Tightening Regret Lower and Upper Bounds in Restless Rising Bandits

Cristiano Migali, Marco Mussi, Gianmarco Genalti, and Alberto Maria Metelli
{cristiano.migali, marco.mussi, gianmarco.genalti, albertomaria.metelli}@polimi.it

Open Problem

Restless rising and **rising concave bandits** are two classes of non-stationary *MABs* where the expected rewards are respectively *non-decreasing* and *non-decreasing concave* functions of time.

	Existing Regret	
	Lower Bounds	Upper Bounds
Rising	$\Omega\left(T^{\frac{1}{2}}\right)$ *Stationary MABs*	$O\left(T^{\frac{2}{3}}\right)$ *Besbes et al. 2014*
Rising Concave	$\Omega\left(T^{\frac{1}{2}}\right)$ *Stationary MABs*	$O\left(T^{\frac{2}{3}}\right)$ *Besbes et al. 2014*

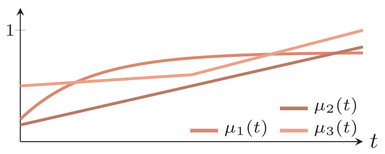
Setting

Reward

$$R_t = X_{I_t,t} \quad \text{with } \mathbb{E}[X_{i,t}] = \mu_i(t)$$

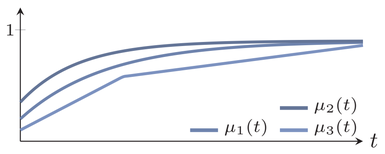
Restless Rising Bandits

Non-Decreasing:
$\gamma_i(t) := \mu_i(t+1) - \mu_i(t) \geqslant 0$



Restless Rising Concave Bandits

Non-Decreasing:
$\gamma_i(t) \geqslant 0$
Concave:
$\gamma_i(t) \geqslant \gamma_i(t+1)$



Variation Budget

$$V_T \geqslant \sum_{t=1}^{T-1} \max_{i \in [\![K]\!]} \gamma_i(t)$$

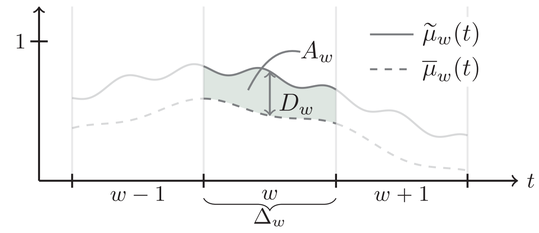
Goal

Minimize
$$R(T) := \mathbb{E}\left[\sum_{t=1}^{T} \left(\mu^*(t) - \mu_{I_t}(t)\right)\right]$$

Lower Bounds

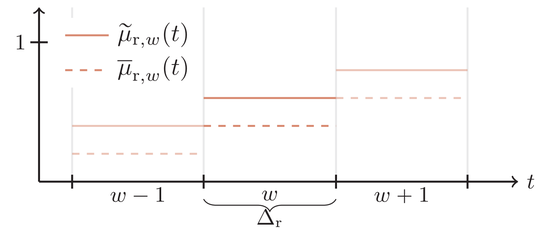
General Recipe

$$R(T) = \Omega\left(\sum_{w=1}^{W} \left(1 - \sqrt{\frac{D_w}{K}}\right) A_w\right)$$



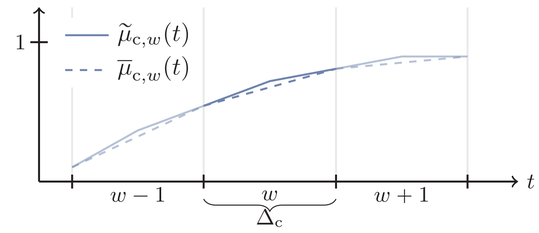
Restless **Rising** Bandits

$$R(T) = \Omega\left(\sigma^{\frac{2}{3}} T^{\frac{2}{3}} K^{\frac{1}{3}} \min\{1, V_T\}^{\frac{1}{3}}\right)$$



Restless **Rising Concave** Bandits

$$R(T) = \Omega\left(\sigma^{\frac{4}{5}} T^{\frac{3}{5}} K^{\frac{2}{5}} \min\{1, V_T\}^{\frac{1}{5}}\right)$$

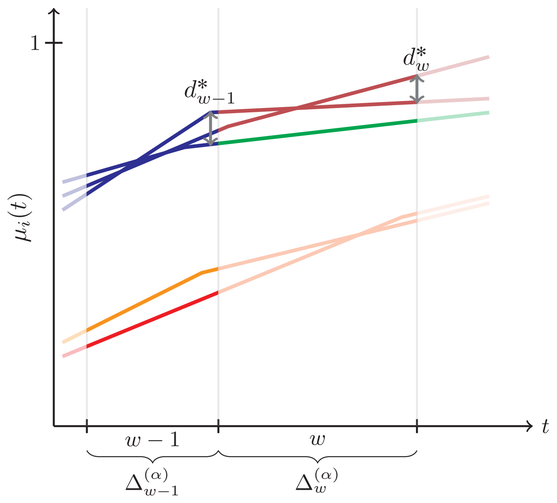


Algorithm for Restless Rising Concave Bandits

Algorithm 1 RC-BE(α).

Let $w \leftarrow 1$ be the **window index**
RESTART:
Let $\mathcal{A} \leftarrow [\![K]\!]$ be the set of **alive arms**,
$d \leftarrow 1$ the **round index** in the **current window**,
$\hat{S}_i \leftarrow 0$ for $i \in [\![K]\!]$ the **cumulative reward**
While $d \leqslant \Delta_w^{(\alpha)} := \lceil w^\alpha \rceil$:
If $|\mathcal{A}| > 1$:
Play each arm in $\mathcal{A}$ once,
increment d (**stop** if $d > \Delta_w^{(\alpha)}$)
Remove from $\mathcal{A}$ all i
s.t. $\hat{S}_i + B_w^{(\alpha)} < \hat{S}^* := \max_{i \in [\![K]\!]} \hat{S}_i$
If $\mathcal{A} = \{\hat{i}^*\}$
Commit to the remaining arm $\hat{i}^*$
Increment d
Increment w
GOTO RESTART

Exploration Commitment



Regret Upper Bound

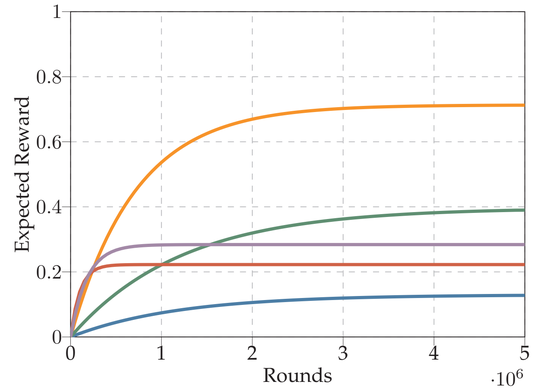
$$R(T) = \widetilde{\mathcal{O}}\left(\sigma^{\frac{9}{11}} T^{\frac{7}{11}} K^{\frac{15}{11}} V_T^{\frac{2}{11}}\right)$$

Numerical Simulations

Experimental Setup

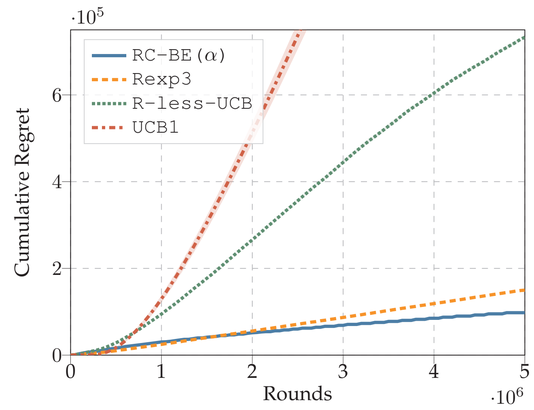
Comparison of RC-BE(α) against SOTA algorithms.

Problem Instance - Arms



Results

We compare the algorithms in terms of **empirical cumulative regret**



Key Finding: we observe that RC-BE(α) is the algorithm that achieves the **lowest regret** at the horizon.

References

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.

Su Jia, Qian Xie, Nathan Kallus, and Peter I. Frazier. Smooth non-stationary bandits. In *International Conference on Machine Learning (ICML)*, 2023.

Alberto M. Metelli, Francesco Trovò, Matteo Pirola, and Marcello Restelli. Stochastic rising bandits. In *International Conference on Machine Learning (ICML)*, 2022.