
Towards Minimax Online Learning with Unknown Time Horizon

Haipeng Luo

Department of Computer Science, Princeton University, Princeton, NJ 08540

HAIPENGL@CS.PRINCETON.EDU

Robert E. Schapire

Department of Computer Science, Princeton University, Princeton, NJ 08540

SCHAPIRE@CS.PRINCETON.EDU

Abstract

We consider online learning when the time horizon is unknown. We apply a minimax analysis, beginning with the fixed horizon case, and then moving on to two unknown-horizon settings, one that assumes the horizon is chosen randomly according to some known distribution, and the other which allows the adversary full control over the horizon. For the random horizon setting with restricted losses, we derive a fully optimal minimax algorithm. And for the adversarial horizon setting, we prove a nontrivial lower bound which shows that the adversary obtains strictly more power than when the horizon is fixed and known. Based on the minimax solution of the random horizon setting, we then propose a new adaptive algorithm which “pretends” that the horizon is drawn from a distribution from a special family, but no matter how the actual horizon is chosen, the *worst-case* regret is of the optimal rate. Furthermore, our algorithm can be combined and applied in many ways, for instance, to online convex optimization, follow the perturbed leader, exponential weights algorithm and first order bounds. Experiments show that our algorithm outperforms many other existing algorithms in an online linear optimization setting.

1. Introduction

We study online learning problems with unknown time horizon with the aim of developing algorithms and approaches for the realistic case that the number of time steps is initially unknown.

We first adopt the standard Hedge setting (Freund & Schapire, 1997) where the learner chooses a distribution over N actions on each round, and the losses for each action are then selected by an adversary. The

learner incurs loss equal to the expected loss of the actions in terms of the distribution it chose for this round, and its goal is to minimize the regret, the difference between its cumulative loss and that of the best action after T rounds.

Various algorithms are known to achieve the optimal (up to a constant) upper bound $O(\sqrt{T \ln N})$ on the regret. Most of them assume that the horizon T is known ahead of time, especially those which are minimax optimal. When the horizon is unknown, the so-called doubling trick (Cesa-Bianchi et al., 1997) is a general technique to make a learning algorithm adaptive and still achieve $O(\sqrt{T \ln N})$ regret uniformly for any T . The idea is to first guess a horizon, and once the actual horizon exceeds this guess, double it and restart the algorithm. Although, in theory, it is widely applicable, the doubling trick is aesthetically inelegant, and intuitively wasteful, since it repeatedly restarts itself, entirely forgetting all the preceding information. Other approaches have also been proposed, as we discuss shortly.

In this paper, we study the problem of learning with unknown horizon in a game-theoretic framework. We consider a number of variants of the problem, and make progress toward a minimax solution. Based on this approach, we give a new general technique which can also make other minimax or non-minimax algorithms adaptive and achieve low regret in a very general online learning setting. The resulting algorithm is still not exactly optimal, but it makes use of all the previous information on each round and achieves much lower regret in experiments.

We view the Hedge problem as a repeated game between the learner and the adversary. Abernethy et al. (2008b), and Abernethy & Warmuth (2010) proposed an exact minimax optimal solution for a slightly different game with binary losses, assuming that the loss of the best action is at most some fixed constant. They derived the solution under a very simple type of loss space; that is, on each round only one action suffers one unit loss. We call this the basis vector loss space. As a preliminary of this paper, we also derive a similar minimax solution under this simple loss space for

our setting where the horizon T is fixed and known to the learner ahead of time.

We then move on to the primary interest of this paper, that is, the case when the horizon is unknown to the learner. We study this unknown horizon setting in the minimax framework, with the aim of ultimately deriving game-theoretically optimal algorithms. Two types of models are studied. The first one assumes the horizon is chosen according to some known distribution, and the learner’s goal is to minimize the expected regret. We show the exact minimax solution for the basis vector loss space in this case. It turns out that the distribution the learner should choose on each round is simply the conditional expectation of the distributions the learner would have chosen for the fixed horizon case.

The second model we study gives the adversary the power to decide the horizon on the fly, which is possibly the most adversarial case. In this case, we no longer use the regret as the performance measure. Otherwise the adversary would obviously choose an infinite horizon. Instead, we use a scaled regret to measure the performance. Specifically, we scale the regret at time t by the optimal regret under fixed horizon t . The exact optimal solution in this case is unfortunately not found and remains an open problem, even for the extremely simple case. However, we give a lower bound for this setting to show that the optimal regret is strictly greater than the one in the fixed horizon game. That is, the adversary does obtain strictly more power if allowed to pick the horizon.

We then propose our new adaptive algorithm based on the minimax solution in the random horizon setting. One might doubt how realistic a random horizon is in practice. Even if the true horizon is indeed drawn from a fixed distribution, how can we know this distribution? We address these problems at the same time. Specifically, we prove that no matter how the horizon is chosen, if we assume it is drawn from a distribution from a special family, and let the learner play in a way similar to the one in the random horizon setting, then the *worst-case* regret at any time T (not the expected regret) can still be of the optimal order. In other words, although the learner is behaving as if the horizon is random, its regret will be small even if the horizon is actually controlled by an adversary. Moreover, the results hold for not just the Hedge problem, but a general online learning setting that includes many interesting problems.

Our idea can be combined not only with the minimax algorithm, but also the “follow the perturbed leader” algorithm and the exponential weights algorithm. In addition, our technique can not only deal with unknown horizon, but also other unknown information such as the loss of the best action, thus leading to a first order regret bound that depends on the loss of the best action (Cesa-Bianchi & Lugosi,

2006). Like the doubling trick, this seems to be a quite general way to make an algorithm adaptive. Furthermore, we conduct experiments showing that our algorithm outperforms many existing algorithms, including the doubling trick, in an online linear optimization setting within an ℓ_2 ball where our algorithm has an explicit closed form.

The rest of the paper is organized as follows. We define the Hedge setting formally in Section 2, and derive the minimax solution for the fixed horizon setting as the preliminary of this paper in Section 3. In Section 4, we study two unknown horizon settings in the minimax framework. We then turn to a general online learning setting and present our new adaptive algorithm in Section 5. Implementation issues, experiments, and applications are discussed in Section 6. We omit most of the proofs due to space limitations, but all details can be found in the supplementary material.

Related work Besides the doubling trick, other adaptive algorithms have been studied (Auer et al., 2002; Gentile, 2003; Yaroshinsky et al., 2004; Chaudhuri et al., 2009). Auer et al. (2002) showed that for algorithms such as the exponential weights algorithm (Littlestone & Warmuth, 1994; Freund & Schapire, 1997; 1999), where a learning rate η should be set as a function of the horizon, typically in the form $\sqrt{(b \ln N)/T}$ for some constant b , one can simply set the learning rate adaptively as $\sqrt{(b \ln N)/t}$, where t is the current number of rounds. In other words, this algorithm always pretends the current round is the last round. Although this idea works with the exponential weights algorithm, we remark that assuming the current round is the last round does not always work. Specifically, one can show that it will fail if applied to the minimax algorithm (see Section 6.4). In another approach to online learning with unknown horizon, Chaudhuri et al. (2009) proposed an adaptive algorithm based on a novel potential function reminiscent of the half-normal distribution.

Other performance measures different from the usual regret were studied before. Foster & Vohra (1998) introduced internal regret comparing the loss of an online algorithm to the loss of a modified algorithm which consistently replaces one action by another. Herbster & Warmuth (1995), and Bousquet & Warmuth (2003) compared the learner’s loss with the best k -shifting expert, while Hazan & Seshadhri (2007) studied the usual regret within any time interval. To the best of our knowledge, the form of scaled regret that we study is new. Lower bounds on anytime regret in terms of the quadratic variations for any loss sequence (instead of the worst case sequence this paper considers) were studied by Gofer & Mansour (2012).

2. Repeated Games

We first consider the following repeated game between a learner and an adversary. The learner has access to N actions. On each round $t = 1, \dots, T$, (1) the learner chooses a distribution \mathbf{P}_t over the actions; (2) the adversary reveals the loss vector $\mathbf{Z}_t = (Z_{t,1}, \dots, Z_{t,N}) \in \mathbf{LS}$, where $Z_{t,i}$ is the loss for action i for this round, and the *loss space* \mathbf{LS} is a subset of $[0, 1]^N$; (3) the learner suffers loss $\ell_t = \mathbf{P}_t \cdot \mathbf{Z}_t$ for this round.

Notice that the adversary can choose the losses on round t with full knowledge of the history $\mathbf{P}_{1:t}$ and $\mathbf{Z}_{1:t-1}$, that is, all the previous choices of the learner and the adversary (we use notation $a_{1:t}$ to denote the multiset $\{a_1, \dots, a_t\}$). We also denote the cumulative loss up to round t for the learner and the actions by $L_t = \sum_{t'=1}^t \ell_{t'}$ and $\mathbf{M}_t = \sum_{t'=1}^t \mathbf{Z}_{t'}$ respectively. The goal for the learner is to minimize the difference between its total loss and that of the best action at the end of the game. In other words, the goal of the learner is to minimize $\mathbf{Reg}(L_T, \mathbf{M}_T)$, where we define the regret function $\mathbf{Reg}(L, \mathbf{M}) \triangleq L - \min_i M_i$, for $L \in \mathbb{R}$ and $\mathbf{M} \in \mathbb{R}^N$. The number of rounds T is called the *horizon*.

Regarding the loss space \mathbf{LS} , perhaps the simplest one is $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$, the N standard basis vectors in N dimensions. Playing with this loss space means that on each round, the adversary chooses one single action to incur one unit loss. In order to show the intuition of our main results, we mainly focus on this basis vector loss space in Sections 3 and 4, but we return to the most general case $[0, 1]^N$ later.

3. Minimax Solution for Fixed Horizon

Although our primary interest in this paper is the case when the horizon is unknown to the learner, we first present some preliminary results on the setting where the horizon is known to both the learner and the adversary ahead of time. These will later be useful for the unknown horizon case.

If we treat the learner as an algorithm \mathbf{Alg} that takes the information of previous rounds as inputs, and outputs a distribution $\mathbf{P}_t = \mathbf{Alg}(\mathbf{P}_{1:t-1}, \mathbf{Z}_{1:t-1})$ that the learner is going to play with, then finding the optimal solution in this fixed horizon setting can be viewed as solving the minimax expression

$$\inf_{\mathbf{Alg}} \sup \mathbf{Reg}(L_T, \mathbf{M}_T). \quad (1)$$

Alternatively, we can recursively define:

$$\begin{aligned} V(\mathbf{M}, 0) &\triangleq -\min_i M_i; \\ V(\mathbf{M}, r) &\triangleq \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}} (\mathbf{P} \cdot \mathbf{Z} + V(\mathbf{M} + \mathbf{Z}, r-1)), \end{aligned}$$

where $\mathbf{M} \in \mathbb{R}^N$ is a loss vector, r is a nonnegative integer, and $\Delta(N)$ is the N dimensional simplex. By a simple

argument, one can show that the value of $V(\mathbf{M}, r)$ is the regret of a game with r rounds starting from the situation that each action has initial loss M_i , and assuming both the learner and the adversary will play optimally. In fact, the value of Eq. (1) is exactly $V(\mathbf{0}, T)$, and the optimal learner algorithm is the one that chooses the \mathbf{P}^* which realizes the minimum in the definition of $V(\mathbf{M}, r)$ when the actions' cumulative loss vector is \mathbf{M} and there are r rounds left. We call $V(\mathbf{0}, T)$ the *value* of the game.

As a concrete illustration of these ideas, we now consider the basis vector loss space¹, that is, $\mathbf{LS} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$. It turns out that under this loss space, the value function V has a nice closed form. Similar to the results from Cesa-Bianchi et al. (1997) and Abernethy et al. (2008b), we show that V can be expressed in terms of a random walk. Suppose $R(\mathbf{M}, r)$ is the expectation of the loss of the best action if the adversary chooses each \mathbf{e}_i uniformly randomly for the remaining r rounds, starting from loss vector \mathbf{M} . Formally, $R(\mathbf{M}, r)$ can be defined in a recursive way: $R(\mathbf{M}, 0) \triangleq \min_i M_i$; $R(\mathbf{M}, r) \triangleq \frac{1}{N} \sum_{i=1}^N R(\mathbf{M} + \mathbf{e}_i, r-1)$. The connection between V and R , and the optimal algorithm are then shown by the following theorem.

Theorem 1. *If $\mathbf{LS} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$, then for any vector \mathbf{M} and integer $r \geq 0$,*

$$V(\mathbf{M}, r) = \frac{r}{N} - R(\mathbf{M}, r).$$

Let $c_N = \frac{1}{N} \sqrt{2(N-1) \ln N}$. Then the value of the game satisfies

$$V(\mathbf{0}, T) \leq c_N \sqrt{T}. \quad (2)$$

Moreover, on round t , the optimal learner algorithm is the one that chooses weight $P_{t,i} = V(\mathbf{M}_{t-1}, r) - V(\mathbf{M}_{t-1} + \mathbf{e}_i, r-1)$ for each action i , where \mathbf{M}_{t-1} is the current cumulative loss vector and r is the number of remaining rounds, that is, $r = T - t + 1$.

Theorem 1 tells us that under the basis vector loss space, the best way to play is to assume that the adversary is playing uniformly randomly, because r/N and $R(\mathbf{M}, r)$ are exactly the expected losses for the learner and for the best action respectively. In practice, computing $R(\mathbf{M}, r)$ needs exponential time. However, we can estimate it by sampling (see similar work in Abernethy et al., 2008b). Note that c_N is decreasing when $N \geq 4$ (with maximum value about 0.72). So contrary to the $O(\sqrt{T \ln N})$ regret bound for the general loss space $[0, 1]^N$ which is increasing in N , here $V(\mathbf{0}, T)$ is of order $O(\sqrt{T})$.

¹For other loss spaces, finding minimax solutions seems difficult. However, we show the relation of the values of the game for different loss spaces in the supplementary file, see Theorem 10.

4. Playing without Knowing the Horizon

We turn now to the case in which the horizon T is unknown to the learner, which is often more realistic in practice. There are several ways of modeling this setting. For example, the horizon can be chosen ahead of time according to some fixed distribution, or it can even be chosen by the adversary. We will discuss these two variants separately.

4.1. Random Horizon

Suppose the horizon T is chosen according to some fixed distribution Q which is known to both the learner and the adversary. Before the game starts, a random T is drawn, and neither the learner nor the adversary knows the actual value of T . The game stops after T rounds, and the learner aims to minimize the expectation of the regret. Using our earlier notation, the problem can be formally defined as

$$\inf_{\text{Alg}} \sup_{\mathbf{Z}_{1:\infty}} \mathbb{E}_{T \sim Q} [\mathbf{Reg}(L_T, \mathbf{M}_T)],$$

where we assume the expectation is always finite. We sometimes omit the subscript $T \sim Q$ for simplicity.

Continuing the example in Section 3 of the basis vector loss space, we can again show the exact minimax solution, which has a strong connection with the one for the fixed horizon setting.

Theorem 2. *If $\mathbf{LS} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$, then*

$$\begin{aligned} & \inf_{\text{Alg}} \sup_{\mathbf{Z}_{1:\infty}} \mathbb{E}_{T \sim Q} [\mathbf{Reg}(L_T, \mathbf{M}_T)] \\ &= \mathbb{E}_{T \sim Q} [\inf_{\text{Alg}} \sup_{\mathbf{Z}_{1:T}} \mathbf{Reg}(L_T, \mathbf{M}_T)]. \end{aligned} \quad (3)$$

Moreover, on round t , the optimal learner plays with the distribution $\mathbf{P}_t = \mathbb{E}_{T \sim Q} [\mathbf{P}_t^T | T \geq t]$, where \mathbf{P}_t^T is the optimal distribution the learner would play if the horizon is T , that is, $P_{t,i}^T = V(\mathbf{M}_{t-1}, T-t+1) - V(\mathbf{M}_{t-1} + \mathbf{e}_i, T-t)$.

Eq. (3) tells us that if the horizon is drawn from some distribution, then even though the learner does not know the actual horizon before playing the game, as long as the adversary does not know this information either, it can still do as well as the case when they are both aware of the horizon.

However, so far this model does not seem to be quite useful in practice for several reasons. First of all, the horizon might not be chosen according to a distribution. Even if it is, this distribution is probably unknown. Secondly, what we really care about is the performance which holds uniformly for any horizon, instead of the expected regret. Last but not least, one might conjecture that the similar result stated in Theorem 2 should hold for other more general loss spaces, which is in fact not true (see Example 1 in the supplementary file), making the result seem even less useful.

Fortunately, we address all these problems and develop new adaptive algorithms based on the result in this section. We discuss these in Section 5 after first introducing the fully adversarial model.

4.2. Adversarial Horizon

The most adversarial setting is the one where the horizon is completely controlled by the adversary. That is, we let the adversary decide whether to continue or stop the game on each round according to the current situation. However, notice that the value of the game is increasing in the horizon. So if the adversary can determine the horizon and its goal is still to maximize the regret, then the problem would not make sense because the adversary would clearly choose to play the game forever and never stop leading to infinite regret. One reasonable way to address this issue is to scale the regret by the value of the fixed horizon game $V(\mathbf{0}, T)$, so that the scaled regret $\mathbf{Reg}(L_T, \mathbf{M}_T)/V(\mathbf{0}, T)$ indicates how many times worse is the regret compared to the one that is optimal given the horizon. Under this setting, the corresponding minimax expression is

$$\tilde{V} = \inf_{\text{Alg}} \sup_T \sup_{\mathbf{Z}_{1:T}} \frac{\mathbf{Reg}(L_T, \mathbf{M}_T)}{V(\mathbf{0}, T)}. \quad (4)$$

Unfortunately, finding the minimax solution to this setting seems to be quite challenging, even for the simplest case $N = 2$. It is clear, however, that \tilde{V} is at most some constant due to the existence of adaptive algorithms such as the doubling trick, which can achieve the optimal regret bound up to a constant without knowing T . Another clear fact is $\tilde{V} \geq 1$, since it is impossible for the learner to do better than the case when it is aware of the horizon. Below, we derive a nontrivial lower bound that is greater than 1, thus proving that the adversary does gain strictly more power when it can stop the game whenever it wants.

Theorem 3. *If $N = 2$ and $\mathbf{LS} = [0, 1]^2$, then $\tilde{V} \geq \sqrt{2}$. That is, for every algorithm, there exists an adversary and a horizon T such that the regret of the learner after T rounds is at least $\sqrt{2}V(\mathbf{0}, T)$.*

5. A New General Adaptive Algorithm

We study next how the random-horizon algorithm of Section 4.1 can be used when the horizon is entirely unknown and furthermore, for a much more general class of online learning problems. In Theorem 2, we proposed an algorithm that simply takes the conditional expectation of the distributions we would have played if the horizon were given. Notice that even though it is derived from the random horizon setting, it can still be used in any setting as an adaptive algorithm in the sense that it does not require the horizon as a parameter. However, to use this algorithm,

we should ask two questions: What distribution should we use? And what can we say about the algorithm's performance for an arbitrary horizon instead of in expectation?

As a first attempt, suppose we use a uniform distribution over $1, \dots, T_0$, where T_0 is a huge integer. From what we observe in some numerical calculations, $\mathbb{E}[\mathbf{P}_t^T | T \geq t]$ tends to be a uniform distribution in this case. Clearly it cannot be a good algorithm if for each round, it just places equal weights for each action regardless of the actions' behaviors. In fact, one can verify that the exponential distribution (that is, $\Pr[T = t] \propto \alpha^t$ for some constant $0 < \alpha < 1$) also does not work. These examples show that even though this algorithm gives us the optimal expected regret, it can still suffer a big regret for a particular trial of the game, which we definitely want to avoid.

Nevertheless, it turns out that there does exist a family of distributions that can guarantee the regret to be of order $O(\sqrt{T})$ for any T . Moreover, this is true for a very general online learning problem that includes the Hedge setting we have been discussing. Before stating our results, we first formally describe this general setting, which is sometimes called the *online convex optimization* problem (Zinkevich, 2003; Shalev-Shwartz, 2011). Let S be a compact convex set, and \mathcal{F} be a set of convex functions defined on S . On each round $t = 1, \dots, T$: (1) the learner chooses a point $\mathbf{x}_t \in S$; (2) the adversary chooses a loss function $f_t \in \mathcal{F}$; (3) the learner suffers loss $f_t(\mathbf{x}_t)$ for this round. The regret after T rounds is defined by

$$\mathbf{Reg}(\mathbf{x}_{1:T}, f_{1:T}) = \sum_{t=1}^T f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in S} \sum_{t=1}^T f_t(\mathbf{x}).$$

It is clear that the Hedge problem is a special case of the above setting with S being the probability simplex, and \mathcal{F} being a set of linear functions defined by a point in the loss space, that is, $\mathcal{F} = \{f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w} : \mathbf{w} \in \mathbf{LS}\}$. Similarly, to study the minimax algorithm we define the following $V_{S, \mathcal{F}}$ function of the multiset \mathcal{M} of loss functions we have encountered and the number of remaining rounds r :

$$V_{S, \mathcal{F}}(\mathcal{M}, 0) \triangleq - \min_{\mathbf{x} \in S} \sum_{f \in \mathcal{M}} f(\mathbf{x});$$

$$V_{S, \mathcal{F}}(\mathcal{M}, r) \triangleq \min_{\mathbf{x} \in S} \max_{f \in \mathcal{F}} (f(\mathbf{x}) + V_{S, \mathcal{F}}(\mathcal{M} \uplus \{f\}, r - 1)),$$

where \uplus denotes multiset union. We omit the subscript of $V_{S, \mathcal{F}}$ whenever there is no confusion. Let \mathbf{x}_t^T be the output of the minimax algorithm on round t . In other words, \mathbf{x}_t^T realizes the minimum in the definition of $V(f_{1:t-1}, T - t + 1)$. We will adapt the idea in Section 4.1 and study the adaptive algorithm that outputs $\mathbb{E}_{T \sim Q}[\mathbf{x}_t^T | T \geq t] \in S$ on round t for a distribution Q on the horizon. One mild assumption needed is

Assumption 1. $\forall \mathcal{M}$ and $r > 0$, $V(\mathcal{M}, r) \geq V(\mathcal{M}, 0)$.

Roughly speaking, this assumption implies that the game is in the adversary's favor: playing more rounds leads to greater regret. It holds for the Hedge setting with basis vector loss space (see Property 7 in the supplementary file). In fact, it also holds as long as \mathcal{F} contains the zero function $f_0(\mathbf{x}) \equiv 0$. To see this, simply observe that

$$\begin{aligned} V(\mathcal{M}, r) &= \min_{\mathbf{x} \in S} \max_{f \in \mathcal{F}} (f(\mathbf{x}) + V(\mathcal{M} \uplus \{f\}, r - 1)) \\ &\geq V(\mathcal{M} \uplus \{f_0\}, r - 1) \\ &\geq \dots \geq V(\mathcal{M} \uplus \{f_0, \dots, f_0\}, 0) = V(\mathcal{M}, 0). \end{aligned}$$

So the assumption is mild and will hold for all the examples we consider.

Below, we first give a general upper bound on the regret that holds for any distribution and has no dependence on the choices of the adversary. After that we will show what the appropriate distributions are to make this bound $O(\sqrt{T})$.

Theorem 4. Let $\bar{V}_t(\mathcal{M}) = \mathbb{E}_{T \sim Q}[V(\mathcal{M}, T - t + 1) | T \geq t]$ and $q_t = \Pr_{T \sim Q}[T = t | T \geq t]$. Suppose Assumption 1 holds, and on round t the learner chooses $\mathbf{x}_t = \mathbb{E}_{T \sim Q}[\mathbf{x}_t^T | T \geq t]$ where \mathbf{x}_t^T is the output of the minimax algorithm as described above. Then for any T_s , the regret after T_s rounds is at most

$$\bar{V}_1(\emptyset) + \sum_{t=1}^{T_s} q_t \bar{V}_{t+1}(\emptyset).$$

To prove Theorem 4, we first show the following lemma.

Lemma 1. For any $r \geq 0$ and multiset \mathcal{M}_1 and \mathcal{M}_2 ,

$$V(\mathcal{M}_1 \uplus \mathcal{M}_2, r) - V(\mathcal{M}_1, 0) \leq V(\mathcal{M}_2, r). \quad (5)$$

Proof. If $r = 0$, then Eq. (5) holds since

$$\min_{\mathbf{x} \in S} \sum_{f \in \mathcal{M}_1} f(\mathbf{x}) + \min_{\mathbf{x} \in S} \sum_{f \in \mathcal{M}_2} f(\mathbf{x}) \leq \min_{\mathbf{x} \in S} \sum_{f \in \mathcal{M}_1 \uplus \mathcal{M}_2} f(\mathbf{x}).$$

Now assume Eq. (5) holds for $r - 1$. By induction one has

$$\begin{aligned} &V(\mathcal{M}_1 \uplus \mathcal{M}_2, r) - V(\mathcal{M}_1, 0) \\ &= \min_{\mathbf{x} \in S} \max_{f \in \mathcal{F}} (f(\mathbf{x}) + V(\mathcal{M}_1 \uplus \mathcal{M}_2 \uplus \{f\}, r - 1)) - V(\mathcal{M}_1, 0) \\ &\leq \min_{\mathbf{x} \in S} \max_{f \in \mathcal{F}} (f(\mathbf{x}) + V(\mathcal{M}_2 \uplus \{f\}, r - 1)) = V(\mathcal{M}_2, r), \end{aligned}$$

concluding the proof. \square

Proof of Theorem 4. By definition of \mathbf{x}_t^T , we have

$$\begin{aligned} &V(f_{1:t-1}, T - t + 1) \\ &= \max_{f \in \mathcal{F}} (f(\mathbf{x}_t^T) + V(f_{1:t-1} \uplus \{f\}, T - t)) \\ &\geq f_t(\mathbf{x}_t^T) + V(f_{1:t}, T - t). \end{aligned}$$

Therefore, by convexity and the fact that $\Pr[T = t' | T \geq t] = (1 - q_t) \Pr[T = t' | T \geq t + 1]$ for any $t' > t$, the loss of the algorithm on round t is

$$\begin{aligned} f_t(\mathbf{x}_t) &= f_t(\mathbb{E}[\mathbf{x}_t^T | T \geq t]) \leq \mathbb{E}[f_t(\mathbf{x}_t^T) | T \geq t] \\ &\leq \mathbb{E}[V(f_{1:t-1}, T - t + 1) - V(f_{1:t}, T - t) | T \geq t] \\ &= \bar{V}_t(f_{1:t-1}) - q_t V(f_{1:t}, 0) - (1 - q_t) \bar{V}_{t+1}(f_{1:t}) \\ &\leq \bar{V}_t(f_{1:t-1}) - \bar{V}_{t+1}(f_{1:t}) + q_t \bar{V}_{t+1}(\emptyset), \end{aligned}$$

where the last equality holds because $\bar{V}_{t+1}(f_{1:t}) - V(f_{1:t}, 0) = \mathbb{E}[V(f_{1:t}, T - t) - V(f_{1:t}, 0) | T \geq t + 1] \leq \mathbb{E}[V(\emptyset, T - t) | T \geq t + 1] = \bar{V}_{t+1}(\emptyset)$ by Lemma 1. We conclude the proof by summing up $f_t(\mathbf{x}_t)$ over $t = 1, \dots, T_s$ and pointing out that $\bar{V}_{T_s+1}(f_{1:T_s}) = \mathbb{E}[V(f_{1:T_s}, T - T_s) | T \geq T_s + 1] \geq \mathbb{E}[V(f_{1:T_s}, 0) | T \geq T_s + 1] = -\min_{\mathbf{x} \in S} \sum_{t=1}^{T_s} f_t(\mathbf{x}_t)$ by Assumption 1. \square

As a direct corollary, we now show an appropriate choice of Q . We assume that the optimal regret under the fixed horizon setting is of order $O(\sqrt{T})$. That is:

Assumption 2. For any T , $V(\emptyset, T) \leq c_N \sqrt{T}$ for some constant c_N that might depend on N .

This is proven to be true in the literature for all the examples we consider, especially when \mathcal{F} contains linear functions.

Theorem 5. Under Assumption 2 and the same conditions of Theorem 4, if $\Pr[T = t] \propto 1/t^d$ where $d > \frac{3}{2}$ is a constant, then for any T_s , the regret after T_s rounds is at most

$$\frac{\Gamma(d - \frac{3}{2})}{\Gamma(d)} (d - 1)^2 c_N \sqrt{\pi T_s} + o(\sqrt{T_s}),$$

where Γ is the gamma function. Choosing $d \approx 2.35$ approximately minimizes the main term in the bound, leading to regret approximately $3c_N \sqrt{T_s} + o(\sqrt{T_s})$.

Theorem 5 tells us that pretending that the horizon is drawn from the distribution $\Pr[T = t] \propto 1/t^d$ ($d > 3/2$) can always achieve low regret, even if the actual horizon T_s is chosen adversarially. Also notice that the constant 3 in the bound for the term $c_N \sqrt{T_s}$ is less than the one for the doubling trick with the fixed horizon optimal algorithm, which is $2 + \sqrt{2}$ (Cesa-Bianchi & Lugosi, 2006). We will see in Section 6.1 an experiment showing that our algorithm performs much better than the doubling trick.

It is straightforward to apply our new algorithm to different instances of the online convex optimization framework. Examples include Hedge with basis vector loss space, predicting with expert advice (Cesa-Bianchi et al., 1997), online linear optimization within an ℓ_2 ball (Abernethy et al., 2008a) or an ℓ_∞ ball (McMahan & Abernethy, 2013). These are examples where minimax algorithms for fixed

horizon are already known. In theory, however, our algorithm is still applicable when the minimax algorithm is unknown, such as Hedge with the general loss space $[0, 1]^N$.

6. Implementation and Applications

In this section, we discuss the implementation issue of our new algorithm, and also show that the idea of using a “pretend prior distribution” is much more applicable in online learning than we have discussed so far.

6.1. Closed Form of the Algorithm

Among the examples listed at the end of Section 5, we are especially interested in online linear optimization within an ℓ_2 ball since our algorithm enjoys an explicit closed form in this case. Specifically, we consider the following problem (all the norms are ℓ_2 norms): take $S = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\| \leq 1\}$, and $\mathcal{F} = \{f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w} : \mathbf{w} \in S\}$. In other words, the adversary also chooses a point in S on each round, which we denote by \mathbf{w}_t . Abernethy et al. (2008a) showed a simple but exact minimax optimal algorithm for the fixed horizon setting (for $N > 2$): on each round t , choose

$$\mathbf{x}_t^T = -\mathbf{W}_{t-1} / \sqrt{\|\mathbf{W}_{t-1}\|^2 + (T - t + 1)}, \quad (6)$$

where $\mathbf{W}_t = \sum_{t'=1}^t \mathbf{w}_{t'}$. This strategy guarantees the regret to be at most \sqrt{T} . To make this algorithm adaptive, we again assign a distribution over the horizon. However, in order to get an explicit form for $\mathbb{E}[\mathbf{x}_t^T | T \geq t]$, a continuous distribution on T is necessary. It does not seem to make sense at first glance since the horizon is always an integer, but keep in mind that the random variable T is merely an artifact of our algorithm, and Eq. (6) is well defined with $T \geq t$ being a real number. As long as the output of the learner is in the set S , our algorithm is valid. The analysis for our algorithm also holds with minor changes. Specifically, we show the following:

Theorem 6. Let $T \geq 1$ be a continuous random variable with probability density $f(T) \propto 1/T^2$. If the learner chooses $\mathbf{x}_t = \mathbb{E}[\mathbf{x}_t^T | T \geq t]$ on round t , where \mathbf{x}_t^T is defined by Eq. (6), then the regret after T_s rounds is at most $\pi \sqrt{T_s} + o(\sqrt{T_s})$ for any T_s . Moreover, \mathbf{x}_t has the following explicit form

$$\mathbf{x}_t = \begin{cases} \left(\frac{t \tanh^{-1}(\sqrt{1-t/c})}{(c-t)^{3/2}} - \frac{\sqrt{c}}{c-t} \right) \mathbf{W}_{t-1} & \text{if } c \neq t \\ -\frac{2t}{3c^{3/2}} \mathbf{W}_{t-1} & \text{else,} \end{cases} \quad (7)$$

where $c = 1 + \|\mathbf{W}_{t-1}\|^2$.

The algorithm we are proposing in Eq. (7) looks quite inexplicable if one does not realize that it comes from the expression $\mathbb{E}[\mathbf{x}_t^T | T \geq t]$ with an appropriate distribution. Yet

the algorithm not only enjoys a low theoretic regret bound as shown in Theorem 6, but also achieves very good performance in simulated experiments.

To show this, we conduct an experiment that compares the regrets of four algorithms at any time step within 1000 rounds against an adversary that chooses points in S uniformly at random ($N = 10$). The results are shown in Figure 1, where each data point is the maximum regret over 1000 randomly generated adversaries for the corresponding algorithm and horizon. The four algorithms are: the minimax algorithm in Eq. (6) (OPT); the one we proposed in Theorem 6 (DIST); online gradient descent, a general algorithm for online optimization (see Zinkevich, 2003) (OGD); and the doubling trick with the minimax algorithm (DOUBLE). Note that OPT is not really an adaptive algorithm: it “cheats” by knowing the horizon $T = 1000$ in advance, and thus performs best at the end of the game. We include this algorithm merely as a baseline. Figure 1 shows that our algorithm DIST achieves consistently much lower regret than any other adaptive algorithm, including OGD which seems to enjoy a better constant in the regret bound ($2\sqrt{2T_s}$, see Zinkevich, 2003). Moreover, for the first 450 rounds or so, our algorithm performs even better than OPT, implying that using the optimal algorithm with a large guess on the horizon is inferior to our algorithm. Finally, we remark that although the doubling trick is widely applicable in theory, in experiments it is beaten by most of the other algorithms.

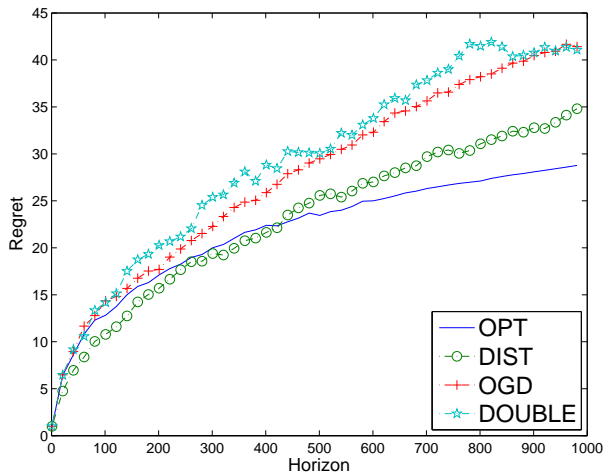


Figure 1. Comparison of four algorithms.

6.2. Randomized Play and Efficient Implementation

Implementation is an issue for our algorithm when there is no closed form for $\mathbb{E}[\mathbf{x}_t^T | T \geq t]$, which is usually the case. One way to address this problem is to compute the sum of

the first sufficient number of terms in the series, which can be a good estimate since the weight for each term decreases rapidly.

However, there is another more natural way to deal with the implementation issue when we are in a similar setting but allowed to play randomly. Specifically, consider a modified Hedge setting where on each round t , the learner can bet on one and only one action I_t , and then the loss vector $\mathbf{Z}_t \in [0, 1]^N$ is revealed with the learner suffering loss Z_{t,I_t} for this round. It is well known that in this kind of problem, randomization is necessary for the learner to achieve sub-linear regret. That is, I_t is a random variable and \mathbf{Z}_t is decided without knowing the actual draw of I_t . In addition, suppose \mathbf{P}_t , the conditional distribution of I_t given the past, only depends on $\mathbf{Z}_{1:t-1}$, and the learner achieves sub-linear regret in the usual Hedge setting (sometimes called *pseudo-regret*):

$$\sum_{t=1}^T \mathbf{P}_t \cdot \mathbf{Z}_t - \min_i M_{T,i} \leq c_N \sqrt{T} \quad (8)$$

(recall $\mathbf{M}_t = \sum_{t'=1}^t \mathbf{Z}_{t'}$) for any $\mathbf{Z}_{1:T}$ and a constant c_N . Then the learner also achieves sub-linear regret with high probability in the randomized setting. That is, with probability at least $1 - \delta$, the actual regret satisfies:

$$\sum_{t=1}^T Z_{t,I_t} - \min_i M_{T,i} \leq c_N \sqrt{T} + \sqrt{\frac{T}{2} \ln \frac{1}{\delta}}. \quad (9)$$

We refer the interested reader to Lemma 4.1 of Cesa-Bianchi & Lugosi (2006) for more details.

Therefore, in this setting we can implement our algorithm in an efficient way: on round t , first draw a horizon $T \geq t$ according to distribution $\Pr[T = t'] \propto 1/t'^d$, then draw I_t according to \mathbf{P}_t^T . It is clear that the marginal distribution of I_t of this process is exactly $\mathbb{E}[\mathbf{P}_t^T | T \geq t]$. Hence, Eq. (8) is satisfied by Theorem 5 and as a result Eq. (9) holds.

6.3. Combining with the FPL algorithm

Even if we have an efficient randomized implementation, or sometimes even have a closed form of the output, it is still too constrained if we can only apply our technique to minimax algorithms since they are usually difficult to derive and sometimes even inefficient to implement. It turns out, however, that the “pretend prior distribution” idea is applicable for many other non-minimax algorithms, which we will discuss from this section on.

Continuing the randomized setting discussed in the previous section, we study the well-known “follow the perturbed leader (FPL)” algorithm (Kalai & Vempala, 2005), which chooses $I_t \in \arg \min_i (M_{t-1,i} + \xi_{t,i})$ where $\xi_t \in R^N$ is a random variable drawn from some distribution. This distribution sometimes requires the horizon T as a parameter. If

this is the case, applying our technique would have a simple *Bayesian interpretation*: put a prior distribution on an unknown parameter of another distribution. Working out the marginal distribution of ξ_t would then give an adaptive variant of FPL.

For simplicity, consider drawing ξ_t^T uniformly at random from the hypercube $[0, \Delta_T]^N$ (see Chapter 4.3 of Cesa-Bianchi & Lugosi, 2006). If $\Delta_T = \sqrt{TN}$, then the pseudo-regret is upper bounded by $2\sqrt{TN}$ (whose dependence on N is not optimal). Now again let $T \geq 1$ be a continuous random variable with probability density $f(T) \propto 1/T^d$ ($d > 3/2$), and ξ_t be obtained by first drawing T given $T \geq t$, and then drawing a point uniformly from $[0, \Delta_T]^N$. We show the following:

Lemma 2. *If $\Delta_t = \sqrt{btN}$ for some constant $b > 0$, the marginal density function of ξ_t is*

$$f_t(\xi) \propto \begin{cases} 0 & \text{if } \min_i \xi_i < 0 \\ \min \left\{ 1, \left(\frac{\Delta_t}{\|\xi\|_\infty} \right)^{2d-2+N} \right\} & \text{else.} \end{cases} \quad (10)$$

The normalization factor is $\frac{d-1}{d-1+N/2} \Delta_t^{-N}$.

Theorem 7. *Suppose on round t , the learner chooses*

$$I_t \in \arg \min_i (M_{t-1,i} + \xi_{t,i}),$$

where ξ_t is a random variable with density function (10). Then the pseudo-regret after T_s rounds is at most

$$\left(\frac{d-1}{\sqrt{b}(d-1/2)} + \frac{\sqrt{b}(d-1)^2}{d-3/2} \right) 2\sqrt{T_s N}.$$

Choosing $b = \frac{d-3/2}{(d-1/2)(d-1)}$ and $d = 1 + \frac{\sqrt{3}}{2}$ minimizes the main term in the bound, leading to about $4.6\sqrt{T_s N}$.

By the exact same argument, the actual regret is bounded by the same quantity plus $\sqrt{\frac{T}{2} \ln \frac{1}{\delta}}$ with probability $1 - \delta$.

6.4. Generalizing the Exponential Weights Algorithm

Now we come back to the usual Hedge setting and consider another popular non-minimax algorithm (note that it is trivial to generalize the results to the randomized setting). When dealing with the most general loss space $[0, 1]^N$, the minimax algorithm is unknown even for the fixed horizon setting. However, generalizing the weighted majority algorithm of Littlestone & Warmuth (1994), Freund & Schapire (1997; 1999) presented an algorithm using exponential weights that can deal with this general loss space and achieve the $O(\sqrt{T \ln N})$ bound on the regret. The algorithm takes the horizon T as a parameter, and on round t , it simply chooses $P_{t,i} \propto \exp(-\eta M_{t-1,i})$, where $\eta = \sqrt{(8 \ln N)/T}$ is the learning rate. It is shown

that the regret of this algorithm is at most $\sqrt{(T \ln N)/2}$. Auer et al. (2002) proposed a way to make this algorithm adaptive by simply setting a time-varying learning rate $\eta = \sqrt{(8 \ln N)/t}$, where t is the current round, leading to a regret bound of $\sqrt{T \ln N}$ for any T (see Chapter 2.5 of Bubeck, 2011). In other words, the algorithm always treats the current round as the last round. Below, we show that our “pretend distribution” idea can also be used to make this exponential weights algorithm adaptive, and is in fact a generalization of the adaptive learning rate algorithm by Auer et al. (2002).

Theorem 8. *Let $\mathbf{LS} = [0, 1]^N$, $\Pr[T = t] \propto 1/t^d$ ($d > 3/2$) and $\eta_T = \sqrt{(b \ln N)/T}$, where b is a constant. If on round t , the learner assigns weight $\mathbb{E}_{T \sim Q}[P_{t,i}^T | T \geq t]$ to each action i , where $P_{t,i}^T \propto \exp(-\eta_T M_{t-1,i})$, then for any T_s , the regret after T_s rounds is at most*

$$\left(\frac{\sqrt{b}(d-1)}{4(d-1/2)} + \frac{d-1}{(d-3/2)\sqrt{b}} \right) \sqrt{T_s \ln N} + o(\sqrt{T_s \ln N}).$$

Setting $b = \frac{4d-2}{d-3/2}$ minimizes the main term, which approaches 1 as $d \rightarrow \infty$.

Note that if $d \rightarrow \infty$, our algorithm simply becomes the one of Auer et al. (2002), because $\Pr[T = \tau | T \geq t]$ is 1 if $\tau = t$ and 0 otherwise. Therefore, our algorithm can be viewed as a generalization of the idea of treating the current round as the last round. However, we emphasize that the way we deal with unknown horizon is more applicable in the sense that if we try to make a minimax algorithm adaptive by treating each round as the last round, one can construct an adversary that leads to linear—and therefore grossly suboptimal—regret, whereas our approach yields nearly optimal regret. (See Example 2 and 3 in the supplementary file for details.)

6.5. First Order Regret Bound

So far all the regret bounds we have discussed are in terms of the horizon, which are also called *zeroth order bounds*. More refined bounds have been studied in the literature (Cesa-Bianchi & Lugosi, 2006). For example, the *first order bound* for Hedge, that depends on the loss of the best action m^* at the end of the game, usually is of order $O(\sqrt{m^* \ln N})$. Again, using the exponential weights algorithm with a slightly different learning rate $\eta = \ln(1 + \sqrt{(2 \ln N)/m^*})$, one can show that the regret is at most $\sqrt{2m^* \ln N} + \ln N$. Here, m^* is prior information on the loss sequence similar to the horizon. To avoid exploiting this information that is unavailable in practice, one can again use techniques like the doubling trick or the time-varying learning rate. Alternatively, we show that the “pretend distribution” technique can also be used here. Again it makes more sense to assign a continuous distribution on the loss of the best action instead of a discrete one.

Theorem 9. Let $\mathbf{LS} = [0, 1]^N$, $m_t = \min_i M_{t,i} + 1$, $\eta_m = \sqrt{(\ln N)/m}$, and $m \geq 1$ be a continuous random variable with probability density $f(m) \propto 1/m^d$ ($d > 3/2$). If on round t , the learner assigns weight $\mathbb{E}[P_{t,i}^m | m \geq m_{t-1}]$ to each action i , where $P_{t,i}^m \propto \exp(-\eta_m M_{t-1,i})$, then for any T_s , the regret after T_s rounds is at most

$$\frac{3(d-7/6)(d-1)}{(d-3/2)(d-1/2)} \sqrt{m^* \ln N} + (1 + (d-1) \ln(m^* + 1)) \ln N + o(\sqrt{m^* \ln N}),$$

where $m^* = \min_i M_{T_s,i}$ is the loss of the best action after T_s rounds. Setting $d = 5/2 + \sqrt{2}$ minimizes the main term, which becomes $(3/2 + \sqrt{2})\sqrt{m^* \ln N}$.

References

- Abernethy, Jacob and Warmuth, Manfred K. Repeated games against budgeted adversaries. In *Advances in Neural Information Processing Systems 24*, 2010.
- Abernethy, Jacob, Bartlett, Peter L., Rakhlin, Alexander, and Tewari, Ambuj. Optimal strategies and minimax lower bounds for online convex games. In *Proceedings of the 21st Annual Conference on Learning Theory*, 2008a.
- Abernethy, Jacob, Warmuth, Manfred K., and Yellin, Joel. Optimal strategies from random walks. In *Proceedings of the 21st Annual Conference on Learning Theory*, 2008b.
- Auer, Peter, Cesa-Bianchi, Nicolò, and Gentile, Claudio. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.
- Berend, Daniel and Kontorovich, Aryeh. On the concentration of the missing mass. *Electron. Commun. Probab.*, 18:no. 3, 1–7, 2013. ISSN 1083-589X. doi: 10.1214/ECP.v18-2359.
- Bousquet, Olivier and Warmuth, Manfred K. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research*, 3:363–396, 2003.
- Bubeck, Sébastien. Introduction to online optimization. Lecture notes, available at <http://www.princeton.edu/~sbubeck/BubeckLectureNotes>, 2011.
- Cesa-Bianchi, Nicolò and Lugosi, Gábor. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
- Cesa-Bianchi, Nicolò, Freund, Yoav, Haussler, David, Helmbold, David P., Schapire, Robert E., and Warmuth, Manfred K. How to use expert advice. *Journal of the ACM*, 44(3):427–485, May 1997.
- Chaudhuri, Kamalika, Freund, Yoav, and Hsu, Daniel. A parameter-free hedging algorithm. *Advances in Neural Information Processing Systems 23*, 2009.
- Foster, Dean P. and Vohra, Rakesh V. Asymptotic calibration. *Biometrika*, 85(2):379–390, 1998.
- Freund, Yoav and Schapire, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.
- Freund, Yoav and Schapire, Robert E. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

Gentile, Claudio. The robustness of the p-norm algorithms. *Machine Learning*, 53(3):265–299, 2003.

Gofer, Eyal and Mansour, Yishay. Lower bounds on individual sequence regret. In *Algorithmic Learning Theory*, pp. 275–289. Springer, 2012.

Hazan, Elad and Seshadhri, C. Adaptive algorithms for online decision problems. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007.

Herbster, Mark and Warmuth, Manfred. Tracking the best expert. In *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 286–294, 1995.

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

Lehmer, D. H. Interesting series involving the central binomial coefficient. *The American Mathematical Monthly*, 92(7):449–457, 1985.

Littlestone, Nick and Warmuth, Manfred K. The weighted majority algorithm. *Information and Computation*, 108: 212–261, 1994.

McMahan, H. Brendan and Abernethy, Jacob. Minimax optimal algorithms for unconstrained linear optimization. In *Advances in Neural Information Processing Systems 27*, 2013.

Rockafellar, R. Tyrrell. *Convex Analysis*. Princeton University Press, 1970.

Shalev-Shwartz, Shai. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2011.

Yaroshinsky, Rani, El-Yaniv, Ran, and Seiden, Steven S. How to better use expert advice. *Machine Learning*, 55 (3):271–309, 2004.

Zinkevich, Martin. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning*, 2003.

Through all the proofs, we denote the set $\{1, \dots, m\}$ by $[m]$.

A. Proof of Theorem 1

We first state a few properties of the function R :

Proposition 1. For any vector \mathbf{M} of N dimensions and integer r ,

Property 1. $R(\mathbf{M}, r) = a + R((M_1 - a, \dots, M_N - a), r)$ for any real number a and $r \geq 0$.

Property 2. $R(\mathbf{M}, r)$ is non-decreasing in M_i for each $i = 1, \dots, N$.

Property 3. If $r > 0$, $R(\mathbf{M}, r) - R(\mathbf{M}, r - 1) \leq 1/N$.

Property 4. If $r > 0$, and $P_i = \frac{1}{N} + R(\mathbf{M} + \mathbf{e}_i, r - 1) - R(\mathbf{M}, r)$ for each $i = 1, \dots, N$, then $\mathbf{P} = (P_1, \dots, P_N)$ is a distribution in the simplex $\Delta(N)$.

Proof of Proposition 1. We omit the proof for Property 1 and 2, since it is straightforward. We prove Property 3 by induction. For the base case $r = 1$, let $S = \{j : M_j = \min_i M_i\}$. If $|S| = 1$, then $R(\mathbf{M} + \mathbf{e}_i, 0)$ is $R(\mathbf{M}, 0)$ for $i \notin S$ and $R(\mathbf{M}, 0) + 1$ otherwise. If $|S| > 1$, then $R(\mathbf{M} + \mathbf{e}_i, 0)$ is simply $R(\mathbf{M}, 0)$ for all i . In either case, we have

$$\begin{aligned} R(\mathbf{M}, 1) &= \frac{1}{N} \sum_{i=1}^N R(\mathbf{M} + \mathbf{e}_i, 0) \leq \frac{1}{N} (1 + \sum_{i=1}^N R(\mathbf{M}, 0)) \\ &= \frac{1}{N} + R(\mathbf{M}, 0), \end{aligned}$$

proving the base case. Now for $r > 1$, by definition of R and induction,

$$\begin{aligned} &R(\mathbf{M}, r) - R(\mathbf{M}, r - 1) \\ &= \frac{1}{N} \sum_{j=1}^N (R(\mathbf{M} + \mathbf{e}_j, r - 1) - R(\mathbf{M} + \mathbf{e}_j, r - 2)) \\ &\leq \frac{1}{N} \sum_{j=1}^N \frac{1}{N} = \frac{1}{N}, \end{aligned}$$

completing the induction. For Property 4, it suffices to prove $P_i \geq 0$ for each i and $\sum_{i=1}^N P_i = 1$. The first part can be shown using Property 2 and 3:

$$\begin{aligned} P_i &= \frac{1}{N} + R(\mathbf{M} + \mathbf{e}_i, r - 1) - R(\mathbf{M}, r) \\ &\geq \frac{1}{N} + R(\mathbf{M}, r - 1) - \left(\frac{1}{N} + R(\mathbf{M}, r - 1)\right) = 0. \end{aligned}$$

The second part is also easy to show by definition of R :

$$\begin{aligned} \sum_{i=1}^N P_i &= 1 + \sum_{i=1}^N R(\mathbf{M} + \mathbf{e}_i, r - 1) - NR(\mathbf{M}, r) \\ &= 1 + NR(\mathbf{M}, r) - NR(\mathbf{M}, r) = 1. \end{aligned}$$

□

Proof of Theorem 1. First inductively prove $V(\mathbf{M}, r) = r/N - R(\mathbf{M}, r)$ for any $r \geq 0$. The base case $r = 0$ is trivial by definition. For $r > 0$,

$$\begin{aligned} V(\mathbf{M}, r) &= \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \text{LS}} (\mathbf{P} \cdot \mathbf{Z} + V(\mathbf{M} + \mathbf{Z}, r - 1)) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} (P_i + V(\mathbf{M} + \mathbf{e}_i, r - 1)) \\ &\quad (\text{LS} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} \left(P_i + \frac{r-1}{N} - R(\mathbf{M} + \mathbf{e}_i, r - 1) \right) \\ &\quad (\text{by induction}) \end{aligned}$$

Denote $P_i + (r-1)/N - R(\mathbf{M} + \mathbf{e}_i, r - 1)$ by $g(\mathbf{P}, i)$. Notice that the average of $g(\mathbf{P}, i)$ over all i is irrelevant to \mathbf{P} : $\frac{1}{N} \sum_{i=1}^N g(\mathbf{P}, i) = r/N - R(\mathbf{M}, r)$. Therefore, $\max_i g(\mathbf{P}, i) \geq r/N - R(\mathbf{M}, r)$ for any \mathbf{P} , and

$$V(\mathbf{M}, r) = \min_{\mathbf{P}} \max_i g(\mathbf{P}, i) \geq r/N - R(\mathbf{M}, r). \quad (11)$$

On the other hand, from Proposition 1, we know that $P_i^* = 1/N + R(\mathbf{M} + \mathbf{e}_i, r - 1) - R(\mathbf{M}, r)$ ($i \in [N]$) is a valid distribution. Also,

$$\begin{aligned} V(\mathbf{M}, r) &= \min_{\mathbf{P}} \max_i g(\mathbf{P}, i) \leq \max_i g(\mathbf{P}^*, i) \\ &= \max_i \left(\frac{r}{N} - R(\mathbf{M}, r) \right) = \frac{r}{N} - R(\mathbf{M}, r). \end{aligned} \quad (12)$$

So from Eq. (11) and (12) we have $V(\mathbf{M}, r) = r/N - R(\mathbf{M}, r)$, and also $P_i^* = 1/N + R(\mathbf{M} + \mathbf{e}_i, r - 1) - R(\mathbf{M}, r) = V(\mathbf{M}, r) - V(\mathbf{M} + \mathbf{e}_i, r - 1)$ realizes the minimum, and thus is the optimal strategy.

It remains to prove $V(\mathbf{0}, T) \leq c_N \sqrt{T}$. Let $\mathbf{Z}_1, \dots, \mathbf{Z}_T$ be independent uniform random variables taking values in $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$. By what we proved above,

$$V(\mathbf{0}, T) = \frac{T}{N} - \mathbb{E}[\min_{i \in [N]} \sum_{t=1}^T Z_{t,i}] = \mathbb{E}[\max_{i \in [N]} \sum_{t=1}^T (1/N - Z_{t,i})].$$

Let $y_{t,i} = 1/N - Z_{t,i}$. Then each $y_{t,i}$ is a random variable that takes value $1/N$ with probability $1 - 1/N$ and $1/N - 1$ with probability $1/N$. Also, for a fixed i , $y_{1,i}, \dots, y_{T,i}$ are independent (note that this is not true for $y_{t,1}, \dots, y_{t,N}$ for a fixed t). It is shown in Lemma 3.3 of Berend & Kontorovich (2013) that each $y_{t,i}$ satisfies

$$\mathbb{E}[\exp(\lambda y_{t,i})] \leq \exp\left(\frac{\lambda^2 \sigma^2}{2}\right), \quad \forall \lambda > 0,$$

where $\sigma^2 = (N-1)/N^2$ is the variance of $y_{t,i}$. So if we let $Y_i = \sum_{t=1}^T y_{t,i}$, by the independence of each term, we have $\forall \lambda > 0$,

$$\mathbb{E}[\exp(\lambda Y_i)] = \mathbb{E}\left[\prod_{t=1}^T \exp(\lambda y_{t,i})\right] = \prod_{t=1}^T \mathbb{E}[\exp(\lambda y_{t,i})]$$

$$\leq \exp\left(\frac{\lambda^2 \sigma^2 T}{2}\right).$$

Now using Lemma A.13 from Cesa-Bianchi & Lugosi (2006), we arrive at

$$\mathbb{E}[\max_{i \in [N]} Y_i] \leq \sigma \sqrt{2T \ln N} = c_N \sqrt{T}.$$

We conclude the proof by pointing out

$$V(\mathbf{0}, T) = \mathbb{E}[\max_{i \in [N]} \sum_{t=1}^T (1/N - Z_{t,i})] = \mathbb{E}[\max_{i \in [N]} Y_i] \leq c_N \sqrt{T}. \quad \square$$

As a direct corollary of Proposition 1 and Theorem 1, below we list a few properties of the function V for later use.

Proposition 2. *If $\text{LS} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$, then for any vector \mathbf{M} and integer r ,*

Property 5. $V(\mathbf{M}, r) = V((M_1 - a, \dots, M_N - a), r) - a$ for any real number a and $r \geq 0$.

Property 6. $V(\mathbf{M}, r)$ is non-increasing in M_i for each $i = 1, \dots, N$.

Property 7. $V(\mathbf{M}, r)$ is non-decreasing in r .

B. Proof of Theorem 2

Proof. Define $\bar{V}_t(\mathbf{M}) = \mathbb{E}[V(\mathbf{M}, T - t + 1) | T \geq t]$ and $q(t', t) = \Pr[T = t' | T \geq t]$. We will prove an important property of the \bar{V} function:

$$\begin{aligned} \bar{V}_t(\mathbf{M}) &= \min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} (P_i + q(t, t) V(\mathbf{M} + \mathbf{e}_i, 0) + \\ &\quad (1 - q(t, t)) \bar{V}_{t+1}(\mathbf{M} + \mathbf{e}_i)). \end{aligned} \quad (13)$$

This equation shows that $\bar{V}_t(\mathbf{M})$ is the conditional expectation of the regret given $T \geq t$, starting from cumulative loss vector \mathbf{M} and assuming both the learner and the adversary are optimal. This is similar to the function V in the fixed horizon case, and again the value of the game $\inf_{\text{Alg}} \sup_{\mathbf{Z}_{1:\infty}} \mathbb{E}_{T \sim Q}[\mathbf{Reg}(L_T, \mathbf{M}_T)]$ is simply $\bar{V}_1(\mathbf{0})$.

To prove Eq. (13), we plug the definition of $\bar{V}_{t+1}(\mathbf{M} + \mathbf{e}_i)$ into the right hand side and get $\min_{\mathbf{P}} \max_i g(\mathbf{P}, i)$ where $g(\mathbf{P}, i) = P_i + q(t, t) V(\mathbf{M} + \mathbf{e}_i, 0) + (1 - q(t, t)) \mathbb{E}[V(\mathbf{M} + \mathbf{e}_i, T - t) | T \geq t + 1]$. Using the fact that for any $t' \geq t + 1$,

$$\begin{aligned} &(1 - q(t, t)) q(t', t + 1) \\ &= \Pr[T > t | T \geq t] \Pr[T = t' | T \geq t + 1] \\ &= \Pr[T = t' | T \geq t] = q(t', t), \end{aligned}$$

$g(\mathbf{P}, i)$ can be simplified in the following way:

$$\begin{aligned}
 & g(\mathbf{P}, i) & (14) \\
 & = P_i + q(t, t)V(\mathbf{M} + \mathbf{e}_i, 0) + \\
 & \quad (1 - q(t, t)) \sum_{T=t+1}^{\infty} (q(T, t+1)V(\mathbf{M} + \mathbf{e}_i, T - t)) \\
 & = P_i + q(t, t)V(\mathbf{M} + \mathbf{e}_i, 0) + \\
 & \quad \sum_{T=t+1}^{\infty} (q(T, t)V(\mathbf{M} + \mathbf{e}_i, T - t)) \\
 & = P_i + \mathbb{E}[V(\mathbf{M} + \mathbf{e}_i, T - t) | T \geq t]. & (15)
 \end{aligned}$$

Also, the average of $g(\mathbf{P}, i)$ over all i is independent of \mathbf{P} :

$$\begin{aligned}
 & \frac{1}{N} \sum_{i=1}^N g(\mathbf{P}, i) \\
 & = \frac{1}{N} + \frac{1}{N} \sum_i^N \mathbb{E}[V(\mathbf{M} + \mathbf{e}_i, T - t) | T \geq t] \\
 & = \mathbb{E} \left[\frac{1}{N} + \frac{1}{N} \sum_i^N V(\mathbf{M} + \mathbf{e}_i, T - t) | T \geq t \right] \\
 & = \mathbb{E} \left[\frac{1}{N} + \frac{1}{N} \sum_i^N \left(\frac{T-t}{N} - R(\mathbf{M} + \mathbf{e}_i, T - t) \right) | T \geq t \right] \\
 & = \mathbb{E} \left[\frac{T-t+1}{N} - R(\mathbf{M}, T-t+1) | T \geq t \right] \\
 & \hspace{15em} \text{(by definition of R)}
 \end{aligned}$$

$$= \mathbb{E}[V(\mathbf{M}, T-t+1) | T \geq t],$$

which implies

$$\min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} g(\mathbf{P}, i) \geq \mathbb{E}[V(\mathbf{M}, T-t+1) | T \geq t]. \quad (16)$$

On the other hand, let $\mathbf{P}^* = \mathbb{E}[\mathbf{P}^T | T \geq t]$, where $P_i^T = V(\mathbf{M}, T-t+1) - V(\mathbf{M} + \mathbf{e}_i, T-t)$. \mathbf{P}^* is a valid distribution since \mathbf{P}^T is a distribution for any T . Also, by plugging into Eq. (15),

$$\begin{aligned}
 g(\mathbf{P}^*, i) & = \mathbb{E}[V(\mathbf{M}, T-t+1) - V(\mathbf{M} + \mathbf{e}_i, T-t) | T \geq t] \\
 & \quad + \mathbb{E}[V(\mathbf{M} + \mathbf{e}_i, T-t) | T \geq t] \\
 & = \mathbb{E}[V(\mathbf{M}, T-t+1) | T \geq t].
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} g(\mathbf{P}, i) & \leq \max_{i \in [N]} g(\mathbf{P}^*, i) \\
 & = \mathbb{E}[V(\mathbf{M}, T-t+1) | T \geq t]. & (17)
 \end{aligned}$$

Eq. (16) and (17) show that $\min_{\mathbf{P}} \max_i g(\mathbf{P}, i) = \mathbb{E}[V(\mathbf{M}, T-t+1) | T \geq t]$, which agrees with the left hand side of Eq. (13). We thus prove $\inf \sup_{\mathbf{Alg} \mathbf{Z}_{1:\infty}} \mathbb{E}_{T \sim Q} [\mathbf{Reg}(L_T, \mathbf{M}_T)] = \mathbb{E}[V(\mathbf{0}, T) | T \geq 1] = \mathbf{Alg} \mathbf{Z}_{1:T} [\inf \sup \mathbf{Reg}(L_T, \mathbf{M}_T)]$, and \mathbf{P}^* is the optimal strategy. \square

C. Proof of Theorem 3

To prove Theorem 3, we need to find out what $V(\mathbf{0}, T)$ is under the general loss space $[0, 1]^2$. Note that Theorem 1 only tells us the case of the basis vector loss space. Fortunately, it turns out that they are the same if $N = 2$. To be more specific, we will show later in Theorem 10 that if $N = 2$ and $\mathbf{LS} = [0, 1]^2$, then $V(\mathbf{0}, T) = T/2 - R(\mathbf{0}, T)$, which can be further simplified as

$$\begin{aligned}
 V(\mathbf{0}, T) & = \frac{T}{2} - \frac{1}{2^T} \sum_{m=0}^T \binom{T}{m} \min\{m, T-m\} \\
 & = \frac{T}{2^T} \binom{T-1}{\lfloor \frac{T}{2} \rfloor}.
 \end{aligned}$$

We can now prove Theorem 3 using this explicit scaling factor, denoted by $S(T)$ for simplicity.

Proof of Theorem 3. Again, solving Eq. (4) is equivalent to finding the value function \tilde{V} defined on each state of the game, similar to the functions V and \bar{V} we had before. The difference is that \tilde{V} should be a function of not only the index of the current round t and the cumulative loss vector \mathbf{M} , but also the cumulative loss L for the learner. Moreover, to obtain a base case for the recursive definition, it is convenient to first assume that T is at most T_0 , where T_0 is some fixed integer. Under these conditions, we define $\tilde{V}_t^{T_0}(L, \mathbf{M})$ recursively as:

$$\begin{aligned}
 \tilde{V}_{T_0}^{T_0}(L, \mathbf{M}) & \triangleq \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}} \frac{\mathbf{Reg}(L + \mathbf{P} \cdot \mathbf{Z}, \mathbf{M} + \mathbf{Z})}{V(\mathbf{0}, T_0)}, \\
 \tilde{V}_t^{T_0}(L, \mathbf{M}) & \triangleq \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}} \max \left\{ \frac{\mathbf{Reg}(L + \mathbf{P} \cdot \mathbf{Z}, \mathbf{M} + \mathbf{Z})}{V(\mathbf{0}, t)}, \right. \\
 & \quad \left. \tilde{V}_{t+1}^{T_0}(L + \mathbf{P} \cdot \mathbf{Z}, \mathbf{M} + \mathbf{Z}) \right\},
 \end{aligned}$$

which is the scaled regret starting from round t with cumulative loss L for the learner and \mathbf{M} for the actions, assuming both the learner and the adversary will play optimally from this round on. The value of the game \tilde{V} is now $\lim_{T_0 \rightarrow +\infty} \tilde{V}_1^{T_0}(\mathbf{0}, \mathbf{0})$.

To simplify this value function, we will need three facts. First, the base case can be related to $V(\mathbf{M}, 1)$:

$$\begin{aligned}
 & \tilde{V}_{T_0}^{T_0}(L, \mathbf{M}) \\
 & = \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}} \frac{\mathbf{Reg}(L + \mathbf{P} \cdot \mathbf{Z}, \mathbf{M} + \mathbf{Z})}{V(\mathbf{0}, T_0)} \\
 & = \left(L + \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}} \mathbf{Reg}(\mathbf{P} \cdot \mathbf{Z}, \mathbf{M} + \mathbf{Z}) \right) / V(\mathbf{0}, T_0) \\
 & = \left(L + \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}} \mathbf{P} \cdot \mathbf{Z} + V(\mathbf{M} + \mathbf{Z}, 0) \right) / V(\mathbf{0}, T_0) \\
 & = \frac{L + V(\mathbf{M}, 1)}{V(\mathbf{0}, T_0)}.
 \end{aligned}$$

Second, for any L and \mathbf{M} , one can inductively show that

$$\tilde{V}_t^{T_0}(L, \mathbf{M}) = \tilde{V}_t^{T_0}(L - R(\mathbf{M}, 0), \mathbf{M}'), \quad (18)$$

where $M'_i = M_i - R(\mathbf{M}, 0)$. (We omit the details since it is straightforward.)

Third, when $\mathbf{M} = \mathbf{0}$, by symmetry, one has with $\mathbf{P}_u = (\frac{1}{N}, \dots, \frac{1}{N})$

$$\begin{aligned} & \tilde{V}_t^{T_0}(L, \mathbf{0}) \\ &= \max_{\mathbf{Z} \in \mathbf{LS}} \max \left\{ \frac{\mathbf{Reg}(L + \mathbf{P}_u \cdot \mathbf{Z}, \mathbf{Z})}{V(\mathbf{0}, t)}, \tilde{V}_{t+1}^{T_0}(L + \mathbf{P}_u \cdot \mathbf{Z}, \mathbf{Z}) \right\} \\ &\geq \max \left\{ \frac{L + \frac{1}{N}}{V(\mathbf{0}, t)}, \tilde{V}_{t+1}^{T_0}(L + \frac{1}{N}, \mathbf{e}_1) \right\}. \end{aligned} \quad (19)$$

Now we can make use of the condition $N = 2$ to lower bound \tilde{V} . The key point is to consider a restricted adversary who can only place one unit more loss on one of the action than the other, if not stopping the game. Clearly the value of this restricted game serves as a lower bound of \tilde{V} . Specifically, consider the value of $\tilde{V}_t^{T_0}(L, \mathbf{e}_1)$ for $t \leq T_0 - 2$:

$$\begin{aligned} & \tilde{V}_t^{T_0}(L, \mathbf{e}_1) \\ &\geq \min_{p \in [0, 1]} \max \left\{ \frac{\mathbf{Reg}(L + p, 2\mathbf{e}_1)}{S(t)}, \tilde{V}_{t+1}^{T_0}(L + 1 - p, \mathbf{e}_1 + \mathbf{e}_2) \right\} \\ &\quad \text{(restricted adversary)} \\ &= \min_{p \in [0, 1]} \max \left\{ \frac{L + p}{S(t)}, \tilde{V}_{t+1}^{T_0}(L - p, \mathbf{0}) \right\} \quad \text{(by Eq. (18))} \\ &\geq \min_{p \in [0, 1]} \max \left\{ \frac{L + p}{S(t)}, \frac{L + 1/2 - p}{S(t+1)}, \tilde{V}_{t+2}^{T_0}(L + \frac{1}{2} - p, \mathbf{e}_1) \right\} \\ &\quad \text{(by Eq. (19))} \\ &\geq \min_{p \in \mathbb{R}} \max \left\{ \frac{L + p}{S(t)}, \tilde{V}_{t+2}^{T_0}(L + \frac{1}{2} - p, \mathbf{e}_1) \right\} \end{aligned}$$

Therefore, if we assume T_0 is even without loss of generality and define function $G_t^{T_0}(L)$ recursively as:

$$\begin{aligned} G_{T_0}^{T_0}(L) &\triangleq \tilde{V}_{T_0}^{T_0}(L, \mathbf{e}_1) = \frac{L + V(\mathbf{e}_1, 1)}{S(T_0)} = \frac{L}{S(T_0)} \\ G_t^{T_0}(L) &\triangleq \min_{p \in \mathbb{R}} \max \left\{ \frac{L + p}{S(t)}, G_{t+2}^{T_0}(L + \frac{1}{2} - p) \right\}, \end{aligned}$$

then it is clear that $\tilde{V}_t^{T_0}(L, \mathbf{e}_1) \geq G_t^{T_0}(L)$, and thus by (19),

$$\tilde{V}_1^{T_0}(0, \mathbf{0}) \geq \max\{1, \tilde{V}_2^{T_0}(\frac{1}{2}, \mathbf{e}_1)\} \geq \max\{1, G_2^{T_0}(\frac{1}{2})\}.$$

It remains to compute $G_2^{T_0}(\frac{1}{2})$. By some elementary computations and the fact that for two linear functions $h_1(p)$ and $h_2(p)$ of different signs of slopes,

$\min_p \max\{h_1(p), h_2(p)\} = h_1(p^*)$ where p^* is such that $h_1(p^*) = h_2(p^*)$, one can inductively prove that for $t = 2, 4, \dots, T_0$,

$$G_t^{T_0}(L) = \frac{2^{\frac{T_0-t}{2}}(L + \frac{1}{2}) - \frac{1}{2}}{S(T_0) + \sum_{k=1}^{(T_0-t)/2} (2^{k-1}S(T_0 - 2k))}.$$

Plugging $S(t) = \frac{t}{2^t} \binom{t-1}{\lfloor t/2 \rfloor}$ and letting $T_0 \rightarrow \infty$, we arrive at

$$\begin{aligned} & \lim_{T_0 \rightarrow \infty} G_2^{T_0}(1/2) \\ &= \lim_{T_0 \rightarrow \infty} \left(\sum_{k=1}^{T_0/2-1} \left(2^{k-T_0/2} S(T_0 - 2k) \right) \right)^{-1} \\ &= \lim_{T_0 \rightarrow \infty} \left(\sum_{k=1}^{T_0/2-1} \left(\frac{S(2k)}{2^k} \right) \right)^{-1} \\ &= \left(\sum_{j=0}^{\infty} \frac{j}{8^j} \binom{2j}{j} \right)^{-1}. \end{aligned}$$

Define $G(x) = \sum_{j=0}^{\infty} \binom{2j}{j} x^j$ and $F(x) = x \cdot G'(x)$. Note that what we want to compute above is exactly $1/F(\frac{1}{8})$. Lehmer (1985) showed that $G(x) = (1 - 4x)^{-1/2}$. Therefore, $F(x) = 2x \cdot (1 - 4x)^{-3/2}$ and

$$\lim_{T_0 \rightarrow \infty} G_2^{T_0}(1/2) = 1/F(1/8) = \sqrt{2}.$$

We conclude the proof by pointing out

$$\begin{aligned} \tilde{V} &= \lim_{T_0 \rightarrow \infty} \tilde{V}_1^{T_0}(0, \mathbf{0}) \\ &\geq \max\{1, \lim_{T_0 \rightarrow \infty} G_2^{T_0}(1/2)\} = \sqrt{2}. \end{aligned}$$

□

As we mentioned at the beginning of this section, the last thing we need to show is that the value $V(\mathbf{0}, T)$ is the same under the two loss spaces. In fact, we will prove stronger results in the following theorem claiming that this is true only if $N = 2$.

Theorem 10. *Let $\mathbf{LS}_1, \mathbf{LS}_2, \mathbf{LS}_3$ be the three loss spaces $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}, \{0, 1\}^N$ and $[0, 1]^N$ respectively, and $V_{\mathbf{LS}}(\mathbf{0}, T)$ be the value of the game $V(\mathbf{0}, T)$ under the loss space \mathbf{LS} . If $N > 2$, we have for any T ,*

$$V_{\mathbf{LS}_1}(\mathbf{0}, T) < V_{\mathbf{LS}_2}(\mathbf{0}, T) = V_{\mathbf{LS}_3}(\mathbf{0}, T).$$

However, the three values above are the same if $N = 2$.

Proof. We first inductively show that for any \mathbf{M} and r , $V_{\mathbf{LS}_2}(\mathbf{M}, r) = V_{\mathbf{LS}_3}(\mathbf{M}, r)$ and $V_{\mathbf{LS}_3}(\mathbf{M}, r)$ is convex in \mathbf{M} . For the base case $r = 0$, by definition, $V_{\mathbf{LS}_2}(\mathbf{M}, 0) = V_{\mathbf{LS}_3}(\mathbf{M}, 0) = -\min_i M_i$. Also, for any two loss vectors \mathbf{M} and \mathbf{M}' , and $\lambda \in [0, 1]$,

$$\begin{aligned} & V_{\mathbf{LS}_3}(\lambda\mathbf{M} + (1-\lambda)\mathbf{M}', 0) \\ &= -\min_i (\lambda M_i + (1-\lambda)M'_i) \\ &\leq -\min_i (\lambda M_i) - \min_i ((1-\lambda)M'_i) \\ &= \lambda V_{\mathbf{LS}_3}(\mathbf{M}, 0) + (1-\lambda)V_{\mathbf{LS}_3}(\mathbf{M}', 0), \end{aligned}$$

showing $V_{\mathbf{LS}_3}(\mathbf{M}, 0)$ is convex in \mathbf{M} . For $r > 0$,

$$V_{\mathbf{LS}_3}(\mathbf{M}, r) = \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}_3} (\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_3}(\mathbf{M} + \mathbf{Z}, r - 1)).$$

Notice that $\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_3}(\mathbf{M} + \mathbf{Z}, r - 1)$ is equal to $\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_2}(\mathbf{M} + \mathbf{Z}, r - 1)$ and is convex in \mathbf{Z} by induction. Therefore the maximum is always achieved at one of the corner points of \mathbf{LS}_3 , which is in \mathbf{LS}_2 . In other words,

$$\begin{aligned} V_{\mathbf{LS}_3}(\mathbf{M}, r) &= \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}_2} (\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_2}(\mathbf{M} + \mathbf{Z}, r - 1)) \\ &= V_{\mathbf{LS}_2}(\mathbf{M}, r). \end{aligned}$$

On the other hand, by introducing a distribution Q over all the elements in \mathbf{LS}_2 , we have

$$\begin{aligned} & V_{\mathbf{LS}_3}(\mathbf{M}, r) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}_2} (\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_3}(\mathbf{M} + \mathbf{Z}, r - 1)) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max_Q \mathbb{E}_{\mathbf{Z} \sim Q} [\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_3}(\mathbf{M} + \mathbf{Z}, r - 1)] \\ &= \max_Q \min_{\mathbf{P} \in \Delta(N)} \mathbb{E}_{\mathbf{Z} \sim Q} [\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_3}(\mathbf{M} + \mathbf{Z}, r - 1)] \\ &= \max_Q \left(\mathbb{E}_{\mathbf{Z} \sim Q} V_{\mathbf{LS}_3}(\mathbf{M} + \mathbf{Z}, r - 1) + \min_{\mathbf{P} \in \Delta(N)} \mathbf{P} \cdot \mathbb{E}_{\mathbf{Z} \sim Q} [\mathbf{Z}] \right) \end{aligned}$$

where we switch the min and max by Corollary 37.3.2 of [Rockafellar \(1970\)](#). Note that the last expression is the maximum over a family of linear combinations of convex functions in \mathbf{M} , which is still a convex function in \mathbf{M} , completing the induction step. To conclude, $V_{\mathbf{LS}_2}(\mathbf{0}, T) = V_{\mathbf{LS}_3}(\mathbf{0}, T)$ for any N and T .

We next prove if $N = 2$, $V_{\mathbf{LS}_1}(\mathbf{0}, T) = V_{\mathbf{LS}_2}(\mathbf{0}, T)$. Again, we inductively prove $V_{\mathbf{LS}_1}(\mathbf{M}, r) = V_{\mathbf{LS}_2}(\mathbf{M}, r)$ for any \mathbf{M} and r . The base case is clear. For $r > 0$, let $P_i^* = V_{\mathbf{LS}_1}(\mathbf{M}, r) - V_{\mathbf{LS}_1}(\mathbf{M} + \mathbf{e}_i, r - 1)$ ($i = 1, 2$). By induction,

$$\begin{aligned} & V_{\mathbf{LS}_2}(\mathbf{M}, r) \\ &= \min_{\mathbf{P} \in \Delta(2)} \max_{\mathbf{Z} \in \mathbf{LS}_2} (\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_1}(\mathbf{M} + \mathbf{Z}, r - 1)) \\ &\leq \max_{Z_1, Z_2 \in \{0, 1\}} (P_1^* Z_1 + P_2^* Z_2 + V_{\mathbf{LS}_1}(\mathbf{M} + (Z_1, Z_2), r - 1)) \end{aligned}$$

$$\begin{aligned} &= \max\{V_{\mathbf{LS}_1}(\mathbf{M}, r - 1), 1 + V_{\mathbf{LS}_1}(\mathbf{M} + (1, 1), r - 1), \\ &\quad V_{\mathbf{LS}_1}(\mathbf{M}, r)\} \\ &= \max\{V_{\mathbf{LS}_1}(\mathbf{M}, r - 1), V_{\mathbf{LS}_1}(\mathbf{M}, r)\} \\ &\quad \text{(by Property 5 in Proposition 2)} \\ &= V_{\mathbf{LS}_1}(\mathbf{M}, r). \quad \text{(by Property 7 in Proposition 2)} \end{aligned}$$

However, it is clear that $V_{\mathbf{LS}_2}(\mathbf{M}, r) \geq V_{\mathbf{LS}_1}(\mathbf{M}, r)$. Therefore, $V_{\mathbf{LS}_1}(\mathbf{M}, r) = V_{\mathbf{LS}_2}(\mathbf{M}, r)$.

Finally, to prove $V_{\mathbf{LS}_1}(\mathbf{0}, T) < V_{\mathbf{LS}_2}(\mathbf{0}, T)$ for $N > 2$, we inductively prove $V_{\mathbf{LS}_1}((T-r)\mathbf{e}_1, r) < V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1, r)$ for $r = 1, \dots, T$. For the base case $r = 1$, $V_{\mathbf{LS}_1}((T-1)\mathbf{e}_1, 1) = 1/N - R((T-1)\mathbf{e}_1, 1) = 1/N$, while

$$\begin{aligned} & V_{\mathbf{LS}_2}((T-1)\mathbf{e}_1, 1) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}_2} (\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_2}((T-1)\mathbf{e}_1 + \mathbf{Z}, 0)) \\ &\geq \min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} (1 - P_i + V_{\mathbf{LS}_2}((T-1)\mathbf{e}_1 + \mathbf{1} - \mathbf{e}_i, 0)) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max\{-P_1, 1 - P_2, \dots, 1 - P_N\}. \end{aligned}$$

We claim that the value of the last minimax expression above, denoted by v , is $(N-2)/(N-1)$, which is strictly greater than $1/N$ if $N > 2$ and thus proves the base case. To show that, notice that for any $\mathbf{P} \in \Delta(N)$, there must exist $i \in \{2, \dots, N\}$ such that $P_i \leq 1/(N-1)$ and

$$\max\{-P_1, 1 - P_2, \dots, 1 - P_N\} \geq 1 - P_i \geq \frac{N-2}{N-1},$$

showing $v \geq (N-2)/(N-1)$. On the other hand, the equality is realized by the distribution $\mathbf{P}^* = (0, \frac{1}{N-1}, \dots, \frac{1}{N-1})$.

For $r > 1$, we have

$$\begin{aligned} & V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1, r) \\ &= \min_{\mathbf{P} \in \Delta(N)} \max_{\mathbf{Z} \in \mathbf{LS}_2} (\mathbf{P} \cdot \mathbf{Z} + V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1 + \mathbf{Z}, r - 1)) \\ &\geq \min_{\mathbf{P} \in \Delta(N)} \max_{i \in [N]} (P_i + V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1 + \mathbf{e}_i, r - 1)) \\ &\geq \min_{\mathbf{P} \in \Delta(N)} \frac{1}{N} \sum_{i=1}^N (P_i + V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1 + \mathbf{e}_i, r - 1)) \\ &= \frac{1}{N} + \frac{1}{N} \sum_{i=1}^N V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1 + \mathbf{e}_i, r - 1) \\ &> \frac{1}{N} + \frac{1}{N} \sum_{i=1}^N V_{\mathbf{LS}_1}((T-r)\mathbf{e}_1 + \mathbf{e}_i, r - 1) \\ &= V_{\mathbf{LS}_1}((T-r)\mathbf{e}_1, r). \end{aligned}$$

Here, the last strict inequality holds because for $i = 1$, $V_{\mathbf{LS}_2}((T-r+1)\mathbf{e}_1, r - 1) > V_{\mathbf{LS}_1}((T-r+1)\mathbf{e}_1, r - 1)$ by induction; for $i \neq 1$, it is trivial that $V_{\mathbf{LS}_2}((T-r)\mathbf{e}_1 +$

$\mathbf{e}_i, r-1) \geq V_{\text{LS}_1}((T-r)\mathbf{e}_1 + \mathbf{e}_i, r-1)$. Therefore, we complete the induction step and thus prove $V_{\text{LS}_1}(\mathbf{0}, T) < V_{\text{LS}_2}(\mathbf{0}, T)$. \square

D. Proof of Theorem 5

The proof (and the one of Theorem 8) relies heavily on a common technique to approximate a sum using an integral, which we state without proof as the following claim.

Claim 1. *Let $f(x)$ be a non-increasing nonnegative function defined on \mathbb{R}_+ . Then the following inequalities hold for any integer $0 < j \leq k$.*

$$\int_j^{k+1} f(x) dx \leq \sum_{i=j}^k f(i) \leq \int_{j-1}^k f(x) dx$$

Proof of Theorem 5. By Theorem 4, it suffices to upper bound $\bar{V}_1(\mathbf{0})$ and $\sum_{t=1}^{T_s} q_t \bar{V}_{t+1}(\mathbf{0})$. Let $S_t = \sum_{t'=t}^{\infty} 1/t'^d$. By applying Claim 1 multiple times, we have

$$\begin{aligned} \frac{1}{S_t} &\leq \left(\int_t^{\infty} \frac{dx}{x^d} \right)^{-1} = t^{d-1}(d-1); \quad (20) \\ q_t &= \frac{1}{S_t \cdot t^d} \leq \frac{d-1}{t}; \end{aligned}$$

$$\begin{aligned} \bar{V}_1(\mathbf{0}) &= \mathbb{E}[V(\mathbf{0}, T) | T \geq 1] \\ &\leq \frac{c_N}{S_1} \sum_{T=1}^{\infty} \frac{1}{T^{d-\frac{1}{2}}} \quad (\text{by Theorem 1}) \\ &\leq \frac{c_N}{S_1} \left(1 + \int_1^{\infty} \frac{dx}{x^{d-\frac{1}{2}}} \right) \quad (\text{by Claim 1}) \\ &= \frac{c_N(d-\frac{1}{2})}{S_1(d-\frac{3}{2})} \\ &\leq \frac{c_N(d-1)(d-\frac{1}{2})}{d-\frac{3}{2}} = O(1). \quad (\text{by Eq. (20)}) \end{aligned}$$

For

$$\begin{aligned} \bar{V}_{t+1}(\mathbf{0}) &= \mathbb{E}[V(\mathbf{0}, T-t) | T \geq t+1] \\ &= \frac{c_N}{S_{t+1}} \sum_{k=1}^{\infty} \frac{\sqrt{k}}{(t+k)^d}, \end{aligned}$$

Claim 1 does not readily apply since the function $g(k) = \sqrt{k}/(t+k)^d$ is increasing on $[0, t/(2d-1)]$ and then decreasing on $[t/(2d-1), \infty)$. However, we can still apply the claim to these two parts separately. Let $x_0 = \lceil t/(2d-1) \rceil$ and $x_1 = \lfloor t/(2d-1) \rfloor$. For simplicity, assume $1 \leq x_0 < x_1$ and $g(x_0) \leq g(x_1)$ (other cases hold similarly). Then we have

$$\bar{V}_{t+1}(\mathbf{0}) = \frac{c_N}{S_{t+1}} \left(g(x_1) + \sum_{k=1}^{x_0} g(k) + \sum_{k=x_1+1}^{\infty} g(k) \right)$$

$$\begin{aligned} &\leq \frac{c_N}{S_{t+1}} \left(g(x_1) + \int_0^{x_1} g(x) dx + \int_{x_1}^{\infty} g(x) dx \right) \\ &= \frac{c_N}{S_{t+1}} \left(g(x_1) + \frac{\Gamma(d-\frac{3}{2})}{2\Gamma(d)} \cdot \frac{\sqrt{\pi}}{t^{d-\frac{3}{2}}} \right) \\ &\leq (d-1)c_N \sqrt{\pi} \cdot \frac{\Gamma(d-\frac{3}{2})}{2\Gamma(d)} \cdot \sqrt{t} + o(\sqrt{t}). \end{aligned}$$

So finally we have

$$\begin{aligned} &\sum_{t=1}^{T_s} q_t \bar{V}_{t+1}(\mathbf{0}) \\ &\leq (d-1)^2 c_N \sqrt{\pi} \cdot \frac{\Gamma(d-\frac{3}{2})}{2\Gamma(d)} \sum_{t=1}^{T_s} \left(\frac{1}{\sqrt{t}} + o\left(\frac{1}{\sqrt{t}}\right) \right) \\ &\leq \frac{\Gamma(d-\frac{3}{2})}{\Gamma(d)} (d-1)^2 c_N \sqrt{\pi T_s} + o(\sqrt{T_s}), \end{aligned}$$

which proves the theorem. \square

E. Proof of Theorem 6

Proof. Let $\Phi_t^T = \sqrt{\|\mathbf{W}_{t-1}\|^2 + (T-t+1)}$ be the potential function for this setting. The key property of the minimax algorithm Eq. (6) shown by Abernethy et al. (2008a) is the following:

$$\mathbf{x}_t^T \cdot \mathbf{w}_t \leq \Phi_t^T - \Phi_{t+1}^T.$$

Based on this property, the loss of our algorithm after T_s rounds is

$$\begin{aligned} \sum_{t=1}^{T_s} \mathbb{E}[\mathbf{x}_t^T | T \geq t] \cdot \mathbf{w}_t &= \sum_{t=1}^{T_s} \mathbb{E}[\mathbf{x}_t^T \cdot \mathbf{w}_t | T \geq t] \\ &\leq \sum_{t=1}^{T_s} \mathbb{E}[\Phi_t^T - \Phi_{t+1}^T | T \geq t]. \end{aligned}$$

Now define $U_t = \mathbb{E}[\Phi_t^T | T \geq t]$ and $q_t = \Pr[T < t+1 | T \geq t]$. By the fact that $f_{T \geq t}(t') = (1 - q_t) f_{T \geq t+1}(t')$ for any $t' \geq t+1$, where $f_{T \geq t}$ and $f_{T \geq t+1}$ are conditional density functions, we have

$$\begin{aligned} &\sum_{t=1}^{T_s} \mathbb{E}[\mathbf{x}_t^T | T \geq t] \cdot \mathbf{w}_t \\ &\leq \sum_{t=1}^{T_s} (U_t - \mathbb{E}[\Phi_{t+1}^T | T \geq t]) \\ &= \sum_{t=1}^{T_s} \left(U_t - \int_t^{t+1} \Phi_{t+1}^T f_{T \geq t}(T) dT - (1 - q_t) U_{t+1} \right) \\ &\leq \sum_{t=1}^{T_s} \left(U_t - \Phi_{t+1}^t \int_t^{t+1} f_{T \geq t}(T) dT - (1 - q_t) U_{t+1} \right) \\ &\quad (\because \Phi_{t+1}^T \text{ increases in } T) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{t=1}^{T_s} (U_t - U_{t+1} + q_t(U_{t+1} - \|\mathbf{W}_{t-1}\|)) \\
 &\quad (\because \Phi_{t+1}^t = \|\mathbf{W}_{t-1}\|) \\
 &= U_1 - U_{T_s+1} + \\
 &\quad \sum_{t=1}^{T_s} q_t \mathbb{E} \left[\sqrt{\|\mathbf{W}_{t-1}\|^2 + (T-t)} - \|\mathbf{W}_{t-1}\| \mid T \geq t+1 \right] \\
 &\leq U_1 - U_{T_s+1} + \sum_{t=1}^{T_s} q_t \mathbb{E} \left[\sqrt{T-t} \mid T \geq t+1 \right]. \\
 &\quad (\because \sqrt{a+b} - \sqrt{a} \leq \sqrt{b})
 \end{aligned}$$

Note that $U_{T_s+1} \geq \|\mathbf{W}_T\|$, and thus it remains to plug in the distribution and compute U_1 and $\sum_{t=1}^{T_s} q_t \mathbb{E}[\sqrt{T-t} \mid T \geq t+1]$, which is almost the same process as what we did in the proof of Theorem 5 if one realizes $q_t \leq (d-1)/t$ also holds here. In a word, the regret can be bounded by

$$\frac{\Gamma(d - \frac{3}{2})}{\Gamma(d)} (d-1)^2 \sqrt{\pi T_s} + o(\sqrt{T_s}),$$

which is $\pi\sqrt{T_s} + o(\sqrt{T_s})$ if $d = 2$. The explicit form in Eq. (7) comes from a direct calculation. \square

F. Proof of Lemma 2 and Theorem 7

Proof of Lemma 2. The results follow by a direct calculation. The conditional distribution of ξ_t given T is $\frac{1}{\Delta_T^N} \mathbf{1}\{\xi \in [0, \Delta_T]^N\}$. Let $S_t = \int_t^\infty 1/T^d dT = ((d-1)t^{d-1})^{-1}$. The marginal distribution for ξ that has negative coordinates is clearly 0. Otherwise, with $\bar{t} = \max\{t, \frac{\|\xi\|_\infty^2}{bN}\}$ one has

$$\begin{aligned}
 f_t(\xi) &= \frac{1}{S_t} \int_t^\infty \frac{1}{T^d \Delta_T^N} \mathbf{1}\{\xi \in [0, \Delta_T]^N\} dT \\
 &= \frac{1}{S_t} \int_{\bar{t}}^\infty \frac{1}{T^d \Delta_T^N} dT \\
 &= \frac{(d-1)t^{d-1}}{(\sqrt{bN})^N} \int_{\bar{t}}^\infty \frac{1}{T^{d+N/2}} dT \\
 &= \frac{d-1}{d-1+N/2} \Delta_t^{-N} \min \left\{ 1, \left(\frac{\Delta_t}{\|\xi\|_\infty} \right)^{2d-2+N} \right\}.
 \end{aligned}$$

\square

Proof of Theorem 7. Applying Theorem 4.2 of Cesa-Bianchi & Lugosi (2006), the pseudo-regret of the FPL algorithm is bounded by

$$\mathbb{E}[\max_i \xi_{T_s, i}] + \sum_{t=1}^{T_s} \mathbb{E}[\max_i (\xi_{t-1, i} - \xi_{t, i})]$$

$$+ \sum_{t=1}^{T_s} \int_{\mathbb{R}^N} F_t(\xi) (f_t(\xi) - f_t(\xi - \mathbf{Z}_t)) d\xi,$$

where we define $\xi_0 = \mathbf{0}$ and $F_t(\xi) = Z_{t, I_\xi}$ with $I_\xi \in \arg \min_i (M_{t-1, i} + \xi_i)$. Now the key observation is that the pseudo-regret remains the same if we replace random variables ξ_1, \dots, ξ_{T_s} with $\xi'_1, \dots, \xi'_{T_s}$ as long as ξ_t and ξ'_t have the same marginal distribution for any t . Specifically, we can let $\xi'_{T_s} = \xi_{T_s}$, and for $1 < t \leq T_s$, let $\xi'_{t-1} = \xi'_t$ with probability $S_t/S_{t-1} = (1-1/t)^{d-1}$ (recall $S_t = \int_t^\infty 1/T^d dT$), or with $1 - S_t/S_{t-1}$ probability be obtained by first drawing $T \in [t-1, t]$ according to density $f(T) \propto 1/T^d$, and then drawing a point uniformly in $[0, \Delta_T]^N$. It is clear that ξ_t and ξ'_t have the same marginal distribution. So the pseudo-regret can be in fact bounded by three terms:

$$A = \mathbb{E}[\max_i \xi_{T_s, i}],$$

$$B = \sum_{t=1}^{T_s} \mathbb{E}[\max_i (\xi'_{t-1, i} - \xi'_{t, i})],$$

$$C = \sum_{t=1}^{T_s} \int_{\mathbb{R}^N} F_t(\xi) (f_t(\xi) - f_t(\xi - \mathbf{Z}_t)) d\xi.$$

A can be further bounded by

$$\frac{1}{S_{T_s}} \int_{T_s}^\infty \frac{\Delta_T}{T^d} dT = \frac{d-1}{d-3/2} \sqrt{bT_s N}.$$

For B , by construction of ξ'_t , we have

$$\begin{aligned}
 B &\leq \sum_{t=2}^{T_s} \left(\frac{\Delta_t}{S_{t-1}} \int_{t-1}^t \frac{dT}{T^d} + \frac{S_t}{S_{t-1}} \cdot 0 \right) \\
 &= \sum_{t=2}^{T_s} \frac{\Delta_t}{t^{d-1}} (t^{d-1} - (t-1)^{d-1}) \\
 &\leq \sum_{t=2}^{T_s} \frac{\Delta_t}{t^{d-1}} \cdot (d-1)t^{d-2} \quad (\text{by convexity}) \\
 &\leq 2(d-1) \sqrt{bT_s N}.
 \end{aligned}$$

For C , let $H = \{\xi : f_t(\xi) > f_t(\xi - \mathbf{Z}_t)\}$. Since $0 \leq F_t(\xi) \leq 1$, we have $C \leq \sum_{t=1}^{T_s} \int_H f_t(\xi) d\xi$. Now observe that when $\min_i \xi_i \geq 0$, $f_t(\xi)$ is non-increasing in each ξ_i . So the only possibility that $f_t(\xi) > f_t(\xi - \mathbf{Z}_t)$ holds is when there exists an i such that ξ_i is strictly smaller than $Z_{t, i}$. That is

$$H = \{\xi : \min_i \xi_i \geq 0 \text{ and } \exists i, \text{ s.t. } \xi_i < Z_{t, i}\}$$

So we have

$$C \leq \sum_{t=1}^{T_s} \frac{1}{S_t} \int_t^\infty \frac{dT}{T^d} \int_H \frac{\mathbf{1}\{\xi \in [0, \Delta_T]^N\}}{\Delta_T^N} d\xi$$

$$\begin{aligned}
 &\leq \sum_{t=1}^{T_s} \frac{1}{S_t} \int_t^\infty \frac{N}{T^d} \frac{Z_{t,i} \Delta_T^{N-1}}{\Delta_T^N} dT \\
 &\leq \frac{d-1}{d-1/2} \sqrt{\frac{N}{b}} \sum_{t=1}^{T_s} \frac{1}{\sqrt{t}} \\
 &\leq \frac{2(d-1)}{\sqrt{b}(d-1/2)} \sqrt{T_s N}.
 \end{aligned}$$

Combining A , B and C proves the theorem. \square

G. Proof of Theorem 8

Proof. We will first show that

$$\text{Reg}(L_{T_s}, \mathbf{M}_{T_s}) \leq \underbrace{(\ln N) \cdot \mathbb{E} \left[\frac{1}{\eta_T} | T \geq T_s + 1 \right]}_A + \underbrace{\frac{1}{8} \sum_{t=1}^{T_s} \mathbb{E}[\eta_T | T \geq t]}_B. \quad (21)$$

Let $\Phi_t^T = \frac{1}{\eta_T} \ln \left(\sum_{i=1}^N \exp(-\eta_T M_{t-1,i}) \right)$. The key point of the proof for the non-adaptive version of the exponential weights algorithm is to use Φ_t^T as a ‘‘potential’’ function, and bound the change in potential before and after a single round (Cesa-Bianchi & Lugosi, 2006). Specifically, they showed that

$$\mathbf{P}_t^T \cdot \mathbf{Z}_t \leq \frac{\eta_T}{8} + \Phi_t^T - \Phi_{t+1}^T.$$

We also base our proof on this inequality. The total loss of the learner after T_s rounds is

$$\begin{aligned}
 L_{T_s} &= \sum_{t=1}^{T_s} \mathbb{E}[\mathbf{P}_t^T | T \geq t] \cdot \mathbf{Z}_t = \sum_{t=1}^{T_s} \mathbb{E}[\mathbf{P}_t^T \cdot \mathbf{Z}_t | T \geq t] \\
 &\leq B + \sum_{t=1}^{T_s} \mathbb{E}[\Phi_t^T - \Phi_{t+1}^T | T \geq t].
 \end{aligned}$$

Define $U_t = \mathbb{E}[\Phi_t^T | T \geq t]$. We do the following transformation:

$$\begin{aligned}
 &\mathbb{E}[\Phi_t^T - \Phi_{t+1}^T | T \geq t] \\
 &= U_t - E_T[\Phi_{t+1}^T | T \geq t] \\
 &= U_t - q_t \Phi_{t+1}^t - (1 - q_t) U_{t+1} \\
 &= U_t - U_{t+1} + q_t (U_{t+1} - \Phi_{t+1}^t) \\
 &= U_t - U_{t+1} + q_t \cdot \mathbb{E}[\Phi_{t+1}^T - \Phi_{t+1}^t | T \geq t + 1] \\
 &= U_t - U_{t+1} + q_t \cdot \mathbb{E}[F_{T,t}(\mathbf{M}_t) | T \geq t + 1],
 \end{aligned}$$

where we define

$$F_{T,t}(\mathbf{M}) = \frac{\ln \left(\sum_i \exp(-\eta_T M_i) \right)}{\eta_T} - \frac{\ln \left(\sum_i \exp(-\eta_t M_i) \right)}{\eta_t}.$$

A key observation is

$$\max_{\substack{\mathbf{M} \in \mathbb{R}_+^N \\ \eta_T < \eta_t}} F_{T,t}(\mathbf{M}) = \frac{\ln N}{\eta_T} - \frac{\ln N}{\eta_t}, \quad (22)$$

which can be verified by a standard derivative analysis that we omit. (An alternative approach using KL-divergence can be found in Chapter 2.5 of Bubeck, 2011.)

We further define another potential function $\bar{\Phi}_t^T = (\ln N)/\eta_T$ and also $\bar{U}_t = \mathbb{E}[\bar{\Phi}_t^T | T \geq t]$. Note that the new potential $\bar{\Phi}_t^T$ has no dependence on t and thus $\bar{\Phi}_t^T = \bar{\Phi}_{t'}^T$ for any t, t' . We now have

$$\begin{aligned}
 &\sum_{t=1}^{T_s} \mathbb{E}[\Phi_t^T - \Phi_{t+1}^T | T \geq t] \\
 &= \sum_{t=1}^{T_s} (U_t - U_{t+1} + q_t \cdot \mathbb{E}[\Phi_{t+1}^T - \Phi_{t+1}^t | T \geq t + 1]) \\
 &= U_1 - U_{T_s+1} + \underbrace{\sum_{t=1}^{T_s} (q_t \cdot \mathbb{E}[\Phi_{t+1}^T - \Phi_{t+1}^t | T \geq t + 1])}_C \\
 &\leq U_1 - U_{T_s+1} + \sum_{t=1}^{T_s} \left(q_t \cdot \mathbb{E} \left[\frac{\ln N}{\eta_T} - \frac{\ln N}{\eta_t} | T \geq t + 1 \right] \right) \\
 &\quad \text{(by Eq. (22))} \\
 &= \bar{U}_1 - \bar{U}_{T_s+1} + \underbrace{\sum_{t=1}^{T_s} (q_t \cdot \mathbb{E}[\bar{\Phi}_{t+1}^T - \bar{\Phi}_{t+1}^t | T \geq t + 1])}_D \\
 &\quad + \bar{U}_{T_s+1} - U_{T_s+1}. \quad (\because U_1 = \bar{U}_1)
 \end{aligned} \quad (23)$$

Notice that D has the exact same form as C except for a different definition of the potential, and also Eq. (23) is an equality. Therefore, by a reverse transformation, we have

$$\begin{aligned}
 &\sum_{t=1}^{T_s} \mathbb{E}[\Phi_t^T - \Phi_{t+1}^T | T \geq t] \\
 &= \sum_{t=1}^{T_s} \mathbb{E}[\bar{\Phi}_t^T - \bar{\Phi}_{t+1}^T | T \geq t] + \bar{U}_{T_s+1} - U_{T_s+1} \\
 &= \bar{U}_{T_s+1} - U_{T_s+1} \quad (\because \bar{\Phi}_t^T = \bar{\Phi}_{t+1}^T)
 \end{aligned}$$

\bar{U}_{T_s+1} is exactly A in Eq. (21), and U_{T_s+1} can be related to the loss of the best action:

$$\begin{aligned}
 U_{T_s+1} &= \mathbb{E} \left[\frac{1}{\eta_T} \ln \sum_{i=1}^N \exp(-\eta_T M_{T_s,i}) \mid T \geq T_s + 1 \right] \\
 &\geq \mathbb{E} \left[\frac{1}{\eta_T} \ln \exp(-\eta_T R(M_{T_s}, 0)) \mid T \geq T_s + 1 \right] \\
 &= -R(M_{T_s}, 0).
 \end{aligned}$$

The regret is therefore

$$\begin{aligned} \text{Reg}(L_{T_s}, \mathbf{M}_{T_s}) &= L_{T_s} - R(M_{T_s}, 0) \\ &\leq A + B - U_{T_s+1} - R(M_{T_s}, 0) \\ &\leq A + B, \end{aligned}$$

proving Eq. (21).

The rest of the proof is merely to plug in the distribution and $\eta_T = \sqrt{(b \ln N)/T}$, and upper bound Eq. (21) using Claim 1. Adopting the notation $S_t = \sum_{t'=t}^{\infty} 1/t'^d$ and the result of Eq. (20) in the proof of Theorem 5, we have

$$\begin{aligned} A &= \frac{\sqrt{\ln N}}{S_{T_s+1}\sqrt{b}} \sum_{T=T_s+1}^{\infty} \frac{1}{T^{d-1/2}} \\ &\leq \frac{(d-1)\sqrt{\ln N}}{\sqrt{b}} (T_s+1)^{d-1} \\ &\quad \left(\int_{T_s+1}^{\infty} \frac{dx}{x^{d-1/2}} + \frac{1}{(T_s+1)^{d-1/2}} \right) \\ &= \frac{d-1}{(d-3/2)\sqrt{b}} \sqrt{T_s \ln N} + o(\sqrt{T_s \ln N}); \end{aligned}$$

$$\begin{aligned} B &= \frac{\sqrt{b \ln N}}{8} \sum_{t=1}^{T_s} \frac{1}{S_t} \sum_{T=t}^{\infty} \frac{1}{T^{d+1/2}} \\ &\leq \frac{(d-1)\sqrt{b \ln N}}{8} \sum_{t=1}^{T_s} t^{d-1} \left(\int_t^{\infty} \frac{dx}{x^{d+1/2}} + \frac{1}{t^{d+1/2}} \right) \\ &\leq \frac{(d-1)\sqrt{b \ln N}}{8} \sum_{t=1}^{T_s} \left(\frac{1}{(d-1/2)\sqrt{t}} + \frac{1}{t^{d+3/2}} \right) \\ &\leq \frac{\sqrt{b}(d-1)}{4(d-1/2)} \sqrt{T_s \ln N} + o(\sqrt{T_s \ln N}). \end{aligned}$$

Combining the bounds above for A and B proves the theorem. \square

H. Proof of Theorem 9

Proof. The main idea resembles the one of Theorem 8, but the details are much more technical. Let us first define several notations:

$$S_t \triangleq \int_{m_t}^{\infty} \frac{dm}{m^d} = \frac{1}{(d-1)m_t^{d-1}},$$

$$\begin{aligned} q_t &\triangleq \Pr[m < m_t | m \geq m_{t-1}] = \frac{1}{S_{t-1}} \int_{m_{t-1}}^{m_t} \frac{dm}{m^d} \\ &= 1 - \left(\frac{m_{t-1}}{m_t} \right)^{d-1}, \end{aligned}$$

$$Y_t^m \triangleq \sum_{i=1}^N \exp(-\eta_m M_{t-1,i}),$$

$$\Phi_t^m \triangleq \left(1 + \frac{1}{\eta_m} \right) \ln Y_t^m, \quad U_t \triangleq \mathbb{E}[\Phi_t^m | m \geq m_{t-1}].$$

The proof starts from the following property of the exponential weights algorithm (Cesa-Bianchi & Lugosi, 2006):

$$\begin{aligned} \mathbf{P}_t^m \cdot \mathbf{Z}_t &\leq \frac{1}{1 - e^{-\eta_m}} (\ln Y_t^m - \ln Y_{t+1}^m) \\ &\leq \Phi_t^m - \Phi_{t+1}^m. \quad (\because \eta_m \geq \ln(1 + \eta_m)) \end{aligned}$$

By the fact that $f_{m \geq m_{t-1}}(m') = (1 - q_t)f_{m \geq m_t}(m')$ for any $m' \geq m_t$, where $f_{m \geq m_{t-1}}$ and $f_{m \geq m_t}$ are conditional density functions, the loss of the learner after T_s rounds L_{T_s} is

$$\begin{aligned} &\sum_{t=1}^{T_s} \mathbb{E}[\mathbf{P}_t^m \cdot \mathbf{Z}_t | m \geq m_{t-1}] \\ &\leq \sum_{t=1}^{T_s} \mathbb{E}[\Phi_t^m - \Phi_{t+1}^m | m \geq m_{t-1}] \\ &= \sum_{t=1}^{T_s} \left(U_t - \int_{m_{t-1}}^{m_t} \Phi_{t+1}^m f_{m \geq m_{t-1}}(m) dm + (1 - q_t)U_{t+1} \right) \\ &\leq \sum_{t=1}^{T_s} \left(U_t - \Phi_{t+1}^{m_{t-1}} \int_{m_{t-1}}^{m_t} f_{m \geq m_{t-1}}(m) dm + (1 - q_t)U_{t+1} \right) \\ &= U_1 - U_{T_s+1} + \sum_{t=1}^{T_s} q_t (U_{t+1} - \Phi_{t+1}^{m_{t-1}}), \end{aligned}$$

Here the last inequality holds because Φ_t^m is increasing in m . To show this, we consider the following

$$\begin{aligned} &\left(1 + \frac{1}{\eta} \right) \ln \sum_{i=1}^N \exp(-\eta a_i) \\ &= \left(1 + \frac{1}{\eta} \right) \left(-\eta a_1 + \ln \sum_{i=1}^N \exp(-\eta(a_i - a_1)) \right) \\ &= -(\eta + 1)a_1 + \left(1 + \frac{1}{\eta} \right) \ln \sum_{i=1}^N \exp(-\eta(a_i - a_1)), \end{aligned}$$

where η, a_1, \dots, a_N are positive numbers. Since $\ln \sum_i \exp(-\eta(a_i - a_1)) \geq 0$, the expression above is decreasing in η , which along with the fact that η_m decreases in m shows that Φ_t^m increases in m .

We now compute U_1 and U_{T_s+1} :

$$\begin{aligned} U_1 &= \mathbb{E}[(1 + \sqrt{m/\ln N}) \ln N | m \geq 1] \\ &= \ln N + \frac{d-1}{d-3/2} \sqrt{\ln N} \end{aligned}$$

$$U_{T_s+1} = \mathbb{E} \left[\left(1 + \frac{1}{\eta_m} \right) \ln \sum_i \exp(-\eta_m M_{T_s,i}) \mid m \geq m_{T_s} \right]$$

$$\begin{aligned}
 &\geq \mathbb{E}[(1 + 1/\eta_m)(-\eta_m m^*) \mid m \geq m_{T_s}] \\
 &= -m^* (1 + \mathbb{E}[\eta_m \mid m \geq m_{T_s}]) \\
 &= -m^* \left(1 + \frac{d-1}{d-1/2} \sqrt{\frac{\ln N}{m_{T_s}}}\right) \\
 &\geq -m^* - \frac{d-1}{d-1/2} \sqrt{m^* \ln N} \\
 &\quad (\because m_{T_s} = m^* + 1)
 \end{aligned}$$

For $U_{t+1} - \Phi_{t+1}^{m_{t-1}} = \mathbb{E}[\Phi_{t+1}^m - \Phi_{t+1}^{m_{t-1}} \mid m \geq m_t]$, we first upper bound the part inside the expectation:

$$\begin{aligned}
 &\Phi_{t+1}^m - \Phi_{t+1}^{m_{t-1}} \\
 &= \left(\frac{\ln Y_{t+1}^m}{\eta_m} - \frac{\ln Y_{t+1}^{m_{t-1}}}{\eta_{m_{t-1}}} \right) + (\eta_{m_{t-1}} - \eta_m) \min_i M_{t,i} \\
 &\quad + \ln \frac{\sum e^{-\eta_m (M_{t,i} - \min_i M_{t,i})}}{\sum e^{-\eta_{m_{t-1}} (M_{t,i} - \min_i M_{t,i})}}.
 \end{aligned}$$

The first term above is at most $\left(\frac{1}{\eta_m} - \frac{1}{\eta_{m_{t-1}}}\right) \ln N = \sqrt{\ln N}(\sqrt{m} - \sqrt{m_{t-1}})$ by Eq. (22). The second term is at most $\sqrt{\ln N} \left(\frac{1}{\sqrt{m_{t-1}}} - \frac{1}{\sqrt{m}}\right) m_{t-1}$ since $\min_i M_{t,i} = m_t - 1 \leq m_{t-1}$, and the last term is at most $\ln N$ since the numerator is at most N while the denominator is at least 1. Therefore, we have

$$\begin{aligned}
 &U_{t+1} - \Phi_{t+1}^{m_{t-1}} \\
 &\leq \ln N + \sqrt{\ln N} \cdot \mathbb{E}[\sqrt{m} - \frac{m_{t-1}}{\sqrt{m}} \mid m \geq m_t] \\
 &= \ln N + \sqrt{\ln N} \left(\frac{d-1}{d-3/2} \sqrt{m_t} - \frac{d-1}{d-1/2} \frac{m_{t-1}}{\sqrt{m_t}} \right) \\
 &\leq \ln N + \sqrt{\ln N} \left(\frac{d-1}{d-3/2} \sqrt{m_t} - \frac{d-1}{d-1/2} \frac{m_t - 1}{\sqrt{m_t}} \right) \\
 &= \ln N + \frac{(d-1)\sqrt{m_t \ln N}}{(d-3/2)(d-1/2)} + \frac{d-1}{d-1/2} \sqrt{\frac{\ln N}{m_t}}.
 \end{aligned}$$

It remains to compute $\sum_{t=1}^{T_s} q_t (U_{t+1} - \Phi_{t+1}^{m_{t-1}})$, which, using the above, can be done by computing $A = \sum_{t=1}^{T_s} q_t$, $B = \sum_{t=1}^{T_s} q_t \sqrt{m_t}$ and $C = \sum_{t=1}^{T_s} q_t / \sqrt{m_t}$. By inequality $1 - x \leq -\ln x$ for any $x > 0$, we have

$$\begin{aligned}
 A &= \sum_{t=1}^{T_s} \left(1 - \left(\frac{m_{t-1}}{m_t}\right)^{d-1}\right) \\
 &\leq -(d-1) \sum_{t=1}^{T_s} (\ln m_{t-1} - \ln m_t) \\
 &= (d-1) \ln(m^* + 1).
 \end{aligned}$$

For B , we first show $q_t \sqrt{m_t} \leq 2(d-1)(\sqrt{m_t} - \sqrt{m_{t-1}})$,

which is equivalent to

$$\frac{q_t \sqrt{m_t}}{\sqrt{m_t} - \sqrt{m_{t-1}}} = \frac{\left(\frac{m_t}{m_{t-1}}\right)^{d-1} - 1}{\left(\frac{m_t}{m_{t-1}}\right)^{d-1} - \left(\frac{m_t}{m_{t-1}}\right)^{d-3/2}} \leq 2(d-1)$$

if $m_t \neq m_{t-1}$ (it is trivial otherwise). Define $h(x) = (x^{d-1} - 1)/(x^{d-1} - x^{d-3/2})$ for $x \in [1, 2]$ (note that m_t/m_{t-1} is within this interval). One can verify that $h'(x) < 0$ and thus $h(x) \leq \lim_{x \rightarrow 1} h(x) = 2(d-1)$. So we prove $q_t \sqrt{m_t} \leq 2(d-1)(\sqrt{m_t} - \sqrt{m_{t-1}})$ and

$$\begin{aligned}
 B &\leq 2(d-1) \sum_{t=1}^{T_s} (\sqrt{m_t} - \sqrt{m_{t-1}}) \\
 &= 2(d-1)(\sqrt{m_{T_s}} - 1) \leq 2(d-1)\sqrt{m^*}.
 \end{aligned}$$

A simple comparison of B and C shows $C = o(\sqrt{m^*})$. We finally conclude the proof by combining all we have

$$\begin{aligned}
 &\mathbf{Reg}(L_{T_s}, \mathbf{M}_{T_s}) \\
 &\leq U_1 - U_{T_s+1} + \sum_{t=1}^{T_s} q_t (U_{t+1} - \Phi_{t+1}^{m_{t-1}}) - m^* \\
 &= (1 + (d-1) \ln(m^* + 1)) \ln N \\
 &\quad + \left(\frac{d-1}{d-1/2} + \frac{2(d-1)^2}{(d-3/2)(d-1/2)} \right) \sqrt{m^* \ln N} \\
 &\quad + o(\sqrt{m^* \ln N}) \\
 &= \frac{3(d-7/6)(d-1)}{(d-3/2)(d-1/2)} \sqrt{m^* \ln N} \\
 &\quad + (1 + (d-1) \ln(m^* + 1)) \ln N + o(\sqrt{m^* \ln N}).
 \end{aligned}$$

□

I. Examples

The first example shows that the results stated in Theorem 2 can not generalize to other loss spaces.

Example 1. Consider the following Hedge setting: $N = 3$, $\mathbf{LS} = \{\mathbf{1} - \mathbf{e}_1, \mathbf{1} - \mathbf{e}_2, \mathbf{1} - \mathbf{e}_3\}$ where $\mathbf{1} = (1, 1, 1)$. Suppose the adversary picked $\mathbf{1} - \mathbf{e}_1$ and $\mathbf{1} - \mathbf{e}_2$ for the first two rounds and we are now on round $t = 3$ with $\mathbf{M}_2 = (1, 1, 2)$. Also the conditional distribution of the horizon given $T \geq 3$ is $\Pr[T = 3] = \Pr[T = 4] = 1/2$. Let \mathbf{P}^* be the minimax strategy for this round and \mathbf{P}^T be the minimax strategy assuming the horizon to be T . Then $\mathbf{P}^* \neq \mathbb{E}[\mathbf{P}^T \mid T \geq 3]$, and also

$$\begin{aligned}
 &\inf_{\mathbf{Alg}} \sup_{\mathbf{Z}_{3:\infty}} \mathbb{E}[\mathbf{Reg}(L_T, \mathbf{M}_T) \mid T \geq 3] \\
 &\neq \mathbb{E}[\inf_{\mathbf{Alg}} \sup_{\mathbf{Z}_{3:T}} \mathbf{Reg}(L_T, \mathbf{M}_T) \mid T \geq 3].
 \end{aligned} \tag{24}$$

Proof. Recall the V function we had in Section 3. Ignoring the loss for the learner for the first two rounds (which is the same for both sides of Eq. (24)), we point out that the right hand side of Eq. (24) is essentially

$$\frac{1}{2}V(\mathbf{M}_2, 1) + \frac{1}{2}V(\mathbf{M}_2, 2),$$

and the left hand side, denoted by V' , is

$$\min_{\mathbf{P}} \max_{\mathbf{Z}} (\mathbf{P} \cdot \mathbf{Z} + \frac{1}{2}V(\mathbf{M}_2 + \mathbf{Z}, 0) + \frac{1}{2}V(\mathbf{M}_2 + \mathbf{Z}, 1)).$$

Also \mathbf{P}^* and \mathbf{P}^T are the distributions that realize the minimum in the definition of V' and $V(\mathbf{M}_2, T-2)$ respectively. Below we show the values of these quantities without giving full details:

$$\begin{aligned} V(\mathbf{M}_2, 1) &= \min_{\mathbf{P}} \max_i \{1 - P_i + V(\mathbf{M}_2 + \mathbf{1} - \mathbf{e}_i, 0)\} \\ &= \min_{\mathbf{P}} \max\{-P_1, -P_2, -P_3 - 1\} \\ &= -1/2, \end{aligned}$$

with $\mathbf{P}^3 = (1/2, 1/2, 0)$;

$$\begin{aligned} V(\mathbf{M}_2, 2) &= \min_{\mathbf{P}} \max_i \{1 - P_i + V(\mathbf{M}_2 + \mathbf{1} - \mathbf{e}_i, 1)\} \\ &= \min_{\mathbf{P}} \max\{-P_1, -P_2, -P_3 - 1/3\} \\ &= -4/9, \end{aligned}$$

with $\mathbf{P}^4 = (4/9, 4/9, 1/9)$;

$$\begin{aligned} V' &= \min_{\mathbf{P}} \max \left(1 - P_i + \frac{1}{2}V(\mathbf{M}_2 + \mathbf{1} - \mathbf{e}_i, 0) \right. \\ &\quad \left. + \frac{1}{2}V(\mathbf{M}_2 + \mathbf{1} - \mathbf{e}_i, 1) \right) \\ &= \min_{\mathbf{P}} \max\{-P_1, -P_2, -P_3 - 2/3\} \\ &= -1/2, \end{aligned}$$

with $\mathbf{P}^* = (1/2, 1/2, 0)$. We thus conclude that

$$\mathbb{E}[\mathbf{P}^T | T \geq 3] = (17/36, 17/36, 1/18) \neq \mathbf{P}^*$$

and

$$\mathbb{E}[V(\mathbf{M}_2, T-2) | T \geq 3] = -17/36 \neq V'.$$

□

The next two examples show that the idea of “treating the current round as the last round” does not work for minimax algorithms.

Example 2. Consider the following Hedge setting: $N = 2$, $\mathbf{LS} = [0, 1]^2$ and the horizon T is an even number. Suppose on round t , the learner chooses \mathbf{P}_t using the minimax algorithm assuming horizon $T = t$. Then the adversary can make the regret after T rounds to be $T/4$ by choosing \mathbf{e}_1 and \mathbf{e}_2 alternatively.

Proof. As shown in Theorem 10, when $N = 2$, the minimax algorithm with $\mathbf{LS} = [0, 1]^2$ is the same as the one with $\mathbf{LS} = \{\mathbf{e}_1, \mathbf{e}_2\}$, which we already know from Theorem 1. If the learner treats the current round as the last round, then $P_{t,1}$ is

$$\begin{aligned} &V(\mathbf{M}_{t-1}, 1) - V(\mathbf{M}_{t-1} + \mathbf{e}_1, 0) \\ &= \frac{1}{2}(1 + \min\{M_{t-1,1} + 1, M_{t-1,2}\} \\ &\quad - \min\{M_{t-1,1}, M_{t-1,2} + 1\}). \end{aligned}$$

Hence, for any round t where t is odd, we have $\mathbf{M}_{t-1} = (\frac{t-1}{2}, \frac{t-1}{2})$ and thus $P_{t,1} = P_{t,2} = 1/2$ and the learner suffers loss $1/2$. For any round t where t is even, we have $\mathbf{M}_{t-1} = (\frac{t}{2}, \frac{t}{2} - 1)$ and thus $P_{t,1} = 0, P_{t,2} = 1$ and the learner suffers loss 1 since the adversary will choose \mathbf{e}_2 for this round. Finally, at the end of T rounds, the loss of the best action is clearly $T/2$. So the regret would be $3T/4 - T/2 = T/4$. □

Example 3. Consider the online linear optimization problem described in Section 6.1. If horizon T is even and the learner predicts using the minimax algorithm Eq (6) with T replaced with t . Then the adversary can make the regret to be $\sqrt{2}T/4$ after T rounds by choosing \mathbf{e}_1 and $-\mathbf{e}_1$ alternatively.

Proof. For any round t where t is odd, we have $\mathbf{W}_{t-1} = \mathbf{0}$ and thus $\mathbf{x}_t = \mathbf{0}$. So the loss for this round is 0. For any round t where t is even, we have $\mathbf{W}_{t-1} = \mathbf{e}_1$ and thus $\mathbf{x}_t = -\frac{\sqrt{2}}{2}\mathbf{e}_1$. So the loss for this round is $\sqrt{2}/2$ since the adversary will pick $-\mathbf{e}_1$. At the end of T rounds, since $\mathbf{W}_T = \mathbf{0}$, the regret will simply be $\sqrt{2}T/4$. □