

# Combinatorial Bandits<sup>☆</sup>

Nicolò Cesa-Bianchi

*Università degli Studi di Milano, Italy*

Gábor Lugosi<sup>1</sup>

*ICREA and Pompeu Fabra University, Spain*

---

## Abstract

We study sequential prediction problems in which, at each time instance, the forecaster chooses a vector from a given finite set  $\mathcal{S} \subseteq \mathbb{R}^d$ . At the same time, the opponent chooses a “loss” vector in  $\mathbb{R}^d$  and the forecaster suffers a loss that is the inner product of the two vectors. The goal of the forecaster is to achieve that, in the long run, the accumulated loss is not much larger than that of the best possible element in  $\mathcal{S}$ . We consider the “bandit” setting in which the forecaster only has access to the losses of the chosen vectors (i.e., the entire loss vectors are not observed). We introduce a variant of a strategy by Dani, Hayes, and Kakade achieving a regret bound that, for a variety of concrete choices of  $\mathcal{S}$ , is of order  $\sqrt{nd \ln |\mathcal{S}|}$  where  $n$  is the time horizon. This is not improvable in general and is better than previously known bounds. The examples we consider are all such that  $\mathcal{S} \subseteq \{0, 1\}^d$ , and we show how the combinatorial structure of these classes can be exploited to improve the regret bounds. We also point out computationally efficient implementations for various interesting choices of  $\mathcal{S}$ .

*Keywords:* online prediction, adversarial bandit problems, online linear optimization

---

<sup>☆</sup>A preliminary version appeared in the Proceedings of the 22nd Annual Conference on Learning Theory (COLT 2009). The authors gratefully acknowledge partial support by the PASCAL2 Network of Excellence under EC grant no. 216886. This publication only reflects the authors’ views.

*Email addresses:* nicolo.cesa-bianchi@unimi.it (Nicolò Cesa-Bianchi), lugosi@upf.es (Gábor Lugosi)

<sup>1</sup>Supported by the Spanish Ministry of Science and Technology grant MTM2009-09063.

---

## 1. Introduction

Consider a sequential prediction problem in which a forecaster is to choose, at every time instance  $t = 1, \dots, n$ , an element from a set  $\mathcal{S}$  of  $N$  actions (or experts). After making a choice, the forecaster suffers a loss corresponding to the chosen action. The goal of the forecaster is to achieve that the accumulated loss is not much larger than that of the best possible fixed action, chosen in hindsight. The difference between the achieved and optimal cumulative losses is called the *regret*. It is well known (see [1] for a survey) that randomized prediction strategies exist that guaranteeing that the expected regret of the forecaster is bounded by a constant times  $\sqrt{n \ln N}$ , regardless of the sequence of losses, as long as they are bounded. The logarithmic dependence on the number of actions allows one to compete with very large classes of actions. However, large classes raise nontrivial computational issues. The construction of computationally efficient forecasters for various cases of structured classes of experts is a thoroughly studied problem. Once again, we refer to [1] for a survey.

An interesting variant of the sequential prediction problem is the *adversarial multi-armed bandit* problem in which the forecaster only observes the loss of the chosen action and uses the randomized choices to gather information. It was shown by Auer et al. [2] that an expected regret of the order of  $\sqrt{nN \ln N}$  is achievable in this case. There has been a flurry of activity to address versions of the adversarial bandit problem for large and structured classes of experts, see Awerbuch and Kleinberg [3], McMahan and Blum [4], Dani and Hayes [5], György, Linder, Lugosi, and Ottucsák [6], Dani, Hayes, and Kakade [7], Abernethy, Hazan, and Rakhlin [8], Bartlett, Dani, Hayes, Kakade, and Tewari [9], Abernethy and Rakhlin [10].

Most of the effort has been focused on two main issues: (1) obtaining regret bounds as small as possible; (2) constructing computationally feasible forecasters.

In this paper we build on the methodology of Dani, Hayes, and Kakade [7], who introduced a general forecaster with close-to-optimal regret bounds. By a simple generalization of their forecaster we obtain improved regret bounds in many cases, when the finite class of experts has a certain combinatorial structure. We also show that in some interesting cases nontrivial efficient algorithms exist.

The paper is organized as follows. In Section 2 we formulate the problem. In Section 3 we discuss the relationship of our results to earlier work. The general prediction strategy is defined and the main performance bound is established in Section 4. Various applications are described in Section 5, including a multitask bandit problem, learning permutations, learning spanning trees of a complete graph, and learning balanced cut sets.

## 2. Statement of the problem

In the bandit linear optimization problem [7, 8, 9] a finite <sup>2</sup> set  $\mathcal{S} \subseteq \mathbb{R}^d$  of elements  $\mathbf{v}(k)$  for  $k = 1, \dots, N$  is given (this is the set of “experts” or “actions”). The forecaster plays a repeated game with an opponent such that, at each round of the game, the forecaster chooses an index between  $\{1, \dots, N\}$  and the forecaster chooses a loss vector  $\boldsymbol{\ell}_t \in \mathbb{R}^d$ . For all  $k = 1, \dots, N$  denote  $c_t(k) = \boldsymbol{\ell}_t^\top \mathbf{v}(k)$ . If the index chosen by the forecaster at time  $t$  is  $K_t$ , then the only information given to the forecaster is the value of  $c_t(K_t)$ . The game is described as follows:

For each step  $t = 1, 2, \dots$

1. The opponent secretly chooses a loss vector  $\boldsymbol{\ell}_t \in \mathbb{R}^d$
2. The forecaster chooses  $K_t \in \{1, \dots, N\}$
3. The cost  $c_t(K_t) = \boldsymbol{\ell}_t^\top \mathbf{v}(K_t)$  is announced to the forecaster.

The forecaster’s goal is to control the *regret*

$$\widehat{L}_n - \min_{k=1, \dots, N} L_n(k) = \sum_{t=1}^n c_t(K_t) - \min_{k=1, \dots, N} \sum_{t=1}^n \boldsymbol{\ell}_t^\top \mathbf{v}(k) .$$

Similarly to [7] we assume that  $|\boldsymbol{\ell}_t^\top \mathbf{v}| \leq 1$  for all  $\mathbf{v} \in \mathcal{S}$  and  $t$ . If 1 is replaced by an arbitrary known positive constant, then the bound on the regret of our forecasting strategy (Theorem 1) must be multiplied by the same scaling constant.

The forecaster is allowed to use randomization. More precisely, at every time instance  $t$ , the forecaster chooses a distribution  $p_{t-1}(1), \dots, p_{t-1}(N)$  over

---

<sup>2</sup>If  $\mathcal{S}$  is infinite but bounded, then [7, Lemma 3.1] shows that it can be approximated with a finite class of size order of  $(dn)^{d/2}$ , causing any forecaster, working on the finite class, to suffer an extra regret w.r.t.  $\mathcal{S}$  of order  $\sqrt{dn}$ .

the set  $\{1, \dots, N\}$  (i.e.,  $p_{t-1}(k) \geq 0$  for all  $k = 1, \dots, N$  and  $\sum_{k=1}^N p_{t-1}(k) = 1$ ) and draws an index  $K_t = k$  with probability  $p_{t-1}(k)$ . Thus, the regret is a random variable. In this paper we investigate the behavior of the *expected regret*

$$\max_{k=1, \dots, N} \mathbb{E} \left[ \widehat{L}_n - L_n(k) \right]$$

where the expectation is with respect to the forecaster’s internal randomization. If the opponent is *oblivious*, that is, the actions of the opponent do not depend on the past actions of the forecaster (see, e.g., [1, Chapter 4] for a formal definition and discussion), then  $L_n(k)$  is not a random variable and the expected regret is simply

$$\mathbb{E} \widehat{L}_n - \min_{k=1, \dots, N} L_n(k) .$$

In this paper we do not restrict ourselves to oblivious opponents.

The most important parameters of the problem are the time horizon  $n$ , the dimension  $d$ , the cardinality  $N$  of the action set  $\mathcal{S}$ , and the maximum “size” of any expert

$$B = \max_{\mathbf{v} \in \mathcal{S}} \|\mathbf{v}\|$$

where  $\|\cdot\|$  indicates the Euclidean norm.

The *combinatorial bandit problem* is a special case of the bandit linear optimization problem where we restrict  $\mathcal{S}$  to be a subset of the binary hypercube  $\{0, 1\}^d$ . This fact allows us to exploit the combinatorial structure of the class of experts in a transparent way. Arguably, the most interesting examples of online linear optimization fit in the combinatorial framework. In the rest of the paper we only consider the “combinatorial” case  $\mathcal{S} \subseteq \{0, 1\}^d$  though the general forecasting strategy and regret bound below extend to arbitrary sets  $\mathcal{S} \subseteq \mathbb{R}^d$  in a straightforward manner.

### 3. Relation to previous work

When  $d = N$  and  $\mathbf{v}(1), \dots, \mathbf{v}(N)$  are the standard basis vectors, then the model is identical to the adversarial bandit problem introduced by Auer et al. [2], who proved a regret bound of the order of  $\sqrt{nN \ln N}$  that holds not only in expectation but also with high probability. (We refer to Audibert and Bubeck [11] for recent improvements of this result.) A well-studied instance of our general framework is the *path planning* problem, in which  $d$  is the number

of edges of a fixed graph and  $\mathbf{v}(1), \dots, \mathbf{v}(N)$  represent all paths between two fixed vertices of the graph. More precisely, each  $\mathbf{v}(k) \in \mathcal{S} \subset \{0, 1\}^d$  is the incidence vector of a path: a component of  $\mathbf{v}(k)$  equals 1 if and only if the corresponding edge is present in the path represented by  $\mathbf{v}(k)$ . At each time instance the forecaster chooses a path and suffers a loss that is the sum of the losses over the individual edges of the chosen path. Takimoto and Warmuth [12] and Kalai and Vempala [13] exhibit computationally efficient forecasters in the “full-information” case, that is, when the forecaster has access to the losses over every edge of the graph.

The partial information setting considered in this paper was first studied by Awerbuch and Kleinberg [3] who proved a regret bound of order  $n^{2/3}$  for the restricted model of oblivious opponent. McMahan and Blum [4] achieved a regret bound of order  $n^{3/4}$  for the general model.

Both [3] and [4] study the somewhat more general framework of *online linear optimization*, introduced by Kalai and Vempala [13]. György et al. [6] considered the problem of path planning in a less demanding partial information framework, when the loss of every edge on the chosen path is revealed to the forecaster. They exhibit a computationally efficient forecaster achieving a regret of order  $\sqrt{nd \ln N}$  with high probability. Even though [6] only considers the path planning problem, it is not difficult to extend their results to the more general setup of this paper. However, the model considered here, that is, when the forecaster only receives information about the total loss of the chosen action, is more challenging. Dani, Hayes, and Kakade [7] were the first to prove an expected regret bound with the optimal  $\sqrt{n}$  dependence on the time horizon. Their bound is of the form  $B\sqrt{nd \ln N}$ . Bartlett, Dani, Hayes, Kakade, Rakhlin, and Tewari [9] show that this bound also holds with high probability. The forecaster of [7] is based on exponential weights and can be computed efficiently whenever efficient implementations of the exponentially weighted average forecaster are available. This is certainly possible for the path planning problem, but there are various other interesting examples —see the discussion of the examples in Section 5 below. Abernethy, Hazan, and Rakhlin [8] consider a very different approach which allows one to construct computationally efficient forecasters for a large variety of problems and has an expected regret of the order of  $d\sqrt{n\theta \ln n}$ , where the parameter  $\theta$  depends on the class of actions  $\mathcal{S}$  (which is supposed to be a convex set). This requires the construction of a *self-concordant* function tailored to the problem at hand. Even though the existence of such a function is guaranteed, its construction (and estimation of the parameter  $\theta$ ) may be a nontrivial task

**Algorithm:** COMBAND

**Parameters:** Finite action set  $\mathcal{S} \subseteq \{0, 1\}^d$ , sampling distribution  $\mu$  over  $\mathcal{S}$ , mixing coefficient  $\gamma > 0$ , learning rate  $\eta > 0$

**Initialization:**  $q_0 =$  uniform distribution on  $\mathcal{S}$

**For**  $t = 1, 2, \dots$

1. Let  $p_{t-1} = (1 - \gamma)q_{t-1} + \gamma\mu$
2. Draw action  $K_t$  from  $p_{t-1}$
3. Incur and observe cost  $c_t(K_t) = \boldsymbol{\ell}_t^\top \mathbf{v}(K_t)$
4. Let  $P_{t-1} = \mathbb{E}[\mathbf{V} \mathbf{V}^\top]$  where  $\mathbf{V}$  has law  $p_{t-1}$
5. Let  $\tilde{\boldsymbol{\ell}}_t = c_t(K_t)P_{t-1}^+ \mathbf{v}(K_t)$
6. Update  $q_t(k) \propto q_{t-1}(k) \exp\left(-\eta \tilde{\boldsymbol{\ell}}_t^\top \mathbf{v}(k)\right)$  for all  $k = 1, \dots, N$ .

Figure 1: The bandit forecaster COMBAND described in Section 4.

in some applications. Abernethy and Rakhlin [10] extend this to analogous regret bounds that hold with high probability.

In this paper we revisit the approach of Dani, Hayes, and Kakade [7]. Like [7], we construct unbiased estimates of each loss component  $\ell_{t,i}$ ,  $i = 1, \dots, d$  and define an exponentially weighted average forecaster based on these estimates. The main difference is in the *exploration* part of the algorithm. Following Awerbuch and Kleinberg [3], Dani, Hayes and Kakade construct a *barycentric spanner* of the set  $\mathcal{S}$  and ensure exploration by mixing the exponential weights with the uniform distribution on spanners. Instead, we use a mixing term derived from a possibly different distribution over  $\mathcal{S}$ . (We mostly consider uniform sampling though other distributions may be advantageous in some examples.) This allows us to achieve an expected regret bound of the order of  $\sqrt{nd \ln N}$  whenever the smallest eigenvalue of a certain matrix associated with  $\mathcal{S}$  (and the sampling distribution) is not too small. The largest part of our efforts is dedicated to show that this smallest eigenvalue can indeed be handled by exploiting the combinatorial structure of the class of experts in a number of interesting cases. Note that the bound  $\sqrt{nd \ln |\mathcal{S}|}$  is not improvable in general when  $\mathcal{S} \subseteq \{0, 1\}^d$ . This follows from a result of [7], as it is shown in Section 5.2 below.

#### 4. The forecasting strategy

The algorithm COMBAND maintains a weight vector defined, at each time  $t$ , by  $w_{t,i} = \exp(-\eta \tilde{L}_{t,i})$  for  $i = 1, \dots, d$ , where  $\tilde{L}_{t,i} = \tilde{\ell}_{1,i} + \dots + \tilde{\ell}_{t,i}$  is a cumulative pseudo-loss, see (1) below, and  $\eta > 0$  is a fixed parameter. Initially,  $w_{0,i} = 1$  for all  $i$ . These weights define corresponding weights  $\bar{w}_t(1), \dots, \bar{w}_t(N) \in \mathbb{R}$  over the elements of  $\mathcal{S}$  in the natural way:

$$\bar{w}_t(k) = \prod_{i: v_i(k)=1} w_{t,i}.$$

Let  $\bar{W}_t = \sum_{k=1}^N \bar{w}_t(k)$  and let  $q_t(k) = \bar{w}_t(k)/\bar{W}_t$ . Note that  $q_0$  is the uniform distribution on  $\mathcal{S}$  because we set  $w_{0,i} = 1$  for all  $i$ . At each time  $t$ , COMBAND plays  $\mathbf{v}(K_t) \in \mathcal{S}$ , where  $K_t$  is drawn from the distribution  $p_{t-1} = (1-\gamma)q_{t-1} + \gamma\mu$  on  $1, \dots, N$ . Here  $\mu$  is any distribution on  $\{1, \dots, N\}$  and  $\gamma > 0$  is a parameter. An equivalent description of the algorithm, without the explicit use of the weights  $w_{t,i}$ , is given in Figure 1. Thus,  $p_{t-1}$  is a mixture of the exponentially weighted distribution  $q_{t-1}$  representing *exploitation* and the fixed distribution  $\mu$  that is responsible of *exploration*. The choice of  $\mu$  is crucial for the performance of the algorithm, and one of the main purposes of the paper is to take a step towards understanding how  $\mu$  should be selected in each problem (i.e., for each set  $\mathcal{S}$ ). We show that in many applications choosing  $\mu$  to be the uniform distribution leads to close-to-optimal performance.

The vector of pseudo-losses  $\tilde{\ell}_t = (\tilde{\ell}_{t,1}, \dots, \tilde{\ell}_{t,d})$  is defined by

$$\tilde{\ell}_t = c_t(K_t) P_{t-1}^+ \mathbf{v}(K_t) \tag{1}$$

where  $P^+$  is the pseudo-inverse of the  $d \times d$  correlation matrix  $\mathbb{E}[\mathbf{V} \mathbf{V}^\top]$  for  $\mathbf{V} \in \mathcal{S}$  distributed according to  $p_{t-1}$ . (Throughout the paper, we use an index  $k = 1, \dots, N$  and its corresponding element  $\mathbf{v}(k) \in \mathcal{S}$  interchangeably.) We also use the notation  $\tilde{c}_t(k) = \tilde{\ell}_t^\top \mathbf{v}(k)$ .

As we mentioned before, COMBAND can be viewed as a generalization of the GEOMETRICHEDGE algorithm of Dani, Hayes and Kakade. The only substantial difference is that we perform exploration by drawing actions from a distribution  $\mu$  over the entire set  $\mathcal{S}$  (step 1 in Figure 1) instead of drawing from a barycentric spanner. This fact gives us a finer control on the loss estimates  $\tilde{\ell}_{t,i}$  in which the factor  $\|P_{t-1}^+\|$  occurs —see (1) above. Indeed, while [7] only achieves  $\|P_{t-1}^+\| \leq d/\gamma$  due to the mix of the barycentric

spanners in  $P_t$ , we can afford the more detailed bound  $\|P_{t-1}^+\| \leq 1/(\gamma\lambda_{\min})$ , where  $\lambda_{\min}$  is the smallest nonzero eigenvalue of the correlation matrix of the initial sampling distribution  $\mu$ . In concrete cases, the computation of tight lower bounds on  $\lambda_{\min}$  allows us to obtain better regret bounds. The COMBAND performance bound stated below indicates that choosing  $\mu$  to ensure that  $\lambda_{\min}$  is as large as possible guarantees better bounds.

**Theorem 1.** *Let  $\mathcal{S}$  be a finite subset of  $\{0, 1\}^d$  and let  $M = \mathbb{E}[\mathbf{V} \mathbf{V}^\top]$  where  $\mathbf{V} \in \mathcal{S}$  is a random vector distributed according to an arbitrary distribution  $\mu$  such that  $\mathcal{S}$  is in the vector space spanned by the support of  $\mu$ . If COMBAND is run with parameters  $\mathcal{S}$ ,  $\mu$ ,*

$$\gamma = \frac{B}{\lambda_{\min}} \sqrt{\frac{\ln N}{n \left( \frac{d}{B^2} + \frac{2}{\lambda_{\min}} \right)}} \quad \text{and} \quad \eta = \frac{1}{B} \sqrt{\frac{\ln N}{n \left( \frac{d}{B^2} + \frac{2}{\lambda_{\min}} \right)}}$$

where  $N = |\mathcal{S}|$ ,  $\lambda_{\min}$  is the smallest nonzero eigenvalue of  $M$ , and  $B \geq \|\mathbf{v}\|$  for all  $\mathbf{v} \in \mathcal{S}$ , then its expected regret after  $n$  steps satisfies

$$\max_{k=1, \dots, N} \mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 2 \sqrt{\left( \frac{2B^2}{d\lambda_{\min}} + 1 \right) nd \ln N} .$$

The proof of Theorem 1, which is based on an appropriate modification of the performance bound of Dani, Hayes, and Kakade [7], is given in Appendix A.

The theorem shows that the success of the forecaster crucially depends on the value of the smallest nonzero eigenvalue  $\lambda_{\min}$  of the correlation matrix  $M$  corresponding to  $\mu$ . In Section 5 we work out various examples in which, for the uniform distribution  $\mu$ ,  $B^2/(d\lambda_{\min}) = \mathcal{O}(1)$ . In all these cases we obtain

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] = \mathcal{O} \left( \sqrt{nd \ln N} \right) . \quad (2)$$

Rewriting the above condition as  $\lambda_{\min} = \Omega(B^2/d)$ , and observing that  $M$  has trace bounded by  $B^2$ , reveals that we achieve (2) whenever the eigenvalues of  $M$  tend to be equal.

Inequality (2) improves on the bound of Dani, Hayes, and Kakade [7] by a factor of  $B$  and on the bound of Abernethy, Hazan, and Rakhlin [8]

by a factor of  $\sqrt{(d\theta \ln n)/(\ln(N))}$ .<sup>3</sup> Computationally, both COMBAND and GEOMETRICHEDGE face the problem of sampling from distributions defined over  $\mathcal{S}$ . In many cases this can be done efficiently, as we discuss in Section 5. The algorithm of [8], instead, works in a completely different way. It performs a randomized gradient descent in the convex hull of  $\mathcal{S}$ , translating each point  $\mathbf{x}_t$  in the convex hull into a distribution over  $\mathcal{S}$ . This is done in such a way that sampling  $K_t$  from this distribution ensures  $\mathbb{E}[\boldsymbol{\ell}_t^\top \mathbf{v}(K_t)] = \boldsymbol{\ell}_t^\top \mathbf{x}_t$ . The efficiency of this procedure depends on the specific choice of  $\mathcal{S}$  (for the path planning problem efficient procedures exist). Moreover, in order to guarantee a good regret, gradient descent is implemented using a *self-concordant* function tailored to the problem. Even if the existence of such a function is guaranteed, its construction may be a non-trivial issue in some applications.

REMARK: CHOICE OF SAMPLING DISTRIBUTION. The upper bound of Theorem 1 suggests a way of choosing the distribution  $\mu$  used for random sampling in the exploration phase: the larger the smallest nonzero eigenvalue  $\lambda_{\min}(M)$ , the tighter the upper bound. In many cases for the uniform distribution  $\mu$  one has  $\lambda_{\min} = \Omega(B^2/d)$  and the order of magnitude of the bound of Theorem 1 cannot be improved for any other distribution. In Section 5 we show several such examples. However, the uniform distribution may be a very bad choice in some case. Indeed, in Section 5.9 we show that in some instances of the path planning problem  $\lambda_{\min}$  may be exponentially small as a function of  $d$ . On the other hand,  $\lambda_{\min} = \Omega(1/d)$  is achievable for *all* classes  $\mathcal{S}$ . Indeed, if  $\mu$  is uniformly distributed over the  $d$  vectors of a *barycentric spanner* (i.e., a collection of  $d$  vectors such that every  $\mathbf{v} \in \mathcal{S}$  can be expressed as a linear combination of these vectors with coefficients between  $-1$  and  $1$ ), then  $\lambda_{\min} \geq 1/d$  as shown in [7]. This choice, while safe, is sub-optimal in general. A more general approach is to determine  $\mu$  so that the value of  $\lambda_{\min}$  is maximized. This may be cast as a *semidefinite programming* problem —see [14, Problem 4.43].

REMARK: REGRET BOUNDS THAT HOLD WITH HIGH PROBABILITY. Theorem 1 bounds the largest *expected* regret  $\max_k \mathbb{E}[\widehat{L}_n - L_n(k)]$  where expectation is taken with respect to the randomized choices of the forecaster.

---

<sup>3</sup>In all applications of Section 5,  $\ln N = \mathcal{O}(\sqrt{d} \ln d)$ . Hence the improvement on [8] is at least by a factor of  $d^{1/4} \sqrt{\theta \ln(n)/\ln(d)}$ , where  $\theta$  is known to be bounded by a polynomial function of  $d$  but may be difficult to determine in specific cases.

However, one may argue that it is more important to bound the realized regret  $\max_k (\widehat{L}_n - L_n(k))$  with high probability. Bartlett, Dani, Hayes, Kakade, Rakhlin, and Tewari [9] showed how one can guarantee that the performance bound of Dani, Hayes, and Kakade [7] holds not only in expectation but also with high probability. The same argument can be used in our case as well. The straightforward but technical details are omitted.

## 5. Applications

In order to apply Theorem 1 to concrete classes  $\mathcal{S}$  we need to find lower bounds on the smallest eigenvalue  $\lambda_{\min} = \lambda_{\min}(M)$  of the linear transformation

$$M = \sum_{k=1}^N \mathbf{v}(k) \mathbf{v}(k)^\top \mu(k)$$

restricted to the vector space  $U$  spanned by the elements  $\mathbf{v}(1), \dots, \mathbf{v}(N)$  of  $\mathcal{S}$ . Since  $\mu$  has support  $\mathcal{S}$ , Lemma 13 implies that this smallest eigenvalue is strictly positive. Thus we want to bound

$$\lambda_{\min} = \min_{\mathbf{x} \in U: \|\mathbf{x}\|=1} \mathbf{x}^\top M \mathbf{x} .$$

In all of our examples (with the exception of Section 5.9) we assume that  $\mu$  is uniform over the set  $\mathcal{S}$ . It is convenient to consider a random vector  $\mathbf{V}$ , distributed according to  $\mu$  over  $\mathcal{S}$ . Then we have

$$\lambda_{\min} = \min_{\mathbf{x} \in U: \|\mathbf{x}\|=1} \mathbb{E} \mathbf{x}^\top \mathbf{V} \mathbf{V}^\top \mathbf{x} .$$

Since  $\mathbf{x}^\top \mathbf{V} \mathbf{V}^\top \mathbf{x} = (\mathbf{V}^\top \mathbf{x})^2$  we have the following simple property.

**Lemma 2.**

$$\lambda_{\min} = \min_{\mathbf{x} \in U: \|\mathbf{x}\|=1} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] .$$

In what follows we write any  $\mathbf{x} \in U$  as  $\mathbf{x} = \sum_{k=1}^N a(k) \mathbf{v}(k)$  where we let  $\sum_k a(k) = \alpha$ .

### 5.1. A multitask bandit problem

In this first example we consider the case when the decision maker acts in  $m$  games in parallel. For simplicity, assume that in each one of the  $m$  games, the decision maker selects one of  $R$  possible actions (a possibly different action in each game). After selecting the  $m$  actions, only the sum of the losses suffered in the  $m$  games is observed. If the loss of each action in each game is bounded between 0 and  $1/m$ , then the condition  $|\boldsymbol{\ell}_t^\top \mathbf{v}| \leq 1$  is satisfied.

**Proposition 3.** *For the multitask bandit problem,  $\lambda_{\min} = 1/R$ .*

In this case  $B = \sqrt{m}$ ,  $d = mR$ ,  $B^2/(d\lambda_{\min}) = 1$ , and  $N = R^m$ . Therefore the optimal regret bound (2) holds and becomes

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 2m\sqrt{3nR \ln R} .$$

Thus, when playing  $m$  games in parallel, the price of getting information about the sum of the losses in spite of the losses suffered separately in each game is just a factor of  $m$  in the regret bound. In this special case COMBAND can be implemented efficiently since it suffices to sample actions independently in each one of the  $R$  games.

PROOF. We can write the elements of  $\mathcal{S} \subseteq \{0, 1\}^d$  as vectors  $\mathbf{v}(k) \in \{0, 1\}^d$ ,  $k = 1, \dots, R^m$ , with components  $v_{j,i}(k)$ ,  $j = 1, \dots, m$ ,  $i = 1, \dots, R$ . These vectors satisfy

$$\sum_{i=1}^R v_{j,i}(k) = 1 \tag{3}$$

for each  $j = 1, \dots, m$  and  $k = 1, \dots, N = R^m$ . According to Lemma 2, we want to lower bound  $\mathbb{E}[(\mathbf{V}^\top \mathbf{x})^2]$  uniformly over  $\mathbf{x}$  in the span of  $\mathcal{S}$ , where  $\mathbf{V}$  is uniformly distributed over  $\mathcal{S}$ . We denote the components of  $\mathbf{V}$  by  $V_{j,i}$ ,  $j = 1, \dots, m$ ,  $i = 1, \dots, R$  and the corresponding components of  $\mathbf{x}$  by  $x_{j,i}$ . We calculate  $\mathbb{E}[(\mathbf{V}^\top \mathbf{x})^2] = \text{VAR}[\mathbf{V}^\top \mathbf{x}] + \mathbb{E}^2[\mathbf{V}^\top \mathbf{x}]$  where  $\mathbf{x} = \sum_{k=1}^N a(k)\mathbf{v}(k)$  is such that  $\|\mathbf{x}\| = 1$ . By (3), for each  $j = 1, \dots, m$ ,

$$\sum_{i=1}^R x_{j,i} = \sum_{k=1}^N a(k) \sum_{i=1}^R v_{j,i}(k) = \sum_{k=1}^N a(k) = \alpha .$$

Thus,

$$\mathbb{E} \mathbf{V}^\top \mathbf{x} = \sum_{j=1}^m \sum_{i=1}^R x_{j,i} \mathbb{E} V_{j,i} = \sum_{j=1}^m \frac{1}{R} \sum_{i=1}^R x_{j,i} = \frac{m}{R} \alpha .$$

On the other hand, since the  $R$ -vectors  $(V_{j,1}, \dots, V_{j,R})$  are independent for  $j = 1, \dots, m$ ,

$$\begin{aligned} \text{VAR}[\mathbf{V}^\top \mathbf{x}] &= \sum_{j=1}^m \text{VAR} \left[ \sum_{i=1}^R x_{j,i} V_{j,i} \right] \\ &= \sum_{j=1}^m \left( \mathbb{E} \left[ \left( \sum_{i=1}^R x_{j,i} V_{j,i} \right)^2 \right] - \mathbb{E}^2 \left[ \sum_{i=1}^R x_{j,i} V_{j,i} \right] \right) \\ &= \sum_{j=1}^m \left( \frac{1}{R} \sum_{i=1}^R x_{j,i}^2 - \left( \frac{1}{R} \sum_{i=1}^R x_{j,i} \right)^2 \right) \\ &= \frac{1}{R} - \frac{m}{R^2} \alpha^2 . \end{aligned}$$

Thus,

$$\mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] = \frac{1}{R} + \frac{m(m-1)}{R^2} \alpha^2 \geq \frac{1}{R}$$

with equality whenever  $\alpha = 0$ .

## 5.2. The hypercube

Suppose next that  $\mathcal{S} = \{0, 1\}^d$  is the entire binary hypercube. This example is interesting because in this case the upper bound of Theorem 1 is optimal up to a constant factor. Indeed, Dani, Hayes, and Kakade [7] shows that there exists an absolute constant  $\kappa$  such that no forecaster can achieve an expected regret smaller than  $\kappa d \sqrt{n}$  for all sequences of loss vectors satisfying  $|\ell_t^\top \mathbf{v}| \leq 1$  for all  $\mathbf{v} \in \{0, 1\}^d$ .

To apply Theorem 1, note that  $N = 2^d$ ,  $B = \sqrt{d}$ , and  $\lambda_{\min} = 1/4$ . This last identity follows simply by Lemma 2 because if  $\mathbf{V} = (V_1, \dots, V_d)$  is uniformly distributed over  $\{0, 1\}^d$  then  $V_1, \dots, V_d$  are independent Bernoulli  $(1/2)$

random variables and then for all  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$  with  $\|\mathbf{x}\| = 1$ ,

$$\begin{aligned} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] &= \mathbb{E} \left[ \sum_{i=1}^d V_i^2 x_i^2 \right] + \mathbb{E} \left[ \sum_{i \neq j} V_i V_j x_i x_j \right] \\ &= \frac{1}{2} \sum_{i=1}^d x_i^2 + \frac{1}{4} \sum_{i \neq j} x_i x_j \\ &= \frac{1}{4} \|\mathbf{x}\|^2 + \frac{1}{4} \left( \sum_{i=1}^d x_i \right)^2 \\ &\geq \frac{1}{4} \end{aligned}$$

with equality whenever  $\sum_{i=1}^d x_i = 0$ . Thus, Theorem 1 implies that for all sequences of loss vectors with  $|\ell_t^\top \mathbf{v}| \leq 1$  for all  $\mathbf{v} \in \{0, 1\}^d$ ,

$$\max_{k=1, \dots, N} \mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 6d\sqrt{n \ln 2}$$

matching the lower bound of [7].

### 5.3. Perfect matchings: learning permutations

Consider the complete bipartite graph  $K_{m,m}$  and let  $\mathcal{S}$  contain all perfect matchings. Thus,  $d = m^2$  (the number of edges of  $K_{m,m}$ ),  $\mathcal{S}$  has  $N = m!$  members, and each perfect matching has  $m$  edges and therefore  $B = \sqrt{m}$ . Each  $\mathbf{v}(k) \in \mathcal{S}$  may be represented by an  $m \times m$  permutation matrix  $[v_{i,j}(k)]_{m \times m}$ ; that is, a zero-one matrix such that  $\sum_{j=1}^m v_{i,j}(k) = 1$  for all  $i = 1, \dots, m$  and  $\sum_{i=1}^m v_{i,j}(k) = 1$  for all  $j = 1, \dots, m$ . Online learning of perfect matchings (or, equivalently, permutations) was considered by Helmbold and Warmuth [15] who introduced a computationally efficient forecaster with good regret bounds in the full-information setting. Koolen, Warmuth, and Kivinen [20] extend this to general classes. However, proving good regret guarantees for an adaptation of their method to the bandit setting remains a challenge.

Here we show that COMBAND performs well for this problem and point out that it has a computationally efficient implementation. The next proposition shows that the term  $\lambda_{\min}$  in Theorem 1 is sufficiently large.

Let  $[V_{i,j}]_{m \times m}$  be chosen uniformly at random from the collection

$$[v_{i,j}(k)]_{m \times m} \quad k = 1, \dots, N$$

representing a random permutation (i.e., perfect matching).

**Proposition 4.** *For the perfect matchings on  $K_{m,m}$ ,*

$$\lambda_{\min} = \frac{1}{m-1} .$$

It follows from the proposition that  $B^2/d\lambda_{\min} \leq 1$ , and therefore the optimal bound (2) holds and it takes the form

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 2m\sqrt{3n \ln(m!)}$$

under the condition  $|\boldsymbol{\ell}_t^\top \mathbf{v}| \leq 1$ , which is fulfilled if the loss corresponding to every edge of  $K_{m,m}$  is bounded between 0 and  $1/m$ .

The fact that COMBAND can be implemented efficiently follows from a beautiful and deep result of Jerrum, Sinclair, and Vigoda [16] who were the first to describe a polynomial-time randomized algorithm for approximating the permanent of a matrix with non-negative entries. To see the connection, observe that the sum of the weights  $\overline{W}_t = \sum_{k=1}^{m!} \overline{w}_t(k)$  is just the permanent of a matrix with entries  $\exp(-\eta \widetilde{L}_{t,(i,j)})$ ,  $i, j \in \{1, \dots, m\}$  where  $\widetilde{L}_{t,(i,j)}$  is the estimated cumulative loss of edge  $(i, j)$ . The algorithm of Jerrum, Sinclair, and Vigoda is based on random sampling perfect matchings from the (approximate) distribution given by the  $\overline{w}_t(k)$  which is exactly what we need to draw a random perfect matching according to the exponentially weighted average distribution.

PROOF. By Lemma 2, we need a lower bound for

$$\mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] = \mathbb{E} \left[ \left( \sum_{i=1}^m \sum_{j=1}^m V_{i,j} x_{i,j} \right)^2 \right]$$

where  $\mathbf{x} = \sum_{k=1}^N a(k) \mathbf{v}(k)$  is such that  $\sum_{i,j=1}^m x_{i,j}^2 = 1$ . Observe that for any fixed  $i$ ,

$$\sum_{j=1}^m x_{i,j} = \sum_{k=1}^N a(k) \sum_{j=1}^m v_{i,j}(k) = \sum_{k=1}^N a_k = \alpha$$

and similarly, for any fixed  $j$ ,  $\sum_{i=1}^m x_{i,j} = \sum_{k=1}^N a_k = \alpha$ . Since

$$\mathbb{P}\{V_{i,j} = 1, V_{i',j'} = 1\} = \begin{cases} \frac{1}{m} & \text{if } i = i' \text{ and } j = j', \\ \frac{1}{m(m-1)} & \text{if } i \neq i' \text{ and } j \neq j', \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\begin{aligned}
\mathbb{E}\left[(\mathbf{V}^\top \mathbf{x})^2\right] &= \mathbb{E}\left[\left(\sum_{i,j=1}^m V_{i,j} x_{i,j}\right)^2\right] \\
&= \sum_{i,j=1}^m \sum_{i',j'=1}^m x_{i,j} x_{i',j'} \mathbb{P}\{V_{i,j} = 1, V_{i',j'} = 1\} \\
&= \frac{1}{m} \sum_{i,j=1}^m x_{i,j}^2 + \frac{1}{m(m-1)} \sum_{i,j=1}^m \sum_{i':i' \neq i} \sum_{j':j' \neq j} x_{i,j} x_{i',j'} \\
&= \frac{1}{m} + \frac{1}{m(m-1)} \sum_{i,j=1}^m \sum_{i':i' \neq i} \sum_{j':j' \neq j} x_{i,j} x_{i',j'} .
\end{aligned}$$

The second term on the right-hand side may be written as

$$\begin{aligned}
&\sum_{i,j=1}^m \sum_{i':i' \neq i} \sum_{j':j' \neq j} x_{i,j} x_{i',j'} = \sum_{i,j=1}^m \sum_{i',j'=1}^m x_{i,j} x_{i',j'} \\
&\quad - \sum_{i,j=1}^m \sum_{j'=1}^m x_{i,j} x_{i',j'} - \sum_{i,j=1}^m \sum_{i'=1}^m x_{i,j} x_{i',j'} + \frac{1}{m} \sum_{i,j=1}^m x_{i,j}^2 \\
&= \left(\sum_{i,j=1}^m x_{i,j}\right)^2 - \sum_{i=1}^m \left(\sum_{j=1}^m x_{i,j}\right)^2 \\
&\quad - \sum_{j=1}^m \left(\sum_{i=1}^m x_{i,j}\right)^2 + 1 \\
&= \left(m \sum_{k=1}^N a(k)\right)^2 - 2m \left(\sum_{k=1}^N a(k)\right)^2 + 1 .
\end{aligned}$$

Summarizing, we have that for all  $\mathbf{x} = \sum_{k=1}^N a(k) \mathbf{v}(k)$  such that  $\|\mathbf{x}\| = 1$ ,

$$\begin{aligned}
\mathbb{E}\left[(\mathbf{V}^\top \mathbf{x})^2\right] &= \frac{1}{m} + \frac{1}{m(m-1)} \left((m\alpha)^2 - 2m\alpha^2 + 1\right) \\
&= \frac{1}{m-1} + \frac{m-2}{m-1} \alpha^2
\end{aligned}$$

which is at least  $1/(m-1)$  with equality whenever  $\alpha = 0$ .

#### 5.4. Spanning trees

Next we consider an online decision problem in which, at each time instance, the decision maker chooses a spanning tree in a graph of  $m$  nodes. The loss of a spanning tree is the sum of the losses over the edges of the tree. Such a problem is meaningful in certain mobile communication networks, in which a minimum-cost subnetwork is to be selected at each time frame to assure connectedness of the whole network. This problem fits in our general framework if we let  $\mathcal{S}$  be the family of all spanning trees of the complete graph  $K_m$ . If all edge losses are in  $[0, 1/(m-1)]$  then  $|\ell_t^\top \mathbf{v}| \leq 1$  holds. Thus,  $d = \binom{m}{2}$ ,  $B = \sqrt{m-1}$ , and by Cayley's formula there are  $N = m^{m-2}$  spanning trees.

In order to estimate  $\lambda_{\min}$  for this case, we start with a general lemma that applies for all sufficiently "symmetric" classes  $\mathcal{S}$ . More precisely, we consider the case when the elements of  $\mathcal{S} \subseteq \{0, 1\}^d$  are the incidence vectors of certain subsets of the edges of a complete graph  $K_m$  (i.e.,  $d = \binom{m}{2}$  in these cases). If  $i$  and  $j$  are distinct edges of  $K_m$ , we write  $i \sim j$  when  $i$  and  $j$  are adjacent (i.e., they have a common endpoint) and  $i \not\sim j$  when  $i$  and  $j$  are disjoint.

We require that  $\mathcal{S}$  is sufficiently symmetric, so that if  $\mathbf{V}$  is drawn uniformly at random from  $\mathcal{S}$ , then the probability  $\mathbb{P}\{V_i = 1, V_j = 1\}$  can take at most three different values depending on whether  $i = j$ ,  $i \sim j$ , or  $i \not\sim j$ .

In such cases, if  $\mathbf{x} = (x_1, \dots, x_d)$  is any vector in  $\mathbb{R}^d$ , then

$$\begin{aligned} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] &= \sum_{i=1}^d \sum_{j=1}^d x_i x_j \mathbb{P}\{V_i = 1, V_j = 1\} \\ &= C_1 \sum_{i=1}^d x_i^2 + C_2 \sum_{i,j:i \sim j} x_i x_j + C_3 \sum_{i,j:i \not\sim j} x_i x_j \end{aligned} \quad (4)$$

where

$$\begin{aligned} C_1 &\stackrel{\text{def}}{=} \mathbb{P}\{V_i = 1\} && \forall i = 1, \dots, d \\ C_2 &\stackrel{\text{def}}{=} \mathbb{P}\{V_i = 1, V_j = 1\} && \forall i, j = 1, \dots, d \text{ s.t. } i \sim j \\ C_3 &\stackrel{\text{def}}{=} \mathbb{P}\{V_i = 1, V_j = 1\} && \forall i, j = 1, \dots, d \text{ s.t. } i \not\sim j \end{aligned}$$

are quantities independent of  $i, j$ .

This property is true for collections  $\mathcal{S}$  of "symmetric" subsets of  $K_m$ , such as spanning trees, balanced cuts, planar graphs, Hamiltonian cycles, cliques

of a certain size, etc. The following result provides a general lower bound for the smallest eigenvalue of the associated matrix  $M$ .

**Lemma 5.** *If (4) holds and  $\mathbf{x} \in \mathbb{R}^d$  has unit norm, then*

$$\mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] \geq C_1 - C_3 - |C_2 - C_3| m - \frac{(C_2 - C_3)^2}{C_3}.$$

PROOF. Since  $\|\mathbf{x}\| = 1$ , we have

$$\begin{aligned} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] &= C_1 + C_2 \sum_{i,j:i \sim j} x_i x_j + C_3 \sum_{i,j:i \not\sim j} x_i x_j \\ &= C_1 - C_3 + (C_2 - C_3) \sum_{i,j:i \sim j} x_i x_j + C_3 \sum_{i,j=1}^d x_i x_j. \end{aligned}$$

Denote the summation over all pairs of adjacent edges by

$$A_m = \sum_{i,j:i \sim j} x_i x_j \quad \text{and let} \quad B_m = \left( \sum_{i=1}^d x_i \right)^2.$$

With this notation, we have

$$\mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] = C_1 - C_3 + (C_2 - C_3) A_m + C_3 B_m. \quad (5)$$

Next we need an appropriate estimate for  $A_m$ . By the Cauchy-Schwarz inequality, and using the fact that  $\|\mathbf{x}\| = 1$ ,

$$\begin{aligned} |A_m| &= \left| \sum_{i=1}^d x_i \sum_{j:i \sim j} x_j \right| \\ &\leq \sqrt{\sum_{i=1}^d \left( \sum_{j:i \sim j} x_j \right)^2} \\ &= \sqrt{\sum_{i=1}^d \left( \sum_{j,l:j \sim i, l \sim i} x_j x_l \right)} \\ &= \sqrt{(m-2) \sum_{i,j:i \sim j} x_i x_j + 4 \sum_{i,j:i \not\sim j} x_i x_j}. \end{aligned} \quad (6)$$

The last equality holds because a pair of edges is counted  $m - 2$  times if they are adjacent ( $m - 2$  is the number of edges adjacent to both) and 4 times if they are not adjacent. We may write the argument of the square root in (6) as

$$\begin{aligned} (m - 2) \sum_{i,j:i \sim j} x_i x_j + 4 \sum_{i,j:i \not\sim j} x_i x_j &= (m - 6) \sum_{i,j:i \sim j} x_i x_j + 4 \sum_{i,j} x_i x_j - 4 \\ &\leq m |A_m| + 4 B_m . \end{aligned} \tag{7}$$

Thus, substituting (7) in (6), and using  $B_m \geq 0$ , we get

$$|A_m| \leq \sqrt{m |A_m| + 4 B_m} .$$

Solving the above for  $|A_m|$  and overapproximating gives

$$|A_m| \leq m + 2\sqrt{B_m}$$

which, substituted into (5) yields

$$\mathbb{E}[(\mathbf{V}^\top \mathbf{x})^2] \geq C_1 - C_3 - |C_2 - C_3| \left( m + 2\sqrt{B_m} \right) + C_3 B_m .$$

Observing that

$$\begin{aligned} C_3 B_m - 2|C_2 - C_3| \sqrt{B_m} &= \left( \sqrt{C_3 B_m} - \frac{|C_2 - C_3|}{\sqrt{C_3}} \right)^2 - \frac{(C_2 - C_3)^2}{C_3} \\ &\geq -\frac{(C_2 - C_3)^2}{C_3} \end{aligned}$$

concludes the proof.

Interestingly, the proof above does not use that fact that  $\mathbf{x}$  is in the space spanned by the incidence vectors of  $\mathcal{S}$ . Thus, the matrix  $\mathbb{E}[\mathbf{V} \mathbf{V}^\top]$  is positive definite whenever the lower bound of Lemma 5 is positive. This also implies that the matrix  $P_t$ , which is used to define the pseudo-losses (1), is positive definite, and thus  $P_t^+$  can be replaced by  $P_t^{-1}$ .

Now we may use Lemma 5 to bound  $\lambda_{\min}$  in the case of spanning trees of the complete graph  $K_m$ . All we need is to calculate the values of  $C_1, C_2$ , and  $C_3$ . We do it by applying the theory of electric networks.

**Lemma 6.** *If  $\mathbf{V}$  is the incidence vector of a uniform random spanning tree of  $K_m$ , then*

$$\begin{aligned}\mathbb{P}\{V_i = 1\} &= \frac{2}{m} \\ \mathbb{P}\{V_i = 1, V_j = 1\} &= \frac{3}{m^2} \quad \text{if } i \sim j \\ \mathbb{P}\{V_i = 1, V_j = 1\} &= \frac{4}{m^2} \quad \text{if } i \not\sim j.\end{aligned}$$

PROOF. Since every spanning tree has  $m - 1$  edges,

$$\mathbb{P}\{V_1 = 1\} + \cdots + \mathbb{P}\{V_d = 1\} = m - 1$$

where  $d = \binom{m}{2}$ . By symmetry,  $\mathbb{P}\{V_i = 1\} = 2/m$  for all  $i = 1, \dots, d$ . The other two cases can be handled by the ‘‘Transfer Current’’ theorem of Burton and Pemantle [17], see also Lyons and Peres [18], which implies that for any  $i \neq j$ ,

$$\mathbb{P}\{V_i = 1, V_j = 1\} = \frac{4}{m^2} - Y(i, j)^2$$

where  $Y(i, j)$  is the voltage difference across the edge  $j$  when a unit current is imposed between the endpoints of edge  $i$ . (For the basic notions of electric networks we refer, e.g., to the books of Doyle and Snell [19] and Lyons and Peres [18].)

First note that if  $i$  and  $j$  are not adjacent then  $Y(i, j) = 0$ . This holds because, by symmetry, every vertex not belonging to edge  $i$  has the same voltage, so there is no current flowing through edge  $j$ . Thus,  $\mathbb{P}\{V_i = 1, V_j = 1\} = 4/m^2$  in this case.

In order to address the case when edges  $i$  and  $j$  are adjacent,  $i \sim j$ , note that, by a result of Kirchoff (1847), the voltage difference between the endpoints of  $i$  equals the probability  $2/m$  that  $i$  belongs to a random spanning tree (see, e.g., the remark to Corollary 4.4 in [18]). By the above considerations, there is current flow only along paths of length two between the endpoints of  $i$ , that is paths that go through edges  $j \sim i$ . Hence the voltage difference between the endpoints of  $j$  is half the voltage difference between the endpoints of  $i$ , that is  $|Y(i, j)| = 1/m$ .

**Corollary 7.** *For the spanning trees of  $K_m$ ,*

$$\lambda_{\min} \geq \frac{1}{m} - \frac{17}{4m^2}.$$

Since  $d = \binom{m}{2}$  and  $B = \sqrt{m-1}$ , the inequality above implies that  $B^2/(d\lambda_{\min}) < 7$  whenever  $m \geq 6$ , and therefore the optimal bound (2) holds. Since  $N = m^{m-2}$ , the performance bound of COMBAND in this case implies

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 4m^{3/2} \sqrt{2n \ln m} \quad \text{for } m \geq 6.$$

Finding computationally efficient algorithms for generating random spanning trees has been an intensive area of research. Although some of these algorithms may be successfully used in practical implementations, we are not aware of any algorithm that guarantees an efficient implementation of COMBAND under all circumstances. Instead of surveying the vast literature, we mention the celebrated method of Propp and Wilson [21], who present an algorithm that, given a graph with non-negative weights  $w_{(i,j)}$  over the edges, samples a random spanning tree from a distribution such that the probability of any spanning tree  $k$  is proportional to  $\bar{w}_t(k) = \prod_{(i,j) \in k} w_{(i,j)}$ . The expected running time of the algorithm is bounded by the cover time of an associated Markov chain that is defined as a random walk over the graph in which the transition probabilities are proportional to the edge weights. If we apply Propp and Wilson's algorithm with weights  $w_{(i,j)} = \exp(-\eta \widetilde{L}_{t,(i,j)})$  over the complete graph  $K_m$ , then we obtain an implementation of the exponentially weighted average forecaster. Unfortunately, there is no guarantee that the cover time is bounded by a polynomial of  $m$ , though in practice we expect a fast running time in most cases. It is an interesting open problem to find an efficient sampling algorithm for all possible assignments of weights.

### 5.5. Cut sets

In this section we consider balanced cuts of the complete graph  $K_{2m}$ . A balanced cut is the collection of all edges between a set of  $m$  vertices and its complement. Thus, each balanced cut has  $m^2$  edges and there are  $N = \binom{2m}{m}$  balanced cuts.

Our starting point in estimating  $\lambda_{\min}$  is (5). First, we compute  $C_1, C_2$ , and  $C_3$ .

**Lemma 8.** *If  $\mathbf{V}$  is the incidence vector of a uniform random  $m$ -cut in  $K_{2m}$ ,*

then

$$\begin{aligned}\mathbb{P}\{V_i = 1\} &= \frac{m}{2m-1} \\ \mathbb{P}\{V_i = 1, V_j = 1\} &= \frac{m(m-1)}{(2m-1)(2m-2)} \quad \text{if } i \sim j \\ \mathbb{P}\{V_i = 1, V_j = 1\} &= \frac{2m(m-1)^2}{(2m-1)(2m-2)(2m-3)} \quad \text{if } i \not\sim j.\end{aligned}$$

PROOF. The sample space is all choices of  $m$ -subsets of  $2m$  vertices (note that each  $m$ -cut is counted twice). Fix an edge  $i = (i_-, i_+)$ . Then the number of  $m$ -subsets that contain  $i_-$  and do not contain  $i_+$  is clearly  $\binom{2m-2}{m-1}$ . By symmetry, this is also the number of  $m$ -subsets that contain  $i_+$  and do not contain  $i_-$ . Therefore

$$\mathbb{P}\{V_i = 1\} = 2 \times \frac{\binom{2m-2}{m-1}}{\binom{2m}{m}} = \frac{m}{2m-1}.$$

Now fix two edges  $i$  and  $j$  that share a vertex, say  $i_- = j_-$ . The number of  $m$ -subsets that contain  $i_- = j_-$  and do not contain neither  $i_+$  nor  $j_+$  is  $\binom{2m-3}{m-1}$ . This is the same as the number of  $m$ -subsets that do not contain  $i_- = j_-$  and contain both  $i_+$  and  $j_+$ . Hence, if  $i \sim j$ ,

$$\mathbb{P}\{V_i = 1, V_j = 1\} = 2 \times \frac{\binom{2m-3}{m-1}}{\binom{2m}{m}} = \frac{m(m-1)}{(2m-1)(2m-2)}.$$

Finally, fix two disjoint edges  $i$  and  $j$ . The number of  $m$ -subsets that contain  $i_+, j_+$  and do not contain neither  $i_-$  nor  $j_-$  is  $\binom{2m-4}{m-2}$ . By symmetry, this is also the number of  $m$ -subsets that contain  $i_-, j_-$  and do not contain neither  $i_+$  nor  $j_+$ , which is the same as the number of those that contain  $i_-, j_+$  and not  $i_+$  or  $j_-$ , etc. Hence, for  $i \not\sim j$ ,

$$\mathbb{P}\{V_i = 1, V_j = 1\} = 4 \times \frac{\binom{2m-4}{m-2}}{\binom{2m}{m}} = \frac{2m(m-1)^2}{(2m-1)(2m-2)(2m-3)}$$

concluding the proof.

Now we may make use of the fact that each balanced cut has the same number of edges. Thus, if  $\mathbf{x} = \sum_{k=1}^{\binom{2m}{m}} a(k) \mathbf{v}(k)$  is a linear combination of the

incidence vectors of all balanced cuts with  $\|\mathbf{x}\| = 1$ , we have  $\sum_i x_i = m^2\alpha$  where  $\alpha = \sum_{k=1}^{\binom{2m}{m}} a(k)$ , which implies that  $B_m = m^4\alpha^2$ .

To compute  $A_m$ , observe that for any fixed  $i$ , the number of edges in any balanced cut adjacent to  $i$  is  $2m$  if the cut doesn't contain  $i$  and  $2(m-1)$  otherwise, that is,

$$\sum_{j:j\sim i} v_i(k) = \begin{cases} 2(m-1) & \text{if } v_i(k) = 1 \\ 2m & \text{if } v_i(k) = 0 \end{cases}$$

so

$$\begin{aligned} \sum_{j:j\sim i} x_j &= \sum_{k=1}^N a(k) \sum_{j:j\sim i} v_i(k) = \sum_{k=1}^N a(k) (2m - 2v_i(k)) \\ &= 2m\alpha - 2 \sum_{k=1}^N a(k)v_i(k) = 2m\alpha - 2x_i . \end{aligned}$$

Therefore, we have

$$A_m = \sum_{i,j:i\sim j} x_i x_j = \sum_i x_i \sum_{j:j\sim i} x_j = m^3\alpha^2 - 2 .$$

Substituting these values in (5), we have, for  $m \geq 2$ ,

$$\begin{aligned} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] &= \frac{1}{4} + \frac{8m-7}{4(2m-1)(2m-3)} \\ &\quad + \alpha^2 \frac{m^4(m-1)(2m^2-2m-1)}{(2m-1)(2m-2)(2m-3)} . \end{aligned}$$

The minimum is achieved for  $\alpha = 0$ , which proves the following.

**Proposition 9.** *For the balanced cuts in  $K_{2m}$ , if  $m \geq 2$  then*

$$\lambda_{\min} = \frac{1}{4} + \frac{8m-7}{4(2m-1)(2m-3)} .$$

In this case we have  $d = \binom{2m}{2}$ ,  $B = m$ , and  $N = \binom{2m}{m} \leq 4^m$ . By Proposition 9 we clearly have  $B^2/(d\lambda_{\min}) \leq 2$  for all  $m \geq 2$ , and therefore the optimal bound (2) applies and it takes the form

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 2 m^{3/2} \sqrt{10n \ln 4}$$

which holds whenever all edge losses are between 0 and  $1/m^2$  (and therefore  $|\ell_t^\top \mathbf{v}| \leq 1$ ). In this case computationally efficient implementations also exist. Such an implementation may be based on an algorithm of Randall and Wilson [22] who, building on Jerrum and Sinclair [23], show how to sample efficiently spin configurations of a ferromagnetic Ising model. The straightforward details are omitted.

### 5.6. Hamiltonian cycles

In our next example we consider the set  $\mathcal{S}$  of all Hamiltonian cycles in  $K_m$ , that is all  $N = (m-1)!/2$  cycles that visit each vertex exactly once and returns to the starting vertex. The corresponding randomized prediction problem may be thought of as an online version of the traveling salesman problem. This problem is computationally notoriously difficult and one cannot expect polynomial-time implementations. Nevertheless, we show that small regret bounds are achievable by COMBAND. To this end, we calculate  $\lambda_{\min}$ .

**Proposition 10.** *If  $m \geq 4$ , then for the class of all Hamiltonian cycles in  $K_m$   $\lambda_{\min} = 2/(m-1)$ .*

Since  $d = \binom{m}{2}$ ,  $N = (m-1)!/2$ , and  $B = \sqrt{m}$ , we have  $B^2/(d\lambda_{\min}) = 1$ . Thus the optimal bound (2) applies achieving

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 2m \sqrt{\frac{3}{2} n \ln(m!)} .$$

PROOF. Once again, our analysis is based on (5). First we calculate the values of the constants  $C_1, C_2, C_3$ . Since each Hamiltonian cycle has  $m$  edges, if  $\mathbf{V}$  is a random Hamiltonian cycle, then  $C_1 = \mathbb{P}\{V_i = 1\} = 2/(m-1)$ . Also, since the degree of every vertex in a Hamiltonian cycle is 2, for any two adjacent edges  $i \sim j$ ,  $C_2 = \mathbb{P}\{V_i = 1, V_j = 1\} = 1/\binom{m-1}{2}$ . On the other hand, if  $i \not\sim j$ , then

$$\begin{aligned} \mathbb{P}\{V_i = 1, V_j = 1\} &= \mathbb{P}\{V_i = 1\} \mathbb{P}\{V_j = 1 | V_i = 1\} \\ &= \frac{2}{m-1} \times \frac{m-3}{\binom{m}{2} - 2(m-2) - 1} \end{aligned}$$

because there are  $\binom{m}{2} - 2(m-2) - 1$  edges in  $K_m$  that are not adjacent to  $i$  and all of them are equally likely to be any of the remaining  $m-3$  edges of the cycle  $\mathbf{V}$ . Thus,  $C_3 = 4/(m-1)(m-2)$ .

Now let  $\mathbf{x} = \sum_{k=1}^N a(k)\mathbf{v}(k)$  be a linear combination of the incidence vectors of all Hamiltonian cycles such that  $\|\mathbf{x}\| = 1$ . The crucial observation is the following: since every  $\mathbf{v}(k)$  has  $m$  edges, and the degree of every vertex equals 2, we have

$$\sum_i x_i = \sum_{k=1}^N a(k) \sum_i v_i(k) = m\alpha .$$

This implies that

$$B_m = \left( \sum_{i=1}^d x_i \right)^2 = m^2\alpha^2 .$$

Observe that for any fixed  $i$ , the number of edges in any Hamiltonian cycle adjacent to  $i$  is 4 if the cycle doesn't contain  $i$  and 2 otherwise, that is,

$$\sum_{j:j\sim i} v_i(k) = \begin{cases} 2 & \text{if } v_i(k) = 1 \\ 4 & \text{if } v_i(k) = 0 \end{cases}$$

Thus,

$$\begin{aligned} \sum_{j:j\sim i} x_j &= \sum_{k=1}^N a(k) \sum_{j:j\sim i} v_i(k) = \sum_{k=1}^N a(k) (4 - 2v_i(k)) \\ &= 4\alpha - 2 \sum_{k=1}^N a(k)v_i(k) = 4\alpha - 2x_i . \end{aligned}$$

Using this, we have

$$A_m = \sum_i x_i \sum_{j:j\sim i} x_j = \sum_i x_i (4\alpha - 2x_i) = 4m\alpha^2 - 2 \sum_i x_i^2 = 4m\alpha^2 - 2 .$$

Substituting these values in (5), we have

$$\begin{aligned} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] &= \frac{2(m-4)}{(m-1)(m-2)} + \frac{2(2m^2\alpha^2 - 4m\alpha^2 + 2)}{(m-1)(m-2)} \\ &= \frac{2}{m-1} + \frac{4m\alpha^2}{m-1} \geq \frac{2}{m-1} . \end{aligned}$$

with equality achieved for  $\sum_k a(k) = 0$ .

### 5.7. Stars

Here we consider a problem related to that of Section 5.4. Suppose that in a fully connected communication network, the decision maker wishes to select a “central” node such that the sum of the losses associated to all edges adjacent to the node is minimal. This leads us to considering the class of all *stars*. A star is a subgraph of  $K_m$  which contains all  $m - 1$  edges incident on a fixed vertex. Thus, there are  $m$  different stars in  $K_m$ . Consider the set  $\mathcal{S}$  of all stars and let  $\mathbf{V}$  be the incidence vector of a random star, chosen uniformly.

**Proposition 11.** *For the stars in  $K_m$ ,*

$$\lambda_{\min} = \frac{m - 3}{2(m - 2)} + \frac{1}{m} .$$

Here  $d = \binom{m}{2}$ ,  $N = m$ , and  $B = \sqrt{m - 1}$ . Thus we have  $B^2/(d\lambda_{\min}) \leq \frac{1}{2}$  and the optimal bound (2) applies with

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] \leq 2m\sqrt{n \ln m} .$$

The implementation of COMBAND is trivially efficient in this case.

PROOF. Clearly,  $\mathbb{P}\{V_i = 1\} = 2/m$ ,  $\mathbb{P}\{V_i = 1, V_j = 1\} = 1/m$  if  $i \sim j$  and  $\mathbb{P}\{V_i = 1, V_j = 1\} = 0$  if  $i \not\sim j$ . Therefore,

$$\mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] = \frac{2}{m} + \frac{A_m}{m}$$

where  $A_m = \sum_{i,j:i \sim j} x_i x_j$ . Let  $\mathbf{x} = \sum_{k=1}^m a_k \mathbf{v}_k$  be such that  $\|\mathbf{x}\| = 1$ . This means that

$$1 = \sum_{i=1}^d \left( \sum_{k=1}^m a_k v_i^{(k)} \right)^2 = \sum_{k=1}^m \sum_{k'=1}^m a_k a_{k'} \sum_{i=1}^d v_i^{(k)} v_i^{(k')} .$$

Since

$$\sum_{i=1}^d v_i^{(k)} v_i^{(k')} = \begin{cases} 1 & \text{if } k \neq k' \\ m - 1 & \text{if } k = k', \end{cases}$$

we have

$$(m - 2) \sum_{k=1}^m a_k^2 + \left( \sum_{k=1}^m a_k \right)^2 = 1 . \tag{8}$$

Now

$$A_m = \sum_{i,j:i \sim j} \left( \sum_{k=1}^m a_k v_i^{(k)} \right) \left( \sum_{k=1}^m a_k v_j^{(k)} \right) = \sum_{k,k'=1}^m a_k a_{k'} \left( \sum_{i,j:i \sim j} v_i^{(k)} v_j^{(k')} \right).$$

Observe that

$$\sum_{i,j:i \sim j} v_i^{(k)} v_j^{(k')} = \begin{cases} 2(m-1) & \text{if } k \neq k' \\ \binom{m-1}{2} & \text{if } k = k'. \end{cases}$$

So

$$A_m = \left( \binom{m-1}{2} - 1 \right) \sum_{k=1}^m a_k^2 + 2(m-1) \left( \sum_{k=1}^m a_k \right)^2.$$

Expressing  $\sum_{k=1}^m a_k^2$  from (8), and substituting in the expression above, we obtain

$$A_m = \frac{m(m-3)}{2(m-2)} + \left( \sum_{k=1}^m a_k \right)^2 \left( 2(m-1) - \frac{m(m-3)}{2(m-2)} \right) \geq \frac{m(m-3)}{2(m-2)}.$$

In conclusion,

$$\lambda_{\min} \geq \frac{2}{m} + \frac{m-3}{2(m-2)}$$

with equality for  $\sum_k a_k = 0$ .

### 5.8. $m$ -sized subsets

Consider  $\mathcal{S}$  to be the set of all  $\mathbf{v} \in \{0,1\}^d$  such that  $\sum_{i=1}^d v_i = m$  for some fixed  $m$  with  $1 \leq m < d$ .

**Proposition 12.** *For the  $m$ -sized subsets,*

$$\lambda_{\min} = \frac{m(d-m)}{d(d-1)}.$$

We have  $B = \sqrt{m}$ ,  $N = \binom{d}{m}$ . Then

$$\frac{B^2}{d\lambda_{\min}} = \frac{d-1}{d-m}.$$

Thus the optimal bound (2) applies whenever  $m = o(d)$ . In this case the regret bound has the form

$$\mathbb{E} \left[ \widehat{L}_n - L_n(k) \right] = \mathcal{O} \left( \sqrt{nm d \ln d} \right).$$

Note that also in this case COMBAND can be implemented efficiently using dynamic programming (see, e.g., Takimoto and Warmuth [12]).

PROOF. Pick  $\mathbf{x} \in U$  such that  $\|\mathbf{x}\| = 1$ . Note that

$$\sum_{i=1}^d x_i = \sum_{k=1}^N a(k) \sum_{i=1}^d v_i(k) = m \sum_{k=1}^N a(k) = m \alpha .$$

Since for any  $i$ ,

$$\mathbb{P}\{V_i = 1\} = \frac{\binom{d-1}{m-1}}{\binom{d}{m}} = \frac{m}{d}$$

and for any  $i \neq j$

$$\mathbb{P}\{V_i = 1, V_j = 1\} = \frac{\binom{d-2}{m-2}}{\binom{d}{m}} = \frac{m(m-1)}{d(d-1)}$$

we can write

$$\begin{aligned} \mathbb{E} \left[ (\mathbf{V}^\top \mathbf{x})^2 \right] &= \sum_{i=1}^d \sum_{j=1}^d x_i x_j \mathbb{P}\{V_i = 1, V_j = 1\} \\ &= \frac{m}{d} \sum_{i=1}^d x_i^2 + \frac{m(m-1)}{d(d-1)} \sum_{i,j:i \neq j} x_i x_j \\ &= \left( \frac{m}{d} - \frac{m(m-1)}{d(d-1)} \right) \sum_{i=1}^d x_i^2 + \frac{m(m-1)}{d(d-1)} \sum_{i,j} x_i x_j \\ &= \left( \frac{m}{d} - \frac{m(m-1)}{d(d-1)} \right) + \frac{m(m-1)}{d(d-1)} m^2 \alpha^2 \\ &= \frac{m(d-m)}{d(d-1)} + \frac{m^3(m-1)}{d(d-1)} \alpha^2 \geq \frac{m(d-m)}{d(d-1)} \end{aligned}$$

with equality whenever  $\alpha = 0$ .

### 5.9. Path planning

The path planning problem, described in Section 3, is one of the most important motivating examples of the bandit linear optimization problem. As mentioned in the introduction, a regret of the order of  $\sqrt{nd \ln N}$  is achievable if the loss of each edge of the chosen path is revealed to the forecaster

(here  $d$  denotes the number of edges of the graph). If only the total loss of the selected path becomes known to the decision maker (as in the model considered in this paper), then the results of Dani, Hayes, and Kakade [7] imply a regret bound of the order of  $B\sqrt{nd\ln N}$  where  $B^2$  is the length of the longest path in the collection. We conjecture that this bound is sub-optimal. However, optimal sampling is a non-trivial issue in general. To see why uniform sampling does not work, consider the case when the graph is the  $m \times m$  square grid (i.e., the vertex set is identified with pairs of integers  $(i, j)$  with  $i, j \in \{1, \dots, m\}$  and vertices  $(i, j)$  and  $(i', j')$  are joined by an edge if and only if  $|i - i'| + |j - j'| = 1$ ) and the class  $\mathcal{S}$  of paths is the set of all monotone paths between vertex  $(1, 1)$  and  $(m, m)$  (there are  $\binom{2m-2}{m-1}$  of them, all of length  $2m - 2$ ). If  $\mu$  is uniform on  $\mathcal{S}$ , then the edges adjacent to vertices  $(1, m)$  and  $(m, 1)$  are in the sampled path with probability that is exponentially small in  $m$ . Thus, there is no chance to achieve a regret bound that depends only polynomially on the number of edges. (Just consider a sequence of loss vectors such that, for all  $t$ , all edge losses are  $1/(2m - 2)$  except for the ones adjacent to vertex  $(1, m)$  which are equal to zero.) Designing a general nearly optimal sampling distribution for the path planning problem is an interesting open problem.

## 6. Conclusions

In this work we have investigated the problem of bandit online linear optimization when the action set  $\mathcal{S}$  is a finite subset of  $\{0, 1\}^d$ , the action vectors  $\mathbf{v} \in \mathcal{S}$  satisfy  $\|\mathbf{v}\| \leq B$ , and the loss vectors  $\ell_t$  satisfy  $|\ell_t^\top \mathbf{v}| \leq 1$ . We introduced and analyzed a new randomized forecasting strategy, COMBAND, closely related to the GEOMETRICHEDGE algorithm of [7].

Although the regret of COMBAND can not be improved in general, in some interesting cases (like the path planning problem) COMBAND has a suboptimal performance because a uniform initial sampling distribution  $\mu$  causes the smallest nonzero eigenvalue  $\lambda_{\min}$  to get too small. In general,  $\mu$  can be chosen in order to maximize  $\lambda_{\min}$  by solving a semidefinite program. We conjecture that for the path planning problem this choice of  $\mu$  is polytime computable, and COMBAND, run with this  $\mu$ , has optimal regret  $\sqrt{nd\ln N}$ .

## Appendix A. Proof of Theorem 1

First we need some auxiliary results.

**Lemma 13.** *Let  $\mathbf{V}$  be a random vector whose distribution is finitely supported in  $\mathbb{R}^d$ . Let  $M = \mathbb{E}[\mathbf{V} \mathbf{V}^\top]$ . Then  $M M^+ \mathbf{v} = \mathbf{v}$  for all  $\mathbf{v} \in \mathbb{R}^d$  such that  $\mathbb{P}\{\mathbf{V} = \mathbf{v}\} > 0$ .*

PROOF. To prove the statement we show that for all  $\mathbf{x} \in \mathbb{R}^d$  such that  $M \mathbf{x} = \mathbf{0}$  and for all  $\mathbf{v} \in \mathbb{R}^d$  such that  $\mathbb{P}\{\mathbf{V} = \mathbf{v}\} > 0$ , it must be the case that  $\mathbf{x}^\top \mathbf{v} = 0$ . Pick any  $\mathbf{x} \in \mathbb{R}^d$  such that  $M \mathbf{x} = \mathbf{0}$ . This implies  $\mathbf{x}^\top M \mathbf{x} = 0$ . Using the definition of  $M$  we obtain  $0 = \mathbf{x}^\top M \mathbf{x} = \mathbb{E}[(\mathbf{x}^\top \mathbf{V})^2]$ . But then it must be the case that  $\mathbf{x}^\top \mathbf{v} = 0$  for all  $\mathbf{v}$  such that  $\mathbb{P}\{\mathbf{V} = \mathbf{v}\} > 0$ .

Let  $Q_t = \mathbb{E}[\mathbf{V} \mathbf{V}^\top]$  where  $\mathbf{V}$  has law  $q_t$ . Note that  $Q_t$  is always positive semidefinite since it is a convex combination of positive semidefinite matrices  $\mathbf{v}(k) \mathbf{v}(k)^\top$ .

**Corollary 14.**  *$P_t P_t^+ \mathbf{v} = \mathbf{v}$  for all  $t$  and all  $\mathbf{v}$  in the linear span of  $\mathcal{S}$ .*

PROOF. Since  $P_t = (1 - \gamma)Q_t + \gamma M$ , for all  $t$  and  $\mathbf{v}(k) \in \mathcal{S}$ ,  $p_t(k) > 0$ . Thus, Lemma 13 implies the result.

**Lemma 15.** *Let  $\mathbf{V}$  be a random element of  $\mathbb{R}^d$  and let  $P = \mathbb{E}[\mathbf{V} \mathbf{V}^\top]$ . Then  $\mathbb{E}[\mathbf{V}^\top P^+ \mathbf{V}] = \text{rank}(P)$ .*

PROOF. By the spectral theorem,

$$P = \sum_{i=1}^d \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$$

where  $\lambda_i \geq 0$  and  $\mathbf{u}_1, \dots, \mathbf{u}_d$  is an orthonormal basis. Then, for any  $\mathbf{v} \in \mathbb{R}^d$ ,

$$\mathbf{v}^\top P^+ \mathbf{v} = \sum_{i: \lambda_i > 0} \mathbf{v}^\top \frac{\mathbf{u}_i \mathbf{u}_i^\top}{\lambda_i} \mathbf{v} = \sum_{i: \lambda_i > 0} \frac{1}{\lambda_i} \mathbf{u}_i^\top \mathbf{v} \mathbf{v}^\top \mathbf{u}_i.$$

This implies

$$\begin{aligned} \mathbb{E}[\mathbf{V}^\top P^+ \mathbf{V}] &= \sum_{i: \lambda_i > 0} \frac{1}{\lambda_i} \mathbf{u}_i^\top \mathbb{E}[\mathbf{V} \mathbf{V}^\top] \mathbf{u}_i = \sum_{i, j: \lambda_i, \lambda_j > 0} \frac{\lambda_j}{\lambda_i} \mathbf{u}_i^\top \mathbf{u}_j \mathbf{u}_j^\top \mathbf{u}_i \\ &= \sum_{i: \lambda_i > 0} (\mathbf{u}_i^\top \mathbf{u}_i)^2 = \text{rank}(P). \end{aligned}$$

PROOF OF THEOREM 1. Let  $\mathbb{E}_t$  be the expectation operator conditioned on the first  $t - 1$  random draws  $K_1, \dots, K_{t-1}$  (i.e., expectation with respect to the distribution  $p_{t-1}$ ). Recall that  $c_t(k) = \boldsymbol{\ell}_t^\top \mathbf{v}(k)$  for  $k = 1, \dots, N$ , so  $\mathbb{E}_t c_t(K_t) \mathbf{v}(K_t) = P_{t-1} \boldsymbol{\ell}_t$ . Since  $\tilde{c}_t(k) = \tilde{\boldsymbol{\ell}}_t^\top \mathbf{v}(k)$ , and since Corollary 14 gives  $\mathbb{E}_t \tilde{\boldsymbol{\ell}}_t = \boldsymbol{\ell}_t^1$  where  $\boldsymbol{\ell}_t^1$  is the orthogonal projection of  $\boldsymbol{\ell}_t$  to the linear space spanned by  $\mathcal{S}$ , we obtain  $\mathbb{E}_t \tilde{c}_t(k) = c_t(k)$  for all  $k = 1, \dots, N$ .

For each  $k \in \{1, \dots, N\}$  define the cumulative pseudo-loss  $\tilde{L}_n(k) = \tilde{c}_1(k) + \dots + \tilde{c}_n(k)$ . Since for every  $k^* \in \{1, \dots, N\}$ ,  $\overline{W}_n = \sum_{k=1}^N \overline{w}_n(k) \geq \overline{w}_n(k^*) = \exp(-\eta \tilde{L}_n(k^*))$ , we have

$$\ln \frac{\overline{W}_n}{\overline{W}_0} \geq -\eta \tilde{L}_n(k^*) - \ln N. \quad (\text{A.1})$$

On the other hand, assuming that  $\eta |\tilde{c}_t(k)| \leq 1$  for all  $t$  and  $k$  (this condition will be verified later), and using  $e^x \leq 1 + x + x^2$  for  $|x| \leq 1$  and  $\ln(1 + y) \leq y$  for  $y > -1$  gives

$$\begin{aligned} \ln \frac{\overline{W}_t}{\overline{W}_{t-1}} &= \ln \sum_{k=1}^N \frac{p_{t-1}(k) - \gamma \mu(k)}{1 - \gamma} \exp(-\eta \tilde{c}_t(k)) \\ &\leq \ln \sum_{k=1}^N \frac{p_{t-1}(k) - \gamma \mu(k)}{1 - \gamma} \left(1 - \eta \tilde{c}_t(k) + \eta^2 \tilde{c}_t(k)^2\right) \\ &\leq -\frac{\eta}{1 - \gamma} \sum_{k=1}^N p_{t-1}(k) \tilde{c}_t(k) + \frac{\eta \gamma}{1 - \gamma} \sum_{k=1}^N \tilde{c}_t(k) \mu(k) \\ &\quad + \frac{\eta^2}{1 - \gamma} \sum_{k=1}^N p_{t-1}(k) \tilde{c}_t(k)^2. \end{aligned} \quad (\text{A.2})$$

The last term on the right-hand side can be written as follows

$$\begin{aligned}
\sum_{k=1}^N p_{t-1}(k) \tilde{c}_t(k)^2 &= \sum_{k=1}^N p_{t-1}(k) \left( \sum_{i=1}^d v_i(k) \tilde{\ell}_{t,i} \right) \left( \sum_{j=1}^d v_j(k) \tilde{\ell}_{t,j} \right) \\
&= \sum_{k=1}^N p_{t-1}(k) \left( \sum_{i,j=1}^d v_i(k) v_j(k) \tilde{\ell}_{t,i} \tilde{\ell}_{t,j} \right) \\
&= \sum_{i,j=1}^d \tilde{\ell}_{t,i} \tilde{\ell}_{t,j} \left( \sum_{k=1}^N v_i(k) v_j(k) p_{t-1}(k) \right) \\
&= \sum_{i,j=1}^d \tilde{\ell}_{t,i} P_{t-1}(i,j) \tilde{\ell}_{t,j} \\
&= \tilde{\boldsymbol{\ell}}_t^\top P_{t-1} \tilde{\boldsymbol{\ell}}_t \\
&= c_t(K_t) \mathbf{v}(K_t)^\top P_{t-1}^+ P_{t-1} P_{t-1}^+ \mathbf{v}(K_t) c_t(K_t) \\
&\leq \mathbf{v}(K_t)^\top P_{t-1}^+ \mathbf{v}(K_t)
\end{aligned}$$

where we used the assumption  $|c(K_t)| \leq 1$ . Summing for  $t = 1, \dots, n$  both sides of the inequality (A.2) gives

$$\begin{aligned}
\ln \frac{\overline{W}_n}{\overline{W}_0} &\leq -\frac{\eta}{1-\gamma} \sum_{t=1}^n \sum_{k=1}^N p_{t-1}(k) \tilde{c}_t(k) + \frac{\eta\gamma}{1-\gamma} \sum_{t=1}^n \sum_{k=1}^N \tilde{c}_t(k) \mu(k) \\
&\quad + \frac{\eta^2}{1-\gamma} \sum_{t=1}^n \mathbf{v}(K_t)^\top P_{t-1}^+ \mathbf{v}(K_t) .
\end{aligned}$$

Combining the above with (A.1), multiplying both sides by  $(1-\gamma)/\eta > 0$ , and using  $(1-\gamma)(\ln N)/\eta \leq (\ln N)/\eta$ , gives

$$\begin{aligned}
\sum_{t=1}^n \sum_{k=1}^N p_{t-1}(k) \tilde{c}_t(k) &\leq (1-\gamma) \tilde{L}_n(k^*) + \frac{\ln N}{\eta} + \gamma \sum_{t=1}^n \sum_{k=1}^N \tilde{c}_t(k) \mu(k) \\
&\quad + \eta \sum_{t=1}^n \mathbf{v}(K_t)^\top P_{t-1}^+ \mathbf{v}(K_t) . \tag{A.3}
\end{aligned}$$

We now take expectation on both sides and use  $\mathbb{E} \tilde{c}_t(k) = c_t(k)$  for all  $t$  and  $k$ . For the first and third term on the right-hand side this gives

$$\mathbb{E} \tilde{L}_n(k^*) = \mathbb{E} L_n(k^*) \quad \text{and} \quad \mathbb{E} \left[ \sum_{t=1}^n \sum_{k=1}^N \tilde{c}_t(k) \mu(k) \right] \leq n . \tag{A.4}$$

The expectation of the term on the left-hand side is

$$\begin{aligned}
\mathbb{E} \left[ \sum_{t=1}^n \sum_{k=1}^N p_{t-1}(k) \tilde{c}_t(k) \right] &= \mathbb{E} \left[ \sum_{t=1}^n \sum_{k=1}^N p_{t-1}(k) \mathbb{E}_t \tilde{c}_t(k) \right] \\
&= \mathbb{E} \left[ \sum_{t=1}^n \sum_{k=1}^N p_{t-1}(k) c_t(k) \right] \\
&= \mathbb{E} \left[ \sum_{t=1}^n \mathbb{E}_t c_t(K_t) \right] \\
&= \mathbb{E} \left[ \sum_{t=1}^n c_t(K_t) \right]. \tag{A.5}
\end{aligned}$$

Finally, we handle the expectation of the last term on the right-hand side of (A.3). Applying Lemma 15,

$$\mathbb{E}_t \left[ \mathbf{V}^\top P_{t-1}^+ \mathbf{V} \right] \leq d \tag{A.6}$$

where  $\mathbf{V}$  is distributed according to  $p_{t-1}$  and  $\mathbb{E}_t[\mathbf{V} \mathbf{V}^\top] = P_{t-1}$ . Substituting (A.4), (A.5), and (A.6) into (A.3) gives, for every  $k^* \in \{1, \dots, N\}$ ,

$$\begin{aligned}
\mathbb{E} \left[ \sum_{t=1}^n c_t(K_t) - L_n(k^*) \right] &\leq -\gamma \mathbb{E} L_n(k^*) + \frac{\ln N}{\eta} + \gamma n + d \eta n \\
&\leq \frac{\ln N}{\eta} + 2\gamma n + d \eta n \tag{A.7}
\end{aligned}$$

where we used  $|c_t(k^*)| \leq 1$  to bound  $-\mathbb{E} L_n(k^*) \leq n$ .

In order to enforce the condition  $\eta |\tilde{c}_t(k)| \leq 1$  we write

$$\begin{aligned}
|\tilde{c}_t(k)| &= |\mathbf{v}(k)^\top \tilde{\boldsymbol{\ell}}_t| \leq |c_t(K_t)| |\mathbf{v}(k)^\top P_{t-1}^+ \mathbf{v}(K_t)| \leq \|P_{t-1}^+\| \max_{\mathbf{v} \in \mathcal{S}} \|\mathbf{v}\|^2 \\
&\leq \frac{B^2}{\lambda_{\min}(P_{t-1})}
\end{aligned}$$

where  $\lambda_{\min}(P_{t-1})$  is the smallest nonzero eigenvalue of  $P_{t-1}$ , and we used, once more,  $|c_t(K_t)| \leq 1$  and  $\|\mathbf{v}\|^2 \leq B^2$ . Let  $\lambda_{\min} = \lambda_{\min}(M)$ . By Weyl's inequality,  $\lambda_{\min}(P_{t-1}) \geq \gamma \lambda_{\min}$ , which in turn implies that  $|\tilde{c}_t(k)| \leq B^2/(\gamma \lambda_{\min})$ .

Hence we choose  $\eta = \gamma \lambda_{\min}/B^2$  and (A.7) becomes

$$\mathbb{E} \left[ \sum_{t=1}^n c_t(K_t) - L_n(k) \right] \leq \frac{B^2 \ln N}{\gamma \lambda_{\min}} + \gamma \lambda_{\min} \left( \frac{d}{B^2} + \frac{2}{\lambda_{\min}} \right) n .$$

Letting

$$\gamma = \frac{B}{\lambda_{\min}} \sqrt{\frac{\ln N}{n \left( \frac{d}{B^2} + \frac{2}{\lambda_{\min}} \right)}}$$

finally yields

$$\mathbb{E} \left[ \sum_{t=1}^n c_t(K_t) - L_n(k) \right] \leq 2 \sqrt{\left( \frac{2B^2}{d \lambda_{\min}} + 1 \right) nd \ln N}$$

which ends the proof of Theorem 1.

## Acknowledgements

Thanks to Sham Kakade for enlightening discussions and to Sebastián Bubeck for pointing out a problem in a preliminary version of this paper.

- [1] N. Cesa-Bianchi, G. Lugosi, Prediction, Learning, and Games, Cambridge University Press, 2006.
- [2] P. Auer, N. Cesa-Bianchi, Y. Freund, R. Schapire, The nonstochastic multiarmed bandit problem, SIAM Journal on Computing 32 (1) (2002) 48–77.
- [3] B. Awerbuch, R. Kleinberg, Adaptive routing with end-to-end feedback: distributed learning and geometric approaches, in: Proceedings of the 36th ACM Symposium on the Theory of Computing, ACM Press, 2004.
- [4] H. McMahan, A. Blum, Online geometric optimization in the bandit setting against an adaptive adversary, in: Proceedings of the 17th Annual Conference on Learning Theory, Springer, 2004, pp. 109–123.
- [5] V. Dani, T. Hayes, Robbing the bandit: less regret in online geometric optimization against an adaptive adversary, in: Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms, ACM/SIAM, 2006, pp. 937–943.

- [6] A. György, T. Linder, G. Lugosi, G. Ottucsák, The on-line shortest path problem under partial monitoring, *Journal of Machine Learning Research* 8 (2007) 2369–2403.
- [7] V. Dani, T. Hayes, S. Kakade, The price of bandit information for online optimization, in: *Advances in Neural Information Processing Systems* 22, MIT Press, 2008, pp. 345–352.
- [8] J. Abernethy, E. Hazan, A. Rakhlin, Competing in the dark: An efficient algorithm for bandit linear optimization, in: *Proceedings of the 21st Annual Conference on Learning Theory*, Omnipress, 2008, pp. 263–274.
- [9] P. Bartlett, V. Dani, T. Hayes, S. Kakade, A. Rakhlin, A. Tewari, High-probability regret bounds for bandit online linear optimization, in: *Proceedings of the 21st Annual Conference on Learning Theory*, Omnipress, 2008, pp. 335–342.
- [10] J. Abernethy, A. Rakhlin, Beating the adaptive bandit with high probability, *Tech. Rep. UCB/EECS-2009-10*, University of California at Berkeley (2009).
- [11] J.-Y. Audibert, S. Bubeck, Minimax policies for adversarial and stochastic bandits, in: *Proceedings of the 22nd Annual Conference on Learning Theory*, Omnipress, 2009.
- [12] E. Takimoto, M. Warmuth, Path kernels and multiplicative updates, *Journal of Machine Learning Research* 4 (5) (2004) 773–818.
- [13] A. Kalai, S. Vempala, Efficient algorithms for online decision problems, *Journal of Computer and System Sciences* 71 (3) (2005) 291–307.
- [14] S. Boyd, L. Vanderberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [15] D. Helmbold, M. Warmuth, Learning permutations with exponential weights, in: *Proceedings of the 20th Annual Conference on Learning Theory*, Springer, 2007, pp. 469–483.
- [16] M. Jerrum, A. Sinclair, E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries, *Journal of the ACM* 51 (2004) 671–697.

- [17] R. Burton, R. Pemantle, Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-impedances, *The Annals of Probability* 21 (1993) 1329–1371.
- [18] R. Lyons, Y. Peres, Probability on trees and networks, manuscript (2008).
- [19] P. Doyle, J. Snell, Random walks and electric networks, Vol. 22 of *Carus Mathematical Monographs*, Mathematical Association of America, 1984.
- [20] W.M. Koolen, M.K. Warmuth, J. Kivinen, Hedging structured concepts, in: *Proceedings of 23rd Annual Conference on Learning Theory*, Omnipress, 2010, pp. 239–254.
- [21] J. Propp, D. Wilson, How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph, *Journal of Algorithm* 27 (1998) 170–217.
- [22] D. Randall, D. Wilson, Sampling spin configurations of an Ising system, in: *Proceedings of the 10th ACM-SIAM Symposium on Discrete Algorithms*, ACM/SIAM, 1999, pp. 959–960.
- [23] M. Jerrum, A. Sinclair, Polynomial-time approximation algorithms for the Ising model, *SIAM Journal on Computing* 22 (1993) 1087–1116.