Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 04: Nonsmooth optimization.

university of groningen

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

Can you think of a convex function that is not continuous?

Convexity, continuity and subdifferentiability

Can you think of a convex function that is not continuous?

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

Can you think of a convex function that is not continuous?

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,

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

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 $v_n \in \partial f(x_n)$

Lemma

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. Then,

$$\frac{1}{n+1}\sum_{k=0}^{n} f(x_k) - \min(f) \leq \frac{\alpha M^2}{2} + \frac{\|x_0 - p\|^2}{2\alpha(n+1)}$$

for every $p \in S$ and $n \geq 0$.

The subgradient method

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

Corollary

Set $\bar{x}_n = \frac{1}{n+1}\sum_{k=0}^{n} x_k$. *Then,*

$$\begin{aligned} \min_{k=1,\dots,n}\big(f(x_k) - \min(f)\big) &\leq \frac{\alpha M^2}{2} + \frac{dist(x_0, S)^2}{2\alpha(n+1)} \\ f(\bar{x}_n) - \min(f) &\leq \frac{\alpha M^2}{2} + \frac{dist(x_0, S)^2}{2\alpha(n+1)} \end{aligned}$$

The subgradient method

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

Corollary

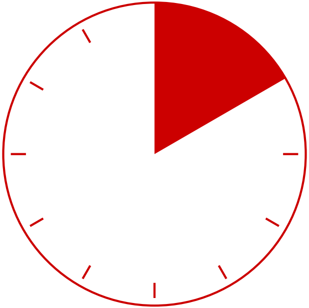
Set $\bar{x}_n = \frac{1}{n+1}\sum_{k=0}^{n} x_k$. *Then,*

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

Question

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

Break



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

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

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 $\text{epi}(f)$ 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. If f and g are closed, and $\lambda \geq 0$, then $\lambda f + g$ 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. If f and g are closed, and $\lambda \geq 0$, then $\lambda f + g$ is closed.
3. 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 $\mathrm{epi}(f)$ is closed.

Examples

1. Every continuous function is closed.
2. If f and g are closed, and $\lambda \geq 0$, then $\lambda f + g$ is closed.
3. 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 $\mathrm{epi}(\iota_C) = C \times [0, +\infty)$.

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. If f and g are closed, and $\lambda \geq 0$, then $\lambda f + g$ is closed.
3. 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.

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. If f and g are closed, and $\lambda \geq 0$, then $\lambda f + g$ is closed.
3. 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 $f : \mathbb{R}^N \to \mathbb{R}$, $\min\{f(x) : x \in C\} = \min\left\{f(x) + \iota_C(x) : x \in \mathbb{R}^N\right\}$.

Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 04: Nonsmooth optimization.

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 closed and convex, and $h : \mathbb{R}^N \to \mathbb{R}$ is L-smooth and convex.

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 closed and convex, and $h : \mathbb{R}^N \to \mathbb{R}$ is L-smooth and convex.

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:

$$x_{n+1} = \operatorname{prox}_{\alpha g}\left(x_n - \alpha \nabla h(x_n)\right)$$

Proximal-gradient algorithm

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

$$\begin{aligned} x_{n+1} &= \mathrm{prox}_{\alpha g}\left(x_n - \alpha \nabla h(x_n)\right) \\ &= \mathrm{Argmin}\left\{ g(x) + \frac{1}{2\alpha}\left\| x - \left(x_n - \alpha \nabla h(x_n)\right)\right\|^2 \right\}. \end{aligned}$$

Proximal-gradient algorithm

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

$$\begin{aligned} x_{n+1} &= \operatorname{prox}_{\alpha g}\big(x_n - \alpha \nabla h(x_n)\big) \\ &= \operatorname{Argmin}\left\{ g(x) + \frac{1}{2\alpha}\left\| x - \big(x_n - \alpha \nabla h(x_n)\big) \right\|^2 \right\}. \end{aligned}$$

This subproblem has a unique solution characterized by

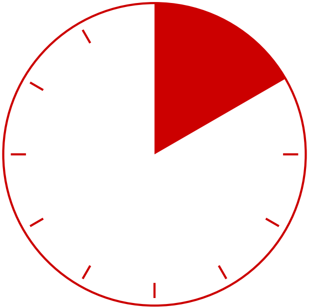
$$0 \in \partial g(x_{n+1}) + \nabla h(x_n) + \frac{1}{\alpha}(x_{n+1} - x_n).$$

Example revisited

$\ell^1 + \ell^2$ minimization

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

Break



Convergence of proximal-gradient sequences

Theorem

Let $f = g + h$, *where* $g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ *is closed and convex, and* $h : \mathbb{R}^N \to \mathbb{R}$ *is L-smooth and convex. Take* $\alpha \in (0, 1/L]$ *and define* (x_n) *by*

$$x_{n+1} = \operatorname{prox}_{\alpha g}\big(x_n - \alpha \nabla h(x_n)\big), \qquad n \geq 0.$$

If $S \neq \emptyset$, x_n *converges to an* $\hat{x} \in S$, *and*

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

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

Proof

$f(x_n) + g(x_n)$ is nonincreasing

$$\begin{cases} f(x_{n+1}) & \leq & f(x_n) + \nabla f(x_n) \cdot (x_{n+1} - x_n) + \frac{L}{2} \|x_{n+1} - x_n\|^2 \\ g(x_{n+1}) & \leq & g(x_n) + \left(\frac{x_n - x_{n+1}}{\alpha} - \nabla f(x_n) \right) \cdot (x_{n+1} - x_n). \end{cases}$$

Proof

$f(x_n) + g(x_n)$ is nonincreasing

$$\begin{cases} f(x_{n+1}) & \leq & f(x_n) + \nabla f(x_n) \cdot (x_{n+1} - x_n) + \frac{L}{2}\|x_{n+1} - x_n\|^2 \\ g(x_{n+1}) & \leq & g(x_n) + \left(\frac{x_n - x_{n+1}}{\alpha} - \nabla f(x_n)\right) \cdot (x_{n+1} - x_n). \end{cases}$$

Convergence rate

$$\begin{cases} f(x_n) & \leq & f(p) + \nabla f(x_n) \cdot (x_n - p) \\ g(x_{n+1}) & \leq & g(p) + \left(\frac{x_n - x_{n+1}}{\alpha} - \nabla f(x_n)\right) \cdot (x_{n+1} - p). \end{cases}$$

Proof

$f(x_n) + g(x_n)$ is nonincreasing

$$\left\{\begin{array}{rcl} f(x_{n+1}) & \leq & f(x_n) + \nabla f(x_n)\cdot(x_{n+1}-x_n) + \frac{L}{2}\|x_{n+1}-x_n\|^2 \\ g(x_{n+1}) & \leq & g(x_n) + \left(\frac{x_n - x_{n+1}}{\alpha} - \nabla f(x_n)\right)\cdot(x_{n+1}-x_n). \end{array}\right.$$

Convergence rate

$$\left\{\begin{array}{rcl} f(x_n) & \leq & f(p) + \nabla f(x_n)\cdot(x_n - p) \\ g(x_{n+1}) & \leq & g(p) + \left(\frac{x_n - x_{n+1}}{\alpha} - \nabla f(x_n)\right)\cdot(x_{n+1}-p). \end{array}\right.$$

Compatibility

$$\left\|x_{n+1} - x_n - \alpha\big(\nabla f(x_{n+1}) - \nabla f(x_n)\big)\right\|^2 \leq \|x_{n+1}-x_n\|^2.$$

Polyak-Łojasiewicz inequality and gradient descent

A function $\phi : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ satisfies the Polyak-Łojasiewicz inequality with constant μ if

$$2\mu\big(\phi(x) - \min(\phi)\big) \leq \|v\|^2 \quad \text{for every } x \in \mathbb{R}^N \text{ and } v \in \partial\phi(x).$$

Proposition

Let $f : \mathbb{R}^N \to \mathbb{R}$ be $L-$smooth and convex, and let $g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be closed and convex. Let $\phi := f + g$ satisfy the Polyak-Łojasiewicz inequality with constant $\mu > 0$, and iterate $x_{k+1} = \operatorname{prox}_{\gamma g}\left(x_k - \gamma \nabla f(x_k)\right)$, with $\gamma < 2/L$. Then, for every $k \geq 0$, we have

$$f(x_k) - \min(f) \leq \left(\frac{1}{1 + \mu\alpha(2 - \alpha L)}\right)^k \big(f(x_0) - \min(f)\big).$$

Nesterov's acceleration

The Fast Iterative Shrinkage-Thresholding Algorithm results from combining Nesterov's acceleration to the proximal-gradient method:

FISTA (Beck-Teboulle, 2009)

$$\begin{cases} y_k &= x_k + \theta_k \left(x_k - x_{k-1}\right) \\ x_{k+1} &= \mathrm{prox}_{\alpha g}\left(y_k - \alpha \nabla f(y_k)\right). \end{cases}$$