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

A generalized approach for nonsmooth convex optimization

Let $f:\mathbb{R}^d \to \mathbb{R}$ be a convex function and $\mathscr{X} \subset \mathbb{R}^d$ a nonempty closed convex set, such that there exists $x_* \in \mathscr{X}$ satisfying

$$f(x_*) = \min_{x\in\mathscr{X}} f(x).$$

1) Prove that for all $x, x' \in \mathbb{R}^d$, $y \in \partial f(x)$ and $y' \in \partial f(x')$,

$$\langle y' - y, x' - x\rangle \geqslant 0.$$

2) Let $(x_t)_{t\geqslant 1}$ be a sequence in $\mathbb{R}^d$, $(\beta_t)_{t\geqslant 1}$ a sequence in $[0,1]$, $(\gamma_t)_{t\geqslant 1}$ a positive sequence and for $t \geqslant 1$, consider

$$\Gamma_t = \sum_{s=1}^{t} \gamma_s, \qquad \bar{x}_t = \frac{\sum_{s=1}^{t} \gamma_s x_s}{\Gamma_t}, \qquad z_t = \beta_t \bar{x}_t + (1-\beta_t) x_t, \qquad \text{and} \qquad g_t \in \partial f(z_t).$$

Let $t \geqslant 1$.

a) Prove that

$$\bar{x}_t - \bar{x}_{t-1} = \frac{\gamma_t}{\Gamma_{t-1}}(x_t - \bar{x}_t) \quad \text{and} \quad x_t - z_t = \frac{\beta_t}{1-\beta_t}(z_t - \bar{x}_t).$$

b) Prove that

$$\Gamma_t f(\bar{x}_t) - \Gamma_{t-1} f(\bar{x}_{t-1}) - \gamma_t f(x_*) \leqslant \langle \gamma_t g_t, x_t - x_* \rangle \,.$$

Indications: make $f(z_t)$ appear, introduce $\tilde{g}_t \in \partial f(\bar{x}_t)$ and use Question 1 with points $\bar{x}_t$ and z_t.

c) Deduce that for $T \geqslant 1$,

$$\min_{1 \leqslant t \leqslant T} f(x_t) - f(x_*) \leqslant \frac{\sum_{t=1}^{T} \langle \gamma_t g_t, x_t - x_* \rangle}{\Gamma_T}.$$

3) Let $\|\cdot\|$ be a norm in $\mathbb{R}^d$ and assume that f is L-Lipschitz continuous for $\|\cdot\|$. Using the tools from the course and the previous question, define algorithms for the minimization of f and derive guarantees.

4) Extend to *stochastic* nonsmooth convex optimization.

5) Perform numerical experiments for e.g. an SVM with a moderate size dataset and compare the performance of various choices for the sequences $(\gamma_t)_{t\geqslant 1}$ and $(\beta_t)_{t\geqslant 1}$.