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

a finer analysis for convex optimization with step-sizes

Let $f : \mathbb{R}^d \to \mathbb{R}$ a differentiable convex function that admits a global minimizer $x_* \in \mathbb{R}^d$, $(\gamma_t)_{t \geqslant 1}$ a positive nonincreasing sequence and $x_1 \in \mathbb{R}^d$. First consider gradient descent

$$x_{t+1} = x_t - \gamma_t \nabla f(x_t), \qquad t \geqslant 1.$$

1) Prove that for all $t \geqslant 1$,

$$0 \leqslant \|x_t - x_*\|_2^2 - \|x_{t+1} - x_*\|_2^2 + \gamma_t^2 \|\nabla f(x_t)\|_2^2 .$$

2) For all $t \geqslant 1$, deduce an upper bound on $\|x_{t+1} - x_*\|_2^2$.

3) For all $\mathrm{T} \geqslant 1$, deduce a more precise[1] upper bound on

$$\sum_{t=1}^{\mathrm{T}} (f(x_t) - f(x_*)).$$

[1] more precise than the analysis carried in the course for e.g. Mirror Descent with nonincreasing step-sizes

4) Extend the above to a constrained convex optimization setting and to more general algorithms (e.g. Projected Gradient Descent, Mirror Descent, UMD).

5) Deduce a finer guarnatee for AdaGrad-Norm in the context of (constrained) (Lipschitz) convex optimization.

6) Bonus. — Inspired by the above, can you define a variant of AdaGrad-Norm (possibly with step-sizes being nonincreasing on successive time intervals only) with improved guarantees in the context of Lipschitz convex optimization?