Evaluation

Online learning

links with optimization and games

Université Paris–Saclay

adagrad with normalization

Let $d \geqslant 1$, $(u_t)_{t\geqslant 1}$ a sequence in $\mathbb{R}^d$, $x_1 \in \mathbb{R}^d$, $\alpha \geqslant 0$ and $\gamma > 0$. Consider

$$x_{t+1} = x_t + \frac{\gamma}{\sqrt{\sum_{s=1}^{t} \|u_t\|^{2(1-\alpha)}}} \frac{u_t}{\|u_t\|^{\alpha}}, \qquad t \geqslant 1$$

1) Prove that $(x_t)_{t\geqslant 1}$ corresponds to AdaGrad-Norm iterates with some alternative payoff vectors.

2) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable convex function that admits a global minimizer $x_* \in \mathbb{R}^d$. Consider the iterates $(x_t)_{t\geqslant 1}$ defined as above with $u_t = -\nabla f(x_t)$. Let $T \geqslant 1$ and

$$\bar{x}_T = \frac{\sum_{t=1}^{T} \|\nabla f(x_t)\|^{-\alpha} x_t}{\sum_{t=1}^{T} \|\nabla f(x_t)\|^{-\alpha}}.$$

   a) Establish an upper bound on $f(\bar{x}_T) - f(x_*)$.

   b) Let $G > 0$ and assume in this question that f is G-Lipschitz continuous for $\|\cdot\|_2$. Establish convergence guarantees in the cases $\alpha = 1$ and $\alpha = 2$.

c) Let $L > 0$ and assume in this question that f is L-smooth for $\|\cdot\|_2$. Establish convergence guarantees in the cases $\alpha = 1$ and $\alpha = 2$.

d) What guarantees can you establish for other values of α?

3) Bonus. — Define and analyze a variant of AdaGrad-Diagonal with normalization.