t))\right].\end{aligned}$$

Next, appealing to the minimax theorem, the last quantity is equal to

$$\sup_{p_1}\inf_{f_1}\mathop{\mathbb{E}}_{\substack{x_1\sim p_1\\ s_1\sim\sigma}}\dots\sup_{p_T}\inf_{f_T}\mathop{\mathbb{E}}_{\substack{x_T\sim p_T\\ s_T\sim\sigma}}\left[\sum_{t=1}^T f_t(\omega(x_t,s_t)) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(\omega(x_t,s_t))\right]$$

Using the technique of [1, 10], we can rewrite the last quantity as

$$= \sup_{p_1}\mathop{\mathbb{E}}_{\substack{x_1\sim p_1\\ s_1\sim\sigma}}\dots\sup_{p_T}\mathop{\mathbb{E}}_{\substack{x_T\sim p_T\\ s_T\sim\sigma}}\left[\sum_{t=1}^T \inf_{f_t}\mathbb{E}_{x'_t,s'_t} f_t(\omega(x'_t,s'_t)) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(\omega(x_t,s_t))\right]$$

where x'_t has the same distribution as x_t conditioned on the history up to time t. Further, the s'_t sequence is i.i.d. with distribution σ. Rewriting the above, we arrive at

$$\sup_{p_1}\mathop{\mathbb{E}}_{\substack{x_1\sim p_1\\ s_1\sim\sigma}}\dots\sup_{p_T}\mathop{\mathbb{E}}_{\substack{x_T\sim p_T\\ s_T\sim\sigma}}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \inf_{f_t}\mathbb{E}_{x'_t,s'_t} f_t(\omega(x'_t,s'_t)) - \sum_{t=1}^T f(\omega(x_t,s_t))\right\}\right]$$

$$\le \sup_{p_1}\mathop{\mathbb{E}}_{\substack{x_1\sim p_1\\ s_1\sim\sigma}}\dots\sup_{p_T}\mathop{\mathbb{E}}_{\substack{x_T\sim p_T\\ s_T\sim\sigma}}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \mathbb{E}_{x'_t,s'_t} f(\omega(x'_t,s'_t)) - \sum_{t=1}^T f(\omega(x_t,s_t))\right\}\right]$$

$$\le \sup_{p_1}\mathop{\mathbb{E}}_{\substack{x_1,x'_1\sim p_1\\ s_1,s'_1\sim\sigma}}\dots\sup_{p_T}\mathop{\mathbb{E}}_{\substack{x_T,x'_T\sim p_T\\ s_T,s'_T\sim\sigma}}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T f(\omega(x'_t,s'_t)) - \sum_{t=1}^T f(\omega(x_t,s_t))\right\}\right]$$

where we've substituted f_t with a suboptimal choice f, and then used Jensen's inequality. The expectation over x_t, x'_t can be upper bounded by the suprema, yielding

$$\sup_{x_1,x'_1}\mathop{\mathbb{E}}_{s_1,s'_1\sim\sigma}\mathbb{E}_{\epsilon_1}\dots\sup_{x_T,x'_T}\mathop{\mathbb{E}}_{s_T,s'_T\sim\sigma}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\left\{\sum_{t=1}^T \epsilon_t(f(\omega(x'_t,s'_t)) - f(\omega(x_t,s_t)))\right\}\right]$$

$$\le 2\sup_{x_1}\mathop{\mathbb{E}}_{s_1\sim\sigma}\mathbb{E}_{\epsilon_1}\dots\sup_{x_T}\mathop{\mathbb{E}}_{s_T\sim\sigma}\mathbb{E}_{\epsilon_T}\left[\sup_{f\in\mathcal{F}}\sum_{t=1}^T \epsilon_t f(\omega(x_t,s_t)\right]$$

□

Proof of Lemma 11. Let us calculate the probability that for no distinct $t, t' \in [T]$ do we have z_t+s_t and $z_{t'}+s_{t'}$ in the same "bin" $[\theta_i, \theta_{i+1})$. We can deal with the boundary behavior by ensuring that $\mathcal{F}$ is in fact a set of thresholds that is $\gamma/2$-away from 0 or 1, but we will omit this discussion for the sake of clarity. The probability that no two elements z_t+s_t and $z_{t'}+s_{t'}$ fall into the same bin clearly depends on the behavior of the adversary in choosing x_t's. Keeping in mind that the distribution of all s_t's is uniform on $[-\gamma/2, \gamma/2]$, we see that the probability of a collision is maximized when z_t is chosen to be constant throughout the T rounds. To see this, let us recast the problem as throwing balls into bins. Observe that the choice of z_t defines the set of γT^a bins into which the ball z_t+s_t falls. To maximize the probability of a "collision", the set of bins should be kept the same for all T rounds.

Now, for z_t's constant throughout the game, we have reduced the problem to that of T balls falling uniformly into $\gamma T^a > T$ bins. The probability of two elements $z_t + s_t$ and $z_t + s_{t'}$ falling into the same bin is

$$P\,(\text{no two balls fall into same bin}) = \frac{\gamma T^a(\gamma T^a-1)\cdots(\gamma T^a - T)}{\gamma T^a\cdot\gamma T^a\cdots\gamma T^a}$$
$$\ge \left(\frac{\gamma T^a - T}{\gamma T^a}\right)^T = \left(1-\frac{1}{\gamma T^{a-1}}\right)^{\frac{\gamma T^{a-1}}{\gamma T^{a-2}}}$$

The last term is approximately $\exp\left\{-1/(\gamma T^{a-2})\right\}$ for large T, so

$$P\,(\text{no two balls fall into same bin}) \ge 1 - \frac{1}{\gamma T^{a-2}}$$

using $e^{-x} \ge 1-x$. □

Proof of Proposition 12. The idea for the proof is the following. By discretizing the interval into bins of size well below the noise level, we can guarantee with high probability that no two smoothed choices $z_t + s_t$ of the adversary fall into the same bin. If this is the case, then the supremum of Theorem 10 can be taken over a discretized set of thresholds. Now, for each fixed threshold f, $\epsilon_t f(\omega(x_t, s_t))$ forms a martingale difference sequence, yielding the desired bound.

For any $f_\theta \in \mathcal{F}$, define

$$M_t^\theta = \epsilon_t f_\theta(\omega(x_t, s_t)) = \epsilon_t \left| y_t - \mathbf{1}\left\{ z_t + s_t < \theta \right\} \right| .$$

Note that $\{M_t^\theta\}_t$ is a zero-mean martingale difference sequence, that is $\mathbb{E}[M_t | z_{1:t}, y_{1:t}, s_{1:t}] = 0$. We conclude that for any fixed $\theta \in [0, 1]$,

$$P\left(\sum_{t=1}^{T} M_t^\theta \geq \epsilon \right) \leq \exp\left\{ -\frac{\epsilon^2}{2T} \right\}$$

by Azuma-Hoeffding's inequality. Let $\mathcal{F}' = \{f_{\theta_1}, \ldots, f_{\theta_N}\} \subset \mathcal{F}$ be obtained by discretizing the interval $[0, 1]$ into $N = T^a$ bins $[\theta_i, \theta_{i+1})$ of length T^{-a}, for some $a \geq 3$. Then

$$P\left(\max_{f_\theta \in \mathcal{F}'} \sum_{t=1}^{T} M_t^\theta \geq \epsilon \right) \leq N \exp\left\{ -\frac{\epsilon^2}{2T} \right\} .$$

Observe that the maximum over the discretization coincides with the supremum over the class $\mathcal{F}$ if no two elements $z_t + s_t$ and $z_{t'} + s_{t'}$ fall into the same interval $[\theta_i, \theta_{i+1})$. Indeed, in this case all the possible values of $\mathcal{F}$ on the set $\{z_1 + s_1, \ldots, z_T + s_T\}$ are obtained by choosing the discrete thresholds in $\mathcal{F}'$.

By Lemma 11,

$$\begin{aligned}
&P\left(\sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \epsilon_t f(\omega(x_t, s_t)) \geq \epsilon \right) \\
&\leq P\left(\sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \epsilon_t f(\omega(x_t, s_t)) \geq \epsilon \,\wedge\, \text{none of } (z_t + s_t)\text{'s fall into same bin} \right) \\
&+ P\left(\text{some of } (z_t + s_t)\text{'s fall into same bin} \right) \\
&= P\left(\max_{f_\theta \in \mathcal{F}'} \sum_{t=1}^{T} M_t^\theta \geq \epsilon \,\wedge\, \text{none of } (z_t + s_t)\text{'s fall into same bin} \right) + \frac{1}{\gamma T^{a-2}} \\
&\leq P\left(\max_{f_\theta \in \mathcal{F}'} \sum_{t=1}^{T} M_t^\theta \geq \epsilon \right) + \frac{1}{\gamma T^{a-2}} \\
&\leq T^a \exp\left\{ -\frac{\epsilon^2}{2T} \right\} + \frac{1}{\gamma T^{a-2}} \ .
\end{aligned}$$

Using the above and the fact that for any $f \in \mathcal{F}$, $|\sum_{t=1}^{T} \epsilon_t f(\omega(x_t, s_t))| \leq T$ we can conclude that

$$\begin{aligned}
\mathcal{V}_T &\leq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \sum_{t=1}^{T} \epsilon_t f(\omega(x_t, s_t)) \right] \\
&\leq \epsilon + T^{a+1} \exp\left\{ -\frac{\epsilon^2}{2T} \right\} + \frac{T^{3-a}}{\gamma} \ .
\end{aligned}$$

Setting $\epsilon = \sqrt{2(a+1) T \log T}$ we conclude that

$$\mathcal{V}_T \leq 1 + \sqrt{2(a+1) T \log T} + \frac{T^{3-a}}{\gamma} \ .$$

Pick $a = 3 + \frac{\log(1/\gamma)}{\log T}$ (this choice is fine because $\gamma T^{a-1} = T^2$ which grows with T as needed for the previous approximation). Hence we see that

$$\begin{aligned}\mathcal{V}_T &\leq 2 + \sqrt{2\left(4 + \frac{\log(1/\gamma)}{\log T}\right) T \log T} \\ &= 2 + \sqrt{2T\left(4\log T + \log(1/\gamma)\right)}\,.\end{aligned}$$

□

Proof of Proposition 13. As in the one dimensional case, we divide the surface of the sphere into bins (e.g. via tessellation of the sphere), with diameter T^{-a}, for some $a > 1$. Then the volume of each bin is at most $O(T^{-(d-1)a})$. Once again, the choice of z_t is deciding on the set of $\Omega(\gamma^{d-1}T^{(d-1)a})$ bins. The probability of two perturbed values in the sequence falling into the same bin is maximized when z_t is kept constant. In this case, with the same calculation as for the one-dimensional case, the probability of a collision is at most $O(\gamma^{1-d}T^{2-(d-1)a})$.

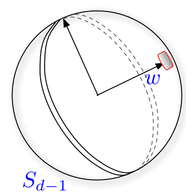

Figure 1: As w varies over the small bin, only a small number of bins change the side of the hyperplane $\langle w, z\rangle$

It remains to show that for any $w \in S_{d-1}$, we can pass to the center of the associated bin at the cost of a small number of bins changing the side of the hyperplane. It is not hard to see that all such bins form a narrow "ring". The number of bins is thus $O(\mathrm{v}_{d-2} \cdot T^a)$, where v_{d-2} is the volume of a $d-2$-dimensional "ring" on the sphere S_{d-1}.

The final result is obtained by choosing $a = \frac{\log 1/\gamma}{\log T} + \frac{3}{d-1}$, similarly to the proof of Proposition 12.

□

B Application: The I.I.D. Adversary

In this section, we consider an adversary who is restricted to draw the moves from a fixed distribution p throughout the game. That is, the time-invariant restrictions are $\mathcal{P}_t(x_{1:t-1}) = \{p\}$. A reader will notice that the definition of the value in (1) forces the restrictions $\mathcal{P}_{1:T}$ to be known to the player before the game. This, in turn, means that the distribution p is known to the learner. In some sense, the problem becomes not interesting, as there is no learning to be done. This is indeed an artifact of the minimax formulation in the *extensive form*. To circumvent the problem, we are forced to define a new value of the game in terms of *strategies*. Such a formulation does allow us to "hide" the distribution from the player since we can talk about "mappings" instead of making the information explicit. We then show two novel results. First, the regret-minimization game with i.i.d. data when the player does *not* observe the distribution p is equivalent (in terms of learnability) to the classical batch learning problem. Second, for supervised learning, when it comes to minimizing regret, the knowledge of p does not help the learner for some distributions.

Let us first define some relevant quantities. Let $\mathbf{s} = \{s_t\}_{t=1}^T$ be a T-round strategy for the player, with $s_t : (\mathcal{F} \times \mathcal{X})^{t-1} \to \mathcal{Q}$. The game where the player does not observe the i.i.d. distribution of the adversary will be called a *distribution-blind* i.i.d. game, and its minimax value will be called the *distribution-blind minimax value*:

$$\mathcal{V}_T^{\text{blind}} \triangleq \inf_{\mathbf{s}} \sup_{p} \left[\mathbb{E}_{x_1,\ldots,x_T\sim p}\mathbb{E}_{f_1\sim s_1}\ldots\mathbb{E}_{f_T\sim s_T(x_{1:T-1},f_{1:T-1})}\left\{\sum_{t=1}^T f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right\}\right]$$

Furthermore, define the value for a general (not necessarily supervised) setting:

$$\mathcal{V}_T^{\text{batch}} \triangleq \inf_{\hat{f}_T} \sup_{p\in\mathcal{P}} \left\{ \mathbb{E}\hat{f}_T - \inf_{f\in\mathcal{F}} \mathbb{E}f \right\}$$

For a distribution p, the value (1) of the online i.i.d. game, as defined through the restrictions $\mathcal{P}_t = \{p\}$ for all t, will be written as $\mathcal{V}_T(\{p\})$. For the non-blind game, we say that the problem is online learnable in the i.i.d. setting if $\sup_p \mathcal{V}_T(\{p\}) \to 0$.

We now proceed to study relationships between online and batch learnability.

Theorem 14. *For a given function class $\mathcal{F}$, online learnability in the distribution-blind game is equivalent to batch learnability. That is,* $\mathcal{V}_T^{blind}/T \to 0$ *if and only if* $\mathcal{V}_T^{batch} \to 0$.

At this point, the reader might wonder if the game formulation studied in the rest of the paper, with the restrictions known to the player, is any easier than batch and distribution-blind learning. In the next section, we show that this is not the case for supervised learning.

B.1 Distribution-Blind vs Non-Blind Supervised Learning

In the supervised game, at time t, the player picks a function $f_t \in [-1,1]^{\mathcal{X}}$, the adversary provides input-target pair (x_t, y_t), and the player suffers loss $|f_t(x_t) - y_t|$. The value of the online supervised learning game for general restrictions $\mathcal{P}_{1:T}$ is defined as

$$\mathcal{V}_T^{\text{sup}}(\mathcal{P}_{1:T}) \triangleq \inf_{q_1\in\mathcal{Q}} \sup_{p_1\in\mathcal{P}_1} \mathop{\mathbb{E}}_{f_1,(x_1,y_1)} \cdots \inf_{q_T\in\mathcal{Q}} \sup_{p_T\in\mathcal{P}_T} \mathop{\mathbb{E}}_{f_T,(x_T,y_T)} \left[\sum_{t=1}^T |f_t(x_t) - y_t| - \inf_{f\in\mathcal{F}} \sum_{t=1}^T |f(x_t) - y_t| \right]$$

where (x_t, y_t) has distribution p_t. As before, the value of an i.i.d. supervised game with a distribution $p_{X\times Y}$ will be written as $\mathcal{V}_T^{\text{sup}}(p_{X\times Y})$. The distribution-blind supervised value is defined as

$$\mathcal{V}_T^{\text{blind, sup}} \triangleq \inf_{\mathbf{s}} \sup_p \left[\mathbb{E}_{z_{1:T}\sim p} \mathbb{E}_{f_1\sim s_1} \ldots \mathbb{E}_{f_T\sim s_T(z_{1:T-1}, f_{1:T-1})} \left\{ \sum_{t=1}^T |f_t(x_t) - y_t| - \inf_{f\in\mathcal{F}} \sum_{t=1}^T |f(x_t) - y_t| \right\} \right]$$

where we use the shorthand $z_t = (x_t, y_t)$ for each t, and the batch supervised value for the absolute loss is defined as

$$\mathcal{V}_T^{\text{batch, sup}} = \inf_{\hat{f}} \sup_{P_{X\times Y}} \left\{ \mathbb{E}|Y - \hat{f}(X)| - \inf_{f\in\mathcal{F}} |Y - f(X)| \right\} \tag{27}$$

The following relationships hold:

Lemma 15. *In the supervised case,*

$$\frac{1}{4} T \mathcal{V}_T^{\textit{batch, sup}} \le \sup_{p_X} \mathfrak{R}_T(\mathcal{F}, p_X) \le \sup_{p_X} \mathcal{V}_T^{\textit{sup}}(\{p_X \times U_Y\}) \le \sup_{p_{X\times Y}} \mathcal{V}_T^{\textit{sup}}(\{p_{X\times Y}\}) \le \mathcal{V}_T^{\textit{blind, sup}}$$

where $\mathfrak{R}_T(\mathcal{F}, p_X)$ is the classical Rademacher complexity, and U_Y is the Rademacher distribution.

Theorem 14, specialized to the supervised setting, says that $\frac{1}{T}\mathcal{V}_T^{\text{blind, sup}} \to 0$ if and only if $\mathcal{V}_T^{\text{batch, sup}} \to 0$. Since $\sup_{p_{X\times Y}} \frac{1}{T}\mathcal{V}_T^{\text{sup}}(\{p_{X\times Y}\})$ is sandwiched between these two values, we conclude the following.

Corollary 16. *Either the supervised problem is learnable in the batch sense (and, by Theorem 14, in the distribution-blind online sense), in which case $\sup_{p_{X\times Y}} \mathcal{V}_T^{sup}(\{p_{X\times Y}\}) = o(T)$. Or, the problem is not learnable in the batch (and the distribution-blind sense), in which case it is not learnable for all distributions in the online sense: $\sup_{p_{X\times Y}} \mathcal{V}_T^{sup}(\{p_{X\times Y}\})$ does not grow sublinearly.*

B.2 Proofs

Proof of Theorem 14. With a proof along the lines of Proposition 2 we establish that

$$\frac{1}{T}\mathcal{V}_T^{\text{blind}} = \inf_{\mathbf{s}} \sup_p \left\{ \frac{1}{T} \sum_{t=1}^T \mathbb{E}_{x_1,\ldots,x_t\sim p} \mathbb{E}_{f_t\sim s_t(x_{1:t-1}, f_{1:t-1})} [f_t(x_t)] - \mathbb{E}_{x_1,\ldots,x_T\sim p} \left[\inf_{f\in\mathcal{F}} \frac{1}{T} \sum_{t=1}^T f(x_t) \right] \right\}$$

$$\ge \inf_{\mathbf{s}} \sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_T\sim p} \left[\frac{1}{T} \sum_{t=1}^T \mathbb{E}_{f_t\sim s_t(x_1,\ldots,x_{t-1})} [\mathbb{E}_{x\sim p}[f_t(x)]] \right] - \inf_{f\in\mathcal{F}} \mathbb{E}_{x_1,\ldots,x_T\sim p} \left[\frac{1}{T} \sum_{t=1}^T f(x_t) \right] \right\}$$

where in the second line we passed to strategies that do not depend on their own randomizations. The argument for this can be found in the proof of Proposition 2. The last expression can be conveniently written as

$$\frac{1}{T}\mathcal{V}_T^{\text{blind}} \geq \inf_{\mathbf{s}} \sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\mathbb{E}_{r\sim \text{Unif}[T-1]} \mathbb{E}_{f\sim s_{r+1}(x_1,\ldots,x_r)} \left[\mathbb{E}_{x\sim p}\left[f(x)\right]\right] - \inf_{f\in\mathcal{F}} \mathbb{E}_{x\sim p}\left[f(x)\right]\right]\right\}$$

The above implies that if $\mathcal{V}_T^{\text{blind}} = o(T)$ (i.e. the problem is learnable against an i.i.d adversary in the online sense without knowing the distribution p), then the problem is learnable in the classical batch sense. Specifically, there exists a strategy $\mathbf{s} = \{s_t\}_{t=1}^T$ with $s_t : \mathcal{X}^{t-1} \mapsto \mathcal{Q}$ such that

$$\sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\mathbb{E}_{r\sim \text{Unif}[1\ldots T]} \mathbb{E}_{f\sim s_{r+1}(x_1,\ldots,x_r)} \left[\mathbb{E}_{x\sim p}\left[f(x)\right]\right]\right] - \inf_{f\in\mathcal{F}} \mathbb{E}_{x\sim p}\left[f(x)\right]\right\} = o(1).$$

This strategy can be used to define a consistent (randomized) algorithm $\hat{f}_T : \mathcal{X}^T \mapsto \mathcal{F}$ as follows. Given an i.i.d. sample $x_1,\ldots,x_T$, draw a random index r from $1,\ldots,T$, and define $\hat{f}_T$ as a random draw from distribution $s_r(x_1,\ldots,x_{r-1})$. We have proven that $\mathcal{V}_T^{\text{batch}} \to 0$ as T increases, which the requirement of Eq. (27) in the general non-supervised case. Note that the rate of this convergence is upper bounded by the rate of decay of $\frac{1}{T}\mathcal{V}_T^{\text{blind}}$ to zero.

To show the reverse direction, say a problem is learnable in the classical batch sense. That is, $\mathcal{V}_T^{\text{batch}} \to 0$. Hence, there exists a randomized strategy $\mathbf{s} = (s_1, s_2, \ldots)$ such that $s_t : \mathcal{X}^{t-1} \mapsto \mathcal{Q}$ and

$$\sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_{t-1}\sim p}\left[\mathbb{E}_{f\sim s_t(x_1,\ldots,x_{t-1})} \mathbb{E}_{x\sim p}\left[f(x)\right]\right] - \inf_{f\in\mathcal{F}} \mathbb{E}_{x\sim p}\left[f(x)\right]\right\} = o(1)$$

as $t \to \infty$. Hence we have that

$$\begin{aligned}
&\sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\frac{1}{T}\sum_{t=1}^T \mathbb{E}_{f\sim s_t(x_1,\ldots,x_{t-1})} \mathbb{E}_{x\sim p}\left[f(x)\right] - \inf_{f\in\mathcal{F}} \mathbb{E}_{x\sim p}\left[f(x)\right]\right]\right\} \\
&\leq \frac{1}{T}\sum_{t=1}^T \sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\mathbb{E}_{f\sim s_t(x_1,\ldots,x_{t-1})} \mathbb{E}_{x\sim p}\left[f(x)\right] - \inf_{f\in\mathcal{F}} \mathbb{E}_{x\sim p}\left[f(x)\right]\right]\right\} = o(1)
\end{aligned}$$

because a Cesàro average of a convergent sequence also converges to the same limit.

As shown in [12], the problem is learnable in the batch sense if and only if

$$\mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\inf_{f\in\mathcal{F}} \frac{1}{T}\sum_{t=1}^T f(x_t)\right] \to \inf_{f\in\mathcal{F}} \mathbb{E}_{x\sim p}\left[f(x)\right]$$

and this rate is uniform for all distributions. Hence we have that

$$\sup_p \left\{ \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\frac{1}{T}\sum_{t=1}^T \mathbb{E}_{f\sim s_t(x_1,\ldots,x_{t-1})} \mathbb{E}_{x\sim p}\left[f(x)\right] - \inf_{f\in\mathcal{F}} \frac{1}{T}\sum_{t=1}^T f(x_t)\right]\right\} = o(1)$$

We conclude that if the problem is learnable in the i.i.d. batch sense then

$$\begin{aligned}
o(T) &= \sup_p \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\sum_{t=1}^T \mathbb{E}_{f\sim s_t(x_1,\ldots,x_{t-1})} \mathbb{E}_{x\sim p}\left[f(x)\right] - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right] \\
&= \sup_p \mathbb{E}_{x_1,\ldots,x_T\sim p}\left[\sum_{t=1}^T \mathbb{E}_{f_t\sim s_t(x_1,\ldots,x_{t-1})} f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right] \\
&= \sup_p \mathbb{E}_{x_1,\ldots,x_T\sim p} \mathbb{E}_{f_1\sim s_1}\ldots\mathbb{E}_{f_T\sim s_T(x_{1:T-1})}\left\{\sum_{t=1}^T f_t(x_t) - \inf_{f\in\mathcal{F}}\sum_{t=1}^T f(x_t)\right\} \\
&\geq \mathcal{V}_T^{\text{blind}}
\end{aligned} \tag{28}$$

Thus we have shown that if a problem is learnable in the batch sense then it is learnable versus all i.i.d. adversaries in the online sense, provided that the distribution is not known to the player.

□

Proof of Lemma 15. The first statement follows from the well-known classical symmetrization argument:

$$\begin{aligned}
\mathcal{V}_T^{\text{batch, sup}} &= \inf_{\hat f} \sup_{p_{X\times Y}} \left\{ \mathbb{E}|y-\hat f(x)| - \inf_{f\in\mathcal{F}} \mathbb{E}|y-f(x)| \right\} \\
&\le \sup_{p_{X\times Y}} \left\{ \mathbb{E}|y-\tilde f(x)| - \inf_{f\in\mathcal{F}} \mathbb{E}|y-f(x)| \right\} \\
&\le 2 \sup_{p_{X\times Y}} \mathbb{E} \sup_{f\in\mathcal{F}} \left| \frac{1}{T}\sum_{t=1}^{T} |y_t - f(x_t)| - \mathbb{E}|y-f(x)| \right| \\
&\le 4 \sup_{p_X} \mathbb{E}_{x_{1:T}} \mathbb{E}_{\epsilon_{1:T}} \sup_{f\in\mathcal{F}} \frac{1}{T}\sum_{t=1}^{T} \epsilon_t f(x_t)
\end{aligned}$$

where the first inequality is obtained by choosing the empirical minimizer $\tilde f$ as an estimator.

The second inequality of the Lemma follows from the lower bound proved in Section D. Lemma 20 implies that the game with i.i.d. restrictions $\mathcal{P}_t = \{p_X \times U_Y\}$ for all t satisfies

$$\mathcal{V}_T^{\text{sup}}(\{p_X \times U_Y\}) \ge \mathfrak{R}_T(\mathcal{F}, p_X)$$

for any p_X.

Now, clearly, the distribution-blind supervised game is harder than the game with the knowledge of the distribution. That is,

$$\sup_{p_{X\times Y}} \mathcal{V}_T^{\text{sup}}(\{p_{X\times Y}\}) \le \mathcal{V}_T^{\text{blind, sup}}$$

□

C Application: Hybrid Learning

In Section B, we studied the relationship between batch and online learnability in the i.i.d. setting, focusing on the supervised case in Section B.1. We now provide a more in-depth study of the value of the supervised game beyond the i.i.d. setting.

As shown in [10], the value of the supervised game with the *worst-case adversary* is upper and lower bounded (to within $O(\log^{3/2} T)$) by *sequential* Rademacher complexity. This complexity can be linear in T if the function class has infinite Littlestone's dimension, rendering worst-case learning futile. This is the case with a class of threshold functions on an interval, which has a Vapnik-Chervonenkis dimension of 1. Surprisingly, it was shown in [6] that for the classification problem with i.i.d. x's and adversarial labels y, online regret can be bounded whenever VC dimension of the class is finite. This suggests that it is the manner in which x is chosen that plays the decisive role in supervised learning. We indeed show that this is the case. Irrespective of the way the labels are chosen, if x_t are chosen i.i.d. then regret is (to within a constant) given by the classical Rademacher complexity. If x_t's are chosen adversarially, it is (to within a logarithmic factor) given by the sequential Rademacher complexity.

We remark that the algorithm of [6] is "distribution-blind" in the sense of last section. The results we present below are for non-blind games. While the equivalence of blind and non-blind learning was shown in the previous section for the i.i.d. supervised case, we hypothesize that it holds for the hybrid supervised learning scenario as well.

Let the loss class be $\phi(\mathcal{F}) = \{(x,y) \mapsto \phi(f(x),y) \ : \ f \in \mathcal{F}\}$ for some Lipschitz function $\phi : \mathbb{R}\times\mathcal{Y} \mapsto \mathbb{R}$ (i.e. $\phi(f(x),y) = |f(x)-y|$). Let $\mathcal{P}_{1:T}$ be the restrictions on the adversary. Theorem 3 then states that

$$\mathcal{V}_T^{\text{sup}}(\mathcal{P}_{1:T}) \le 2 \sup_{\mathbf{p}\in\mathfrak{P}} \mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p})$$

where the supremum is over all joint distributions $\mathbf{p}$ on the sequences $((x_1,y_1),\dots,(x_T,y_T))$, such that $\mathbf{p}$ satisfies the restrictions $\mathcal{P}_{1:T}$. The idea is to pass from a complexity of $\phi(\mathcal{F})$ to that of the class $\mathcal{F}$ via a Lipschitz composition lemma, and then note that the resulting complexity does not depend on y-variables. If this can be done, the complexity associated only with the choice of x is then an upper bound on the value of the game. The results of this section, therefore, hold whenever a Lipschitz composition lemma can be proved for the distribution-dependent Rademacher complexity.

The following lemma gives an upper bound on the distribution-dependent Rademacher complexity in the "hybrid" scenario, i.e. the distribution of x_t's is i.i.d. from a fixed distribution p but the distribution of y_t's is arbitrary (recall that adversarial choice of the player translates into vacuous restrictions $\mathcal{P}_t$ on the mixed strategies). Interestingly, the upper bound is a blend of the classical Rademacher complexity (on the x-variable) and the worst-case sequential Rademacher complexity for the y-variable. This captures the hybrid nature of the problem.

Lemma 17. *Fix a class $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ and a function $\phi : \mathbb{R} \times \mathcal{Y} \mapsto \mathbb{R}$. Given a distribution p over $\mathcal{X}$, let $\mathfrak{P}$ consist of all joint distributions $\mathbf{p}$ such that the conditional distribution $p_t^{x,y}(x_t, y_t|x^{t-1}, y^{t-1}) = p(x_t) \times p_t(y_t|x^{t-1}, y^{t-1}, x_t)$ for some conditional distribution p_t. Then,*

$$\sup_{\mathbf{p}\in\mathfrak{P}} \mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p}) \leq \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{\mathbf{y}} \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), \mathbf{y}_t(\epsilon)) \right] .$$

Armed with this result, we can appeal to the following Lipschitz composition lemma. It says that the distribution-dependent sequential Rademacher complexity for the hybrid scenario with a Lipschitz loss can be upper bounded via the classical Rademacher complexity of the function class on the x-variable only. That is, we can "erase" the Lipschitz loss function together with the (adversarially chosen) y variable. The lemma is an analogue of the classical contraction principle initially proved by Ledoux and Talagrand [7] for the i.i.d. process.

Lemma 18. *Fix a class $\mathcal{F} \subseteq [-1, 1]^{\mathcal{X}}$ and a function $\phi : [-1, 1] \times \mathcal{Y} \mapsto \mathbb{R}$. Assume, for all $y \in \mathcal{Y}$, $\phi(\cdot, y)$ is a Lipschitz function with a constant L. Let $\mathfrak{P}$ be as in Lemma 17. Then, for any $\mathbf{p} \in \mathfrak{P}$,*

$$\mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p}) \leq L\, \mathfrak{R}_T(\mathcal{F}, p) .$$

Lemma 17 in tandem with Lemma 18 imply that the value of the game with i.i.d. x's and adversarial y's is upper bounded by the classical Rademacher complexity.

For the case of adversarially-chosen x's and (potentially) adversarially chosen y's, the necessary Lipschitz composition lemma is proved in [10] with an extra factor of $O(\log^{3/2} T)$. We summarize the results in the following Corollary.

Corollary 19. *For stochastic-adversarial supervised learning with absolute loss,*

(1) If x_t are chosen adversarially, then irrespective of the way y_t's are chosen,

$$\mathcal{V}_T^{sup} \leq 2\mathfrak{R}(\mathcal{F}) \times O(\log^{3/2}(T)),$$

where $\mathfrak{R}(\mathcal{F})$ is the (worst-case) sequential Rademacher complexity [10]. A matching lower bound of $\mathfrak{R}(\mathcal{F})$ is attained by choosing y_t's as i.i.d. Rademacher random variables.

(2) If x_t are chosen i.i.d. from p, then irrespective of the way y_t's are chosen,

$$\mathcal{V}_T^{sup} \leq 2\mathfrak{R}(\mathcal{F}, p),$$

where $\mathfrak{R}(\mathcal{F}, p)$ defined in (6) is the classical Rademacher complexity. The matching lower bound of $\mathfrak{R}(\mathcal{F}, p)$ is obtained by choosing y_t's as i.i.d. Rademacher random variables.

The lower bounds stated in Corollary 19 are proved in the Appendix.

C.1 Proofs

Proof of Lemma 17. We want to bound the supremum (as $\mathbf{p}$ ranges over $\mathfrak{P}$) of the distribution-dependent Rademacher complexity:

$$\sup_{\mathbf{p}\in\mathfrak{P}} \mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p}) = \sup_{\mathbf{p}\in\mathfrak{P}} \mathop{\mathbb{E}}_{((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}')))\sim\boldsymbol{\rho}} \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(\mathbf{x}_t(\epsilon)), \mathbf{y}_t(\epsilon)) \right]$$

for an associated process $\boldsymbol{\rho}$ defined in Section 3. To elucidate the random process $\boldsymbol{\rho}$, we expand the succinct tree notation and write the above quantity as

$$\sup_{\mathbf{p}} \mathbb{E}_{x_1,x_1'\sim p} \mathbb{E}_{\substack{y_1\sim p_1(\cdot|x_1)\\ y_1'\sim p_1(\cdot|x_1')}} \mathbb{E}_{\epsilon_1} \mathbb{E}_{x_2,x_2'\sim p} \mathbb{E}_{\substack{y_2\sim p_2(\cdot|\chi_1(\epsilon_1),x_2)\\ y_2'\sim p_2(\cdot|\chi_1(\epsilon_1),x_2')}} \mathbb{E}_{\epsilon_2} \quad \ldots$$
$$\ldots \quad \mathbb{E}_{x_T,x_T'\sim p} \mathbb{E}_{\substack{y_T\sim p_T(\cdot|\chi_1(\epsilon_1),\ldots,\chi_{T-1}(\epsilon_{T-1}),x_T)\\ y_T'\sim p_T(\cdot|\chi_1(\epsilon_1),\ldots,\chi_{T-1}(\epsilon_{T-1}),x_T')}} \mathbb{E}_{\epsilon_T} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), y_t) \right]$$

where $\chi_t(\epsilon_t)$ now selects the pair (x_t, y_t) or (x_t', y_t'). By passing to the supremum over y_t, y_t' for all t, we arrive at

$$\sup_{\mathbf{p}\in\mathfrak{P}} \mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p}) \le \sup_{\mathbf{p}} \mathbb{E}_{x_1,x_1'\sim p} \sup_{y_1,y_1'} \mathbb{E}_{\epsilon_1} \mathbb{E}_{x_2,x_2'\sim p} \sup_{y_2,y_2'} \mathbb{E}_{\epsilon_2} \ldots \mathbb{E}_{x_T,x_T'\sim p} \sup_{y_T,y_T'} \mathbb{E}_{\epsilon_T} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), y_t) \right]$$
$$= \mathbb{E}_{x_1\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \mathbb{E}_{x_2\sim p} \sup_{y_2} \mathbb{E}_{\epsilon_2} \ldots \mathbb{E}_{x_T\sim p} \sup_{y_T} \mathbb{E}_{\epsilon_T} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), y_t) \right]$$

where the sequence of x_t''s and y_t''s has been eliminated. By moving the expectations over x_t's outside the suprema (and thus increasing the value), we upper bound the above by:

$$\le \mathbb{E}_{x_1,\ldots,x_T\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \sup_{y_2} \mathbb{E}_{\epsilon_2} \ldots \sup_{y_T} \mathbb{E}_{\epsilon_T} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), y_t) \right]$$
$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{\mathbf{y}} \mathbb{E}_{\epsilon} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), \mathbf{y}_t(\epsilon)) \right]$$

□

Proof of Lemma 18. First without loss of generality assume $L = 1$. The general case follow from this by simply scaling ϕ appropriately. By Lemma 17,

$$\mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p}) \le \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{\mathbf{y}} \mathbb{E}_{\epsilon} \left[\sup_{f\in\mathcal{F}} \sum_{t=1}^T \epsilon_t \phi(f(x_t), \mathbf{y}_t(\epsilon)) \right] \tag{29}$$

The proof proceeds by sequentially using the Lipschitz property of $\phi(f(x_t), \mathbf{y}_t(\epsilon))$ for decreasing t, starting from $t = T$. Towards this end, define

$$R_t = \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{\mathbf{y}} \mathbb{E}_{\epsilon} \left[\sup_{f\in\mathcal{F}} \sum_{s=1}^t \epsilon_s \phi(f(x_s), \mathbf{y}_s(\epsilon)) + \sum_{s=t+1}^T \epsilon_s f(x_s) \right] .$$

Since the mappings $\mathbf{y}_{t+1}, \ldots, \mathbf{y}_T$ do not enter the expression, the supremum is in fact taken over the trees $\mathbf{y}$ of depth t. Note that $R_0 = \mathfrak{R}(\mathcal{F}, p)$ is precisely the classical Rademacher complexity (without the dependence on $\mathbf{y}$), while R_T is the upper bound on $\mathfrak{R}_T(\phi(\mathcal{F}), \mathbf{p})$ in Eq. (29). We need to show $R_T \le R_0$ and we will show this by proving $R_t \le R_{t-1}$ for all $t \in [T]$. So, let us fix $t \in [T]$ and start with R_t:

$$R_t = \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{\mathbf{y}} \mathbb{E}_{\epsilon} \left[\sup_{f\in\mathcal{F}} \sum_{s=1}^t \epsilon_s \phi(f(x_s), \mathbf{y}_s(\epsilon)) + \sum_{s=t+1}^T \epsilon_s f(x_s) \right]$$
$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \ldots \sup_{y_t} \mathbb{E}_{\epsilon_t} \mathbb{E}_{\epsilon_{t+1:T}} \left[\sup_{f\in\mathcal{F}} \sum_{s=1}^t \epsilon_s \phi(f(x_s), y_s) + \sum_{s=t+1}^T \epsilon_s f(x_s) \right]$$
$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \ldots \sup_{y_t} \mathbb{E}_{\epsilon_{t+1:T}} \quad S(x_{1:T}, y_{1:t}, \epsilon_{1:t-1}, \epsilon_{t+1:T})$$

with

$$S(x_{1:T}, y_{1:t}, \epsilon_{1:t-1}, \epsilon_{t+1:T}) = \mathbb{E}_{\epsilon_t}\left[\sup_{f\in\mathcal{F}} \sum_{s=1}^{t} \epsilon_s \phi(f(x_s), y_s) + \sum_{s=t+1}^{T} \epsilon_s f(x_s)\right]$$
$$= \frac{1}{2}\left\{\sup_{f\in\mathcal{F}} \sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) + \phi(f(x_t), y_t) + \sum_{s=t+1}^{T} \epsilon_s f(x_s)\right\}$$
$$+ \frac{1}{2}\left\{\sup_{f\in\mathcal{F}} \sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) - \phi(f(x_t), y_t) + \sum_{s=t+1}^{T} \epsilon_s f(x_s)\right\}$$

The two suprema can be combined to yield

$$2S(x_{1:T}, y_{1:t}, \epsilon_{1:t-1}, \epsilon_{t+1:T})$$
$$= \sup_{f,g\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s(\phi(f(x_s), y_s) + \phi(g(x_s), y_s)) + \phi(f(x_t), y_t) - \phi(g(x_t), y_t) + \sum_{s=t+1}^{T} \epsilon_s(f(x_s) + g(x_s))\right\}$$
$$\le \sup_{f,g\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s(\phi(f(x_s), y_s) + \phi(g(x_s), y_s)) + |f(x_t) - g(x_t)| + \sum_{s=t+1}^{T} \epsilon_s(f(x_s) + g(x_s))\right\} \quad (*)$$
$$= \sup_{f,g\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s(\phi(f(x_s), y_s) + \phi(g(x_s), y_s)) + f(x_t) - g(x_t) + \sum_{s=t+1}^{T} \epsilon_s(f(x_s) + g(x_s))\right\} \quad (**)$$

The first inequality is due to the Lipschitz property, while the last equality needs a justification. First, it is clear that the term $(**)$ is upper bounded by $(*)$. The reverse direction can be argued as follows. Let a pair (f^*, g^*) achieve the supremum in $(*)$. Suppose first that $f^*(x_t) \ge g^*(x_t)$. Then (f^*, g^*) provides the same value in $(**)$ and, hence, the supremum is no less than the supremum in $(*)$. If, on the other hand, $f^*(x_t) < g^*(x_t)$, then the pair (g^*, f^*) provides the same value in $(**)$.

We conclude that

$$S(x_{1:T}, y_{1:t}, \epsilon_{1:t-1}, \epsilon_{t+1:T})$$
$$\le \frac{1}{2}\sup_{f,g\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s(\phi(f(x_s), y_s) + \phi(g(x_s), y_s)) + f(x_t) - g(x_t) + \sum_{s=t+1}^{T} \epsilon_s(f(x_s) + g(x_s))\right\}$$
$$= \frac{1}{2}\left\{\sup_{f\in\mathcal{F}} \sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) + f(x_t) + \sum_{s=t+1}^{T} \epsilon_s f(x_s)\right\} + \frac{1}{2}\left\{\sup_{f\in\mathcal{F}} \sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) - f(x_t) + \sum_{s=t+1}^{T} \epsilon_s f(x_s)\right\}$$
$$= \mathbb{E}_{\epsilon_t} \sup_{f\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) + \epsilon_t f(x_t) + \sum_{s=t+1}^{T} \epsilon_s f(x_s)\right\}$$

Thus,

$$R_t = \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \ldots \sup_{y_t} \mathbb{E}_{\epsilon_{t+1:T}} \; S(x_{1:T}, y_{1:t}, \epsilon_{1:t-1}, \epsilon_{t+1:T})$$
$$\le \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \ldots \sup_{y_t} \mathbb{E}_{\epsilon_{t:T}} \sup_{f\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) + \sum_{s=t}^{T} \epsilon_s f(x_s)\right\}$$
$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_T\sim p} \sup_{y_1} \mathbb{E}_{\epsilon_1} \ldots \sup_{y_{t-1}} \mathbb{E}_{\epsilon_{t-1}} \mathbb{E}_{\epsilon_{t:T}} \sup_{f\in\mathcal{F}} \left\{\sum_{s=1}^{t-1} \epsilon_s \phi(f(x_s), y_s) + \sum_{s=t}^{T} \epsilon_s f(x_s)\right\}$$
$$= R_{t-1}$$

where we have removed the supremum over y_t as it no longer appears in the objective. This concludes the proof.

□

D Lower Bounds

We now give two lower bounds on the value $\mathcal{V}_T^{\text{sup}}$, defined with the absolute value loss function $\phi(f(x), y) = |f(x) - y|$. The lower bounds hold whenever the adversary's restrictions $\{\mathcal{P}_t\}_{t=1}^T$ allow the labels to be i.i.d. coin flips. That is, for the purposes of proving the lower bound, it is enough to choose a joint probability $\mathbf{p}$ (an oblivious strategy for the adversary) such that each conditional probability distribution on the pair (x, y) is of the form $p_t(x|x_1, \ldots, x_{t-1}) \times b(y)$ with $b(-1) = b(1) = 1/2$. Pick any such $\mathbf{p}$.

Our first lower bound will hold whenever the restrictions $\mathcal{P}_t$ are history-independent. That is, $\mathcal{P}_t(x_{1:t-1}) = \mathcal{P}_t(x'_{1:t-1})$ for any $x_{1:t-1}, x'_{1:t-1} \in \mathcal{X}^{t-1}$. Since the worst-case (all distributions) and i.i.d. (single distribution) are both history-independent restrictions, the lemma can be used to provide lower bounds for these cases. The second lower bound holds more generally, yet it is weaker than that of Lemma 20.

Lemma 20. *Let $\mathfrak{P}$ be the set of all $\mathbf{p}$ satisfying the history-independent restrictions $\{\mathcal{P}_t\}$ and $\mathfrak{P}' \subseteq \mathfrak{P}$ the subset that allows the label y_t to be an i.i.d. Rademacher random variable for each t. Then*

$$\mathcal{V}_T^{sup}(\mathcal{P}_{1:T}) \geq \sup_{\mathbf{p} \in \mathfrak{P}'} \mathfrak{R}_T(\mathcal{F}, \mathbf{p})$$

In particular, Lemma 20 gives matching lower bounds for Corollary 19.

Lemma 21. *Let $\mathfrak{P}$ be the set of all $\mathbf{p}$ satisfying the restrictions $\{\mathcal{P}_t\}$ and let $\mathfrak{P}' \subseteq \mathfrak{P}$ be the subset that allows the label y_t to be an i.i.d. Rademacher random variable for each t. Then*

$$\mathcal{V}_T^{sup}(\mathcal{P}_{1:T}) \geq \sup_{\mathbf{p} \in \mathfrak{P}'} \mathbb{E}_{(\mathbf{x}, \mathbf{x}') \sim \boldsymbol{\rho}} \mathbb{E}_{\epsilon} \left[\sup_{f \in \mathcal{F}} \sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(-\mathbf{1})) \right]$$

Proof of Lemma 20. Notice that $\mathbf{p}$ defines the stochastic process $\boldsymbol{\rho}$ as in (4) where the i.i.d. y_t's now play the role of the ϵ_t's. More precisely, at each time t, two copies x_t and x'_t are drawn from the marginal distribution $p_t(\cdot|\chi_1(y_1), \ldots, \chi_{t-1}(y_{t-1}))$, then a Rademacher random variable y_t is drawn i.i.d. and it indicates whether x_t or x'_t is to be used in the subsequent conditional distributions via the selector $\chi_t(y_t)$. This is a well-defined process obtained from $\mathbf{p}$ that produces a sequence of $(x_1, x'_1, y_1), \ldots, (x_T, x'_T, y_T)$. The x' sequence is only used to define conditional distributions below, while the sequence $(x_1, y_1), \ldots, (x_T, y_T)$ is presented to the player. Since restrictions are history-independent, the stochastic process is following the protocol which defines $\boldsymbol{\rho}$.

For any $\mathbf{p}$ of the form described above, the value of the game in (2) can be lower-bounded via Proposition 2.

$$\begin{aligned} \mathcal{V}_T^{\text{sup}} &\geq \mathbb{E}\left[\sum_{t=1}^T \inf_{f_t \in \mathcal{F}} \mathbb{E}_{(x_t, y_t)}\left[|y_t - f_t(x_t)| \,\Big|\, (x, y)_{1:t-1}\right] - \inf_{f \in \mathcal{F}} \sum_{t=1}^T |y_t - f(x_t)|\right] \\ &= \mathbb{E}\left[\sum_{t=1}^T 1 - \inf_{f \in \mathcal{F}} \sum_{t=1}^T |y_t - f(x_t)|\right] \end{aligned}$$

A short calculation shows that the last quantity is equal to

$$\mathbb{E} \sup_{f \in \mathcal{F}} \sum_{t=1}^T (1 - |y_t - f(x_t)|) = \mathbb{E} \sup_{f \in \mathcal{F}} \sum_{t=1}^T y_t f(x_t).$$

The last expectation can be expanded to show the stochastic process:

$$\begin{aligned} &\mathbb{E}_{x_1, x'_1 \sim p_1} \mathbb{E}_{y_1} \mathbb{E}_{x_2, x'_2 \sim p_2(\cdot|\chi_1(y_1))} \mathbb{E}_{y_2} \ldots \mathbb{E}_{x_T, x'_T \sim p_T(\cdot|\chi_1(y_1), \ldots, \chi_{T-1}(y_{T-1}))} \mathbb{E}_{y_T} \sup_{f \in \mathcal{F}} \sum_{t=1}^T y_t f(x_t) \\ &= \mathbb{E}_{(\mathbf{x}, \mathbf{x}') \sim \boldsymbol{\rho}} \mathbb{E}_{\epsilon} \left[\sup_{f \in \mathcal{F}} \sum_{t=1}^T \epsilon_t f(\mathbf{x}_t(\epsilon))\right] \\ &= \mathfrak{R}_T(\mathcal{F}, \mathbf{p}) \end{aligned}$$

Since this lower bound holds for any $\mathbf{p}$ which allows the labels to be independent ± 1 with probability $1/2$, we conclude the proof. $\square$

Proof of Lemma 21. For the purposes of this proof, the adversary presents y_t an i.i.d. Rademacher random variable on each round. Unlike the previous lemma, only the $\{x_t\}$ sequence is used for defining conditional distributions. Hence, the $\mathbf{x}'$ tree is immaterial and the lower bound is only concerned with the left-most path. The rest of the proof is similar to that of Lemma 20:

$$\begin{aligned}\mathcal{V}_T^{\text{sup}} &\geq \mathbb{E}\left[\sum_{t=1}^{T} \inf_{f_t \in \mathcal{F}} \mathbb{E}_{(x_t,y_t)}\left[|y_t - f_t(x_t)| \,\Big|\, (x,y)_{1:t-1}\right] - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} |y_t - f(x_t)|\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T} 1 - \inf_{f\in\mathcal{F}} \sum_{t=1}^{T} |y_t - f(x_t)|\right]\end{aligned}$$

As before, this expression is equal to

$$\begin{aligned}&\mathbb{E} \sup_{f\in\mathcal{F}} \sum_{t=1}^{T} y_t f(x_t) = \mathbb{E}_{x_1\sim p_1}\mathbb{E}_{y_1}\mathbb{E}_{x_2\sim p_2(\cdot|x_1)}\mathbb{E}_{y_2}\ldots\mathbb{E}_{x_T\sim p_T(\cdot|x_1,\ldots,x_{T-1})}\mathbb{E}_{y_T} \sup_{f\in\mathcal{F}} \sum_{t=1}^{T} y_t f(x_t) \\ &= \mathbb{E}_{(\mathbf{x},\mathbf{x}')\sim\boldsymbol{\rho}}\mathbb{E}_{\epsilon}\left[\sup_{f\in\mathcal{F}} \sum_{t=1}^{T} \epsilon_t f(\mathbf{x}_t(-\mathbf{1}))\right]\end{aligned}$$

$\square$