# Evaluation

# Online learning

## links with optimization and games

Université Paris–Saclay

### an alternative approach for smooth convex optimization

Let $\mathscr{X} \subset \mathbb{R}^d$ be closed convex set, $f : \mathbb{R}^d \to \mathbb{R}$ a differentiable convex function that admits a minimizer $x_*$ on $\mathscr{X}$, meaning:

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

We define an class of algorithms for solving this problem.

Let H be a mirror map compatible with $\mathscr{X}$, $(\eta_t)_{t \geqslant 1}$ a positive sequence, $h_1 = \eta_1^{-1}(\mathrm{H} - \min_{\mathscr{X}} \mathrm{H}) + \mathrm{I}_{\mathscr{X}}$ and

$$h_{t+1/2} = h_{t+1} = \frac{\mathrm{H} - \min_{\mathscr{X}} \mathrm{H}}{\eta_{t+1}} + \mathrm{I}_{\mathscr{X}}, \qquad t \geqslant 1.$$

Let $((x_t, y_t))_{t \geqslant 1}$ be a UMD sequence associated with regularizers $(h_t)_{t \in 1 + \frac{1}{2}\mathbb{N}}$ and sequence $(-\nabla f(x_t))_{t \geqslant 1}$ satisfying $y_1 = 0$ and for all $t \geqslant 1$,

$$\forall x \in \operatorname{dom} h, \qquad \langle y_{t+1} - y_t + \nabla f(x_t), x - x_{t+1} \rangle \geqslant 0.$$

1) Explicitely write at least two different sequences $(x_t)_{t \geqslant 1}$ that satisfy the above.

2) Assume that for all $t \geqslant 1$,

$$D_f(x_{t+1}, x_t) \leqslant h_{t+1}(x_{t+1}) - h_t(x_t) - \langle y_t, x_{t+1} - x_t \rangle \,. \tag{1}$$

Let $T \geqslant 1$ and $\bar{x}_{T+1} = \frac{1}{T}\sum_{t=2}^{T+1} x_t$. Establish a guarantee on

$$f\left(\bar{x}_{T+1}\right) - f(x_*).$$

*Hint: Get some inspiration from Proposition 6.2.1 from the lecture notes.*

3) Let $\|\cdot\|$ be a norm and $K, L > 0$. We now assume that for $\|\cdot\|$, $f$ is L-smooth and $h$ is K-strongly convex.

a) Deduce a guarantee in this case for a simple choice of $(\eta_t)_{t\geqslant 1}$.

b) Let $T \geqslant 1$ be given. Describe a possible trial-and-error heuristic for choosing, given $x_t$, a suitable value for $\eta_{t+1}$. It could start with a low and safe value for $\eta_1$ and then progressively increase so that $\eta_T$ ends up getting as close as possible to

$$\frac{D_h(x_{T+1}, x_T;\ \eta_T y_T)}{D_f(x_{T+1}, x_T)}$$

while ensuring that condition (1) always remains satisfied for all $1 \leqslant t \leqslant T$.

c) Now consider for $h$ the Euclidean regularizer on $\mathcal{X}$. Perform numerical experiments for e.g. a logistic regression (or propose another smooth objective function). Derive a smoothness constant for the objective function. Compare the simple choice $(\eta_t)_{t\geqslant 1}$ from question a) and the heuristic from question b). Compare the case $\mathcal{X} = \mathbb{R}^d$ and the case where $\mathcal{X}$ is a closed Euclidean ball. In the latter case, compare the two different algorithms written in question 1).