Optimization
– Basics –

Nelly Pustelnik

CNRS, Laboratoire de Physique de l'ENS de Lyon, France

(several slides in this part traced back Tutorial ICASSP 2014 written in collaboration with **Jean-Christophe Pesquet** from Centre de Vision Numérique, CentraleSupelec, University Paris-Saclay, Inria, France.)

Hilbert spaces

A (real) **Hilbert space** $\mathcal{H}$ is a complete real vector space endowed with an inner product $\langle \cdot \mid \cdot \rangle$. The associated norm is

$$(\forall x \in \mathcal{H}) \qquad \|x\| = \sqrt{\langle x \mid x \rangle}.$$

- Particular case: $\mathcal{H} = \mathbb{R}^N$ (Euclidean space with dimension N).
- Course dedicated to finite dimension.

Norm and adjoint

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.

A linear operator $L\colon \mathcal{H} \to \mathcal{G}$ is **bounded** (or continuous) if

$$\|L\| = \sup_{\|x\|_{\mathcal{H}} \leq 1} \|Lx\|_{\mathcal{G}} < +\infty$$

Norm and adjoint

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.

A linear operator $L\colon \mathcal{H} \to \mathcal{G}$ is **bounded** (or continuous) if

$$\|L\| = \sup_{\|x\| \leq 1} \|Lx\| < +\infty$$

Norm and adjoint

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
A linear operator $L\colon \mathcal{H} \to \mathcal{G}$ is **bounded** (or continuous) if

$$\|L\| = \sup_{\|x\| \leq 1} \|Lx\| < +\infty$$

- In finite dimension, every linear operator is bounded.

Norm and adjoint

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
A linear operator $L \colon \mathcal{H} \to \mathcal{G}$ is **bounded** (or continuous) if

$$\|L\| = \sup_{\|x\| \leq 1} \|Lx\| < +\infty$$

- In finite dimension, every linear operator is bounded.

$\mathcal{B}(\mathcal{H}, \mathcal{G})$: Banach space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$.

Norm and adjoint

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its **adjoint** L^* is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

$$(\forall (x, y) \in \mathcal{H} \times \mathcal{G}) \qquad \langle y \mid Lx \rangle_{\mathcal{G}} = \langle L^* y \mid x \rangle_{\mathcal{H}} .$$

Example:

$$\text{If} \qquad L \colon \mathcal{H} \to \mathcal{H}^n \colon x \mapsto (x, \ldots, x)$$

$$\text{then} \qquad L^* \colon \mathcal{H}^n \to \mathcal{H} \colon y = (y_1, \ldots, y_n) \mapsto \sum_{i=1}^{n} y_i$$

Proof: $\langle Lx \mid y \rangle = \langle (x, \ldots, x) \mid (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} \langle x \mid y_i \rangle = \left\langle x \mid \sum_{i=1}^{n} y_i \right\rangle$

Norm and adjoint

Let $\mathcal{H}$ and $\mathcal{G}$ be two Hilbert spaces.
Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Its **adjoint** L^* is the operator in $\mathcal{B}(\mathcal{G}, \mathcal{H})$ defined as

$$(\forall (x, y) \in \mathcal{H} \times \mathcal{G}) \qquad \langle y \mid Lx \rangle = \langle L^* y \mid x \rangle \, .$$

Example:

$$\text{If} \qquad L \colon \mathcal{H} \to \mathcal{H}^n \colon x \mapsto (x, \ldots, x)$$

$$\text{then} \qquad L^* \colon \mathcal{H}^n \to \mathcal{H} \colon y = (y_1, \ldots, y_n) \mapsto \sum_{i=1}^{n} y_i$$

Proof: $\langle Lx \mid y \rangle = \langle (x, \ldots, x) \mid (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} \langle x \mid y_i \rangle = \left\langle x \mid \sum_{i=1}^{n} y_i \right\rangle$

Norm and adjoint

- **About** L^*:
  - ⊙ **Compute gradient and proximity operator** operations (Parts III and IV)
  - ⊙ Dual formulation (cf. Part VI)
  - ⊙ Finite dimensions: If $L \in \mathcal{B}(\mathbb{R}^N, \mathbb{R}^K)$ then $L^* = L^\top$.
  - ⊙ Check the correct implementation by using its definition

$$(\forall (x, y) \in \mathbb{R}^N \times \mathbb{R}^K) \qquad \langle Lx \mid y \rangle = \langle x \mid L^* y \rangle$$

- **About** $\|L\|$:
  - ⊙ Required for **gradient-based** algorithms;
  - ⊙ We have $\|L^*\| = \|L\|$;
  - ⊙ **Normalized power method** $\rightarrow$ Morgane Bergot course

Functional analysis: definitions

$$\text{Find} \quad \widehat{x} \in \operatorname*{Argmin}_{x \in \mathcal{H}} \quad f(x)$$

Class of functions $f \in \Gamma_0(\mathcal{H})$:

- Proper function
- Lower semi-continuous function
- Convex function

Reminder about norms in finite dimension

- **Vectors**
  - Let $x = (x_i)_{1 \leq i \leq N} \in \mathbb{R}^N$.
  - **ℓ_1-norm**: $\|x\|_1 = \sum_i |x_i|$.
  - **ℓ_2-norm**: $\|x\|_2 = \sqrt{\sum_i x_i^2}$.

- **Matrices**
  - Let A be a real symmetric $N \times N$ matrix.
  - **Spectral/eigen decomposition**: A can be factored as

    $$A = Q \Lambda Q^\top$$

    where $Q \in \mathbb{R}^{N \times N}$ is orthogonal (i.e. $\mathrm{Q}^\top Q = \mathrm{Id}$) and $\Lambda = \mathrm{diag}(\lambda_1, \ldots \lambda_N)$ (where the real numbers λ_i are the eigenvalues of A).
    The column of Q form an orthonormal set of eigenvectors of A.
  - **Spectral norm**: $\|A\|_2 = \max_i |\lambda_i|$.
  - **Frobenius norm**: $\|A\|_F = \sqrt{\sum_i \lambda_i^2}$.

Functional analysis: definitions

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ where $\mathcal{H}$ is a Hilbert space.

- The **domain** of f is $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$.
- The function f is **proper** if $\operatorname{dom} f \neq \varnothing$.

Functional analysis: definitions

Let $C \subset \mathcal{H}$.
The **indicator function of C** is

$$(\forall x \in \mathcal{H}) \qquad \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Epigraph

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$. The **epigraph** of f is

$$\operatorname{epi} f = \big\{(x, \zeta) \in \operatorname{dom} f \times \mathbb{R} \;\big|\; f(x) \leq \zeta\big\}$$

Lower semi-continuity

Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$.
f is a **lower semi-continuous** function on $\mathcal{H}$ if and only if $\operatorname{epi} f$ is closed

Lower semi-continuity

Let $f : \mathcal{H} \to \,]-\infty, +\infty]$.
f is a **lower semi-continuous** function on $\mathcal{H}$ if and only if $\operatorname{epi} f$ is closed

- Examples:
  - ⊙ **Do not allow for strict constraints** e.g. $Ax < b$ or $x > 0$;
  - ⊙ Allow for inequality or equality constraints e.g. $Ax = b$, $Ax \leq b$ or $x > 0$;

Lower semi-continuity

Let $f : \mathcal{H} \to\ ]-\infty, +\infty]$.
f is a **lower semi-continuous** function on $\mathcal{H}$ if and only if $\operatorname{epi} f$ is closed

- Examples:
  - ⊙ **Do not allow for strict constraints** e.g. $Ax < b$ or $x > 0$;
  - ⊙ Allow for inequality or equality constraints e.g. $Ax = b$, $Ax \leq b$ or $x > 0$;

- Properties:
  - ⊙ Every continuous function on $\mathcal{H}$ is l.s.c.
  - ⊙ **Every finite sum of l.s.c. functions is l.s.c.**

Convex set

$C \subset \mathcal{H}$ is a **convex set** if

$$(\forall (x,y) \in C^2)(\forall \alpha \in]0,1[) \qquad \alpha x + (1-\alpha) y \in C$$

Convex function: definitions

$f : \mathcal{H} \to \,]-\infty, +\infty]$ is a **convex function** if

$$\big(\forall (x, y) \in \mathcal{H}^2\big)(\forall \alpha \in]0, 1[) \qquad f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

Convex functions: definition

$f : \mathcal{H} \to \,]-\infty, +\infty]$ is convex $\Leftrightarrow$ its epigraph is convex.

Convex functions: definition

$f : \mathcal{H} \to \,]-\infty, +\infty]$ is convex $\Leftrightarrow$ its epigraph is convex.

- Properties :
  - ⊙ Composition of an increasing convex funct. and a convex funct. is convex.
  - ⊙ If $f : \mathcal{H} \to \,]-\infty, +\infty]$ is convex, then $\operatorname{dom} f$ is convex.
  - ⊙ $f : \mathcal{H} \to [-\infty, +\infty[$ is concave if $-f$ is convex.
  - ⊙ Every finite **sum of convex functions is convex**.
- $\Gamma_0(\mathcal{H})$: class of convex, l.s.c., and proper functions from $\mathcal{H}$ to $]-\infty, +\infty]$.
- $\iota_C \in \Gamma_0(\mathcal{H}) \Leftrightarrow C$ is a nonempty closed convex set.

Strictly convex functions

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
f is **strictly convex** if

$$(\forall x \in \operatorname{dom} f)(\forall y \in \operatorname{dom} f)(\forall \alpha \in]0,1[)$$
$$x \neq y \quad \Rightarrow \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y).$$

Functional analysis: minimizers

$$\text{Find} \quad \widehat{x} \in \underset{x \in C}{\operatorname{Argmin}} \quad f(x)$$

- Class of functions $f \in \Gamma_0(\mathcal{H})$:

- **Minimizers**
  - ⊙ Local versus global minimizers
  - ⊙ Coercivity and existence
  - ⊙ Convex function

Minimizers

Let C be a nonempty set of a Hilbert space $\mathcal{H}$.
Let $f : C \to]-\infty, +\infty]$ be a proper function and let $\widehat{x} \in C$.

- $\widehat{x} \in \operatorname{dom} f$ is a **local minimizer** of f if there exists an open neigborhood O of $\widehat{x}$ such that

$$(\forall x \in O \cap C) \qquad f(\widehat{x}) \leq f(x).$$

- $\widehat{x}$ is a **(global) minimizer** of f if

$$(\forall x \in C) \qquad f(\widehat{x}) \leq f(x).$$

Minimizers

Let C be a nonempty set of a Hilbert space $\mathcal{H}$.
Let $f : C \to \left]-\infty, +\infty\right]$ be a proper function and let $\widehat{x} \in C$.

- $\widehat{x}$ is a **strict local minimizer** of f if there exists an open neigborhood O of $\widehat{x}$ such that

$$(\forall x \in (O \cap C) \setminus \{\widehat{x}\}) \qquad f(\widehat{x}) < f(x).$$

- $\widehat{x}$ is a **strict (global) minimizer** of f if

$$(\forall x \in C \setminus \{\widehat{x}\}) \qquad f(\widehat{x}) < f(x).$$

Minimizers of a convex function

Theorem: Let $\mathcal{H}$ be a Hilbert space. Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ be a **proper convex** function such that $\mu = \inf f > -\infty$.

- $\{x \in \mathcal{H} \mid f(x) = \mu\}$ is convex.
- Every local minimizer of f is a global minimizer.
- If f is strictly convex, then there exists at most one minimizer.

Minimizers of a convex function

Theorem: Let $\mathcal{H}$ be a Hilbert space. Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ be a **proper convex** function such that $\mu = \inf f > -\infty$.

- $\{x \in \mathcal{H} \mid f(x) = \mu\}$ is convex.
- Every local minimizer of f is a global minimizer.
- If f is strictly convex, then there exists at most one minimizer.

Existence of a minimizer

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$.
f is **coercive** if $\lim_{\|x\| \to +\infty} f(x) = +\infty$.

Existence and uniqueness of a minimizer

Theorem: Let $\mathcal{H}$ be a Hilbert space and C a **closed convex** subset of $\mathcal{H}$. Let $f \in \Gamma_0(\mathcal{H})$ such that $\operatorname{dom} f \cap C \neq \varnothing$.
If f **is coercive** or C **is bounded**, then there exists $\widehat{x} \in C$ such that

$$f(\widehat{x}) = \inf_{x \in C} f(x).$$

If, moreover, f is strictly convex, this minimizer $\widehat{x}$ is unique.

Functional analysis: minimizers

$$\text{Find} \quad \widehat{x} \in \operatorname*{Argmin}_{x \in C} \quad f(x)$$

- Class of functions $f \in \Gamma_0(\mathcal{H})$:
- Minimizers
- **Differentiability and optimality condition**

Differentiable functions

If $f\colon \mathbb{R}^N \to \mathbb{R}$ is differentiable function in $x \in \mathbb{R}^N$, the **gradient of f at x** is $\nabla f(x) \in \mathbb{R}^N$ and its components are the partial derivatives of f:

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_j} \right)_{1 \leq i \leq N}$$

- Example: Let $x \in \mathbb{R}^N$, $z \in \mathbb{R}^K$ and $A \in \mathbb{R}^{K \times N}$ and $f(x) = \frac{1}{2}\|Ax - z\|^2$, then

$$\nabla f(x) = A^*(Ax - z)$$

Differentiable functions

Let $f\colon \mathbb{R}^N \to \,]-\infty, +\infty]$ be a proper differentiable function in the neighborhood of $x \in \mathbb{R}^N$.

The **directional derivative** of f at x with respect to the direction $y \in \mathbb{R}^N$ is defined as:

$$\langle \nabla f(x) \mid y \rangle = \lim_{\alpha \to 0} \frac{f(x + \alpha y) - f(x)}{\alpha}.$$

Optimality condition

1st order necessary and sufficient condition (P. Fermat)

Let $f \in \Gamma_0(\mathcal{H})$ be continuously differentiable function on $\mathcal{H}$.
$\widehat{x}$ is a global minimizer of f i.e

$$\widehat{x} \in \mathrm{Argmin}_{x \in \mathbb{R}^N} f(x) \qquad \Leftrightarrow \qquad \nabla f(\widehat{x}) = 0$$

- More details about optimality conditions here :
  [Jean-Charles Gilbert course]
  [Nocedal-Wright, 1999]

- Limitations :
  - ⊙ Lead to a N equations - N unknown problem.
  - ⊙ Closed form expression for only few cases.
  - ⊙ If no closed form expression exists, an iterative procedure is required.

Optimality condition

- Example: **Solving mean squares**

Find $\widehat{x} = \mathrm{Argmin}_{x \in \mathbb{R}^N} \|Ax - y\|_2^2$ with $\begin{cases} A \in \mathbb{R}^{N \times N} \text{ full rank} \\ y \in \mathbb{R}^M \end{cases}$

→ Optimality condition:

$$\nabla f(\widehat{x}) = 0 \quad \Leftrightarrow \quad A^\top (A\widehat{x} - y) = 0$$

$$\boxed{\widehat{x} = (A^\top A)^{-1}(A^\top y)}$$

→ **Closed form expression** but sometimes difficult to invert $A^\top A$.

Optimality condition

- Example: **Logistic based criterion:**

$$\text{Find} \quad \widehat{x} \in \operatorname{Argmin}_{x\in\mathbb{R}} \log\big(1+\exp(-yx)\big) \qquad \text{with} \qquad y \in \mathbb{R}$$

$\rightarrow$ Optimality condition:

$$\nabla f(\widehat{x}) = 0 \quad \Leftrightarrow \quad \boxed{\frac{-y\exp(-y\widehat{x})}{1+\exp(-y\widehat{x})} = 0}$$

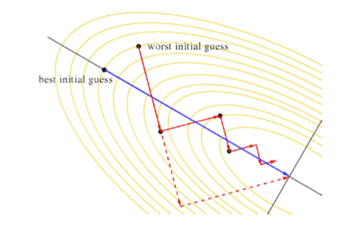
$\rightarrow$ **No closed form expression**. An iterative procedure is required.

Gradient descent

Gradient descent

Let $f : \mathbb{R}^N \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^N$. Let $x_0 \in \mathbb{R}^N$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla f(x_n).$$

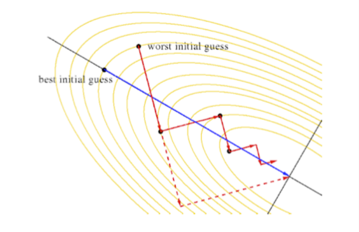


Gradient descent

Gradient descent

Let $f : \mathbb{R}^N \to \mathbb{R}$ be continuously differentiable on $\mathbb{R}^N$. Let $x_0 \in \mathbb{R}^N$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla f(x_n).$$



- An iterative method consists to build a sequence $(x_n)_{n\in\mathbb{N}}$ such that, at each iteration n

$$f(x_{n+1}) < f(x_n)$$

- How to proove convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ to $\widehat{x} \in \mathrm{Argmin}\, f(x)$.
- Choose γ_n for convergence ? For faster convergence ?

Hessian matrix

The second derivative of a real-valued function $f: \mathbb{R}^N \to \mathbb{R}$ or the Hessian matrix of f at x , denoted $\nabla^2 f(x) \in \mathbb{R}^{N\times N}$, is given by

$$\nabla^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{1\leq i\leq N, 1\leq j\leq N}$$

- Let A be a real symmetric $N \times N$ matrix.
- **Spectral/eigen decomposition**: A can be factored as

$$A = Q\Lambda A^\top$$

  where $Q \in \mathbb{R}^{N\times N}$ is orthogonal (i.e. $\mathrm{Q}^\top Q = \mathrm{Id}$) and $\Lambda = \mathrm{diag}(\lambda_1, \ldots \lambda_N)$ (where the real numbers λ_i are the eigenvalues of A The column of Q form an orthonormal set of eigenvectors of A.
- **Spectral norm**: $\|A\|_2 = \max_i |\lambda_i|$.
- **Frobenius norm**: $\|A\|_F = \sqrt{\sum_i \lambda_i^2}$.

L-smooth

A function $f \in \Gamma_0(\mathbb{R}^N)$ is Lipschitz smooth with constant L or L-smooth if its gradient is Lipschitz continuous with constant L:

$$(\forall (x, y) \in \mathbb{R}^N \times \mathbb{R}^N) \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

<u>Remark</u>:

- If f is twice differentiable, a function is L smooth if

$$(\forall x \in \mathbb{R}^N) \quad \nabla^2 f(x) \leq L \cdot \mathrm{Id}$$

- Particular case: $f = \|A \cdot -z\|_2^2$ is β-smooth with $\beta = \mathrm{vp}_{\max}(A^\top A) = \|A\|^2$.

Iterative scheme

Problem: Let $f \in \Gamma_0(\mathbb{R}^N)$, find $\quad \widehat{x} \in \underset{x \in \mathbb{R}^N}{\mathrm{Argmin}}\, f(x)$.

- If f is L-smooth with $L > 0$, the (explicit) **gradient method**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n \nabla f(x_n)$$

$\rightarrow$ Convergence insured when $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2L^{-1}$.

Iterative scheme

Problem: Let $f \in \Gamma_0(\mathbb{R}^N)$, find $\quad \widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}}\, f(x)$.

- If f is L-smooth with $L > 0$, the (explicit) **gradient method**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n \nabla f(x_n)$$

$\rightarrow$ Convergence insured when $0 < \inf_{n\in\mathbb{N}} \gamma_n$ et $\sup_{n\in\mathbb{N}} \gamma_n < 2L^{-1}$.

- If f nonsmooth, the (explicit) **subgradient method** **(be defined next)**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n u_n \quad \text{with} \quad u_n \in \partial f(x_n)$$

$\rightarrow$ Convergence insured when $\gamma_n \in]0, +\infty[$ such that $\sum_{n=0}^{+\infty} \gamma_n^2 < +\infty$ and $\sum_{n=0}^{+\infty} \gamma_n = +\infty$. [Shor, 1979].

Iterative scheme

Problem: Let $f \in \Gamma_0(\mathbb{R}^N)$, find $\quad \widehat{x} \in \underset{x \in \mathbb{R}^N}{\mathrm{Argmin}}\, f(x)$.

- If f is L-smooth with $L > 0$, the (explicit) **gradient method**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n \nabla f(x_n)$$

$\rightarrow$ Convergence insured when $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2L^{-1}$.

- If f nonsmooth, the (explicit) **subgradient method (be defined next)**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n u_n \quad \text{with} \quad u_n \in \partial f(x_n)$$

$\rightarrow$ Convergence insured when $\gamma_n \in]0, +\infty[$ such that $\sum_{n=0}^{+\infty} \gamma_n^2 < +\infty$ and $\sum_{n=0}^{+\infty} \gamma_n = +\infty$. [Shor, 1979].

- If f nonsmooth, the **implicit subgradient method** is

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n u_n \quad \text{with} \quad u_n \in \partial f(x_{n+1})$$

$\rightarrow$ Convergence insured when $\sum_{n=0}^{+\infty} \gamma_n = +\infty$.

Iterative scheme

Problem: Let $f \in \Gamma_0(\mathbb{R}^N)$, find $\quad \widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}}\, f(x)$.

- If f is L-smooth with $L > 0$, the (explicit) **gradient method**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n \nabla f(x_n)$$

→ Convergence insured when $0 < \inf_{n \in \mathbb{N}} \gamma_n$ et $\sup_{n \in \mathbb{N}} \gamma_n < 2L^{-1}$.

- If f nonsmooth, the (explicit) **subgradient method** **(be defined next)**:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n u_n \quad \text{with} \quad u_n \in \partial f(x_n)$$

→ Convergence insured when $\gamma_n \in]0, +\infty[$ such that $\sum_{n=0}^{+\infty} \gamma_n^2 < +\infty$ and $\sum_{n=0}^{+\infty} \gamma_n = +\infty$. [Shor, 1979].

- If f nonsmooth, the **implicit subgradient method** is

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n u_n \quad \text{with} \quad u_n \in \partial f(x_{n+1})$$

→ Convergence insured when $\sum_{n=0}^{+\infty} \gamma_n = +\infty$ ⇒ **Proximity operator**.