Efficient Online Learning via Randomized Rounding

Nicolò Cesa-Bianchi
DSI, Università degli Studi di Milano
Italy
nicolo.cesa-bianchi@unimi.it

Ohad Shamir
Microsoft Research New England
USA
ohadsh@microsoft.com

Abstract

Most online algorithms used in machine learning today are based on variants of mirror descent or follow-the-leader. In this paper, we present an online algorithm based on a completely different approach, which combines "random playout" and randomized rounding of loss subgradients. As an application of our approach, we provide the first computationally efficient online algorithm for collaborative filtering with trace-norm constrained matrices. As a second application, we solve an open question linking batch learning and transductive online learning.

1 Introduction

Online learning algorithms, which have received much attention in recent years, enjoy an attractive combination of computational efficiency, lack of distributional assumptions, and strong theoretical guarantees. However, it is probably fair to say that at their core, most of these algorithms are based on the same small set of fundamental techniques, in particular mirror descent and regularized follow-the-leader (see for instance [14]).

In this work we revisit, and significantly extend, an algorithm which uses a completely different approach. This algorithm, known as the *Minimax Forecaster*, was introduced in [9, 11] for the setting of prediction with static experts. It computes minimax predictions in the case of known horizon, binary outcomes, and absolute loss. Although the original version is computationally expensive, it can easily be made efficient through randomization.

We extend the analysis of [9] to the case of non-binary outcomes and arbitrary convex and Lipschitz loss functions. The new algorithm is based on a combination of "random playout" and randomized rounding, which assigns random binary labels to future unseen instances, in a way depending on the loss subgradients. Our resulting *Randomized Rounding (R^2) Forecaster* has a parameter trading off regret performance and computational complexity, and runs in polynomial time (for T predictions, it requires computing $\mathcal{O}(T^2)$ empirical risk minimizers in general, as opposed to $\mathcal{O}(T)$ for generic follow-the-leader algorithms). The regret of the R^2 Forecaster is determined by the Rademacher complexity of the comparison class. The connection between online learnability and Rademacher complexity has also been explored in [2, 1]. However, these works focus on the information-theoretically achievable regret, as opposed to computationally efficient algorithms. The idea of "random playout", in the context of online learning, has also been used in [16, 3], but we apply this idea in a different way.

We show that the R^2 Forecaster can be used to design the first efficient online learning algorithm for collaborative filtering with trace-norm constrained matrices. While this is a well-known setting, a straightforward application of standard online learning approaches, such as mirror descent, appear to give only trivial performance guarantees. Moreover, our regret bound matches the best currently known sample complexity bound in the batch distribution-free setting [21].

As a different application, we consider the relationship between batch learning and transductive online learning. This relationship was analyzed in [16], in the context of binary prediction with respect to classes of bounded VC dimension. Their main result was that efficient learning in a statistical setting implies efficient learning in the transductive online setting, but at an inferior rate of $T^{3/4}$ (where T is the number of rounds). The main open question posed by that paper is whether a better rate can be obtained. Using the R^2 Forecaster, we improve on those results, and provide an efficient algorithm with the optimal $\sqrt{T}$ rate, for a wide class of losses. This shows that efficient batch learning not only implies efficient transductive online learning (the main thesis of [16]), but also that the same rates can be obtained, and for possibly non-binary prediction problems as well.

We emphasize that the R^2 Forecaster requires computing many empirical risk minimizers (ERM's) at each round, which might be prohibitive in practice. Thus, while it does run in polynomial time whenever an ERM can be efficiently computed, we make no claim that it is a "fully practical" algorithm. Nevertheless, it seems to be a useful tool in showing that *efficient* online learnability is possible in various settings, often working in cases where more standard techniques appear to fail. Moreover, we hope the techniques we employ might prove useful in deriving practical online algorithms in other contexts.

2 The Minimax Forecaster

We start by introducing the sequential game of prediction with expert advice —see [10]. The game is played between a forecaster and an adversary, and is specified by an outcome space $\mathcal{Y}$, a prediction space $\mathcal{P}$, a nonnegative loss function $\ell : \mathcal{P} \times \mathcal{Y} \to \mathbb{R}$, which measures the discrepancy between the forecaster's prediction and the outcome, and an expert class $\mathcal{F}$. Here we focus on classes $\mathcal{F}$ of *static experts*, whose prediction at each round t does not depend on the outcome in previous rounds. Therefore, we think of each $\mathbf{f} \in \mathcal{F}$ simply as a sequence $\mathbf{f} = (f_1, f_2, \dots)$ where each $f_t \in \mathcal{P}$. At each step $t = 1, 2, \dots$ of the game, the forecaster outputs a prediction $p_t \in \mathcal{P}$ and simultaneously the adversary reveals an outcome $y_t \in \mathcal{Y}$. The forecaster's goal is to predict the outcome sequence almost as well as the best expert in the class $\mathcal{F}$, irrespective of the outcome sequence $\mathbf{y} = (y_1, y_2, \dots)$. The performance of a forecasting strategy A is measured by the worst-case regret

$$\mathcal{V}_T(A, \mathcal{F}) = \sup_{\mathbf{y} \in \mathcal{Y}^T} \left(\sum_{t=1}^{T} \ell(p_t, y_t) - \inf_{\mathbf{f} \in \mathcal{F}} \sum_{t=1}^{T} \ell(f_t, y_t) \right) \tag{1}$$

viewed as a function of the horizon T.

Consider now the special case where the horizon T is fixed and known in advance, the outcome space is $\mathcal{Y} = \{-1, +1\}$, the prediction space is $\mathcal{P} = [-1, +1]$, and the loss is the absolute loss $\ell(p, y) = |p - y|$. To simplify notation, let $L(\mathbf{f}, \mathbf{y}) = \sum_{t=1}^{T} |f_t - y_t|$. We will denote the regret in this special case as $\mathcal{V}_T^{\text{abs}}(A, \mathcal{F})$.

The Minimax Forecaster —which is based on work presented in [9] and [11], see also [10] for an exposition— is derived by an explicit analysis of the minimax regret $\inf_A \mathcal{V}_T^{\text{abs}}(A, \mathcal{F})$, where the infimum is over all forecasters A producing at round t a prediction p_t as a function of $p_1, y_1, \dots p_{t-1}, y_{t-1}$. For general online learning problems, the analysis of this quantity is intractable. However, for the specific setting we focus on (absolute loss and binary outcomes), one can get both an explicit expression for the minimax regret, as well as an explicit algorithm, provided $\inf_{\mathbf{f} \in \mathcal{F}} \sum_{t=1}^{T} \ell(f_t, y_t)$ can be efficiently computed for any sequence $y_1, \dots, y_T$. This procedure is akin to performing empirical risk minimization (ERM) in statistical learning. A full development of the analysis is out of scope, but is outlined in Appendix A of the supplementary material. In a nutshell, the idea is to begin by calculating the optimal prediction in the last round T, and then work backwards, calculating the optimal prediction at round $T-1$, $T-2$ etc. Remarkably, the value of $\inf_A \mathcal{V}_T^{\text{abs}}(A, \mathcal{F})$ is *exactly* the Rademacher complexity $\mathcal{R}_T(\mathcal{F})$ of the class $\mathcal{F}$, which is known to play a crucial role in understanding the sample complexity in statistical learning [5]. In this paper, we define it as[1]:

$$\mathcal{R}_T(\mathcal{F}) = \mathbb{E}\left[\sup_{\mathbf{f}\in\mathcal{F}} \sum_{t=1}^{T} \sigma_t f_t\right] \tag{2}$$

where $\sigma_1, \ldots, \sigma_T$ are i.i.d. Rademacher random variables, taking values $-1, +1$ with equal probability. When $\mathcal{R}_T(\mathcal{F}) = o(T)$, we get a minimax regret $\inf_A \mathcal{V}_T^{\text{abs}}(A, \mathcal{F}) = o(T)$ which implies a vanishing per-round regret.

In terms of an explicit algorithm, the optimal prediction p_t at round t is given by a complicated-looking recursive expression, involving exponentially many terms. Indeed, for general online learning problems, this is the most one seems able to hope for. However, an apparently little-known fact is that when one deals with a class $\mathcal{F}$ of fixed binary sequences as discussed above, then one can write the optimal prediction p_t in a much simpler way. Letting $Y_1, \ldots, Y_T$ be i.i.d. Rademacher random variables, the optimal prediction at round t can be written as

$$p_t \;=\; \mathbb{E}\left[\inf_{\mathbf{f}\in\mathcal{F}} L\left(\mathbf{f}, y_1 \cdots y_{t-1}\,(-1)\,Y_{t+1} \cdots Y_T\right) - \inf_{\mathbf{f}\in\mathcal{F}} L\left(\mathbf{f}, y_1 \cdots y_{t-1}\,1\,Y_{t+1} \cdots Y_T\right)\right]. \tag{3}$$

In words, the prediction is simply the expected difference between the minimal cumulative loss over $\mathcal{F}$, when the adversary plays -1 at round t and random values afterwards, and the minimal cumulative loss over $\mathcal{F}$, when the adversary plays $+1$ at round t, and the same random values afterwards. We refer the reader to Appendix A of the supplementary material for how this is derived. We denote this optimal strategy (for absolute loss and binary outcomes) as the Minimax Forecaster (MF):

Algorithm 1 Minimax Forecaster (MF)

```
for $t = 1$ to $T$ do
  Predict $p_t$ as defined in Eq. (3)
  Receive outcome $y_t$ and suffer loss $|p_t - y_t|$
end for
```

The relevant guarantee for MF is summarized in the following theorem.

Theorem 1. *For any class $\mathcal{F} \subseteq [-1, +1]^T$ of static experts, the regret of the Minimax Forecaster (Algorithm 1) satisfies $\mathcal{V}_T^{\text{abs}}(\text{MF}, \mathcal{F}) = \mathcal{R}_T(\mathcal{F})$.*

2.1 Making the Minimax Forecaster Efficient

The Minimax Forecaster described above is not computationally efficient, as the computation of p_t requires averaging over exponentially many ERM's. However, by a martingale argument, it is not hard to show that it is in fact sufficient to compute only two ERM's per round.

Algorithm 2 Minimax Forecaster with efficient implementation (MF*)

```
for $t = 1$ to $T$ do
  For $i = t+1, \ldots, T$, let $Y_i$ be a Rademacher random variable
  Let $p_t := \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, y_1 \ldots y_{t-1}\,(-1)\,Y_{t+1} \ldots Y_T) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, y_1 \ldots y_{t-1}\,1\,Y_{t+1} \ldots Y_T)$
  Predict $p_t$, receive outcome $y_t$ and suffer loss $|p_t - y_t|$
end for
```

Theorem 2. *For any class $\mathcal{F} \subseteq [-1, +1]^T$ of static experts, the regret of the randomized forecasting strategy MF* (Algorithm 2) satisfies*

$$\mathcal{V}_T^{\text{abs}}(\text{MF}^*, \mathcal{F}) \le \mathcal{R}_T(\mathcal{F}) + \sqrt{2T\ln(1/\delta)}$$

[1] In the statistical learning literature, it is more common to scale this quantity by $1/T$, but the form we use here is more convenient for stating cumulative regret bounds.

with probability at least $1-\delta$. Moreover, if the predictions $\mathbf{p}=(p_1,\ldots,p_T)$ are computed reusing the random values $Y_1,\ldots,Y_T$ computed at the first iteration of the algorithm, rather than drawing fresh values at each iteration, then it holds that

$$\mathbb{E}\left[L(\mathbf{p},\mathbf{y})-\inf_{\mathbf{f}\in\mathcal{F}}L(\mathbf{f},\mathbf{y})\right]\le\mathcal{R}_T(\mathcal{F})\qquad \textit{for all } \mathbf{y}\in\{-1,+1\}^T.$$

Proof sketch. To prove the second statement, note that $\bigl|\mathbb{E}[p_t]-y_t\bigr|=\mathbb{E}\bigl[|p_t-y_t|\bigr]$ for any fixed $y_t\in\{-1,+1\}$ and p_t bounded in $[-1,+1]$, and use Thm. 1. To prove the first statement, note that $|p_t-y_t|-\bigl|\mathbb{E}_{p_t}[p_t]-y_t\bigr|$ for $t=1,\ldots,T$ is a martingale difference sequence with respect to $p_1,\ldots,p_T$, and apply Azuma's inequality. □

The second statement in the theorem bounds the regret only in expectation and is thus weaker than the first one. On the other hand, it might have algorithmic benefits. Indeed, if we reuse the same values for $Y_1,\ldots,Y_T$, then the computation of the infima over $\mathbf{f}$ in MF* are with respect to an outcome sequence which changes only at one point in each round. Depending on the specific learning problem, it might be easier to re-compute the infimum after changing a single point in the outcome sequence, as opposed to computing the infimum over a different outcome sequence in each round.

3 The R^2 Forecaster

The Minimax Forecaster presented above is very specific to the absolute loss $\ell(f,y)=|f-y|$ and for binary outcomes $\mathcal{Y}=\{-1,+1\}$, which limits its applicability. We note that extending the forecaster to other losses or different outcome spaces is not trivial: indeed, the recursive unwinding of the minimax regret term, leading to an explicit expression and an explicit algorithm, does not work as-is for other cases. Nevertheless, we will now show how one can deal with general (convex, Lipschitz) loss functions and outcomes belonging to any real interval $[-b,b]$.

The algorithm we propose essentially uses the Minimax Forecaster as a subroutine, by feeding it with a carefully chosen sequence of binary values z_t, and using predictions f_t which are scaled to lie in the interval $[-1,+1]$. The values of z_t are based on a randomized rounding of values in $[-1,+1]$, which depend in turn on the loss subgradient. Thus, we denote the algorithm as the Randomized Rounding (R^2) Forecaster.

To describe the algorithm, we introduce some notation. For any scalar $f\in[-b,b]$, define $\widetilde{f}=f/b$ to be the scaled versions of f into the range $[-1,+1]$. For vectors $\mathbf{f}$, define $\widetilde{\mathbf{f}}=(1/b)\mathbf{f}$. Also, we let $\partial_{p_t}\ell(p_t,y_t)$ denote any subgradient of the loss function ℓ with respect to the prediction p_t. As before, we define $L(\widetilde{\mathbf{f}},\mathbf{y})=\sum_{t=1}^{T}|\tilde{f}_t-y_t|$. The pseudocode of the R^2 Forecaster is presented as Algorithm 3 below, and its regret guarantee is summarized in Thm. 3. The proof is presented in Appendix B of the supplementary material.

Theorem 3. *Suppose ℓ is convex and ρ-Lipschitz in its first argument. For any $\mathcal{F}\subseteq[-b,b]^T$ the regret of the R^2 Forecaster (Algorithm 3) satisfies*

$$\mathcal{V}_T(R^2,\mathcal{F})\le\rho\,\mathcal{R}_T(\mathcal{F})+\rho\,b\left(\sqrt{\frac{1}{\eta}}+2\right)\sqrt{2T\ln\left(\frac{2T}{\delta}\right)}\tag{4}$$

with probability at least $1-\delta$.

The prediction p_t which the algorithm computes is an empirical approximation to

$$b\,\mathbb{E}_{Y_{t+1},\ldots,Y_T}\left[\inf_{\mathbf{f}\in\mathcal{F}}L\left(\widetilde{\mathbf{f}},z_1\ldots z_{t-1}\,0\,Y_{t+1}\ldots Y_T\right)-\inf_{\mathbf{f}\in\mathcal{F}}L\left(\widetilde{\mathbf{f}},z_1\cdots z_{t-1}\,1\,Y_{t+1}\cdots Y_T\right)\right]$$

by repeatedly drawing independent values to $Y_{t+1},\ldots,Y_T$ and averaging. The accuracy of the approximation is reflected in the precision parameter η. A larger value of η improves the regret bound, but also increases the runtime of the algorithm. Thus, η provides a trade-off

Algorithm 3 The R^2 Forecaster

```
Input: Upper bound $b$ on $|f_t|, |y_t|$ for all $t = 1, \dots, T$ and $\mathbf{f} \in \mathcal{F}$; upper bound $\rho$ on $\sup_{p,y \in [-b,b]} \left|\partial_p \ell(p, y)\right|$; precision parameter $\eta \geq \frac{1}{T}$.
for $t = 1$ to $T$ do
  $p_t := 0$
  for $j = 1$ to $\eta T$ do
    For $i = t, \dots, T$, let $Y_i$ be a Rademacher random variable
    Draw $\Delta := \inf_{\mathbf{f} \in \mathcal{F}} L\left(\widetilde{\mathbf{f}}, z_1 \dots z_{t-1}\,(-1)\,Y_{t+1} \dots Y_T\right) - \inf_{\mathbf{f} \in \mathcal{F}} L\left(\widetilde{\mathbf{f}}, z_1 \dots z_{t-1}\,1\,Y_{t+1} \dots Y_T\right)$
    Let $p_t := p_t + \frac{b}{\eta T} \Delta$
  end for
  Predict $p_t$
  Receive outcome $y_t$ and suffer loss $\ell(p_t, y_t)$
  Let $r_t := \frac{1}{2}\big(1 - \frac{1}{\rho} \partial_{p_t} \ell(p_t, y_t)\big) \in [0, 1]$
  Let $z_t := 1$ with probability $r_t$, and $z_t := -1$ with probability $1 - r_t$
end for
```

between the computational complexity of the algorithm and its regret guarantee. We note that even when η is taken to be a constant fraction, the resulting algorithm still runs in polynomial time $\mathcal{O}(T^2 c)$, where c is the time to compute a single ERM. In subsequent results pertaining to this Forecaster, we will assume that η is taken to be a constant fraction.

We end this section with a remark that plays an important role in what follows.

Remark 1. *The predictions of our forecasting strategies do not depend on the ordering of the predictions of the experts in $\mathcal{F}$. In other words, all the results proven so far also hold in a setting where the elements of $\mathcal{F}$ are functions $f : \{1, \dots, T\} \to \mathcal{P}$, and the adversary has control on the permutation $\pi_1, \dots, \pi_T$ of $\{1, \dots, T\}$ that is used to define the prediction $f(\pi_t)$ of expert f at time t.*[2] *Also, Thm. 1 implies that the value of $\mathcal{V}_T^{\text{abs}}(\mathcal{F})$ remains unchanged irrespective of the permutation chosen by the adversary.*

4 Application 1: Transductive Online Learning

The first application we consider is a rather straightforward one, in the context of transductive online learning [6]. In this model, we have an arbitrary sequence of labeled examples $(x_1, y_1), \dots, (x_T, y_T)$, where only the set $\{x_1, \dots, x_T\}$ of unlabeled instances is known to the learner in advance. At each round t, the learner must provide a prediction p_t for the label of y_t. The true label y_t is then revealed, and the learner incurs a loss $\ell(p_t, y_t)$. The learner's goal is to minimize the transductive online regret $\sum_{t=1}^{T} \big(\ell(p_t, y_t) - \inf_{f \in \mathcal{F}} \ell(f(x_t), y_t)\big)$ with respect to a fixed class of predictors $\mathcal{F}$ of the form $\{x \mapsto f(x)\}$.

The work [16] considers the binary classification case with zero-one loss. Their main result is that if a class $\mathcal{F}$ of binary functions has bounded VC dimension d, and there exists an efficient algorithm to perform empirical risk minimization, then one can construct an efficient randomized algorithm for transductive online learning, whose regret is at most $\mathcal{O}(T^{3/4}\sqrt{d \ln(T)})$ in expectation. The significance of this result is that efficient batch learning (via empirical risk minimization) implies efficient learning in the transductive online setting. This is an important result, as online learning can be computationally harder than batch learning —see, e.g., [8] for an example in the context of Boolean learning.

A major open question posed by [16] was whether one can achieve the optimal rate $\mathcal{O}(\sqrt{dT})$, matching the rate of a batch learning algorithm in the statistical setting. Using the R^2 Forecaster, we can easily achieve the above result, as well as similar results in a strictly more general setting. This shows that efficient batch learning not only implies efficient

[2] Formally, at each step t: (1) the adversary chooses and reveals the next element π_t of the permutation; (2) the forecaster chooses $p_t \in \mathcal{P}$ and simultaneously the adversary chooses $y_t \in \mathcal{Y}$.

transductive online learning (the main thesis of [16]), but also that the same rates can be obtained, and for possibly non-binary prediction problems as well.

Theorem 4. *Suppose we have a computationally efficient algorithm for empirical risk minimization (with respect to the zero-one loss) over a class $\mathcal{F}$ of $\{0,1\}$-valued functions with VC dimension d. Then, in the transductive online model, the efficient randomized forecaster* MF* *achieves an expected regret of $\mathcal{O}(\sqrt{dT})$ with respect to the zero-one loss.*
Moreover, for an arbitrary class $\mathcal{F}$ of $[-b,b]$-valued functions with Rademacher complexity $\mathcal{R}_T(\mathcal{F})$, and any convex ρ-Lipschitz loss function, if there exists a computationally efficient algorithm for empirical risk minimization, then the R^2 Forecaster is computationally efficient and achieves, in the transductive online model, a regret of $\rho\mathcal{R}_T(\mathcal{F})+\mathcal{O}(\rho b\sqrt{T\ln(T/\delta)})$ with probability at least $1-\delta$.

Proof. Since the set $\{x_1,\dots,x_T\}$ of unlabeled examples is known, we reduce the online transductive model to prediction with expert advice in the setting of Remark 1. This is done by mapping each function $f \in \mathcal{F}$ to a function $f:\{1,\dots,T\}\to\mathcal{P}$ by $t\mapsto f(x_t)$, which is equivalent to an expert in the setting of Remarks 1. When $\mathcal{F}$ maps to $\{0,1\}$, and we care about the zero-one loss, we can use the forecaster MF* to compute randomized predictions and apply Thm. 2 to bound the expected transductive online regret with $\mathcal{R}_T(\mathcal{F})$. For a class with VC dimension d, $\mathcal{R}_T(\mathcal{F}) \le \mathcal{O}(\sqrt{dT})$ for some constant $c>0$, using Dudley's chaining method [12], and this concludes the proof of the first part of the theorem. The second part is an immediate corollary of Thm. 3. □

We close this section by contrasting our results for online transductive learning with those of [7] about standard online learning. If $\mathcal{F}$ contains $\{0,1\}$-valued functions, then the optimal regret bound for online learning is order of $\sqrt{d'T}$, where d' is the Littlestone dimension of $\mathcal{F}$. Since the Littlestone dimension of a class is never smaller than its VC dimension, we conclude that online learning is a harder setting than online transductive learning.

5 Application 2: Online Collaborative Filtering

We now turn to discuss the application of our results in the context of collaborative filtering with trace-norm constrained matrices, presenting what is (to the best of our knowledge) the first computationally efficient online algorithms for this problem.

In collaborative filtering, the learning problem is to predict entries of an unknown $m \times n$ matrix based on a subset of its observed entries. A common approach is norm regularization, where we seek a low-norm matrix which matches the observed entries as best as possible. The norm is often taken to be the trace-norm [22, 19, 4], although other norms have also been considered, such as the max-norm [18] and the weighted trace-norm [20, 13].

Previous theoretical treatments of this problem assumed a stochastic setting, where the observed entries are picked according to some underlying distribution (e.g., [23, 21]). However, even when the guarantees are distribution-free, assuming a fixed distribution fails to capture important aspects of collaborative filtering in practice, such as non-stationarity [17]. Thus, an online adversarial setting, where no distributional assumptions whatsoever are required, seems to be particularly well-suited to this problem domain.

In an online setting, at each round t the adversary reveals an index pair (i_t, j_t) and secretely chooses a value y_t for the corresponding matrix entry. After that, the learner selects a prediction p_t for that entry. Then y_t is revealed and the learner suffers a loss $\ell(p_t, y_t)$. Hence, the goal of a learner is to minimize the regret with respect to a fixed class $\mathcal{W}$ of prediction matrices, $\sum_{t=1}^{T}\ell(p_t,y_t) - \inf_{W\in\mathcal{W}}\sum_{t=1}^{T}\ell\big(W_{i_t,j_t},y_t\big)$. Following reality, we will assume that the adversary picks a different entry in each round. When the learner's performance is measured by the regret after all $T=mn$ entries have been predicted, the online collaborative filtering setting reduces to prediction with expert advice as discussed in Remark 1.

As mentioned previously, $\mathcal{W}$ is often taken to be a convex class of matrices with bounded trace-norm. Many convex learning problems, such as linear and kernel-based predictors, as well as matrix-based predictors, can be learned efficiently both in a stochastic and an online setting, using mirror descent or regularized follow-the-leader methods. However, for reasonable choices of $\mathcal{W}$, a straightforward application of these techniques can lead to algorithms with trivial bounds. In particular, in the case of $\mathcal{W}$ consisting of $m \times n$ matrices with trace-norm at most r, standard online regret bounds would scale like $\mathcal{O}\big(r\sqrt{T}\big)$. Since for this norm one typically has $r = \mathcal{O}\big(\sqrt{mn}\big)$, we get a per-round regret guarantee of $\mathcal{O}(\sqrt{mn/T})$. This is a trivial bound, since it becomes "meaningful" (smaller than a constant) only after all $T = mn$ entries have been predicted.

On the other hand, based on general techniques developed in [15] and greatly extended in [1], it can be shown that online learnability *is* information-theoretically possible for such $\mathcal{W}$. However, these techniques do not provide a computationally efficient algorithm. Thus, to the best of our knowledge, there is currently no efficient (polynomial time) online algorithm, which attain non-trivial regret. In this section, we show how to obtain such an algorithm using the R^2 Forecaster.

Consider first the transductive online setting, where the set of indices to be predicted is known in advance, and the adversary may only choose the order and values of the entries. It is readily seen that the R^2 Forecaster can be applied in this setting, using any convex class $\mathcal{W}$ of fixed matrices with bounded entries to compete against, and any convex Lipschitz loss function. To do so, we let $\{i_k, j_k\}_{k=1}^{T}$ be the set of entries, and run the R^2 Forecaster with respect to $\mathcal{F} = \{t \mapsto W_{i_t,j_t} \ : \ W \in \mathcal{W}\}$, which corresponds to a class of experts as discussed in Remark 1.

What is perhaps more surprising is that the R^2 Forecaster can also be applied in a *non-transductive* setting, where the indices to be predicted are not known in advance. Moreover, the Forecaster doesn't even need to know the horizon T in advance. The key idea to achieve this is to utilize the non-asymptotic nature of the learning problem —namely, that the game is played over a finite $m \times n$ matrix, so the time horizon is necessarily bounded.

The algorithm we propose is very simple: we apply the R^2 Forecaster as if we are in a setting with time horizon $T = mn$, which is played over *all* entries of the $m \times n$ matrix. By Remark 1, the R^2 Forecaster does not need to know the order in which these $m \times n$ entries are going to be revealed. Whenever $\mathcal{W}$ is convex and ℓ is a convex function, we can find an ERM in polynomial time by solving a convex problem. Hence, we can implement the R^2 Forecaster efficiently.

To show that this is indeed a viable strategy, we need the following lemma, whose proof is presented in Appendix C of the supplementary material.

Lemma 1. *Consider a (possibly randomized) forecaster A for a class $\mathcal{F}$ whose regret after T steps satisfies $\mathcal{V}_T(A,\mathcal{F}) \le G$ with probability at least $1-\delta > \frac{1}{2}$. Furthermore, suppose the loss function is such that* $\inf_{p' \in \mathcal{P}} \sup_{y \in \mathcal{Y}} \inf_{p \in \mathcal{P}} \big(\ell(p,y) - \ell(p',y)\big) \ge 0$. *Then*

$$\max_{t=1,\ldots,T} \mathcal{V}_t(A,\mathcal{F}) \le G \qquad \textit{with probability at least } 1-\delta.$$

Note that a simple sufficient condition for the assumption on the loss function to hold, is that $\mathcal{P} = \mathcal{Y}$ and $\ell(p,y) \ge \ell(y,y)$ for all $p, y \in \mathcal{P}$.

Using this lemma, the following theorem exemplifies how we can obtain a regret guarantee for our algorithm, in the case of $\mathcal{W}$ consisting of the convex set of matrices with bounded trace-norm and bounded entries. For the sake of clarity, we will consider $n \times n$ matrices.

Theorem 5. *Let ℓ be a loss function which satisfies the conditions of Lemma 1. Also, let $\mathcal{W}$ consist of $n \times n$ matrices with trace-norm at most $r = \mathcal{O}(n)$ and entries at most $b = \mathcal{O}(1)$, suppose we apply the R^2 Forecaster over time horizon n^2 and all entries of the matrix. Then with probability at least $1-\delta$, after T rounds, the algorithm achieves an average per-round regret of at most*

$$\mathcal{O}\left(\frac{n^{3/2} + n\sqrt{\ln(n/\delta)}}{T}\right) \qquad \textit{uniformly over } T = 1,\ldots,n^2.$$

Proof. In our setting, where the adversary chooses a different entry at each round, [21, Theorem 6] implies that for the class $\mathcal{W}'$ of all matrices with trace-norm at most $r = \mathcal{O}(n)$, it holds that $\mathcal{R}_T(\mathcal{W}')/T \leq \mathcal{O}(n^{3/2}/T)$. Therefore, $\mathcal{R}_{n^2}(\mathcal{W}') \leq \mathcal{O}(n^{3/2})$. Since $\mathcal{W} \subseteq \mathcal{W}'$, we get by definition of the Rademacher complexity that $\mathcal{R}_{n^2}(\mathcal{W}) = \mathcal{O}(n^{3/2})$ as well. By Thm. 3, the regret after n^2 rounds is $\mathcal{O}(n^{3/2} + n\sqrt{\ln(n/\delta)})$ with probability at least $1 - \delta$. Applying Lemma 1, we get that the cumulative regret at the end of any round $T = 1, \ldots, n^2$ is at most $\mathcal{O}(n^{3/2} + n\sqrt{\ln(n/\delta)})$, as required. □

This bound becomes non-trivial after $n^{3/2}$ entries are revealed, which is still a vanishing proportion of all n^2 entries. While the regret might seem unusual compared to standard regret bounds (which usually have rates of $1/\sqrt{T}$ for general losses), it is a natural outcome of the non-asymptotic nature of our setting, where T can never be larger than n^2. In fact, this is the same rate one would obtain in a batch setting, where the entries are drawn from an arbitrary distribution. Moreover, an assumption such as boundedness of the entries is required for currently-known guarantees even in a batch setting —see [21] for details.

Acknowledgments

The first author acknowledges partial support by the PASCAL2 NoE under EC grant FP7-216886.

References

[1] K. Sridharan A. Rakhlin and A. Tewari. Online learning: Random averages, combinatorial parameters, and learnability. In *NIPS*, 2010.

[2] J. Abernethy, P. Bartlett, A. Rakhlin, and A. Tewari. Optimal strategies and minimax lower bounds for online convex games. In *COLT*, 2009.

[3] J. Abernethy and M. Warmuth. Repeated games against budgeted adversaries. In *NIPS*, 2010.

[4] F. Bach. Consistency of trace-norm minimization. *Journal of Machine Learning Research*, 9:1019–1048, 2008.

[5] P. Bartlett and S. Mendelson. Rademacher and Gaussian complexities: Risk bounds and structural results. In *COLT*, 2001.

[6] S. Ben-David, E. Kushilevitz, and Y. Mansour. Online learning versus offline learning. *Machine Learning*, 29(1):45–63, 1997.

[7] S. Ben-David, D. Pál, and S. Shalev-Shwartz. Agnostic online learning. In *COLT*, 2009.

[8] A. Blum. Separating distribution-free and mistake-bound learning models over the boolean domain. *SIAM J. Comput.*, 23(5):990–1000, 1994.

[9] N. Cesa-Bianchi, Y. Freund, D. Haussler, D. Helmbold, R. Schapire, and M. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, May 1997.

[10] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games.* Cambridge University Press, 2006.

[11] T. Chung. Approximate methods for sequential decision making using expert advice. In *COLT*, 1994.

[12] R. M. Dudley. *A Course on Empirical Processes, École de Probabilités de St. Flour, 1982*, volume 1097 of *Lecture Notes in Mathematics.* Springer Verlag, 1984.

[13] R. Foygel, R. Salakhutdinov, O. Shamir, and N. Srebro. Learning with the weighted trace-norm under arbitrary sampling distributions. In *NIPS*, 2011.

[14] E. Hazan. The convex optimization approach to regret minimization. In S. Nowozin S. Sra and S. Wright, editors, *Optimization for Machine Learning.* MIT Press, To Appear.

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

[16] S. Kakade and A. Kalai. From batch to transductive online learning. In *NIPS*, 2005.

[17] Y. Koren. Collaborative filtering with temporal dynamics. In *KDD*, 2009.

[18] J. Lee, B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp. Practical large-scale optimization for max-norm regularization. In *NIPS*, 2010.

[19] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2007.

[20] R. Salakhutdinov and N. Srebro. Collaborative filtering in a non-uniform world: Learning with the weighted trace norm. In *NIPS*, 2010.

[21] O. Shamir and S. Shalev-Shwartz. Collaborative filtering with the trace norm: Learning, bounding, and transducing. In *COLT*, 2011.

[22] N. Srebro, J. Rennie, and T. Jaakkola. Maximum-margin matrix factorization. In *NIPS*, 2004.

[23] N. Srebro and A. Shraibman. Rank, trace-norm and max-norm. In *COLT*, 2005.

A Derivation of the Minimax Forecaster

In this appendix, we outline how the Minimax Forecaster is derived, as well as its associated guarantees. This outline closely follows the exposition in [10, Chapter 8], to which we refer the reader for some of the technical derivations.

First, we note that the Minimax Forecaster as presented in [10] actually refers to a slightly different setup than ours, where the outcome space is $\mathcal{Y} = \{0, 1\}$ and the prediction space is $\mathcal{P} = [0, 1]$, rather than $\mathcal{Y} = \{-1, +1\}$ and $\mathcal{P} = [-1, +1]$. We will first derive the forecaster for the first setting, and then show how to convert it to the second setting.

Our goal is to find a predictor which minimizes the worst-case regret,

$$\max_{\mathbf{y}\in\{0,1\}^T} \Big(L(\mathbf{p}, \mathbf{y}) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}) \Big)$$

where $\mathbf{p} = (p_1, \ldots, p_T)$ is the prediction sequence.

For convenience, in the following we sometimes use the notation $\mathbf{y}^t$ to denote a vector in $\{0, 1\}^t$. The idea of the derivation is to work backwards, starting with computing the optimal prediction at the last round T, then deriving the optimal prediction at round $T - 1$ and so on. In the last round T, the first $T - 1$ outcomes $\mathbf{y}^{T-1}$ have been revealed, and we want to find the optimal prediction p_T. Since our goal is to minimize worst-case regret with respect to the absolute loss, we just need to compute p_T which minimizes

$$\max\Big\{ L(\mathbf{p}^{T-1}, \mathbf{y}^{T-1}) + p_T - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{T-1}0) \;,\; L(\mathbf{p}^{T-1}, \mathbf{y}^{T-1}) + (1 - p_T) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{T-1}1) \Big\} \;.$$

In our setting, it is not hard to show that $\left|\inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{t-1}0) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{t-1}1)\right| \leq 1$ (see [10, Lemma 8.1]). Using this, we can compute the optimal p_T to be

$$p_T = \frac{1}{2}\Big(A_T(\mathbf{y}^{T-1}1) - A_T(\mathbf{y}^{T-1}0) + 1 \Big) \tag{5}$$

where $A_T(\mathbf{y}^T) = -\inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^T)$.

Having determined p_T, we can continue to the previous prediction p_{T-1}. This is equivalent to minimizing

$$\max\Big\{ L(\mathbf{p}^{T-2}, \mathbf{y}^{T-2}) + p_{T-1} + A_{T-1}(\mathbf{y}^{T-2}0) \;,\; L(\mathbf{p}^{T-1}, \mathbf{y}^{T-1}) + (1 - p_{T-1}) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{T-1}1) \Big\}$$

where

$$A_{t-1}(\mathbf{y}^{t-1}) = \min_{p_t\in[0,1]} \max \Big\{ p_t - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{t-1}0) \;,\; (1 - p_t) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^{t-1}1) \Big\}. \tag{6}$$

Note that by plugging in the value of p_T from Eq. (5), we also get the following equivalent formulation for $A_{T-1}(\mathbf{y}^{T-1})$:

$$A_{T-1}(\mathbf{y}^{T-1}) = \frac{1}{2}\Big(A_T(\mathbf{y}^{T-1}0) + A_T(\mathbf{y}^{T-1}1) + 1 \Big).$$

Again, it is possible to show that the optimal value of p_{T-1} is

$$p_{T-1} = \frac{1}{2}\Big(A_{T-1}(\mathbf{y}^{T-2}1) - A_T(\mathbf{y}^{T-2}0) + 1 \Big).$$

Repeating this procedure, one can show that at any round t, the minimax optimal prediction is

$$p_t = \frac{1}{2}\Big(A_t(\mathbf{y}^{t-1}1) - A_t(\mathbf{y}^{t-1}0) + 1 \Big) \tag{7}$$

where A_t is defined recursively as $A_T(\mathbf{y}^T) = -\inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \mathbf{y}^T)$ and

$$A_{t-1}(\mathbf{y}^{t-1}) = \frac{1}{2}\Big(A_t(\mathbf{y}^{t-1}0) + A_t(\mathbf{y}^{t-1}1) + 1 \Big). \tag{8}$$

for all t.

At first glance, computing p_t from Eq. (7) might seem tricky, since it requires computing $A_t(\mathbf{y}^t)$ whose recursive expansion in Eq. (8) involves exponentially many terms. Luckily, the recursive expansion has a simple structure, and it is not hard to show that

$$A_t(\mathbf{y}^t) ~=~ \frac{T-t}{2} - \frac{1}{2^T}\sum_{\mathbf{y}\in\{0,1\}^T}\left(\inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f},\mathbf{y}^tY^{T-t})\right) ~=~ \frac{T-t}{2} - \mathbb{E}\Big[\inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f},\mathbf{y}^tY^{T-t})\Big] \tag{9}$$

where Y^{T-t} is a sequence of $T-t$ i.i.d. Bernoulli random variables, which take values in $\{0,1\}$ with equal probability. Plugging this into the formula for the minimax prediction in Eq. (7), we get that[3]

$$p_t = \frac{1}{2}\left(\mathbb{E}\left[\inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f},\mathbf{y}^{t-1}0Y^{T-t}) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f},\mathbf{y}^{t-1}1Y^{T-t})\right] + 1\right). \tag{10}$$

This prediction rule constitutes the Minimax Forecaster as presented in [10].

After deriving the algorithm, we turn to analyze its regret performance. To do so, we just need to note that A_0 equals the worst-case regret —see the recursive definition at Eq. (6). Using the alternative explicit definition in Eq. (9), we get that the worst-case regret equals

$$\frac{T}{2} - \mathbb{E}\left[\inf_{\mathbf{f}\in\mathcal{F}}\sum_{t=1}^{T}|f_t - Y_t|\right] ~=~ \mathbb{E}\left[\sup_{\mathbf{f}\in\mathcal{F}}\sum_{t=1}^{T}\left(\frac{1}{2} - |f_t - Y_t|\right)\right] ~=~ \mathbb{E}\left[\sup_{\mathbf{f}\in\mathcal{F}}\sum_{t=1}^{T}\left(f_t - \frac{1}{2}\right)\sigma_t\right]$$

where σ_t are i.i.d. Rademacher random variables (taking values of -1 and $+1$ with equal probability). Recalling the definition of Rademacher complexity, Eq. (2), we get that the regret is bounded by the Rademacher complexity of the shifted class, which is obtained from $\mathcal{F}$ by taking every $\mathbf{f}\in\mathcal{F}$ and replacing every coordinate f_t by $f_t - 1/2$.

Finally, it remains to show how to convert the forecaster and analysis above to the setting discussed in this paper, where the outcomes are in $\{-1,+1\}$ rather than $\{0,1\}$ and the predictions are in $[-1,+1]$ rather than $[0,1]$. To do so, consider a learning problem in this new setting, with some class $\mathcal{F}$. For any vector $\mathbf{y}$, define $\widetilde{\mathbf{y}}$ to be the shifted vector $(\mathbf{y}+\mathbf{1})/2$, where $\mathbf{1}=(1,\ldots,1)$ is the all-ones vector. Also, define $\widetilde{\mathcal{F}}$ to be the shifted class $\widetilde{\mathcal{F}} = \{(\mathbf{f}+\mathbf{1})/2 ~:~ \mathbf{f}\in\mathcal{F}\}$. It is easily seen that $L(\mathbf{f},\mathbf{y}) = 2L(\widetilde{\mathbf{f}},\widetilde{\mathbf{y}})$ for any $\mathbf{f},\mathbf{y}$. As a result, if we look at the prediction p_t given by our forecaster in Eq. (3), then $\widetilde{p_t} = (p_t+1)/2$ is the minimax optimal prediction given by Eq. (10) with respect to the class $\widetilde{\mathcal{F}}$ and the outcomes $\widetilde{\mathbf{y}}^T$. So our analysis above applies, and we get that

$$\begin{aligned}
\max_{\mathbf{y}\in\{-1,+1\}^T}\left(L(\mathbf{p},\mathbf{y}) - \inf_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f},\mathbf{y})\right) ~&=~ \max_{\widetilde{\mathbf{y}}\in[0,1]^T} 2\left(L(\widetilde{\mathbf{p}},\widetilde{\mathbf{y}}) - \inf_{\widetilde{\mathbf{f}}\in\widetilde{\mathcal{F}}} L(\widetilde{\mathbf{f}},\widetilde{\mathbf{y}})\right)\\
&=~ 2\mathbb{E}\left[\sup_{\widetilde{\mathbf{f}}\in\widetilde{\mathcal{F}}}\sum_{t=1}^{T}\left(\widetilde{f}_t - \frac{1}{2}\right)\sigma_t\right]\\
&=~ \mathbb{E}\left[\sup_{\mathbf{f}\in\mathcal{F}}\sum_{t=1}^{T}\sigma_t f_t\right]
\end{aligned}$$

which is exactly the Rademacher complexity of the class $\mathcal{F}$.

B Proof of Thm. 3

Let $Y(t)$ denote the set of Bernoulli random variables chosen at round t. Let $\mathbb{E}_{z_t}$ denote expectation with respect to z_t, conditioned on $z_1, Y(1), \ldots, z_{t-1}, Y(t-1)$ as well as $Y(t)$. Let $\mathbb{E}_{Y(t)}$ denote the expectation with respect to the random drawing of $Y(t)$, conditioned on $z_1, Y(1), \ldots, z_{t-1}, Y(t-1)$.

We will need two simple observations. First, by convexity of the loss function, we have that for any p_t, f_t, y_t, $\ell(p_t,y_t) - \ell(f_t,y_t) \le (p_t - f_t)\,\partial_{p_t}\ell(p_t,y_t)$. Second, by definition of r_t and

[3] This fact appears in an implicit form in [9] —see also [10, Exercise 8.4].

z_t, we have that for any fixed p_t, f_t,

$$\begin{aligned}
\frac{1}{\rho b}(p_t - f_t)\partial_{p_t}\ell(p_t, y_t) &= \frac{1}{b}(p_t - f_t)(1 - 2r_t) \\
&= \frac{1}{b}r_t(f_t - p_t) + \frac{1}{b}(1 - r_t)(p_t - f_t) \\
&= r_t(\widetilde{f}_t - \widetilde{p}_t) + (1 - r_t)(\widetilde{p}_t - \widetilde{f}_t) \\
&= r_t\left((1 - \widetilde{p}_t) - \left(1 - \widetilde{f}_t\right)\right) + (1 - r_t)\left((\widetilde{p}_t + 1) - \left(\widetilde{f}_t + 1\right)\right) \\
&= \mathbb{E}_{z_t}\left[|\widetilde{p}_t - z_t| - \left|\widetilde{f}_t - z_t\right|\right] \ .
\end{aligned}$$

The last transition uses the fact that $\widetilde{p}_t, \widetilde{f}_t \in [-1, +1]$. By these two observations, we have

$$\sum_{t=1}^{T}\left(\ell(p_t, y_t) - \ell(f_t, y_t)\right) \leq \sum_{t=1}^{T}(p_t - f_t)\,\partial_{p_t}\ell(p_t, y_t) = \rho\, b\ \sum_{t=1}^{T}\mathbb{E}_{z_t}\left[|\widetilde{p}_t - z_t| - \left|\widetilde{f}_t - z_t\right|\right] \ . \tag{11}$$

Now, note that $|\widetilde{p}_t - z_t| - |\widetilde{f}_t - z_t| - \mathbb{E}_{z_t}\left[|\widetilde{p}_t - z_t| - |\widetilde{f}_t - z_t|\right]$ for $t = 1, \ldots, T$ is a martingale difference sequence: for any values of $z_1, Y(1), \ldots, z_{t-1}, Y(t-1), Y(t)$ (which fixes $\widetilde{p}_t$), the conditional expectation of this expression over z_t is zero. Using Azuma's inequality, we can upper bound Eq. (11) with probability at least $1 - \delta/2$ by

$$\rho\, b\ \sum_{t=1}^{T}\left(|\widetilde{p}_t - z_t| - |\widetilde{f}_t - z_t|\right) + \rho\, b\sqrt{8T\ln(2/\delta)}. \tag{12}$$

The next step is to relate Eq. (12) to $\rho\, b\sum_{t=1}^{T}\left(\left|\mathbb{E}_{Y(t)}[\widetilde{p}_t] - z_t\right| - |\widetilde{f}_t - z_t|\right)$. It might be tempting to appeal to Azuma's inequality again. Unfortunately, there is no martingale difference sequence here, since z_t is itself a random variable whose distribution is influenced by $Y(t)$. Thus, we need to turn to coarser methods. Eq. (12) can be upper bounded by

$$\rho\, b\ \sum_{t=1}^{T}\left(\left|\mathbb{E}_{Y(t)}[\widetilde{p}_t] - z_t\right| - |\widetilde{f}_t - z_t|\right) + \rho\, b\ \sum_{t=1}^{T}\left|\widetilde{p}_t - \mathbb{E}_{Y(t)}[\widetilde{p}_t]\right| + \rho\, b\sqrt{8T\ln(2/\delta)}. \tag{13}$$

Recall that $\widetilde{p}_t$ is an average over ηT i.i.d. random variables, with expectation $\mathbb{E}_{Y(t)}[\widetilde{p}_t]$. By Hoeffding's inequality, this implies that for any $t = 1, \ldots, T$, with probability at least $1 - \delta/2T$ over the choice of $Y(t)$, $\left|\widetilde{p}_t - \mathbb{E}_{Y(t)}[\widetilde{p}_t]\right| \leq \sqrt{2\ln(2T/\delta)/(\eta T)}$. By a union bound, it follows that with probability at least $1 - \delta/2$ over the choice of $Y(1), \ldots, Y(T)$,

$$\sum_{t=1}^{T}\left|\widetilde{p}_t - \mathbb{E}_{Y(t)}[\widetilde{p}_t]\right| \leq \sqrt{\frac{2T\ln(2T/\delta)}{\eta}} \ .$$

Combining this with Eq. (13), we get that with probability at least $1 - \delta$,

$$\rho\, b\sum_{t=1}^{T}\left(\left|\mathbb{E}_{Y(t)}[\widetilde{p}_t] - z_t\right| - |\widetilde{f}_t - z_t|\right) + \rho\, b\sqrt{\frac{2T\ln(2T/\delta)}{\eta}} + \rho\, b\sqrt{8T\ln(2/\delta)} \ . \tag{14}$$

Finally, by definition of $\widetilde{p}_t = p_t/b$, we have

$$\mathbb{E}_{Y(t)}[\widetilde{p}_t] = \mathbb{E}_{Y(t)}\left[\inf_{\mathbf{f}\in\mathcal{F}} L\left(\widetilde{\mathbf{f}}, z_1 \ldots z_{t-1}\,(-1)\,Y_{t+1} \ldots Y_T\right) - \inf_{\mathbf{f}\in\mathcal{F}} L\left(\widetilde{\mathbf{f}}, z_1 \ldots z_{t-1}\,1\,Y_{t+1} \ldots Y_T\right)\right] \ .$$

This is exactly the Minimax Forecaster's prediction at round t, with respect to the sequence of outcomes $z_1, \ldots, z_{t-1} \in \{-1, +1\}$, and the class $\widetilde{\mathcal{F}} := \left\{\widetilde{\mathbf{f}} : \mathbf{f} \in \mathcal{F}\right\} \subseteq [-1, 1]^T$. Therefore, using Thm. 1, we can upper bound Eq. (14) by

$$\rho\, b\,\mathcal{R}_T(\widetilde{\mathcal{F}}) + \rho\, b\sqrt{\frac{2T\ln(2T/\delta)}{\eta}} + \rho\, b\sqrt{8T\ln(2/\delta)} \ .$$

By definition of $\widetilde{\mathcal{F}}$ and Rademacher complexity, it is straightforward to verify that $\mathcal{R}_T(\widetilde{\mathcal{F}}) = \frac{1}{b}\mathcal{R}_T(\mathcal{F})$. Using that to rewrite the bound, and slightly simplifying for readability, the result stated in the theorem follows.

C Proof of Lemma 1

The proof assumes that the infimum and supremum of certain functions over $\mathcal{Y}, \mathcal{F}$ are attainable. If not, the proof can be easily adapted by finding attainable values which are ϵ-close to the infimum or supremum, and then taking $\epsilon \to 0$.

For the purpose of contradiction, suppose there exists a strategy for the adversary and a round $r \leq T$ such that at the end of round r, the forecaster suffers a regret $G' > G$ with probability larger than δ. Consider the following modified strategy for the adversary: the adversary plays according to the aforementioned strategy until round r. It then computes

$$f^* = \operatorname*{argmin}_{f \in \mathcal{F}} \sum_{t=1}^{r} \ell(f_t, y_t) \ .$$

At all subsequent rounds $t = r+1, r+2, \ldots, T$, the adversary chooses

$$y_t^* = \operatorname*{argmax}_{y \in \mathcal{Y}} \inf_{p \in \mathcal{P}} \big(\ell(p, y) - \ell(f_t^*, y)\big) \ .$$

By the assumption on the loss function,

$$\ell(p_t, y_t^*) - \ell(f_t^*, y_t^*) \geq \inf_{p \in \mathcal{P}} \big(\ell(p, y_t^*) - \ell(f_t^*, y_t^*)\big) = \sup_{y \in \mathcal{Y}} \inf_{p \in \mathcal{P}} \big(\ell(p, y) - \ell(f_t^*, y)\big) \geq 0 \ .$$

Thus, the regret over all T rounds, with respect to f^*, is

$$\sum_{t=1}^{r} \big(\ell(p_t, y_t) - \ell(f_t^*, y_t)\big) + \sum_{t=r+1}^{T} \big(\ell(p_t, y_t^*) - \ell(f_t^*, y_t^*)\big) \geq \sum_{t=1}^{r} \ell(p_t, y_t) - \inf_{f \in \mathcal{F}} \sum_{t=1}^{r} \ell(f_t, y_t) + 0$$

which is at least G' with probability larger than δ. On the other hand, we know that the learner's regret is at most most G with probability at least $1 - \delta$. Thus we have a contradiction and the proof is concluded.