Online-to-PAC Conversions: Generalization Bounds via Regret Analysis

Gábor Lugosi GABOR.LUGOSI@GMAIL.COM
ICREA, Universitat Pompeu Fabra, and Barcelona School of Economics, Barcelona, Spain

Gergely Neu GERGELY.NEU@GMAIL.COM
Universitat Pompeu Fabra, Barcelona, Spain

Abstract

We present a new framework for deriving bounds on the generalization bound of statistical learning algorithms from the perspective of online learning. Specifically, we construct an online learning game called the "generalization game", where an online learner is trying to compete with a fixed statistical learning algorithm in predicting the sequence of generalization gaps on a training set of i.i.d. data points. We establish a connection between the online and statistical learning setting by showing that the existence of an online learning algorithm with bounded regret in this game implies a bound on the generalization error of the statistical learning algorithm, up to a martingale concentration term that is independent of the complexity of the statistical learning method. This technique allows us to recover several standard generalization bounds including a range of PAC-Bayesian and information-theoretic guarantees, as well as generalizations thereof.

Keywords: statistical learning, generalization error, online learning, regret analysis

1. Introduction

We study the standard model of statistical learning. We are given a training sample of n i.i.d. data points $S_n = (Z_1, \ldots, Z_n)$ drawn from a distribution μ over a measurable *instance space* $\mathcal{Z}$. A *learning algorithm* $\mathcal{A} : \mathcal{Z}^n \to \mathcal{W}$ maps the training sample to an output $W_n = \mathcal{A}(S_n)$ taking values in a measurable set $\mathcal{W}$ (called the *hypothesis class*) in a potentially randomized way. More precisely, a randomized learning algorithm assigns, to any n-tuple of samples from $\mathcal{Z}$, a probability distribution over $\mathcal{W}$ and draws a sample from that distribution, conditionally independently of S_n. The resulting random element is denoted by W_n.

We study the performance of the learning algorithm measured by a *loss function* $\ell : \mathcal{W} \times \mathcal{Z} \to \mathbb{R}_+$. Two key objects of interest are the *training error* $L(w, S_n) = \frac{1}{n}\sum_{i=1}^n \ell(w, Z_i)$ and the *test error* $\mathbb{E}\left[\ell(w, Z')\right]$ of a hypothesis $w \in \mathcal{W}$, where the random element Z' has the same distribution as the Z_i, and is independent of S_n. The *generalization error* of the algorithm is defined as

$$\mathrm{gen}(W_n, S_n) = \mathbb{E}\left[\ell(W_n, Z') \middle| W_n\right] - L(W_n, S_n)\,.$$

The quantity $\mathrm{gen}(W_n, S_n)$ expresses the difference between the expected loss (i.e., risk) of a learning algorithm and its empirical counterpart, computed over the same sample S_n that was used to train the algorithm. In other words, the generalization error measures the extent of *overfitting* occurring during training. As such, understanding (and more specifically, upper-bounding) the generalization error has been in the center of focus of statistical learning theory ever since its inception. Over the past half century, numerous approaches have been proposed to tackle this challenge. Key ideas include uniform convergence arguments (Vapnik and Chervonenkis, 1974), distribution-dependent complexity measures like the Rademacher or Gaussian complexities (Bartlett et al., 2002; Koltchinskii, 2001; Bartlett and Mendelson, 2002), or various notions of stability that can guarantee small generalization error (Devroye and Wagner, 1979; Bousquet and Elisseeff, 2002; Mukherjee et al., 2006; Shalev-Shwartz et al., 2010). The most relevant to our work is the family of so-called PAC-Bayesian generalization bounds, a topic to which we return shortly (Shawe-Taylor and Williamson, 1997; McAllester, 1998; Audibert, 2004; Catoni, 2007).

In this work, we establish a connection between the statistical learning model described above and the model of online learning (Cesa-Bianchi and Lugosi, 2006; Orabona, 2019). Online learning models sequential games between an online learner and its environment, where in each round $t = 1, 2, \ldots, n$, the following steps are repeated: (1) the online learner picks a distribution over hypotheses $P_t \in \Delta_{\mathcal{W}}$; (2) the environment picks a cost function $c_t : \mathcal{W} \to \mathbb{R}$; (3) the online learner incurs cost $\mathbb{E}_{\widetilde{W}_t \sim P_t}\left[c_t\big(\widetilde{W}_t\big)\right]$; (4) the online learner observes the cost function c_t. Importantly, the two players make their choices in parallel, and their actions are revealed to each other only at the end of the round. Typically, no assumptions about are made about the environment, and, in particular, it is allowed to have full knowledge of the online learner's decision-making policy. The performance of an online learning algorithm Π_n is then measured in terms of its *regret* against a comparator point $P^* \in \Delta_{\mathcal{W}}$, defined as

$$\mathrm{regret}_{\Pi_n}(P^*) = \sum_{t=1}^{n} \left(\mathbb{E}_{\widetilde{W}_t \sim P_t}\left[c_t\left(\widetilde{W}_t\right)\right] - \mathbb{E}_{W^* \sim P^*}\left[c_t\left(W^*\right)\right] \right).$$

Notably, the comparator P^* is allowed to depend on the entire sequence of costs chosen by the environment, and is typically picked as the distribution minimizing the cumulative costs. In the last few decades, numerous algorithms with strong regret bounds have been proposed for the above setting and a variety of its generalizations.

Our main contribution is a new framework for bounding the generalization error of any statistical learning algorithm via a reduction to online learning. Specifically, we construct an online learning game (that we call the *generalization game*) where the regret of the learner against a well-chosen comparator point can be shown to be equal to the generalization error of the statistical learning algorithm under investigation, up to a perturbation term whose magnitude is independent of the complexity of the hypothesis class or the statistical learning method. Using this construction, we show that the *existence* of an online learning algorithm with small regret against the comparator in question implies a bound on the generalization error of the statistical learning algorithm. To emphasize the connection between PAC learning, online learning, and online-to-batch conversions, we refer to our framework as *online-to-PAC conversion*.

Various links have been previously established between online learning and a variety of concentration inequalities. It is well known that the relatively simple setting of testing and mean estimation of scalar-valued random variables is intimately connected with sequential betting—see, e.g., Shafer and Vovk (2001), Waudby-Smith and Ramdas (2020), and especially Orabona and Jun (2021) for an impressive literature review on the subject. Our work draws direct inspiration from Kakade, Sridharan, and Tewari (2008), who used online learning techniques for bounding the Rademacher complexity of linear function classes. An even earlier example of a similar flavor is the work of Zhang (2002), who proved bounds on the covering numbers of linear function classes via a reduction to online binary classification and in particular the classical Perceptron mistake bound (cf. his Theorem 4). In the broader context of concentration of empirical processes, Rakhlin and Sridharan (2017) showed an *equivalence* between tail inequalities on the supremum of a collection of martingales and existence of an online learning algorithm with bounded regret in a game tailor-made to the collection of martingales in question. While many of these results can be adapted to provide bounds on the generalization error of statistical learning methods, they all suffer from being overly conservative by considering a supremum over a collection of random variables, which only serves as a possibly loose proxy to the quantity of interest. Our key observation is that the general approach taken in the above works results can be significantly strengthened when specialized to statistical learning, and online learning techniques can be employed to bound the generalization error in an algorithm-dependent fashion rather than the unnecessarily pessimistic worst-case fashion.

Our online-to-PAC conversion scheme allows us to recover a range of previously known generalization bounds, as well as to establish some new ones. As our most elementary example, we show that the classical PAC-Bayesian generalization bound of McAllester (1998) can be directly recovered from our framework by employing the standard exponentially weighted average algorithm of Littlestone and Warmuth (1994) as the online learning method in the generalization game. To illustrate the power of our reduction, we derive a variety of extensions to this fundamental theorem, including data-dependent bounds that approach zero at a fast rate when the empirical risk is zero, and a parameter-free bound that shaves off a logarithmic factor that appears in all other PAC-Bayesian bounds that we are aware of. We also provide a much more general family of generalization bounds that replaces the relative entropy appearing in the classical PAC-Bayesian bounds with an appropriately chosen strongly convex function of the conditional distribution of the output W_n given the input S_n, via an application of the standard Follow-the-Regularized-Leader (FTRL) algorithm for online learning (see, e.g., Orabona, 2019 and the references therein). Furthermore, we provide an empirical version of the latter bounds by adapting the idea of "optimistic" updates as proposed by Rakhlin and Sridharan (2013a,b). As an example application of these techniques, we provide a new generalization bound that replaces the relative entropy factor and the subgaussianity constant appearing in the classical PAC-Bayes bounds with the squared Wasserstein-2 distance, and a Sobolev-type norm of the loss function. Finally, we provide an extension to our framework that allows the use of data-dependent priors and regularizers in the vein of the "almost exchangeable priors" of Audibert (2004) and Catoni (2007), which also allows us to recover several classical PAC-learning bounds.

The present paper is a significant expansion of our earlier work published at COLT 2022 (Lugosi and Neu, 2022). In this previous work, we have proved only a small subset of the results in the present paper, using a considerably more complicated analysis that only yielded bounds that hold on expectation. Indeed, the analysis in Lugosi and Neu (2022) relied on direct convex-analytic calculations inspired by the analysis of a specific online learning algorithm (FTRL), which is now only one of the many applications captured by our general framework (cf. Section 3.2). We have managed to significantly generalize these earlier results after learning about the work of Kakade, Sridharan, and Tewari (2008) that several colleagues have brought to our attention after hearing about our work at COLT 2022. We are grateful to the COLT community for helping us make this connection.

The rest of the paper is organized as follows. In Section 2, we provide further details about the setup we consider and introduce our online-to-PAC conversion framework. In Section 3, we provide a long list of applications of the basic framework, including a variety of PAC-Bayesian generalization bounds in Section 3.1, and generalized versions thereof in Section 3.2. We provide further extensions of the framework in Section 4 and conclude in Section 5.

Notation. For a distribution over hypotheses $P \in \Delta_{\mathcal{W}}$ and a bounded measurable function $f : \mathcal{W} \to \mathbb{R}$, we write $\langle P, f \rangle$ to refer to the expectation of $\mathbb{E}_{W \sim P}[f(W)]$. We use $\|\cdot\|$ to denote a norm on the Banach space $\mathcal{Q} = \{aP - bP' : P, P' \in \Delta_{\mathcal{W}}, a, b \in \mathbb{R}\}$, and $\|\cdot\|_*$ to denote the corresponding dual norm on the dual space $\mathcal{Q}^*$ of measurable functions on $\mathcal{W}$, defined for each $f \in \mathcal{Q}^*$ as $\|f\|_* = \sup_{Q \in \mathcal{Q}: \|Q\| \leq 1} \langle Q, f \rangle$.

2. Online-to-PAC Conversions

We start by describing the key construction underlying our online-to-PAC conversion framework: an online learning game that we call the *generalization game* that connects the statistical learning setting with the online setting. Precisely, we consider a sequential interaction scheme between an online learner and an adversary, where the following steps are repeated in a sequence of rounds $t = 1, 2, \ldots, n$:

1. the online learner picks a distribution $P_t \in \Delta_{\mathcal{W}}$;
2. the adversary selects a cost function $c_t : \mathcal{W} \to \mathbb{R}$ defined for each $w \in \mathcal{W}$ as
$$c_t(w) = \ell(w, Z_t) - \mathbb{E}_{Z' \sim \mu}[\ell(w, Z')] \ ;$$
3. the online learner incurs cost $\langle P_t, c_t \rangle = \mathbb{E}_{W \sim P_t}[c_t(W)]$;
4. the adversary reveals the value of Z_t to the online learner.

We emphasize that in this setup, the online learner is allowed to know the loss function ℓ and the distribution μ of the data points Z_t, and therefore by revealing the value of Z_t, the online learner may compute the entire cost function $c_t(\cdot)$. However, the online learner is *not* allowed any type of access to the realization of the future data points $Z_t, \ldots, Z_n$ before making its decision about P_t.

We use $\mathcal{F}_t$ to denote the sigma-algebra induced by the sequence of random variables generated and used by both players (including the random data points $Z_1, \dots, Z_t$ and all potential randomization utilized by the online learner) up until the end of round t. Formally, an *online learning algorithm* $\Pi_n = \{P_t\}_{t=1}^n$ is a sequence of functions such that P_t maps the sequence of past outcomes $(z_1, \dots, z_{t-1}) \in \mathcal{Z}^{t-1}$ to $\Delta_{\mathcal{W}}$, the set of all probability distributions over the hypothesis class $\mathcal{W}$. For the brevity of notation, we abbreviate $P_t = P_t(Z_1, \dots, Z_{t-1})$. We denote by $\mathcal{P}_n$ the class of all online learning algorithms over sequences of length n.

A random variable of crucial importance, associated to the online learning algorithm Π_n, is

$$M_{\Pi_n} = \frac{1}{n} \sum_{t=1}^{n} \langle P_t, c_t \rangle \ .$$

Notice that for any online learning algorithm, M_{Π_n} is a normalized sum of martingale differences, due to the conditional independence of P_t and c_t:

$$\begin{aligned} \mathbb{E}\left[\langle P_t, c_t \rangle \middle| \mathcal{F}_{t-1} \right] &= \mathbb{E}_{\widetilde{W}_t \sim P_t}\left[c_t\big(\widetilde{W}_t\big) \middle| \mathcal{F}_{t-1} \right] \\ &= \mathbb{E}_{\widetilde{W}_t \sim P_t}\left[\ell\big(\widetilde{W}_t, Z_t\big) - \ell\big(\widetilde{W}_t, Z'\big) \middle| \mathcal{F}_{t-1} \right] = 0 \ . \end{aligned}$$

Indeed, the online learning protocol guarantees that $\widetilde{W}_t$ is chosen before Z_t is revealed to the online learner, and thus $(\widetilde{W}_t, Z_t)$ has the same conditional distribution as $(\widetilde{W}_t, Z')$ given the history $\mathcal{F}_{t-1}$.

As is usual in online learning, the goal of the online learner is to accumulate a total cost that is not much worse than an appropriately chosen comparator distribution $P^* \in \Delta_{\mathcal{W}}$. Specifically, the performance metric for evaluating the online learner's performance is the *regret* defined against the comparator P^* as

$$\mathrm{regret}_{\Pi_n}(P^*) = \sum_{t=1}^{n} \langle P_t - P^*, c_t \rangle \ .$$

Many online learning algorithms come with performance guarantees that hold with probability one for all cost sequences c_t, against comparators P^* that may depend on the cost sequence in an arbitrary way. We exploit this property below by choosing a very specific comparator point that will allow us to establish our main result.

In particular, we choose the comparator point P^* as the conditional distribution of the output W_n given the input S_n. We denote this distribution by $P_{W_n|S_n}$ and remark that when the statistical learning algorithm does not use randomization, then the distribution $P_{W_n|S_n}$ puts all its mass to a single point W_n. For randomized algorithms, we study the generalization error in expectation with respect to the additional randomization of $W_n|S_n$. To this end, we introduce

$$\overline{\mathrm{gen}}(W_n, S_n) = \mathbb{E}\left[\mathrm{gen}(W_n, S_n) | S_n \right].$$

Naturally, for algorithms that do not use randomization, we have $\overline{\mathrm{gen}}(W_n, S_n) = \mathrm{gen}(W_n, S_n)$.

The following theorem characterizes the generalization error in terms of the regret of an online learning algorithm.

Theorem 1 *The generalization error of any learning algorithm $W_n = \mathcal{A}(S_n)$ satisfies that, for any online learning algorithm $\Pi_n \in \mathcal{P}_n$,*

$$\overline{\mathrm{gen}}(W_n, S_n) = \frac{\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})}{n} - M_{\Pi_n} \ .$$

In particular, for any class $\mathcal{C}_n \subset \mathcal{P}_n$ of online learning algorithms,

$$\overline{\mathrm{gen}}(W_n, S_n) = \inf_{\Pi_n \in \mathcal{C}_n} \left(\frac{\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})}{n} - M_{\Pi_n} \right) \ .$$

Proof Let $\Pi_n = \{P_t\}_{t=1}^n$ be an arbitrary online learning algorithm. Recalling the notation $\langle P, f \rangle = \mathbb{E}_{W \sim P}[f(W)]$, we rewrite the conditional expectation of the generalization error as follows:

$$\begin{aligned}\overline{\text{gen}}(W_n, S_n) &= \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\ell(W_n, Z_t) - \ell(W_n, Z')\right| S_n\right] = -\frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[c_t(W_n)\right| S_n\right] \\ &= -\frac{1}{n}\sum_{t=1}^n \left\langle P_{W_n|S_n}, c_t \right\rangle \\ &= \frac{1}{n}\sum_{t=1}^n \left\langle P_t - P_{W_n|S_n}, c_t \right\rangle - \frac{1}{n}\sum_{t=1}^n \left\langle P_t, c_t \right\rangle \\ &= \frac{\text{regret}_{\Pi_n}(P_{W_n|S_n})}{n} - \frac{1}{n}\sum_{t=1}^n \left\langle P_t, c_t \right\rangle \ .\end{aligned}$$

Recalling the definition $M_{\Pi_n} = \frac{1}{n}\sum_{t=1}^n \langle P_t, c_t \rangle$ concludes the proof. ■

While the claim and its proof are strikingly simple (and perhaps even trivial in hindsight), it has never appeared in previous literature with such clarity—at least to our knowledge. A result that comes close is Theorem 1 of Kakade et al. (2008), which bounds the Rademacher complexity of linear function classes using a similar technique, which in fact served as direct inspiration for our proof above. The key difference between their approach and ours is that we directly bound the generalization error via an algorithm-specific choice of comparator point, as opposed to the simple worst-case choice taken by Kakade, Sridharan, and Tewari (2008). We may use the simple observation of Theorem 1 in various ways to obtain upper bounds for the generalization error. The simplest way is to choose a single online learning algorithm and use martingale concentration arguments to bound the second term on the right-hand side. Then one may invoke known regret bounds for specific online learning algorithms. For example, the following corollary is immediate.

Corollary 2 *Consider an arbitrary online learning algorithm $\Pi_n \in \mathcal{P}_n$ and suppose that there exists $\sigma > 0$ such that $\sup_{w \in \mathcal{W}} \mathbb{E}\left[\ell(w, Z')^2\right] \leq \sigma^2$. Then, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \leq \frac{\text{regret}_{\Pi_n}(P_{W_n|S_n})}{n} + \sqrt{\frac{2\sigma^2 \log\left(\frac{1}{\delta}\right)}{n}} \ .$$

The proof follows from applying a standard concentration result for the lower tail of a sum of nonnegative random variables, which form martingales after centering. In particular, we note that

$$\mathbb{E}_{Z_t \sim \mu} \left\langle P_t, \ell(\cdot, Z_t) \right\rangle^2 = \mathbb{E}_{Z_t \sim \mu}\left[\mathbb{E}_{W \sim P_t} \ell(W, Z_t)\right]^2 \leq \sigma^2 \ ,$$

and use a standard result (stated as Lemma 26 in Appendix B.1) to upper bound $-M_{\Pi_n}$. Notice that this argument does not require the loss function to be uniformly bounded from above, or have light tails, yet it still yields a subgaussian bound for the lower tail of the martingale M_{Π_n}. Alternatively, one may assume that that the loss function is (uniformly) subgaussian in the sense that there exists $\Sigma > 0$ such that for all $\lambda < 0$,

$$\sup_{w \in \mathcal{W}} \mathbb{E}_{Z' \sim \mu}\left[e^{\lambda(\ell(w, Z') - \mathbb{E}_{Z \sim \mu} \ell(w, Z))}\right] \leq e^{\lambda^2 \Sigma^2 / 2} \ .$$

In that case, the second term in the upper bound of Corollary 2 becomes $\sqrt{\frac{\Sigma^2 \log\left(\frac{1}{\delta}\right)}{2n}}$, by standard exponential martingale inequalities.

Notably, the result of Corollary 2 holds with high probability *simultaneously for all randomized statistical learning algorithms* against which the online learner has bounded regret. This property mirrors the key property of PAC-Bayesian generalization bounds that hold uniformly over "posteriors" in the same sense, which thus justifies the name "online-to-PAC conversion". An extension to bounds that hold uniformly over time can be easily achieved by noting that online learning algorithms typically come with time-uniform bounds that hold with probability one over all data sequences, and thus it only remains to gain uniform control over the martingale term, which can be done via standard techniques.

In words, the key idea of the online-to-PAC conversion scheme is the following. In the generalization game, the cost function selected by the adversary corresponds to the generalization gap on example Z_t, and has zero expectation for every fixed $w \in \mathcal{W}$. As such, no matter what strategy it follows, the online learner's cost has zero expectation in each round, and its total cost is thus a martingale. Adding and subtracting an appropriately renormalized version of this zero-mean term to the generalization error establishes a connection between the regret of the online learner in the generalization game and the generalization error.

Notably, the goal of the online learner in this game is *not* to achieve low total cost, but rather to achieve a cumulative cost that is *comparable* to the generalization error of the statistical learning method. Indeed, since the costs are all zero-mean, minimizing the cost is a hopeless task by definition, and the online learner can only hope to do "not much worse" than an ideal comparator that may depend on the realization of the data sequence. Our comparator is chosen as the strategy that plays the decision of the statistical learning method in each round. Thus, the implication of the result is that whenever one can prove the *existence* an online learner that predicts the sequence of i.i.d. generalization gaps of the statistical learning method well, the statistical learning algorithm can be guaranteed to generalize well to test data.

We highlight that the relationship between the generalization error, the regret, and the total cost of the learner holds with *equality for all online learners*, which may appear counterintuitive at first sight. Indeed, the peculiarity of the generalization game implies several surprising facts. For instance, it is easy to see that whenever the generalization error $\overline{\text{gen}}(W_n, S_n)$ decays to zero as n increases, *all online learning algorithms have sublinear regret against* $P_{W_n|S_n}$, due to both sides of the bound of Theorem 1 necessarily vanishing for large n. Likewise, if one can establish a sublinear regret bound for any particular online learning algorithm, then it implies that the regret of all other online learning methods are also sublinear in this game. As an aside, we note that this observation can be used to prove *lower bounds* on the regret of online learning algorithms, and fact in our decomposition is closely related to a common technique for proving lower bounds via "Yao's minimax principle" (Yao, 1977). As long as one is interested in proving upper bounds on the generalization bound (as we are in the present paper), it is often more helpful to apply Theorem 1 with the infimum being taken over regret *bounds* instead of regrets.

The results above suggest a strategy to obtain upper bounds for the generalization error: choose an appropriate online learning algorithm and bound its regret against $P_{W_n|S_n}$. We show several examples on the remaining pages of this paper. However, sometimes one may obtain better bounds by considering a class of online learning algorithms. For example, if one chooses a finite class $\mathcal{C}_n$ of online learning algorithms, the argument of Corollary 2, combined with the union bound, immediately implies the following.

Corollary 3 *Consider a finite set $\mathcal{C}_n \subset \mathcal{P}_n$ of N online learning algorithms and suppose that there exists $\sigma > 0$ such that $\sup_{w\in\mathcal{W}} \mathbb{E}_{Z'\sim\mu}\ell(w, Z')^2 \le \sigma^2$. Then, with probability at least $1-\delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \le \min_{\Pi_n \in \mathcal{C}_n} \frac{\text{regret}_{\Pi_n}(P_{W_n|S_n})}{n} + \sqrt{\frac{2\sigma^2 \log\left(\frac{N}{\delta}\right)}{n}} .$$

The reason why it is often beneficial to consider larger classes of online learning algorithms is that the regret bounds may be different for different training samples and Corollary 3 allows us to use the best of these bounds. Corollary 3 is based on the naive upper bound

$$\inf_{\Pi_n \in \mathcal{C}_n} \left(\frac{\text{regret}_{\Pi_n}(P_{W_n|S_n})}{n} - M_{\Pi_n} \right) \le \inf_{\Pi_n \in \mathcal{C}_n} \frac{\text{regret}_{\Pi_n}(P_{W_n|S_n})}{n} + \sup_{\Pi_n \in \mathcal{C}_n} (-M_{\Pi_n}) ,$$

and the second term on the right-hand side in Corollary 3 is simply an upper bound for $\max_{\Pi_n \in \mathcal{C}_n}(-M_{\Pi_n})$. The price for the minimum over $\mathcal{C}_n$ is an additional term of order $\sigma\sqrt{(\log N)/n}$ in the upper bound. In principle, one may be able to improve on this naive bound by a more careful balancing of the regret and martingale terms. For example, if Π_n^* denotes an online learning algorithm that (approximately) minimizes $\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})$ over $\Pi_n \in \mathcal{C}_n$, then it suffices to bound $-M_{\Pi_n^*}$. Since Π_n^* depends on the random sample, obtaining sharp upper bounds for $-M_{\Pi_n^*}$ is a nontrivial matter. In some cases, one may be able to improve on the simple union bound that we used above. This requires a deeper understanding of the martingale process $\{M_{\Pi_n} : \Pi_n \in \mathcal{C}_n\}$. We leave this exciting question for future research.

3. Applications

In what follows, we instantiate the general bound of Theorem 1 using a variety of concrete choices for the online learning algorithm. For the sake of completeness, we include the proofs of the regret bounds we make use of in Appendix A. The purpose of this section is not necessarily to provide results that beat the state of the art, but rather to illustrate the use of our framework and demonstrate its flexibility. In Section 3.1 we start with deriving some classical PAC-Bayes-flavored generalization bounds, including some well-known ones and some others that we have not seen in the related literature. The impatient reader interested in going beyond PAC-Bayes can skip ahead to Section 3.2 that includes a range of new generalization guarantees derived from a general family of online learning methods known as "follow-the-regularized-leader".

3.1. PAC-Bayesian generalization bounds via multiplicative weights

The most elementary application of our framework is based on using the classical exponential weighted average (EWA)—or multiplicative weights—algorithm of Littlestone and Warmuth (1994) as the online learner's strategy in the generalization game (see also Vovk, 1990 and Freund and Schapire, 1997). The most basic version of this method is initialized with some fixed "prior" distribution $P_1 \in \Delta_{\mathcal{W}}$, and then sequentially calculates its updates by solving the optimization problem

$$P_{t+1} = \operatorname*{arg\,min}_{P \in \Delta_{\mathcal{W}}} \left\{ \langle P, c_t \rangle - \frac{1}{\eta} \mathcal{D}_{\mathrm{KL}}\left(P \| P_t\right) \right\} ,$$

where η is a positive *learning-rate* parameter, and $\mathcal{D}_{\mathrm{KL}}\left(P \| Q\right)$ is the relative entropy between the distributions P and Q. The minimizer can be shown to exist and satisfies

$$\frac{\mathrm{d}P_{t+1}}{\mathrm{d}P_t}(w) = \frac{e^{-\eta c_t(w)}}{\int_{\mathcal{W}} e^{-\eta c_t(w')} \mathrm{d}P_t(w')} .$$

In what follows, we derive a range of generalization bounds using the classical regret analysis of this algorithm and some of its variants. Throughout the section, we suppose that the loss function is uniformly bounded in the range $[0,1]$, but we note that a relaxation to more general subgaussian losses is also possible at the expense of a slightly more involved technical analysis.

3.1.1. A VANILLA PAC-BAYES BOUND

A direct application of the classical regret analysis of the multiplicative weights algorithm gives the following generalization bound via the reduction of Corollary 2:

Corollary 4 *Suppose that there exists $\sigma > 0$ such that $\sup_{w \in \mathcal{W}} \mathbb{E}\left[\ell(w, Z')^2\right] \leq \sigma^2$. Then, for any fixed $\eta > 0$, any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\mathrm{gen}}(W_n, S_n) \leq \frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{\eta n} + \frac{\eta}{2n} \sum_{t=1}^{n} \left\| \ell(\cdot, Z_t) - \mathbb{E}\left[\ell(\cdot, Z')\right] \right\|_{\infty}^2 + \sqrt{\frac{\sigma^2 \log(1/\delta)}{2n}} .$$

Note that the first two terms appearing on the right hand side depend on the realization of the data set S_n, and are thus random variables. Also notice however that the second term is not empirically observable as it features the test error $\mathbb{E}[\ell(\cdot, Z')]$. Under the additional assumption that the loss function ℓ is almost surely bounded in $[0,1]$, the data-dependent quantity $\frac{1}{n}\sum_{t=1}^n \|\ell(\cdot, Z_t) - \mathbb{E}[\ell(\cdot, Z')]\|_\infty^2$ can be simply bounded by 1. This result essentially recovers the original PAC-Bayes bound of McAllester (1998). The proof of the regret bound serving as the foundation of our result is included in Appendix A.1.

Notice that the bound of Corollary 4 only holds for a fixed η, and optimizing the bound requires choosing a data-dependent η due to the randomness of $\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)$. Such learning-rate choices are disallowed by our framework, as they would require the learner to access information about future data points $Z_t, \dots, Z_n$ when picking its decision P_t. However, we may use the second statement of Theorem 1 with $\mathcal{C}_n$ containing exponentially weighted average algorithms with a given prior P_1 and a range of different values of the learning parameter η. In particular, an application of Corollary 3 on an appropriately chosen finite range of learning rates gives the following result.

Corollary 5 *Suppose that $\ell(w,z) \in [0,1]$ for all w, z. Fix $\epsilon \in (0,1]$. Then, for any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1-\delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \le \left(1 + \frac{\epsilon^2}{2}\right)\sqrt{\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{2n}} + \sqrt{\frac{\log\log(4\sqrt{n}) + \log(1/\delta) + \log\frac{2}{\epsilon}}{2n}}\,.$$

We include the simple proof for didactic purposes below.

Proof Recall that if the exponentially weighted average algorithm is run with tuning parameter η, then its regret may be upper bounded by

$$\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{\eta n} + \frac{\eta}{8}\,.$$

Observe first that the upper bound is trivial whenever $\frac{\eta}{8} > 1$. To proceed, denote the optimal (data-dependent) learning rate as $\eta^* = \sqrt{8\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)/n}$. Running the exponentially weighted average algorithm with this choice of the tuning parameter yields the regret bound

$$\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{\eta^* n} + \frac{\eta^*}{8} = \sqrt{\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{2n}}\,.$$

If $\eta^* < 2/\sqrt{n}$, then $\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right) \le 1/2$ the regret bound is at most $1/\sqrt{4n}$, which is absorbed by the second term in the stated bound. Hence, it is sufficient to consider learning rates that are at least as large as $2/\sqrt{n}$.

Let $a > 0$ and let $\mathcal{C}_n$ contain all exponentially weighted average online learning algorithms with prior P_1 and learning rate $\eta \in \{a^i : i \in \mathbb{N}\} \cap [2/\sqrt{n}, 8]$. By Corollary 3, with probability at least $1-\delta$, for all $W_n = \mathcal{A}(S_n)$, we have

$$\overline{\text{gen}}(W_n, S_n) \le \min_{\eta \in \{a^i : i\in\mathbb{N}\}\cap[2/\sqrt{n}, 8]} \left(\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{\eta n} + \frac{\eta}{8}\right) + \sqrt{\frac{\log\frac{\log_a(2\sqrt{n})}{\delta}}{2n}}\,.$$

Since the optimal choice of η in the set $\{a^i : i \in \mathbb{N}\}$ is at most a factor of a away from η^*, we get that

$$\min_{\eta \in \{a^i : i\in\mathbb{N}\}\cap[2/\sqrt{n}, 8]} \left(\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{\eta n} + \frac{\eta}{8}\right) \le \frac{a + 1/a}{2}\sqrt{\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{2n}}\,,$$

yielding the bound

$$\overline{\text{gen}}(W_n, S_n) \le \frac{a + 1/a}{2}\sqrt{\frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)}{2n}} + \sqrt{\frac{\log\log(4\sqrt{n}) + \log(1/\delta) - \log\log a}{2n}}\,.$$

Now we may choose the value of a. If $a = 1+\epsilon$ for some $\epsilon \in (0,1]$, then $(a+1/a)/2 \le 1+\epsilon^2/2$. Moreover, $\log a \ge \epsilon/2$, which implies the stated inequality. ■

3.1.2. A PARAMETER-FREE BOUND

The bound of Corollary 5 is derived from aggregating a number of different parametric bounds over a grid of learning rates η. This procedure adds a $\sqrt{\frac{\log\log n}{n}}$ term to the generalization bound via a union bound. It is natural to ask if it is possible to avoid this overhead by tapping into the literature on *parameter-free online learning* algorithms that avoid using learning rates altogether (Chaudhuri et al.; Chernov and Vovk, 2010; Luo and Schapire, 2015; Koolen and Van Erven, 2015; Orabona and Pál, 2016). Due to the flexibility of our online-to-PAC framework, this question can be easily answered in the positive. The following theorem instantiates a generalization bound that can be derived from Corollary 6 of Orabona and Pál (2016) (or, equivalently, Theorem 9 of van der Hoeven et al., 2018).

Corollary 6 *Suppose that $\ell(w,z) \in [0,1]$ for all w, z. Then, for any $P_1 \in \Delta_{\mathcal{W}}$, with probability at least $1-\delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \le \sqrt{\frac{3\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right) + 9}{n}} + \sqrt{\frac{\log\frac{1}{\delta}}{2n}}.$$

Notably, this bound does not feature any logarithmic factors of n. We are not aware of PAC-Bayesian guarantees with such property, with the exception of Theorem 1 of McAllester (2013) that only holds for countable hypothesis classes. We find it plausible that similar guarantees can be derived from the other works we have listed above. In particular, Theorem 9 of Chernov and Vovk (2010) provides essentially the same regret guarantee as the one we have used above, and additionally holds uniformly over time.

3.1.3. A DATA-DEPENDENT BOUND

We now provide a tighter data-dependent bound derived from a slightly more sophisticated version of the standard exponentially weighted average forecaster. Under some conditions, this bound will decay to zero at a fast rate, which necessitates some adjustments to the basic setup that will allow us to bound the martingale term $-M_{\Pi_n}$ more effectively. In particular, we will employ the skewed cost surrogate $c_t(w) = \ell(w,Z_t) - \mathbb{E}\left[\ell(w,Z')\right] + \eta\left(\ell(w,Z_t)\right)^2$, which turns $-M_{\Pi_n}$ into a supermartingale that can be upper-bounded much more tightly. The price paid for this adjustment is a data-dependent term appearing in the generalization bound that will be shown to be strictly dominated by the regret bound arising from the online-to-PAC conversion we will employ. We note that this technique can be used more generally to achieve fast rates for other algorithms.

The online learning algorithm itself will make use of two tricks familiar from the online learning literature: optimistic updates as introduced by Rakhlin and Sridharan (2013a,b) and second-order adjustments as used in several works on adaptive online learning (e.g., Cesa-Bianchi et al., 2007; Gaillard et al., 2014; Koolen and Van Erven, 2015). In particular, the online learning algorithm uses a *guess* $g_t \in \mathbb{R}^{\mathcal{W}}$ of the cost function c_t, and calculates two sequences of updates. The first is a sequence of auxiliary distributions initialized at $\widetilde{P}_1$ and updated as

$$\frac{\mathrm{d}\widetilde{P}_{t+1}}{\mathrm{d}\widetilde{P}_t}(w) = \frac{e^{-\eta c_t(w) - \eta^2(c_t(w)-g_t(w))^2}}{\int_{\mathcal{W}} e^{-\eta c_t(w') - \eta^2(c_t(w')-g_t(w'))^2}\mathrm{d}P_t(w')},$$

and the main update is calculated as

$$\frac{\mathrm{d}P_{t+1}}{\mathrm{d}\widetilde{P}_{t+1}}(w) = \frac{e^{-\eta g_{t+1}(w)}}{\int_{\mathcal{W}} e^{-\eta g_{t+1}(w')}\mathrm{d}P_t(w')}.$$

We state a regret guarantee for this algorithm in Appendix A.2. Instantiating the method with the choice $g_t(w) = -\mathbb{E}\left[\ell(w, Z')\right]$ and plugging the regret bound into our online-to-PAC conversion scheme gives the following generalization bound:

Corollary 7 *Suppose that $\ell(w, z) \in [0, 1]$ for all w, z. For any $\widetilde{P}_1 \in \Delta_{\mathcal{W}}$ and any $\eta \in \left[0, \frac{1}{4}\right]$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \leq \frac{\mathcal{D}_{\text{KL}}\left(P_{W_n|S_n} \,\middle\|\, \widetilde{P}_1\right)}{\eta n} + \frac{3\eta}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\ell(W_n, Z_t)\right)^2\middle| S_n\right] + \frac{\log\frac{1}{\delta}}{\eta n}.$$

Proof We first observe that the generalization error can be bounded as

$$\begin{aligned}
\overline{\text{gen}}(W_n, S_n) &= -\frac{1}{n}\sum_{t=1}^{n}\left\langle P_{W_n|S_n}, c_t\right\rangle + \frac{\eta}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\ell(W_n, Z_t)\right)^2\middle| S_n\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\left\langle P_t - P_{W_n|S_n}, c_t\right\rangle - \frac{1}{n}\sum_{t=1}^{n}\left\langle P_t, c_t\right\rangle + \frac{\eta}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\ell(W_n, Z_t)\right)^2\middle| S_n\right] \\
&\leq \frac{\mathcal{D}_{\text{KL}}\left(P_{W_n|S_n} \,\middle\|\, \widetilde{P}_1\right)}{\eta n} + \frac{3\eta}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\ell(W_n, Z_t)\right)^2\middle| S_n\right] - \frac{1}{n}\sum_{t=1}^{n}\left\langle P_t, c_t\right\rangle,
\end{aligned}$$

where we used the regret bound of Theorem 23 in the last step (noting that the condition $\eta c_t(w) \leq \frac{1}{2}$ is satisfied under our condition on η), and that

$$\left(c_t(w) - g_t(w)\right)^2 = \left(\ell(w, Z_t) + \eta\left(\ell(w, Z_t)\right)^2\right)^2 \leq 2\left(\ell(w, Z_t)\right)^2$$

holds under our conditions on η. To continue, we notice that the sum $-\sum_{t=1}^{n}\left\langle P_t, c_t\right\rangle$ is a supermartingale that decays at a fast rate. In particular, we rewrite this term and apply Lemma 27 with $X_t = \left\langle P_t, \ell(\cdot, Z_t)\right\rangle$ and $\lambda = \eta$ to obtain the following bound that holds with probability at least $1 - \delta$:

$$\begin{aligned}
-\frac{1}{n}\sum_{t=1}^{n}\left\langle P_t, c_t\right\rangle &= \frac{1}{n}\sum_{t=1}^{n}\left(\left\langle P_t, \mathbb{E}\left[\ell(\cdot, Z')\right]\right\rangle - \left\langle P_t, \ell(\cdot, Z_t)\right\rangle - \eta\left\langle P_t, \left(\ell(\cdot, Z_t)\right)^2\right\rangle\right) \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\left(\left\langle P_t, \mathbb{E}\left[\ell(\cdot, Z')\right]\right\rangle - \left\langle P_t, \ell(\cdot, Z_t)\right\rangle - \eta\left(\left\langle P_t, \ell(\cdot, Z_t)\right\rangle\right)^2\right) \leq \frac{\log\frac{1}{\delta}}{\eta n},
\end{aligned}$$

where the first inequality is Jensen's. This concludes the proof. ■

By further upper bounding the quadratic term appearing in the upper bound by the training error, we obtain the following relaxation of the bound:

Corollary 8 *Suppose that $\ell(w, z) \in [0, 1]$ for all w, z. For any $\widetilde{P}_1 \in \Delta_{\mathcal{W}}$ and any $\eta \in \left[0, \frac{1}{2}\right]$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\mathbb{E}\left[\ell(W_n, Z')\middle| S_n\right] \leq \frac{\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\ell(W_n, Z_t)\middle| S_n\right]}{1 - \eta} + \frac{\mathcal{D}_{\text{KL}}\left(P_{W_n|S_n}\middle\|\widetilde{P}_1\right)}{\eta n} + \frac{\log\frac{1}{\delta}}{2\eta n}.$$

Several similar data-dependent bounds have been proposed in the PAC-Bayesian literature. For instance, the "PAC-Bayes-Bernstein inequality" of Seldin et al. (2012), which features the variance of the losses instead of their second moment, can be recovered from the same regret analysis as above by setting $g_t = 0$. Later results of Tolstikhin and Seldin (2013) and Mhammedi et al. (2019) have replaced this unobservable quantity by the empirical variance of the training loss. While these results were proved using sophisticated concentration inequalities combined with PAC-Bayesian "change of measure" arguments, our result directly follows from a combination of a few standard techniques from online learning. The main merit of these results (shared by our result above) is that they imply a fast rate of order $1/n$ when the training error is zero.

3.2. Generalized PAC-Bayesian bounds via Following the Regularized Leader

We now provide a range of entirely new generalization bounds derived from a general class of online learning algorithms known as *Follow the Regularized Leader* (FTRL, see, e.g., Rakhlin, 2009; Shalev-Shwartz, 2012; Orabona, 2019). FTRL algorithms are defined using a convex regularization function $h : \Delta_{\mathcal{W}} \to \mathbb{R}$. For the sake of this paper, we concentrate on regularizers that are α-*strongly convex* with respect to a norm $\|\cdot\|$ defined on the set of signed measures on $\mathcal{W}$, in the sense that the following inequality is satisfied for all $P, P' \in \Delta_{\mathcal{W}}$ and all $\lambda \in [0, 1]$:

$$h(\lambda P + (1-\lambda)P') \le \lambda h(P) + (1-\lambda)h(P') - \frac{\alpha\lambda(1-\lambda)}{2} \left\|P - P'\right\|^2. \tag{1}$$

We also assume that h is proper in the sense that it never takes the value $-\infty$ and is not identically equal to $+\infty$, and that it is lower semicontinuous on its effective domain. Given such a regularization function and a positive learning-rate parameter η, we can define the distribution P_t picked by FTRL in round t as

$$P_t = \operatorname*{arg\,min}_{P \in \Delta_{\mathcal{W}}} \left\{ \left\langle P, \sum_{k=1}^{t-1} c_k \right\rangle + \frac{1}{\eta} h(P) \right\}.$$

The existence of the minimum is guaranteed by the compactness of $\Delta_{\mathcal{W}}$ under $\|\cdot\|$, and its uniqueness is ensured by the strong convexity of h. In what follows, we derive generalization bounds for this vanilla version of FTRL as well as some of its straightforward variants, and we will instantiate the bounds with some interesting regularization functions.

3.2.1. STRONGLY CONVEX REGULARIZERS

We first state a generalization bound obtained via an application of the classical FTRL analysis for strongly convex regularizers.

Theorem 9 *Suppose that h is α-strongly convex with respect to the norm $\|\cdot\|$, and that there exists $\sigma > 0$ such that $\sup_{w\in\mathcal{W}} \mathbb{E}_{Z'\sim\mu} \ell(w, Z')^2 \le \sigma^2$. Then, for any $\eta > 0$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \le \frac{h(P_{W_n|S_n}) - h(P_1)}{\eta n} + \frac{\eta}{2\alpha n} \sum_{t=1}^{n} \left\|\ell(\cdot, Z_t) - \mathbb{E}\left[\ell(\cdot, Z')\right]\right\|_*^2 + \sqrt{\frac{\sigma^2 \log(1/\delta)}{2n}}.$$

We provide the proof of the regret bound forming the basis of this theorem in Appendix A.3 (cf. Theorem 9). A straightforward covering argument over the choice of η gives the following bound.

Corollary 10 *Suppose that h is α-strongly convex with respect to the norm $\|\cdot\|$. Furthermore, suppose that $\ell(w, z) \in [0, 1]$ for all w, z and that $\left\|\ell(\cdot, Z_t) - \mathbb{E}\left[\ell(\cdot, Z')\right]\right\|_* \le B$ for some positive B. Fix $\epsilon \in (0, 1]$. Then, for any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \le \left(1 + \frac{\epsilon^2}{2}\right) \sqrt{\frac{B^2 \left(h\left(P_{W_n|S_n}\right) - h(P_1)\right)}{2\alpha n}} + \sqrt{\frac{\log\log(4\sqrt{n}) + \log(1/\delta) + \log\frac{2}{\epsilon}}{2n}}.$$

The proof follows the same arguments as the proof of Corollary 5. The above two bounds respectively recover the PAC-Bayesian generalization bounds of Corollaries 4 and 5 when setting $h(P) = \mathcal{D}_{\text{KL}}\left(P \| P_1\right)$, which is known to be 1-strongly convex with respect to the total variation norm (whose dual norm is the supremum norm). We provide further examples at the end of this section.

3.2.2. EMPIRICAL BOUNDS VIA OPTIMISTIC FTRL

One downside of the bound claimed in the previous section is that it depends on the dual norms of $\ell(\cdot, Z_t) - \mathbb{E}[\ell(\cdot, Z')]$, which involves the unobservable quantity $\mathbb{E}[\ell(\cdot, Z')]$. It is often more desirable to provide generalization bounds that are fully empirical and can be evaluated without having to estimate the test error (which would indeed defeat the purpose of proving generalization bounds in the first place). In this section, we provide a simple remedy to this issue by considering an *optimistic* version of FTRL. Optimistic online learning algorithms were first proposed by Rakhlin and Sridharan (2013a,b) as algorithms that can take advantage of a *guess* g_t of the cost function c_t. Such methods provide tighter guarantees whenever c_t and g_t are close in an appropriate sense, and typically retain the worst-case guarantees of FTRL when this is not the case. In our setting, it is natural to pick g_t as the predictable part of c_t corresponding to the test loss: $g_t(w) = -\mathbb{E}[\ell(w, Z')]$. Indeed, the only unpredictable part of the cost function from the perspective of the online learner is the empirical loss $\ell(w, Z_t)$ on the t-th data point, as the other additive term remains the same in all rounds.

We now derive a generalization bound by considering the optimistic version of FTRL that picks the following distribution P_t in each round t:

$$P_t = \operatorname*{arg\,min}_{P \in \Delta_{\mathcal{W}}} \left\{ \left\langle P, g_t + \sum_{k=1}^{t-1} c_k \right\rangle + \frac{1}{\eta} h(P) \right\} ,$$

The generalization bound derived using this algorithm with the above choice of g_t is stated in the following theorem.

Theorem 11 *Suppose that h is α-strongly convex with respect to the norm $\|\cdot\|$. Then, for any $\eta > 0$, any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1-\delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\mathrm{gen}}(W_n, S_n) \le \frac{h(P_{W_n|S_n}) - h(P_1)}{\eta n} + \frac{\eta}{2\alpha n} \sum_{t=1}^{n} \left\| \ell(\cdot, Z_t) \right\|_*^2 + \sqrt{\frac{\sigma^2 \log(1/\delta)}{n}} .$$

The proof makes use of a standard regret bound for optimistic FTRL that we state and prove as Theorem 25 in Appendix A.3 for completeness. As can be seen from the statement, this bound improves over that of Theorem 9 in that it replaces the norm of the generalization gap with the norm of the training loss, which can be empirically measured. An optimally tuned version similar to Theorem 10 can be derived using similar techniques as before.

3.2.3. EXAMPLE: p-NORM REGULARIZATION

From the perspective of convex analysis, the family of p-norm distances is a natural candidate for defining regularization functions. Concretely, we define the weighted p-norm distance between the signed measures $P, P' \in \Delta_{\mathcal{W}}$ and base measure P_1 as the L_p distance between their Radon–Nykodim derivatives with respect to P_1:

$$\|P - P'\|_{p,P_1} = \left(\int_{\mathcal{W}} \left(\frac{\mathrm{d}P}{\mathrm{d}P_1} - \frac{\mathrm{d}P'}{\mathrm{d}P_1} \right)^p \mathrm{d}P_1 \right)^{1/p} . \tag{2}$$

The corresponding dual norm is the L_q-norm defined for all f as

$$\|f\|_{q,P_1,*} = \left(\int_{\mathcal{W}} f^q \mathrm{d}P_1 \right)^{1/q} ,$$

with $q > 1$ such that $1/p + 1/q = 1$. It is useful to note that the distance $\|P - P_1\|_{p,P_1}^p$ is the f-divergence corresponding to $\varphi(x) = (x-1)^p$, which is known under several different names such as Hellinger divergence of order p, p-Tsallis divergence or simply α-divergence with $\alpha = p$ (see, e.g., Sason and Verdú, 2016; Nielsen and Nock, 2011). The case $p = 2$ is often given special attention, and the corresponding squared norm can be seen to match Pearson's χ^2-divergence (Pearson, 1900). We denote this divergence by $\mathcal{D}_{\chi^2}$ below.

Powers of the norm defined above exhibit different strong-convexity properties depending on the value of p, with two distinct regimes $p \in (1, 2]$ and $p > 2$. The following corollary summarizes the results obtained in these two regimes when respectively setting $h(P) = \|P - P_1\|_{p,P_1}^2$ and $h(P) = \|P - P_1\|_{p,P_1}^p$:

Corollary 12 *Fix $p > 1$ and q such that $1/p + 1/q = 1$, and suppose that there exists $B > 0$ such that $\sup_{w\in\mathcal{W}} \mathbb{E}_{Z'\sim\mu} \left(\ell(w, Z')\right)^{\min\{2,q\}} \le B$. Then, for any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies one the following bounds:*

(a) For $p \in (1, 2]$,

$$\overline{\text{gen}}(W_n, S_n) \le \frac{\left\|P_{W_n|S_n} - P_1\right\|_{p,P_1}^2}{\eta n} + \frac{\eta}{(p-1)n}\sum_{t=1}^{n}\left\|\ell(\cdot, Z_t)\right\|_{q,P_1,*}^2 + \sqrt{\frac{B^2\log(1/\delta)}{2n}}\,.$$

(b) For $p \ge 2$ and q,

$$\overline{\text{gen}}(W_n, S_n) \le \frac{\left\|P_{W_n|S_n} - P_1\right\|_{p,P_1}}{\eta n} + 2\left(\frac{\eta}{2n}\right)^{q-1}\sum_{t=1}^{n}\left\|\ell(\cdot, Z_t)\right\|_{q,P_1,*}^q + B\left(\frac{\log(1/\delta)}{n}\right)^{1-1/q}\,.$$

These bounds are fully empirical in the sense that they depend on the realization of the data sequence, up to an additional martingale concentration term of order $B\left(\log(1/\delta)/n\right)^{1-1/q}$. The perhaps unexpectedly mild dependence on the confidence level δ is due to the fact that the martingale average $-M_{\Pi_n}$ that needs to be probabilistically controlled is the *lower tail* of a sum of nonnegative random variables, which can be effectively bounded even for heavy-tailed random variables. Upper-bounding the data-dependent terms using Markov's inequality, one can recover the results of Bégin et al. (2016) and Alquier and Guedj (2018) that were proved under the much stronger assumptions that the losses are bounded or that they always have finite variance. Rodríguez-Gálvez et al. (2021) derive a comparable bound on the *expected* generalization error in the special case $p = 2$. Our bounds in the regime $p > 2$ do not require such assumptions and remain meaningful when the losses are heavy tailed and the q-th moment of the random loss is bounded only for some $q < 2$. In such cases, our result implies a slow rate of $n^{-(1-1/q)}$ for the generalization error, which is expected when dealing with concentration of heavy-tailed random variables (Gnedenko and Kolmogorov, 1954). In the regime $p \in (1, 2]$, our bound interpolates between the guarantee for $p = 2$ and the standard PAC-Bayesian bound of Corollary 4 as p approaches 1, at least in terms of dependence on the L_q-norm of the loss function. In terms of dependence on the divergence measures, this interpolation fails as p tends to 1, as the squared L_p-divergence converges to the squared total variation distance which is not strongly convex. Accordingly, the bound blows up in this regime and Corollary 4 gives a strictly better bound. All of these guarantees require the boundedness of $\|P_n - P_1\|_{p,P_1}$, which becomes a more and more stringent condition as p increases.

The results in Corollary 12 are direct consequences of Theorem 9. The case $p = 2$ is the simplest and can be proved by picking $h(P) = \mathcal{D}_{\chi^2}(P\|P_1)$. Being a squared 2-norm, h is obviously 1-strongly convex with respect to $\|P - P_0\|_{2,P_0}$ as it satisfies the condition of Equation (1) with equality. A similar argument works for the regime $p \in (1, 2]$, where the choice $h(P) = \|P - P'\|_{p,P_0}^2$ exhibits $2(p-1)$-strong convexity with respect to the norm $\|\cdot\|_{p,P_0}$ (see, e.g., Proposition 3 in Ball et al., 1994, that also establishes that strong convexity does not hold for $p > 2$).

The case $p \ge 2$ is slightly more complex and it requires minor adjustments to the proof of Theorem 9. In this range we consider the regularizer $h(P) = \|P - P_1\|_{p,P_1}^p$. While this function is not strongly convex, it satisfies the following weaker notion of *p-uniform convexity*:

$$h(P) \ge h(P') + \langle g, P - P'\rangle + \frac{\alpha}{2}\|P - P'\|_{p,P_1}^p$$

with $\alpha = 2$, where $g \in \partial h(P)$. We refer to Ball et al. (1994) who attribute this result to Clarkson (1936). Following the proof of Lemma 29, we can show that Φ satisfies the following *q-uniform smoothness* condition:

$$\mathcal{B}_\Phi\left(f \| f'\right) \leq \frac{1}{\alpha^{q-1}} \left\| f - f' \right\|_{q,P_0,*}^q .$$

Replacing the inequality used in the last step of the proof of Theorem 9 with the above proves the regret bound of FTRL used in the corollary. It thus remains to bound the martingale term $-M_{\Pi_n}$, which can be done via an application of Lemma 28 presented in Appendix B.1. Indeed, applying this result with $X_t = \mathbb{E}\left[\ell(\widetilde{W}_t, Z_t)\middle| \mathcal{F}_{t-1}, S_n\right]$ implies that for any $\lambda > 0$, we have, with probability at least $1-\delta$,

$$-M_{\Pi_n} \leq \lambda^{q-1} B^q + \frac{\log \frac{1}{\delta}}{\lambda n} .$$

Setting $\lambda = B\left(\frac{\log\frac{1}{\delta}}{n}\right)^{1/q}$ concludes the proof.

3.2.4. EXAMPLE: SMOOTHED RELATIVE-ENTROPY REGULARIZATION

Let us now consider the special case where the hypothesis space is $\mathcal{W} = \mathbb{R}^d$. In this case, a common idea in the PAC-Bayesian literature is to smooth the posterior distribution $P_{W_n|S_n}$ by adding Gaussian noise to an otherwise deterministic output hypothesis W_n^* to ensure the boundedness of $\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)$. The effect of this perturbation is then typically addressed by analyzing the gap $\mathbb{E}\left[\ell(W_n, Z') - \ell(W_n^*, Z') | S_n\right]$. Making sure that this gap does not end up dominating the bound generally necessitates using perturbation levels that go to zero for large n. In this section, we provide an alternative smoothing method that allows using constant perturbation levels for a class of smooth functions.

Our approach is based on an FTRL variant based on a smoothed version of the relative entropy as regularization function. In order to construct this regularizer, we define the Gaussian smoothing operator G_γ that acts on any distribution $P \in \Delta_{\mathcal{W}}$ as $G_\gamma P = \int_{\mathcal{W}} \mathcal{N}(w, \gamma^2 I) \mathrm{d}P(w)$, where $\mathcal{N}(w, \gamma^2 I)$ is the d-dimensional Gaussian distribution with mean w and covariance $\gamma^2 I$. Using this operator, we define the γ-smoothed relative entropy as $\mathcal{D}_\gamma\left(P \| P'\right) = \mathcal{D}_{\mathrm{KL}}\left(G_\gamma P \| G_\gamma P'\right)$ and set $h(P) = \mathcal{D}_\gamma\left(P \| P_1\right)$. Similarly, we define the smoothed total variation distance between P and P' as $\left\|P - P'\right\|_\gamma = \left\|G_\gamma P - G_\gamma P'\right\|_{\mathrm{TV}}$. Both of these divergences have the attractive property that they remain meaningfully bounded under much milder assumptions than their unsmoothed counterparts, even when the supports of P and P' are disjoint.

It is straightforward to verify that the Bregman divergence associated with h satisfies

$$\mathcal{B}_h\left(P \| P'\right) = \mathcal{D}_\gamma\left(P \| P'\right) \geq \frac{1}{2} \left\|G_\gamma\left(P - P'\right)\right\|_{\mathrm{TV}}^2 = \frac{1}{2} \left\|P - P'\right\|_\gamma^2 ,$$

thus implying 1-strong convexity in terms of the smoothed total variation distance. The dual norm of the smoothed TV distance is defined as $\|f\|_{\gamma,*} = \sup_{\|P-P'\|_\gamma \leq 1} \langle f, P - P' \rangle$, which, together with the above arguments and Theorem 11, immediately implies the following result:

Corollary 13 *Suppose that there exists $\sigma > 0$ such that $\sup_{w\in\mathcal{W}} \mathbb{E}_{Z'\sim\mu} \ell(w, Z')^2 \leq \sigma^2$. Then, for any $\eta > 0$, any $\gamma \geq 0$, any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\mathrm{gen}}(W_n, S_n) \leq \frac{\mathcal{D}_\gamma\left(P_{W_n|S_n} \middle\| P_0\right)}{\eta n} + \frac{\eta}{n} \sum_{t=1}^n \left\|\ell(\cdot, Z_t)\right\|_{\gamma,*}^2 + \sqrt{\frac{2\sigma^2 \log\frac{1}{\delta}}{n}} .$$

It remains to be shown that the dual norm $\|\ell(\cdot, z)\|_{\gamma,*}$ can be bounded meaningfully. By the intuitive properties of the smoothed total variation distance, one can reasonably expect this norm to capture the smoothness properties of the loss function, and that it is small whenever $\ell(\cdot, z)$ is bounded and highly smooth. In what follows, we provide an upper bound on this norm that holds for a class infinitely smooth functions. Specifically, we say that a function f is infinitely smooth if all of its higher-order directional derivatives exist and satisfy $D^j f(w|v_1, v_2, \ldots, v_j) \leq \beta_j$ for all directions $v_1, v_2, \ldots, v_j$, all $w \in \mathcal{W}$, and all j. For such functions, the following lemma provides an upper bound on $\|f\|_{\gamma,*}$:

Lemma 14 *Suppose that f is infinitely smooth in the above sense. Then, the dual norm $\|f\|_{\gamma,*}$ satisfies $\|f\|_{\gamma,*} \leq \sum_{j=0}^{\infty} \left(\gamma\sqrt{d}\right)^j \beta_j$.*

The proof is based on a successive smoothing argument and is provided in Appendix B.3. With the help of this lemma, we can thus obtain the following result:

Corollary 15 *Suppose that $\ell(\cdot, z)$ is infinitely smooth for all z with $\beta_j \leq \beta$ for all $j \geq 0$, that there exists $\sigma > 0$ such that $\sup_{w\in\mathcal{W}} \mathbb{E}_{Z'\sim\mu}\ell(w, Z')^2 \leq \sigma^2$, and that $\gamma < 1/\sqrt{d}$. Then, for any $\eta > 0$, any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \leq \frac{\mathcal{D}_\gamma\left(P_{W_n|S_n} \middle\| P_0\right)}{\eta n} + \frac{\eta\beta}{1-\gamma\sqrt{d}} + \sqrt{\frac{2\sigma^2 \log\frac{1}{\delta}}{n}}\,.$$

The result follows from applying Corollary 13 and upper-bounding

$$\|\ell(\cdot, Z_t)\|_{\gamma,*}^2 \leq \beta \sum_{j=0}^{\infty} \left(\gamma\sqrt{d}\right)^j = \frac{\beta}{1-\gamma\sqrt{d}}.$$

Setting the smoothing level as $\gamma = 1/2\sqrt{d}$ guarantees that the second term is of order $\eta\beta$. To our knowledge, PAC-Bayesian guarantees of similar form can only be derived for perturbation levels that decay to zero as n increases, which severely limits the potential gains that can be achieved by smoothing.

In order to obtain a more explicit bound, we note that the smoothed relative entropy can be upper-bounded in terms of the squared Wasserstein-2 distance as $\mathcal{D}_\gamma\left(P\|P'\right) \leq \frac{1}{2\gamma^2}\mathbb{W}_2^2(P, P')$. For completeness, we give the precise definition of the Wasserstein distance $\mathbb{W}_2$ and a direct proof of this result in Appendix B.4. The following corollary states the resulting bound when setting $\gamma = 1/(2\sqrt{d})$.

Corollary 16 *Suppose that $\ell(\cdot, z)$ is infinitely smooth for all z with $\beta_j \leq \beta$ for all $j \geq 0$, and that there exists $\sigma > 0$ such that $\sup_{w\in\mathcal{W}} \mathbb{E}_{Z'\sim\mu}\ell(w, Z')^2 \leq \sigma^2$. Then, for any $\eta > 0$, any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, with probability at least $1 - \delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\text{gen}}(W_n, S_n) \leq \frac{2d\mathbb{W}_2^2\left(P_{n|S}, P_1\right)}{\eta n} + 2\beta\eta + \sqrt{\frac{\sigma^2 \log(1/\delta)}{n}}\,.$$

We are not aware of any directly comparable results in the literature. Zhang et al. (2018), Wang et al. (2019) and Rodríguez-Gálvez et al. (2021) provide vaguely similar guarantees that depend on the Wasserstein-1 distance and only require bounded first derivatives, but it is not clear if these bounds are decreasing with the sample size n in general. Whenever all hypotheses satisfy $\|w\|_2 \leq R$ for some R, the result stated above implies an upper bound on the expected generalization error that scales as $R\sqrt{\beta d/n}$ whenever all hypotheses satisfy $\|w\|_2 \leq R$ for some R, which is directly comparable with what one might obtain via a straightforward uniform convergence argument involving the covering number of Lipschitz functions on a bounded domain (see, e.g., Dudley, 1984). The dependence on the dimension d of such guarantees can be relaxed or completely removed when assuming more structure about the loss function (Bartlett, 1998; Williamson et al., 2000; Zhang, 2002). Whether such arguments can be applied to remove the dependence on d from the above bound is a curious problem we leave open for future research.

4. Extensions

The online-to-PAC conversion framework naturally lends itself to a number of extensions that allow one to prove more and more advanced generalization bounds. In this section, we provide a few such extensions and briefly explain how they can be applied to tighten the bounds presented in the previous sections.

4.1. Conditional online-to-PAC conversions

One limitation of standard PAC-Bayesian bounds is that the prior P_1 is not allowed to depend in any way on the training data S_n. This entails several further limitations, for instance that the bounds can become vacuous even in the simplest setting of learning one-dimensional threshold classifiers (see, e.g., Livni and Moran, 2020). Our framework presented in Section 2 shares the same limitation: the online learner is only allowed causal access to the training data, and in particular, its sequential predictions cannot depend on the entire data set. As such, the further limitations of PAC-Bayes are also inherited. In the context of PAC-Bayesian and information-theoretic generalization, this issue has been successfully addressed by Audibert (2004) and Catoni (2007) via the notion of "almost exchangeable priors", which allows the recovery of all classical PAC-learning bounds from a PAC-Bayesian framework. This idea has recently been rediscovered and popularized by the works of Steinke and Zakynthinou (2020) and Hellström and Durisi (2020) (see also Haghifam et al., 2021; Grünwald et al., 2021).

As we show below, our framework can readily address the issue at hand via a simple extension that we call *conditional online-to-PAC conversion*. Similarly to the frameworks described above, we define a supersample of $2n$ data points sampled i.i.d. from the distribution μ, denoted as $\widehat{S}_n = (Z_1^{+1}, Z_1^{-1}, Z_2^{+1}, Z_2^{-1}, \dots, Z_n^{+1}, Z_n^{-1})$. Based on these samples, we define the *conditional generalization game* where the following steps are repeated in a sequence of rounds $t = 1, 2, \dots, n$:

1. The online learner picks a distribution $P_t \in \Delta_{\mathcal{W}}$;
2. the adversary draws an index $I_t \in \{-1, +1\}$ uniformly at random, and selects the cost function $c_t : \mathcal{W} \to \mathbb{R}$ defined for each $w \in \mathcal{W}$ as
$$c_t(w) = \ell(w, Z_t^{I_t}) - \frac{1}{2}\left(\ell(w, Z_t^{+1}) - \ell(w, Z_t^{-1})\right) \ ;$$
3. the online learner incurs cost $\langle P_t, c_t \rangle$;
4. the adversary reveals the index I_t to the online learner.

In this setup, the online learner is allowed even more knowledge than in the standard generalization game: besides knowing the loss function ℓ and the data distribution μ, the online learner is also allowed to know the entire supersample $\widehat{S}_n$. Thus, revealing the index I_t in round t reveals the entire cost function c_t to the learner. The outcome of the game is associated with the training and test loss of the statistical learning algorithm by setting the training set as $S_n = \left\{Z_t^{I_t}\right\}_{t=1}^n$ and the test set as $S_n' = \left\{Z_t^{-I_t}\right\}_{t=1}^n$.

We treat the additional formalities by introducing the following notation. We use $\widehat{\mathcal{F}}_t$ to denote the sigma-algebra induced by the *entire data set* $\widehat{S}_n$, and the sequence of random variables generated and used by both players (including the random coin flips $I_1, \dots, I_t$ and all potential randomization utilized by the online learner) up until the end of round t. In this protocol, an online learning algorithm $\Pi_n = \{P_t\}_{t=1}^n$ is a sequence of functions such that P_t maps the sequence of past coin flips $(i_1, \dots, i_{t-1}) \in \{-1, 1\}^{t-1}$ *and data sets* $\widehat{s}_n \in \mathcal{Z}^{2n}$ to $\Delta_{\mathcal{W}}$, the set of all probability distributions over the hypothesis class $\mathcal{W}$. For the brevity of notation, we abbreviate $P_t = P_t(I_1, \dots, I_{t-1}, \widehat{S}_n)$, and we denote by $\widehat{\mathcal{P}}_n$ the class of all online learning algorithms over sequences of length n in this protocol.

In the conditional generalization game defined above, our technique gives an upper bound on the *empirical generalization error* defined as
$$\widehat{\text{gen}}(W_n, \widehat{S}_n) = \frac{1}{n}\sum_{t=1}^n \left(\ell(W_n, Z_t') - \ell(W_n, Z_t)\right).$$

In order to bound the generalization error, we need to control the gap between the two quantities, denoted as

$$\Delta(W_n, \widehat{S}_n) = \mathrm{gen}(W_n, S_n) - \widehat{\mathrm{gen}}(W_n, \widehat{S}_n) = \frac{1}{n}\sum_{t=1}^{n} \left(\mathbb{E}\left[\ell(W_n, Z')|\, W_n\right] - \ell(W_n, Z_t')\right).$$

It is important to note that W_n only depends on $S_n = (Z_1, \dots, Z_n)$ and therefore, conditioned on S_n, the quantity $\Delta(W_n, \widehat{S}_n)$ is the difference between the empirical mean of nonnegative i.i.d. random variables and its expectation. Hence, the lower tail of $\Delta(W_n, \widehat{S}_n)$ can be controlled using standard tools. As before, the total cost accumulated by the online learner is denoted by $M_{\Pi_n} = \frac{1}{n}\sum_{t=1}^{n}\langle P_t, c_t\rangle$. Importantly, this quantity is a normalized sum of martingale differences, when conditioned on $\widehat{S}_n$, due to the conditional independence of P_t and c_t:

$$\begin{aligned}\mathbb{E}\big[\langle P_t, c_t\rangle \,\big|\mathcal{F}_{t-1}, \widehat{S}_n\big] &= \mathbb{E}_{\widetilde{W}_t \sim P_t}\big[c_t\big(\widetilde{W}_t\big)\big|\mathcal{F}_{t-1}, \widehat{S}_n\big] \\ &= \mathbb{E}_{\widetilde{W}_t \sim P_t}\left[\ell\big(\widetilde{W}_t, Z_t^{I_t}\big) - \frac{1}{2}\Big(\ell\big(\widetilde{W}_t, Z_t^{+1}\big) + \ell\big(\widetilde{W}_t, Z_t^{-1}\big)\Big)\Big|\mathcal{F}_{t-1}, \widehat{S}_n\right] = 0\,.\end{aligned}$$

Indeed, the online learning protocol guarantees that $\widetilde{W}_t$ is chosen before the index I_t is revealed to the online learner, and thus the conditional expectation of $\ell\big(\widetilde{W}_t, Z_t^{I_t}\big)$ is $\frac{1}{2}\big(\ell\big(\widetilde{W}_t, Z_t^{+1}\big) + \ell\big(\widetilde{W}_t, Z_t^{-1}\big)\big)$. Note that M_{Π_n} is *not* a martingale average without conditioning on $\widehat{S}_n$, as each W_t may depend on the entire data set, including future instances $Z_s^{\pm 1}$ with $s > t$.

The following result connects the generalization error to the regret in a similar way as Theorem 1 does for the basic generalization game:

Theorem 17 *The generalization error of any learning algorithm $W_n = \mathcal{A}(S_n)$ satisfies that, for any online learning algorithm $\Pi_n \in \widehat{\mathcal{P}}_n$,*

$$\overline{\mathrm{gen}}(W_n, S_n) = \frac{\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})}{n} - M_{\Pi_n} + \mathbb{E}\left[\Delta(W_n, \widehat{S}_n)\Big|\, \widehat{S}_n\right]\,.$$

In particular, for any class $\mathcal{C}_n \subset \widehat{\mathcal{P}}_n$ of online learning algorithms,

$$\overline{\mathrm{gen}}(W_n, S_n) = \inf_{\Pi_n \in \mathcal{C}_n}\left(\frac{\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})}{n} - M_{\Pi_n}\right) + \mathbb{E}\left[\Delta(W_n, \widehat{S}_n)\Big|\, \widehat{S}_n\right]\,.$$

Proof For simplicity, we denote the training points as $Z_t = Z_t^{I_t}$ and the test points as $Z_t' = Z_t^{-I_t}$. Recalling the notation $\langle P, f\rangle = \mathbb{E}_{W\sim P}\left[f(W)\right]$, we rewrite the conditional expectation of the empirical generalization error as follows:

$$\begin{aligned}\mathbb{E}\left[\widehat{\mathrm{gen}}(W_n, \widehat{S}_n)\Big|\, \widehat{S}_n\right] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\ell(W_n, Z_t) - \ell(W_n, Z_t')|\, \widehat{S}_n\right] = -\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[c_t(W_n)|\, \widehat{S}_n\right] \\ &= -\frac{1}{n}\sum_{t=1}^{n}\left\langle P_{W_n|\widehat{S}_n}, c_t\right\rangle = -\frac{1}{n}\sum_{t=1}^{n}\left\langle P_{W_n|S_n}, c_t\right\rangle \\ &= \frac{1}{n}\sum_{t=1}^{n}\left\langle P_t - P_{W_n|S_n}, c_t\right\rangle - \frac{1}{n}\sum_{t=1}^{n}\langle P_t, c_t\rangle \\ &= \frac{\mathrm{regret}_n(P_{W_n|S_n})}{n} - \frac{1}{n}\sum_{t=1}^{n}\langle P_t, c_t\rangle\,,\end{aligned}$$

where in the second line we observed that $P_{W_n|\widehat{S}_n} = P_{W_n|S_n}$ holds due to the construction of W_n that only depends on the training data S_n. Recalling the definitions of M_{Π_n} and $\Delta(W_n, \widehat{S}_n)$ completes the proof. ■

The following corollary instantiates the bound for a single algorithm under the assumption that the loss function is bounded in $[0, 1]$.

Corollary 18 *Consider an arbitrary online learning algorithm $\Pi_n \in \widehat{\mathcal{P}}_n$ and suppose that $\ell(w,z) \in [0,1]$ for all w, z. Then, with probability at least $1-\delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\overline{\mathrm{gen}}(W_n, S_n) \leq \frac{\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})}{n} + \sqrt{\frac{2\log\left(\frac{1}{\delta}\right)}{n}}\,.$$

The bound follows from applying the Hoeffding–Azuma inequality for the martingale averages M_{Π_n} and for $\mathbb{E}\left[\Delta(W_n, \widehat{S}_n)\middle|\widehat{S}_n\right]$. In principle, the latter term can be bounded under the milder condition that the second moment of the loss function is bounded for all w (due to it being a lower tail of a nonnegative martingale), but boundedness of the loss function is needed now to ensure that the martingale increments constituting M_{Π_n} are bounded almost surely. Indeed, notice that M_{Π_n} is a martingale *only when conditioned on the entire supersample* $\widehat{S}_n$, and thus it needs to be controlled for all realizations of the data sequence.

The advantage of the conditional online-to-PAC conversion framework is that it allows the online learner to have prior knowledge of the supersample $\widehat{S}_n$, which includes the training data. In particular, the online learner can now use data-dependent learning rates and regularization functions. To illustrate the use of this framework, we provide a simple application that leads to a conditional PAC-Bayesian generalization bound. In particular, using the standard multiplicative weights algorithm for the online learner leads to the following generalization bound:

Corollary 19 *Suppose that $\ell(w,z) \in [0,1]$ for all w, z and let $\widehat{S}_n$ be a set of $2n$ i.i.d. data points drawn from μ. Then, for any $\eta > 0$ and any $P_1 \in \Delta_{\mathcal{W}}$ that potentially depend on $\widehat{S}_n$, and any $n > 1$, with probability at least $1-\delta$, the generalization error of every statistical learning algorithm $W_n = \mathcal{A}(S_n)$ simultaneously satisfies*

$$\mathbb{E}\left[\mathrm{gen}(W_n, S_n)\right| S_n] \leq \frac{\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n}\middle\|P_1\right)}{\eta n} + \frac{\eta}{8} + \sqrt{\frac{2\log(1/\delta)}{n}}\,.$$

The power of this result resides in the fact that the learning rate η and the prior P_1 are allowed to depend on the supersample $\widehat{S}_n$. Such priors are called *almost exchangeable* by Audibert (2004), who proved an analogous result using classic PAC-Bayesian methodology. With a special choice of prior, and relaxing the high-probability bound above to only hold on expectation, this result also recovers the conditional information-theoretic bound of Steinke and Zakynthinou (2020) (see also Grünwald et al., 2021).

More generally, we can obtain similar conditional versions of *all* generalization bounds derived in earlier sections of this work, including the data-dependent bounds of Section 3.1.3 and the parameter-free bound of Section 3.1.2. Furthermore, the conditional generalization game allows the online learner to use a conditional version of FTRL, where η and h can both depend on the supersample $\widehat{S}_n$, for instance by setting h as a convex divergence measure between P and a data-dependent prior P_1.

4.2. Bounds on the expected generalization error

Besides the high-confidence guarantees provided in the rest of this paper, it is straightforward to derive bounds on the expected generalization error using our framework. Such relaxations of PAC-Bayesian guarantees have been extensively studied in the last few years under the moniker "information-theoretic generalization bounds" (Russo and Zou, 2016, 2019; Xu and Raginsky, 2017). Here, we derive a generalized version of the bounds proposed in these works using our online-to-PAC conversion scheme. In particular, the following guarantee can be deduced directly from Theorem 1:

Corollary 20 *Consider any statistical learning algorithm $W_n = \mathcal{A}(S_n)$ and a class $\mathcal{C}_n \subset \mathcal{P}_n$ of online learning algorithms. Then, the generalization error of $\mathcal{A}$ satisfies*

$$\mathbb{E}\left[\mathrm{gen}(W_n, S_n)\right] = \inf_{\Pi_n \in \mathcal{C}_n} \frac{\mathbb{E}\left[\mathrm{regret}_{\Pi_n}(P_{W_n|S_n})\right]}{n}.$$

While this bound is decidedly weaker than its previously discussed versions, it has some merits that make it worthy of consideration. Most notably, since the right-hand side features an infimum over *deterministic* quantities, one can directly obtain generalization bounds without requiring any sophisticated techniques to control suprema of empirical processes. We illustrate this benefit below, as well as explain the connections with the related literature.

As a simple application of Corollary 20, the following analogue of Corollary 4 can be easily derived by choosing the multiplicative weights algorithm as the online learning method:

Corollary 21 *Suppose that $\ell(w,z) \in [0,1]$ for all w, z. Then, for any fixed $\eta > 0$, any $P_1 \in \Delta_{\mathcal{W}}$ and any $n > 1$, the expected generalization error of any statistical learning algorithm $W_n = \mathcal{A}(S_n)$ satisfies the bound*

$$\mathbb{E}\left[\operatorname{gen}(W_n, S_n)\right] \leq \frac{\mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)\right]}{\eta n} + \frac{\eta}{8}\,.$$

In particular, letting P_{W_n} denote the marginal distribution of W_n, we have

$$\mathbb{E}\left[\operatorname{gen}(W_n, S_n)\right] \leq \inf_{P_1 \in \Delta_{\mathcal{W}}} \sqrt{\frac{\mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)\right]}{2n}} = \sqrt{\frac{\mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_{W_n}\right)\right]}{2n}}\,.$$

The divergence $\mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_{W_n}\right)\right]$ appearing in the second bound is the mutual information between W_n and S_n, and thus this result recovers the bounds of Xu and Raginsky (2017) that are stated in terms of the same quantity. Our result can be verified directly after making the following simple observations. First, notice that the value of η minimizing the first bound for a given P_1 is non-random, and thus we can avoid the covering argument required for the proof of Corollary 4. Second, by the variational characterization of the mutual information, we have $\inf_{P_1 \in \Delta_{\mathcal{W}}} \mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_1\right)\right] = \mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}\left(P_{W_n|S_n} \middle\| P_{W_n}\right)\right]$ (Kemperman, 1974).

All other guarantees stated in earlier sections can be adjusted analogously. Most importantly, all results in the preliminary version of this work (Lugosi and Neu, 2022) can be exactly recovered by adapting the results in Section 3.2, using FTRL as the online learning algorithm. "Conditional" analogues to the same guarantees can be derived via the construction proposed in Section 4.1, and in particular the "conditional mutual-information" bounds of Steinke and Zakynthinou (2020) can be recovered via the same argument as we used above for Corollary 21.

5. Conclusion

Our new online-to-PAC conversion scheme establishes a link between online and statistical learning that provides a flexible framework for proving generalization bounds using regret analysis. In the present paper, we provide a short list of applications of this technique to derive generalization bounds, recovering several state-of-the-art results and improving them in several minor ways. These results are most likely only scratching the surface of what this framework is able to achieve, and in fact we feel that we have opened more questions in this work than what we have managed to answer. We discuss some of the numerous exciting directions for future work below.

In recent years, several connections have emerged between regret analysis in online learning, generalization bounds, and concentration inequalities. Early forerunners of such results are Zhang (2002) and Kakade et al. (2008) who respectively used online learning techniques to bound covering numbers and Rademacher complexities of linear function classes, both well-studied proxies of the generalization error in statistical learning. Some years later, a sequence of works by Rakhlin and Sridharan (2017) and (Foster et al., 2015, 2017, 2018) established a deep connection between uniform convergence of collections of martingales and the existence of online algorithms with bounded regret, effectively showing that any worst-case regret bound for a class of online learning games can be turned into a martingale concentration inequality and vice versa. Our result can be seen as a variant of the above results that is more tightly adapted to analyzing the generalization error of statistical learning algorithms, our key idea being a more specific choice of comparator point in the definition of regret. A potentially closely related line of work proposes to derive concentration bounds on the means of random variables using sequential betting strategies (Waudby-Smith and Ramdas, 2020; Orabona and Jun, 2021). This can be seen as a reduction to another online learning problem with the logarithmic loss function (cf. Chapter 9 of Cesa-Bianchi and Lugosi, 2006). This is to be contrasted with the linear loss functions used in our work and all of the previously mentioned ones, and we wonder if a closer connection could be made between these approaches.

Our techniques are quite different from those that have been traditionally used for proving generalization bounds. Instead of combinatorial arguments used for studying suprema of empirical processes, our results make use of regret analysis, which itself traditionally builds on tools from convex analysis and optimization. This strikes us as an entirely new approach to study this fundamental problem of statistical learning, and also as an unexpected new application of convex analysis and optimization that may open interesting research directions in both of these fields. Indeed, much of the online learning and convex optimization literature is focused on questions of computational efficiency that are entirely absent in our setup: since we only need to prove the *existence* of online learning algorithms with appropriate regret guarantees, we never have to worry about implementation issues. We believe that this aspect can open up a new and interesting direction of research not only within statistical learning theory, but more broadly in convex optimization.

Our analysis framework appears to be flexible enough to go beyond the standard statistical learning framework that assumes i.i.d. data. The fact that the key part of our bounds are controlled almost surely via regret analysis is encouraging in that it suggests that at least some probabilistic assumptions can be dropped, but this still leaves us with designing appropriate notions of generalization for non-i.i.d. data. It seems straightforward (and natural) to generalize our results to stationary data sequences by adjusting the definition of the test error $\mathbb{E}\left[\ell(w, Z')\right]$ to involve an expectation taken with respect to the stationary distribution of Z'_t. We wonder if it is possible to go more significantly beyond the standard model by dropping even more probabilistic assumptions on the data sequence, and adapt our framework to deal with problems of "out-of-distribution" generalization.

Finally, we note that we expect that our framework will be able to capture several more concepts used in the statistical learning theory literature to explain generalization. Such ideas include stability (Devroye and Wagner, 1979; Bousquet and Elisseeff, 2002; Mukherjee et al., 2006; Shalev-Shwartz et al., 2010; Hardt et al., 2016), differential privacy (Dwork et al., 2006a,b; Chaudhuri et al., 2011; Bassily et al., 2014), or various margin and noise conditions (Bartlett et al., 2002; Bartlett and Mendelson, 2006; van Erven et al., 2015). We leave such extensions open for future work.

Acknowledgments

G. Lugosi acknowledges the support of Ayudas Fundación BBVA a Proyectos de Investigación Científica 2021 and the Spanish Ministry of Economy and Competitiveness, Grant PGC2018-101643-B-I00 and FEDER, EU. G. Neu was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 950180). Part of this work was done while the second author was visiting the Simons Institute for the Theory of Computing. The authors wish to thank the numerous colleagues who helped shape the results presented in this paper either via comments, discussions, or pointers to related literature. In no particular order, we thank Tamás Linder, Csaba Szepesvári, Peter Grünwald, Peter Bartlett, Wojciech Kotłowski, Borja Rodríguez Gálvez, Gautam Goel, Ishaq Aden-Ali, Abhishek Shetty, Nikita Zhivotovskiy, Ilja Kuzborskij, Francesco Orabona, Matus Telgarsky, Ohad Shamir, Roi Livni, and Adam Block.

References

P. Alquier and B. Guedj. Simpler PAC-Bayesian bounds for hostile data. *Machine Learning*, 107(5):887–902, 2018.

J.-Y. Audibert. *PAC-Bayesian statistical learning theory*. PhD thesis, Université Paris VI, 2004.

K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones mathematicae*, 115(1):463–482, 1994.

P. Bartlett and S. Mendelson. Rademacher and Gaussian complexities: risk bounds and structural results. *Journal of Machine Learning Research*, 3:463–482, 2002.

P. Bartlett and S. Mendelson. Empirical minimization. *Probability Theory and Related Fields*, 135:311–334, 2006.

P. Bartlett, S. Boucheron, and G. Lugosi. Model selection and error estimation. *Machine Learning*, 48:85–113, 2002.

P. L. Bartlett. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Transactions on Information Theory*, 44(2):525–536, 1998.

R. Bassily, A. Smith, and A. Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *Foundations of Computer Science (FOCS)*, pages 464–473, 2014.

L. Bégin, P. Germain, F. Laviolette, and J.-F. Roy. PAC-Bayesian bounds based on the Rényi divergence. In *Artificial Intelligence and Statistics*, pages 435–444, 2016.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, 2013.

O. Bousquet and A. Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2:499–526, 2002.

O. Catoni. PAC-Bayesian supervised classification. *Lecture Notes-Monograph Series. IMS*, 1277, 2007.

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

N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2-3):321–352, 2007.

K. Chaudhuri, Y. Freund, and D. J. Hsu. A parameter-free hedging algorithm. pages 297–305.

K. Chaudhuri, C. Monteleoni, and A. D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

A. Chernov and V. Vovk. Prediction with advice of unknown number of experts. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 117–125, 2010.

J. A. Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, 1936.

L. Devroye and T. J. Wagner. Distribution-free inequalities for the deleted and holdout error estimates. *IEEE Transactions on Information Theory*, 25(2):202–207, 1979.

R. M. Dudley. A course on empirical processes. In *Ecole d'été de Probabilités de Saint-Flour XII-1982*, pages 1–142. Springer, 1984.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 486–503, 2006a.

C. Dwork, F. McSherry, K. Nissim, and A. D. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference (TCC)*, pages 265–284, 2006b.

D. J. Foster, A. Rakhlin, and K. Sridharan. Adaptive online learning. *Advances in Neural Information Processing Systems*, 28, 2015.

D. J. Foster, A. Rakhlin, and K. Sridharan. Zigzag: A new approach to adaptive online learning. In *Conference on Learning Theory*, pages 876–924, 2017.

D. J. Foster, A. Rakhlin, and K. Sridharan. Online learning: Sufficient statistics and the Burkholder method. In *Conference On Learning Theory*, pages 3028–3064, 2018.

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.

P. Gaillard, G. Stoltz, and T. van Erven. A second-order bound with excess losses. In *Conference on Learning Theory*, pages 176–196, 2014.

B. V. Gnedenko and A. N. Kolmogorov. Limit distributions for sums of independent random variables. *Am. J. Math*, 105, 1954.

P. Grünwald, T. Steinke, and L. Zakynthinou. PAC-Bayes, MAC-Bayes and conditional mutual information: Fast rate bounds that handle general VC classes. In *Conference on Learning Theory*, pages 2217–2247, 2021.

M. Haghifam, G. K. Dziugaite, S. Moran, and D. Roy. Towards a unified information-theoretic framework for generalization. *Advances in Neural Information Processing Systems*, 34:26370–26381, 2021.

M. Hardt, B. Recht, and Y. Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *International Conference on Machine Learning (ICML)*, pages 1225–1234, 2016.

F. Hellström and G. Durisi. Generalization bounds via information density and conditional information density. *IEEE Journal on Selected Areas in Information Theory*, 1(3):824–839, 2020.

S. M. Kakade, K. Sridharan, and A. Tewari. On the complexity of linear prediction: Risk bounds, margin bounds, and regularization. *Advances in neural information processing systems*, 21, 2008.

J. Kemperman. On the Shannon capacity of an arbitrary channel. In *Indagationes Mathematicae (Proceedings)*, volume 77, pages 101–115. North-Holland, 1974.

V. Koltchinskii. Rademacher penalties and structural risk minimization. *IEEE Transactions on Information Theory*, 47:1902–1914, 2001.

W. M. Koolen and T. Van Erven. Second-order quantile methods for experts and combinatorial games. In *Conference on Learning Theory*, pages 1155–1175, 2015.

N. Littlestone and M. Warmuth. The weighted majority algorithm. *Information and Computation*, 108: 212–261, 1994.

R. Livni and S. Moran. A limitation of the PAC-Bayes framework. *Advances in Neural Information Processing Systems*, 33:20543–20553, 2020.

G. Lugosi and G. Neu. Generalization bounds via convex analysis. In *Proceedings of the 35th Conference on Learning Theory (COLT)*, pages 3524–3546, 2022.

H. Luo and R. E. Schapire. Achieving all with no parameters: Adanormalhedge. pages 1286–1304, 2015.

D. McAllester. A PAC-Bayesian tutorial with a dropout bound. *arXiv preprint arXiv:1307.2118*, 2013.

D. A. McAllester. Some PAC-Bayesian theorems. In *Proceedings of the eleventh annual conference on Computational Learning Theory (COLT)*, pages 230–234, 1998.

Z. Mhammedi, P. Grünwald, and B. Guedj. PAC-Bayes un-expected Bernstein inequality. *Advances in Neural Information Processing Systems*, 32, 2019.

S. Mukherjee, P. Niyogi, T. Poggio, and R. Rifkin. Learning theory: stability is sufficient for generalization and necessary and sufficient for consistency of empirical risk minimization. *Advances in Computational Mathematics*, 25:161–193, 2006.

G. Neu, G. K. Dziugaite, M. Haghifam, and D. M. Roy. Information-theoretic generalization bounds for stochastic gradient descent. In *Proceedings of the 34th Conference on Learning Theory (COLT)*, pages 3526–3545, 2021.

F. Nielsen and R. Nock. On Rényi and Tsallis entropies and divergences for exponential families. *arXiv preprint arXiv:1105.3259*, 2011.

F. Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

F. Orabona and K.-S. Jun. Tight concentrations and confidence sequences from the regret of universal portfolio. *arXiv preprint arXiv:2110.14099*, 2021.

F. Orabona and D. Pál. Coin betting and parameter-free online learning. In *Advances in Neural Information Processing Systems*, pages 577–585, 2016.

K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

A. Rakhlin. Lecture notes on online learning. 2009.

A. Rakhlin and K. Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019, 2013a.

A. Rakhlin and K. Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013b.

A. Rakhlin and K. Sridharan. On equivalence of martingale tail bounds and deterministic regret inequalities. In *Conference on Learning Theory*, pages 1704–1722. PMLR, 2017.

B. Rodríguez-Gálvez, G. Bassi, R. Thobaben, and M. Skoglund. Tighter expected generalization error bounds via Wasserstein distance. In *Advances in Neural Information Processing Systems*, 2021.

D. Russo and J. Zou. Controlling bias in adaptive data analysis using information theory. In *Artificial Intelligence and Statistics*, pages 1232–1240, 2016.

D. Russo and J. Zou. How much does your data exploration overfit? controlling bias via information usage. *IEEE Transactions on Information Theory*, 66(1):302–323, 2019.

I. Sason and S. Verdú. f-divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, 2016.

Y. Seldin, F. Laviolette, N. Cesa-Bianchi, J. Shawe-Taylor, and P. Auer. PAC-Bayesian inequalities for martingales. *IEEE Transactions on Information Theory*, 58(12):7086–7093, 2012.

G. Shafer and V. Vovk. *Probability and finance: it's only a game!*, volume 491. John Wiley & Sons, 2001.

S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2012.

S. Shalev-Shwartz, O. Shamir, N. Srebro, and K. Sridharan. Learnability, stability and uniform convergence. *Journal of Machine Learning Research*, 11:2635–2670, 2010.

J. Shawe-Taylor and R. C. Williamson. A PAC analysis of a Bayesian estimator. In *Proceedings of the tenth annual conference on Computational learning theory*, pages 2–9, 1997.

T. Steinke and L. Zakynthinou. Reasoning about generalization via conditional mutual information. In *Proceedings of the 33rd Conference on Learning Theory (COLT)*, 2020.

I. O. Tolstikhin and Y. Seldin. PAC-Bayes-empirical-Bernstein inequality. *Advances in Neural Information Processing Systems*, 26, 2013.

D. van der Hoeven, T. van Erven, and W. Kotłowski. The many faces of exponential weights in online learning. In *Conference On Learning Theory*, pages 2067–2092, 2018.

T. van Erven, P. D. Grünwald, N. A. Mehta, M. D. Reid, and R. C. Williamson. Fast rates in statistical and online learning. *Journal of Machine Learning Research*, 16:1793–1861, 2015.

V. Vapnik and A. Chervonenkis. *Theory of Pattern Recognition*. Nauka, Moscow, 1974. (in Russian); German translation: *Theorie der Zeichenerkennung*, Akademie Verlag, Berlin, 1979.

C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2003.

V. Vovk. Aggregating strategies. In *Proceedings of the third annual workshop on Computational learning theory (COLT)*, pages 371–386, 1990.

H. Wang, M. Diaz, J. C. S. Santos Filho, and F. P. Calmon. An information-theoretic view of generalization via Wasserstein distance. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 577–581. IEEE, 2019.

I. Waudby-Smith and A. Ramdas. Estimating means of bounded random variables by betting. *arXiv preprint arXiv:2010.09686*, 2020.

R. C. Williamson, A. J. Smola, and B. Schölkopf. Entropy numbers of linear function classes. In *Proceedings of the 13th Conference on Learning Theory (COLT)*, pages 309–319, 2000.

A. Xu and M. Raginsky. Information-theoretic analysis of generalization capability of learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2524–2533, 2017.

A. C. Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science*, pages 222–227, 1977.

J. Zhang, T. Liu, and D. Tao. An optimal transport view on generalization. *arXiv preprint arXiv:1811.03270*, 2018.

T. Zhang. Covering number bounds of certain regularized linear function classes. *Journal of Machine Learning Research*, 2(Mar):527–550, 2002.

Appendix A. Regret bounds

A.1. Exponentially weighted averages

Theorem 22 *Consider the exponentially weighted average forecaster defined via the iteration*

$$P_{t+1} = \operatorname*{arg\,min}_{P \in \Delta_{\mathcal{W}}} \left\{ \langle P, c_t \rangle - \frac{1}{\eta} \mathcal{D}_{\mathrm{KL}} \left(P \| P_t \right) \right\} .$$

For any sequence of cost functions $c_1, c_2, \ldots, c_n$, the regret of this method satisfies

$$\mathrm{regret}_n(P^*) \leq \frac{\mathcal{D}_{\mathrm{KL}}\left(P^* \| P_1\right)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\|c_t\|_\infty^2 .$$

Proof The proof is based on studying a potential function Φ defined for all $c \in \mathbb{R}^{\mathcal{W}}$ as

$$\Phi(c) = \frac{1}{\eta}\log\int_{\mathcal{W}} e^{-\eta c(w)} \mathrm{d}P_1(w).$$

In particular, we consider $\Phi(\sum_{t=1}^n c_t)$ and notice that it is related to the total cost of the comparator P^* as follows:

$$\Phi\left(\sum_{t=1}^{n} c_t\right) = \frac{1}{\eta}\log\int_{\mathcal{W}} e^{-\eta\sum_{t=1}^n c_t(W)}\mathrm{d}P_1(w) \geq -\sum_{t=1}^{n}\left\langle P^*, c_t\right\rangle - \frac{1}{\eta}\mathcal{D}_{\mathrm{KL}}\left(P^* \| P_1\right),$$

where the inequality is the Donsker–Varadhan variational formula (cf. Section 4.9 in Boucheron et al., 2013). On the other hand, we have

$$\begin{aligned}
\Phi\left(\sum_{t=1}^{n} c_t\right) &= \sum_{t=1}^{n}\left(\Phi\left(\sum_{k=1}^{t} c_k\right) - \Phi\left(\sum_{k=1}^{t-1} c_k\right)\right) \\
&= \sum_{t=1}^{n}\frac{1}{\eta}\log\frac{\int_{\mathcal{W}} e^{-\eta\sum_{k=1}^{t} c_k(w)}\mathrm{d}P_1(w)}{\int_{\mathcal{W}} e^{-\eta\sum_{k=1}^{t-1} c_t(w)}\mathrm{d}P_1(w)} \\
&= \sum_{t=1}^{n}\frac{1}{\eta}\log\int_{\mathcal{W}}\frac{e^{-\eta\sum_{k=1}^{t-1} c_k(w)}}{\int_{\mathcal{W}} e^{-\eta\sum_{k=1}^{t-1} c_k(w)}\mathrm{d}P_1(w)}\cdot e^{-\eta c_t(w)}\mathrm{d}P_1(w) \\
&= \sum_{t=1}^{n}\frac{1}{\eta}\log\int_{\mathcal{W}} e^{-\eta c_t(w)}\mathrm{d}P_t(w)
\end{aligned}$$

Finally, we notice that the term appearing on the right-hand side can be bounded using Hoeffding's lemma (see, e.g., Lemma 2.2 in Boucheron et al., 2013):

$$\frac{1}{\eta}\log\int_{\mathcal{W}} e^{-\eta c_t(w)}\mathrm{d}P_t(w) = -\left\langle P_t, c_t\right\rangle + \frac{1}{\eta}\int_{\mathcal{W}} e^{-\eta(c_t(w)-\langle P_t, c_t\rangle)}\mathrm{d}P_t(w) \leq -\left\langle P_t, c_t\right\rangle + \frac{\eta\left\|c_t\right\|_\infty^2}{2}. \quad (3)$$

Plugging this inequality back into the previous calculations concludes the proof. ∎

A.2. Optimistic Second-Order EWA

Let us now consider an "optimistic" version of a EWA-based method that uses a guess g_t of c_t when playing its action P_t. This algorithm calculates two sequences of updates: first, an auxiliary distribution $\widetilde{P}_t$ as

$$\frac{\mathrm{d}\widetilde{P}_{t+1}}{\mathrm{d}\widetilde{P}_t}(w) = \frac{e^{-\eta c_t(w)-\eta^2(c_t(w)-g_t(w))^2}}{\int_{\mathcal{W}} e^{-\eta c_t(w')-\eta^2(c_t(w)-g_t(w))^2}\mathrm{d}P_t(w')}‘,$$

and the actual update calculated as

$$\frac{\mathrm{d}P_{t+1}}{\mathrm{d}\widetilde{P}_{t+1}}(w) = \frac{e^{-\eta g_{t+1}(w)}}{\int_{\mathcal{W}} e^{-\eta g_{t+1}(w')}\mathrm{d}P_t(w')}.$$

The second-order adjustment appearing in the auxiliary update sequence is commonly used in the online learning literature to achieve data-dependent bounds (Cesa-Bianchi et al., 2007; Gaillard et al., 2014; Koolen and Van Erven, 2015), and is the component that enables us to prove a strong comparator-dependent regret bound. The idea of using an auxiliary update sequence is inspired by the "optimistic online learning" algorithms of Rakhlin and Sridharan (2013a,b), and allows us to achieve fully data-dependent bounds by eliminating the test error from the generalization bounds. The algorithm defined above satisfies the following regret bound.

Theorem 23 *For any $\eta \in \left[0, \frac{1}{2}\right]$ and any sequence of cost functions $c_1, c_2, \dots, c_n$ and predictions $g_1, g_2, \dots, g_n$, the regret of the optimistic second-order EWA forecaster defined above satisfies*

$$\sum_{t=1}^{n} \langle P_t - P^*, c_t \rangle \leq \frac{\mathcal{D}_{\mathrm{KL}}\left(P^* \| P_1\right)}{\eta} + \eta \sum_{t=1}^{n} \left\langle P^*, (c_t - g_t)^2 \right\rangle.$$

We are going to use this theorem with $c_t(w) = \ell(w, Z_t) - \mathbb{E}\left[\ell(w, Z')\right]$ and $g_t(w) = -\mathbb{E}\left[\ell(w, Z')\right] \leq 0$ so the last term on the right-hand side is negative, which will help us deduce a fast rate from the regret bound.

Proof The proof follows from similar arguments as used in the proof of Theorem 22. In particular, we introduce the auxiliary notation $\widetilde{c}_t(w) = c_t(w) - \eta(c_t(w) - g_t(w))^2$ and study the potential $\Phi(\sum_{t=1}^{n} \widetilde{c}_t)$. Lower-bounding the potential gives

$$\Phi\left(\sum_{t=1}^{n} \widetilde{c}_t\right) = \frac{1}{\eta} \log \int_{\mathcal{W}} e^{-\eta \sum_{t=1}^{n} \widetilde{c}_t(W)} \mathrm{d}\widetilde{P}_1(w) \geq -\sum_{t=1}^{n} \langle P^*, \widetilde{c}_t \rangle - \frac{1}{\eta} \mathcal{D}_{\mathrm{KL}}\left(P^* \middle\| \widetilde{P}_1\right).$$

On the other hand, we have

$$\Phi\left(\sum_{t=1}^{n} \widetilde{c}_t\right) = \sum_{t=1}^{n} \frac{1}{\eta} \log \int_{\mathcal{W}} e^{-\eta \widetilde{c}_t(w)} \mathrm{d}\widetilde{P}_t(w).$$

Noticing that $\frac{\mathrm{d}\widetilde{P}_t}{\mathrm{d}P_t}(w) = \frac{e^{\eta g_t(w)}}{\int_{\mathcal{W}} e^{\eta g_t(w')} \mathrm{d}P_t(w')}$, we can upper bound each term in the above sum as

$$\begin{aligned}
\frac{1}{\eta} \log \int_{\mathcal{W}} e^{-\eta \widetilde{c}_t(w)} \mathrm{d}\widetilde{P}_t(w) &= \frac{1}{\eta} \log \int_{\mathcal{W}} e^{-\eta c_t(w) - \eta^2 (c_t(w) - g_t(w))^2} \frac{\mathrm{d}\widetilde{P}_t}{\mathrm{d}P_t}(w) \mathrm{d}P_t(w) \\
&= \frac{1}{\eta} \log \int_{\mathcal{W}} e^{-\eta (c_t(w) - g_t(w)) - \eta^2 (c_t(w) - g_t(w))^2} \mathrm{d}P_t(w) - \frac{1}{\eta} \log \int_{\mathcal{W}} e^{\eta g_t(w)} \mathrm{d}P_t(w) \\
&\leq \frac{1}{\eta} \log \int_{\mathcal{W}} \left(1 - \eta\left(c_t(w) - g_t(w)\right)\right) \mathrm{d}P_t(w) - \langle P_t, g_t \rangle \\
&\leq - \langle P_t, c_t \rangle,
\end{aligned}$$

where we have used the inequality $e^{-x-x^2} \leq 1 - x$ that holds for all $x \leq \frac{1}{2}$ (which is ensured by our choice of $\eta \leq \frac{1}{2}$ and the boundedness of the cost function), and Jensen's inequality that implies $\langle P_t, g_t \rangle \leq \frac{1}{\eta} \log \int_{\mathcal{W}} e^{\eta g_t(w)} \mathrm{d}P_t(w)$. Putting the two bounds together proves the statement. ■

A.3. Follow the Regularized Leader

We recall that the predictions of the FTRL algorithm are defined as

$$P_t = \operatorname*{arg\,min}_{P \in \Delta_{\mathcal{W}}} \left\{ \left\langle P, \sum_{k=1}^{t-1} c_k \right\rangle + \frac{1}{\eta} h(P) \right\}.$$

For simplicity, we use the notation $C_t = \sum_{k=1}^{t} c_k$ hereafter. We first show that, under the conditions we have assumed in the main text (properness and strong convexity), the minimum exists and is unique.

For simplicity, we use the notation $\Psi_t = \langle \cdot, C_{t-1} \rangle + \frac{1}{\eta} h$. We denote the effective domain of h by $\Gamma_h = \{P \in \Delta_{\mathcal{W}} : h(P) < +\infty\}$ and note that the condition that h is proper and lower semicontinuous implies that Ψ_t is also proper and lower semicontinuous. Together with the compactness of $\Delta_{\mathcal{W}}$, this implies the existence of a minimum. In order to show unicity, let us suppose that P_t and P'_t are both minimizers of Ψ_t. Then, by convexity, we have for all $\lambda \in [0,1]$ that $\lambda P_t + (1-\lambda)P'_t \in \arg\min_P \Psi_t(P)$. However, by strong convexity of h, we know that Ψ_t is also strongly convex and thus

$$\begin{aligned}\min_{P \in \Delta_{\mathcal{W}}} \Psi_t(P) = \Psi_t(\lambda P_t + (1-\lambda)P'_t) &\leq \lambda \Psi_t(P_t) + (1-\lambda)\Psi_t(P'_t) - \frac{\alpha\lambda(1-\lambda)}{2} \|P_t - P'_t\|^2 \\ &= \min_{P \in \Delta_{\mathcal{W}}} \Psi_t(P) - \frac{\alpha\lambda(1-\lambda)}{2} \|P_t - P'_t\|^2 .\end{aligned}$$

Hence, equality is only possible when $\|P_t - P'_t\| = 0$, or, equivalently, $P_t = P'_t$. This shows that the minimum is indeed unique.

The analysis below uses a few other concepts from convex analysis. A key notion is the Legendre–Fenchel conjugate of the convex function h denoted as $h^* : \mathbb{R}^{\mathcal{W}} \to \mathbb{R}$, mapping a function f to $h^*(f) = \max_{P \in \Delta_{\mathcal{W}}} \{\langle P, f \rangle - h(P)\}$. The subdifferential of a convex functional $U : \mathbb{R}^{\mathcal{W}} \to \mathbb{R}$ at $f \in \mathbb{R}^{\mathcal{W}}$ is defined as the set of signed measures

$$\partial U(f) = \left\{P \in \overline{\Delta}_{\mathcal{W}} : U(g) \geq U(f) + \langle P, g - f \rangle \quad (\forall g \in \mathbb{R}^{\mathcal{W}})\right\},$$

and the associated (generalized) Bregman divergence is defined as

$$\mathcal{D}_U\left(g \| f\right) = U(g) - U(f) + \sup_{P \in \partial U(f)} \langle P, f - g \rangle ,$$

where the supremum is introduced to resolve the ambiguity of the subdifferential. Notice that this is a convex function of g, being a sum of a convex function and a supremum of affine functions, and that $\mathcal{B}_\Phi\left(g \| f\right) \geq 0$ for all f and g due to convexity of U.

Having established these basic facts, we are now ready to state a regret bound for the above algorithm.

Theorem 24 *Suppose that h is α-strongly convex with respect to the norm $\|\cdot\|$. Then, for any $\eta > 0$ and any sequence of cost functions $c_1, c_2, \dots, c_n$, the regret of the FTRL algorithm defined above satisfies*

$$\operatorname{regret}_n(P^*) \leq \frac{h\left(P^*\right) - h\left(P_1\right)}{\eta} + \frac{\eta}{2\alpha} \sum_{t=1}^n \|c_t\|_*^2 .$$

Proof The proof is based on studying a potential function Φ defined for all $c \in \mathbb{R}^{\mathcal{W}}$ as

$$\Phi(c) = \max_{P \in \Delta_{\mathcal{W}}} \left\{-\langle P, c \rangle - \frac{1}{\eta} h(P)\right\} = \frac{1}{\eta} h^*(-\eta c) .$$

In particular, we consider $\Phi(C_n)$ and notice that it is related to the total cost of the comparator P^* as follows:

$$\Phi\left(C_n\right) = \max_{P \in \Delta_{\mathcal{W}}} \left\{-\langle P, C_n \rangle - \frac{1}{\eta} h(P)\right\} \geq -\sum_{t=1}^n \langle P^*, c_t \rangle - \frac{h(P^*)}{\eta} .$$

On the other hand, we have

$$\begin{aligned}\Phi\left(C_n\right) &= \sum_{t=1}^n \left(\Phi\left(C_t\right) - \Phi\left(C_{t-1}\right)\right) + \Phi(0) \\ &= \sum_{t=1}^n \left(\mathcal{B}_\Phi\left(C_t \| C_{t-1}\right) - \sup_{P \in \partial\Phi(C_{t-1})} \langle P, C_{t-1} - C_t \rangle\right) - \frac{h(P_1)}{\eta} \\ &\leq \sum_{t=1}^n \Big(\mathcal{B}_\Phi\left(C_t \| C_{t-1}\right) - \langle P_t, c_t \rangle\Big) - \frac{h(P_1)}{\eta},\end{aligned}$$

where in the last step we have used the fact that $-P_t \in \partial\Phi(C_{t-1})$. Indeed, this follows from the definition of the algorithm:

$$\begin{aligned}\Phi(C_{t-1}) - \langle P_t, c_t\rangle &= \max_{P\in\Delta_{\mathcal{W}}}\left\{-\langle P, C_{t-1}\rangle - \frac{1}{\eta}h(P)\right\} - \langle P_t, c_t\rangle = -\langle P_t, C_t\rangle - \frac{1}{\eta}h(P_t)\\ &\le \max_{P\in\Delta_{\mathcal{W}}}\left\{-\langle P, C_t\rangle - \frac{1}{\eta}h(P)\right\} = \Phi(C_t)\,.\end{aligned}$$

Putting the above inequalities together, we obtain the following bound on the regret:

$$\sum_{t=1}^{n}\langle P^* - P_t, c_t\rangle \le \frac{h(P^*) - h(P_1)}{\eta} + \sum_{t=1}^{n}\mathcal{B}_{\Phi}\left(C_t \| C_{t-1}\right).$$

Finally, we note that Φ is the Legendre–Fenchel conjugate of $P \mapsto \frac{1}{\eta}h(\eta P)$, which is an η/α strongly convex function of its argument. Thus, we use a classic duality property between the regularizer h and h^* (proved in Appendix B.2 for completeness) to show that

$$\mathcal{B}_{\Phi}\left(C_t \| C_{t-1}\right) \le \frac{\eta\left\|c_t\right\|_*^2}{2\alpha}.$$

This completes the proof. ■

We also consider an "optimistic" version of FTRL that makes use of a sequence of hints $g_t \in \mathbb{R}^{\mathcal{W}}$, by choosing its updates according to the assignment

$$P_t = \operatorname*{arg\,min}_{P\in\Delta_{\mathcal{W}}}\left\{\left\langle P, g_t + \sum_{k=1}^{t-1} c_k\right\rangle + \frac{1}{\eta}h(P)\right\}\,.$$

A similar method has been proposed and analyzed by Rakhlin and Sridharan (2013a,b). The following performance guarantee is easily obtained by a series of simple adjustments to the proof of Theorem 24 presented above.

Theorem 25 *Suppose that h is α-strongly convex with respect to the norm $\|\cdot\|$. Then, for any $\eta > 0$ and any sequence of cost functions $c_1, c_2, \dots, c_n$ and predictions $g_1, g_2, \dots, g_n$, the regret of the optimistic FTRL algorithm defined above satisfies*

$$\operatorname{regret}_n(P^*) \le \frac{h\left(P^*\right) - h\left(P_1\right)}{\eta} + \frac{\eta}{2\alpha}\sum_{t=1}^{n}\left\|c_t - g_t\right\|_*^2\,.$$

Proof The proof is based on studying a potential function Φ defined for all $c \in \mathbb{R}^{\mathcal{W}}$ as

$$\Phi(c) = \max_{P\in\Delta_{\mathcal{W}}}\left\{-\langle P, c\rangle - \frac{1}{\eta}h(P)\right\} = \frac{1}{\eta}h^*(-\eta c)\,.$$

In particular, we consider $\Phi(C_n)$ and notice that it is related to the total cost of the comparator P^* as follows:

$$\Phi\left(C_n\right) = \max_{P\in\Delta_{\mathcal{W}}}\left\{-\langle P, C_n\rangle - \frac{1}{\eta}h(P)\right\} \ge -\sum_{t=1}^{n}\langle P^*, c_t\rangle - \frac{h(P^*)}{\eta}\,.$$

On the other hand, we have

$$\Phi\left(C_n\right) = \sum_{t=1}^{n}\left(\Phi\left(C_t\right) - \Phi\left(C_{t-1} + g_t\right) + \Phi\left(C_{t-1} + g_t\right) - \Phi\left(C_{t-1}\right)\right) + \Phi(0). \tag{4}$$

We bound the two key terms arising in the above expression as follows, recalling that $-P_t \in \partial\Phi(C_{t-1} + g_t)$. First, we have

$$\begin{aligned}\Phi(C_t) - \Phi(C_{t-1} + g_t) &\leq \mathcal{B}_\Phi(C_t \| C_{t-1} + g_t) - \sup_{P \in \partial\Phi(C_{t-1}+g_t)} \langle P, C_{t-1} - C_t + g_t \rangle \\ &\leq \mathcal{B}_\Phi(C_t \| C_{t-1} + g_t) - \langle P_t, c_t - g_t \rangle \ .\end{aligned}$$

The remaining term is treated by exploiting the convexity of Φ that guarantees

$$\Phi(C_{t-1}) \geq \Phi(C_{t-1} + g_t) - \sup_{P \in \partial\Phi(C_{t-1}+g_t)} \langle P, g_t \rangle \geq \Phi(C_{t-1} + g_t) + \langle P_t, g_t \rangle$$

by the definition of the subdifferential $\partial\Phi(C_{t-1} + g_t)$. Putting these inequalities together with Equation (4), we obtain

$$\Phi(C_n) \leq \sum_{t=1}^{n} \Big(\mathcal{B}_\Phi(C_t \| C_{t-1} + g_t) - \langle P_t, c_t \rangle \Big) - \frac{h(P_1)}{\eta} \ ,$$

The proof is concluded by putting everything together and using the smoothness property of Φ implied by the strong convexity of h (cf. Appendix B.2):

$$\mathcal{B}_\Phi(C_t \| C_{t-1} + g_t) \leq \frac{\eta \left\| c_t - g_t \right\|_*^2}{2\alpha} \ .$$

■

Appendix B. Technical tools

B.1. Martingale concentration inequalities

Here we provide a simple concentration inequality to control the lower tails of sums of nonnegative random variables with bounded second moments, used several times in the proofs.

Lemma 26 *Let $(X_t)_{t=1}^n$ be a sequence non-negative random variables and for $t \geq 0$, let $\mathcal{F}_t$ denote the σ-algebra generated by $X_1, \dots, X_t$. Assume that X_t has finite conditional mean $\mu_t = \mathbb{E}[X_t | \mathcal{F}_{t-1}]$ and second moment $\sigma_t^2 = \mathbb{E}\left[X_t^2 \middle| \mathcal{F}_{t-1}\right]$. Then, for any $\lambda > 0$, the following bound holds with probability at least $1 - \delta$:*

$$\sum_{t=1}^{n} (\mu_t - X_t) \leq \frac{\lambda}{2} \sum_{t=1}^{n} \sigma_t^2 + \frac{\log \frac{1}{\delta}}{\lambda}.$$

Proof We use the notation $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_{t-1}]$. We start by noticing that, for any $\lambda > 0$, we have

$$\mathbb{E}_t\left[e^{-\lambda X_t}\right] \leq \mathbb{E}_t\left[1 - \lambda X_t + \frac{\lambda^2 X_t^2}{2}\right] \leq e^{-\lambda\mu_t + \lambda^2\sigma_t^2/2},$$

where we have used the inequality $e^y \leq 1 + y + \frac{y^2}{2}$ that holds for all $y \leq 0$. Using this repeatedly gives

$$\mathbb{E}\left[e^{\lambda \sum_{t=1}^n \left(\mu_t - X_t - \lambda\sigma_t^2/2\right)}\right] \leq 1,$$

and thus an application of Markov's inequality yields

$$\mathbb{P}\left[\sum_{t=1}^{n} \left(\mu_t - X_t - \lambda\sigma_t^2/2\right) \geq \varepsilon\right] = \mathbb{P}\left[e^{\lambda \sum_{t=1}^n \left(\mu_t - X_t - \lambda\sigma_t^2/2\right)} \geq e^{\lambda\varepsilon}\right] \leq e^{-\lambda\varepsilon}.$$

Thus, setting $\varepsilon = \log(1/\delta)/\lambda$ and reordering the terms proves the statement. ■

The following simple bound provides an empirical variant of the above bound that holds for bounded random variables.

Lemma 27 *Let $(X_t)_{t=1}^n$ be a sequence random variables supported on $[0,1]$ and for $t \geq 0$, let $\mathcal{F}_t$ denote the σ-algebra generated by $X_1, \dots, X_t$. Assume that X_t has finite conditional mean $\mu_t = \mathbb{E}\left[X_t \middle| \mathcal{F}_{t-1}\right]$. Then, for any $\lambda \in \left[0, \frac{1}{2}\right]$, the following bound holds with probability at least $1-\delta$:*

$$\sum_{t=1}^n (\mu_t - X_t) \leq \lambda \sum_{t=1}^n X_t^2 + \frac{\log \frac{1}{\delta}}{\lambda}.$$

Proof We use the notation $\mathbb{E}_t\left[\cdot\right] = \mathbb{E}\left[\cdot|\mathcal{F}_{t-1}\right]$. We start by noticing that, for any $\lambda > 0$, we have

$$\mathbb{E}_t\left[e^{-\lambda X_t - \lambda^2 X_t^2}\right] \leq \mathbb{E}_t\left[1 - \lambda X_t\right] \leq e^{-\lambda \mu_t},$$

where we have used the inequality $e^{-y-y^2} \leq 1 - y$ that holds for all $y \leq \frac{1}{2}$, which is ensured by the conditions on X_t and λ. Using this repeatedly gives

$$\mathbb{E}\left[e^{\lambda \sum_{t=1}^n \left(\mu_t - X_t - \lambda X_t^2\right)}\right] \leq 1,$$

and thus an application of Markov's inequality yields

$$\mathbb{P}\left[\sum_{t=1}^n \left(\mu_t - X_t - \lambda X_t^2\right) \geq \varepsilon\right] = \mathbb{P}\left[e^{\lambda \sum_{t=1}^n \left(\mu_t - X_t - \lambda X_t^2\right)} \geq e^{\lambda\varepsilon}\right] \leq e^{-\lambda\varepsilon}.$$

Thus, setting $\varepsilon = \log(1/\delta)/\lambda$ and reordering the terms proves the statement. ■

Finally, we also supply the following simple extension that applies to heavy-tailed random variables with bounded q-th moments.

Lemma 28 *Let $(X_t)_{t=1}^n$ be a sequence non-negative random variables with finite conditional mean $\mu_t = \mathbb{E}\left[X_t \middle| \mathcal{F}_{t-1}\right]$ and q-th moment $B_t = \left(\mathbb{E}\left[X_t^q \middle| \mathcal{F}_{t-1}\right]\right)^{1/q}$ for some $q \in (1,2]$. Then, for any $\lambda > 0$, the following bound holds with probability at least $1-\delta$:*

$$\sum_{t=1}^n (\mu_t - X_t) \leq \lambda^{q-1} \sum_{t=1}^n B_t^q + \frac{\log \frac{1}{\delta}}{\lambda}.$$

Proof We start by noticing that, for any $\lambda > 0$, we have

$$\mathbb{E}_t\left[e^{-\lambda X_t}\right] \leq \mathbb{E}_t\left[1 - \lambda X_t + \lambda^q X_t^q\right] \leq e^{-\lambda\mu_t + \lambda^q B_t^q},$$

where we have used the inequality $e^y \leq 1 + y + y^q$ that holds for all $y \leq 0$. Using this repeatedly gives

$$\mathbb{E}\left[e^{\lambda \sum_{t=1}^n (\mu_t - X_t - \lambda^q B_t^q)}\right] \leq 1,$$

and thus an application of Markov's inequality yields

$$\mathbb{P}\left[\sum_{t=1}^n \left(\mu_t - X_t - \lambda^{q-1} B_t^q\right) \geq \varepsilon\right] = \mathbb{P}\left[e^{\lambda \sum_{t=1}^n \left(\mu_t - X_t - \lambda^{q-1} B_t^q\right)} \geq e^{\lambda\varepsilon}\right] \leq e^{-\lambda\varepsilon}.$$

Thus, setting $\varepsilon = \log(1/\delta)/\lambda$ and reordering the terms proves the statement. ■

B.2. Strong-convexity / smoothness duality

Lemma 29 *Let $f, f' \in \mathbb{R}^{\mathcal{W}}$ and let $P \in \partial h^*(f)$ and $P' \in \partial h^*(f')$. Suppose that h is α-strongly convex with respect to $\|\cdot\|$. Then, Φ satisfies*

$$\mathcal{B}_{\Phi}\left(f \| g\right) \leq \frac{1}{\alpha} \left\|f - f'\right\|_*^2 .$$

Proof Let $s_P \in \partial h(P)$ and $s_{P'} \in \partial h(P')$. Then, by first-order optimality of P and P', we have

$$\begin{aligned} \langle s_P - f, P - P' \rangle &\leq 0 \\ \langle s_{P'} - f', P' - P \rangle &\leq 0. \end{aligned}$$

Summing the two inequalities, we get

$$\langle s_{P'} - s_P, P - P' \rangle \leq \langle P' - P, f' - f \rangle .$$

Now, using the strong convexity of h, we get

$$\begin{aligned} h(P) &\geq h(P') + \langle s_{P'}, P - P' \rangle + \frac{\alpha}{2} \left\|P - P'\right\|^2 \\ h(P') &\geq h(P) + \langle s_P, P' - P \rangle + \frac{\alpha}{2} \left\|P - P'\right\|^2 . \end{aligned}$$

Summing these two inequalities then gives

$$\alpha \left\|P - P'\right\|^2 \leq \langle s_P - s_{P'}, P - P' \rangle .$$

Combining both inequalities above, we obtain

$$\alpha \left\|P' - P\right\|^2 \leq \langle P - P', f - f' \rangle \leq \left\|P - P'\right\| \left\|f - f'\right\|_* ,$$

and reordering yields

$$\left\|P - P'\right\| \leq \frac{1}{\alpha} \left\|f - f'\right\|_* . \tag{5}$$

By the mean value theorem, there exists an $f_\lambda = \lambda f + (1 - \lambda) f'$ with $\lambda \in [0, 1]$ such that $P_\lambda \in \partial \Phi(f_\lambda)$ and

$$\begin{aligned} h^*(f) &= h^*(f') + \langle P_\lambda, f - f' \rangle \\ &= h^*(f') + \langle P', f - f' \rangle + \langle P_\lambda - P', f - f' \rangle \\ &\leq \Phi(f') + \langle P', f - f' \rangle + \frac{1}{\alpha} \left\|f_\lambda - f'\right\|_* \left\|f - f'\right\|_* \qquad \text{(by Equation (5))} \\ &= \Phi(f') + \langle P', f - f' \rangle + \frac{\lambda}{\alpha} \left\|f - f'\right\|_*^2 . \\ &\leq \Phi(f') + \langle P', f - f' \rangle + \frac{1}{\alpha} \left\|f - f'\right\|_*^2 . \end{aligned}$$

The proof is completed by recalling that $P' \in \partial h^*(f')$ and the definition of the Bregman divergence, and reordering the terms. ■

B.3. The proof of Lemma 14

For clarity, we start by formalizing the notion of directional derivatives of f via the following recursive definition. We let B_1 denote the Euclidean unit ball $B_1 = \left\{v \in \mathbb{R}^d : \|v\|_2 = 1\right\}$, and define $D^0 f = f$ and for each subsequent $j > 0$, we define $D^j f : \mathcal{W} \times B_1^j$ as

$$D^j f(w | v_1, v_2, \ldots, v_j) = \lim_{c \to 0} \frac{D^{j-1} f(w + c v_j | v_1, v_2, \ldots, v_{j-1}) - D^{j-1} f(w | v_1, v_2, \ldots, v_{j-1})}{c} .$$

Notice that D^j is linear in f.

The proof of the lemma itself is based on the following successive smoothing argument: we begin by smoothing the original function f using the conjugate of the smoothing operator G_σ^*, then smoothing out the residual $f - G_\sigma^* f$ and continue indefinitely. As we show, the residuals decay rapidly at a rate determined by the higher-order derivatives of the original function f. To make this argument precise, we let $f_0 = f$ and recursively define $f_{j+1} = f_j - G_\sigma^* f_j$, so that we can write

$$\begin{aligned}
\langle P - P', f \rangle &= \langle P - P', G_\sigma^* f \rangle + \langle P - P', f - G_\sigma^* f \rangle = \langle G_\sigma (P - P'), f_0 \rangle + \langle P - P', f_1 \rangle \\
&= \langle G_\sigma (P - P'), f_0 \rangle + \langle P - P', G_\sigma^* f_1 \rangle + \langle P - P', f_1 - G_\sigma^* f_1 \rangle \\
&= \langle G_\sigma (P - P'), f_0 \rangle + \langle G_\sigma (P - P'), f_1 \rangle + \langle P - P', f_2 \rangle + \dots \\
&= \sum_{j=0}^{\infty} \langle G_\sigma (P - P'), f_j \rangle \le \|P - P'\|_\sigma \sum_{j=0}^{\infty} \|f_j\|_\infty ,
\end{aligned}$$

where the last step follows from Hölder's inequality. This shows that the dual norm is indeed upper bounded as follows:

$$\|f\|_{\sigma,*} = \sup_{\|P-P'\|_\sigma \le 1} \langle P - P', f \rangle \le \sum_{j=0}^{\infty} \|f_j\|_\infty .$$

It remains to relate $\|f_j\|_\infty$ to the derivatives of the original function f. To this end, let ξ denote a Gaussian vector distributed as $\mathcal{N}(0, \sigma^2 I)$, and note that for all j, we have

$$\begin{aligned}
\|f_j\|_\infty &= \sup_w \left| f_{j-1}(w) - \mathbb{E}\left[f_{j-1}(w + \xi) \right] \right| \\
&\le \sup_w \mathbb{E}\left[\|\xi\|_2 \cdot \left| \frac{f_{j-1}(w) - f_{j-1}(w + \xi)}{\|\xi\|_2} \right| \right] \\
&\le \mathbb{E}\left[\|\xi\|_2 \right] \sup_w \sup_{v_1 \in B_1} \left| D^1 f_{j-1}(w|v_1) \right| \\
&\le \left(\sigma\sqrt{d} \right) \sup_w \sup_{v_1 \in B_1} \left| \mathbb{E}\left[D^1 f_{j-2}(w|v_1) - D^1 f_{j-2}(w + \xi|v_1) \right] \right| \\
&\le \left(\sigma\sqrt{d} \right) \sup_w \sup_{v_1 \in B_1} \mathbb{E}\left[\|\xi\|_2 \cdot \left| \frac{D^1 f_{j-2}(w|v_1) - D^1 f_{j-2}(w + \xi|v_1)}{\|\xi\|_2} \right| \right] \\
&\le \left(\sigma\sqrt{d} \right) \mathbb{E}\left[\|\xi\|_2 \right] \sup_w \sup_{v_1, v_2 \in B_1} \left| D^2 f_{j-2}(w|v_1, v_2) \right| \\
&\le \dots \le \left(\sigma\sqrt{d} \right)^j \sup_w \sup_{v_1, v_2, \dots, v_j \in B_1} \left| D^j f(w|v_1, v_2, \dots, v_j) \right| \le \left(\sigma\sqrt{d} \right)^j \beta_j .
\end{aligned}$$

Here, we have used the bound $\mathbb{E}\left[\|\xi\|_2 \right] \le \sigma\sqrt{d}$ several times. Putting this together with the previous bound proves the claim. ■

B.4. Wasserstein distance and smoothed relative entropy

This section provides some results supporting the claims made in Section 3.2.4. We first give a precise definition for the Wasserstein distance between two distributions $P, P' \in \Delta_{\mathcal{W}}$. For the sake of concreteness, we only give the defintion for the distance metric given by the Euclidean distance on $\mathbb{R}^d$, and refer the reader to the book of Villani (2003) for a more general treatment. Letting $\Pi(P, P')$ denote the set of joint distributions on $\mathcal{W} \times \mathcal{W}$ with marginals P and P', the squared Wasserstein-2 distance between P and P' is defined as

$$\mathbb{W}_2(P, P') = \inf_{\pi \in \Pi(P,P')} \int_{\mathcal{W} \times \mathcal{W}} \|w - w'\|_2^2 \, \mathrm{d}\pi(w, w').$$

The following lemma (whose proof is largely based on the proof of Lemma 4 of Neu et al., 2021) provides a bound on the smoothed relative entropy in terms of the squared Wasserstein-2 distance:

Lemma 30 *Let W and W' be two random variables on $\mathbb{R}^d$ with respective laws P and P'. For any $\sigma > 0$, the smoothed relative entropy between P and P' is bounded as*

$$\mathcal{D}_\gamma\left(P \| P'\right) \leq \frac{1}{2\sigma^2}\mathbb{E}\left[\left\|W - W'\right\|_2^2\right].$$

Proof Let us consider a fixed coupling $\pi \in \Pi(P, P')$ and observe that the smoothed distributions $G_\sigma P$ and $G_\sigma P'$ can be respectively written as

$$G_\sigma P = \int_{\mathcal{W}\times\mathcal{W}} \mathcal{N}(w, \sigma^2 I)\mathrm{d}\pi(w, w') \quad \text{and} \quad G_\sigma P' = \int_{\mathcal{W}\times\mathcal{W}} \mathcal{N}(w', \sigma^2 I)\mathrm{d}\pi(w, w').$$

Using this observation, we can write

$$\begin{aligned}
\mathcal{D}_\gamma\left(P \| P'\right) &= \mathcal{D}\left(\int_{\mathcal{W}\times\mathcal{W}} \mathcal{N}(w, \sigma^2 I)\mathrm{d}\pi(w, w') \middle\| \int_{\mathcal{W}\times\mathcal{W}} \mathcal{N}(w', \sigma^2 I)\mathrm{d}\pi(w, w')\right) \\
&\leq \int_{\mathcal{W}\times\mathcal{W}} \mathcal{D}\left(\mathcal{N}(w, \sigma^2 I) \middle\| \mathcal{N}(w', \sigma^2 I)\right)\mathrm{d}\pi(w, w') \\
&= \frac{1}{2\sigma^2}\int_{\mathcal{W}\times\mathcal{W}} \left\|W - W'\right\|_2^2 \mathrm{d}\pi(w, w'),
\end{aligned}$$

where the second line uses Jensen's inequality and the joint convexity of $\mathcal{D}\left(\cdot\|\cdot\right)$ in its arguments, and the last line follows from noticing that $\mathcal{D}\left(\mathcal{N}(x, \Sigma)\|\mathcal{N}(y, \Sigma)\right) = \frac{1}{2}\left\|x - y\right\|_{\Sigma^{-1}}^2$ for any x, y and any symmetric positive definite covariance matrix Σ. The result then follows from taking the infimum with respect to π on the right-hand side. ■