Modern Optimization

Jean-Luc Bouchot

School of Mathematics and Statistics
Beijing Institute of Technology
jlbouchot@bit.edu.cn

Spring 2021

1. Constrained optimization: Projected methods

Outline

Remark 2.1

Let $\Omega \subset \mathbb{R}^d$ be a closed convex body, we are interested in the following optimization problems:

$$\min f_0(x)$$
$$\text{s.t. } x \in \Omega.$$

Definition 2.1

The **projected gradient descent** *is defined as the following sequence of rules*

$$y^{k+1} = x^k - \gamma_k \nabla f(x^k)$$
$$x^{k+1} = P_\Omega(y^{k+1}),$$

where P_Ω corresponds to the projection onto the convex set Ω (POCS: Projection Onto Convex Sets is not unrelated...).

Remark 2.2

Let us review the main ideas behind the projected gradient descent:

- The first step is a basic gradient step
- The second step corrects the gradient step if it reaches a point out of the feasible set
- γ_k the step size may or may not vary
- The projection step might not be cheap!:

$$P_\Omega(x) := \underset{v \in \Omega}{\operatorname{argmin}} \|x - v\|.$$

Proposition 2.1 (Left as exercise)

Let Ω be a closed convex body. Then the projection P_Ω are well defined (and unique for all $x \in \mathbb{R}^d$).

Remark 2.3

Note that the first step assumes a point $x^0 \in \Omega$. This means simply doing a first projection on the input point (or even 0).

Exercise 2.1

Let $\Omega = B^2_{x^*}(R)$ be the ℓ^2 ball centered at a given point $x^* \in \mathbb{R}^d$ and of radius $R > 0$. What is $P_\Omega(x)$ for any $x \in \mathbb{R}^d$.

Lemma 1

Let $\Omega \subseteq \mathbb{R}^d$ *be a closed convex body. Let* $x \in \Omega$ *and* $y \in \mathbb{R}^d$*. It holds*

1. $\langle x - P_\Omega(y), y - P_\Omega(y) \rangle \leq 0$.
2. $\|x - P_\Omega(y)\|^2 + \|y - P_\Omega(y)\|^2 \leq \|x - y\|^2$.

Theorem 2.1

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ *be a convex differentiable function. Assume furthermore that* $\Omega \subseteq \mathrm{dom}(f)$ *is a closed convex subset,* x^* *is a minimizer of* f *over* Ω*,* $\|x^0 - x^*\| \leq R$ *for some* $R > 0$ *and* $x^0 \in \Omega$*. If the gradient of* f *is bounded:* $\|\nabla f(x)\| \leq B$ *for all* $x \in \Omega$*, then choosing a gradient step of*

$$\gamma := \frac{R}{B\sqrt{K}}$$

ensures that the iterates generated by the projected gradient descent starting at x^0 *satisfy*

$$\frac{1}{K} \sum_{k=0}^{K-1} \Big(f(x^k) - f(x^*) \Big) \leq \frac{RB}{\sqrt{K}}.$$

Lemma 2 (Descent direction of projected gradient)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be a convex differentiable L-smooth function over a closed and convex set $\Omega \subseteq \mathrm{dom}(f)$. Given a constant stepsize

$$\gamma = \frac{1}{L},$$

the sequence of iterates of the projected gradient, starting at $x^0 \in \Omega$ satisfies

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|^2 + \frac{L}{2}\|y^{k+1} - x^{k+1}\|^2.$$

Lemma 3

Let x^k be the sequence of iterates generated by the projected gradient descent of an L-smooth convex differentiable function f over a closed convex domain Ω. Then, using a fixed gradient step

$$\gamma = \frac{1}{L}$$

we have

$$f(x^{k+1}) \leq f(x^k) - \frac{L}{2}\|x^{k+1} - x^k\|^2$$

Theorem 2.2

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ be a convex differentiable L-smooth function over a closed and convex set $\Omega \subseteq \mathrm{dom}(f)$ and assume the existence of a minimizer $x^* \in \Omega$ of f in Ω. Given a constant stepsize

$$\gamma = \frac{1}{L},$$

the sequence of iterates of the projected gradient, starting at $x^0 \in \Omega$ satisfies

$$f(x^K) \leq f(x^*) + \frac{L}{2K}\|x^0 - x^*\|^2, \quad K > 0.$$