Online Optimization in $\mathcal{X}$-Armed Bandits

Sébastien Bubeck[1], Rémi Munos[1], Csaba Szepesvári[2], Gilles Stoltz[3]

[1] INRIA Lille - Nord Europe, SequeL project, France; [2] Department of Computing Science, University of Alberta Edmonton T6G 2E8, Canada; [3]Ecole Normale Supérieure, CNRS, France and HEC Paris, CNRS, France

Abstract

We consider the problem of online optimization of a noisy function in an arbitrary measurable space. We prove that if we know the local smoothness of the function around its maximum then one can perform efficient optimization. In particular one can obtain an expected cumulative regret of order $\sqrt{n}$ no matter the "dimension" of the space.

Framework

- A measurable space $\mathcal{X}$ equipped with a *dissimilarity* ℓ, that is, a non-negative mapping $\ell : \mathcal{X}^2 \to \mathbb{R}$ satisfying $\ell(x, x) = 0$.
- An $\mathcal{X}$-armed bandit (or environment on $\mathcal{X}$), $M : \mathcal{X} \to \mathcal{P}_1([0, 1])$ (the set of probability measures on $[0, 1]$).
- Let $f(x)$ be the expectation of $M(x)$.
- When one pulls a point $x \in \mathcal{X}$ one receives an independent reward $Y \sim M(x)$.
- Goal: search online where is the maximum $f^* = \sup_{x \in \mathcal{X}} f(x)$ of f. That is, pull sequentially points $X_1, \ldots, X_n$ so as to minimize the cumulative regret

$$R_n = \sum_{t=1}^{n} f^* - f(X_t).$$

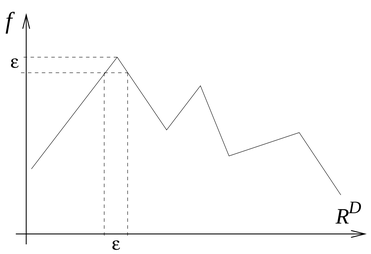
Assumption 1. *The mean-payoff function f is weakly Lipschitz with respect to ℓ, i.e., for all $x, y \in \mathcal{X}$,*

$$f^* - f(y) \leq f^* - f(x) + \max\{f^* - f(x),\, \ell(x, y)\} \ .$$

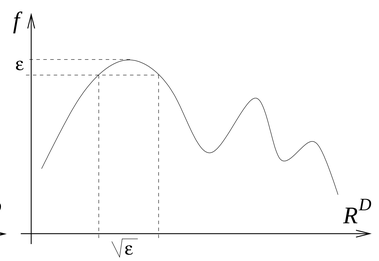
Basically this assumption implies that f is locally Lipschitz around any maxima and that there is no brutal decrease in the neighborhood of any maxima.

Near-optimality dimension

Definition 1. *Let $\mathcal{X}_\varepsilon = \{x \in \mathcal{X}, f^* - f(x) \leq \varepsilon\}$ be the set of ε–optimal arms. The near-optimality dimension of f is defined as the smallest $d \geq 0$ such that $\mathcal{X}_\varepsilon$ can be packed with $O(\varepsilon^{-d})$ balls of radius ε for ε sufficiently small.*



For $\ell(x, y) = ||x - y||$ we get $d = 0$.



For $\ell(x, y) = ||x - y||$ we get $d = 1/2$ but with $\ell(x, y) = ||x - y||^2$ we get $d = 0$.

HOO (Hierarchical Optimistic Optimization)

Input: Tree of coverings

- (h, i) is the i–th node of depth h and corresponds to a subset $\mathcal{P}_{h,i} \subset \mathcal{X}$;
- the root corresponds the whole domain, i.e., $\mathcal{X} = \mathcal{P}_{0,1}$;
- any parent node is covered by its two children
- the diameter (measured with ℓ) of the domains shrinks as the depth increases:

$$\mathrm{diam}(\mathcal{P}_{h,i}) \leq \rho^h; \rho \in (0, 1).$$

We also require that $\mathcal{P}_{h,i}$ contains a ball (with respect to ℓ) of diameter $c\rho^h$ for a fixed $c > 0$.

Global strategy given B–values for each node:

- Start with all nodes "turned off".
- Follow a path from the root to a turned-off node (h, i), where at each node along the path you select the child with the largest B–value.
- Pull a point in $\mathcal{P}_{h,i}$ and turn on the node (h, i).

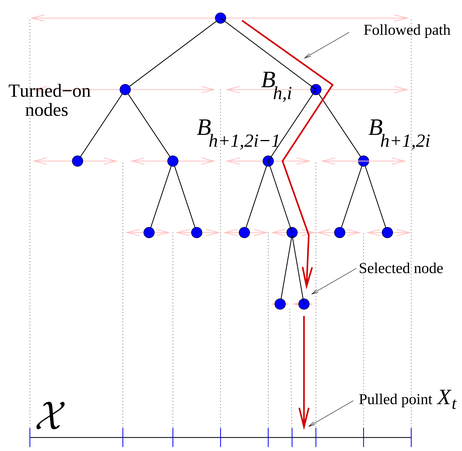
Definition of B–values:

- Let $N_{h,i}(n)$ be the number of times (up to time n) we followed a path going through (h, i).
- Let $\widehat{\mu}_{h,i}(n)$ be the empirical average of rewards collected when we followed a path going through (h, i).
- We consider the following upper confidence bound for each turned-on node :

$$U_{h,i}(n) = \widehat{\mu}_{h,i}(n) + \sqrt{\frac{2 \ln n}{N_{h,i}(n)}} + \rho^h.$$

- For turned-off nodes we set the B–values to infinity and for each turned-on node (h, i) we set:

$$B_{h,i}(n) = \min\Big\{U_{h,i}(n),\ \max\big\{B_{h+1,2i-1}(n),\ B_{h+1,2i}(n)\big\}\Big\}.$$



Theoretical Results

Theorem 1 (Main result)**.** *Let d be the near-optimality dimension of f. There exists a constant $C(d)$ such that for all $n \geq 1$, HOO strategy satisfy*

$$\mathbb{E} R_n \leq C(d)\, n^{(d+1)/(d+2)} \left(\ln n\right)^{1/(d+2)} .$$

Theorem 2 (Minimax Optimality)**.** *Let $c > 0$ such that for all $\varepsilon \leq 1/4$ there exist $c\,\varepsilon^{-D} \geq 2$ disjoint balls of radius ε in $\mathcal{X}$. Then for all $n \geq 4^{D-1}\, c/\ln(4/3)$, all strategies are bound to suffer a regret of at least*

$$\sup \mathbb{E}\, R_n \geq \frac{1}{4} \left(\frac{1}{4} \sqrt{\frac{c}{4 \ln(4/3)}}\right)^{2/(D+2)} n^{(D+1)/(D+2)},$$

where the supremum is taken over all environments with weakly Lipschitz payoff functions. Moreover HOO matches this rate up to logarithmic terms.

Example

We consider $\mathcal{X} = [0, 1]^D$ and a tree of dyadic partitions. Let $\alpha \in [0, \infty)$ and assume that for any maximum x^* of f:

$$f(x^*) - f(x) = \Theta(||x - x^*||^\alpha) \text{ as } x \to x^*.$$

We run the algorithm with $\ell_\beta(x, y) = ||x - y||^\beta$.

- **Known smoothness: $\beta = \alpha$.** We get a near optimality dimension $d = 0$. Thus in this case the regret of HOO is $\widetilde{O}(\sqrt{n})$**,i.e., the rate is independent of the dimension** D. Previously this rate had be obtained only for $D = 1$ or when $\alpha \leq 1$ (and with algorithms arguably less easy to implement).
- **Smoothness underestimated: $\beta < \alpha$.** The near optimality dimension is $d = D\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)$ and the regret is $\widetilde{O}(n^{(d+1)/(d+2)})$.
- **Smoothness overestimated: $\beta > \alpha$.** No guarantee since the Weak Lipschitz assumption is violated.

Numerical Example