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

adagrad-diagonal: stronger adaptivity to smoothness

Let $d \geqslant 1$, $f : \mathbb{R}^d \to \mathbb{R}$ a convex function that admits a global minimizer $x_* \in \mathbb{R}^d$, in other words

$$f(x_*) = \min_{x \in \mathbb{R}^d} f(x).$$

Let M be a symmetric positive definite matrix of size d, $L > 0$ and assume that f is L-smooth for $\|\cdot\|_M$.

1) Prove that for all $x \in \mathbb{R}^d$,

$$\frac{1}{2L} \|\nabla f(x)\|^2_{M^{-1}} \leqslant f(x) - f(x_*).$$

2) Let $\gamma > 0$, $x_1 \in \mathbb{R}^d$ and for $t \geqslant 1$, define x_{t+1} as

$$x_{t+1,i} = x_{t,i} - \frac{\gamma}{\sqrt{\sum_{s=1}^t g_{s,i}^2}} g_{t,i}, \qquad 1 \leqslant i \leqslant d, \quad t \geqslant 1,$$

where $g_t = \nabla f(x_t)$, and with convention $0/0 = 0$.

Assume that M is diagonal.

a) Let $T \geqslant 1$. Prove that

$$\sum_{t=1}^{T} \langle g_t, x_t - x_* \rangle \leqslant \left(\frac{\max_{1 \leqslant t \leqslant T} \|x_t - x_*\|_\infty^2}{2\gamma} + \gamma \right) \sum_{i=1}^{d} \sqrt{\sum_{t=1}^{T} g_{t,i}^2},$$

b) For the minimization of f, derive a guarantee that is adaptive to the smoothness of f (for $\|\cdot\|_M$).

3) Bonus. — Prove that $(x_t)_{t \geqslant 1}$ is bounded.

4) Bonus. — In a context of stochastic smooth convex optimization, derive a guarantee that is adaptive to both smoothness and noise.