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

composite optimization

We consider a convex "composite" optimization problem where the objective function writes $f + \phi$, with $f : \mathbb{R}^d \to \mathbb{R}$ being a convex differentiable function, and $\phi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ a proper, convex, lower semicontinuous function. We assume that the proximal operator associated with ϕ:

$$\mathrm{Prox}_{\phi}(x) = \operatorname*{arg\,min}_{x' \in \mathbb{R}^d} \left\{ \phi(x') + \frac{1}{2} \|x' - x\|_2^2 \right\}, \qquad x \in \mathbb{R}^d.$$

is easily computable.

Exemple. — If $\phi = \lambda \|x\|_1$ for some $\lambda > 0$,

$$\mathrm{Prox}_{\phi}(x)_i = \begin{cases} x_i - \lambda & \text{if } x_i \geqslant \lambda \\ 0 & \text{if } |x_i| \leqslant \lambda \\ x_i + \lambda & \text{if } x_i \leqslant -\lambda \end{cases}, \qquad 1 \leqslant i \leqslant d,\ x \in \mathbb{R}^d.$$

The most popular algorithm for this setting is the *proximal gradient method*, aka *forward-backward splitting*: let $x_0 \in \mathbb{R}^d$ such that $\partial\phi(x_0) \neq \varnothing$, $(\gamma_t)_{t \geqslant 0}$ a positive sequence, and

$$x_{t+1} = \mathrm{Prox}_{\gamma_t \phi}\left(x_t - \gamma_t \nabla f(x_t)\right), \qquad t \geqslant 0.$$

1) Prove that the proximal gradient method is an extension of projected gradient descent.

2) Prove that the proximal gradient method is an instance of UMD iterates. *Hint. — For $t \geqslant 1$, consider $h_t(x) = \frac{1}{2}\|x\|_2^2 + \gamma_t \mathrm{D}_\phi(x, x_t;\ g_t)$ ($x \in \mathbb{R}^d$), where $g_t \in \partial\phi(x_t)$.*

3) $\mathrm{L} > 0$. In the case where f is L-smooth for $\|\cdot\|_2$, using the tools from the course, establish for the proximal gradient method a convergence guarantee that extends the classical guarantee for projected gradient descent.

4) Let $\|\cdot\|$ be an arbitrary norm in $\mathbb{R}^d$ and assume that f is L-smooth for $\|\cdot\|$. Propose for this context an algorithm that extends the proximal gradient method and its convergence guarantee.

5) Propose an alternative algorithm for the context of the previous question and derive corresponding guarantees.