Prediction by random-walk perturbation

Luc Devroye LUCDEVROYE@GMAIL.COM
School of Computer Science
McGill University

Gábor Lugosi GABOR.LUGOSI@GMAIL.COM
ICREA and Department of Economics
Universitat Pompeu Fabra

Gergely Neu GERGELY.NEU@GMAIL.COM
MTA SZTAKI Institute for Computer Science and Control
and
Department of Computer Science and Information Theory
Budapest University of Technology and Economics

Editor:

Abstract

We propose a version of the follow-the-perturbed-leader online prediction algorithm in which the cumulative losses are perturbed by independent symmetric random walks. The forecaster is shown to achieve an expected regret of the optimal order $O(\sqrt{n \log N})$ where n is the time horizon and N is the number of experts. More importantly, it is shown that the forecaster changes its prediction at most $O(\sqrt{n \log N})$ times, in expectation. We also extend the analysis to online combinatorial optimization and show that even in this more general setting, the forecaster rarely switches between experts while having a regret of near-optimal order. This is the first forecaster with such a proven property.

Keywords: Online learning, Follow the Perturbed Leader, random walk

1. Preliminaries

In this paper we study the problem of online prediction with expert advice, see Cesa-Bianchi and Lugosi (2006). The problem may be described as a repeated game between a *forecaster* and an adversary–the *environment*. At each time instant $t = 1, \ldots, n$, the forecaster chooses one of the N available actions (often called *experts*) and suffers a loss $\ell_{i,t} \in [0, 1]$ corresponding to the chosen action i. We consider the so-called *oblivious adversary* model in which the environment selects all losses before the prediction game starts and reveals the losses $\ell_{i,t}$ at time t after the forecaster has made its prediction. The losses are deterministic but the forecaster may randomize: at time t, the forecaster chooses a probability distribution $\boldsymbol{p}_t$ over the set of N actions and draws a random action I_t according to the distribution $\boldsymbol{p}_t$. The prediction protocol is described in Figure 1.

The usual goal for the standard prediction problem is to devise an algorithm such that the cumulative loss $\widehat{L}_n = \sum_{t=1}^n \ell_{I_t,t}$ is as small as possible, in expectation and/or with high probability (where probability is with respect to the forecaster's randomization). Since we do not make any assumption on how the environment generates the losses $\ell_{i,t}$, we cannot hope to minimize the above the cumulative loss. Instead, a meaningful goal is to minimize the performance gap between our algorithm and the strategy that selects the best action chosen in hindsight. This performance gap is called the *regret* and is defined formally as

$$R_n = \max_{i \in \{1,2,\ldots,N\}} \sum_{t=1}^{n} (\ell_{I_t,t} - \ell_{i,t}) = \widehat{L}_n - L_n^*,$$

where have we also introduced the notation $L_n^* = \min_{i \in \{1,2,\ldots,N\}} \sum_{t=1}^n \ell_{i,t}$. Minimizing the regret defined above is a well studied problem. It is known that no matter what algorithm the forecaster uses,

$$\liminf_{n,N \to \infty} \sup \frac{\mathbb{E} R_n}{\sqrt{(n/2) \ln N}} \geq 1$$

where the supremum is taken with respect to all possible loss assignments with losses in $[0,1]$ (see, e.g., Cesa-Bianchi and Lugosi (2006)). On the other hand, several prediction algorithms are known whose expected regret is of optimal order $O(\sqrt{n \log N})$ and many of them achieve a regret of this order with high probability. Perhaps the most popular one is the exponentially weighted average forecaster (a variant of weighted majority algorithm of Littlestone and Warmuth (1994), and aggregating strategies of Vovk (1990), also known as *Hedge* by Freund and Schapire (1997)). The exponentially weighted average forecaster assigns probabilities to the actions that are inversely proportional to an exponential function of the loss accumulated by each action up to time t.

Another popular forecaster is the *follow the perturbed leader* (FPL) algorithm of Hannan, 1957. Kalai and Vempala (2003) showed that Hannan's forecaster, when appropriately modified, indeed achieves an expected regret of optimal order. At time t, the FPL forecaster adds a random perturbation $Z_{i,t}$ to the cumulative loss $L_{i,t-1} = \sum_{s=1}^{t-1} \ell_{i,s}$ of each action and chooses an action that minimizes the sum $L_{i,t-1} + Z_{i,t}$. If the vector of random variables $\boldsymbol{Z}_t = (Z_{1,t}, \ldots, Z_{N,t})$ have joint density $(\eta/2)^N e^{-\eta \|z\|_1}$ for $\eta \sim \sqrt{\log N/n}$, then the expected regret of the forecaster is of order $O(\sqrt{n \log N})$ (Kalai and Vempala (2003), see also Cesa-Bianchi and Lugosi (2006), Hutter and Poland (2004), Poland (2005)). This is true whether $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ are independent or not. It they are independent, then one may show that the regret is concentrated around its expectation. Another interesting choice is when $\boldsymbol{Z}_1 = \cdots = \boldsymbol{Z}_n$, that is, the same perturbation is used over time. Even though this forecaster has an expected regret of optimal order, the regret is much less concentrated and may fail with reasonably high probability.

Small regret is not the only desirable feature of an online forecasting algorithm. In many applications, on would like to define forecasters that do not change their prediction too often. Examples of such problems include the online buffering problem described by Geulen et al. (2010) and the online lossy source coding problem of György and Neu (2011). A more abstract problem where the number of abrupt switches in the behavior is costly is the problem of online learning in Markovian decision processes, as described by Even-Dar et al. (2009) and Neu et al. (2010).

Parameters: set of actions $\mathcal{I} = \{1, 2, \ldots, N\}$, number of rounds n;
The environment chooses the losses $\ell_{i,t} \in [0, 1]$ for all $i \in \{1, 2, \ldots, N\}$ and $t = 1, \ldots, n$.
For all $t = 1, 2, \ldots, n$, repeat

1. The forecaster chooses a probability distribution $\boldsymbol{p}_t$ over $\{1, 2, \ldots, N\}$.
2. The forecaster draws an action I_t randomly according to $\boldsymbol{p}_t$
3. The environment reveals $\ell_{i,t}$ for all $i \in \{1, 2, \ldots, N\}$.
4. The forecaster suffers loss $\ell_{I_t,t}$.

Figure 1: Prediction with expert advice.

To be precise, define the number of action switches up to time n by

$$C_n = |\{1 < t \leq n : I_{t-1} \neq I_t\}| \ .$$

In particular, we are interested in defining randomized forecasters that achieve a regret R_n of the order $O(\sqrt{n \log N})$ while keeping the number of action switches C_n as small as possible. However, the usual forecasters with small regret–such as the exponentially weighted average forecaster or the FPL forecaster with i.i.d. perturbations–may switch actions a large number–typically $\Theta(n)$–times. Therefore, the design of special forecasters with small regret and small number of action switches is called for.

The first paper to explicitly attack this problem is by Geulen et al. (2010), who propose a variant of the exponentially weighted average forecaster called the "shrinking dartboard" algorithm and prove that it provides an expected regret of $O(\sqrt{n \log N})$, while guaranteeing that the number of switches is at most $O(\sqrt{n \log N})$ with high probability. A less conscious attempt to solve the problem is due to Kalai and Vempala (2005b); they show that the simplified version of the FPL algorithm with identical perturbations (as described above) guarantees an $O(\sqrt{n \log N})$ bound on both the expected regret and the expected number of switches. propose a method based on FPL in which perturbations are defined by independent symmetric random walks. We show that this, intuitively appealing, forecaster has similar regret and switch-number guarantees as shrinking dartboard and FPL with identical perturbations. A further important advantage of the new forecaster is that it may be used simply in the more general problem of *online combinatorial*—or, more generally, *linear—optimization*. We postpone the definitions and the statement of the results to Section 4 below.

2. The algorithm

To address the problem described in the previous section, we propose a variant of the Follow the Perturbed Leader (FPL) algorithm. The proposed forecaster perturbs the loss of each action at every time instant by a symmetric coin flip and chooses an action with minimal cumulative perturbed loss. More precisely, the algorithm draws the independent random variables $X_{i,t}$ that take values $\pm 1/2$ with equal probabilities and $X_{i,t}$ is added to each loss

$\ell_{i,t-1}$. At time t action i is chosen that minimizes $\sum_{s=1}^{t} (\ell_{i,t-1} + X_{i,t})$ (where we define $\ell_{i,0} = 0$).

Algorithm 1 The algorithm.

Initialization: set $L_{i,0} = 0$ and $Z_{i,0} = 0$ for all $i = 1, 2, \ldots, N$.
For all $t = 1, 2, \ldots, n$**, repeat**

1. Draw $X_{i,t}$ for all $i = 1, 2, \ldots, N$ such that
$$X_{i,t} = \begin{cases} \frac{1}{2} & \text{with probability } \frac{1}{2} \\ -\frac{1}{2} & \text{with probability } \frac{1}{2}. \end{cases}$$
2. Let $Z_{i,t} = Z_{i,t-1} + X_{i,t}$ for all $i = 1, 2, \ldots, N$.
3. Choose action
$$I_t = \arg\min_i \left(L_{i,t-1} + Z_{i,t} \right).$$
4. Observe losses $\ell_{i,t}$ for all $i = 1, 2, \ldots, N$, suffer loss $\ell_{I_t,t}$.
5. Set $L_{i,t} = L_{i,t-1} + \ell_{i,t}$ for all $i = 1, 2, \ldots, N$.

Equivalently, the forecaster may be thought of as an FPL algorithm in which the cumulative losses $L_{i,t-1}$ are perturbed by $Z_{i,t} = \sum_{i=1}^{t} X_{i,t}$. Since for each fixed i, $Z_{i,1}, Z_{i,2}, \ldots$ is a symmetric random walk, cumulative losses of the N actions are perturbed by N independent symmetric random walks. This is the way the algorithm is presented in Algorithm 1.

A simple variation is when one replaces random coin flips by independent standard normal random variables. Both have similar performance guarantees and we choose $\pm(1/2)$-valued perturbations for mathematical convenience. In Section 4 we switch to normally distributed perturbations–again driven by mathematical simplicity. In practice both versions are expected to have a similar behavior.

Conceptually, the difference between standard FPL and the proposed version is the way the perturbations are generated: while common versions of FPL use perturbations that are generated in an i.i.d. fashion, the perturbations of the algorithm proposed here are dependent. This will enable us to control the number of action switches during the learning process. Note that the standard deviation of these perturbations is still of order $\sqrt{t}$ just like for the standard FPL forecaster with optimal parameter settings.

To obtain intuition why this approach will solve our problem, first consider a problem with $N = 2$ actions and an environment that generates equal losses, say $\ell_{i,t} = 0$ for all i and t, for all actions. When using i.i.d. perturbations, FPL switches actions with probability $1/2$ in each round, thus yielding $C_t = t/2 + O(\sqrt{t})$ with overwhelming probability. The same holds for the exponentially weighted average forecaster. On the other hand, when using the random-walk perturbations described above, we only switch between the actions when the leading random walk is changed, that is, when the difference of the two random walks–which is also a symmetric random walk–hits zero. It is a well known that the number of occurrences of this event up to time t is $O_p(\sqrt{t})$, see, Feller (1968). As we show below, this is the worst case for the number of switches.

3. Performance bounds

The next theorem summarizes our performance bounds for the proposed forecaster.

Theorem 1 *The expected regret and expected number of switches of actions of the forecaster of Algorithm 1 satisfy, for all possible loss sequences (under the oblivious-adversary model),*

$$\mathbb{E}R_n \le 2\mathbb{E}C_n \le 8\sqrt{2n \log N} + 16 \log n + 16 \ .$$

Remark. Even though we only prove bounds for the expected regret and the expected number of switches, it is of great interest to understand upper tail probabilities. However, this is a highly nontrivial problem. One may get an intuition by considering the case when $N = 2$ and all losses are equal to zero. In this case the algorithm switches actions whenever a symmetric random walk returns to zero. This distribution is well understood and the probability that this occurs more than $x\sqrt{n}$ times during the first n steps is roughly $2\mathbb{P}\{N > 2x\} \le 2e^{-2x^2}$ where N is a standard normal random variable (see (Feller, 1968, Section III.4)). Thus, in this case we see that bot the number of switches and the regret are bounded by $O\left(\sqrt{n \log(1/\delta)}\right)$, with probability at least $1-\delta$. However, proving analog bounds for the general case remains a challenge.

To prove the theorem, we first show that the regret can be bounded in terms of the number of action switches. Then we turn to analyzing the expected number of action switches.

3.1 Regret and number of switches

The next simple lemma shows that the regret of the forecaster may be bounded in terms of the number of times the forecaster switches actions.

Lemma 2 *Fix any $i \in \{1, 2, \ldots, N\}$. Then*

$$\widehat{L}_n - L_{i,n} \le 2C_n + Z_{i,n+1} - \sum_{t=1}^{n+1} X_{I_{t-1},t} \ .$$

Proof We apply Lemma 3.1 of Cesa-Bianchi and Lugosi (2006) (sometimes referred to as the *"be-the-leader"* lemma) for the sequence $(\ell_{\cdot,t-1} + X_{\cdot,t})_{t=1}^{\infty}$ with $\ell_{j,0} = 0$ for all $j \in \{1, 2, \ldots, N\}$, obtaining

$$\begin{aligned} \sum_{t=1}^{n+1} (\ell_{I_t,t-1} + X_{I_t,t}) &\le \sum_{t=1}^{n+1} (\ell_{i,t-1} + X_{i,t}) \\ &= L_{i,n} + Z_{i,n+1} \ . \end{aligned}$$

Reordering terms, we get

$$\sum_{t=1}^{n} \ell_{I_t,t} \le L_{i,n} + \sum_{t=1}^{n} \left(\ell_{I_t,t} - \ell_{I_{t+1},t}\right) + Z_{i,n} - \sum_{t=1}^{n+1} X_{I_t,t} \ . \qquad (1)$$

The last term can be rewritten as

$$-\sum_{t=1}^{n+1} X_{I_t,t} = -\sum_{t=1}^{n+1} X_{I_{t-1},t} + \sum_{t=1}^{n+1} \left(X_{I_{t-1},t} - X_{I_t,t} \right) \ .$$

Now notice that $X_{I_{t-1},t} - X_{I_t,t}$ and $\ell_{I_{t-1},t-1} - \ell_{I_t,t-1}$ are both zero when $I_t = I_{t-1}$ and are upper bounded by 1 otherwise. That is, we get that

$$\sum_{t=1}^{n+1} \left(\ell_{I_{t-1},t-1} - \ell_{I_t,t-1} \right) + \sum_{t=1}^{n+1} \left(X_{I_{t-1},t} - X_{I_t,t} \right) \le 2 \sum_{t=1}^{n+1} \mathbb{I}\left\{ I_{t-1} \ne I_t \right\} = 2C_n \ .$$

Putting everything together gives the statement of the lemma. ∎

3.2 Bounding the number of switches

Next we analyze the number of switches C_n. In particular, we upper bound the marginal probability $\mathbb{P}\left[I_{t+1} \ne I_t \right]$ for each $t \ge 1$. We define the *lead pack* A_t as the set of actions that, at time t, have a positive probability of taking the lead at time $t+1$:

$$A_t = \left\{ i \in \{1, 2, \ldots, N\} : L_{i,t-1} + Z_{i,t} \le \min_j \left(L_{j,t-1} + Z_{j,t} \right) + 2 \right\} \ .$$

We bound the probability of lead change as

$$\mathbb{P}\left[I_t \ne I_{t+1} \right] \le \frac{1}{2} \mathbb{P}\left[|A_t| > 1 \right] \ .$$

The key to the proof of the theorem is the following lemma that gives an upper bound for the probability that the lead pack contains more than one action. It implies, in particular, that

$$\mathbb{E}\left[C_n \right] \le 4\sqrt{2n \log N} + 8 \log n + 8 \ ,$$

which is what we need to prove the expected-value bounds of Theorem 1.

Lemma 3

$$\mathbb{P}\left[|A_t| > 1 \right] \le 4 \sqrt{2 \frac{\log N}{t}} + \frac{8}{t} \ .$$

Proof Define $p_t(k) = \mathbb{P}\left[Z_{i,t} = \frac{k}{2} \right]$ for all $k = -t, \ldots, t$ and we let S_t denote the set of leaders at time t (so that the forecaster picks $I_t \in S_t$ arbitrarily):

$$S_t = \left\{ j \in \{1, 2, \ldots, N\} : L_{j,t-1} + Z_{j,t} = \min_i \left\{ L_{i,t-1} + Z_{i,t} \right\} \right\} \ .$$

Let us start with analyzing $\mathbb{P}[|A_t| = 1]$:

$$\begin{aligned}
\mathbb{P}[|A_t| = 1] &= \sum_{k=-t}^{t}\sum_{j=1}^{N} p_t(k)\mathbb{P}\left[\min_{i\in\{1,2,\ldots,N\}\setminus j}\{L_{i,t-1}+Z_{i,t}\} \geq L_{j,t-1}+\frac{k}{2}+2\right] \\
&\geq \sum_{k=-t}^{t-4}\sum_{j=1}^{N} p_t(k+4)\mathbb{P}\left[\min_{i\in\{1,2,\ldots,N\}\setminus j}\{L_{i,t-1}+Z_{i,t}\} \geq L_{j,t-1}+\frac{k+4}{2}\right]\frac{p_t(k)}{p_t(k+4)} \\
&= \sum_{k=-t+4}^{t}\sum_{j=1}^{N} p_t(k)\mathbb{P}\left[\min_{i\in\{1,2,\ldots,N\}\setminus j}\{L_{i,t-1}+Z_{i,t}\} \geq L_{j,t-1}+\frac{k}{2}\right]\frac{p_t(k-4)}{p_t(k)}\ .
\end{aligned}$$

Before proceeding, we need to make two observations. First of all,

$$\begin{aligned}
\sum_{j=1}^{N} p_t(k)\mathbb{P}\left[\min_{i\in\{1,2,\ldots,N\}\setminus j}\{L_{i,t-1}+Z_{i,t}\} \geq L_{j,t-1}+\frac{k}{2}\right] &\geq \mathbb{P}\left[\exists j\in S_t : Z_{j,t}=\frac{k}{2}\right] \\
&\geq \mathbb{P}\left[\min_{j\in S_t} Z_{j,t}=\frac{k}{2}\right],
\end{aligned}$$

where the first inequality follows from the union bound and the second from the fact that the latter event implies the former. Also notice that $Z_{i,t}+\frac{t}{2}$ is binomially distributed with parameters t and $1/2$ and therefore

$$p_t(k) = \binom{t}{\frac{t+k}{2}}\frac{1}{2^t}\ .$$

Hence

$$\begin{aligned}
\frac{p_t(k-4)}{p_t(k)} &= \frac{\left(\frac{t+k}{2}\right)!\left(\frac{t-k}{2}\right)!}{\left(\frac{t+k}{2}-2\right)!\left(\frac{t-k}{2}+2\right)!} \\
&= \frac{(t+k)(t+k-2)}{(t-k+2)(t-k+4)} \\
&= \frac{t^2+2tk+k^2-2k-2t}{t^2-2tk+k^2-6k+6t+8} \\
&= 1+\frac{4(t+1)(k-2)}{(t-k+2)(t-k+4)}\ .
\end{aligned}$$

It can be easily verified that

$$\frac{4(t+1)(k-2)}{(t-k+2)(t-k+4)} \geq \frac{4(t+1)(k-2)}{(t+2)(t+4)}$$

holds for all $k\in[-t,t]$. Putting these observations together, we get

$$\begin{aligned}
\mathbb{P}[|A_t| = 1] &\geq \sum_{j}\sum_{k=-t+4}^{t} p_t(k)\mathbb{P}\left[\min_{i\in\{1,2,\ldots,N\}\setminus j}\{L_{i,t-1}+Z_{l,t}\} \geq L_{j,t-1}+\frac{k}{2}\right]\frac{p_t(k-4)}{p_t(k)} \\
&\geq \sum_{k=-t+4}^{t}\mathbb{P}\left[\min_{j\in S_t} Z_{j,t}=\frac{k}{2}\right]\frac{p_t(k-4)}{p_t(k)}\ .
\end{aligned}$$

This implies

$$
\begin{aligned}
\mathbb{P}\left[|A_t| > 1\right] \leq & 1 - \sum_{k=-t+4}^{t} \mathbb{P}\left[\min_{j \in S_t} Z_{j,t} = \frac{k}{2}\right] \frac{p_t(k-4)}{p_t(k)} \\
\leq & 1 - \sum_{k=-t}^{t} \mathbb{P}\left[\min_{j \in S_t} Z_{j,t} = \frac{k}{2}\right] \left(1 + \frac{4(t+1)(k-2)}{(t+2)(t+4)}\right) \\
= & \sum_{k=-t}^{t} \mathbb{P}\left[\min_{j \in S_t} Z_{j,t} = \frac{k}{2}\right] \left(\frac{4(2-k)(t+1)}{(t+2)(t+4)}\right) \\
= & \frac{8(t+1)}{(t+2)(t+4)} - 8\frac{t+1}{(t+2)(t+4)} \mathbb{E}\left[\min_{j \in S_t} Z_{j,t}\right] \\
\leq & \frac{8}{t} + \frac{8}{t}\mathbb{E}\left[\max_{j \in \{1,2,\ldots,N\}} Z_{j,t}\right] .
\end{aligned}
$$

Now using

$$
\mathbb{E}\left[\max_{j} Z_{j,t}\right] \leq \sqrt{\frac{t \log N}{2}}
$$

implies

$$
\mathbb{P}\left[|A_t| > 1\right] \leq 4\sqrt{\frac{2 \log N}{t}} + \frac{8}{t}
$$

as desired. ■

4. Online combinatorial optimization

In this section we study the case of online linear optimization (see, among others, Gentile and Warmuth (1998), Kivinen and Warmuth (2001), Grove et al. (2001), Takimoto and Warmuth (2003), Kalai and Vempala (2005a), Warmuth and Kuzmin (2008), Helmbold and Warmuth (2009), Hazan et al. (2010), Koolen et al. (2010), Audibert et al. (2011)). This is a similar prediction problem as the one described in the introduction but here each action i is represented by a vector $\boldsymbol{v}_i \in \mathbb{R}^d$. The loss corresponding to action i at time t equals $\boldsymbol{v}_i^\top \boldsymbol{\ell}_t$ where $\boldsymbol{\ell}_t \in [0,1]^d$ is the so-called *loss vector*. Thus, given a set of actions $\mathcal{S} = \{\boldsymbol{v}_i : i = 1, 2, \ldots, N\} \subseteq \mathbb{R}^d$, at every time instant t, the forecaster chooses, in a possibly randomized way, a vector $\boldsymbol{V}_t \in \mathcal{S}$ and suffers loss $\boldsymbol{V}_t^\top \boldsymbol{\ell}_t$. We denote by $\widehat{L_n} = \sum_{t=1}^{n} \boldsymbol{V}_t^\top \boldsymbol{\ell}_t$ the cumulative loss of the forecaster and the regret becomes

$$
\widehat{L_n} - \min_{\boldsymbol{v} \in \mathcal{S}} \boldsymbol{v}^\top \boldsymbol{L}_t
$$

where $\boldsymbol{L}_t = \sum_{s=1}^{t} \boldsymbol{\ell}_t$ is the cumulative loss vector. Of course, one may treat each $\boldsymbol{v}_i \in \mathcal{S}$ as a separate action and the results of the previous section hold but one may gain important computational advantage by taking the structure of the action set into account. In particular, as Kalai and Vempala (2005a) emphasize, FPL-type forecasters may often be computed efficiently. In this section we propose such a forecaster which adds a random-walk perturbation to each *component* of the loss vector. To gain simplicity in the presentation, we restrict our attention to the case of *online combinatorial optimization* in which $\mathcal{S} \subset \{0,1\}^d$, that is, each action is represented a binary vector. This special case arguably contains most important applications such a the *online shortest path* problem. In this example, a fixed directed acyclic graph of d edges is given with two distinguished vertices u and w. The forecaster, at every time instant t, chooses a directed path from u to w. Such a path is represented by it binary incidence vector $\boldsymbol{v} \in \{0,1\}^d$. The components of the loss vector $\boldsymbol{\ell}_t \in [0,1]^d$ represent losses assigned to the d edges and $\boldsymbol{v}^\top \boldsymbol{\ell}_t$ is the total loss assigned to the path $\boldsymbol{v}$.

Another (non-essential) simplifying assumption is that every action $\boldsymbol{v} \in \mathcal{S}$ has the same number of 1's: $\|\boldsymbol{v}\|_1 = m$ for all $\boldsymbol{v} \in \mathcal{S}$. The value of m plays an important role in the bounds below.

The proposed prediction algorithm is defined as follows. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent Gaussian random vectors taking values in $\mathbb{R}^d$ such that the components of each $\boldsymbol{X}_t$ are i.i.d. normal $X_{i,t} \sim \mathcal{N}(0, \eta^2)$ for some fixed $\eta > 0$ whose value will be specified later. Denote

$$\boldsymbol{Z}_t = \sum_{s=1}^{t} \boldsymbol{X}_t \ .$$

The forecaster at time t, chooses the action

$$\boldsymbol{V}_t = \arg\min_{\boldsymbol{v} \in \mathcal{S}} \left\{ \boldsymbol{v}^\top \left(\boldsymbol{L}_{t-1} + \boldsymbol{Z}_t \right) \right\},$$

where $\boldsymbol{L}_t = \sum_{s=1}^{t} \boldsymbol{\ell}_t$ for $t \geq 1$ and $\boldsymbol{L}_0 = (0, \ldots, 0)^\top$.

The next theorem bounds the performance of the proposed forecaster. Again, we are not only interested in the regret but also the number of switches $\sum_{t=1}^{n} \mathbb{I}\{\boldsymbol{V}_{t+1} \neq \boldsymbol{V}_t\}$. The regret of similar order–roughly $m\sqrt{dn}$–as that of the standard FPL forecaster, up to a logarithmic factor. Moreover, the expected number of switches is $O\left(m^2 (\log d)^{5/2} \sqrt{n}\right)$. Remarkably, the dependence on d is only polylogarithmic and it is the weight m of the actions that plays an important role.

Theorem 4 *The expected regret and the expected number of action switches satisfy (under the oblivious adversary model)*

$$\mathbb{E}\widehat{L}_n - \boldsymbol{v}^\top \boldsymbol{L}_n \leq m\sqrt{n} \left(\frac{2d}{\eta} + \eta\sqrt{2\log d} \right) + \frac{md(\log n + 1)}{\eta^2}$$

and

$$\begin{aligned} \mathbb{E}\sum_{t=1}^{n} \mathbb{I}\{\boldsymbol{V}_{t+1} \neq \boldsymbol{V}_t\} \quad &\leq \quad \sum_{t=1}^{n} \frac{m^2 \left(1 + 2\eta\left(2\log d + \sqrt{2\log d} + 1\right) + \eta^2 \left(2\log d + \sqrt{2\log d} + 1\right)^2\right)}{4\eta^2 t} \\ &\quad + \sum_{t=1}^{n} \frac{m^2 \left(1 + \eta\left(2\log d + \sqrt{2\log d} + 1\right)\right)\sqrt{2\log d}}{\eta^2\sqrt{t}}. \end{aligned}$$

In particular, setting $\eta = \sqrt{\frac{2d}{\sqrt{2\log d}}}$ *yields*

$$\mathbb{E}\widehat{L}_n - \boldsymbol{v}^\top \boldsymbol{L}_n \leq 4m\sqrt{dn}\sqrt[4]{\log d} + m(\log n + 1)\sqrt{\log d}.$$

and

$$\mathbb{E}\sum_{t=1}^{n}\mathbb{I}\left\{\boldsymbol{V}_{t+1}\neq\boldsymbol{V}_t\right\}=O\left(m^2(\log d)^{5/2}\sqrt{n}\right).$$

The proof of the regret bound is quite standard, similar to Audibert et al. (2011), and it is omitted. The more interesting part is the bound for the expected number of action switches $\mathbb{E}\sum_{t=1}^{n}\mathbb{I}\left\{\boldsymbol{V}_{t+1}\neq\boldsymbol{V}_t\right\}=\sum_{t=1}^{n}\mathbb{P}\left[\boldsymbol{V}_{t+1}\neq\boldsymbol{V}_t\right]$. It follows from the lemma below and the well-known fact that the expected value of the maximum of the square of d independent standard normal random variables is at most $2\log d+\sqrt{2\log d}+1$ (see, e.g., Boucheron et al. (2013)). Thus, it suffices to prove the following:

Lemma 5 *For each $t=1,2,\ldots,n$,*

$$\mathbb{P}\left[\boldsymbol{V}_{t+1}\neq\boldsymbol{V}_t\,|\boldsymbol{X}_{t+1}\right]\leq\frac{m^2\left\|\boldsymbol{h}_t\right\|_\infty^2}{4\eta^2 t}+\frac{m^2\left\|\boldsymbol{h}_t\right\|_\infty\mathbb{E}\left[\left\|\boldsymbol{Z}_t\right\|_\infty\right]}{\eta^2 t}.$$

Proof We use the notation $\mathbb{P}_t\left[\cdot\right]=\mathbb{P}\left[\cdot\,|\boldsymbol{X}_{t+1}\right]$ and $\mathbb{E}_t\left[\cdot\right]=\mathbb{E}\left[\cdot\,|\boldsymbol{X}_{t+1}\right]$. Also, let

$$\boldsymbol{h}_t=\boldsymbol{\ell}_t+\boldsymbol{X}_{t+1}\qquad\text{and}\qquad\boldsymbol{H}_t=\sum_{s=0}^{t-1}\boldsymbol{h}_t.$$

Define the set $A_t(c)$ as the "lead pack of width c":

$$A_t(c)=\left\{\boldsymbol{w}\in\mathcal{S}:(\boldsymbol{w}-\boldsymbol{V}_t)^\top\boldsymbol{H}_t\leq c\right\}.$$

where c is a positive number that we choose later. (It is allowed to depend on $\boldsymbol{X}_{t+1}$.) Observe that $c\geq\max_{\boldsymbol{w}\in\mathcal{S}}\left|(\boldsymbol{w}-\boldsymbol{V}_t)^\top\boldsymbol{h}_t\right|$ guarantees that no action outside $A_t(c)$ can take the lead at time $t+1$, since if $\boldsymbol{w}\notin A_t$, then

$$(\boldsymbol{w}-\boldsymbol{V}_t)^\top\boldsymbol{H}_t\geq\max_{\boldsymbol{w}\in\mathcal{S}}\left|(\boldsymbol{w}-\boldsymbol{V}_t)^\top\boldsymbol{h}_t\right|$$

so $(\boldsymbol{w}-\boldsymbol{V}_t)^\top\boldsymbol{H}_{t+1}\geq 0$ and $\boldsymbol{w}$ cannot be the new leader. For $\boldsymbol{w}\in A_t$, we use the trivial bound $\mathbb{P}_t\left[\boldsymbol{V}_{t+1}=\boldsymbol{w}\right]\leq 1$, thus we have the bound

$$\mathbb{P}_t\left[\boldsymbol{V}_{t+1}\neq\boldsymbol{V}_t\right]\leq\mathbb{P}_t\left[|A_t(c)|>1\right],$$

which leaves us with the problem of bounding $\mathbb{P}_t\left[|A_t(c)|>1\right]$. Similarly to the proof of Lemma 3, we start analyzing $\mathbb{P}_t\left[|A_t(c)|=1\right]$:

$$\begin{aligned}\mathbb{P}_t\left[|A_t(c)|=1\right]&=\sum_{\boldsymbol{v}\in\mathcal{S}}\mathbb{P}_t\left[\forall\boldsymbol{w}\neq\boldsymbol{v}:(\boldsymbol{w}-\boldsymbol{v})^\top\boldsymbol{H}_t\geq c\right]\\&=\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}}f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall\boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\top\boldsymbol{H}_t\geq y+c\,\middle|\,\boldsymbol{v}^\top\boldsymbol{H}_t=y\right]dy,\end{aligned}\tag{2}$$

where $f_{\boldsymbol{v}}$ is the distribution of $\boldsymbol{v}^\top \boldsymbol{H}_t$. Next we crucially use the fact that the conditional distributions of correlated Gaussian random variables are also Gaussian. In particular, defining $k(\boldsymbol{w}, \boldsymbol{v}) = (m - \|\boldsymbol{w} - \boldsymbol{v}\|_1)$, the covariances are given as

$$cov\left(\boldsymbol{w}^\top \boldsymbol{H}_t, \boldsymbol{v}^\top \boldsymbol{H}_t\right) = \eta^2(m - \|\boldsymbol{w} - \boldsymbol{v}\|_1)t = \eta^2 k(\boldsymbol{w}, \boldsymbol{v})t.$$

Let us organize all actions $\boldsymbol{w} \in \mathcal{S}\backslash v$ into a vector $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{N-1})$. The conditional distribution of $\boldsymbol{W}^\top \boldsymbol{H}_t$ is an $(N-1)$-variate Gaussian distribution with mean

$$\mu_{\boldsymbol{v}}(y) = \left(w_1^\top \boldsymbol{L}_{t-1} + y\frac{k(\boldsymbol{w}, \boldsymbol{v})}{m}, \ldots, w_{N-1}^\top \boldsymbol{L}_{t-1} + y\frac{k(\boldsymbol{w}, \boldsymbol{v})}{m}\right)^\top$$

and covariance matrix $\Sigma_{\boldsymbol{v}}$, given that $\boldsymbol{v}^\top \boldsymbol{H}_t = y$. Defining $\boldsymbol{K} = (k(\boldsymbol{w}_1, \boldsymbol{v}), \ldots, k(\boldsymbol{w}_{N-1}, \boldsymbol{v}))^\top$ and using the notation $\varphi(x) = \frac{1}{\sqrt{(2\pi)^{N-1}|\Sigma_v|}} \exp(-\frac{x^2}{2})$, we get that

$$\begin{aligned}
&\mathbb{P}_t\left[\forall \boldsymbol{w} \neq \boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c \,\middle|\, \boldsymbol{v}^\top \boldsymbol{H}_t = y\right] \\
&\quad = \int_{z_i = y+c}^{\infty} \cdots \int \phi\left(\sqrt{(\boldsymbol{z} - \mu_{\boldsymbol{v}}(y))^\top \Sigma_y^{-1} (\boldsymbol{z} - \mu_{\boldsymbol{v}}(y))}\right) d\boldsymbol{z} \\
&\quad = \int_{z_i = y + c\left(1 + \frac{k(\boldsymbol{w}_i, \boldsymbol{v})}{m}\right)}^{\infty} \cdots \int \phi\left(\sqrt{\left(\boldsymbol{z} - \mu_{\boldsymbol{v}}(y) - \frac{c}{m}\boldsymbol{K}\right)^\top \Sigma_y^{-1} \left(\boldsymbol{z} - \mu_{\boldsymbol{v}}(y) - \frac{c}{m}\boldsymbol{K}\right)}\right) d\boldsymbol{z} \\
&\quad = \int_{z_i = y + c\left(1 + \frac{k(\boldsymbol{w}_i, \boldsymbol{v})}{m}\right)}^{\infty} \cdots \int \phi\left(\sqrt{(\boldsymbol{z} - \mu_{\boldsymbol{v}}(y+c))^\top \Sigma_y^{-1} (\boldsymbol{z} - \mu_{\boldsymbol{v}}(y+c))}\right) d\boldsymbol{z} \\
&\quad = \mathbb{P}_t\left[\forall \boldsymbol{w} \neq \boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c\left(1 + \frac{k}{m}\right) \,\middle|\, \boldsymbol{v}^\top \boldsymbol{H}_t = y + c\right],
\end{aligned}$$

where we used $\mu_{y+c} = \mu_y + \frac{c}{m}\boldsymbol{K}$. Using this, we rewrite (2) as

$$
\begin{aligned}
\mathbb{P}_t\left[|A_t(c)| = 1\right] =& \sum_{\boldsymbol{v}\in\mathcal{S}} \int_{y\in\mathbb{R}} f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\, \boldsymbol{v}^\top\boldsymbol{H}_t = y\right] \\
&- \sum_{\boldsymbol{v}\in\mathcal{S}} \int_{y\in\mathbb{R}} \left(f_{\boldsymbol{v}}(y) - f_{\boldsymbol{v}}(y-c)\right)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\, \boldsymbol{v}^\top\boldsymbol{H}_t = y\right] \\
\geq& \sum_{\boldsymbol{v}\in\mathcal{S}} \int_{y\in\mathbb{R}} f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c\frac{m-1}{m}\,\middle|\, \boldsymbol{v}^\top\boldsymbol{H}_t = y\right] \\
&- \sum_{\boldsymbol{v}\in\mathcal{S}} \int_{y\in\mathbb{R}} \left(f_{\boldsymbol{v}}(y) - f_{\boldsymbol{v}}(y-c)\right)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\, \boldsymbol{v}^\top\boldsymbol{H}_t = y\right] \\
=& \mathbb{P}_t\left[\left|A_t\left(c\frac{m-1}{m}\right)\right| = 1\right] \\
&- \sum_{\boldsymbol{v}\in\mathcal{S}} \int_{y\in\mathbb{R}} \left(f_{\boldsymbol{v}}(y) - f_{\boldsymbol{v}}(y-c)\right)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v} : \boldsymbol{w}^\top \boldsymbol{H}_t \geq y + c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\, \boldsymbol{v}^\top\boldsymbol{H}_t = y\right],
\end{aligned}
$$

where we used $k(\boldsymbol{w},\boldsymbol{v}) \leq m-1$ in the inequality. After reorganizing, we obtain an upper bound for

$$
\mathbb{P}_t\left[|A_t(c)| > 1\right] - \mathbb{P}_t\left[\left|A_t\left(c\frac{m-1}{m}\right)\right| > 1\right],
$$

which is the probability that there are actions in the "outer ring" of the lead pack. Introducing the notation

$$
B_t(a,b) = \left\{\boldsymbol{w}\in\mathcal{S} : (\boldsymbol{w} - \boldsymbol{V}_t)^\top \boldsymbol{H}_t \in [a,b]\right\},
$$

we may write

$$
\mathbb{P}_t\left[|A_t(c)| > 1\right] - \mathbb{P}_t\left[\left|A_t\left(c\frac{m-1}{m}\right)\right| > 1\right] = \mathbb{P}_t\left[\left|B_t\left(c\frac{m-1}{m}, c\right)\right| > 0\right].
$$

To treat the remaining term, we use that $\boldsymbol{v}^\top\boldsymbol{H}_t$ is Gaussian with mean $\boldsymbol{v}^\top\boldsymbol{L}_{t-1}$ and standard deviation $\eta\sqrt{mt}$ and obtain

$$
\begin{aligned}
f_{\boldsymbol{v}}(y) - f_{\boldsymbol{v}}(y-c) =& f_{\boldsymbol{v}}(y)\left(1 - \frac{f_{\boldsymbol{v}}(y-c)}{f_{\boldsymbol{v}}(y)}\right) \\
\leq& f_{\boldsymbol{v}}(y)\left(\frac{c^2}{2\eta^2 mt} - \frac{c(y - \boldsymbol{v}^\top\boldsymbol{L}_{t-1})}{\eta^2 mt}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}} \left(f_{\boldsymbol{v}}(y)-f_{\boldsymbol{v}}(y-c)\right)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\top\boldsymbol{H}_t\geq y+c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\,\boldsymbol{v}^\top\boldsymbol{H}_t=y\right]\\
&\quad\leq\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}} f_{\boldsymbol{v}}(y)\left(\frac{c^2}{2\eta^2mt}-\frac{c(y-\boldsymbol{v}^\top\boldsymbol{L}_{t-1})}{\eta^2mt}\right)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\top\boldsymbol{H}_t\geq y+c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\,\boldsymbol{v}^\top\boldsymbol{H}_t=y\right]\\
&\quad\leq\sum_{\boldsymbol{v}\in\mathcal{S}}\int_y\left(\frac{c^2}{2\eta^2mt}+\left|\frac{c(y-\boldsymbol{v}^\top\boldsymbol{L}_{t-1})}{\eta^2mt}\right|\right)f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\top\boldsymbol{H}_t\geq y+c\frac{k(\boldsymbol{w},\boldsymbol{v})}{m}\,\middle|\,\boldsymbol{v}^\top\boldsymbol{H}_t=y\right]\\
&\quad\leq\sum_{\boldsymbol{v}\in\mathcal{S}}\int_y\left(\frac{c^2}{2\eta^2mt}+\left|\frac{c(y-\boldsymbol{v}^\top\boldsymbol{L}_{t-1})}{\eta^2mt}\right|\right)f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\top\boldsymbol{H}_t\geq y\,\middle|\,\boldsymbol{v}^\top\boldsymbol{H}_t=y\right]\\
&\quad=\frac{c^2}{2\eta^2mt}+\frac{c\mathbb{E}\left[|\boldsymbol{V}_t^\top\boldsymbol{Z}_t|\right]}{\eta^2mt}\leq\frac{c^2}{2\eta^2mt}+\frac{cm\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]}{\eta^2mt},
\end{aligned}
$$

where we used $k(\boldsymbol{w},\boldsymbol{v})\geq 0$ in the third inequality. Eventually, we obtain

$$
\mathbb{P}_t\left[\left|B_t\left(c\frac{m-1}{m},c\right)\right|>0\right]\leq\frac{c^2}{2\eta^2mt}+\frac{c\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]}{\eta^2t}. \tag{3}
$$

Now observe that the lead pack can be decomposed into disjoint layers as

$$
A_t(c)\setminus\boldsymbol{V}_t=\bigcup_{s=0}^{\infty}B_t\left(c\left(\frac{m-1}{m}\right)^{s+1},c\left(\frac{m-1}{m}\right)^{s}\right)
$$

Using the union bound, we obtain

$$
\begin{aligned}
\mathbb{P}_t\left[|A_t(c)|>1\right]&\leq\sum_{s=0}^{\infty}\mathbb{P}_t\left[\left|B_t\left(c\left(\frac{m-1}{m}\right)^{s+1},c\left(\frac{m-1}{m}\right)^{s}\right)\right|>0\right]\\
&\leq\sum_{s=0}^{\infty}\left(\left(\frac{m-1}{m}\right)^{2s}\frac{c^2}{2\eta^2mt}+\left(\frac{m-1}{m}\right)^{s}\frac{c\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]}{\eta^2t}\right)\\
&\leq\frac{m^2}{2m-1}\cdot\frac{c^2}{2\eta^2mt}+\frac{cm\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]}{\eta^2t}
\end{aligned}
$$

Using $c=m\|\boldsymbol{h}_t\|_\infty$ proves the theorem:

$$
\begin{aligned}
\mathbb{P}_t\left[|A_t(c)|>1\right]&\leq\frac{m}{2}\cdot\frac{m^2\|\boldsymbol{h}_t\|_\infty^2}{2\eta^2mt}+\frac{m^2\|\boldsymbol{h}_t\|_\infty\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]}{\eta^2t}\\
&\leq\frac{m^2\|\boldsymbol{h}_t\|_\infty^2}{4\eta^2t}+\frac{m^2\|\boldsymbol{h}_t\|_\infty\sqrt{2\log d}}{\eta^2\sqrt{t}}.
\end{aligned} \tag{4}
$$

■

References

J. Y. Audibert, S. Bubeck, and G. Lugosi. Minimax policies for combinatorial prediction games. In *Conference on Learning Theory*, 2011. URL `http://arxiv.org/abs/1105.4871`.

S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities:A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

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

Eyal Even-Dar, Sham. M. Kakade, and Yishay Mansour. Online Markov decision processes. *Mathematics of Operations Research*, 34(3):726–736, 2009. ISSN 0364-765X. doi: http://dx.doi.org/10.1287/moor.1090.0396.

W. Feller. *An Introduction to Probability Theory and its Applications, Vol. 1*. John Wiley, New York, 1968.

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

C. Gentile and M. Warmuth. Linear hinge loss and average margin. In *Advances in Neural Information Processing Systems (NIPS)*, 1998.

S. Geulen, B. Voecking, and M. Winkler. Regret minimization for online buffering problems using the weighted majority algorithm. In *Proceedings of the Twenty-Third Conference on Computational Learning Theory*, 2010.

A. Grove, N. Littlestone, and D. Schuurmans. General convergence results for linear discriminant updates. *Machine Learning*, 43:173–210, 2001.

András György and Gergely Neu. Near-optimal rates for limited-delay universal lossy source coding. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, 2011.

J. Hannan. Approximation to Bayes risk in repeated play. *Contributions to the theory of games*, 3:97–139, 1957.

E. Hazan, S. Kale, and M. Warmuth. Learning rotations with little regret. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, 2010.

D. P. Helmbold and M. Warmuth. Learning permutations with exponential weights. *Journal of Machine Learning Research*, 10:1705–1736, 2009.

M. Hutter and J. Poland. Prediction with expert advice by following the perturbed leader for general weights. In *Algorithmic Learning Theory*, pages 279–293. Springer, 2004.

A. Kalai and S Vempala. Efficient algorithms for the online decision problem. In B. Schölkopf and M. Warmuth, editors, *Proceedings of the 16th Annual Conference on Learning Theory and the 7th Kernel Workshop, COLT-Kernel 2003*, pages 26–40, New York, USA, Aug. 2003. Springer.

A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71:291–307, 2005a.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *J. Comput. Syst. Sci.*, 71:291–307, October 2005b. ISSN 0022-0000. doi: http://dx.doi.org/10.1016/j.jcss.2004.10.016. URL `http://dx.doi.org/10.1016/j.jcss.2004.10.016`.

J. Kivinen and M. Warmuth. Relative loss bounds for multidimensional regression problems. *Machine Learning*, 45:301–329, 2001.

W. Koolen, M. Warmuth, and J. Kivinen. Hedging structured concepts. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, pages 93–105, 2010.

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

G. Neu, A. György, Cs Szepesvári, and A. Antos. Online Markov decision processes under bandit feedback. In *Advances in Neural Information Processing Systems 23*, 2010.

Jan Poland. FPL analysis for adaptive bandits. In *In 3rd Symposium on Stochastic Algorithms, Foundations and Applications (SAGA'05)*, pages 58–69, 2005.

E. Takimoto and M. Warmuth. Paths kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003.

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

M. Warmuth and D. Kuzmin. Randomized online pca algorithms with regret bounds that are logarithmic in the dimension. *Journal of Machine Learning Research*, 9:2287–2320, 2008.