Lecture 2: Basics on convex smooth/non-smooth optimisation

Luca Calatroni
CR CNRS, Laboratoire I3S
CNRS, UCA, Inria SAM, France

MSc DSAI - UCA

Inverse problems in image processing

January 13 2023

Table of contents

Motivation

Goal: providing theoretical/practical tools (i.e. algorithms) for solving

$$\boxed{\min_{x \in \mathbb{R}^n} \; F(x)}$$

for a function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with suitable properties.

Motivation

Goal: providing theoretical/practical tools (i.e. algorithms) for solving

$$\boxed{\min_{x \in \mathbb{R}^n} \; F(x)}$$

for a function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with suitable properties.

Such minimisation problem often appears in many contexts:

- **Inverse problems**: imaging, variable selection, compressed sensing. . . .
  Example: $F(x) = \frac{1}{2}\|Ax - y\|_2^2$ (least-square problem),. . .
- **Statistical/machine learning**: empirical risk minimisation, regression. . .
- **Numerical analysis/optimisation**: analysis/implementation of fast algorithms for solving large-scale problems. . .
- . . . many more!

Recalling last lecture

Convolution problem: $Y = g * X + N$ with Y, X, N, $\mathbb{R}^{N\times N}$ (2D objects)

- Represent this using a Fourier (circulant) matrix $A \in \mathbb{R}^{n\times n}$ with $n = N^2$

$$\boxed{y = Ax + n}$$

  where $y = \text{vec}(Y)$, $x = \text{vec}(X)$ and $n = \text{vec}(N)$.
- You have looked at the condition number of A and commented on the small size of the singular values of A, which creates the problem
- **Not possible** to perform the "naive" inversion:

$$x = A^{-1}(y - n) \qquad \textbf{THIS IS WRONG!}$$

- **Minimisation approach**: look instead at the minimisation problem

$$\boxed{\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - y\|^2}$$

  My lectures: how to solve it numerically.

References

Some standard reference books/surveys:

- R. Tyller Rockafeller, *Convex Analysis*, Princeton University Press, 1970.
- S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- A. Beck, *First-order methods in optimization*, Volume 25, MOS-SIAM series on Optimization, 2017.
- A. Chambolle, T. Pock, *An introduction to continuous optimization for imaging*, Acta Numerica, 2016

. . . will try mostly to follow these books.

My contacts

- **Who**: Luca Calatroni (my website)
- **What**: Junior research scientist (CR) CNRS.
- **Where**: I3S laboratory, 2000 Route des lucioles, Sophia-Antipolis, France.
- **When**: by videoconference/appointement.
- **How**: e-mail calatroni@i3s.unice.fr

Research interests

Inverse imaging problems, smooth/non-smooth & convex/non-convex optimisation, biomedical imaging applications, digital restoration of artworks, computational neurosciences. . .

Preliminaries & basic notions

- **Linear algebra**: vector spaces, norms, scalar product
- **Basic calculus**: the space $\mathbb{R}^n$: dot product $\langle v, w \rangle = v^T w$, ℓ_p norms $(p \geq 1)$ $\|x\|_p$.
  Will use $\|x\| = \|x\|_2 = \left(\sum x_i^2\right)^{1/2}$ to denote the standard Euclidean norm.
- **Analysis**: sequences, convergence, limits, continuity, differentiability...
- Convex subsets $S \subset \mathbb{R}^n$:

$$(\forall x, y \in S) \quad (\forall \alpha \in [0,1]) \quad z := \alpha x + (1-\alpha) y \in S$$

- Linear operators $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$; their matrix representation $A \in \mathbb{R}^{m \times n}$, adjoint operators A^T. Example: convolution with kernel g: $g * X \leftrightarrow Ax$, with $x = \mathrm{vec}(X)$.
- Operator norm $\|A\| = \max\{\|Ax\| : \|x\| \leq 1\}$

Proper functions

Minimal property to have well-defined minimisation problems.

Proper function

A function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said *proper* iff

$$\exists x \in \mathbb{R}^n \quad \text{such that} \quad F(x) \neq +\infty.$$

We define $\mathcal{P} := \{F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \text{ s.t. } F \text{ is proper}\}$.

Proper functions

Minimal property to have well-defined minimisation problems.

Proper function

A function $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said *proper* iff

$$\exists x \in \mathbb{R}^n \quad \text{such that} \quad F(x) \neq +\infty.$$

We define $\mathcal{P} := \{F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \text{ s.t. } F \text{ is proper}\}$.

Why this? We are interested in solving

$$\boxed{\min_{x \in \mathbb{R}^n} F(x)}$$

so we need to exclude atypical functions identically equal to $+\infty$. The set:

$$\operatorname{dom}(F) := \{x \in \mathbb{R}^n : F(x) < +\infty\}$$

is called the *effective domain* of F.

Some examples

- For $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$, consider $F(x) = \frac{1}{2}\|Ax - y\|_2^2 \to$ least-square minimisation problem.
- In the case of Poisson noise statistics $F(x) = \sum_{i=1}^m \left((Ax)_i - y_i \log{(Ax)_i}\right)$
- *Regularised problems*: for $\lambda > 0$, $F(x) = \frac{1}{2}\|Ax - y\|_2^2 + \lambda g(x)$ with, for instance:
  - $g(x) = \|x\|_2^2$ or $g(x) = \|\nabla x\|_2^2$;
  - $g(x) = \|x\|_1$ or $g(x) = \|\nabla x\|_1$;
  - $g(x) = \|Wx\|_1$ where W is a wavelet operator (see next lectures. . .)

Some examples

- For $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$, consider $F(x) = \frac{1}{2}\|Ax - y\|_2^2 \to$ least-square minimisation problem.
- In the case of Poisson noise statistics $F(x) = \sum_{i=1}^m \left((Ax)_i - y_i \log{(Ax)_i}\right)$
- *Regularised problems*: for $\lambda > 0$, $F(x) = \frac{1}{2}\|Ax - y\|_2^2 + \lambda g(x)$ with, for instance:
  - $g(x) = \|x\|_2^2$ or $g(x) = \|\nabla x\|_2^2$;
  - $g(x) = \|x\|_1$ or $g(x) = \|\nabla x\|_1$;
  - $g(x) = \|Wx\|_1$ where W is a wavelet operator (see next lectures. . .)

The type of problems considered will be often in the form:

$$\boxed{\min_{x\in\mathbb{R}^n} \ \{F(x) := f(x) + g(x)\}\,,}$$

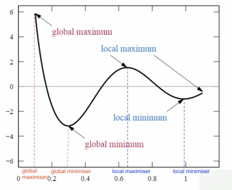
where f, g are proper functions and where f is *nicer* (smoother) than g. Namely:

- $g \equiv 0$: **smooth** optimisation ($\to F(x) = f(x)$, so F is smooth);
- $g \not\equiv 0$: **non-smooth** optimisation.

Global/local minimisers

For $F \in \mathcal{P}$, we denote:

- **global minimiser**: $x^* \in \mathbb{R}^n$: $F(x^*) \leq F(x)$ for $\underline{\text{every } x \in \mathbb{R}^n}$.
- **local minimiser**: $x^* \in \mathbb{R}^n$: there exists $\delta > 0$ and a neighbourhood $B_\delta(x^*)$ such that $F(x^*) \leq F(x)$ for $\underline{\text{every } x \in B_\delta(x^*)}$.

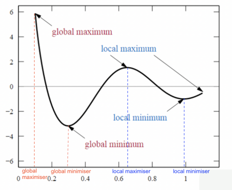


Note: minima of a functional $\neq$ **minimisers** (points in the domain where minima are attained)!

Global/local minimisers

For $F \in \mathcal{P}$, we denote:

- **global minimiser**: $x^* \in \mathbb{R}^n$: $F(x^*) \leq F(x)$ for $\underline{\text{every } x \in \mathbb{R}^n}$.
- **local minimiser**: $x^* \in \mathbb{R}^n$: there exists $\delta > 0$ and a neighbourhood $B_\delta(x^*)$ such that $F(x^*) \leq F(x)$ for $\underline{\text{every } x \in B_\delta(x^*)}$.



Note: minima of a functional $\neq$ **minimisers** (points in the domain where minima are attained)!

Set of minimisers

The **set** of (local, global) minimisers of F is denoted by:

$$\arg\min F = \{x^* \in \mathbb{R}^n : x^* \text{ is a minimiser of } F\}\,.$$

Not necessarily a singleton! (there can be more than one minimiser, it depends on F)

A stupid, but important remark

Optimisation problems are often formulated in two ways:

- min form: looking for minimal values of the functional

$$\min_{x \in \mathbb{R}^n} F(x)$$

- (more instructive) argmin form: looking for minisers of the functional

$$\arg\min_{x \in \mathbb{R}^n} F(x)$$

A stupid, but important remark

Optimisation problems are often formulated in two ways:

- min form: looking for minimal values of the functional

$$\min_{x \in \mathbb{R}^n} F(x)$$

- (more instructive) argmin form: looking for minisers of the functional

$$\arg\min_{x \in \mathbb{R}^n} F(x)$$

In practice, we look most of the times to the second formulation as we are also interested in the element x^* minimising F.

Convexity

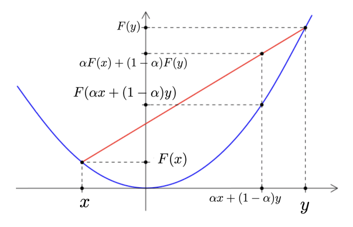
Let $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function, i.e. $F \in \mathcal{P}$.

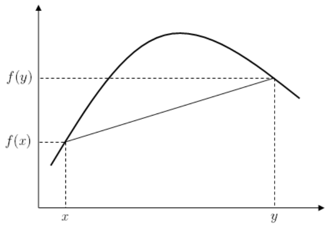
Convex function

$F \in \mathcal{P}$ is said to be *convex* if:

$$(\forall x, y \in \mathbb{R}^n) \quad (\forall \alpha \in [0,1]) \quad F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y).$$

Moreover, F is *strictly convex* if the inequality holds when $x, y \in \mathrm{dom}(F)$, $x \neq y$ and $\alpha \in (0,1)$. We say that $G : \mathbb{R}^n \to [-\infty, +\infty)$ is *concave* is $F = -G$ is convex. If a function is not convex nor concave we say that is *non-convex*.





Convex/concave function

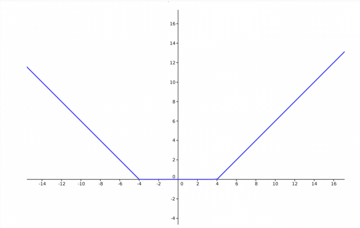
Let $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function, i.e. $F \in \mathcal{P}$.

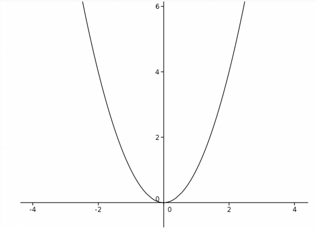
Convex function

$F \in \mathcal{P}$ is said to be *convex* if:

$$(\forall x, y \in \mathbb{R}^n) \quad (\forall \alpha \in [0,1]) \quad F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y).$$

Moreover, F is *strictly convex* if the inequality holds when $x, y \in \mathrm{dom}(F)$, $x \neq y$ and $\alpha \in (0,1)$. We say that $G : \mathbb{R}^n \to [-\infty, +\infty)$ is *concave* is $F = -G$ is convex. If a function is not convex nor concave we say that is *non-convex*.





Convex VS. strictly convex functions

Convex functions

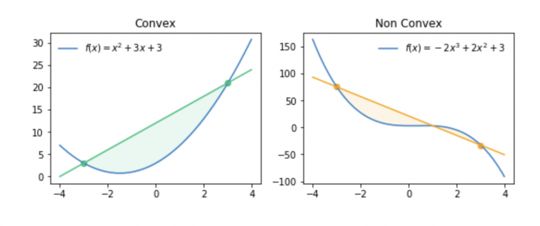
Let $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function, i.e. $F \in \mathcal{P}$.

Convex function

$F \in \mathcal{P}$ is said to be *convex* if:

$$(\forall x, y \in \mathbb{R}^n) \quad (\forall \alpha \in [0,1]) \quad F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y).$$

Moreover, F is *strictly convex* if the inequality holds when $x, y \in \mathrm{dom}(F)$, $x \neq y$ and $\alpha \in (0,1)$. We say that $G : \mathbb{R}^n \to [-\infty, +\infty)$ is *concave* is $F = -G$ is convex. If a function is not convex nor concave we say that is *non-convex*.



Convex VS. non-convex

Convex functions

Let $F : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper function, i.e. $F \in \mathcal{P}$.

Convex function

$F \in \mathcal{P}$ is said to be *convex* if:

$$(\forall x, y \in \mathbb{R}^n) \quad (\forall \alpha \in [0,1]) \quad F(\alpha x + (1-\alpha)y) \leq \alpha F(x) + (1-\alpha)F(y).$$

Moreover, F is *strictly convex* if the inequality holds when $x, y \in \operatorname{dom}(F)$, $x \neq y$ and $\alpha \in (0,1)$. We say that $G : \mathbb{R}^n \to [-\infty, +\infty)$ is *concave* is $F = -G$ is convex. If a function is not convex nor concave we say that is *non-convex*.

Examples:

- $F(x) = \|x\|$ is convex

  $$\|\alpha x + (1-\alpha)y\| \leq \|\alpha x\| + \|(1-\alpha)y\| = \alpha\|x\| + (1-\alpha)\|y\| \qquad \forall x, y \in \mathbb{R}^n$$

- $F(x) = \|x\|^2$ is strictly convex
- $F(x) = \|x\|_p$, $p \in [1, +\infty)$ are convex

Useful properties

Proposition (operations with convex functions)

Let F and G two convex functions and let $\beta > 0$. Then, the sum $F + G$ is a convex function and the function βF is a convex function.

Useful properties

Proposition (operations with convex functions)

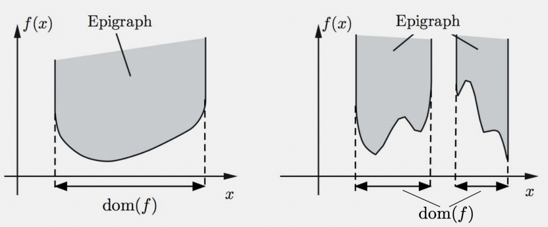
Let F and G two convex functions and let $\beta > 0$. Then, the sum $F + G$ is a convex function and the function βF is a convex function.

Proposition (epigraph of convex functions is convex set)

Let $F \in \mathcal{P}$. Then F is a convex function if and only if the set

$$\mathrm{epi}(F) := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : F(x) \leq t\}$$

is convex.



Continuity and differentiability

Lower semi-continuity

You all know what a continuous function is:

$$(\forall x \in \mathrm{dom}(F)) \quad \lim_{y \to x} F(y) = F(x)$$

However, a weaker notion can be used to guarantee the well-posedness of optimisation problems classically encountered in this context.

Lower semi-continuity

You all know what a continuous function is:

$$(\forall x \in \operatorname{dom}(F)) \quad \lim_{y \to x} F(y) = F(x)$$

However, a weaker notion can be used to guarantee the well-posedness of optimisation problems classically encountered in this context.

Lower semi-continuity

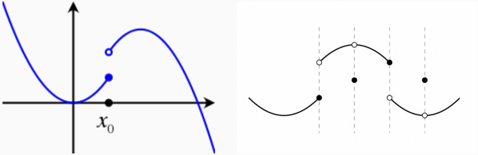
Let $F \in \mathcal{P}$. We say that F is *lower semi-continuous (l.s.c.) at the point* $x \in \mathbb{R}^n$ iff

$$F(x) \leq \liminf_{y \to x} F(y).$$

Using sequences, this means that for every sequence $(x_k)_{k \in \mathbb{N}}$ such that $x_k \to x$:

$$F(x) \leq \liminf_{k \to +\infty} F(x_k) := \lim_{k \to +\infty} \inf \left\{ F(x_j) : j \geq k \right\}.$$

If F is l.s.c. at every $x \in \mathbb{R}^n$, we say that the function is l.s.c.



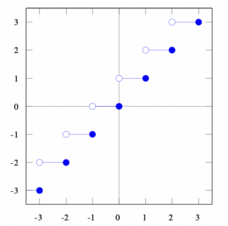
Left: lower l.s.c. **Right**: where the function is lower l.s.c.?

Examples of l.s.c. functions

- The functions

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}, \qquad F(x) = \lceil x \rceil = \min \{k \in \mathbb{Z} : x \leq k\}$$

are l.s.c. (but not continuous).



$F(x) = \lceil x \rceil$

- All continuous functions (l.s.c + u.s.c.).

Coercivity

We need to ensure that the minimum is not attained at the “extreme points” of the domain. . .

We need to ensure that the minimum is not attained at the "extreme points" of the domain...

Coercivity

Let $F \in \mathcal{P}$. We say that F is *coercive* iff

$$\lim_{\|x\| \to +\infty} F(x) = +\infty.$$

This means that the function has a large growth for values of x whose norm is large.

Examples:

- $F : \mathbb{R} \to \mathbb{R}_+$, $F(x) = e^x$ is **not** coercive, but $F : \mathbb{R} \to \mathbb{R}_+$, $F(x) = e^{|x|}$ is.
- $F : \mathbb{R}^2 \to \mathbb{R}_+$, $F(x_1, x_2) = x_1^2 + x_2^2$ is coercive.
- $F : \mathbb{R}^2 \to \mathbb{R}_+$, $F(x_1, x_2) = x_1^2 - 2x_1x_2 + x_2^2 = (x_1 - x_2)^2$ is **not** coercive. ($F(x_1, x_2) = 0$ on the line $x_1 = x_2$, so there $\|x\| = \sqrt{x_1^2 + x_2^2} \to +\infty$, but $F(x) = 0$)

Existence of minimisers

We have everything we need to provide conditions for showing the **extistence** of minimisers for the problem considered.

Theorem (existence of minimisers)

Let $F \in \mathcal{P}$. If F is l.s.c. and coercive, then F admits a minimiser.

Note I: Equivalently said, the problem:

$$\arg\min_{x \in \mathbb{R}^n} F(x)$$

has at least one solution, or, similarly, $\arg\min F \neq \emptyset$ (again: without further conditions, it is composed of many points).

Existence of minimisers

We have everything we need to provide conditions for showing the **extistence** of minimisers for the problem considered.

Theorem (existence of minimisers)

Let $F \in \mathcal{P}$. If F is l.s.c. and coercive, then F admits a minimiser.

Note I: Equivalently said, the problem:

$$\arg\min_{x \in \mathbb{R}^n} F(x)$$

has at least one solution, or, similarly, $\arg\min F \neq \emptyset$ (again: without further conditions, it is composed of many points).

Note II: this result generalises the standard Bolzano-Weirestrass theorem saying that problem

$$\boxed{\min_{x \in C} \; F(x)}$$

for compact C and *continuous* F, admits at least a solution ($\arg\min F \neq \emptyset$).

Uniqueness of minimisers

So far, only existence of minimisers. How to guarantee uniqueness?

Theorem (existence+uniqueness of minimisers)

Let $F \in \mathcal{P}$. If F is l.s.c., coercive and strictly convex, then F admits a unique minimiser.

Note I: Equivalently said, the problem:

$$\arg\min_{x \in \mathbb{R}^n} F(x)$$

has **exactly** one solution, or, similarly, $\arg\min F = \{x^*\}$, a singleton.

Introducing smoothness

Recalling the original problem

$$\boxed{\min_{x \in \mathbb{R}^n} \left\{ F(x) := f(x) + g(x) \right\},}$$

we will assume at least both f and g to be proper, l.s.c. and convex to have at least one solution.

Introducing smoothness

Recalling the original problem

$$\boxed{\min_{x \in \mathbb{R}^n} \; \{F(x) := f(x) + g(x)\}\,,}$$

we will assume at least both f and g to be proper, l.s.c. and convex to have at least one solution.

As said above, most of the problems encountered in this course will have a *nice* component f for which **further regularity** (smoothness) holds. . .

Introducing smoothness

Recalling the original problem

$$\boxed{\min_{x \in \mathbb{R}^n} \ \{F(x) := f(x) + g(x)\},}$$

we will assume at least both f and g to be proper, l.s.c. and convex to have at least one solution.

As said above, most of the problems encountered in this course will have a *nice* component f for which **further regularity** (smoothness) holds...

To simplify things, let us first look at:

$$\boxed{\min_{x \in \mathbb{R}^n} \ \{F(x) := f(x) + \cancel{g(x)}\},}$$

so ignore the presence of the non-nice component g.

We would like to provide a characterisation of the minimisers of a function f in terms of a suitable notion of "∇f". We use Gâteaux differentiability.

Gâteaux differentiability

Let $f \in \mathcal{P}$ and let $x \in \mathbb{R}^n$ s.t. $f(x) < +\infty$. For $v \in \mathbb{R}^n$, we denote the *directional derivative* in x along the direction v the limit

$$f'(x; v) = f'(x)[v] := \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t},$$

when it exists. If there exists w such that:

$$(\forall v \in \mathbb{R}^n) \quad f'(x; v) = \langle w, v \rangle,$$

then we say that f is *Gâteaux differentiable* in x and denote by $\nabla f(x) = w$ the *Gâteaux derivative* (or, simply, the *gradient*) of f at x.

Optimality conditions (smooth case)

We can now express the optimality of point in terms of ∇f.

Theorem (Fermat's rule)

Let $F \in \mathcal{P}$ be convex and differentiable at point x^*. Then:

$$x^* \in \underset{x \in \mathbb{R}^n}{\arg\min}\, f(x) \quad \Longleftrightarrow \quad \nabla f(x^*) = 0,$$

hence solving an optimisation problem corresponds to solve a system of equations.

Exercise: for $f(x) = \frac{1}{2}\|Ax - y\|_2^2$, compute ∇f.

Lipschitz continuity

You may have seen this notion already applied to a function $h : \mathbb{R}^n \to \mathbb{R}$:

$$\exists L \geq 0 : \forall x, y \in \mathbb{R}^n \quad |h(x) - h(y)| \leq L\|x - y\|.$$

It is a condition controlling the growth of h.

Lipschitz continuity

You may have seen this notion already applied to a function $h : \mathbb{R}^n \to \mathbb{R}$:

$$\exists L \geq 0 : \forall x, y \in \mathbb{R}^n \quad |h(x) - h(y)| \leq L\|x - y\|.$$

It is a condition controlling the growth of h.

In the framework of first-order optimisation methods, such condition is rather important when assumed on the gradient of the differentiable function $f \in \mathcal{P}$.

Gradient Lipschitz continuity

Let $f \in \mathcal{P}$ be differentiable. We say that f is a *gradient Lipschitz continuous function* with constant $L \geq 0$ iff:

$$\exists L \geq 0 : \forall x, y \in \mathbb{R}^n \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Exercise: Compute the Lipschitz constant L of the gradient of the function $f(x) = \frac{1}{2}\|Ax - y\|_2^2$.

A first (but useful!) algorithm

Gradient descent (GD) algorithm: ubiquitous in many applications (training of neural nets...) for minimising convex, differentiable and proper functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

Gradient descent algorithm

Input: $x_0 \in \mathbb{R}^n$ (initial guess), $\tau \in \left(0, \frac{2}{L}\right]$ (step-size)

Iterate for $k \geq 0$:

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

till **convergence**.

$\Rightarrow$ the algorithm converges to a (possibly not unique) minimiser of f

A first (but useful!) algorithm

Gradient descent (GD) algorithm: ubiquitous in many applications (training of neural nets...) for minimising convex, differentiable and proper functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

Gradient descent algorithm

Input: $x_0 \in \mathbb{R}^n$ (initial guess), $\tau \in \left(0, \frac{2}{L}\right]$ (step-size)

Iterate for $k \geq 0$:

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

till **convergence**.

$\Rightarrow$ the algorithm converges to a (possibly not unique) minimiser of f

- Choice of τ: important to guarantee convergence (need to be sufficiently small), it relates to the inverse of L.
- Convex assumption: no dependence on x_0.
- Stopping criterion: relative error $\|x_{k+1} - x_k\| \leq \texttt{tol}$ or gradient check $\|\nabla f(x_{k+1})\| \leq \texttt{tol}$ (approaching 0).

Life is not smooth. . .

Unfortunately, in many applications the function g in

$$\boxed{\min_{x \in \mathbb{R}^n} \; \{F(x) := f(x) + g(x)\} \,,}$$

is different from 0. Typically, g is convex, but **non differentiable** so its gradient (and henceforth the one of F) cannot be defined. . .

Subdifferentiability

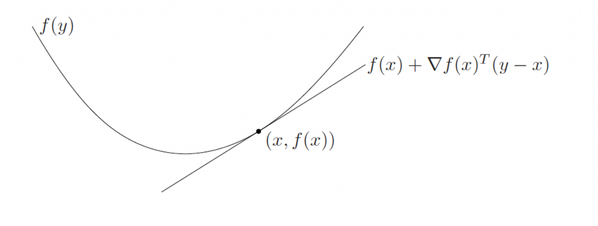
A preliminary observation

One can show that if f is differentiable:

$$f \text{ is convex} \quad \Leftrightarrow \quad (\forall x, y \in \mathbb{R}^n) \quad f(y) \geq \underbrace{f(x) + \nabla f(x)^T(y-x)}_{=:\phi(y;x)}$$

Or, in other words:

- the function $\phi(y;x)$ is an affine lower bound/estimator of f
- the tangent to f is below f at all points.



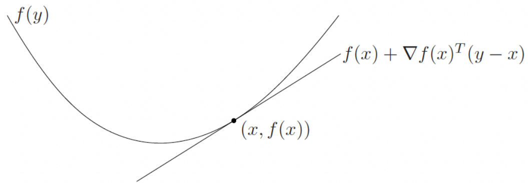
A preliminary observation

One can show that if f is differentiable:

$$f \text{ is convex} \quad \Leftrightarrow \quad (\forall x, y \in \mathbb{R}^n) \quad f(y) \geq \underbrace{f(x) + \nabla f(x)^T (y - x)}_{=:\phi(y;x)}$$

Or, in other words:

- the function $\phi(y; x)$ is an affine lower bound/estimator of f
- the tangent to f is below f at all points.



…what if f is not differentiable (but convex)?

Subdifferential and subgradients

Let us know look at the *non-nice* component g of the problem:

$$\boxed{\min_{x\in\mathbb{R}^n} \ \left\{F(x) := \cancel{f(x)} + g(x)\right\},}$$

Subdifferential and subgradients

Let us know look at the *non-nice* component g of the problem:

$$\min_{x \in \mathbb{R}^n} \quad \{F(x) := \cancel{f(x)} + g(x)\},$$

Subdifferentials and subgradients

Let $g \in \mathcal{P}$ be **convex**. Then, a vector $p \in \mathbb{R}^n$ is a *subgradient* of g at point $x \in \mathrm{dom}(g)$ iff:

$$g(y) \geq g(x) + p^T(y - x), \qquad \forall y \in \mathrm{dom}(g)$$

The set of all subgradients at a point $x \in \mathbb{R}^n$ is called the *subdifferential* of g in x, and it is the denoted by:

$$\partial g(x) = \{p \in \mathbb{R}^n : p \text{ is a subgradient of } g \text{ at point } x\}$$

Subdifferential and subgradients

Let us know look at the *non-nice* component g of the problem:

$$\min_{x \in \mathbb{R}^n} \; \{F(x) := \cancel{f(x)} + g(x)\},$$

Subdifferentials and subgradients

Let $g \in \mathcal{P}$ be **convex**. Then, a vector $p \in \mathbb{R}^n$ is a *subgradient* of g at point $x \in \mathrm{dom}(g)$ iff:

$$g(y) \geq g(x) + p^T(y - x), \qquad \forall y \in \mathrm{dom}(g)$$

The set of all subgradients at a point $x \in \mathbb{R}^n$ is called the *subdifferential* of g in x, and it is the denoted by:

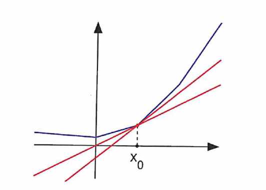
$$\partial g(x) = \{p \in \mathbb{R}^n : p \text{ is a subgradient of } g \text{ at point } x\}$$

Interpretation:

- $p \in \partial g(x)$ if and only if $\phi(y; x) = g(x) + p^T(y - x)$ is a lower affine bound for g.
- $\partial g(x)$ collects all the **slopes** of the "tangent" straight lines passing through x.

Remarks

In general, $\partial g(x)$ contains many elements ("many derivatives at each point").



Multiple subgradients at a non-differentiable point x_0.

However, one can show that <u>if g is differentiable in x</u>, then:

$$\partial g(x) = \{\nabla g(x)\},$$

i.e. the only element in $\partial g(x)$ is the (classical) gradient of g in x.

Exercise: compute $\partial g(x)$ at all $x \in \mathbb{R}$ for the 1D function $g(x) = |x|$ and provide a graphical representation of the result.

Further exercises for more rules on subdifferential calculus.

Questions?

calatroni@i3s.unice.fr