Minimax Policies for Combinatorial Prediction Games

Jean-Yves Audibert
Imagine, Univ. Paris Est, and Sierra,
CNRS/ENS/INRIA, Paris, France
audibert@imagine.enpc.fr

Sébastien Bubeck
Centre de Recerca Matemàtica
Barcelona, Spain
sbubeck@crm.cat

Gábor Lugosi
ICREA and Pompeu Fabra University
Barcelona, Spain
lugosi@upf.es

Abstract

We address the online linear optimization problem when the actions of the forecaster are represented by binary vectors. Our goal is to understand the magnitude of the minimax regret for the worst possible set of actions. We study the problem under three different assumptions for the feedback: full information, and the partial information models of the so-called "semi-bandit", and "bandit" problems. We consider both L_∞-, and L_2-type of restrictions for the losses assigned by the adversary.

We formulate a general strategy using Bregman projections on top of a potential-based gradient descent, which generalizes the ones studied in the series of papers György et al. (2007), Dani et al. (2008), Abernethy et al. (2008), Cesa-Bianchi and Lugosi (2009), Helmbold and Warmuth (2009), Koolen et al. (2010), Uchiya et al. (2010), Kale et al. (2010) and Audibert and Bubeck (2010). We provide simple proofs that recover most of the previous results. We propose new upper bounds for the semi-bandit game. Moreover we derive lower bounds for all three feedback assumptions. With the only exception of the bandit game, the upper and lower bounds are tight, up to a constant factor. Finally, we answer a question asked by Koolen et al. (2010) by showing that the exponentially weighted average forecaster is suboptimal against L_∞ adversaries.

1 Introduction

In the sequential decision making problems considered in this paper, at each time instance $t = 1, \dots, n$, the forecaster chooses, possibly in a randomized way, an action from a given set $\mathcal{S}$ where $\mathcal{S}$ is a subset of the d-dimensional hypercube $\{0,1\}^d$. The action chosen by the forecaster at time t is denoted by $V_t = (V_{1,t}, \dots, V_{d,t}) \in \mathcal{S}$. Simultaneously to the forecaster, the adversary chooses a loss vector $\ell_t = (\ell_{1,t}, \dots, \ell_{d,t}) \in [0, +\infty)^d$ and the loss incurred by the forecaster is $\ell_t^T V_t$. The goal of the forecaster is to minimize the expected cumulative loss $\mathbb{E} \sum_{t=1}^n \ell_t^T V_t$ where the expectation is taken with respect to the forecaster's internal randomization. This problem is an instance of an "online linear optimization" problem[1], see, e.g., Awerbuch and Kleinberg (2004), McMahan and Blum (2004), Kalai and Vempala (2005), György et al. (2007), Dani et al. (2008), Abernethy et al. (2008), Cesa-Bianchi and Lugosi (2009), Helmbold and Warmuth (2009), Koolen et al. (2010), Uchiya et al. (2010) and Kale et al. (2010)

We consider three variants of the problem, distinguished by the type of information that becomes available to the forecaster at each time instance, after taking an action. (1) In the *full information game* the forecaster observes the entire loss vector ℓ_t; (2) in the *semi-bandit game* only those components $\ell_{i,t}$ of ℓ_t are observable for which $V_{i,t} = 1$; (3) in the *bandit game* only the total loss $\ell_t^T V_t$ becomes available to the forecaster.

We refer to these problems as *combinatorial prediction games*. All three prediction games are sketched in Figure 1. For all three games, we define the regret[2] of the forecaster as

$$\overline{R}_n = \mathbb{E} \sum_{t=1}^n \ell_t^T V_t - \min_{v \in \mathcal{S}} \mathbb{E} \sum_{t=1}^n \ell_t^T v.$$

[1] In online linear optimization problems, the action set is often not restricted to be a subset of $\{0,1\}^d$ but can be an arbitrary subset of $\mathbb{R}^d$. However, in the most interesting cases, actions are naturally represented by Boolean vectors and we restrict our attention to this case.

[2] For the full information game, one can directly upper bound the stronger notion of regret $\mathbb{E} \sum_{t=1}^n \ell_t^T V_t - \mathbb{E} \min_{v \in \mathcal{S}} \sum_{t=1}^n \ell_t^T v$ which is always larger than $\overline{R}_n$. However, for partial information games, this requires more work.

Parameters: set of actions $\mathcal{S} \subset \{0,1\}^d$; number of rounds $n \in \mathbb{N}$.

For each round $t = 1, 2, \ldots, n$;

(1) the forecaster chooses $V_t \in \mathcal{S}$ with the help of an external randomization;

(2) simultaneously the adversary selects a loss vector $\ell_t \in [0, +\infty)^d$ (without revealing it);

(3) the forecaster incurs the loss $\ell_t^T V_t$. He observes

- the loss vector ℓ_t in the full information game,
- the coordinates $\ell_{i,t}\mathbb{1}_{V_{i,t}=1}$ in the semi-bandit game,
- the instantaneous loss $\ell_t^T V_t$ in the bandit game.

Goal: The forecaster tries to minimize his cumulative loss $\sum_{t=1}^n \ell_t^T V_t$.

Figure 1: Combinatorial prediction games.

	L_∞			L_2		
	Full Info	Semi-Bandit	Bandit	Full Info	Semi-Bandit	Bandit
Lower Bound	$d\sqrt{n}$	$d\sqrt{n}$	$\mathbf{d^{3/2}\sqrt{n}}$	$\sqrt{dn}$	$\sqrt{dn}$	$d\sqrt{n}$
Upper Bound	$d\sqrt{n}$	$\mathbf{d\sqrt{n}}$	$d^{5/2}\sqrt{n}$	$\sqrt{dn}$	$\mathbf{\sqrt{dn\log d}}$	$d^{3/2}\sqrt{n}$

Table 1: Bounds on R_n proved in this paper (up to constant factor). The new results are set in bold.

In order to make meaningful statements about the regret, one needs to restrict the possible loss vectors the adversary may assign. We work with two different natural assumptions that have been considered in the literature:
L_∞ **assumption:** here we assume that $\|\ell_t\|_\infty \leq 1$ for all $t = 1, \ldots, n$
L_2 **assumption:** assume that $\ell_t^T v \leq 1$ for all $t = 1, \ldots, n$ and $v \in \mathcal{S}$.

Note that, without loss of generality, we may assume that for all $i \in \{1, \ldots, d\}$, there exists $v \in \mathcal{S}$ with $v_i = 1$, and then the L_2 assumption implies the L_∞ assumption.

The goal of this paper is to study the *minimax regret*, that is, the performance of the forecaster that minimizes the regret for the worst possible sequence of loss assignments. This, of course, depends on the set $\mathcal{S}$ of actions. Our aim is to determine the order of magnitude of the minimax regret for the most difficult set to learn. More precisely, for a given game, if we write sup for the supremum over all allowed adversaries (that is, either L_∞ or L_2 adversaries) and inf for the infimum over all forecaster strategies for this game, we are interested in the maximal minimax regret

$$R_n = \max_{\mathcal{S} \subset \{0,1\}^d} \inf \sup \overline{R}_n \, .$$

Note that in this paper we do not restrict our attention to computationally efficient algorithms. The following example illustrates the different games that we introduced above.

Example 1 *Consider the well studied example of* path planning *in which, at every time instance, the forecaster chooses a path from one fixed vertex to another in a graph. At each time, a loss is assigned to every edge of the graph and, depending on the model of the feedback, the forecaster observes either the losses of all edges, the losses of each edge on the chosen path, or only the total loss of the chosen path. The goal is to minimize the total loss for any sequence of loss assignments. This problem can be cast as a combinatorial prediction game in dimension* d *for* d *the number of edges in the graph.*

Our contribution is threefold. First, we propose a variant of the algorithm used to track the best linear predictor (Herbster and Warmuth, 1998) that is well-suited to our combinatorial prediction games. This leads to an algorithm called CLEB that generalizes various approaches that have been proposed. This new point of view on algorithms that were defined for specific games (only the full information game, or only the standard multi-armed bandit game) allows us to generalize them easily to all combinatorial prediction games, leading to new algorithms such as LINPOLY. This algorithmic contribution leads to our second main result, the improvement of the known upper bounds for the semi-bandit game. This point of view also leads to a different proof of the minimax $\sqrt{nd}$ regret bound in the standard d-armed bandit game that is much simpler than the one provided in Audibert and Bubeck (2010). A summary of the bounds proved in this paper can be

	L_∞			L_2		
	Full Info	Semi-Bandit	Bandit	Full Info	Semi-Bandit	Bandit
EXP2	$d^{3/2}\sqrt{n}$	$d^{3/2}\sqrt{n}$	$d^{5/2}\sqrt{n}$	$\sqrt{dn}$	$\mathbf{d\sqrt{n}}$ *	$d^{3/2}\sqrt{n}$
LINEXP	$d\sqrt{n}$	$\mathbf{d\sqrt{n}}$	$\mathbf{d^2n^{2/3}}$	$\mathbf{\sqrt{dn}}$	$\mathbf{d\sqrt{n}}$ *	$\mathbf{d^2n^{2/3}}$
LINPOLY	$\mathbf{d\sqrt{n}}$	$\mathbf{d\sqrt{n}}$	-	$\mathbf{\sqrt{dn}}$	$\mathbf{\sqrt{dn\log d}}$	-

Table 2: Upper bounds on $\overline{R}_n$ for specific forecasters. The new results are in bold. We also show that the bound for EXP2 in the full information game is unimprovable. Note that the bound for (Bandit, LINEXP) is very weak. The bounds with * become $\sqrt{dn\log d}$ if we restrict our attention to sets $\mathcal{S}$ that are "almost symmetric" in the sense that for some k, $\mathcal{S} \subset \left\{v \in \{0,1\}^d : \sum_{i=1}^d v_i \leq k\right\}$ and $\text{Conv}(\mathcal{S}) \cap \left[\frac{k}{2d}; 1\right]^d \neq \emptyset$.

found in Table 1 and Table 2. In addition we prove several lower bounds. First, we establish lower bounds on the minimax regret in all three games and under both types of adversaries, whereas only the cases (L_2/L_∞, Full Information) and (L_2, Bandit) were previously treated in the literature. Moreover we also answer a question of Koolen et al. (2010) by showing that the traditional exponentially weighted average forecaster is suboptimal against L_∞ adversaries.

In particular, this paper leads to the following (perhaps unexpected) conclusions:

- The full information game is as hard as the semi-bandit game. More precisely, in terms of R_n, the price that one pays for the limited feedback of the semi-bandit game compared to the full information game is only a constant factor (or a $\sqrt{\log d}$ factor for the L_2 setting).
- In the full information and semi-bandit game, the traditional exponentially weighted average forecaster is provably suboptimal for L_∞ adversaries while it is optimal for L_2 adversaries in the full information game.
- Denote by $\mathcal{A}_2$ (respectively $\mathcal{A}_\infty$) the set of adversaries that satisfy the L_2 assumption (respectively the L_∞ assumption). We clearly have $\mathcal{A}_2 \subset \mathcal{A}_\infty \subset d\mathcal{A}_2$. We prove that, in the full information game, R_n gains an additional factor of $\sqrt{d}$ at each inclusion. In the semi-bandit game, we show that the same statement remains true up to a logarithmic factor.

Notation. The convex hull of $\mathcal{S}$ is denoted $\text{Conv}(\mathcal{S})$.

2 Combinatorial learning with Bregman projections

In this section we introduce a general forecaster that we call CLEB (Combinatorial LEarning with Bregman projections). Every forecaster investigated in this paper is a special case of CLEB.

Let $\mathcal{D}$ be a convex subset of $\mathbb{R}^d$ with nonempty interior $\text{Int}(\mathcal{D})$ and boundary $\partial\mathcal{D}$.

Definition 1 *We call Legendre any function $F : \mathcal{D} \to \mathbb{R}$ such that*

(i) F is strictly convex and admits continuous first partial derivatives on Int($\mathcal{D}$)
(ii) For any $u \in \partial\mathcal{D}$, for any $v \in$ Int($\mathcal{D}$), we have

$$\lim_{s \to 0, s>0} (u-v)^T \nabla F\big((1-s)u + sv\big) = +\infty.$$

The Bregman divergence $D_F : \mathcal{D} \times \text{Int}(\mathcal{D})$ associated to a Legendre function F is defined by

$$D_F(u,v) = F(u) - F(v) - (u-v)^T \nabla F(v).$$

We consider the algorithm CLEB described in Figure 2. The basic idea is to use a potential-based gradient descent (1) followed by a projection (2) with respect to the Bregman divergence of the potential onto the convex hull of $\mathcal{S}$ to ensure that the resulting weight vector w_{t+1} can be viewed as $w_{t+1} = \mathbb{E}_{V \sim p_{t+1}} V$ for some distribution p_{t+1} on $\mathcal{S}$. The combination of Bregman projections with potential-based gradient descent was first used in Herbster and Warmuth (1998). Online learning with Bregman divergences without the projection step has a long history (see Section 11.11 of Cesa-Bianchi and Lugosi (2006)). As discussed below, CLEB may be viewed as a generalization of the forecasters LINEXP and INF.

The Legendre conjugate F^* of F is defined by $F^*(u) = \sup_{v \in \mathcal{D}} \left\{u^T v - F(v)\right\}$. The following theorem establishes the first step of all upper bounds for the regret of CLEB.

Theorem 2 CLEB *satisfies for any $u \in$ Conv($\mathcal{S}$) $\cap\, \mathcal{D}$,*

$$\sum_{t=1}^n \tilde{\ell}_t^T w_t - \sum_{t=1}^n \tilde{\ell}_t^T u \leq D_F(u, w_1) + \sum_{t=1}^n D_{F^*}(\nabla F(w_t) - \tilde{\ell}_t, \nabla F(w_t)). \tag{3}$$

Parameters:

- a Legendre function F defined on $\mathcal{D}$ with $\text{Conv}(\mathcal{S}) \cap \text{Int}(\mathcal{D}) \neq \emptyset$
- $w_1 \in \text{Conv}(\mathcal{S}) \cap \text{Int}(\mathcal{D})$

For each round $t = 1, 2, \ldots, n$;

(a) Let p_t be a distribution on the set $\mathcal{S}$ such that $w_t = \mathbb{E}_{V \sim p_t} V$.

(b) Draw a random action V_t according to the distribution p_t and observe

- the loss vector ℓ_t in the full information game,
- the coordinates $\ell_{i,t} \mathbb{1}_{V_{i,t}=1}$ in the semi-bandit game,
- the instantaneous loss $\ell_t^T V_t$ in the bandit game.

(c) Estimate the loss ℓ_t by $\tilde{\ell}_t$. For instance, one may take

- $\tilde{\ell}_t = \ell_t$ in the full information game,
- $\tilde{\ell}_{i,t} = \frac{\ell_{i,t}}{\sum_{v \in \mathcal{S}: v_i = 1} p_t(v)} V_{i,t}$ in the semi-bandit game,
- $\tilde{\ell}_t = P_t^+ V_t V_t^T \ell_t$, with $P_t = \mathbb{E}_{v \sim p_t}(vv^T)$ in the bandit game.

(d) Let $w'_{t+1} \in \text{Int}(\mathcal{D})$ satisfying

$$\nabla F(w'_{t+1}) = \nabla F(w_t) - \tilde{\ell}_t. \tag{1}$$

(e) Project the weight vector w'_{t+1} defined by (1) on the convex hull of $\mathcal{S}$:

$$w_{t+1} \in \operatorname*{argmin}_{w \in \text{Conv}(\mathcal{S}) \cap \text{Int}(\mathcal{D})} D_F(w, w'_{t+1}). \tag{2}$$

Figure 2: Combinatorial learning with Bregman projections (CLEB).

Proof: By applying the definition of the Bregman divergences (or equivalently using Lemma 11.1 of Cesa-Bianchi and Lugosi (2006)), we obtain

$$\begin{aligned}\tilde{\ell}_t^T w_t - \tilde{\ell}_t^T u &= (u - w_t)^T \big(\nabla F(w'_{t+1}) - \nabla F(w_t)\big) \\ &= D_F(u, w_t) + D_F(w_t, w'_{t+1}) - D_F(u, w'_{t+1}).\end{aligned}$$

By the Pythagorean theorem (Lemma 11.3 of Cesa-Bianchi and Lugosi (2006)), we have $D_F(u, w'_{t+1}) \geq D_F(u, w_{t+1}) + D_F(w_{t+1}, w'_{t+1})$, hence

$$\tilde{\ell}_t^T w_t - \tilde{\ell}_t^T u \leq D_F(u, w_t) + D_F(w_t, w'_{t+1}) - D_F(u, w_{t+1}) - D_F(w_{t+1}, w'_{t+1}).$$

Summing over t then gives

$$\sum_{t=1}^n \tilde{\ell}_t^T w_t - \sum_{t=1}^n \tilde{\ell}_t^T u \leq D_F(u, w_1) - D_F(u, w_{n+1}) + \sum_{t=1}^n \big(D_F(w_t, w'_{t+1}) - D_F(w_{t+1}, w'_{t+1})\big). \tag{4}$$

By the nonnegativity of the Bregman divergences, we get

$$\sum_{t=1}^n \tilde{\ell}_t^T w_t - \sum_{t=1}^n \tilde{\ell}_t^T u \leq D_F(u, w_1) + \sum_{t=1}^n D_F(w_t, w'_{t+1}).$$

From Proposition 11.1 of Cesa-Bianchi and Lugosi (2006), we have $D_F(w_t, w'_{t+1}) = D_{F^*}\big(\nabla F(w_t) - \tilde{\ell}_t, \nabla F(w_t)\big)$, which concludes the proof. ■

As we will see below, by the equality $\mathbb{E} \sum_{t=1}^n \tilde{\ell}_t^T V_t = \mathbb{E} \sum_{t=1}^n \tilde{\ell}_t^T w_t$, and provided that $\tilde{\ell}_t^T V_t$ and $\tilde{\ell}_t^T u$ are unbiased estimates of $\mathbb{E}\ell_t^T V_t$ and $\mathbb{E}\ell_t^T u$, Theorem 2 leads to an upper bound on the regret $\overline{R}_n$ of CLEB, which allows us to obtain the bounds of Table 2 by using appropriate choices of F. Moreover, if F admits an Hessian, denoted $\nabla^2 F$, that is always invertible, then one can prove that up to a third-order term $\big(\text{in } \tilde{\ell}_t\big)$, the

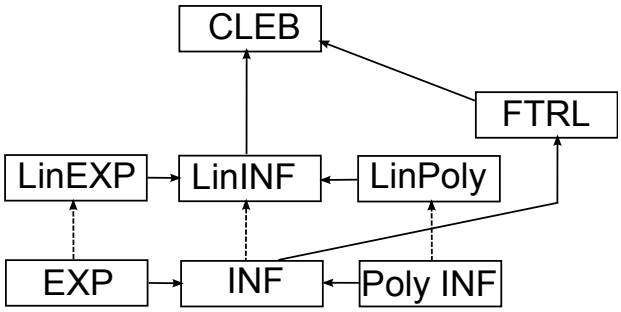

Figure 3: The figure sketches the relationship of the algorithms studied in this paper with arrows representing "is a special case of". Dotted arrows indicate that the link is obtained by "expanding" $\mathcal{S}$, that, is seeing $\mathcal{S}$ as the set of basis vector in $\mathbb{R}^{|\mathcal{S}|}$ rather than seeing it as a (structured) subset of $\{0,1\}^d$ (see Section 3.1). The six algorithms on the bottom use a Legendre function with a diagonal Hessian. On the contrary, the FTRL algorithm (see Section 3.3) may consider Legendre functions more adapted to the geometry of the convex hull of $\mathcal{S}$. POLYINF is the algorithm considered in Audibert and Bubeck (2010).

regret bound can be written as:

$$\sum_{t=1}^{n} \tilde{\ell}_t^T w_t - \sum_{t=1}^{n} \tilde{\ell}_t^T u \lessapprox D_F(u, w_1) + \sum_{t=1}^{n} \tilde{\ell}_t^T \left(\nabla^2 F(w_t)\right)^{-1} \tilde{\ell}_t. \tag{5}$$

In this paper, we restrict our attention to the combinatorial learning setting in which $\mathcal{S}$ is a subset of $\{0,1\}^d$. However, one should note that this specific form of $\mathcal{S}$ plays no role in the definition of CLEB, meaning that the algorithm on Figure 2 can be used to handle general online linear optimization problems, where $\mathcal{S}$ is any subset of $\mathbb{R}^d$.

3 Different instances of CLEB

In this section we describe several instances of CLEB and relate them to existing algorithms. Figure 3 summarizes the relationship between the various algorithms introduced below.

3.1 EXP2 (Expanded Exponentially weighted average forecaster)

The simplest approach to combinatorial prediction games is to consider each vertex of $\mathcal{S}$ as an independent expert, and then apply a strategy designed for the expert problem. We call EXP2 the resulting strategy when one uses the traditional exponentially weighted average forecaster (also called Hedge, Freund and Schapire (1997)), see Figure 4. In the full information game, EXP2 corresponds to Expanded Hedge defined in Koolen et al. (2010), where it was studied under the L_∞ assumption. It was also studied in the full information game under the L_2 assumption in Dani et al. (2008). In the semi-bandit game, EXP2 was studied in György et al. (2007) under the L_∞ assumption. Finally in the bandit game, EXP2 corresponds to the strategy proposed by Dani et al. (2008) and also to the ComBand strategy, studied under the L_∞ assumption in Cesa-Bianchi and Lugosi (2009) and under the L_2 assumption in Cesa-Bianchi and Lugosi (2010). (These last strategies differ in how the losses are estimated.)

EXP2 is a CLEB strategy in dimension $|S|$ that uses $\mathcal{D} = [0,+\infty)^{|S|}$ and the function $F : u \mapsto \frac{1}{\eta}\sum_{i=1}^{|S|} u_i \log(u_i)$, for some $\eta > 0$ (this can be proved by using the fact that the Kullback-Leibler projection on the simplex is equivalent to a L_1-normalization). The following theorem shows the regret bound that one can obtain for EXP2 (for instance with Theorem 5 applied to the case where $\mathcal{S}$ is replaced by $\mathcal{S}' = \left\{u \in \{0,1\}^{|\mathcal{S}|} : \sum_{v\in\mathcal{S}} u_v = 1\right\}$).

Theorem 3 *For the* EXP2 *forecaster, provided that* $\mathbb{E}\tilde{\ell}_t = \ell_t$*, we have*

$$\overline{R}_n \leq \frac{\log(|\mathcal{S}|)}{\eta} + \frac{\eta}{2}\sum_{t=1}^{n}\sum_{v\in\mathcal{S}} \mathbb{E}\left[p_t(v)(\tilde{\ell}_t^T v)^2 \max\left(1, \exp(-\eta\tilde{\ell}_t^T v)\right)\right].$$

3.2 LINEXP (Linear Exponentially weighted average forecaster)

We call LINEXP the CLEB strategy that uses $\mathcal{D} = [0,+\infty)^d$ and the function $F : u \mapsto \frac{1}{\eta}\sum_{i=1}^{d} u_i \log(u_i)$ associated to the Kullback-Leibler divergence, for some $\eta > 0$. In the full information game, LINEXP corresponds to Component Hedge defined in Koolen et al. (2010), where it was studied under the L_∞ assumption.

EXP2*:*

Parameter: Learning rate η.

Let $w_1 = \left(\frac{1}{|\mathcal{S}|}, \ldots, \frac{1}{|\mathcal{S}|}\right) \in \mathbb{R}^{|\mathcal{S}|}$.

For each round $t = 1, 2, \ldots, n$;

(a) Let p_t the distribution on $\mathcal{S}$ such that $p_t(v) = w_{v,t}$ for any $v \in \mathcal{S}$.

(b) Play $V_t \sim p_t$ and observe

- the loss vector ℓ_t in the full information game,
- the coordinates $\ell_{i,t}\mathbb{1}_{V_{i,t}=1}$ in the semi-bandit game,
- the instantaneous loss $\ell_t^T V_t$ in the bandit game.

(c) Estimate the loss vector ℓ_t by $\tilde{\ell}_t$. For instance, one may take

- $\tilde{\ell}_t = \ell_t$ in the full information game,
- $\tilde{\ell}_{i,t} = \frac{\ell_{i,t}}{\sum_{v \in \mathcal{S}: v_i = 1} p_{v,t}} V_{i,t}$ in the semi-bandit game,
- $\tilde{\ell}_t = P_t^+ V_t V_t^T \ell_t$, with $P_t = \mathbb{E}_{v \sim p_t}(vv^T)$ in the bandit game.

(d) Update the weights, for all $v \in \mathcal{S}$,

$$w_{v,t+1} = \frac{\exp(-\eta \tilde{\ell}_t^T v) w_{v,t}}{\sum_{u \in \mathcal{S}} \exp(-\eta \tilde{\ell}_t^T u) w_{u,t}}.$$

Figure 4: EXP2 forecaster.

In the semi-bandit game, LINEXP was studied in Uchiya et al. (2010), Kale et al. (2010) under the L_∞ assumption, and for the particular set $\mathcal{S}$ with all vertices of L_1 norm equal to some value k.

3.3 FTRL (Follow the Regularized Leader)

If $\text{Conv}(\mathcal{S}) \subset \mathcal{D}$ and $w_1 \in \text{argmin}_{w \in \mathcal{D}} F(w)$, steps (d) and (e) are equivalent to

$$w_{t+1} \in \operatorname*{argmin}_{w \in \text{Conv}(\mathcal{S})} \left(\sum_{s=1}^{t} \tilde{\ell}_s^T w + F(w) \right),$$

showing that in this case CLEB can be interpreted as a regularized follow-the-leader algorithm. This type of algorithm was studied in Abernethy and Rakhlin (2009) in the full information and bandit setting (see also the lecture notes Rakhlin and Tewari (2008)). A survey of FTRL strategies for the full information game can be found in Hazan (2010). In the bandit game, FTRL with F being a self-concordant barrier function and a different estimate than the one proposed in Figure 2 was studied in Abernethy et al. (2008).

3.4 LININF (Linear Implicitly Normalized Forecaster)

Let $f : \mathbb{R}^d \to \mathbb{R}$. The function f has a diagonal Hessian if and only if it can be written as $f(u) = \sum_{i=1}^d f_i(u_i)$, for some twice differentiable functions $f_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, d$. The Hessian is called exchangeable when the functions $f_1'', \ldots, f_d''$ are identical. In this case, up to adding an affine function of u (note that this does not alter neither the Bregman divergence nor CLEB), we have $f(u) = \sum_{i=1}^d g(u_i)$ for some twice differentiable function g. In this section, we consider this type of Legendre functions. To underline the surprising link[3] with the Implicitly Normalized Forecaster proposed in Audibert and Bubeck (2010), we consider g of the form $x \mapsto \int_\cdot^x \psi^{-1}(s)ds$, and will refer to the algorithm presented hereafter as LININF.

Definition 4 *Let $\omega \geq 0$. A function $\psi : (-\infty, a) \to \mathbb{R}_+^*$ for some $a \in \mathbb{R} \cup \{+\infty\}$ is called an ω-potential if and only if it is convex, continuously differentiable, and satisfies*

$$\lim_{x \to -\infty} \psi(x) = \omega \qquad \lim_{x \to a} \psi(x) = +\infty$$

$$\psi' > 0 \qquad \int_\omega^{\omega+1} |\psi^{-1}(s)|ds < +\infty.$$

[3] detailed in Appendix A of the extended version Audibert et al. (2011).

Theorem 5 *Let $\omega \geq 0$ and let ψ be an ω-potential function. The function F defined on $\mathcal{D} = [\omega, +\infty)^d$ by $F(u) = \sum_{i=1}^d \int_\omega^{u_i} \psi^{-1}(s)ds$ is Legendre. The associated* CLEB *satisfies, for any $u \in Conv(\mathcal{S}) \cap \mathcal{D}$,*

$$\sum_{t=1}^n \tilde{\ell}_t^T w_t - \sum_{t=1}^n \tilde{\ell}_t^T u \leq D_F(u, w_1) + \frac{1}{2}\sum_{t=1}^n \sum_{i=1}^d \tilde{\ell}_{i,t}^2 \max\Big(\psi'\big(\psi^{-1}(w_{i,t})\big), \psi'\big(\psi^{-1}(w_{i,t}) - \tilde{\ell}_{i,t}\big)\Big), \tag{6}$$

where for any $(u, v) \in \mathcal{D} \times Int(\mathcal{D})$,

$$D_F(u, v) = \sum_{i=1}^d \bigg(\int_{v_i}^{u_i} \psi^{-1}(s)ds - (u_i - v_i)\psi^{-1}(v_i)\bigg). \tag{7}$$

In particular, when the estimates $\tilde{\ell}_{i,t}$ are nonnegative, we have

$$\sum_{t=1}^n \tilde{\ell}_t^T w_t - \sum_{t=1}^n \tilde{\ell}_t^T u \leq D_F(u, w_1) + \sum_{t=1}^n \sum_{i=1}^d \frac{\tilde{\ell}_{i,t}^2}{2(\psi^{-1})'(w_{i,t})}. \tag{8}$$

Proof: It is easy to check that F is a Legendre function and that (7) holds. We also have $\nabla F^*(u) = (\nabla F)^{-1}(u) = \big(\psi(u_1), \ldots, \psi(u_d)\big)$, hence $D_{F^*}(u, v) = \sum_{i=1}^d \bigg(\int_{v_i}^{u_i} \psi(s)ds - (u_i - v_i)\psi(v_i)\bigg)$. From the Taylor-Lagrange expansion, we have $D_{F^*}(u, v) \leq \sum_{i=1}^d \max_{s\in[u_i,v_i]} \frac{1}{2}\psi'(s)(u_i - v_i)^2$. Since the function ψ is convex, we have $\max_{s\in[u_i,v_i]} \psi'(s) \leq \psi'\big(\max(u_i, v_i)\big)$, which gives the desired results. ■

Note that LINEXP is an instance of LININF with $\psi : x \mapsto \exp(\eta x)$. On the other hand, Audibert and Bubeck (2010) recommend the choice $\psi(x) = (-\eta x)^{-q}$ with $\eta > 0$ and $q > 1$ since it leads to the minimax optimal rate $\sqrt{nd}$ for the standard d-armed bandit game $\big($while the best bound for Exp3 is of the order of $\sqrt{nd \log d}\big)$. This corresponds to a function F of the form $F(u) = -\frac{q}{(q-1)\eta}\sum_{i=1}^d u_i^{(q-1)/q}$. We refer to the corresponding CLEB as LINPOLY. In the extended version [Appendix A, Audibert et al. (2011)] we show that a simple application of Theorem 5 proves that LINPOLY with $q = 2$ satisfies $\overline{R}_n \leq 2\sqrt{2nd}$. This improves on the bound $\overline{R}_n \leq 8\sqrt{nd}$ obtained in Theorem 11 of Audibert and Bubeck (2010).

4 Full Information Game

This section details the upper bounds of the forecasters EXP2, LINEXP and LINPOLY under the L_2 and L_∞ assumptions for the full information game. All results are gathered in Table 2 (page 3). The proofs can be found in Appendix B of the extended version Audibert et al. (2011). Up to numerical constants, the results concerning (EXP2, L_2 and L_∞) and (LINEXP, L_∞) appeared or can be easily derived from respectively Dani et al. (2008) and Koolen et al. (2010).

Theorem 6 **(LINEXP, L_∞)** *Under the L_∞ assumption, for* LINEXP *with $\tilde{\ell}_t = \ell_t$, $\eta = \sqrt{2/n}$ and $w_1 = \operatorname{argmin}_{w\in Conv(\mathcal{S})} D_F\big(w, (1, \ldots, 1)^T\big)$, we have*

$$\overline{R}_n \leq d\sqrt{2n}.$$

Theorem 7 **(LINEXP, L_2)** *Under the L_2 assumption, for* LINEXP *with $\tilde{\ell}_t = \ell_t$, $\eta = \sqrt{2d/n}$ and $w_1 = \operatorname{argmin}_{w\in Conv(\mathcal{S})} D_F\big(w, (1, \ldots, 1)^T\big)$, we have*

$$\overline{R}_n \leq \sqrt{2nd}.$$

Theorem 8 **(LINPOLY, L_∞)** *Under the L_∞ assumption, for* LINPOLY *with $\tilde{\ell}_t = \ell_t$, $\eta = \sqrt{\frac{2}{q(q-1)n}}$ and $w_1 = \operatorname{argmin}_{w\in Conv(\mathcal{S})} D_F\big(w, (1, \ldots, 1)^T\big)$, we have*

$$\overline{R}_n \leq d\sqrt{\frac{2qn}{q-1}}.$$

Theorem 9 **(LINPOLY, L_2)** *Under the L_2 assumption, for* LINPOLY *with $\tilde{\ell}_t = \ell_t$, $\eta = \sqrt{\frac{2d}{q(q-1)n}}$ and $w_1 = \operatorname{argmin}_{w\in Conv(\mathcal{S})} D_F\big(w, (1, \ldots, 1)^T\big)$, we have*

$$\overline{R}_n \leq \sqrt{\frac{2qdn}{q-1}}.$$

Theorem 10 (EXP2, L_∞) *Under the L_∞ assumption, for* EXP2 *with $\tilde{\ell}_t = \ell_t$, we have*

$$\overline{R}_n \leq \frac{d \log 2}{\eta} + \frac{\eta n d^2}{2}.$$

In particular for $\eta = \sqrt{\frac{2 \log 2}{nd}}$, we have $\overline{R}_n \leq \sqrt{2d^3 n \log 2}$.

From Theorem 19, the above upper bound is tight, and consequently there exists $\mathcal{S}$ for which the algorithm EXP2 is not minimax optimal in the full information game under the L_∞ assumption.

Theorem 11 (EXP2, L_2) *Under the L_2 assumption, for* EXP2 *with $\tilde{\ell}_t = \ell_t$, we have*

$$\overline{R}_n \leq \frac{d \log 2}{\eta} + \frac{\eta n}{2}.$$

In particular for $\eta = \sqrt{\frac{2d \log 2}{n}}$, we have $\overline{R}_n \leq \sqrt{2dn \log 2}$.

5 Semi-Bandit Game

This section details the upper bounds of the forecasters EXP2, LINEXP and LINPOLY under the L_2 and L_∞ assumptions for the semi-bandit game. These bounds are gathered in Table 2 (page 3). The proofs can be found in Appendix C of the extended version Audibert et al. (2011). Up to the numerical constant, the result concerning (EXP2, L_∞) appeared in György et al. (2007) in the context of the online shortest path problem. Uchiya et al. (2010) and Kale et al. (2010) studied the semi-bandit problem under the L_∞ assumption for action sets of the form $\mathcal{S} = \left\{v \in \{0,1\}^d : \sum_{i=1}^d v_i = k\right\}$ for some value k. Their common algorithm corresponds to LINEXP and the bounds are of order $\sqrt{knd \log(d/k)}$. Our upper bounds for the regret of LINEXP extend these results to more general sets of arms and to the L_2 assumption.

Theorem 12 (LINEXP, L_∞) *Under the L_∞ assumption, for* LINEXP *with $\tilde{\ell}_{i,t} = \ell_{i,t} \frac{V_{i,t}}{w_{i,t}}$, $\eta = \sqrt{2/n}$ and $w_1 = \operatorname{argmin}_{w \in Conv(\mathcal{S})} D_F\big(w, (1, \ldots, 1)^T\big)$, we have*

$$\overline{R}_n \leq d\sqrt{2n}.$$

Since the L_2 assumption implies the L_∞ assumption, we also have $\overline{R}_n \leq d\sqrt{2n}$ under the L_2 assumption.

Let us now detail how LINEXP behaves for almost symmetric action sets as defined below.

Definition 13 *The set $\mathcal{S} \subset \{0,1\}^d$ is called almost symmetric if for some $k \in \{1, \ldots, d\}$, $\mathcal{S} \subset \left\{v \in \{0,1\}^d : \sum_{i=1}^d v_i \leq k\right\}$ and Conv($\mathcal{S}$) $\cap \left[\frac{k}{2d}; 1\right]^d \neq \emptyset$. The integer k is called the order of the symmetry.*

The set $\mathcal{S} = \left\{v \in \{0,1\}^d : \sum_{i=1}^d v_i = k\right\}$ considered in Uchiya et al. (2010) and Kale et al. (2010) is a particular almost symmetric set.

Theorem 14 (LINEXP, **almost symmetric** $\mathcal{S}$) *Let $\mathcal{S}$ be an almost symmetric set of order $k \in \{1, \ldots, d\}$. Consider* LINEXP *with $\tilde{\ell}_{i,t} = \ell_{i,t} \frac{V_{i,t}}{w_{i,t}}$ and $w_1 = \operatorname*{argmin}_{w \in Conv(\mathcal{S})} D_F\big(w, (\frac{k}{d}, \ldots, \frac{k}{d})^T\big)$. Let $\mathcal{L} = \max\big(\log\big(\frac{d}{k}\big), 1\big)$.*

- *Under the L_∞ assumption, taking $\eta = \sqrt{\frac{2k\mathcal{L}}{nd}}$, we have $\overline{R}_n \leq \sqrt{2knd\mathcal{L}}$.*
- *Under the L_2 assumption, taking $\eta = k\sqrt{\frac{\mathcal{L}}{nd}}$, we have $\overline{R}_n \leq 2\sqrt{nd\mathcal{L}}$.*

In particular, it means that under the L_2 assumption, there is a gain in the regret bound of a factor $\sqrt{d/\mathcal{L}}$ when the set of actions is an almost symmetric set of order k.

Theorem 15 (LINPOLY, L_∞) *Under the L_∞ assumption, for* LINPOLY *with $\tilde{\ell}_{i,t} = \ell_{i,t} \frac{V_{i,t}}{w_{i,t}}$, $\eta = \sqrt{\frac{2}{q(q-1)n}}$ and $w_1 = \operatorname{argmin}_{w \in Conv(\mathcal{S})} D_F\big(w, (1, \ldots, 1)^T\big)$, we have*

$$\overline{R}_n \leq d\sqrt{\frac{2qn}{q-1}}.$$

Theorem 16 (LINPOLY, L_2) *Under the L_2 assumption, for* LINPOLY *with $\tilde{\ell}_{i,t} = \ell_{i,t}\frac{V_{i,t}}{w_{i,t}}$, $\eta = \sqrt{\frac{2d^{\frac{1}{q}}}{q(q-1)n}}$ and $w_1 = \operatorname{argmin}_{w \in Conv(\mathcal{S})} D_F\big(w, (1,\dots,1)^T\big)$, we have*

$$\overline{R}_n \leq \sqrt{\frac{2qnd}{q-1}d^{1-\frac{1}{q}}}.$$

In particular, for $q = 1 + (\log d)^{-1}$, we have $\overline{R}_n \leq \sqrt{2nde\log(ed)}$.

Theorem 17 (EXP2, L_∞) *Under the L_∞ assumption, for the* EXP2 *forecaster described in Figure 4 using $\tilde{\ell}_{i,t} = \ell_{i,t}\frac{V_{i,t}}{w_{i,t}}$, we have*

$$\overline{R}_n \leq \frac{d\log 2}{\eta} + \frac{\eta n d^2}{2}.$$

In particular for $\eta = \sqrt{\frac{2\log 2}{nd}}$, we have $\overline{R}_n \leq \sqrt{2d^3 n\log 2}$.

The corresponding lower bound is given in Theorem 19.

Theorem 18 (EXP2, L_2) *Under the L_2 assumption, for* EXP2 *with $\tilde{\ell}_{i,t} = \ell_{i,t}\frac{V_{i,t}}{w_{i,t}}$, we have*

$$\overline{R}_n \leq \frac{d\log 2}{\eta} + \frac{\eta n d}{2}.$$

In particular for $\eta = \sqrt{\frac{2\log 2}{n}}$, we have $\overline{R}_n \leq d\sqrt{2n\log 2}$.

Note that as for LINEXP, we end up upper bounding $\sum_{i=1}^d \ell_{i,t}$ by d. In the case of almost symmetric set $\mathcal{S}$ of order k, this sum can be bounded by $2d/k$, while $\log(|\mathcal{S}|)$ is upper bounded by $k\log(d+1)$. So as for LINEXP, this leads to a regret bound of order $\sqrt{nd\log d}$ when the set of actions is an almost symmetric set.

6 Bandit Game

The upper bounds for EXP2 in the bandit case proposed in Table 2 (page 3) are extracted from Dani et al. (2008). The approach proposed by the authors is to use EXP2 in the space described by a barycentric spanner. More precisely, let $m = \dim(\text{Span}(\mathcal{S}))$ and $e_1,\dots,e_m$ be a barycentric spanner of $\mathcal{S}$; for instance, take $(e_1,\dots,e_m) \in \operatorname{argmax}_{(x_1,\dots,x_m)\in\mathcal{S}^m} |\det_{\text{Span}(\mathcal{S})}(x_1,\dots,x_m)|$ (see Awerbuch and Kleinberg, 2004). We introduce the transformations $T_1 : \mathbb{R}^d \to \mathbb{R}^m$ such that for $x \in \mathbb{R}^d$, $T_1(x) = (x^T e_1,\dots,x^T e_m)^T$, and $T_2 : \mathcal{S} \to [-1,1]^m$ such that for $v \in \mathcal{S}$, $v = \sum_{i=1}^m (T_2(v))_i e_i$. Note that for any $v \in \mathcal{S}$, we have $\ell_t^T v = T_1(\ell_t)^T T_2(v)$. Then the loss estimate for $v \in \mathcal{S}$ is

$$\tilde{\ell}_t^T v = \big(Q_t^+ T_2(V_t)T_2(V_t)^T T_1(\ell_t)\big)^T T_2(v), \text{ where } Q_t = \mathbb{E}_{V\sim p_t} T_2(V)T_2(V)^T.$$

Moreover the authors also add a forced exploration which is uniform over the barycentric spanner.

A concurrent approach is the one proposed in Cesa-Bianchi and Lugosi (2009, 2010). There the authors study EXP2 directly in the original space, with the estimate described in Figure 4, and with an additional forced exploration which is uniform over $\mathcal{S}$. They work out several examples of sets $\mathcal{S}$ for which they improve the regret bound by a factor $\sqrt{d}$ with respect to Dani et al. (2008). Unfortunately there exists sets $\mathcal{S}$ for which this approach fails to provide a bound polynomial in d. In general one needs to replace the uniform exploration over $\mathcal{S}$ by an exploration that is tailored to this set. How to do this in general is still an open question.

The upper bounds for LINEXP in the bandit case proposed in Table 2 (page 3) are derived by using the trick of Dani et al. (2008) (that is, by working with a barycentric spanner). The proof of this result is omitted, since it does not yield the optimal dependency in n. Moreover we can not analyze LINPOLY since (1) is not well defined in this case, because $\tilde{\ell}_t$ can be non-positive. In general we believe that the LININF approach is not sound for the bandit case, and that one needs to work with a Legendre function with non-diagonal Hessian.

The only known CLEB with non-diagonal Hessian is the one proposed in Abernethy et al. (2008), where the authors use a self-concordant barrier function. In this case, they are able to propose a loss estimate related to the structure of the Hessian. This approach is powerful, and under the L_2 assumption leads to a regret upper bound of order $d\sqrt{\theta n\log n}$ for $\theta > 0$ such that $\text{Conv}(\mathcal{S})$ admits a θ-self-concordant barrier function (see Abernethy et al., 2008, section 5). When $\text{Conv}(\mathcal{S})$ admits a $O(1)$-self-concordant barrier function, the upper bound matches the lower bound $O\big(d\sqrt{n}\big)$. The open question is to determine for which sets $\mathcal{S}$, this occurs.

7 Lower Bounds

We start this Section with a result that shows that EXP2 is suboptimal against L_∞ adversaries. This answers a question of Koolen et al. (2010).

Theorem 19 *Let $n \geq d$. There exists a subset $\mathcal{S} \subset \{0,1\}^d$ such that in the full information game, for the* EXP2 *strategy (for any learning rate η), we have*

$$\sup \overline{R}_n \geq 0.02\, d^{3/2}\sqrt{n},$$

where the supremum is taken over all L_∞ adversaries.

Proof: For sake of simplicity we assume here that d is a multiple of 4 and that n is even. We consider the following subset of the hypercube:

$$\mathcal{S} = \Bigg\{ v \in \{0,1\}^d : \sum_{i=1}^{d/2} v_i = d/4 \text{ and } \Big(v_i = 1, \forall i \in \{d/2+1; \ldots, d/2+d/4\} \Big) \text{ or } \Big(v_i = 1, \forall i \in \{d/2+d/4+1, \ldots, d\} \Big) \Bigg\}.$$

That is, choosing a point in $\mathcal{S}$ corresponds to choosing a subset of $d/4$ elements in the first half of the coordinates, and choosing one of the two first disjoint intervals of size $d/4$ in the second half of the coordinates.

We will prove that for any parameter η, there exists an adversary such that Exp (with parameter η) has a regret of at least $\frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right)$, and that there exists another adversary such that its regret is at least $\min\left(\frac{d\log 2}{12\eta}, \frac{nd}{12}\right)$. As a consequence, we have

$$\begin{aligned}
\sup \overline{R}_n &\geq \max\left(\frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right), \min\left(\frac{d\log 2}{12\eta}, \frac{nd}{12}\right)\right) \\
&\geq \min\left(\max\left(\frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right), \frac{d\log 2}{12\eta}\right), \frac{nd}{12}\right) \geq \min\left(A, \frac{nd}{12}\right),
\end{aligned}$$

with

$$\begin{aligned}
A &= \min_{\eta \in [0,+\infty)} \max\left(\frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right), \frac{d\log 2}{12\eta}\right) \\
&\geq \min\left\{ \min_{\eta d \geq 8} \frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right), \min_{\eta d < 8} \max\left(\frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right), \frac{d\log 2}{12\eta}\right)\right\} \\
&\geq \min\left\{ \frac{nd}{16}\tanh(1), \min_{\eta d < 8} \max\left(\frac{nd}{16}\frac{\eta d}{8}\frac{1}{\tanh(1)}, \frac{d\log 2}{12\eta}\right)\right\} \\
&\geq \min\left\{ \frac{nd}{16}\tanh(1), \sqrt{\frac{nd^3\log 2}{128\times 12\times \tanh(1)}}\right\} \geq \min\left(0.04\, nd, 0.02\, d^{3/2}\sqrt{n}\right).
\end{aligned}$$

Let us first prove the lower bound $\frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right)$. We define the following adversary:

$$\ell_{i,t} = \begin{cases} 1 & \text{if} \quad i \in \{d/2+1; \ldots, d/2+d/4\} \text{ and } t \text{ odd}, \\ 1 & \text{if} \quad i \in \{d/2+d/4+1, \ldots, d\} \text{ and } t \text{ even}, \\ 0 & \text{otherwise.} \end{cases}$$

This adversary always put a zero loss on the first half of the coordinates, and alternates between a loss of $d/4$ for choosing the first interval (in the second half of the coordinates) and the second interval. At the beginning of odd rounds, any vertex $v \in \mathcal{S}$ has the same cumulative loss and thus Exp picks its expert uniformly at random, which yields an expected cumulative loss equal to $nd/16$. On the other hand at even rounds the probability distribution to select the vertex $v \in \mathcal{S}$ is always the same. More precisely the probability of selecting a vertex which contains the interval $\{d/2+d/4+1, \ldots, d\}$ (i.e, the interval with a $d/4$ loss at this round) is exactly $\frac{1}{1+\exp(-\eta d/4)}$. This adds an expected cumulative loss equal to $\frac{nd}{8}\frac{1}{1+\exp(-\eta d/4)}$. Finally note that the loss of any fixed vertex is $nd/8$. Thus we obtain

$$\overline{R}_n = \frac{nd}{16} + \frac{nd}{8}\frac{1}{1+\exp(-\eta d/4)} - \frac{nd}{8} = \frac{nd}{16}\tanh\left(\frac{\eta d}{8}\right).$$

We move now to the dependency in $1/\eta$. Here we consider the adversary defined by:

$$\ell_{i,t} = \begin{cases} 1-\varepsilon & \text{if} \quad i \leq d/4, \\ 1 & \text{if} \quad i \in \{d/4+1, \ldots, d/2\}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that against this adversary the choice of the interval (in the second half of the components) does not matter. Moreover by symmetry the weight of any coordinate in $\{d/4+1, \ldots, d/2\}$ is the same (at any round). Finally remark that this weight is decreasing with t. Thus we have the following identities (in the big sums i represents the number of components selected in the first $d/4$ components):

$$\begin{aligned}
\overline{R}_n &= \mathbb{E}\left(\varepsilon \sum_{t=1}^{n} \sum_{i=d/4+1}^{d/2} V_{i,t}\right) = \varepsilon \frac{d}{4} \sum_{t=1}^{n} \mathbb{E} V_{d/2,t} \geq \frac{n\varepsilon d}{4} \mathbb{P}(V_{d/2,n} = 1) \\
&= \frac{n\varepsilon d}{4} \frac{\sum_{v \in \mathcal{S}: v_{d/2}=1} \exp(-\eta n \ell_2^T v)}{\sum_{v \in \mathcal{S}} \exp(-\eta n \ell_2^T v)} \\
&= \frac{n\varepsilon d}{4} \frac{\sum_{i=0}^{d/4-1} \binom{d/4}{i} \binom{d/4-1}{d/4-i-1} \exp(-\eta(nd/4 - in\varepsilon))}{\sum_{i=0}^{d/4} \binom{d/4}{i} \binom{d/4}{d/4-i} \exp(-\eta(nd/4 - in\varepsilon))} \\
&= \frac{n\varepsilon d}{4} \frac{\sum_{i=0}^{d/4-1} \binom{d/4}{i} \binom{d/4-1}{d/4-i-1} \exp(\eta i n\varepsilon)}{\sum_{i=0}^{d/4} \binom{d/4}{i} \binom{d/4}{d/4-i} \exp(\eta i n\varepsilon)} \\
&= \frac{n\varepsilon d}{4} \frac{\sum_{i=0}^{d/4-1} \left(1 - \frac{4i}{d}\right) \binom{d/4}{i} \binom{d/4}{d/4-i} \exp(\eta i n\varepsilon)}{\sum_{i=0}^{d/4} \binom{d/4}{i} \binom{d/4}{d/4-i} \exp(\eta i n\varepsilon)}
\end{aligned}$$

where we used $\binom{d/4-1}{d/4-i-1} = \left(1 - \frac{4i}{d}\right) \binom{d/4}{d/4-i}$ in the last equality. Thus taking $\varepsilon = \min\left(\frac{\log 2}{\eta n}, 1\right)$ yields

$$\overline{R}_n \geq \min\left(\frac{d \log 2}{4\eta}, \frac{nd}{4}\right) \frac{\sum_{i=0}^{d/4-1} \left(1 - \frac{4i}{d}\right) \binom{d/4}{i}^2 \min(2, \exp(\eta n))^i}{\sum_{i=0}^{d/4} \binom{d/4}{i}^2 \min(2, \exp(\eta n))^i} \geq \min\left(\frac{d \log 2}{12\eta}, \frac{nd}{12}\right),$$

where the last inequality follows from Lemma 23 in the extended version Audibert et al. (2011) (see Appendix E). This concludes the proof of the lower bound. ■

The next two theorems give lower bounds under the three feedback assumptions and the two types of adversaries. The cases (L_2, Full Information) and (L_2, Bandit) already appeared in Dani et al. (2008), while the case (L_∞, Full Information) was treated in Koolen et al. (2010) (with more precise lower bounds for subsets $\mathcal{S}$ of particular interest). Note that the lower bounds for the semi-bandit case trivially follow from the ones for the full information game. Thus our main contribution here is the lower bound for (L_∞, Bandit), which is technically quite different from the other cases. We also give explicit constants in all cases.

Theorem 20 *Let $n \geq d$. Against L_∞ adversaries in the cases of full information and semi-bandit games, we have*

$$R_n \geq 0.008 \; d\sqrt{n},$$

and in the bandit game

$$R_n \geq 0.01 \; d^{3/2}\sqrt{n}.$$

Proof: In this proof we consider the following subset of $\{0,1\}^d$:

$$\mathcal{S} = \{v \in \{0,1\}^d : \forall i \in \{1, \ldots, \lfloor d/2 \rfloor\}, v_{2i-1} + v_{2i} = 1\}.$$

Under full information, playing in $\mathcal{S}$ corresponds to playing $\lfloor d/2 \rfloor$ independent standard full information games with 2 experts. Thus we can apply [Theorem 30, Audibert and Bubeck (2010)] to obtain:

$$R_n \geq \lfloor d/2 \rfloor \times 0.03\sqrt{n \log 2} \geq 0.008 \; d\sqrt{n}.$$

We now move to the bandit game, for which the proof is more challenging. For the sake of simplicity, we assume in the following that d is even. Moreover, we restrict our attention to deterministic forecasters, the extension to general forecaster can be done by a routine application of Fubini's theorem.

First step: definitions.

We denote by $I_{i,t} \in \{1,2\}$ the random variable such that $V_{2i,t} = 1$ if and only if $I_{i,t} = 2$. That is, $I_{i,t}$ is the expert chosen at time t in the i^{th} game. We also define the empirical distribution of plays $q_n^i = (q_{1,n}^i, q_{2,n}^i)$ in game i as $q_{j,n}^i = \frac{\sum_{t=1}^n \mathbb{1}_{I_{i,t}=j}}{n}$. Let $J_{i,n}$ be drawn according to q_n^i.

In this proof we consider a set of $2^{d/2}$ adversaries. For $\alpha = (\alpha_1, \ldots, \alpha_{d/2}) \in \{1,2\}^{d/2}$ we define the α-adversary as follows: For any $t \in \{1, \ldots, n\}$, the loss of expert α_i in game i is drawn from a Bernoulli of parameter $1/2$ while the loss of the other expert in game i is drawn from a Bernoulli of parameter $1/2 + \varepsilon$. We note $\mathbb{E}_\alpha$ when we integrate with respect to the reward generation process of the α-adversary. We note $\mathbb{P}_{i,\alpha}$ the law of $J_{i,n}$ when the forecaster plays against the α-adversary. Remark that we have $\mathbb{P}_{i,\alpha}(J_{i,n} = j) = \mathbb{E}_\alpha \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{I_{i,t}=j}$, hence, against the α-adversary we have:

$$\overline{R}_n = \mathbb{E}_\alpha \sum_{t=1}^n \sum_{i=1}^{d/2} \varepsilon \mathbb{1}_{I_{i,t} \neq \alpha_i} = n\varepsilon \sum_{i=1}^{d/2} \left(1 - \mathbb{P}_{i,\alpha}(J_{i,t} = \alpha_i)\right),$$

which implies (since the maximum is larger than the mean)

$$\sup_{\alpha \in \{1,2\}^{d/2}} \overline{R}_n \geq n\varepsilon \sum_{i=1}^{d/2} \left(1 - \frac{1}{2^{d/2}} \sum_{\alpha \in \{1,2\}^{d/2}} \mathbb{P}_{i,\alpha}(J_{i,n} = \alpha_i)\right). \tag{9}$$

Second step: information inequality.

Let $\mathbb{P}_{-i,\alpha}$ be the law of $J_{i,n}$ against the adversary which plays like the α-adversary except that in the i^{th} game, the losses of both coordinates are drawn from a Bernoulli of parameter $1/2 + \varepsilon$ (we call it the $(-i, \alpha)$-adversary). Now we use Pinsker's inequality which gives:

$$\mathbb{P}_{i,\alpha}(J_{i,n} = \alpha_i) \leq \mathbb{P}_{-i,\alpha}(J_{i,n} = \alpha_i) + \sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}_{-i,\alpha}, \mathbb{P}_{i,\alpha})},$$

and thus, (thanks to the concavity of the square root)

$$\frac{1}{2^{d/2}} \sum_{\alpha \in \{1,2\}^{d/2}} \mathbb{P}_{i,\alpha}(J_{i,n} = \alpha_i) \leq \frac{1}{2} + \sqrt{\frac{1}{2^{d/2+1}} \sum_{\alpha \in \{1,2\}^{d/2}} \mathrm{KL}(\mathbb{P}_{-i,\alpha}, \mathbb{P}_{i,\alpha})}. \tag{10}$$

Third step: computation of $\mathrm{KL}(\mathbb{P}_{-i,\alpha}, \mathbb{P}_{i,\alpha})$ *with the chain rule for Kullback-Leibler divergence.*

Note that since the forecaster is deterministic, the sequence of observed losses (up to time n) $W_n \in \{0, 1, \ldots, d\}^n$ uniquely determines the empirical distribution of plays q_n^i, and in particular the law of $J_{i,n}$ conditionally to W_n is the same for any adversary. Thus, if we note $\mathbb{P}_\alpha^n$ (respectively $\mathbb{P}_{-i,\alpha}^n$) the law of W_n when the forecaster plays against the α-adversary (respectively the $(-i, \alpha)$-adversary), then one can easily prove that $\mathrm{KL}(\mathbb{P}_{-i,\alpha}, \mathbb{P}_{i,\alpha}) \leq \mathrm{KL}(\mathbb{P}_{-i,\alpha}^n, \mathbb{P}_\alpha^n)$. Now we use the chain rule for Kullback-Leibler divergence iteratively to introduce the laws $\mathbb{P}_\alpha^t$ of the observed losses W_t up to time t. More precisely, we have,

$$\begin{aligned}
\mathrm{KL}(\mathbb{P}_{-i,\alpha}^n, \mathbb{P}_\alpha^n) &= \mathrm{KL}(\mathbb{P}_{-i,\alpha}^1, \mathbb{P}_\alpha^1) + \sum_{t=2}^n \sum_{w_{t-1} \in \{0,1,\ldots,d\}^{t-1}} \mathbb{P}_{-i,\alpha}^{t-1}(w_{t-1}) \mathrm{KL}(\mathbb{P}_{-i,\alpha}^t(.|w_{t-1}), \mathbb{P}_\alpha^t(.|w_{t-1})) \\
&= \mathrm{KL}\left(\mathcal{B}_\emptyset, \mathcal{B}'_\emptyset\right) \mathbb{1}_{I_{i,1}=\alpha_i} + \sum_{t=2}^n \sum_{w_{t-1}: I_{i,t}=\alpha_i} \mathbb{P}_{-i,\alpha}^{t-1}(w_{t-1}) \mathrm{KL}\left(\mathcal{B}_{w_{t-1}}, \mathcal{B}'_{w_{t-1}}\right),
\end{aligned}$$

where $\mathcal{B}_{w_{t-1}}$ and $\mathcal{B}'_{w_{t-1}}$ are sums of $d/2$ Bernoulli distributions with parameters in $\{1/2, 1/2 + \varepsilon\}$ and such that the number of Bernoullis with parameter $1/2 + \varepsilon$ in $\mathcal{B}_{w_{t-1}}$ is equal to the number of Bernoullis with parameter $1/2 + \varepsilon$ in $\mathcal{B}'_{w_{t-1}}$ plus one. Now using Lemma 24 from the extended version Audibert et al. (2011) (see Appendix E) we obtain $\mathrm{KL}(\mathbb{P}_{-i,\alpha}^n, \mathbb{P}_\alpha^n) \leq \frac{16\,\varepsilon^2}{d} \mathbb{E}_{-i,\alpha} \sum_{t=1}^n \mathbb{1}_{I_{i,t}=\alpha_i}$. Summing and plugging this into (10) we obtain $\frac{1}{2^{d/2}} \sum_{\alpha \in \{1,2\}^{d/2}} \mathbb{P}_{i,\alpha}(J_{i,n} = \alpha_i) \leq \frac{1}{2} + 2\varepsilon\sqrt{\frac{n}{d}}$. To conclude the proof one needs to plug in this last equation in (9) along with straightforward computations. ■

Theorem 21 *Let $n \geq d$. Against L_2 adversaries in the cases of full information and semi-bandit games, we have*

$$R_n \geq 0.05\sqrt{dn},$$

and in the bandit game

$$R_n \geq 0.05 \min(n, d\sqrt{n}).$$

The proof of this last result can be found in Appendix D of the extended version Audibert et al. (2011).

References

J. Abernethy and A. Rakhlin. Beating the adaptive bandit with high probability. In *22nd annual conference on learning theory*, 2009.

J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In Rocco A. Servedio and Tong Zhang, editors, *COLT*, pages 263–274. Omnipress, 2008.

J.-Y. Audibert and S. Bubeck. Regret bounds and minimax policies under partial monitoring. *JMLR*, 11: 2635–2686, 2010.

J.-Y. Audibert, S. Bubeck, and G. Lugosi. Minimax policies for combinatorial prediction games. *Available on Arxiv*, 2011.

B. Awerbuch and R.D. Kleinberg. Adaptive routing with end-to-end feedback: distributed learning and geometric approaches. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 45–53. ACM, 2004.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006. ISBN 0521841089.

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. In *22nd annual conference on learning theory*, 2009.

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Submitted*, 2010.

V. Dani, T. Hayes, and S.M. Kakade. The price of bandit information for online optimization. In *Advances in Neural Information Processing Systems*, volume 20, pages 345–352, 2008.

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*, 55:119–139, 1997.

A. György, T. Linder, G. Lugosi, and G. Ottucsák. The on-line shortest path problem under partial monitoring. *J. Mach. Learn. Res.*, 8:2369–2403, 2007.

E. Hazan. A survey: The convex optimization approach to regret minimization. Working draft, 2010.

D. P. Helmbold and M. K. Warmuth. Learning permutations with exponential weights. *JMLR*, 10:1705–1736, 2009.

M. Herbster and M. K. Warmuth. Tracking the best expert. *Mach. Learn.*, 32:151–178, August 1998. ISSN 0885-6125.

A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71:291307, 2005.

S. Kale, L. Reyzin, and R. Schapire. Non-stochastic bandit slate problems. *Advances in Neural Information Processing Systems*, pages 1054–1062, 2010.

W. M. Koolen, M. K. Warmuth, and J. Kivinen. Hedging structured concepts. In *23rd annual conference on learning theory*, 2010.

H. B. McMahan and A. Blum. Online geometric optimization in the bandit setting against an adaptive adversary. In *In Proceedings of the 17th Annual Conference on Learning Theory*, pages 109–123, 2004.

A. Rakhlin and A. Tewari. Lecture notes on online learning. 2008.

T. Uchiya, A. Nakamura, and M. Kudo. Algorithms for adversarial bandit problems with multiple plays. In *Proc. of the 21st International Conference on Algorithmic Learning Theory*, 2010.