Evaluation
Online learning
links with optimization and games
Université Paris–Saclay

adagrad with cube-root normalization

Let $d \geqslant 1$ an integer, $\mathscr{X} \subset \mathbb{R}^d$ a nonempty closed convex set, $(u_t)_{t\geqslant 1}$ a sequence in $\mathbb{R}^d$, $x_1 \in \mathscr{X}$ and $\gamma > 0$. We define

$$x_{t+1} = \Pi_{\mathscr{X}}\left(x_t + \frac{\gamma}{\sqrt[3]{\sum_{s=1}^{t} \|u_t\|_2^2}} u_t\right), \qquad t \geqslant 1.$$

1) Establish a general regret bound.

2) Deduce guarantees in the context of nonsmooth convex optimization and smooth convex optimization.

3) Extend the analysis to arbitrary exponents $\alpha > 0$ at the denominator:

$$x_{t+1} = \Pi_{\mathscr{X}}\left(x_t + \frac{\gamma}{\left(\sum_{s=1}^{t} \|u_t\|_2^2\right)^{\alpha}} u_t\right), \qquad t \geqslant 1.$$

4) Bonus. — Define and analyze similar variants of AdaGrad-Diagonal.