First Order Descent Methods in Nonsmooth Convex Optimization

Juan PEYPOUQUET
Universidad Técnica Federico Santa María
Universidad de Chile

CIMPA School on Algorithmic Nonsmooth Optimization

Ibarra, September 19-21, 2017

INTRODUCTION TO OPTIMIZATION

Outline

Classes 1 & 2: Introduction. Smooth convex functions.

Classes 3 & 4: Nonsmooth convex functions, and problems with a mixed structure.

Class 5: Group work: Theoretical exercise.

Class 6: Fast convergence: acceleration and geometry.

Classes 7 & 8: Group work: Numerical exercise.

Outline

Classes 1 & 2: Introduction. Smooth convex functions.

Classes 3 & 4: Nonsmooth convex functions, and problems with a mixed structure.

Class 5: Group work: Theoretical exercise.

Class 6: Fast convergence: acceleration and geometry.

Classes 7 & 8: Group work: Numerical exercise.

Outline

Classes 1 & 2: Introduction. Smooth convex functions.

Classes 3 & 4: Nonsmooth convex functions, and problems with a mixed structure.

Class 5: Group work: Theoretical exercise.

Class 6: Fast convergence: acceleration and geometry.

Classes 7 & 8: Group work: Numerical exercise.

Optimization problems

Let H be a real Hilbert space, and let

$$f : \mathrm{dom}(f) \subset H \to \mathbb{R}.$$

We are concerned with the numerical approximation of

- the infimal value that f approaches on some feasible set

$$C \subset \mathrm{dom}(f),$$

- and the points where this value is attained, if there are any.

Optimization problems

Let H be a real Hilbert space, and let

$$f : \mathrm{dom}(f) \subset H \to \mathbb{R}.$$

We are concerned with the numerical approximation of

- the infimal value that f approaches on some feasible set

$$C \subset \mathrm{dom}(f),$$

- and the points where this value is attained, if there are any.

Optimization problems

The value of the problem is

$$\inf(f) = \inf\{f(x) : x \in C\},$$

even if it is not attained, and the solution set is

$$S = \operatorname{argmin}(f) = \{x \in C : f(x) = \min(f)\}.$$

Theorem (Weierstrass)

If $f : C \subset H \to \mathbb{R}$ is continuous and C is compact, then $S \neq \emptyset$.

Optimization problems

The value of the problem is

$$\inf(f) = \inf\{f(x) : x \in C\},$$

even if it is not attained, and the solution set is

$$S = \operatorname{argmin}(f) = \{x \in C : f(x) = \min(f)\}.$$

Theorem (Weierstrass)

If $f : C \subset H \to \mathbb{R}$ is continuous and C is compact, then $S \neq \emptyset$.

Optimization problems

The value of the problem is

$$\inf(f) = \inf\{f(x) : x \in C\},$$

even if it is not attained, and the solution set is

$$S = \operatorname{argmin}(f) = \{x \in C : f(x) = \min(f)\}.$$

Theorem (Weierstrass)

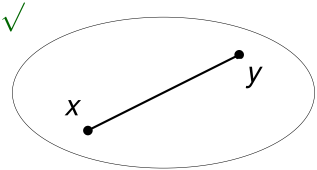
If $f : C \subset H \to \mathbb{R}$ is continuous and C is compact, then $S \neq \emptyset$.

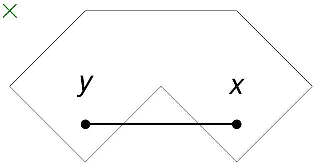
Convex sets

A set $C \subset H$ is convex if, given any two points $x, y \in C$, the segment joining x and y:

$$\text{seg}[x, y] = \{\lambda x + (1-\lambda)y : \lambda \in [0, 1]\}$$

is contained in C.





Differential calculus

The directional derivative of f at x in the direction d is

$$f'(x;d) = \lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t}.$$

We say f is Gâteaux-differentiable at x if the function

$$d \mapsto f'(x;d)$$

is linear and bounded. The gradient of f at x is the vector $\nabla f(x)$ satisfying

$$\langle \nabla f(x), d \rangle = f'(x;d) \quad \text{for all } d.$$

Differential calculus

The directional derivative of f at x in the direction d is

$$f'(x;d) = \lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t}.$$

We say f is Gâteaux-differentiable at x if the function

$$d \mapsto f'(x;d)$$

is linear and bounded. The gradient of f at x is the vector $\nabla f(x)$ satisfying

$$\langle \nabla f(x), d \rangle = f'(x;d) \quad \text{for all } d.$$

Differential calculus

The directional derivative of f at x in the direction d is

$$f'(x;d) = \lim_{t \to 0^+} \frac{f(x+td) - f(x)}{t}.$$

We say f is Gâteaux-differentiable at x if the function

$$d \mapsto f'(x;d)$$

is linear and bounded. The gradient of f at x is the vector $\nabla f(x)$ satisfying

$$\langle \nabla f(x), d \rangle = f'(x;d) \quad \text{for all } d.$$

A necessary condition for optimality

Theorem (Fermat's Rule)

Let $f : H \to \mathbb{R}$ and let $C \subset H$ be convex. If $f(x^*) \leq f(c)$ for every $c \in C$ and f is (Gâteaux) differentiable at x^*, then

$$\langle \nabla f(x^*), c - x^* \rangle \geq 0$$

for all $c \in C$. In particular, if $x^* \in int(C)$, then $\nabla f(x^*) = 0$.

Dynamic Minimization of Smooth Convex Functions

Convex functions

A function $f : H \to \mathbb{R}$ is convex if

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

for every $x, y \in H$ and every $\lambda \in (0,1)$.

This is equivalent to the convexity of the epigraph of f:

$$\text{epi}(f) = \{(x,z) \in H \times \mathbb{R} : f(x) \leq z\},$$

and implies the convexity of every sublevel set

$$[f \leq \gamma] = \{x \in H : f(x) \leq \gamma\}, \qquad \gamma \in \mathbb{R}.$$

Convex functions

A function $f : H \to \mathbb{R}$ is convex if

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

for every $x, y \in H$ and every $\lambda \in (0,1)$.

This is equivalent to the convexity of the epigraph of f:

$$\text{epi}(f) = \{(x,z) \in H \times \mathbb{R} : f(x) \leq z\},$$

and implies the convexity of every sublevel set

$$[f \leq \gamma] = \{x \in H : f(x) \leq \gamma\}, \qquad \gamma \in \mathbb{R}.$$

Convex functions

A function $f : H \to \mathbb{R}$ is convex if

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda) f(y)$$

for every $x, y \in H$ and every $\lambda \in (0,1)$.

This is equivalent to the convexity of the epigraph of f:

$$\operatorname{epi}(f) = \{(x,z) \in H \times \mathbb{R} : f(x) \le z\},$$

and implies the convexity of every sublevel set

$$[f \le \gamma] = \{x \in H : f(x) \le \gamma\}, \qquad \gamma \in \mathbb{R}.$$

The gradient of a convex function

Let $f : H \to \mathbb{R}$ be differentiable. The following are equivalent:

1. f is convex.
2. The gradient inequality holds:

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

3. The gradient is monotone:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

The gradient of a convex function

Let $f : H \to \mathbb{R}$ be differentiable. The following are equivalent:

1. f is convex.
2. The gradient inequality holds:

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

3. The gradient is monotone:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

The gradient of a convex function

Let $f : H \to \mathbb{R}$ be differentiable. The following are equivalent:

1. f is convex.
2. The gradient inequality holds:

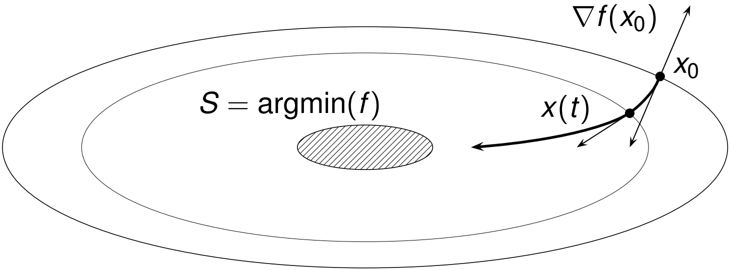
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

3. The gradient is monotone:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

Steepest descent dynamics

The differential equation $\dot{x}(t) = -\nabla f(x(t))$, $x(0) = x_0$, has a unique solution if ∇f is locally Lipschitz-continuous[1].



[1] In this case, f is continuous.

Basic properties

Hypothesis

f is convex, and ∇f is locally Lipschitz-continuous.

- $t \mapsto f(x(t))$ is nonincreasing.
- If f is bounded from below, then $\int_0^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty$.
- $\lim_{t \to +\infty} f(x(t)) = \inf(f)$.

Basic properties

Hypothesis

f is convex, and ∇f is locally Lipschitz-continuous.

- $t \mapsto f(x(t))$ is nonincreasing.
- If f is bounded from below, then $\int_0^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty$.
- $\lim_{t \to +\infty} f(x(t)) = \inf(f)$.

Basic properties

Hypothesis

f is convex, and ∇f is locally Lipschitz-continuous.

- $t \mapsto f(x(t))$ is nonincreasing.
- If f is bounded from below, then $\displaystyle\int_0^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty$.
- $\displaystyle\lim_{t\to+\infty} f(x(t)) = \inf(f)$.

Basic properties

Hypothesis

f is convex, and ∇f is locally Lipschitz-continuous.

- $t \mapsto f(x(t))$ is nonincreasing.
- If f is bounded from below, then $\int_0^{+\infty} \|\dot{x}(t)\|^2 \, dt < +\infty$.
- $\lim_{t \to +\infty} f(x(t)) = \inf(f)$.

Convergence

- If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.
- If $S \neq \emptyset$, then $f(x(t)) - \min(f) \leq \dfrac{\operatorname{dist}(x_0, S)^2}{2t}$.
- If $x^* \in S$, then $t \mapsto \|x(t) - x^*\|$ is nonincreasing and $\lim_{t \to +\infty} \|x(t) - x^*\|$ exists.
- $S \neq \emptyset$ if, and only if, $x(\cdot)$ is bounded.
- If $S \neq \emptyset$, $x(t)$ converges weakly, as $t \to +\infty$, to a point in S.

Convergence

- If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.
- If $S \neq \emptyset$, then $f(x(t)) - \min(f) \leq \dfrac{\mathrm{dist}(x_0, S)^2}{2t}$.
- If $x^* \in S$, then $t \mapsto \|x(t) - x^*\|$ is nonincreasing and $\lim_{t \to +\infty} \|x(t) - x^*\|$ exists.
- $S \neq \emptyset$ if, and only if, $x(\cdot)$ is bounded.
- If $S \neq \emptyset$, $x(t)$ converges weakly, as $t \to +\infty$, to a point in S.

Convergence

- If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.
- If $S \neq \emptyset$, then $f(x(t)) - \min(f) \leq \dfrac{\text{dist}(x_0, S)^2}{2t}$.
- If $x^* \in S$, then $t \mapsto \|x(t) - x^*\|$ is nonincreasing and $\lim_{t \to +\infty} \|x(t) - x^*\|$ exists.
- $S \neq \emptyset$ if, and only if, $x(\cdot)$ is bounded.
- If $S \neq \emptyset$, $x(t)$ converges weakly, as $t \to +\infty$, to a point in S.

Convergence

- If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.
- If $S \neq \emptyset$, then $f(x(t)) - \min(f) \leq \dfrac{\operatorname{dist}(x_0, S)^2}{2t}$.
- If $x^* \in S$, then $t \mapsto \|x(t) - x^*\|$ is nonincreasing and $\lim_{t\to+\infty} \|x(t) - x^*\|$ exists.
- $S \neq \emptyset$ if, and only if, $x(\cdot)$ is bounded.
- If $S \neq \emptyset$, $x(t)$ converges weakly, as $t \to +\infty$, to a point in S.

Convergence

- If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.
- If $S \neq \emptyset$, then $f(x(t)) - \min(f) \leq \dfrac{\text{dist}(x_0, S)^2}{2t}$.
- If $x^* \in S$, then $t \mapsto \|x(t) - x^*\|$ is nonincreasing and $\lim_{t \to +\infty} \|x(t) - x^*\|$ exists.
- $S \neq \emptyset$ if, and only if, $x(\cdot)$ is bounded.
- If $S \neq \emptyset$, $x(t)$ converges weakly, as $t \to +\infty$, to a point in S.

Ingredients for the proof

Gradient inequality:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \qquad \forall x, y \in H.$$

Lemma (Opial)

Let $\emptyset \neq S \subset H$ and $x : [0, +\infty) \to H$. Suppose:

- *If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.*
- *For each $u \in S$, $\lim_{t \to +\infty} \|x(t) - u\|$ exists.*

Then $x(t)$ converges, as $t \to +\infty$, to a point in S.

Ingredients for the proof

Gradient inequality:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \qquad \forall x, y \in H.$$

Lemma (Opial)

Let $\emptyset \neq S \subset H$ and $x : [0, +\infty) \to H$. Suppose:

- *If $t_k \to +\infty$ and $x(t_k) \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.*
- *For each $u \in S$, $\lim_{t \to +\infty} \|x(t) - u\|$ exists.*

Then $x(t)$ converges, as $t \to +\infty$, to a point in S.

Algorithmic Version: The Gradient Method

The gradient method

Begin with $x_0 \in H$ and pick a step size $\lambda > 0$. Given $x_k \in H$, compute x_{k+1} by

$$x_{k+1} = x_k - \lambda \nabla f(x_k).$$

In other words,

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

which is a (forward) finite-difference discretization of

$$\dot{x}(t) = -\nabla f(x(t)).$$

The gradient method

Begin with $x_0 \in H$ and pick a step size $\lambda > 0$. Given $x_k \in H$, compute x_{k+1} by

$$x_{k+1} = x_k - \lambda \nabla f(x_k).$$

In other words,

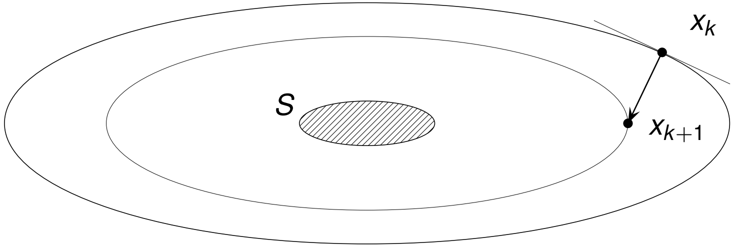
$$\frac{x_{k+1} - x_k}{\lambda} = -\nabla f(x_k),$$

which is a (forward) finite-difference discretization of

$$\dot{x}(t) = -\nabla f(x(t)).$$

The gradient method

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



This may work, or not

Example

Let $f(x) = x^2$ *and* $x_0 = 1$. *Then*

$$x_k = (1 - 2\lambda)^k.$$

In particular,

1. *If* $\lambda > 1$, $\lim_{k \to +\infty} |x_k| = +\infty$;
2. *If* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *but* $\nexists \lim_{k \to +\infty} x_k$; *and*
3. *If* $\lambda < 1$, $\lim_{k \to +\infty} x_k = 0$.

This may work, or not

Example

Let $f(x) = x^2$ *and* $x_0 = 1$. *Then*

$$x_k = (1 - 2\lambda)^k.$$

In particular,

1. *If* $\lambda > 1$, $\lim_{k \to +\infty} |x_k| = +\infty$;
2. *If* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *but* $\nexists \lim_{k \to +\infty} x_k$; *and*
3. *If* $\lambda < 1$, $\lim_{k \to +\infty} x_k = 0$.

This may work, or not

Example

Let $f(x) = x^2$ *and* $x_0 = 1$. *Then*

$$x_k = (1 - 2\lambda)^k.$$

In particular,

1. *If* $\lambda > 1$, $\lim_{k \to +\infty} |x_k| = +\infty$;
2. *If* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *but* $\nexists \lim_{k \to +\infty} x_k$; *and*
3. *If* $\lambda < 1$, $\lim_{k \to +\infty} x_k = 0$.

This may work, or not

Example

Let $f(x) = x^2$ *and* $x_0 = 1$. *Then*

$$x_k = (1 - 2\lambda)^k.$$

In particular,

1. *If* $\lambda > 1$, $\lim_{k\to+\infty} |x_k| = +\infty$;
2. *If* $\lambda = 1$, $x_k = (-1)^k$, $|x_k| \equiv 1$ *but* $\nexists \lim_{k\to+\infty} x_k$; *and*
3. *If* $\lambda < 1$, $\lim_{k\to+\infty} x_k = 0$.

Further properties of the gradient

Let $f : H \to \mathbb{R}$ be convex and differentiable. The following are equivalent:

1. ∇f is Lipschitz-continuous with constant L:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \qquad \forall x, y.$$

2. The descent inequality holds:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|x - y\|^2 \qquad \forall x, y.$$

3. ∇f is cocoercive with constant $1/L$:

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \qquad \forall x, y.$$

Properties

Hypothesis

∇f is Lipschitz-continuous with constant L and $\lambda L \leq 1$.

- The sequence $\big(f(x_k)\big)$ is nonincreasing.
- If f is bounded from below, then $\sum_{k \geq 0} \|x_{k+1} - x_k\|^2 < +\infty$.
- $\lim_{k \to +\infty} f(x_k) = \inf(f)$.
- If $x_{m_k} \rightharpoonup \bar{x}$ as $k \to +\infty$, then $\bar{x} \in S$.

Convergence

- If $S \neq \emptyset$, then $f(x_k) - \min(f) \leq \dfrac{\text{dist}(x_0, S)^2}{2\lambda k}$.
- For each $x^* \in S$, the sequence $(\|x_k - x^*\|)$ is nonincreasing.
- $S \neq \emptyset$ if, and only if, (x_k) is bounded.
- If $S \neq \emptyset$, x_k converges weakly, as $k \to +\infty$, to a point in S.

Exercise

Most properties hold with $\lambda L < 2$, with minor changes.

Convergence

- If $S \neq \emptyset$, then $f(x_k) - \min(f) \leq \dfrac{\text{dist}(x_0, S)^2}{2\lambda k}$.
- For each $x^* \in S$, the sequence $\left(\|x_k - x^*\|\right)$ is nonincreasing.
- $S \neq \emptyset$ if, and only if, (x_k) is bounded.
- If $S \neq \emptyset$, x_k converges weakly, as $k \to +\infty$, to a point in S.

Exercise

Most properties hold with $\lambda L < 2$, with minor changes.

Exercise

Let ∇f be Lipschitz-continuous with constant L. Consider iterations of the form

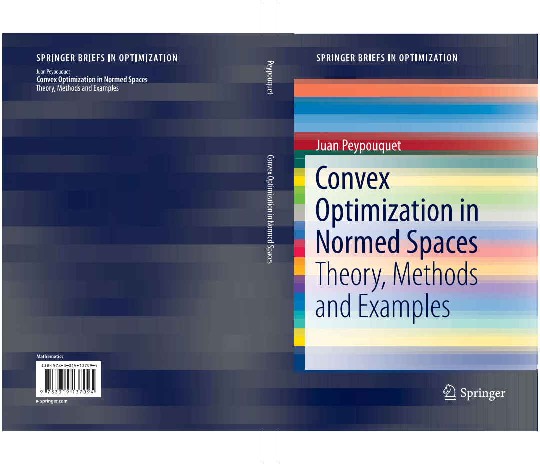
$$x_{k+1} = x_k + \lambda d_k. \tag{1}$$

Suppose there exist $\alpha, \beta > 0$ such that $\beta\lambda L < 2\alpha$ and

$$\alpha\|d_k\|^2 \leq \|\nabla f(x_k)\|^2 \leq -\beta\langle d_k, f(x_k)\rangle.$$

- Analyze the behavior of the sequence (x_k).
- What can you say if we replace λ by λ_k in (1)?

Complementary reading



Preprint available at http://jpeypou.mat.utfsm.cl/