Learning Upper–Lower Value Envelopes to Shape Online RL

A Principled Approach

Authors

Sebastian Reboul
Hélène Halconruy

Domains

Reinforcement Learning

Probability Theory

Key Words

Regret Bounds

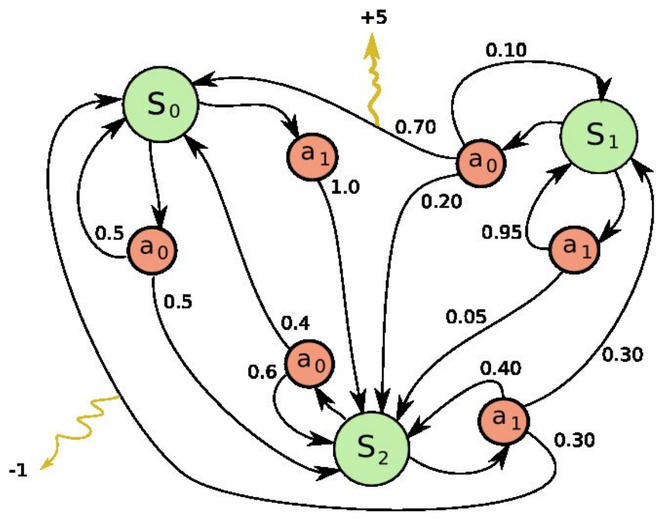
Concentration Inequalities

Partners

Wiremind Cargo

Link to Preprint

arXiv:
2510.19528



Setting

Markov Decision Process (MDP) learning Algorithm.
Objective : Produce a sequence of policies that minimize cumulative regret.

Regret

$$V_h^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{t=h}^{H} r_t(s_t, a_t) \mid s_h = s\right]$$

$$\mathrm{Regret}(T) = \sum_{t=1}^{T}\left(V_1^{*}(s_1^t) - V_1^{\pi^t}(s_1^t)\right)$$

Online RL

1. Interactive
2. Can adapt strategy on the fly
3. Does not have outside help from another player

Offline RL

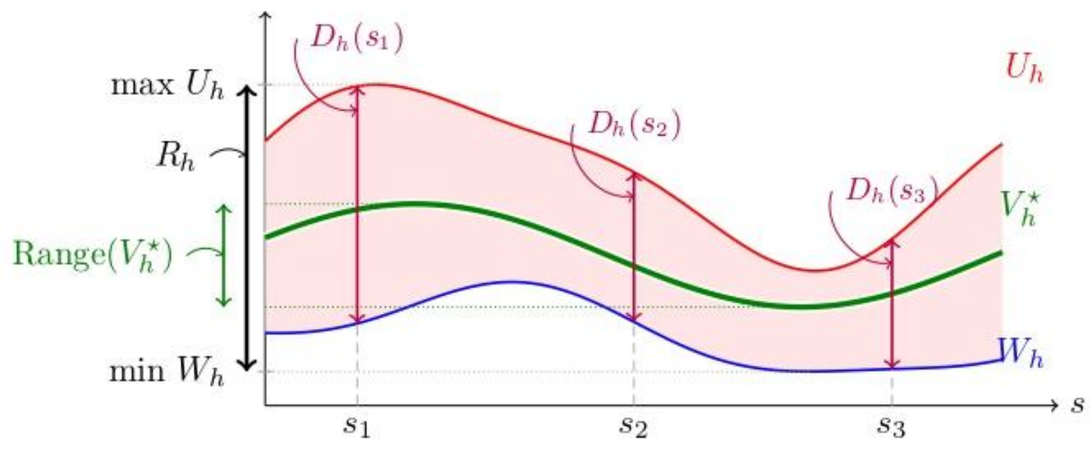
1. Non Interactive
2. Access to games from a single player
3. Works great **if** that player is good

Offline-to-Online

1. First get data from **any** player (K games)
2. Extract information (privacy preserving)
3. Use it to accelerate online learning.

Framework

Run an offline algorithm to produce **Value Envelopes** (high probability bounds on the value function)



Strategy: Pick actions maximising :

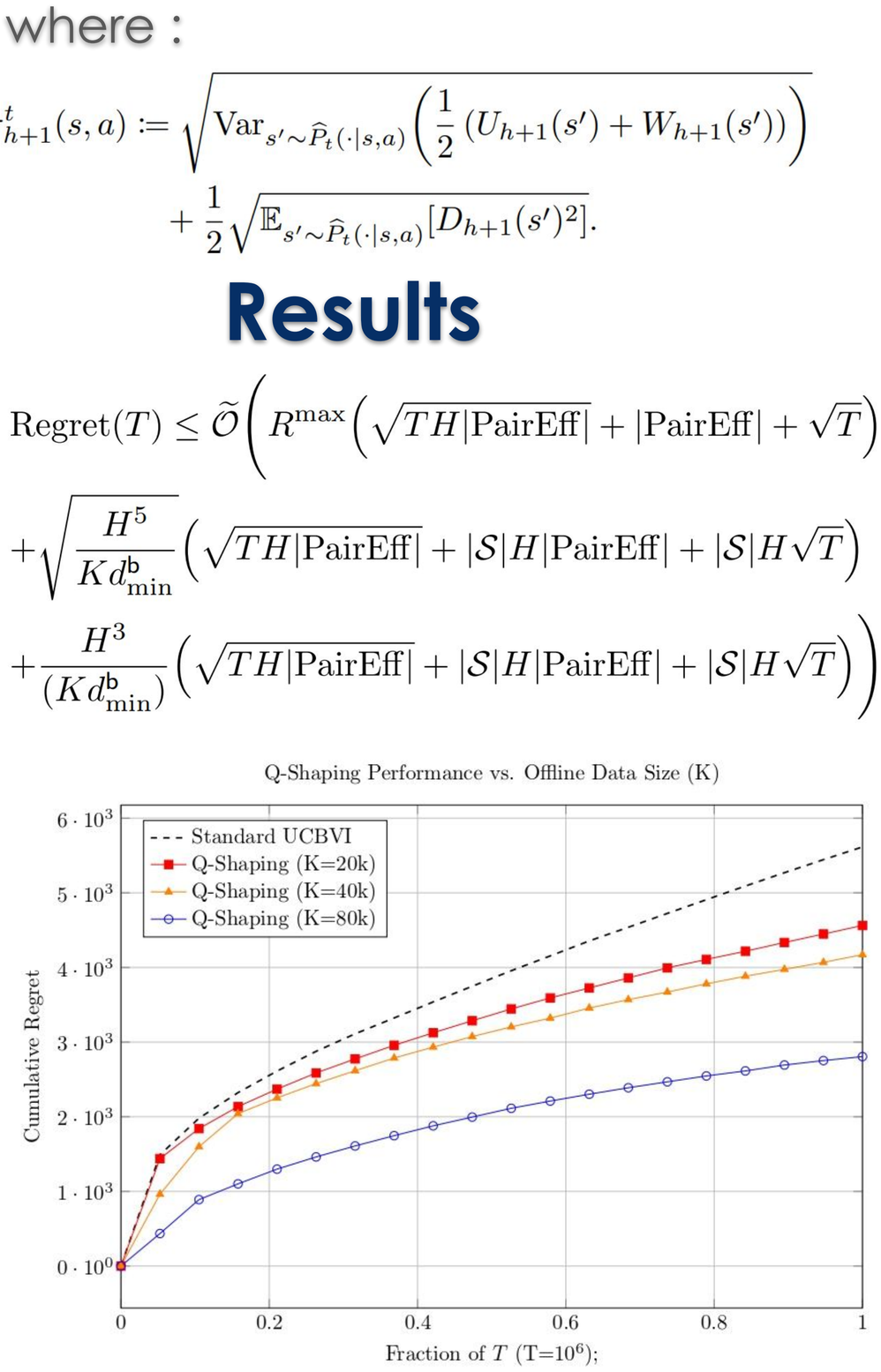
$$b_h^{t,\mathrm{on}}(s,a) = c_1\sigma_{h+1}^t(s,a)\sqrt{\frac{L}{N_h^t(s,a)}} + c_2\frac{R_{h+1}L}{N_h^t(s,a)}$$

where :

$$\sigma_{h+1}^t(s,a) := \sqrt{\mathrm{Var}_{s'\sim\widehat{P}_t(\cdot|s,a)}\left(\frac{1}{2}\left(U_{h+1}(s') + W_{h+1}(s')\right)\right)} + \frac{1}{2}\sqrt{\mathbb{E}_{s'\sim\widehat{P}_t(\cdot|s,a)}[D_{h+1}(s')^2]}.$$

Results

$$\mathrm{Regret}(T) \le \widetilde{\mathcal{O}}\Bigg(R^{\max}\left(\sqrt{TH|\mathrm{PairEff}|} + |\mathrm{PairEff}| + \sqrt{T}\right) + \sqrt{\frac{H^5}{Kd_{\min}^{\mathsf{b}}}}\left(\sqrt{TH|\mathrm{PairEff}|} + |\mathcal{S}|H|\mathrm{PairEff}| + |\mathcal{S}|H\sqrt{T}\right) + \frac{H^3}{(Kd_{\min}^{\mathsf{b}})}\left(\sqrt{TH|\mathrm{PairEff}|} + |\mathcal{S}|H|\mathrm{PairEff}| + |\mathcal{S}|H\sqrt{T}\right)\Bigg)$$



PHD Day

February 2026

Contact sebastian.reboul@telecom-sudparis.eu