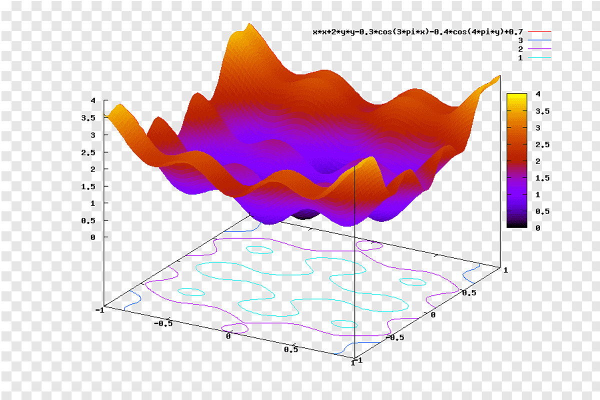

LINMA2471 Optimization models and methods II

Simon Desmidt

Academic year 2024-2025 - Q1

UCLouvain

Table des matières

Gradient Method

An optimization problem is defined as

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function.

1.1 Definitions

— A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous when

$$\|F(y) - F(x)\| \leq L\|y - x\| \qquad \forall x, y \in \mathbb{R}^n$$

where we use the euclidian norm.

— If ∇f is L-Lipschitz then, given $x \in \mathbb{R}^n$,

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 = m_x(y) \qquad \forall y \in \mathbb{R}^n$$

and f is said to be a L-smooth function.

— We say that a differentiable function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is L-smooth for some $L \geq 0$ when, given $x \in \mathbb{R}^n$,

$$\Psi(y) \leq \Psi(x) + \langle \nabla \Psi(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 \qquad \forall y \in \mathbb{R}^n$$

— A convex function $f : \mathbb{R}^n \to \mathbb{R}$ is convex when, given $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

— Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. If f is differentiable, then

$$f(y) \geq f(x) + \nabla f(x)^T(x - y) \qquad \forall x, y \in \mathbb{R}^n$$

— A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex $(\mu > 0)$ if, given $x \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2}\|y - x\|^2 \qquad \forall y \in \mathbb{R}^n$$

— PL inequality for a μ-strongly convex function[1] :

$$f(x) - f(x^*) \leq \frac{1}{2\mu}\|\nabla f(x)\|^2 \qquad \forall x \in \mathbb{R}^n$$

1. x^* is the minimizer of f

1.2 Complexity

The demonstration of the final results here obtained is in the notes, but not explained here.

1.2.1 Hypotheses

— f is convex and differentiable ;
— ∇f is L-Lipschitz ;
— we start from a $x_0 \in \mathbb{R}^n$ that is not a minimizer of f ;

1.2.2 Results

We use the sequence $\{x_k\}_{k\geq 0}$, given a $x_0 \in \mathbb{R}^n$, such that

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$$

Problem class	Goal	Complexity bound
Non-convex f	$\|\nabla f(x_k)\| \leq \varepsilon$	$\mathcal{O}(\varepsilon^{-2})$
Convex f	$f(x_k) - f(x^*) \leq \varepsilon$	$\mathcal{O}(\varepsilon^{-1})$
μ-strongly-convex f	$f(x_k) - f(x^*) \leq \varepsilon$	$\mathcal{O}(\log(\varepsilon^{-1}))$

1.3 GM with Armijo Line Search

The Armijo Line Search consists of changing the constant in the GM in order to be more efficient and be able to make bigger steps in some directions where it is possible.

$$x_{k+1} = x_k - \alpha\nabla f(x_k) \qquad \alpha > 0 \tag{1.2}$$

Algorithm 1 Gradient Method with Armijo Line Search

1: **Step 0 :** Given $x_0 \in \mathbb{R}^n$ and $\alpha_0 > 0$, set $k := 0$.
2: **Step 1 :** Set $\ell := 0$.
3: **Step 1.1 :** Compute $x_k^+ = x_k - (0.5)^\ell \alpha_k \nabla f(x_k)$.
4: **Step 1.2 (Armijo Line Search) :** If

$$f(x_k) - f(x_k^+) \geq \frac{(0.5)^\ell \alpha_k}{2}\|\nabla f(x_k)\|^2 \tag{1}$$

set $\ell_k := \ell$ and go to Step 2. Otherwise, set $\ell := \ell + 1$ and go back to Step 1.1.
5: **Step 3 :** Define $x_{k+1} = x_k^+$, $\alpha_{k+1} = (0.5)^{\ell_k - 1}\alpha_k$, set $k := k+1$ and go back to Step 1.

1.4 Problems with convex constraints

Consider the problem

$$\min_{x\in\mathbb{R}^n} f(x) \text{ such that } x \in \Omega \tag{1.3}$$

where f is L-smooth, and $\Omega \subseteq \mathbb{R}^n$ is nonempty, closed and convex. Given an approximation $x_k \in \Omega$ for a solution of 1.3, a possible generalization of the Gradient Method is to define

$$x_{k+1} = P_\Omega\left(x_k - \frac{1}{L}\nabla f(x_k)\right) \tag{1.4}$$

where P_Ω is the projection of z onto Ω, and we call this method the Projected Gradient Method.
If $\Omega = [a,b]^n$, then the projection of an element z onto Ω is such that its element i is given by :

$$[P_\Omega(z)]_i = \begin{cases} z_i \text{ if } a \leq z_i \leq b \\ a \text{ if } z_i < a \\ b \text{ if } z_i > b \end{cases} \qquad \forall i = 1, \dots, n \tag{1.5}$$

If x^* is a solution of (1.3), then

$$\langle \nabla f(x^*), z - x^* \rangle \geq 0 \qquad \forall z \in \Omega$$

→ N.B. : if $\Omega = \mathbb{R}^n$, then this lemma is true, in particular for $z = x^* - \nabla f(x^*)$. Then it is straightforward that we must have $\nabla f(x^*) = 0$.

1.5 Reduced gradient method

For a L-smooth function for the problem (1.3), we define

$$G_L(x_k) = L(x_k - x_{k+1}) \tag{1.6}$$

where x_{k+1} is given by the general formula[2]

$$x_{k+1} = \arg\min_{y\in\Omega} f(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \frac{L}{2}|y - x_k\|_2^2 \tag{1.7}$$

From this, we can show as we did in the previous sections that there is a lower bound for the method :

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|G_L(x_k)\|_2^2 \tag{1.8}$$

This is the same result we found for the unconstrained gradient method, but with a different gradient definition. This is thus a generalization of the first cases. Furtheremore, by the same process we used before, we can show that the complexity of this Reduced Gradient Method is the same as in the table 1.2.2.

2. This definition of x_{k+1} is true for any type of gradient method, the first case seen being with $\Omega = \mathbb{R}^n$.

1.6 Proximal Gradient Method

We will here consider problems of the form

$$\min_{x \in \mathbb{R}^n} F(x) \equiv f(x) + \phi(x) \tag{1.9}$$

where $f(\cdot)$ is L-smooth and $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, possibly nonsmooth. In this case, the formula for x_{k+1} is

$$x_{k+1} = \arg\min_{y \in \mathbb{R}^n} f(x_k) + \langle \nabla f(x_k), y - x_k \rangle + \frac{L}{2}|y - x_k|_2^2 + l(y) \tag{1.10}$$

which can be re-expressed as

$$x_{k+1} = \arg\min_{y \in \mathbb{R}^n} \frac{1}{2}\|y - (x_k - \frac{1}{L}\nabla f(x_k))\|^2 + \frac{1}{L}l(y) \tag{1.11}$$

Given a convex function h, we define the proximal operator $prox_h : \mathbb{R}^n \to \mathbb{R}^n$ by

$$prox_h(z) = \arg\min_{y \in \mathbb{R}^n} \frac{1}{2}\|y - z\|^2 + h(y) \tag{1.12}$$

Then, we can write

$$x_{k+1} = prox_{\frac{1}{L}\phi}\left(x_k - \frac{1}{L}\nabla f(x_k)\right) \tag{1.13}$$

→ N.B. : if the ϕ function is the indicator function, i.e. $\phi = i_\Omega = \begin{cases} 0 \text{ if } x \in \Omega \\ \infty \text{ otherwise} \end{cases}$,

then $prox_{\frac{1}{L}i_\Omega}(z) = P_\Omega(z)$.