Introduction to Optimization

Lecture 09: Subgradient descent.
The proximal-gradient algorithm.

university of groningen

Reminder

A vector $v \in \mathbb{R}^N$ is a subgradient of f a convex function $f : \mathrm{dom}(f) \subset \mathbb{R}^N \to \mathbb{R}$ at the point x if

$$f(y) \geq f(x) + \langle v, y - x \rangle$$

for all $y \in \mathbb{R}^N$.

Proposition

If $f : \mathrm{dom}(f) \subset \mathbb{R}^N \to \mathbb{R}$ is convex, for each $x \in \mathrm{int}(\mathrm{dom}(f))$, 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, $\emptyset \neq \partial f(x) \subseteq \bar{B}(0, L_x)$.

Proposition

Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex and Lipschitz-continuous with constant M ($|f(x) - f(y)| \leq M\|x - y\|$) with minimizers, and let (x_n) be defined by the subgradient method. Set $\bar{x}_n = \frac{1}{n+1}\sum_{k=0}^{n} x_k$. Then,

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

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 ($|f(x) - f(y)| \leq M\|x-y\|$) with minimizers, and let (x_n) be defined by the subgradient method. Set $\bar{x}_n = \frac{1}{n+1}\sum_{k=0}^{n} x_k$. Then,

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

Question

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

Extended real-valued functions, I

We extend the codomain $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$, with the conventions that $+\infty > \gamma$ for all $\gamma \in \mathbb{R}$

Extended real-valued functions, I

We extend the codomain $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$, with the conventions that $+\infty > \gamma$ for all $\gamma \in \mathbb{R}$, and some algebraic operations are allowed:

- $+\infty + \gamma = +\infty$ for all $\gamma \in \mathbb{R}$,
- $\gamma(+\infty) = +\infty$ for all $\gamma > 0$, and $0(+\infty) = 0$.

Extended real-valued functions, I

We extend the codomain $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$, with the conventions that $+\infty > \gamma$ for all $\gamma \in \mathbb{R}$, and some algebraic operations are allowed:

- $+\infty + \gamma = +\infty$ for all $\gamma \in \mathbb{R}$,
- $\gamma(+\infty) = +\infty$ for all $\gamma > 0$, and $0(+\infty) = 0$.

The (effective) domain and epigraph of a function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is

$$\begin{aligned} \mathrm{dom}(f) &= \{x \in \mathbb{R}^N \ : \ f(x) < +\infty\} \\ \mathrm{epi}(f) &= \{(x, z) \in \mathbb{R}^{N+1} \ : \ f(x) \le z\}, \end{aligned}$$

respectively. We will always assume that $\mathrm{dom}(f) \neq \emptyset$.

Extended real-valued functions, I

We extend the codomain $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$, with the conventions that $+\infty > \gamma$ for all $\gamma \in \mathbb{R}$, and some algebraic operations are allowed:

- $+\infty + \gamma = +\infty$ for all $\gamma \in \mathbb{R}$,
- $\gamma(+\infty) = +\infty$ for all $\gamma > 0$, and $0(+\infty) = 0$.

The (effective) domain and epigraph of a function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is

$$\begin{aligned} \mathrm{dom}(f) &= \{x \in \mathbb{R}^N \ : \ f(x) < +\infty\} \\ \mathrm{epi}(f) &= \{(x,z) \in \mathbb{R}^{N+1} \ : \ f(x) \leq z\}, \end{aligned}$$

respectively. We will always assume that $\mathrm{dom}(f) \neq \emptyset$.

Notice that $\mathrm{dom}(\lambda f + g) = \mathrm{dom}(f) \cap \mathrm{dom}(g)$.

Extended real-valued functions, II

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed if $\mathrm{epi}(f)$ is closed.

Extended real-valued functions, II

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed if $\mathrm{epi}(f)$ is closed.

Examples

1. Every continuous function is closed.

Extended real-valued functions, II

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed if $\mathrm{epi}(f)$ is closed.

Examples

1. Every continuous function is closed.
2. The indicator function of $C \subset \mathbb{R}^N$ ($C \neq \emptyset$) is $\iota_C : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, defined as $\iota_C(x) = 0$ if $x \in C$ and $\iota_C(x) = +\infty$ if $x \notin C$.

Extended real-valued functions, II

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed if $\text{epi}(f)$ is closed.

Examples

1. Every continuous function is closed.
2. The indicator function of $C \subset \mathbb{R}^N$ ($C \neq \emptyset$) is $\iota_C : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, defined as $\iota_C(x) = 0$ if $x \in C$ and $\iota_C(x) = +\infty$ if $x \notin C$.

   Here, $\text{dom}(\iota_C) = C$ and $\text{epi}(\iota_C) = C \times [0, +\infty)$.

Extended real-valued functions, II

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed if $\text{epi}(f)$ is closed.

Examples

1. Every continuous function is closed.
2. The indicator function of $C \subset \mathbb{R}^N$ ($C \neq \emptyset$) is $\iota_C : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, defined as $\iota_C(x) = 0$ if $x \in C$ and $\iota_C(x) = +\infty$ if $x \notin C$.

   Here, $\text{dom}(\iota_C) = C$ and $\text{epi}(\iota_C) = C \times [0, +\infty)$.

   This function is closed if C is closed, and convex if C is convex.

Extended real-valued functions, II

A function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed if $\mathrm{epi}(f)$ is closed.

Examples

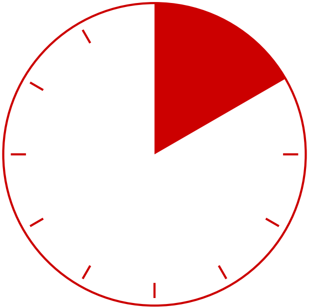
1. Every continuous function is closed.
2. The indicator function of $C \subset \mathbb{R}^N$ ($C \neq \emptyset$) is $\iota_C : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$, defined as $\iota_C(x) = 0$ if $x \in C$ and $\iota_C(x) = +\infty$ if $x \notin C$.

   Here, $\mathrm{dom}(\iota_C) = C$ and $\mathrm{epi}(\iota_C) = C \times [0, +\infty)$.

   This function is closed if C is closed, and convex if C is convex.

If $f : \mathbb{R}^N \to \mathbb{R}$, $\min\{f(x) : x \in C\} = \min\left\{f(x) + \iota_C(x) : x \in \mathbb{R}^N\right\}$.

Break



Closedness and proximity operator

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, and $y \in \mathbb{R}^N$, the function

$$f_y(x) = f(x) + \frac{1}{2}\|x - y\|^2$$

is closed and strongly convex.

Closedness and proximity operator

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, and $y \in \mathbb{R}^N$, the function

$$f_y(x) = f(x) + \frac{1}{2}\|x - y\|^2$$

is closed and strongly convex. Its subdifferential is given by

$$\partial f_y(x) = \partial f(x) + x - y,$$

for each $x \in \mathrm{dom}(f)$.

Closedness and proximity operator

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, and $y \in \mathbb{R}^N$, the function

$$f_y(x) = f(x) + \frac{1}{2}\|x - y\|^2$$

is closed and strongly convex. Its subdifferential is given by

$$\partial f_y(x) = \partial f(x) + x - y,$$

for each $x \in \mathrm{dom}(f)$. The unique minimizer of f_y is denoted by $\mathrm{prox}_f(y)$, and is characterized by

$$y - \mathrm{prox}_f(y) \in \partial f\big(\mathrm{prox}_f(y)\big).$$

The proximal method

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, and fix $\alpha > 0$. From an initial point $x_0 \in \mathbb{R}^N$, define a sequence inductively by

$$x_{n+1} = \text{prox}_{\alpha f}(x_n).$$

The proximal method

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, and fix $\alpha > 0$. From an initial point $x_0 \in \mathbb{R}^N$, define a sequence inductively by

$$x_{n+1} = \mathrm{prox}_{\alpha f}(x_n).$$

Exercise

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, and $S \neq \emptyset$, then x_n converges to a point in S. Moreover,

$$f(x_n) - \min(f) \leq \frac{\mathrm{dist}(x_0, S)^2}{2\alpha n}, \qquad n \geq 1.$$

Proximal-gradient algorithm

Suppose we want to find the minima of $f = g + h$, where $g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is convex and lower-semicontinuous, and $h : \mathbb{R}^N \to \mathbb{R}$ is convex and L-smooth.

Proximal-gradient algorithm

Suppose we want to find the minima of $f = g + h$, where $g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is convex and lower-semicontinuous, and $h : \mathbb{R}^N \to \mathbb{R}$ is convex and L-smooth.

Example

A typical example in image and signal processing, statistics, ML, is

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \rho\|x\|_1$$

for $x \in \mathbb{R}^N$.

Proximal-gradient algorithm

The proximal-gradient method consists in applying proximal iterations while linearizing the smooth function:

$$x_{n+1} = \operatorname{Argmin}\left\{g(x) + h(x_n) + \langle \nabla h(x_n), x - x_n \rangle + \frac{1}{2\gamma}\|x - x_n\|^2\right\}$$

Proximal-gradient algorithm

The proximal-gradient method consists in applying proximal iterations while linearizing the smooth function:

$$
\begin{aligned}
x_{n+1} &= \operatorname{Argmin}\left\{ g(x) + h(x_n) + \langle \nabla h(x_n), x - x_n \rangle + \frac{1}{2\gamma}\|x - x_n\|^2 \right\} \\
&= \operatorname{Argmin}\left\{ g(x) + \frac{1}{2\gamma}\left\| x - \left(x_n - \gamma \nabla h(x_n)\right)\right\|^2 \right\}.
\end{aligned}
$$

Proximal-gradient algorithm

The proximal-gradient method consists in applying proximal iterations while linearizing the smooth function:

$$
\begin{aligned}
x_{n+1} &= \operatorname{Argmin}\left\{g(x) + h(x_n) + \langle \nabla h(x_n), x - x_n\rangle + \frac{1}{2\gamma}\|x - x_n\|^2\right\} \\
&= \operatorname{Argmin}\left\{g(x) + \frac{1}{2\gamma}\left\|x - \left(x_n - \gamma\nabla h(x_n)\right)\right\|^2\right\}.
\end{aligned}
$$

This subproblem has a unique solution characterized by

$$0 \in \partial g(x_{n+1}) + \nabla h(x_n) + \frac{1}{\gamma}(x_{n+1} - x_n).$$

Proximal-gradient algorithm

Note that

$$0 \in \partial g(x_{n+1}) + \nabla h(x_n) + \frac{1}{\gamma}(x_{n+1} - x_n)$$

is equivalent to

$$x_{n+1} + \gamma \partial g(x_{n+1}) \ni x_n - \gamma \nabla h(x_n).$$

Proximal-gradient algorithm

Note that

$$0 \in \partial g(x_{n+1}) + \nabla h(x_n) + \frac{1}{\gamma}(x_{n+1} - x_n)$$

is equivalent to

$$x_{n+1} + \gamma \partial g(x_{n+1}) \ni x_n - \gamma \nabla h(x_n).$$

In turn, this can be rewritten as

$$x_{n+1} = (I + \gamma \partial g)^{-1}(I - \gamma \nabla h)x_n,$$

where we identify a gradient subiteration with respect to h, and then a proximal subiteration with respect to g.

Proximal-gradient algorithm

Note that

$$0 \in \partial g(x_{n+1}) + \nabla h(x_n) + \frac{1}{\gamma}(x_{n+1} - x_n)$$

is equivalent to

$$x_{n+1} + \gamma \partial g(x_{n+1}) \ni x_n - \gamma \nabla h(x_n).$$

In turn, this can be rewritten as

$$x_{n+1} = (I + \gamma \partial g)^{-1}(I - \gamma \nabla h)x_n,$$

where we identify a gradient subiteration with respect to h, and then a proximal subiteration with respect to g. It is a splitting method.

Convergence of proximal-gradient sequences

Theorem

Let $f = g + h$, *where* $g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ *is convex and lower-semicontinuous, and* $h : \mathbb{R}^N \to \mathbb{R}$ *is convex and L-smooth. Take* $\gamma \in (0, 2/L)$ *and define* (x_n) *by*

$$x_{n+1} = \big(I + \gamma \partial g\big)^{-1}(I - \gamma \nabla h)x_n, \qquad n \geq 0.$$

Then, x_n *converges to a minimizer of* f, *and there is* $C > 0$ *such that*

$$f(x_n) - \min(f) \leq \frac{dist(x_0, \operatorname{Argmin}(f))^2}{2\gamma n}, \qquad n \geq 1.$$

Moreover, $\lim\limits_{n \to \infty} n\big(f(x_n) - \min(f)\big) = 0.$