Introduction to Optimization

Lecture 09: Simple constraints and projected gradient. Nonsmooth convex functions and subgradient method.

university of groningen

A quick reminder

Let $C \subset \mathbb{R}^N$ be nonempty, closed and convex. For each $x \in \mathbb{R}^N$, the projection of x onto C, denoted by $P_C(x)$, is the point in C, which is the closest to x.

A quick reminder

Let $C \subset \mathbb{R}^N$ be nonempty, closed and convex. For each $x \in \mathbb{R}^N$, the projection of x onto C, denoted by $P_C(x)$, is the point in C, which is the closest to x.

1. The positive orthant $\mathbb{R}^N_+$.

A quick reminder

Let $C \subset \mathbb{R}^N$ be nonempty, closed and convex. For each $x \in \mathbb{R}^N$, the projection of x onto C, denoted by $P_C(x)$, is the point in C, which is the closest to x.

1. The positive orthant $\mathbb{R}^N_+$. ✍
2. Balls $\bar{B}(x_0, r)$ and boxes $\prod[a_i, b_i]$. ✍

A quick reminder

Let $C \subset \mathbb{R}^N$ be nonempty, closed and convex. For each $x \in \mathbb{R}^N$, the projection of x onto C, denoted by $P_C(x)$, is the point in C, which is the closest to x.

1. The positive orthant $\mathbb{R}^N_+$. ✍
2. Balls $\bar{B}(x_0, r)$ and boxes $\prod [a_i, b_i]$. ✍
3. Affine spaces, such as $\{x_0\} + V$, or $\{x \in \mathbb{R}^N : Ax = b\}$. ✍

The projected gradient method

Consider the problem

$$\min_{x \in C} f(x),$$

where

- $f : \mathbb{R}^N \to \mathbb{R}$, and
- $C \subset \mathbb{R}^N$ is nonempty, closed, convex and easy to project onto.

The set of solutions is denoted by S, and the optimal value by f^*.

The projected gradient method

Consider the problem

$$\min_{x \in C} f(x),$$

where

- $f : \mathbb{R}^N \to \mathbb{R}$, and
- $C \subset \mathbb{R}^N$ is nonempty, closed, convex and easy to project onto.

The set of solutions is denoted by S, and the optimal value by f^*.

Idea: At each iteration, perform a gradient step and then project onto C.

$$x_{n+1} = P_C\big(x_n - \alpha_n \nabla f(x_n)\big).$$

Convergence of the projected gradient method

Theorem

Let f be convex and L-smooth, and suppose $S \neq \emptyset$. Iterate $x_{n+1} = P_C\big(x_n - \alpha \nabla f(x_n)\big)$ with $0 < \alpha \leq \frac{1}{L}$.

Convergence of the projected gradient method

Theorem

Let f be convex and L-smooth, and suppose $S \neq \emptyset$. Iterate $x_{n+1} = P_C\big(x_n - \alpha\nabla f(x_n)\big)$ with $0 < \alpha \leq \frac{1}{L}$.

1. *x_n converges to a point in S, and*

$$f(x_n) - f^* \leq \frac{\mathrm{dist}(x_0, S)^2}{2\alpha n}.$$

2. *If f is $\mu-$strongly convex, then*

$$f(x_n) - f^* \leq \frac{f(x_0) - f^*}{(1 + \alpha\mu)^n}.$$

Break

Nonsmooth convex functions

Let $f : A \subset \mathbb{R}^N \to \mathbb{R}$ be convex. If f is differentiable at x, then

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

for all $y \in A$.

Nonsmooth convex functions

Let $f : A \subset \mathbb{R}^N \to \mathbb{R}$ be convex. If f is differentiable at x, then

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

for all $y \in A$.

A vector $v \in \mathbb{R}^N$ is a subgradient of f at x if

$$f(y) \geq f(x) + v \cdot (y - x)$$

for all $y \in A$.

Nonsmooth convex functions

Let $f : A \subset \mathbb{R}^N \to \mathbb{R}$ be convex. If f is differentiable at x, then

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

for all $y \in A$.

A vector $v \in \mathbb{R}^N$ is a subgradient of f at x if

$$f(y) \geq f(x) + v \cdot (y - x)$$

for all $y \in A$. The subdifferential of f at x, denoted by $\partial f(x)$, is the set of all the subgradients of f at x.

A few simple but important examples

Example (A big ice cream cone)

Let $f : \mathbb{R}^N \to \mathbb{R}$ be defined by $f(x) = \|x\|$.

A few simple but important examples

Example (A big ice cream cone)

Let $f : \mathbb{R}^N \to \mathbb{R}$ be defined by $f(x) = \|x\|$.

Example (The ℓ^1-norm)

Let $\| \cdot \|_1 : \mathbb{R}^N \to \mathbb{R}$ be defined by $\|x\|_1 = |x_1| + \cdots + |x_N|$.

A few simple but important examples

Example (A big ice cream cone)

Let $f : \mathbb{R}^N \to \mathbb{R}$ be defined by $f(x) = \|x\|$.

Example (The ℓ^1-norm)

Let $\|\cdot\|_1 : \mathbb{R}^N \to \mathbb{R}$ be defined by $\|x\|_1 = |x_1| + \cdots + |x_N|$.

Example (The indicator function)

Let $C \subset \mathbb{R}^N$ be nonempty, closed and convex. The indicator function of C is the function $\iota_C : C \to \mathbb{R}$, defined as $\iota_C(x) = 0$ for all $x \in C$.

Convexity, continuity and subdifferentiability

Proposition

If $f : A \subset \mathbb{R}^N \to \mathbb{R}$ *is convex, for each* $x \in \mathrm{int}(A)$*, there exist* $L_x, r_x > 0$ *such that*

$$|f(z) - f(y)| \leq L_x \|z - y\|$$

for all $z, y \in B(x, r_x)$*. Moreover,*

$$\emptyset \neq \partial f(x) \subseteq \bar{B}(0, L_x).$$

Convexity, continuity and subdifferentiability

Proposition

If $f : A \subset \mathbb{R}^N \to \mathbb{R}$ *is convex, for each* $x \in \text{int}(A)$*, there exist* $L_x, r_x > 0$ *such that*

$$|f(z) - f(y)| \leq L_x \|z - y\|$$

for all $z, y \in B(x, r_x)$*. Moreover,*

$$\emptyset \neq \partial f(x) \subseteq \bar{B}(0, L_x).$$

Convex functions are continuous and subdifferentiable in the interior of their domains.

The subgradient method

$x_{n+1} = x_n - \alpha v_n$ with $v_n \in \partial f(x_n)$

The subgradient method

$x_{n+1} = x_n - \alpha v_n$ with $v_n \in \partial f(x_n)$

Proposition

Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and Lipschitz-continuous with constant M,[a] with minimizers, and let (x_n) be defined by the subgradient method. Then,

$$\min_{k=1,\dots,n} \left(f(x_k) - \min(f)\right) \le \frac{\alpha M^2}{2} + \frac{dist(x_0, S)^2}{2\alpha(n+1)}.$$

[a] Recall that this means that $|f(x) - f(y)| \le M\|x - y\|$.

The subgradient method

$x_{n+1} = x_n - \alpha v_n$ with $v_n \in \partial f(x_n)$

Proposition

Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and Lipschitz-continuous with constant M,[a] *with minimizers, and let (x_n) be defined by the subgradient method. Then,*

$$\min_{k=1,\dots,n} \big(f(x_k) - \min(f)\big) \le \frac{\alpha M^2}{2} + \frac{dist(x_0, S)^2}{2\alpha(n+1)}.$$

[a] Recall that this means that $|f(x) - f(y)| \le M\|x - y\|$.

Question

Given $\varepsilon > 0$, after how many iterations can we be sure to have found a point $\hat{x}$ such that $f(\hat{x}) - \min(f) \le \varepsilon$?