Random-Walk Perturbations for Online Combinatorial Optimization

Luc Devroye, Gábor Lugosi and Gergely Neu

***Abstract*—We study online combinatorial optimization problems where a learner is interested in minimizing its cumulative regret in the presence of switching costs. To solve such problems, we propose a version of the follow-the-perturbed-leader algorithm in which the cumulative losses are perturbed by independent symmetric random walks. In the general setting, our forecaster is shown to enjoy near-optimal guarantees on both quantities of interest, making it the best known efficient algorithm for the studied problem. In the special case of prediction with expert advice, we show that the forecaster achieves 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, while guaranteeing that the predictions are switched at most $O(\sqrt{n \log N})$ times, in expectation.**

***Index Terms*—Online learning, Online combinatorial optimization, Follow the Perturbed Leader, Random walk**

I. Preliminaries

In this paper we study the problem of online prediction with expert advice (see [2]), and in particular, online linear optimization (see, e.g., [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]). The problem may be described as a repeated game between a *forecaster* and an adversary—the *environment*. At each time instant $t = 1, \dots, n$, the forecaster chooses one of the N available actions and suffers a loss corresponding to the chosen action i. Each action i is represented by a vector $\boldsymbol{v}_i \in \mathbb{R}^d$, while the losses assigned by the environment at time t are described by the *loss vector* $\boldsymbol{\ell}_t \in [0,1]^d$. Thus, given the set of actions $\mathcal{S} = \{\boldsymbol{v}_i : i = 1, 2, \dots, N\} \subseteq \mathbb{R}^d$ at every time instant t, the forecaster chooses, in a possibly randomized way, a vector $\boldsymbol{V}_t \in \mathcal{S}$ and suffers loss $\boldsymbol{V}_t^\mathsf{T} \boldsymbol{\ell}_t$.

We consider the so-called *oblivious adversary* model in which the environment selects all losses before the prediction game starts and reveals the loss vector $\boldsymbol{\ell}_t$ 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 $\boldsymbol{V}_t$ 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 $\widehat{L}_n = \sum_{t=1}^n \boldsymbol{V}_t^\mathsf{T} \boldsymbol{\ell}_t$ 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 $\boldsymbol{\ell}_t$, we cannot hope to minimize the above 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

$$R_n = \max_{i \in \{1,2,\dots,N\}} \sum_{t=1}^n (\boldsymbol{V}_t - \boldsymbol{v})^\mathsf{T} \boldsymbol{\ell}_t = \widehat{L}_n - L_n^*,$$

where we have also introduced the notation $L_n^* = \min_{\boldsymbol{v} \in \mathcal{S}} \boldsymbol{v}^\mathsf{T} \sum_{t=1}^n \boldsymbol{\ell}_t$.

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 as a binary vector. This special case arguably contains most important applications such as 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 its 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 $\boldsymbol{v}^\mathsf{T} \boldsymbol{\ell}_t$ 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 presented in the paper.

A fundamental special case of the framework above is *prediction with expert advice*. In this setting, we have $m = 1$, $d = N$, and the learner has access to the unit vectors $\mathcal{S} = \{\boldsymbol{e}_i\}_{i=1}^N$ as the decision set. Minimizing the regret in this setting is a well-studied problem (see the book of Cesa-Bianchi

Parameters: set of actions $\mathcal{S} \subseteq \mathbb{R}^d$, number of rounds n;
The environment chooses the loss vector $\boldsymbol{\ell}_t \in [0,1]^d$ for all $t = 1, \dots, n$.
For all $t = 1, 2, \dots, n$**, repeat**
1) The forecaster chooses a probability distribution $\boldsymbol{p}_t$ over $\mathcal{S}$.
2) The forecaster draws an action $\boldsymbol{V}_t$ randomly according to $\boldsymbol{p}_t$.
3) The environment reveals $\boldsymbol{\ell}_t$.
4) The forecaster suffers loss $\boldsymbol{V}_t^\mathsf{T} \boldsymbol{\ell}_t$.

Fig. 1. Online linear optimization.

L. Devroye is with the School of Computer Science, McGill University, Montreal, Canada H3A 2K6 (email: lucdevroye@gmail.com). G. Lugosi is with ICREA and the Department of Economics, Pompeu Fabra University, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain (email: gabor.lugosi@gmail.com). G. Neu is with the SequeL team, INRIA Lille – Nord Europe, 40 avenue Halley, 59650 Villeneuve d'Ascq, France (email: gergely.neu@gmail.com).

A previous version of this paper was published in the Proceedings of the 26th Annual Conference on Learning Theory, 2013 as [1].

and Lugosi [2]). 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 $[0,1]$. 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 [14], and aggregating strategies of Vovk [15], also known as *Hedge* by Freund and Schapire [16]). 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 [17]. Kalai and Vempala [7] 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 vector $\boldsymbol{Z}_t \in \mathbb{R}^d$ to the cumulative loss $\boldsymbol{L}_{t-1} = \sum_{s=1}^{t-1} \boldsymbol{\ell}_s$ of each action and chooses an action $\boldsymbol{v}$ that minimizes $\boldsymbol{v}^\intercal(\boldsymbol{L}_{t-1} + \boldsymbol{Z}_t)$. If the elements of the perturbation vector 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 $O(\sqrt{n\log N})$ ([7], see also [2], [18], [19]). 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, it may fail with reasonably high probability since its regret is much less concentrated.

While the results presented above still hold in the general case where $m > 1$ when treating each $\boldsymbol{v} \in \mathcal{S}$ as a separate action, one may gain important computational advantage by taking the structure of the action set into account. In particular, as [7] emphasize, FPL-type forecasters may often be computed efficiently. Interestingly, this efficiency does not come at the price of inferior regret guarantees: as Neu and Bartók [20] have recently shown, an appropriately tuned version of FPL achieves the same regret of $O(m^{3/2}\sqrt{d\log d})$ as the straightforward extension of the exponentially weighted forecaster. The only known forecaster to achieve better performance than this is Component Hedge proposed by Koolen, Warmuth and Kivinen [11], guaranteeing a minimax optimal regret of $O(m\sqrt{n\log(d/m)})$. However, this forecaster can only be implemented efficiently for some special decision sets, and can still take $\Omega(d^6)$ time to run in the worst case (see [21]). In this paper, we propose an FPL-variant that retains the near-optimal regret guarantees of $O(m^{3/2}\sqrt{d\log d})$, while having nice additional properties discussed below.

Small regret is not the only desirable feature of an online forecasting algorithm. In many applications, one would like to define forecasters that do not change their prediction too often. For instance, consider a sequential routing problem on a computer network where predictions correspond to selecting a path in a graph for each packet to traverse. In this situation, switching between routes might result in out-of-order delivery of packets due to changing delays, and eventually lead to decoding errors. Further examples of such problems include the online buffering problem described by Geulen, Voecking and Winkler [22] and the online lossy source coding problem of György and Neu [23]. 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, Kakade and Mansour [24] and Neu, György, Szepesvári, and Antos [25].

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

$$C_n = |\{1 < t \leq n : \boldsymbol{V}_{t-1} \neq \boldsymbol{V}_t\}| \ .$$

In particular, we are interested in defining randomized forecasters that achieve a regret R_n of near-optimal order 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 of times, typically $\Theta(n)$. Therefore, the design of special forecasters with small regret and small number of action switches is called for.

The first known algorithm to address this issue is the Follow the Lazy Leader (FLL) algorithm proposed by Kalai and Vempala [7]. This algorithm is designed to behave identically to their FPL algorithm *in expectation* (with an $O(m^{3/2}\sqrt{n\log d})$ bound on the regret, as shown by Neu and Bartók [20]), while guaranteeing that the expected number of action switches is $O(d\sqrt{(n/m)\log d})$. The "Shrinking Dartboard" algorithm proposed by Geulen, Voecking and Winkler [22] is based on a similar idea: this algorithm simulates the exponentially weighted forecaster in expectation, guaranteeing a regret of $O(m^{3/2}\sqrt{n\log d})$, while improving the upper bound on the expected number of switches to $O(\sqrt{mn\log d})$. In this paper, we propose a family of methods based on FPL in which perturbations are defined by independent symmetric random walks. We show that these intuitively appealing forecasters have similar regret and switch-number guarantees as Shrinking Dartboard and FLL.

In particular, we first propose an FPL-variant in which perturbations are generated by independent Gaussian random walks for each coordinate of the perturbation vector. We show that this algorithm guarantees a regret of $O(m^{3/2}\sqrt{n\log d})$, while keeping the number of switches bounded by $O(m\sqrt{n}\log d)$. While this bound is inferior to that of the Shrinking Dartboard algorithm by a factor of $\sqrt{m\log d}$, that algorithm can only be efficiently implemented for some special decision sets $\mathcal{S}$—see [11] and [12] for some examples. On the other hand, our algorithm can be efficiently implemented whenever there exists an efficient implementation of the static optimization problem of finding $\arg\min_{\boldsymbol{v}\in\mathcal{S}} \boldsymbol{v}^\intercal\boldsymbol{\ell}$ for any $\boldsymbol{\ell} \in \mathbb{R}^d$. We compare our results to other results known in the literature in Table I.

Notice that our regret bound described above only guarantees that the number of switches is of $O(\sqrt{n}\log N)$ in the setting of prediction with expert advice. In the second half of

	Regret	Switches	Efficient
Shrinking Dartboard [22]	$O(m^{3/2}\sqrt{n \log d})$	$O(\sqrt{mn \log d})$	sometimes
Follow the Lazy Leader [7]	$O(m^{3/2}\sqrt{n \log d})$	$O(d\sqrt{(n/m) \log d})$	always
Prediction by random-walk perturbations	$O(m^{3/2}\sqrt{n \log d})$	$O(m\sqrt{n} \log d)$	always

TABLE I
THE RESULTS PRESENTED IN THIS PAPER VERSUS THE RESULTS OF [7] AND [22].

the paper, we show that this can be improved to $O(\sqrt{n \log N})$ by using symmetric *binary* random walks as perturbations. An interesting property of the resulting algorithm is that the expected regret can be directly upper bounded in terms of the expected number of switches.

We also note that a similar variant of the FPL forecaster was recently derived by Rakhlin, Shamir and Sridharan [26], who use perturbations of the form $Z_{i,t} = \sum_{s=t+1}^{n} X_{i,s}$, where $X_{i,s}$ are i.i.d. random variables with an arbitrary symmetric distribution. Rakhlin, Shamir and Sridharan exploit the fact that these perturbations can serve as a relaxation of the Rademacher complexity of the prediction game, and prove an $O(\sqrt{n \log N})$ bound on the expected regret of the resulting algorithm. While this approach cannot be used for analyzing our FPL-variant, our analysis can be directly applied to provide both regret and switch-number guarantees for their method. Additionally, note that our algorithm does not need to use prior knowledge of the number of rounds.

An interesting open question left for future research is characterizing the minimax rate for the quantity $R_n + C_n$ in terms of m. For all known algorithms (including ours), the upper bounds on this quantity depend on m as $m^{3/2}$, as contributed by the bound on R_n. However, this dependence is known to be suboptimal: the optimal linear dependence of the regret on m is achieved by the Component Hedge algorithm of [11]. We conjecture that the correct minimax rate of $R_n + C_n$ matches the minimax rate of R_n, which is of $\Theta(m\sqrt{n \log(d/m)})$. Such a bound could be achieved by a switch-constrained algorithm that implements Component Hedge on expectation in a similar way as the Shrinking Dartboard implements the exponentially weighted forecaster or FLL implements FPL. However, constructing such an implementation is far from trivial and is left here as an interesting open problem.

II. RANDOM-WALK PERTURBATIONS FOR ONLINE COMBINATORIAL OPTIMIZATION

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 zero-mean Gaussian random variable with variance $\eta^2 > 0$ and chooses an action with minimal cumulative perturbed loss. More precisely, the algorithm draws independent random variables $X_{i,t} \sim \mathcal{N}(0, \eta^2)$ and the vector $\boldsymbol{X}_t = (X_{1,t}, \ldots, X_{d,t})$ is added to the observed loss vector $\boldsymbol{\ell}_{t-1}$. At time t action $\boldsymbol{v} \in \mathcal{S}$ is chosen that minimizes $\sum_{s=1}^{t} \boldsymbol{v}^{\mathsf{T}} (\boldsymbol{\ell}_{t-1} + \boldsymbol{X}_t)$ (where we define $\boldsymbol{\ell}_0$ as the all-zero vector $\mathbf{0}$). Equivalently, the forecaster may be thought of as an FPL algorithm in which the cumulative losses $\boldsymbol{L}_{t-1}$ are perturbed by the independent symmetric random walks $\boldsymbol{Z}_t = \sum_{s=1}^{t} \boldsymbol{X}_s$. This is the way the algorithm is presented in Algorithm 1.

Algorithm 1 Online combinatorial optimization by random-walk perturbations.

Initialization: set $\boldsymbol{L}_0 = \mathbf{0}$ and $Z_0 = \mathbf{0}$.
For all $t = 1, 2, \ldots, n$**, repeat**

1) Draw $\boldsymbol{X}_t$ with i.i.d. Gaussian components
$$X_{i,t} \sim \mathcal{N}(0, \eta^2)$$
2) Let $\boldsymbol{Z}_t = \boldsymbol{Z}_{t-1} + \boldsymbol{X}_t$.
3) Choose action
$$\boldsymbol{V}_t = \arg\min_{\boldsymbol{v} \in \mathcal{S}} \left\{ \boldsymbol{v}^{\mathsf{T}} \left(\boldsymbol{L}_{t-1} + \boldsymbol{Z}_t \right) \right\},$$
where ties are broken in favor of $\boldsymbol{V}_{t-1}$.
4) Observe the loss vector $\boldsymbol{\ell}_t$, suffer loss $\boldsymbol{V}_t^{\mathsf{T}} \boldsymbol{\ell}_t$.
5) Set $\boldsymbol{L}_t = \boldsymbol{L}_{t-1} + \boldsymbol{\ell}_t$.

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 at time t 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 action set $\mathcal{S} = \{(1,0)^{\mathsf{T}}, (0,1)^{\mathsf{T}}\}$ and an environment that generates equal losses, say $\ell_{i,t} = 0$ for all i and t. 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. This distribution is well understood and the probability that this occurs more than $x\sqrt{n}$ times during the first n steps is roughly $2\mathbb{P}\{N > 2x\} \leq 2e^{-2x^2}$ where N is a standard normal random variable (see [27, Section III.4]). Thus, in this case we see that the number of switches is bounded by $O\left(\sqrt{n \log(1/\delta)}\right)$, with probability at least $1 - \delta$. As we show below, assuming all-zero losses is the worst case for the number of switches.

Even though we only prove bounds for the expected regret and the expected number of switches, the above example gives

some intuition about the upper tail probabilities. While the above idea can be extended to the case of non-zero loss sequences to obtain high-confidence switch-number guarantees, proving similar results for the general setting is a highly nontrivial problem. We note that by our Lemma 2 (stated later in Section III), the regret of our algorithm can be directly bounded in terms of the number of switches, thus we can guarantee upper bounds of $O(\sqrt{n})$ on *both* C_n and R_n with high probability. We are not aware of any other algorithm that provides high-confidence guarantees on both quantities of interest even in this simple special case.

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

Theorem 1: The expected regret and the expected number of action switches satisfy (under the oblivious adversary model),

$$\mathbb{E}R_n \leq m\sqrt{2n\log d}\left(\frac{2m}{\eta}+\eta\right)+\frac{m^2(\log n+1)}{\eta^2}$$

and

$$\mathbb{E}C_n \leq \sum_{t=1}^{n}\frac{m\left(1+\eta\sqrt{2\log d}\right)\sqrt{2\log d}}{\eta\sqrt{t}} + \sum_{t=1}^{n}\frac{m\left(1+2\eta\sqrt{2\log d}+\eta^2\left(2\log d+\sqrt{2\log d}+1\right)\right)}{4\eta^2 t}.$$

In particular, setting $\eta = \sqrt{2m}$ yields

$$\mathbb{E}R_n \leq 4m^{3/2}\sqrt{n\log d}+\frac{m(\log n+1)}{2}.$$

and

$$\mathbb{E}C_n = O\left(m\sqrt{n}\log d\right).$$

The proof of the regret bound follows the steps of the proof of Theorem 1 in [20], and is deferred to the appendix. The more interesting part is the bound for the expected number of action switches $\mathbb{E}C_n = \sum_{t=1}^{n}\mathbb{P}\left[\boldsymbol{V}_{t+1} \neq \boldsymbol{V}_t\right]$. For proving a bound on this quantity, we study the evolution of the *lead pack* A_t defined as the set of actions with near-optimal loss given by

$$A_t = \big\{\boldsymbol{w}\in\mathcal{S} : (\boldsymbol{w}-\boldsymbol{V}_t)^{\mathsf{T}}(\boldsymbol{L}_{t-1}+\boldsymbol{Z}_t) \leq \|\boldsymbol{w}-\boldsymbol{V}_t\|_1 \cdot \|\boldsymbol{\ell}_t+\boldsymbol{X}_{t+1}\|_\infty\big\}\ . \tag{1}$$

We sometimes refer to $\|\boldsymbol{\ell}_t+\boldsymbol{X}_{t+1}\|_\infty$ as the *diameter* of the lead pack. Observe that no action outside A_t can take the lead at time $t+1$, since if $\boldsymbol{w}\notin A_t$, then

$$(\boldsymbol{w}-\boldsymbol{V}_t)^{\mathsf{T}}(\boldsymbol{L}_{t-1}+\boldsymbol{Z}_t) > |(\boldsymbol{w}-\boldsymbol{V}_t)^{\mathsf{T}}(\boldsymbol{\ell}_t+\boldsymbol{X}_{t+1})|$$

so $\boldsymbol{w}^{\mathsf{T}}(\boldsymbol{L}_t+\boldsymbol{Z}_{t+1}) > \boldsymbol{V}_t^{\mathsf{T}}(\boldsymbol{L}_t+\boldsymbol{Z}_{t+1})$ and $\boldsymbol{w}$ cannot be the new leader. It follows that we can upper bound the probability of switching as

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

which leaves us with the problem of upper bounding $\mathbb{P}\left[|A_t|>1\right]$. The following lemma gives a bound of this quantity. Putting this statement together with the well-known facts that $\mathbb{E}\left[\|\boldsymbol{X}_1\|_\infty\right] \leq \eta\sqrt{2\log d}$ and $\mathbb{E}\left[\|\boldsymbol{X}_1\|_\infty^2\right] \leq \eta^2\left(2\log d+\sqrt{2\log d}+1\right)$ (see, e.g., [28]) proves the second statement of the theorem.

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

$$\mathbb{P}\left[|A_t|>1\,|\boldsymbol{X}_{t+1}\right] \leq \frac{m\,\|\boldsymbol{\ell}_t+\boldsymbol{X}_{t+1}\|_\infty\sqrt{2\log d}}{\eta\sqrt{t}} + \frac{m\,\|\boldsymbol{\ell}_t+\boldsymbol{X}_{t+1}\|_\infty^2}{2\eta^2 t}.$$

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

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

Furthermore, we use the shorthand notation $c = \|\boldsymbol{h}_t\|_\infty$.

We start by analyzing $\mathbb{P}_t\left[|A_t|=1\right]$:

$$\begin{aligned}
&\mathbb{P}_t\left[|A_t|=1\right] = \\
&= \sum_{\boldsymbol{v}\in\mathcal{S}}\mathbb{P}_t\left[\forall\boldsymbol{w}\neq\boldsymbol{v} : (\boldsymbol{w}-\boldsymbol{v})^{\mathsf{T}}\boldsymbol{H}_t > \|\boldsymbol{w}-\boldsymbol{v}\|_1\, c\right] \\
&= \sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}} f_{\boldsymbol{v}}(y)\mathbb{P}_t\big[\forall\boldsymbol{w}\neq\boldsymbol{v} : \\
&\qquad\qquad \boldsymbol{w}^{\mathsf{T}}\boldsymbol{H}_t > y+\|\boldsymbol{w}-\boldsymbol{v}\|_1\, c\big|\boldsymbol{v}^{\mathsf{T}}\boldsymbol{H}_t=y\big]\,dy,
\end{aligned} \tag{2}$$

where $f_{\boldsymbol{v}}$ is the density of $\boldsymbol{v}^{\mathsf{T}}\boldsymbol{H}_t$.

Next, we crucially use the fact that the conditional distributions of correlated Gaussian random variables are also Gaussian. In particular, let us fix an arbitrary $\boldsymbol{v}\in\mathcal{S}$ and define $k_{\boldsymbol{w}} = (m-\|\boldsymbol{w}-\boldsymbol{v}\|_1)$ for all $\boldsymbol{w}\in\mathcal{S}$. Then, the covariance between the perturbed loss of $\boldsymbol{w}$ and $\boldsymbol{v}$ is given as

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

Organizing all actions $\boldsymbol{w}\in\mathcal{S}\setminus\boldsymbol{v}$ into a matrix $\boldsymbol{W} = (\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_{N-1})$, the conditional distribution of $\boldsymbol{W}^{\mathsf{T}}\boldsymbol{H}_t$ is an $(N-1)$-variate Gaussian distribution with mean vector

$$\mu_{\boldsymbol{v}}(y) = \left(\boldsymbol{w}_i^{\mathsf{T}}\boldsymbol{L}_{t-1}+y\frac{k_{\boldsymbol{w}_i}}{m},\ldots,\boldsymbol{w}_{N-1}^{\mathsf{T}}\boldsymbol{L}_{t-1}+y\frac{k_{\boldsymbol{w}_{N-1}}}{m}\right)^{\mathsf{T}}$$

and covariance matrix $\Sigma_{\boldsymbol{v}}$, given that $\boldsymbol{v}^{\mathsf{T}}\boldsymbol{H}_t = y$. Defining $\boldsymbol{K} = \left(k_{\boldsymbol{w}_1},\ldots,k_{\boldsymbol{w}_{N-1}}\right)^{\mathsf{T}}$ and using the notation

$$\varphi_{\boldsymbol{v}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^{N-1}\left|\Sigma_v\right|}}\exp\left(-\frac{\boldsymbol{x}^{\mathsf{T}}\Sigma_{\boldsymbol{v}}^{-1}\boldsymbol{x}}{2}\right),$$

we get that

$$\begin{aligned}
&\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y+\|\boldsymbol{w}-\boldsymbol{v}\|_1\,c\,|\boldsymbol{v}^\intercal\boldsymbol{H}_t=y\right]\\
&\quad=\int_{z_i=y+(m-k_{\boldsymbol{w}_i})c}^{\infty}\cdots\int\varphi_{\boldsymbol{v}}\left(\boldsymbol{z}-\mu_{\boldsymbol{v}}(y)\right)\,d\boldsymbol{z}\\
&\quad=\int_{z_i=y+mc}^{\infty}\cdots\int\varphi_{\boldsymbol{v}}\left(\boldsymbol{z}-\mu_{\boldsymbol{v}}(y)-c\boldsymbol{K}\right)\,d\boldsymbol{z}\\
&\quad=\int_{z_i=y+mc}^{\infty}\cdots\int\varphi_{\boldsymbol{v}}\left(\boldsymbol{z}-\mu_{\boldsymbol{v}}(y+mc)\right)\,d\boldsymbol{z}\\
&\quad=\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y+mc|\,\boldsymbol{v}^\intercal\boldsymbol{H}_t=y+mc\right],
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{P}_t\left[|A_t|=1\right]\\
&=\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}}f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y|\,\boldsymbol{v}^\intercal\boldsymbol{H}_t=y\right]\,dy\\
&\quad-\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}}\left(f_{\boldsymbol{v}}(y)-f_{\boldsymbol{v}}(y-mc)\right)\cdot\\
&\qquad\qquad\cdot\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y|\,\boldsymbol{v}^\intercal\boldsymbol{H}_t=y\right]\,dy
\end{aligned}$$

Observing that the first term above corresponds to the probability that the leader is unique and that this event holds almost surely under our perturbation scheme, we get that the last term equals $\mathbb{P}_t\left[|A_t|>1\right]$. To treat this term, we use that $\boldsymbol{v}^\intercal\boldsymbol{H}_t$ is Gaussian with mean $\boldsymbol{v}^\intercal\boldsymbol{L}_{t-1}$ and variance $\eta^2 mt$ to obtain

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

Also observe that the density of $\boldsymbol{V}_t^\intercal\boldsymbol{H}_t$ at $y\in\mathbb{R}$ is exactly

$$\sum_{\boldsymbol{v}\in\mathcal{S}}f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y|\,\boldsymbol{v}^\intercal\boldsymbol{H}_t=y\right].$$

Thus,

$$\begin{aligned}
\mathbb{P}_t\left[|A_t|>1\right]&\leq\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}}\left(f_{\boldsymbol{v}}(y)-f_{\boldsymbol{v}}(y-mc)\right)\cdot\\
&\qquad\cdot\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y|\,\boldsymbol{v}^\intercal\boldsymbol{H}_t=y\right]\,dy\\
&\leq\sum_{\boldsymbol{v}\in\mathcal{S}}\int_{y\in\mathbb{R}}f_{\boldsymbol{v}}(y)\mathbb{P}_t\left[\forall \boldsymbol{w}\neq\boldsymbol{v}:\boldsymbol{w}^\intercal\boldsymbol{H}_t>y|\,\boldsymbol{v}^\intercal\boldsymbol{H}_t=y\right]\cdot\\
&\qquad\cdot\left(\frac{mc^2}{2\eta^2 t}-\frac{c(y-\boldsymbol{v}^\intercal\boldsymbol{L}_{t-1})}{\eta^2 t}\right)\,dy\\
&=\frac{mc^2}{2\eta^2 t}-\frac{c\mathbb{E}\left[\boldsymbol{V}_t^\intercal\boldsymbol{Z}_t\right]}{\eta^2 t}\leq\frac{mc^2}{2\eta^2 t}+\frac{mc\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]}{\eta^2 t}.
\end{aligned}$$

Using the definition of c and $\mathbb{E}\left[\|\boldsymbol{Z}_t\|_\infty\right]\leq\eta\sqrt{2t\log d}$ gives the result. ∎

III. Random-walk perturbations for prediction with expert advice

In this section we refine our method to obtain bounds of optimal order in the special case of prediction with expert advice, that is, when the set of actions is the set of $d=N$ unit vectors. In this case, we use I_t to denote the unique coordinate i of $\boldsymbol{V}_t$ with $V_{i,t}=1$. While the straightforward application of Algorithm 1 and Theorem 1 guarantees a regret of optimal order in this case, the switch-number guarantees are of a suboptimal $O(\sqrt{n}\log N)$. In this section, we propose a variant of our algorithm that achieves order-optimal regret guarantees while switching its predictions only $O(\sqrt{n\log N})$ times, similarly to the algorithms of [7] and [22].

The algorithm—presented as Algorithm 2—is obtained by replacing the Gaussian increments in Algorithm 1 by independent random variables that take values $\pm1/2$ with equal probabilities. The benefit of using this perturbation scheme is that the diameter of the lead packs defined in (1) can be upper bounded by a constant, and thus we can eliminate higher moments of $\|\boldsymbol{X}_{t+1}\|_\infty$ in the upper bound on $\mathbb{P}\left[|A_t|>1\right]$. To gain further intuition, notice that choosing any *fixed* (i.e., non-random) lead-pack diameter in the proof of Lemma 1 would allow experts from outside the lead pack to take the lead with positive probability. While this probability can be decreased at the expense of slightly expanding the diameter, its rate of decay is not sufficiently fast for improving our previously presented results. In fact, balancing the two terms constituting the probability of switching gives the exact same result. On the other hand, when using $\pm1/2$-valued random increments, it is possible to set a fixed diameter of 1 that ensures that the new leader comes from the lead pack with probability 1, and thus the extra term vanishes.

Algorithm 2 Random-walk perturbations for prediction with expert advice.

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
$$I_t=\arg\min_i\left(L_{i,t-1}+Z_{i,t}\right),$$
where ties are broken in favor of I_{t-1}.
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 $i=1,2,\ldots,N$.

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

Theorem 2: The expected regret and expected number of switches of actions of the forecaster of Algorithm 2 satisfy, for all possible loss sequences (under the oblivious-adversary

model),

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

While it is possible to perform the analysis of Algorithm 2 similarly to that of Algorithm 1, we take a different path: The proof we present below is based on the observation that the regret of Algorithm 2 can be bounded in terms of the number of action switches. The next simple lemma formalizes this statement.

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 [2] (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

$$\sum_{t=1}^{n} \ell_{I_t,t} \le L_{i,n} + \sum_{t=1}^{n+1} \big(\ell_{I_{t-1},t-1} - \ell_{I_t,t-1}\big) + Z_{i,n} - \sum_{t=1}^{n+1} X_{I_t,t} \ . \tag{3}$$

The last term can be rewritten as

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

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

$$\begin{aligned}&\sum_{t=1}^{n+1} \big(\ell_{I_{t-1},t-1} - \ell_{I_t,t-1}\big) + \sum_{t=1}^{n+1} \big(X_{I_{t-1},t} - X_{I_t,t}\big) \\ &\qquad \le 2\sum_{t=1}^{n+1} \mathbb{1}_{\{I_{t-1} \ne I_t\}} = 2C_n \ .\end{aligned}$$

Putting everything together gives the statement of the lemma. ∎

Next we analyze the number of switches C_n. Similarly to the analysis of Algorithm 1, we study the *lead pack* A_t defined as

$$A_t = \{i \in \{1, \ldots, N\} : L_{i,t-1} + Z_{i,t} < L_{I_t,t-1} + Z_{I_t,t} + 2\} \, .$$

Once again, observe that no action from outside the lead pack has a positive probability of taking the lead at time $t+1$. 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

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

which is what we need to prove the bounds of Theorem 2.

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:

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

The forecaster picks $I_t \in S_t$ arbitrarily when $I_{t-1} \notin S_t$, otherwise it stays with $I_t = I_{t-1}$. Let us start with analyzing the probability of the event $\{A_t = \{I_t\}\} = \{|A_t| = 1\}$:

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

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

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

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 $\frac{t+Z_{1,t}}{2}$ 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,

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

It can be easily verified that

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

holds for all $k \in [-t, t]$. Using our first observation and the bound on $\mathbb{P}\left[|A_t| = 1\right]$, we get

$$\mathbb{P}\left[|A_t| = 1\right] \ge \sum_{k=-t+4}^{t} \mathbb{P}\left[\min_{j\in S_t} Z_{j,t} = \frac{k}{2}\right]\frac{p_t(k-4)}{p_t(k)} \ .$$

Along with our second observation, this implies

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

Now using $\mathbb{E}\left[\max_j Z_{j,t}\right]\le\sqrt{\frac{t\log N}{2}}$ implies

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

as desired. ∎

APPENDIX

Proof of the first statement of Theorem 1: The proof is based on the proof of Theorem 4.2 of [2] and Theorem 3 of [13], with key insights taken from Neu and Bartók [20]. The main difference from those proofs is that the standard deviation of our perturbations changes over time, however, this issue is easy to treat. First, we define an infeasible "forecaster" that peeks one step into the future and uses perturbation $\widehat{\boldsymbol{Z}}_t=\sqrt{t}\boldsymbol{X}_1$:

$$\widehat{\boldsymbol{V}}_t=\arg\min_{w\in\mathcal{S}}\boldsymbol{w}^\mathsf{T}\left(\boldsymbol{L}_t+\widehat{\boldsymbol{Z}}_t\right).$$

Now fix any $\boldsymbol{v}\in\mathcal{S}$. Applying Lemma 3.1 of [2] for the sequence $\left(\boldsymbol{\ell}_{t-1}+\widehat{\boldsymbol{Z}}_t-\widehat{\boldsymbol{Z}}_{t-1}\right)_{t=1}^{\infty}$ with $\boldsymbol{\ell}_0=\boldsymbol{0}$, we get

$$\sum_{t=1}^{n}\widehat{\boldsymbol{V}}_t^\mathsf{T}(\boldsymbol{\ell}_t+(\widehat{\boldsymbol{Z}}_t-\widehat{\boldsymbol{Z}}_{t-1}))\le\boldsymbol{v}^\mathsf{T}(\boldsymbol{L}_n+\widehat{\boldsymbol{Z}}_n).$$

After reordering, we obtain

$$\begin{aligned}\sum_{t=1}^{n}\boldsymbol{V}_t^\mathsf{T}\boldsymbol{\ell}_t \le&\boldsymbol{v}^\mathsf{T}\boldsymbol{L}_n+\boldsymbol{v}^\mathsf{T}\widehat{\boldsymbol{Z}}_n+\sum_{t=1}^{n}(\boldsymbol{V}_t-\widehat{\boldsymbol{V}}_t)^\mathsf{T}\boldsymbol{\ell}_t\\ &-\sum_{t=1}^{n}\widehat{\boldsymbol{V}}_t^\mathsf{T}(\widehat{\boldsymbol{Z}}_t-\widehat{\boldsymbol{Z}}_{t-1})\\ =&\boldsymbol{v}^\mathsf{T}\boldsymbol{L}_n+\boldsymbol{v}^\mathsf{T}\widehat{\boldsymbol{Z}}_n+\sum_{t=1}^{n}(\boldsymbol{V}_t-\widehat{\boldsymbol{V}}_t)^\mathsf{T}\boldsymbol{\ell}_t\\ &+\sum_{t=1}^{n}(\sqrt{t-1}-\sqrt{t})\widehat{\boldsymbol{V}}_t^\mathsf{T}\boldsymbol{X}_1\end{aligned}$$

The last term can be bounded as

$$\begin{aligned}\sum_{t=1}^{n}(\sqrt{t-1}-\sqrt{t})\widehat{\boldsymbol{V}}_t^\mathsf{T}\boldsymbol{X}_1 \le&\sum_{t=1}^{n}(\sqrt{t}-\sqrt{t-1})\left|\widehat{\boldsymbol{V}}_t^\mathsf{T}\boldsymbol{X}_1\right|\\ \le&m\sum_{t=1}^{n}(\sqrt{t}-\sqrt{t-1})\left\|\boldsymbol{X}_1\right\|_\infty\\ \le&m\sqrt{n}\left\|\boldsymbol{X}_1\right\|_\infty.\end{aligned}$$

Taking expectations, we obtain the bound

$$\mathbb{E}\widehat{L}_n-\boldsymbol{v}^\mathsf{T}\boldsymbol{L}_n\le\sum_{t=1}^{n}\mathbb{E}\left[(\boldsymbol{V}_t-\widehat{\boldsymbol{V}}_t)^\mathsf{T}\boldsymbol{\ell}_t\right]+\eta m\sqrt{2n\log d},$$

where we used $\mathbb{E}\left[\|\boldsymbol{X}_1\|_\infty\right]\le\eta\sqrt{2\log d}$.

Thus, we are left with the problem of bounding $\mathbb{E}\left[(\boldsymbol{V}_t-\widehat{\boldsymbol{V}}_t)^\mathsf{T}\boldsymbol{\ell}_t\right]$ for each $t\ge1$. Similarly to [20], we do this by introducing

$$p_t(\boldsymbol{u})=\mathbb{P}\left[\boldsymbol{V}_t=\boldsymbol{u}\right]\quad\text{and}\quad\hat{p}_t(\boldsymbol{u})=\mathbb{P}\left[\widehat{\boldsymbol{V}}_t=\boldsymbol{u}\right]$$

for all $\boldsymbol{u}\in\mathcal{S}$ and studying the relationship between the distributions $\boldsymbol{p}_t$ and $\hat{\boldsymbol{p}}_t$. To this end, let us fix an arbitrary $\boldsymbol{u}\in\mathcal{S}$ and define the "sparse loss vector" $\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})$ with its k-th component being $\tilde{\ell}_{k,t}(\boldsymbol{u})=u_k\ell_{k,t}$. Let

$$\widetilde{\boldsymbol{V}}_t(\boldsymbol{u})=\arg\min_{\boldsymbol{w}\in\mathcal{S}}\boldsymbol{w}^\mathsf{T}\left(\boldsymbol{L}_{t-1}+\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})+\widehat{\boldsymbol{Z}}_t\right)$$

and

$$\tilde{p}_t(\boldsymbol{u})=\mathbb{P}\left[\widetilde{\boldsymbol{V}}_t(\boldsymbol{u})=\boldsymbol{u}\right].$$

As shown in Lemma 2 of [20], $\tilde{p}_t(\boldsymbol{u})\le\hat{p}_t(\boldsymbol{u})$ holds independently of the distribution of the perturbations, given that all components of the loss vector are nonnegative. Now define

$$\boldsymbol{w}_t(\boldsymbol{z})=\arg\min_{\boldsymbol{w}\in\mathcal{S}}\boldsymbol{w}^\mathsf{T}(\boldsymbol{L}_{t-1}+\boldsymbol{z})$$

for all $\boldsymbol{z}\in\mathbb{R}^d$ and let $f_t(\boldsymbol{z})$ be the density of $\boldsymbol{Z}_t$ (which coincides with the density of $\widehat{\boldsymbol{Z}}_t$). For all $\boldsymbol{u}\in\mathcal{S}$, we have

$$\begin{aligned}p_t(\boldsymbol{u})=&\mathbb{E}\left[\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{Z}_t)=\boldsymbol{u}\}}\right]\\ =&\int_{\boldsymbol{z}\in\mathbb{R}^d}f_t(\boldsymbol{z})\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z}\\ =&\int_{\boldsymbol{z}\in\mathbb{R}^d}f_t\big(\boldsymbol{z}+\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\big)\mathbb{1}_{\left\{\boldsymbol{w}_t\left(\boldsymbol{z}+\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right)=\boldsymbol{u}\right\}}\,d\boldsymbol{z}\\ =&\mathbb{E}\left[\mathbb{1}_{\left\{\boldsymbol{w}_t\left(\widehat{\boldsymbol{Z}}_t+\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right)=\boldsymbol{u}\right\}}\right]\\ &+\int_{\boldsymbol{z}\in\mathbb{R}^d}\Big(f_t\big(\boldsymbol{z}+\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\big)-f_t(\boldsymbol{z})\Big)\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z}\\ =&\tilde{p}_t(\boldsymbol{u})+\int_{\boldsymbol{z}\in\mathbb{R}^d}\Big(f_t\big(\boldsymbol{z}+\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\big)-f_t(\boldsymbol{z})\Big)\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z}.\end{aligned}$$

While the first term is upper bounded by $\hat{p}_t(\boldsymbol{u})$, the last one can be upper bounded as

$$\begin{aligned}&\int_{\boldsymbol{z}\in\mathbb{R}^d}f_t(\boldsymbol{z})\left(1-\exp\left(\frac{\left(\boldsymbol{z}-\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right)^\mathsf{T}\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})}{\eta^2t}\right)\right)\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z}\\ &\le-\int_{\boldsymbol{z}\in\mathbb{R}^d}f_t(\boldsymbol{z})\left(\frac{(\boldsymbol{z}-\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u}))^\mathsf{T}\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})}{\eta^2t}\right)\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z}\\ &\le\frac{p_t(\boldsymbol{u})\left\|\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right\|_2^2}{\eta^2t}+\frac{1}{\eta^2t}\int_{\boldsymbol{z}\in\mathbb{R}^d}f_t(\boldsymbol{z})\left|\boldsymbol{z}^\mathsf{T}\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right|\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z}\\ &\le\frac{p_t(\boldsymbol{u})m}{\eta^2t}+\frac{m}{\eta^2t}\int_{\boldsymbol{z}\in\mathbb{R}^d}f_t(\boldsymbol{z})\left\|\boldsymbol{z}\right\|_\infty\mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}}\,d\boldsymbol{z},\end{aligned}$$

where we have used that $\left\|\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right\|_2^2 \le m$ and $\left\|\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})\right\|_1 \le m$ hold by the definition of $\tilde{\boldsymbol{\ell}}_t(\boldsymbol{u})$. Using that $\mathbb{P}\left[\widehat{\boldsymbol{V}}_t = \boldsymbol{u}\right] = \hat{p}_t(\boldsymbol{u})$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{V}_t^\mathsf{T}\boldsymbol{\ell}_t\right] = \sum_{\boldsymbol{u}\in\mathcal{S}} p_t(\boldsymbol{u})\boldsymbol{u}^\mathsf{T}\boldsymbol{\ell}_t \le \sum_{\boldsymbol{u}\in\mathcal{S}} \hat{p}_t(\boldsymbol{u})\boldsymbol{u}^\mathsf{T}\boldsymbol{\ell}_t \\
&+ \sum_{\boldsymbol{u}\in\mathcal{S}} \left(\frac{p_t(\boldsymbol{u})m}{\eta^2 t} + \frac{m}{\eta^2 t} \int\limits_{\boldsymbol{z}\in\mathbb{R}^d} f_t(\boldsymbol{z}) \left\|\boldsymbol{z}\right\|_\infty \mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}} d\boldsymbol{z} \right) \boldsymbol{u}^\mathsf{T}\boldsymbol{\ell}_t \\
&\le \mathbb{E}\left[\widehat{\boldsymbol{V}}_t^\mathsf{T}\boldsymbol{\ell}_t\right] + \frac{m^2}{\eta^2 t} + \frac{m^2}{\eta^2 t} \int\limits_{\boldsymbol{z}\in\mathbb{R}^d} f_t(\boldsymbol{z}) \left\|\boldsymbol{z}\right\|_\infty \sum_{\boldsymbol{u}\in\mathcal{S}} \mathbb{1}_{\{\boldsymbol{w}_t(\boldsymbol{z})=\boldsymbol{u}\}} \, d\boldsymbol{z} \\
&= \mathbb{E}\left[\widehat{\boldsymbol{V}}_t^\mathsf{T}\boldsymbol{\ell}_t\right] + \frac{m^2}{\eta^2 t} + \frac{m^2}{\eta^2 t}\mathbb{E}\left[\left\|\boldsymbol{Z}_t\right\|_\infty\right] \\
&\le \mathbb{E}\left[\widehat{\boldsymbol{V}}_t^\mathsf{T}\boldsymbol{\ell}_t\right] + \frac{m^2}{\eta^2 t} + \frac{m^2}{\eta}\sqrt{\frac{2\log d}{t}},
\end{aligned}$$

where we used $\mathbb{E}\left[\left\|\boldsymbol{Z}_t\right\|_\infty\right] \le \eta\sqrt{2t\log d}$ in the last step.

Putting everything together, we obtain

$$\begin{aligned}
\mathbb{E}\widehat{L}_n - \boldsymbol{v}^\mathsf{T}\boldsymbol{L}_n &\le \sum_{t=1}^n \frac{m^2}{\eta^2 t} + \sum_{t=1}^n \frac{m^2}{\eta}\sqrt{\frac{2\log d}{t}} + \eta m\sqrt{2n\log d} \\
&\le \frac{2m^2\sqrt{2n\log d}}{\eta} + \eta m\sqrt{2n\log d} + \frac{m^2(\log n + 1)}{\eta^2},
\end{aligned}$$

concluding the proof of the statement. ∎

ACKNOWLEDGMENT

This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Spanish Ministry of Science and Technology grant MTM2012-37195, INRIA, the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 270327 (project CompLACS), the Ministry of Higher Education and Research, and by FUI project Hermès. The authors thank László and János Györfi, as well as László Németh for organizing the workshop PICSA 2012 in Jásd, during which some major details of the paper were worked out.

REFERENCES

[1] L. Devroye, G. Lugosi, and G. Neu, "Prediction by random-walk perturbation," in *Proceedings of the 25th Annual Conference on Learning Theory* (S. Shalev-Shwartz and I. Steinwart, eds.), pp. 460–473, 2013.

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

[3] C. Gentile and M. Warmuth, "Linear hinge loss and average margin," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 225–231, 1998.

[4] J. Kivinen and M. Warmuth, "Relative loss bounds for multidimensional regression problems," *Machine Learning*, vol. 45, pp. 301–329, 2001.

[5] A. Grove, N. Littlestone, and D. Schuurmans, "General convergence results for linear discriminant updates," *Machine Learning*, vol. 43, pp. 173–210, 2001.

[6] E. Takimoto and M. Warmuth, "Paths kernels and multiplicative updates," *Journal of Machine Learning Research*, vol. 4, pp. 773–818, 2003.

[7] A. Kalai and S. Vempala, "Efficient algorithms for online decision problems," *Journal of Computer and System Sciences*, vol. 71, pp. 291–307, 2005.

[8] M. Warmuth and D. Kuzmin, "Randomized online PCA algorithms with regret bounds that are logarithmic in the dimension," *Journal of Machine Learning Research*, vol. 9, pp. 2287–2320, 2008.

[9] D. P. Helmbold and M. Warmuth, "Learning permutations with exponential weights," *Journal of Machine Learning Research*, vol. 10, pp. 1705–1736, 2009.

[10] E. Hazan, S. Kale, and M. Warmuth, "Learning rotations with little regret," in *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, pp. 144–154, 2010.

[11] W. Koolen, M. Warmuth, and J. Kivinen, "Hedging structured concepts," in *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, pp. 93–105, 2010.

[12] N. Cesa-Bianchi and G. Lugosi, "Combinatorial bandits," *Journal of Computer and System Sciences*, vol. 78, pp. 1404–1422, 2012.

[13] J. Y. Audibert, S. Bubeck, and G. Lugosi, "Regret in online combinatorial optimization," *Mathematics of Operations Research*, vol. 39, pp. 31–45, 2014.

[14] N. Littlestone and M. Warmuth, "The weighted majority algorithm," *Information and Computation*, vol. 108, pp. 212–261, 1994.

[15] V. Vovk, "Aggregating strategies," in *Proceedings of the third annual workshop on Computational learning theory (COLT)*, pp. 371–386, 1990.

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

[17] J. Hannan, "Approximation to Bayes risk in repeated play," *Contributions to the theory of games*, vol. 3, pp. 97–139, 1957.

[18] M. Hutter and J. Poland, "Prediction with expert advice by following the perturbed leader for general weights," in *ALT*, pp. 279–293, 2004.

[19] J. Poland, "FPL analysis for adaptive bandits," in *In 3rd Symposium on Stochastic Algorithms, Foundations and Applications (SAGA'05)*, pp. 58–69, 2005.

[20] G. Neu and G. Bartók, "An efficient algorithm for learning with semi-bandit feedback," in *Proceedings of the 24th International Conference on Algorithmic Learning Theory* (S. Jain, R. Munos, F. Stephan, and T. Zeugmann, eds.), vol. 8139 of *Lecture Notes in Computer Science*, pp. 234–248, Springer, 2013.

[21] D. Suehiro, K. Hatano, S. Kijima, E. Takimoto, and K. Nagano, "Online prediction under submodular constraints," in *Algorithmic Learning Theory*, vol. 7568 of *Lecture Notes in Computer Science*, pp. 260–274, Springer Berlin Heidelberg, 2012.

[22] S. Geulen, B. Voecking, and M. Winkler, "Regret minimization for online buffering problems using the weighted majority algorithm," in *Proceedings of the 23rd Annual Conference on Learning Theory (COLT 2010)* (A. Kalai and M. Mohri, eds.), pp. 132–143, 2010.

[23] A. György and G. Neu, "Near-optimal rates for limited-delay universal lossy source coding," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011)*, 2011.

[24] E. Even-Dar, S. M. Kakade, and Y. Mansour, "Online Markov decision processes," *Mathematics of Operations Research*, vol. 34, no. 3, pp. 726–736, 2009.

[25] G. Neu, A. György, Cs. Szepesvári, and A. Antos, "Online Markov decision processes under bandit feedback," in *Advances in Neural Information Processing Systems 23* (J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, eds.), pp. 1804–1812, 2011.

[26] S. Rakhlin, O. Shamir, and K. Sridharan, "Relax and randomize : From value to algorithms," in *Advances in Neural Information Processing Systems 25*, pp. 2150–2158, 2012.

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

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

Luc Devroye (b. Belgium) obtained his Ph.D. from the University of Texas in 1976, and joined the School of Computer Science at McGill University in Montreal, Canada, in 1977. His research interests include probability theory as applied to the analysis of algorithms, mathematical statistics, machine learning, pattern recognition, and random number generation.

Gábor Lugosi graduated in electrical engineering at the Technical University of Budapest in 1987, and received his Ph.D. from the Hungarian Academy of Sciences in 1991. Since 1996, he has been at the Department of Economics, Pompeu Fabra University. In 2006 he became an ICREA research professor. His research interests include learning theory, nonparametric statistics, inequalities in probability, random structures, and information theory.

Gergely Neu received his M.Sc. degree in Electrical Engineering and his Ph.D. degree in Technical Informatics from the Budapest University of Technology and Ecomomics (Hungary) in 2008 and 2013, respectively. Since 2013, he is a postdoctoral fellow at the SequeL team of INRIA Lille – Nord Europe. His research interests include reinforcement learning, online learning, and bandit problems.