Minimizing Wide Range Regret with Time Selection Functions

Subhash Khot and Ashok Kumar Ponnuswami
New York University and Georgia Tech
khot@cs.nyu.edu and pashok@cc.gatech.edu

Abstract

We consider the problem of minimizing regret with respect to a given set $\mathcal{S}$ of pairs of time selection functions and modifications rules. We give an online algorithm that has $O(\sqrt{T \log |\mathcal{S}|})$ regret with respect to $\mathcal{S}$ when the algorithm is run for T time steps and there are N actions allowed. This improves the upper bound of $O(\sqrt{TN \log(|\mathcal{I}||\mathcal{F}|)})$ given by Blum and Mansour [BM07a] for the case when $\mathcal{S} = \mathcal{I} \times \mathcal{F}$ for a set $\mathcal{I}$ of time selection functions and a set $\mathcal{F}$ of modification rules. We do so by giving a simple reduction that uses an online algorithm for external regret as a black box.

1 Introduction

We consider the following online optimization problem. At the beginning of each day (a time step), we have to choose one of the N allowed actions. Instead of picking one action deterministically, we may come up with a distribution over the actions. At the end of the day, an adversary, with the knowledge of the distribution we picked, fixes a loss for each action. We give a concrete example from Cesa-Bianchi et al. [CBFH$^+$97]. Suppose we want to predict the probability that it rains on a day based on the predictions of N weather forecasting websites. But we don't know which of these "experts" give good forecasts. We come up with some weights on the websites using an online algorithm and use the weighted prediction as our guess for the probability of raining. At the end of the day, based on whether or not it rained, everyone is incurs a loss depending on how inaccurate their prediction was. Usually it is assumed that the loss for each action is picked from a fixed interval, like $[0, 1]$. For example, we could charge a person who predicts p as the probability of rain $1 - p$ if it rains and p if does not. After T days, we compare the loss incurred by the online algorithm we used to the loss incurred if we had followed a simple strategy (like just picking the same action each day). Our goal is to minimize our *regret* for not following one of the simple strategies. One may also compare the algorithm's performance to the performance if the distribution over actions at each time step were modified using a certain set of rules. We consider the problem of designing algorithms with low regret with respect to a given set of strategies or modification rules.

The most basic regret studied is *external regret*, which is the difference between the loss incurred by the algorithm and the loss incurred by the best action in hindsight. Another kind of regret commonly studied is called *internal regret*. This was introduced by Foster and Vohra [FV98]. Here, we consider the set of modification rules where for each pair (a, b) of actions we have a rule of the kind: Every time the algorithm suggests picking a, pick b instead. The internal regret of the algorithm is the regret of not having applied one of these modification rules. Each rule here can be considered as a function $f_{a,b}$ that maps every action to itself, except action a which gets mapped to b. If we consider the set of modification rules corresponding to all functions mapping the set of actions into itself, we get the notion of *swap regret*. Finally, we can allow any subset of these mappings as the set of allowed modification rules which gives the notion of *wide range regret*. This was defined by Lehrer [Leh03]. Lehrer also associates *time selection* function with each rule that indicates whether a rule is "active" at a give time or not. A related model is that of "sleeping experts" or "specialists" defined in Freund et al. [FSSW97]. Here, at the beginning of time t, each specialist can decide whether are not the current situation is her area of speciality and make a prediction only if it does. In addition, Blum and Mansour [BM07a] consider the case where the experts can be "partially awake". One way to interpret the activeness function is that it measures degree of confidence that the corresponding rule will perform well at a given time. In this case, we weigh the loss incurred by the algorithm and the modified action with the time selection function to calculate the regret.

The first algorithm with external regret sublinear in T was developed by Hannan [Han57]. An algorithm whose external regret has only logarithmic dependence on N was given by Littlestone and Warmuth [LW94] and Cesa-Bianchi et al. [CBFH$^+$97].

Lemma 1 ([CBFH$^+$97]) *There exists an online algorithm with external regret at most $O(\sqrt{T \log N})$ when the losses are picked from $[-1, +1]$. The running time is polynomial in T and N.*

The number of time steps T for which it will be run need not be provided as an input to the above algorithm. Stoltz and Lugosi [SL05] give a general method to convert any "weighted average predictor" algorithm for external regret to a low internal or swap regret algorithm. At a high level, they pretend there is an expert for each modification rule

who always suggests using that rule. At each time step, the expert is charged the loss that would be incurred if his modification rule were actually used. The weighted average predictor would give a distribution over the experts. The distribution over the actual actions is found by computing the fixed point of the expected modification rule picked from the distribution over the experts. This gives algorithms with $O(\sqrt{T \log N})$ internal regret and $O(\sqrt{TN \log N})$ swap regret. Our approach for wide range regret with time selection functions is based on the same idea. A drawback of a swap regret algorithm constructed this way is that it needs to maintain N^N weights. Blum and Mansour [BM07a] give an algorithm that has $O(\sqrt{TN \log N})$ swap regret and runs in time polynomial in N too. They also give an algorithm that has $O(\sqrt{TN \log(KM)})$ regret with respect to K modification rules and M time selection functions. Here, for each modification rule and time selection function, the regret of not having modified the algorithms action by the rule with the losses weighed by the time selection function is considered. In this case, we can think of there being M people who are interested in following an algorithm's predictions. They have varying degrees of importance associated with each day (given by their corresponding time selection function) and want to minimize regret with respect to all the modification rules. The algorithm's goal is to minimize the maximum regret of a person. This is a bit different from the model considered in Lehrer [Leh03]. But with some effort, one can check that the result of Blum and Mansour [BM07a] can be generalized to the model of Lehrer [Leh03]. We refer the reader to [BM07b]for other bounds on the regret minimization and the relation of various kinds of regret to equilibriums in games.

The paper is organized as follows. In the next section, we define the model we work with formally and state our main result. We state the ideas we use from related results in Section 3. We prove our main result of an improved upper bound for wide range regret in Section 4. We conclude with a "first-order" upper bound in Section 5.

2 Our Model and Result

Let the set of actions be $[N] = \{1, 2, \ldots, N\}$. Consider the following T round game between an online algorithm H and an adversary. At the beginning of time $t = 1, 2, \ldots, T$, the algorithm picks a probability vector[1][2] $\boldsymbol{p}^t = (p_1^t, p_2^t, \ldots, p_N^t)$. The adversary then picks the loss vector $\boldsymbol{l}^t = (l_1^t, l_2^t, \ldots, l_N^t)$ for time t. The entries of $\boldsymbol{l}^t$ are picked from a fixed interval. In this paper, we assume the losses are either picked from $[0, 1]$ or from $[-1, +1]$.

Define the regret of H with respect to action $a \in [N]$ to be

$$R_{H,a} = \sum_{t=1}^{T} (\sum_{b \in [N]} p_b^t l_b^t - l_a^t) = \sum_{t=1}^{T} \sum_{b \in [N]} p_b^t (l_b^t - l_a^t).$$

This can be interpreted as the difference between the expected loss of H and the loss of action a. Define the *external regret* of H to be

$$R_{H,ext} = \max_{a \in [N]} R_{H,a}.$$

We now define the model with time selection functions from Blum and Mansour [BM07a]. A time selection function is a function $I : \mathbb{N} \longrightarrow [0, 1]$. Let $\mathcal{I}$ be the set of time selection functions. At the beginning of time t, the adversary sets the values of $I(t)$ for each $I \in \mathcal{I}$. The algorithm then picks $\boldsymbol{p}^t$ after which the adversary now picks $\boldsymbol{l}^t$ as before. Given a modification rule $f : [N] \longrightarrow [N]$, define $\boldsymbol{M}_f$ to be the matrix with a 1 in column $f(i)$ of row i for all i and zeros everywhere else. Define the regret of H with respect to time selection function I and a modification rule f to be

$$\begin{aligned} R_{H,I,f} &= \sum_t I(t) \sum_{a \in [N]} p_a^t (l_a^t - l_{f(a)}^t) \\ &= \sum_t I(t) (\boldsymbol{p}^t \cdot \boldsymbol{l}^t - \boldsymbol{p}^{t\top} \boldsymbol{M}_f \boldsymbol{l}^t). \end{aligned}$$

Informally, we first weigh all the losses at time t by $I(t)$, the significance attached to time t. Then we look at the difference between the expected loss of H and the expected loss if the output of H were modified every time by applying f. That is, we measure the regret of not having played action $f(a)$ every time we played a. Given a set $\mathcal{S}$ of pairs (I, f), where I is a time selection function and f is a modification rule, the *wide range regret* of H with respect to $\mathcal{S}$ is defined as

$$R_{H,\mathcal{S}} = \max_{(I,f) \in \mathcal{S}} R_{H,I,f}.$$

Let $\mathbb{1} : \mathbb{N} \longrightarrow [0, 1]$ be the function that always outputs 1, i.e., $\mathbb{1}(t) = 1$. For simplicity of notation, we will use f to also denote the pair $(\mathbb{1}, f)$ when we are not concerned with time selection functions, in which case we assume that the adversary always sets $\mathbb{1}(t)$ to 1. It is easy to check that external regret is the same $R_{H,\mathcal{F}_{ext}}$ where $\mathcal{F}_{ext} = \{f_a\}_{a \in [N]}$ and $\forall b \in [N] : f_a(b) = a$. The *internal regret* of H is defined to be $R_{H,\mathcal{F}_{int}}$, where $\mathcal{F}_{int} = \{f_{a,b}\}_{a,b \in [N]}$ and $f_{a,b}(a) = b$ while $f_{a,b}(c) = c$ for $c \neq a$. The *swap regret* of H is defined to be $R_{H,\mathcal{F}_{swap}}$, where $\mathcal{F}_{swap}$ is the set of all functions $f : [N] \longrightarrow [N]$.

We prove the following theorem for minimizing wide range regret.

Theorem 2 *There exists an online algorithm H that for any given set $\mathcal{S}$ satisfies*

- $R_{H,\mathcal{S}} = O(\sqrt{T \log |\mathcal{S}|})$ *when the losses are picked from the* $[0, 1]$.
- *The running time of H is polynomial in T, N and $|\mathcal{S}|$.*

Note that this matches (upto a constant) the results for external, internal and swap regret if we are not concerned with time selection functions. A drawback of our apporoach is that if the size of the set $\mathcal{S}$ is large, the running time is high. For example, for swap regret with time selection functions, we may need time polynomial in T and N^N. But for this case, the result of Blum and Mansour already gives a more efficient algorithm with the same regret (upto a constant).

[1] A probability vector is a vector in which the entries are non-negative and sum to 1.

[2] All vectors we consider are column vectors. We will use $^\top$ to denote the transpose.

3 Previous Results

We use ideas from Stoltz and Lugosi [SL05] and Blum and Mansour [BM07a].

We first describe the approach of Stoltz and Lugosi [SL05] for internal regret. The idea is to simulate a low external regret algorithm for $N(N-1)$ imaginary experts. Start with any "weighted average predictor" H_{ext} with low external regret. There are $N(N-1)$ imaginary experts, one for each modification rule $f_{a,b}$. The expert corresponding to $f_{a,b}$ always suggests playing b instead of a. We will specify how the probability weights over the actual actions are calculated from the output of H_{ext} and how the losses are generated for the imaginary experts of H_{ext}.

At time t, suppose H_{ext} outputs probability $q^t_{a,b}$ for the expert corresponding to $f_{a,b}$. Then compute the probability vector $\boldsymbol{p}^t = (p^t_1, p^t_2, \ldots, p^t_N)$ on the actual actions as a fixed point of

$$\boldsymbol{p}^t = \sum_{a,b \in [N]} q^t_{a,b} \boldsymbol{p}^t_{a \to b},$$

where $\boldsymbol{p}^t_{a\to b}$ denotes the probability vector obtained from $\boldsymbol{p}^t$ by changing the weight of action a to zero at putting it on action b. This can also be expressed as

$$\boldsymbol{p}^{t\top} = \sum_{a,b} q^t_{a,b} \boldsymbol{p}^{t\top} \boldsymbol{M}_{f_{a,b}} = \boldsymbol{p}^{t\top} \sum_{a,b} q^t_{a,b} \boldsymbol{M}_{f_{a,b}}.$$

Let the adversary return back $\boldsymbol{l}^t$ as the loss vector at time t. The loss incurred at time t by each of the imaginary experts for $f_{a,b}$ is calculated as

$$l^t_{f_{a,b}} = \boldsymbol{l}^t \cdot \boldsymbol{p}^t_{i\to j} = \boldsymbol{p}^{t\top} \boldsymbol{M}_{f_{a,b}} \boldsymbol{l}^t.$$

This quantity can be thought of as the loss incurred if we followed the expert's suggestion of playing b instead of a. Stoltz and Lugosi [SL05] showed that this achieves low internal regret. For an arbitrary set of modification rules $\mathcal{F}$, we have an expert for each modification rule $f \in \mathcal{F}$ and the probability and loss vectors are now calculated as

$$\boldsymbol{p}^t = \boldsymbol{p}^{t\top} \sum_{f \in \mathcal{F}} q^t_f \boldsymbol{M}_f$$

and

$$l^t_f = \boldsymbol{p}^{t\top} \boldsymbol{M}_f \boldsymbol{l}^t.$$

We now discuss the ideas we use from Blum and Mansour [BM07a]. We start with the case where $\mathcal{S} = \mathcal{I} \times \mathcal{F}_{ext}$ for some $\mathcal{I}$. In this case, there is an expert for each $(I, f_a) \in \mathcal{S}$. There is a weight $w^t_{I,a}$ associated with this expert at the end of time t where

$$w^t_{I,a} = \beta^{-\tilde{R}^t_{I,a}}$$

and

$$\tilde{R}^t_{I,a} = \sum_{t'=1}^{t} I(t')(\beta l^{t'}_H - l^{t'}_a)$$

for some parameter $\beta \in (0,1)$. Above, l^t_H is the actual loss incurred at time t. The quantity $\tilde{R}^t_{I,a}$ is called a "less-strict" external regret. The probability p^t_a associated with the action a at time t is then proportional to $\sum^t_{I\in\mathcal{I}} I(t) w^t_{I,a}$. By optimizing for the parameter β, Blum and Mansour [BM07a] show that this achieves a low external regret with respect to all time selection functions.

To generalize this idea for wide range regret, where $\mathcal{S} = \mathcal{I} \times \mathcal{F}$, they introduce an expert for each $a \in [N]$, $I \in \mathcal{I}$ and $f \in \mathcal{F}$. There is a weight $w^t_{a,I,f}$ for each such expert. Note that this does not simplify to the reduction in the previous paragraph for the case when $\mathcal{F} = \mathcal{F}_{ext}$. Instead, in the next section we obtain a reduction where there are experts only for each $(I, f) \in \mathcal{S}$. Intuitively, this is where we remove the polynomial dependence of wide range regret on N and obtain a slightly simpler reduction.

4 A Reduction from Wide Range Regret to External Regret

We will prove Theorem 2 in this section. We first give an algorithm that when given a low external regret algorithm as a black box uses it to guarantee low wide range regret.

Theorem 3 *Given an algorithm H_{ext} with external regret $R(T,N)$ when the losses are from $[-1,+1]$, one can construct an algorithm H that when given losses from $[0,1]$ satisfies:*

- $R_{H,\mathcal{S}} = R(T, |\mathcal{S}|)$
- *The running time of H is polynomial in the running time of H_{ext}, T, N, and $|\mathcal{S}|$.*

Idea: H will basically simulate an instance of H_{ext} with the elements of $\mathcal{S}$ being the actions. Figure 1 shows the inputs and outputs of H and H_{ext} at time t. At time t, H_{ext} produces some $q^t_{I,f}$ for each $(I,f) \in \mathcal{S}$, where the $q^t_{I,f}$ form a probability distribution over $\mathcal{S}$. H will then use this to come up with a probability vector $\boldsymbol{p}^t = (p^t_1, p^t_2, \ldots, p^t_N)$ on the actual actions. H will basically pick a random (I,f) with probability proportional to $I(t)q^t_{I,f}$. After this, it picks a vector $\boldsymbol{p}^t$ over the actual actions such that $\boldsymbol{p}^t$ is a fixed point of such a random f, i.e., modifying $\boldsymbol{p}^t$ by f in expectation just yields $\boldsymbol{p}^t$. Intuitively, the loss passed to the black box H_{ext} for (I,f) is such that $q^t_{I,f}$ measures the regret with respect to time selection function I of not having modified the output of H using function f. Multiplying this by $I(t)$ takes care of the relevance of (I,f) at time t. Basically, the algorithm makes sure that if the regret with respect to (I,f) was large so far, then that regret doesn't increase at the current step.

Proof: We first specify how H computes $\boldsymbol{p}^t$ and $\boldsymbol{l}'^t$ at time t. To compute $\boldsymbol{p}^t$, get $\boldsymbol{q}^t$ from H_{ext}. If $\sum_{(I,f)\in\mathcal{S}} I(t)q^t_{I,f} = 0$, then output any probability vector $\boldsymbol{p}$. Otherwise define $\boldsymbol{p}^t$ to be any vector satisfying

$$\boldsymbol{p}^{t\top} = \boldsymbol{p}^{t\top} \left(\frac{\sum_{(I,f)\in\mathcal{S}} I(t) q^t_{I,f} \boldsymbol{M}_f}{\sum_{(I,f)\in\mathcal{S}} I(t) q^t_{I,f}} \right). \tag{1}$$

This is well defined since $\sum_{(I,f)} I(t)q^t_{I,f} \neq 0$. Such a vector $\boldsymbol{p}^t$ exists since every row of

$$\frac{\sum_{(I,f)} I(t) q^t_{I,f} \boldsymbol{M}_f}{\sum_{(I,f)} I(t) q^t_{I,f}} \tag{2}$$

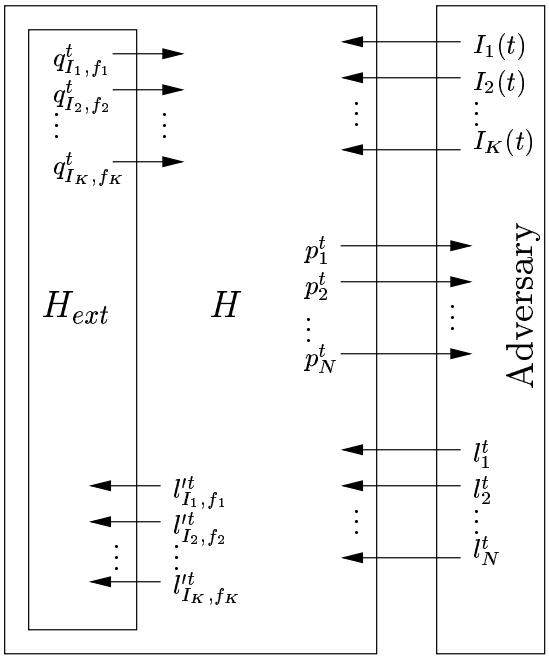

Figure 1: The reduction from wide range to external regret.

is a probability vector because $\boldsymbol{M}_f$ has exactly one 1 in each row. That is, (2) defines the transition matrix of a Markov chain. When H gets back loss vector $\boldsymbol{l}^t$, it computes

$$l'^t_{I,f} = I(t) \sum_{a \in N} p^t_a (l^t_{f(a)} - l^t_a) = I(t) \boldsymbol{p}^{t\top} (\boldsymbol{M}_f - \boldsymbol{I}) \boldsymbol{l}^t$$

where $\boldsymbol{I}$ is the identity matrix. This yields

$$\sum_t l'^t_{I,f} = \sum_t I(t) \boldsymbol{p}^{t\top} (\boldsymbol{M}_f - \boldsymbol{I}) \boldsymbol{l}^t = -R_{H,I,f}. \quad (3)$$

That is, $l'^t_{I,f}$ is exactly the decrease at time t of the regret with respect to (I, f). It is easy to check that $l'^t_{I,f} \in [-1, +1]$.

From the low external regret guarantee of H_{ext}, for all $(I, f) \in \mathcal{S}$:

$$\sum_t \sum_{(J,g) \in \mathcal{S}} q^t_{J,g} l'^t_{J,g} \leq \sum_t l'^t_{I,f} + R(T, |\mathcal{S}|). \quad (4)$$

We will next show that

$$\sum_{(J,g) \in \mathcal{S}} q^t_{J,g} l'^t_{J,g} = 0. \quad (5)$$

Together with (3) and (4), this will show that for all $(I, f) \in \mathcal{S}$,

$$0 \leq -R_{H,I,f} + R(T, |\mathcal{S}|),$$

or $R_{H,I,f} \leq R(T, |\mathcal{S}|)$ which proves the theorem.

We now proceed to prove (5).

$$\sum_{(J,g) \in \mathcal{S}} q^t_{J,g} l'^t_{J,g} = \sum_{(J,g)} q^t_{J,g} J(t) \boldsymbol{p}^{t\top} (\boldsymbol{M}_g - \boldsymbol{I}) \boldsymbol{l}^t$$
$$= \sum_{(J,g)} q^t_{J,g} J(t) \boldsymbol{p}^{t\top} \boldsymbol{M}_g \boldsymbol{l}^t - \sum_{(J,g)} q^t_{J,g} J(t) \boldsymbol{p}^{t\top} \boldsymbol{l}^t$$
$$= \boldsymbol{p}^{t\top} \Big(\sum_{(J,g)} J(t) q^t_{J,g} \boldsymbol{M}_g \Big) \boldsymbol{l}^t - \Big(\sum_{(J,g)} q^t_{J,g} J(t) \Big) (\boldsymbol{p}^{t\top} \boldsymbol{l}^t).$$

(Case 1:) Suppose $\sum_{(J,g)} J(t) q^t_{J,g} \neq 0$. In this case we can use (1) to get

$$\sum_{(J,g) \in \mathcal{S}} q^t_{J,g} l'^t_{J,g} = \Big(\sum_{(J,g)} q^t_{J,g} J(t) \Big) (\boldsymbol{p}^{t\top} \boldsymbol{l}^t) - \Big(\sum_{(J,g)} q^t_{J,g} J(t) \Big) (\boldsymbol{p}^{t\top} \boldsymbol{l}^t) = 0.$$

(Case 2:) Assume $\sum_{(J,g)} J(t) q^t_{J,g} = 0$. Then $J(t) q^t_{J,g} = 0$ for all pairs (J, g) since $J(t)$ and $q^t_{J,g}$ are all non-negative, which implies

$$\sum_{(J,g) \in \mathcal{S}} q^t_{J,g} l'^t_{J,g} = 0.$$

■

It can be seen easily that Theorem 3 and Lemma 1 imply Theorem 2.

5 A First-Order Bound for Wide Range Regret

If we are only concerned with regret bounds as a function of T and N (called "zero-order" bounds in Cesa-Bianchi et al. [CBMS05]), Theorem 2 matches (up to a constant) the known upper bounds for external, internal and swap regret. One can also try to obtain "first-order" bounds, bounds that depend on the sum of payoffs of actions instead of the time. For example, Blum and Mansour [BM07a] show a $O(\sqrt{L_{min} \log(NM)} + \log(NM))$ upper bound for minimizing external regret with respect to a set $\mathcal{I}$ of M time selection functions, where $L_{min} = \max_I \min_a L_{I,a}$ and $L_{I,a} = \sum_t I(t) l^t_a$. For the case when there is at least one "real" expert that does well most of the time, such a bound will be much tighter than a zero-order bound. One can hope to use external regret algorithms with good first-order bounds like the following to come up with good first-order bounds for wide range regret.

Lemma 4 *(Cesa-Bianchi et al. [CBFH+97]) There exists an algorithm with running time polynomial in T and N and external regret $O(\sqrt{L_{min} \log N} + \log N)$ when the losses are picked from $[0, 1]$.*

We need an algorithm that can handle losses from the interval $[-1, +1]$ in Theorem 3. One way to use the algorithm from Lemma 4 is to map the losses $l'^t_{I,f}$ to the interval $[0, 1]$ by a linear transformation. But this also changes the loss of best action and makes the first order bound obtained very weak. Another alternative is to tinker with the quantity that $l'^t_{I,f}$ signifies. If we are concerned only with modification rules (and not time selection functions), we can redefine l'^t_f as

$$l'^t_f = \sum_{a \in N} p^t_a l^t_{f(a)}.$$

But for technical reasons, this can't be done if we are also working with time selection functions. Note that the only term in (4) that depends on I and f is $l'^t_{I,f}$, and hence it must

capture all the terms that depend on *either* I or f in the definition of $R_{H,I,f}$. So we give a method based on the approach of Blum and Mansour [BM07a]. The main idea is to define a *reduced regret* for each pair (I, f).

Theorem 5 *There exists an online algorithm that for any $\mathcal{S}$ satisfies:*

- *The wide range regret with respect to $\mathcal{S}$ is at most $O(\sqrt{L_{min}\log|\mathcal{S}|} + \log|\mathcal{S}|)$, where*

$$L_{min} = \max_{I} \min_{(I,f)\in\mathcal{S}} \sum_{t} I(t)\boldsymbol{p}^{t\top}\boldsymbol{M}_f\boldsymbol{l}^t.$$

- *The running time is polynomial in T, N, and $|\mathcal{S}|$.*

Proof: Define the loss of H with respect to I till time t as

$$L^t_{H,I} = \sum_{t'=1}^{t} I(t)\boldsymbol{p}^t \cdot \boldsymbol{l}^t,$$

and the loss of H with respect to (I, f) till time t as

$$L^t_{H,I,f} = \sum_{t'=1}^{t} I(t)\boldsymbol{p}^{t\top}\boldsymbol{M}_f\boldsymbol{l}^t.$$

We assume that at any time t, not all $I(t)$ are zero. This is without loss of generality since in this case, the losses defined above don't change at time t. For some $\beta \in (0, 1)$ to be fixed later, we basically run a exponentially weighted predictor with a weight for each pair (I, f). The weight of (I, f) at the end of time t is $w^t_{I,f} = \beta^{-\tilde{R}^t_{H,I,f}}$, where

$$\tilde{R}^t_{H,I,f} = \beta L^t_{H,I} - L^t_{H,I,f}.$$

That is, $\tilde{R}^t_{H,I,f}$ is a regret of H with respect to (I, f) where the incurred loss is reduced by a factor β. We define $q^t_{I,f} = w^{t-1}_{I,f}/W^{t-1}$, where $W^t = \sum_{(I,f)\in\mathcal{S}} w^t_{I,f}$ is the sum of the weights.

At time t, the algorithm does the following. It computes $q^t_{I,f}$ as above. The probability vector $\boldsymbol{p}^t$ over the actual actions is picked as in (2). This is well defined since $w^t_{I,f}$ (and hence $q^t_{I,f}$) are all non-zero and at least one of the $I(t)$ is also non-zero (by assumption). Then the algorithm updates all the losses and weights when it gets back $\boldsymbol{l}^t$ from the adversary. We first show that the sum of the weights can not increase at any time.

Claim 6

$$\forall t : \sum_{(I,f)\in\mathcal{S}} w^t_{I,f} \le \sum_{(I,f)\in\mathcal{S}} w^{t-1}_{I,f}$$

Proof: We will use the fact that for any $\beta \in (0, 1)$ and $x \in [0, 1]$, $\beta^x \le 1 - (1-\beta)x$ and $\beta^{-x} \le 1 + (1-\beta)x/\beta$. This gives

$$\begin{aligned}
\sum_{(I,f)\in\mathcal{S}} w^t_{I,f} &= \sum_{(I,f)} w^{t-1}_{I,f}\beta^{I(t)(\boldsymbol{p}^{t\top}\boldsymbol{M}_f\boldsymbol{l}^t - \boldsymbol{p}^t\cdot\boldsymbol{l}^t)} \\
&\le \sum_{(I,f)} \Bigg[w^{t-1}_{I,f}\Big(1-(1-\beta)I(t)\boldsymbol{p}^{t\top}\boldsymbol{M}_f\boldsymbol{l}^t\Big) \\
&\qquad \times \Big(1+(1-\beta)I(t)\boldsymbol{p}^t\cdot\boldsymbol{l}^t\Big)\Bigg] \\
&\le \sum_{(I,f)} w^{t-1}_{I,f} - \left[(1-\beta)W^{t-1}\sum_{(I,f)} q^t_{I,f}I(t)\boldsymbol{p}^{t\top}\boldsymbol{M}_f\boldsymbol{l}^t\right] \\
&\qquad + \left[(1-\beta)W^{t-1}\sum_{(I,f)} q^t_{I,f}I(t)\boldsymbol{p}^t\cdot\boldsymbol{l}^t\right] \\
&= \sum_{(I,f)} w^{t-1}_{I,f} - \left[(1-\beta)W^{t-1}\boldsymbol{p}^{t\top}\Big(\sum_{(I,f)} q^t_{I,f}I(t)\boldsymbol{M}_f\Big)\boldsymbol{l}^t\right] \\
&\qquad + \left[(1-\beta)W^{t-1}\Big(\sum_{(I,f)} q^t_{I,f}I(t)\Big)(\boldsymbol{p}^t\cdot\boldsymbol{l}^t)\right] \\
&= \sum_{(I,f)} w^{t-1}_{I,f}.
\end{aligned}$$

Above, the second inequality follows from the definition of $q^t_{I,f}$ and the last equality follows from (2). ∎

We now get back to the proof of the theorem. The claim implies that for all $(I, f) \in \mathcal{S}$,

$$\beta^{-(\beta L^T_{H,I} - L^T_{H,I,f})} = \beta^{-\tilde{R}^T_{H,I,f}} = w^T_{I,f} \le \sum_{(J,g)\in\mathcal{S}} w^0_{J,g} = |\mathcal{S}|$$

which gives

$$(\beta L^T_{H,I} - L^T_{H,I,f})\log(1/\beta) \le \log|\mathcal{S}|$$

or

$$L_{H,I} \le \frac{L_{H,I,f} + \frac{\log|\mathcal{S}|}{\log(1/\beta)}}{\beta}.$$

Since for a given I, the statement is true for all f such that $(I, f) \in \mathcal{S}$, we can rewrite it as:

$$L_{H,I} \le \frac{L_{H,I,min} + \frac{\log|\mathcal{S}|}{\log(1/\beta)}}{\beta}$$

where

$$L_{H,I,min} = \min_{f:(I,f)\in\mathcal{S}} L_{H,I,f}.$$

Setting β so that

$$\beta^{-1} = 1 + \min\left\{\sqrt{\frac{\log|\mathcal{S}|}{L_{min}}}, \frac{1}{2}\right\}$$

gives the theorem. ∎

Acknowledgments

The authors would like to thank Yishay Mansour for comments on an early draft of the paper.

References

[BM07a] Avrim Blum and Yishay Mansour. From external to internal regret. *J. Mach. Learn. Res.*, 8:1307–1324, 2007.

[BM07b] Avrim Blum and Yishay Mansour. Learning, regret minimization and equilibria. In Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay Vazirani, editors, *Algorithmic Game Theory*, chapter 4. Cambridge University Press, 2007.

[CBFH+97] Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth. How to use expert advice. *J. ACM*, 44(3):427–485, 1997.

[CBMS05] Nicolò Cesa-Bianchi, Yishay Mansour, and Gilles Stoltz. Improved second-order bounds for prediction with expert advice. In *COLT*, pages 217–232, 2005.

[FSSW97] Yoav Freund, Robert E. Schapire, Yoram Singer, and Manfred K. Warmuth. Using and combining predictors that specialize. In *STOC '97: Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 334–343, 1997.

[FV98] Dean Foster and Rakesh V. Vohra. Asymptotic calibration. *Biometrika*, 85:379–390, 1998.

[Han57] J. Hannan. Approximation to bayes risk in repeated plays. In M.Dresher, A. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*, volume 3, pages 97–139. Princeton University Press, 1957.

[Leh03] Ehud Lehrer. A wide range no-regret theorem. *Games and Economic Behavior*, 42(1):101–115, 2003.

[LW94] Nick Littlestone and Manfred K. Warmuth. The weighted majority algorithm. *Inf. Comput.*, 108(2):212–261, 1994.

[SL05] Gilles Stoltz and Gábor Lugosi. Internal regret in on-line portfolio selection. *Mach. Learn.*, 59(1-2):125–159, 2005.