Supplementary material to "Online Optimization in $\mathcal{X}$-Armed Bandits"

Sébastien Bubeck
INRIA Lille, SequeL project, France
sebastien.bubeck@inria.fr

Rémi Munos
INRIA Lille, SequeL project, France
remi.munos@inria.fr

Gilles Stoltz
Ecole Normale Supérieure and HEC Paris
gilles.stoltz@ens.fr

Csaba Szepesvári
Department of Computing Science, University of Alberta
szepesva@cs.ualberta.ca

A Proof of Lemma 1

Proof. We denote by $x^*_{h,i}(\delta)$ an element of $\mathcal{P}_{h,i}$ such that

$$f\big(x^*_{h,i}(\delta)\big) \geq f^*_{h,i} - \delta \,.$$

By the weakly Lipschitz property, it then follows that for all $y \in \mathcal{P}_{h,i}$,

$$f^* - f(y) \leq f^* - f\big(x^*_{h,i}(\delta)\big) + \max\Big\{ f^* - f\big(x^*_{h,i}(\delta)\big),\, \ell\big(x^*_{h,i}(\delta),\, y\big) \Big\} \leq \Delta_{h,i} + \delta + \max\big\{ \Delta_{h,i} + \delta,\, \operatorname{diam} \mathcal{P}_{h,i} \big\} \,.$$

Letting $\delta \to 0$ and substituting the bounds on the suboptimality and on the diameter of $\mathcal{P}_{h,i}$ concludes the proof. $\square$

B Proof of Lemma 2

Proof. We consider a given round $t \in \{1, \ldots, n\}$. If $(H_t, I_t) \in \mathcal{C}(h,i)$, then this is because the child of (k, i^*_k) on the path to (h,i) had a better B–value than its brother $(k+1, i^*_{k+1})$. Since by definition, B–values can only decrease on a path, this entails that $B_{h,i}(t) \geq B_{k+1,i^*_{k+1}}(t)$. This is turns implies, again by definition of the B–values, that $U_{h,i}(t) \geq B_{k+1,i^*_{k+1}}(t)$. Thus,

$$\big\{(H_t, I_t) \in \mathcal{C}(h,i)\big\} \subset \big\{U_{h,i}(t) \geq B_{k+1,i^*_{k+1}}(t)\big\} \subset \big\{U_{h,i}(t) \geq f^*\big\} \cup \big\{B_{k+1,i^*_{k+1}}(t) \leq f^*\big\} \,.$$

But, once again by definition of B–values,

$$\big\{B_{k+1,i^*_{k+1}}(t) \leq f^*\big\} \subset \big\{U_{k+1,i^*_{k+1}}(t) \leq f^*\big\} \cup \big\{B_{k+2,i^*_{k+2}}(t) \leq f^*\big\} \,,$$

and the argument can be iterated. Since at round t not more than t nodes have been played (including the suboptimal (h,i)), we know that (t, i^*_t) and its descendants have U–values and B–values equal to $+\infty$. We thus have proved the inclusion

$$\big\{(H_t, I_t) \in \mathcal{C}(h,i)\big\} \subset \big\{U_{h,i}(t) \geq f^*\big\} \cup \Big(\big\{B_{k+1,i^*_{k+1}}(t) \leq f^*\big\} \cup \ldots \cup \big\{B_{t-1,i^*_{t-1}}(t) \leq f^*\big\} \Big) \,.$$

The result follows by simply distinguishing whether $N_{h,i}(t) > u$ (which can only happen if $t \geq u$) or not. $\square$

C Proof of Lemma 3

Proof. $U_{h,i} \le f^*$ is not true when node (h,i) was never pulled (in this case, by definition, $U_{h,i}(n) = +\infty$). We may thus conduct the study in the sequel on the event $\big\{N_{h,i}(n) \ge 1\big\}$.

Lemma 1 with $c = 0$ gives that $f^* - f(x) \le \nu_1 \rho^h$ holds for any arm $x \in \mathcal{P}_{h,i}$. Hence,

$$\sum_{t=1}^{n} \big(f(X_t) + \nu_1 \rho^h - f^*\big)\, \mathbb{I}_{\{(H_t,I_t)\in\mathcal{C}(h,i)\}} \ge 0$$

and therefore,

$$\begin{aligned}
&\mathbb{P}\big\{U_{h,i}(n) \le f^* \ \text{ and } \ N_{h,i}(n) \ge 1\big\} \\
&\quad = \ \mathbb{P}\left\{\widehat{\mu}_{h,i}(n) + \sqrt{\frac{2\ln n}{N_{h,i}(n)}} + \nu_1\rho^h \le f^* \ \text{ and } \ N_{h,i}(n) \ge 1\right\} \\
&\quad = \ \mathbb{P}\left\{N_{h,i}(n)\,\widehat{\mu}_{h,i}(n) + N_{h,i}(n)\,\big(\nu_1\rho^h - f^*\big) \le -\sqrt{N_{h,i}(n)\,2\ln n} \ \text{ and } \ N_{h,i}(n) \ge 1\right\} \\
&\quad \le \ \mathbb{P}\left\{\sum_{t=1}^{n} \big(f(X_t) - Y_t\big)\mathbb{I}_{\{(H_t,I_t)\in\mathcal{C}(h,i)\}} \ge \sqrt{N_{h,i}(n)\,2\ln n} \ \text{ and } \ N_{h,i}(n) \ge 1\right\} .
\end{aligned}$$

We take care of the last term with a union bound and the Hoeffding-Azuma inequality for martingale differences. To do this properly we need to define a sequence of (random) times when arms in $\mathcal{C}(h,i)$ were pulled:

$$T_j = \min\{\, t \,:\, N_{h,i}(t) = j \,\}, \quad j = 1, 2, \ldots .$$

Note that $1 \le T_1 < T_2 < \ldots$ and hence it holds that $T_j \ge j$. With these notation, $\widetilde{X}_j = X_{T_j}$ is the j–th arm pulled in a domain corresponding to $\mathcal{C}(h,i)$, $\widetilde{Y}_j = Y_{T_j}$ is the corresponding reward, and

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{t=1}^{n} \big(f(X_t) - Y_t\big)\mathbb{I}_{\{(H_t,I_t)\in\mathcal{C}(h,i)\}} \ge \sqrt{N_{h,i}(n)\,2\ln n} \ \text{ and } \ N_{h,i}(n) \ge 1\right\} \\
&\quad = \ \mathbb{P}\left\{\sum_{j=1}^{N_{h,i}(n)} \Big(f(\widetilde{X}_j) - \widetilde{Y}_j\Big) \ge \sqrt{N_{h,i}(n)\,2\ln n} \ \text{ and } \ N_{h,i}(n) \ge 1\right\} \\
&\quad \le \ \sum_{t=1}^{n} \mathbb{P}\left\{\sum_{j=1}^{t} \Big(f(\widetilde{X}_j) - \widetilde{Y}_j\Big) \ge \sqrt{2\,t\ln n}\right\}
\end{aligned}$$

where we used a union bound to get the last inequality.

We now prove that

$$Z_t = \sum_{j=1}^{t} \Big(f(\widetilde{X}_j) - \widetilde{Y}_j\Big)$$

is a martingale difference sequence (with respect to the filtration it generates). This follows, via optional skipping (see [**?**], Theorem 2.3), from the fact that

$$\sum_{t=1}^{n} \big(f(X_t) - Y_t\big)\mathbb{I}_{\{(H_t,I_t)\in\mathcal{C}(h,i)\}}$$

is a martingale, with respect to the filtration $\mathcal{F}_t = \sigma(X_1, Y_1, \ldots, X_t, Y_t)$, and that $\{T_j = k\} \in \mathcal{F}_{k-1}$. Applying the Hoeffding-Azuma inequality (using the bounded ranges), we then get, for each $t \geq 1$,

$$\mathbb{P}\left\{\sum_{j=1}^{t}\Big(f(\widetilde{X}_j) - \widetilde{Y}_j\Big) \geq \sqrt{2\,t\ln n}\right\} \leq \exp\left(-\frac{2\left(\sqrt{2\,t\ln n}\right)^2}{t}\right) = n^{-4}\,,$$

which concludes the proof. □

D Proof of Lemma 4

Proof. Remark that for the u mentioned in the statement of the lemma,

$$\sqrt{\frac{2\ln t}{u}} + \nu_1\rho^h \leq (\Delta_{h,i} + \nu_1\rho^h)/2\,,$$

and therefore,

$$\begin{aligned}
&\mathbb{P}\big\{U_{h,i}(t) > f^* \text{ and } N_{h,i}(t) > u\big\} \\
&\quad= \mathbb{P}\left\{\widehat{\mu}_{h,i}(t) + \sqrt{\frac{2\ln t}{N_{h,i}(t)}} + \nu_1\rho^h > f^*_{h,i} + \Delta_{h,i} \text{ and } N_{h,i}(t) > u\right\} \\
&\quad\leq \mathbb{P}\left\{\widehat{\mu}_{h,i}(t) > f^*_{h,i} + \frac{\Delta_{h,i} - \nu_1\rho^h}{2} \text{ and } N_{h,i}(t) > u\right\} \\
&\quad\leq \mathbb{P}\left\{N_{h,i}(t)\left(\widehat{\mu}_{h,i}(t) - f^*_{h,i}\right) > \frac{\Delta_{h,i} - \nu_1\rho^h}{2}\,u \text{ and } N_{h,i}(t) > u\right\} \\
&\quad= \mathbb{P}\left\{\sum_{s=1}^{t}\big(Y_s - f^*_{h,i}\big)\mathbb{I}_{\{(H_s,I_s)\in\mathcal{C}(h,i)\}} > \frac{\Delta_{h,i} - \nu_1\rho^h}{2}\,u \text{ and } N_{h,i}(t) > u\right\} \\
&\quad\leq \mathbb{P}\left\{\sum_{s=1}^{t}\big(Y_s - f(X_s)\big)\mathbb{I}_{\{(H_s,I_s)\in\mathcal{C}(h,i)\}} > \frac{\Delta_{h,i} - \nu_1\rho^h}{2}\,u \text{ and } N_{h,i}(t) > u\right\}.
\end{aligned}$$

Now it follows again by the optional skipping argument, the Hoeffding-Azuma inequality, and a union bound, that

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{s=1}^{t}\big(Y_s - f(X_s)\big)\mathbb{I}_{\{(H_s,I_s)\in\mathcal{C}(h,i)\}} > \frac{\Delta_{h,i} - \nu_1\rho^h}{2}\,u \text{ and } N_{h,i}(t) > u\right\} \\
&\quad\leq \sum_{s=u+1}^{t}\exp\left(-\frac{2}{s}\left(\frac{(\Delta_{h,i} - \nu_1\rho^h)\,u}{2}\right)^2\right) \leq t\,\exp\left(-\frac{1}{2}\,u\left(\Delta_{h,i} - \nu_1\rho^h\right)^2\right) \leq t\,n^{-4}
\end{aligned}$$

(where we used the stated bound on u to obtain the last inequality). □

E Proof of Theorem 2

We only deal with the case of deterministic strategies. The extension to randomized strategies can be done using Fubini's theorem.

For $\eta \in [0,\, 1/4]$ and $x^* \in \mathcal{X}$, we denote by f_{η,x^*} the mapping defined by

$$f_{\eta,x^*}(x) = \max\big\{\eta - \ell(x,x^*),\, 0\big\}$$

for all $x \in \mathcal{X}$ and by M_{η,x^*} the environment defined by

$$M_{\eta,x^*}(x) = \mathrm{Ber}\left(\frac{1}{2} + f_{\eta,x^*}(x)\right)$$

for all $x \in \mathcal{X}$. We consider K points $x_1,\ldots,x_K$ in $\mathcal{X}$ such that the balls $B_{x_j,\eta}$ with radius η centered at each of the x_j are non-overlapping. Note that $B_{x_j,\eta}$ is the support of f_{η,x^*}. In addition, the mean functions of all the defined environments are 1–Lipschitz and thus are weakly Lipschitz.

We will also need to consider environments on a finite set of arms $\{1,\ldots,K+1\}$. We construct K different product-distributions $\nu_1,\nu_2,\ldots,\nu_K$ for the arms $\{1,\ldots,K+1\}$ as follows. For a given ν_j, the reward distribution associated to the i-th arm is $\nu_{j,i} = \mathrm{Ber}(1/2)$ for all $i \neq j$ and $\nu_{j,j} = \mathrm{Ber}(1/2+\eta)$.

To each (deterministic) strategy φ on $\mathcal{X}$, we associate a random strategy ψ on the finite set of arms $\{1,\ldots,K+1\}$ as follows. Let $t \geq 1$. Since φ is deterministic it associates to each sequence of rewards $\{r_1,..,r_{t-1}\} \in \{0,1\}^{t-1}$ a unique sequence $\{x_1,..,x_t\} \in \mathcal{X}^t$ of arms that φ would have pull under this sequence of rewards. With a slight abuse of notation we can write $\varphi(r_1,..,r_{t-1}) = (x_1,..,x_t)$. Now assume that the historic of ψ at time t is $X_1,R_1,\ldots,X_{t-1},R_{t-1}$ and let $(X'_1,..,X'_t) = \varphi(R_1,..,R_{t-1})$. We then define

$$\begin{array}{ll} \psi_t = \delta_{K+1} & \text{if} \quad X'_t \notin \cup_j B_{x_j,\eta}, \\ \psi_t = \left(1 - \frac{\ell(X'_t,x_j)}{\eta}\right)\delta_{x_j} + \frac{\ell(X'_t,x_j)}{\eta}\delta_{K+1} & \text{if} \quad X'_t \in B_{x_j,\eta}, \end{array}$$

where δ_j is a dirac distribution on j.

We now want to prove that the distributions of the regrets for φ under M_{η,x_j} and for ψ under ν_j are equal for all $j = 1,\ldots,K$. On the one hand, the expectations of the best arms are $1/2+\eta$ under all these environments. On the other hand we can prove recursively that for any $\{r_1,..,r_t\} \in \{0,1\}^t$,

$$\mathbb{P}(R_1 = r_1,..,R_t = r_t) = \mathbb{P}(R'_1 = r_1,..,R'_t = r_t).$$

where $R_1,\ldots,R_t$ (respectively $R'_1,\ldots,R'_t$) is the sequence of rewards obtained by φ under M_{η,x_j} (respectively ψ under ν_j). The result is easy to check for $t=1$ and for $t>1$ it follows from

$$\mathbb{P}(R_1 = r_1,..,R_t = r_t) = \mathbb{P}(R_t = r_t | R_1 = r_1,..,R_t = r_t)\mathbb{P}(R_1 = r_1,..,R_{t-1} = r_{t-1})$$

and the same calculation for R'_t.

As a consequence, the regrets $R_n(\varphi)$ and $R_n(\psi)$ have the same expectation, that is, for all $j = 1,\ldots,K$,

$$\mathbb{E}_j\, R_n(\varphi) = \mathbb{E}'_j\, R_n(\psi) \tag{1}$$

where $\mathbb{E}_j$ denotes the expectation under M_{η,x_j} and $\mathbb{E}'_j$ the one under ν_j.

But it can be extracted from the proof of the lower bound of [**?** , Section 6.9] that for all strategies ψ', all $\eta \in [0,1/4]$, and all integers K,

$$\max_{j=1,\ldots,K} \mathbb{E}'_j\, R_n(\psi') \geq \eta n \left(1 - \frac{1}{K} - \eta\sqrt{4\ln(4/3)\,\frac{n}{K}}\right)\ . \tag{2}$$

By the assumption on packing dimension, there exists $c > 0$ such that $K = c\,\eta^{-d} \geq 2$ is a suitable choice. Substituting this value, we get

$$\max_{j=1,\ldots,K} \mathbb{E}_j\, R_n(\varphi) = \max_{j=1,\ldots,K} \mathbb{E}'_j\, R_n(\psi) \geq \eta n \left(\frac{1}{2} - \eta^{1+d/2}\sqrt{\frac{4\ln(4/3)}{c}\,n}\right)\ .$$

The left-hand side is smaller than the maximal regret with respect to all weak-Lipschitz environments; the right-hand side can be optimized over $\eta \leq 1/4$ to get the claimed bound, by taking

$$\eta = \left(\frac{1}{4} \sqrt{\frac{c}{4 \, \ln(4/3)}} \right)^{2/(d+2)} n^{-1/(d+2)} \, .$$