Master MVA
Optimization Reminders
Part II

Jean-Christophe Pesquet

Center for Visual Computing
Inria, CentraleSupélec
Institut Universitaire de France
jean-christophe@pesquet.eu

Iterating projections

Feasibility problem

Problem

Let $\mathcal{H}$ be a Hilbert space. Let $m \in \mathbb{N} \setminus \{0, 1\}$.

Let $(C_i)_{1 \leq i \leq m}$ be closed convex subsets of $\mathcal{H}$ such that $\bigcap_{i=1}^{m} C_i \neq \varnothing$.

We want to

$$\text{Find } \widehat{x} \in \bigcap_{i=1}^{m} C_i.$$

POCS (Projection Onto Convex Sets) algorithm

For every $n \in \mathbb{N} \setminus \{0\}$, let $i_n - 1$ denote the remainder after division of $n - 1$ by m.

$$\begin{array}{l} \text{Set } x_0 \in \mathcal{H} \\ \text{For } n = 1, \ldots \\ \lfloor \; x_{n+1} = P_{C_{i_n}}(x_n). \end{array}$$

Feasibility problem

Problem

Let $\mathcal{H}$ be a Hilbert space. Let $m \in \mathbb{N} \setminus \{0,1\}$.

Let $(C_i)_{1\leq i\leq m}$ be closed convex subsets of $\mathcal{H}$ such that $\bigcap_{i=1}^{m} C_i \neq \varnothing$.

We want to

$$\text{Find } \widehat{x} \in \bigcap_{i=1}^{m} C_i.$$

POCS (Projection Onto Convex Sets) algorithm

Let $(\lambda_n)_{n\geq 1}$ be a sequence of $[\epsilon_1, 2-\epsilon_2]$ with $(\epsilon_1,\epsilon_2) \in \,]0,+\infty[^2$ such that $\epsilon_1+\epsilon_2 < 2$.
For every $n \in \mathbb{N}\setminus\{0\}$, let $i_n - 1$ denote the remainder after division of $n-1$ by m.

$$\begin{array}{l} \text{Set } x_0 \in \mathcal{H} \\ \text{For } n = 1, \ldots \\ \lfloor \; x_{n+1} = x_n + \lambda_n (P_{C_{i_n}}(x_n) - x_n). \end{array}$$

Convergence of POCS

Theorem

The sequence $(x_n)_{n\in\mathbb{N}}$ generated by the POCS algorithm converges to a point in $\bigcap_{i=1}^{m} C_i$.

Exercise 1

Let $\mathcal{H}$ be a real Hilbert space.

1. Let $c \in \mathcal{H}$ and $\rho \in \left]0, +\infty\right[$. What is the expression of the projection onto a closed ball $B(c, \rho)$ with center c and radius ρ ?
2. We consider three closed balls which are assumed to have a common point. Propose an algorithms to compute such a point.

Lagrange duality

Constrained optimization problem

Let $\mathcal{H}$ be a Hilbert space. Let $f\colon \mathcal{H} \to \,]-\infty, +\infty]$.
Let $(m, q) \in \mathbb{N}^2$. For every $i \in \{1, \ldots, m\}$, let $g_i\colon \mathcal{H} \to \mathbb{R}$ and for every $j \in \{1, \ldots, q\}$, let $h_j\colon \mathcal{H} \to \mathbb{R}$.
Let

$$C = \{x \in \mathcal{H} \mid (\forall i \in \{1, \ldots, m\}) \;\; g_i(x) = 0 \\ (\forall j \in \{1, \ldots, q\}) \;\; h_j(x) \leq 0\}.$$

We want to:

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

Remark: A vector $x \in \mathcal{H}$ is said to be feasible if $x \in \operatorname{dom} f \cap C$.

Definitions

The **Lagrange function** (or Lagrangian) associated with the previous problem is defined as

$$(\forall x \in \mathcal{H})(\forall \mu = (\mu_i)_{1\le i\le m} \in \mathbb{R}^m)(\forall \lambda = (\lambda_j)_{1\le j\le q} \in [0,+\infty[^q)$$
$$\mathcal{L}(x,\mu,\lambda) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{q} \lambda_j h_j(x).$$

The vectors μ and λ are called **Lagrange multipliers**.

Remark: $\operatorname{dom}\mathcal{L} = \operatorname{dom} f \times \mathbb{R}^m \times [0,+\infty[^q$.

Saddle points

$(\widehat{x}, \widehat{\mu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is saddle point of $\mathcal{L}$ if

$$\big(\forall (x, \mu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q\big) \qquad \mathcal{L}(\widehat{x}, \mu, \lambda) \leq \mathcal{L}(\widehat{x}, \widehat{\mu}, \widehat{\lambda}) \leq \mathcal{L}(x, \widehat{\mu}, \widehat{\lambda}).$$

Saddle points

$(\widehat{x}, \widehat{\mu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is saddle point of $\mathcal{L}$ if

$$\big(\forall (x, \mu, \lambda) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q\big) \qquad \mathcal{L}(\widehat{x}, \mu, \lambda) \le \mathcal{L}(\widehat{x}, \widehat{\mu}, \widehat{\lambda}) \le \mathcal{L}(x, \widehat{\mu}, \widehat{\lambda}).$$

Theorem

Let $\underline{\mathcal{L}}$ and $\overline{\mathcal{L}}$ be defined as

$$(\forall (\mu, \lambda) \in \mathbb{R}^m \times [0, +\infty[^q) \quad \underline{\mathcal{L}}(\mu, \lambda) = \inf_{x \in \mathcal{H}} \mathcal{L}(x, \mu, \lambda)$$

$$(\forall x \in \mathcal{H}) \quad \overline{\mathcal{L}}(x) = \sup_{\mu \in \mathbb{R}^m, \lambda \in [0, +\infty[^q} \mathcal{L}(x, \mu, \lambda).$$

$(\widehat{x}, \widehat{\mu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$ if and only if

$$(\forall x \in \mathcal{H}) \quad \overline{\mathcal{L}}(\widehat{x}) \le \overline{\mathcal{L}}(x)$$

$$(\forall (\mu, \lambda) \in \mathbb{R}^m \times [0, +\infty[^q) \quad \underline{\mathcal{L}}(\mu, \lambda) \le \underline{\mathcal{L}}(\widehat{\mu}, \widehat{\lambda})$$

$$\underline{\mathcal{L}}(\widehat{\mu}, \widehat{\lambda}) = \overline{\mathcal{L}}(\widehat{x}).$$

Saddle points

Theorem

Let $\underline{\mathcal{L}}$ and $\overline{\mathcal{L}}$ be defined as

$$(\forall(\mu,\lambda)\in\mathbb{R}^m\times[0,+\infty[^q)\quad \underline{\mathcal{L}}(\mu,\lambda)=\inf_{x\in\mathcal{H}}\mathcal{L}(x,\mu,\lambda)$$
$$(\forall x\in\mathcal{H})\quad \overline{\mathcal{L}}(x)=\sup_{\mu\in\mathbb{R}^m,\lambda\in[0,+\infty[^q}\mathcal{L}(x,\mu,\lambda).$$

$(\widehat{x},\widehat{\mu},\widehat{\lambda})\in\mathcal{H}\times\mathbb{R}^m\times[0,+\infty[^q$ is a saddle point of $\mathcal{L}$ if and only if

$$(\forall x\in\mathcal{H})\quad \overline{\mathcal{L}}(\widehat{x})\le\overline{\mathcal{L}}(x)$$
$$(\forall(\mu,\lambda)\in\mathbb{R}^m\times[0,+\infty[^q)\quad \underline{\mathcal{L}}(\mu,\lambda)\le\underline{\mathcal{L}}(\widehat{\mu},\widehat{\lambda})$$
$$\underline{\mathcal{L}}(\widehat{\mu},\widehat{\lambda})=\overline{\mathcal{L}}(\widehat{x}).$$

Remark: $\overline{\mathcal{L}}$ is called the primal Lagrange function and $\underline{\mathcal{L}}$ the dual Lagrange function.

Sufficient condition for a constrained minimum

Assume that there exists a feasible point.
If $(\widehat{x}, \widehat{\mu}, \widehat{\lambda}) \in \mathcal{H} \times \mathbb{R}^m \times [0, +\infty[^q$ is a saddle point of $\mathcal{L}$,
then $\widehat{x}$ is minimizer of f over C.
In addition, the complementary slackness condition holds:

$$(\forall j \in \{1, \dots, q\}) \qquad \widehat{\lambda}_j h_j(\widehat{x}) = 0.$$

Convex case

Assume that f is a convex function, $(g_i)_{1\le i\le m}$ are affine functions and $(h_j)_{1\le j\le q}$ are convex functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \operatorname{dom} f$ such that

$$
\begin{aligned}
(\forall i \in \{1,\ldots,m\}) \qquad & g_i(\overline{x}) = 0\\
(\forall j \in \{1,\ldots,q\}) \qquad & h_j(\overline{x}) < 0.
\end{aligned}
$$

$\widehat{x}$ is a minimizer of f over C if and only if there exists $\widehat{\mu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0,+\infty[^q$ such that $(\widehat{x},\widehat{\mu},\widehat{\lambda})$ is a saddle point of the Lagrangian.

Convex case

Assume that f is a convex function, $(g_i)_{1\leq i\leq m}$ are affine functions and $(h_j)_{1\leq j\leq q}$ are convex functions. Assume that the Slater condition holds, i.e. there exists $\overline{x} \in \operatorname{dom} f$ such that

$$\begin{aligned} &(\forall i \in \{1,\ldots,m\}) && g_i(\overline{x}) = 0 \\ &(\forall j \in \{1,\ldots,q\}) && h_j(\overline{x}) < 0. \end{aligned}$$

$\widehat{x}$ is a minimizer of f over C if and only if there exists $\widehat{\mu} \in \mathbb{R}^m$ and $\widehat{\lambda} \in [0,+\infty[^q$ such that $(\widehat{x},\widehat{\mu},\widehat{\lambda})$ is a saddle point of the Lagrangian.

Remark: Under the assumptions of the above theorem, if $\widehat{x}$ is a minimizer of f over C then $\mathcal{L}(\cdot,\widehat{\mu},\widehat{\lambda})$ is a convex function which is minimum at $\widehat{x}$. This optimality condition is often used to calculate $\widehat{x}$, in conjunction with the complementary slackness condition.

Differentiable case

Karush-Kuhn-Tucker (KKT) theorem

Assume that f, $(g_i)_{1\leq i\leq m}$, and $(h_j)_{1\leq j\leq q}$ are continuously differentiable on $\mathcal{H}=\mathbb{R}^N$.

Assume that $\widehat{x}$ is a local minimizer of f over C satisfying the following Mangasarian-Fromovitz constraint qualification conditions :

(*i*) $\{\nabla g_i(\widehat{x}) \mid i\in\{1,\ldots,m\}\}$ is a family of linearly independent vectors;

(*ii*) there exists $z\in\mathbb{R}^N$ such that

$$\begin{aligned}(\forall i\in\{1,\ldots,m\}) \qquad & \langle \nabla g_i(\widehat{x}) \mid z\rangle = 0\\ (\forall j\in J(\widehat{x})) \qquad & \langle \nabla h_j(\widehat{x}) \mid z\rangle < 0\end{aligned}$$

where $J(\widehat{x})=\{j\in\{1,\ldots,q\}\mid h_j(\widehat{x})=0\}$ is the set of active inequality constraints at $\widehat{x}$.

Then, there exists $\widehat{\mu}\in\mathbb{R}^N$ and $\widehat{\lambda}\in[0,+\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot,\widehat{\mu},\widehat{\lambda})$ and the complementary slackness condition holds.

Differentiable case

Karush-Kuhn-Tucker (KKT) theorem

Assume that f, $(g_i)_{1\leq i\leq m}$, and $(h_j)_{1\leq j\leq q}$ are continuously differentiable on $\mathcal{H}=\mathbb{R}^N$.

Assume that $\widehat{x}$ is a local minimizer of f over C satisfying the following Mangasarian-Fromovitz constraint qualification conditions:

(*i*) $\{\nabla g_i(\widehat{x}) \mid i \in \{1,\ldots,m\}\}$ is a family of linearly independent vectors;

(*ii*) there exists $z \in \mathbb{R}^N$ such that

$$\begin{aligned}(\forall i \in \{1,\ldots,m\}) \qquad & \langle \nabla g_i(\widehat{x}) \mid z\rangle = 0\\ (\forall j \in J(\widehat{x})) \qquad & \langle \nabla h_j(\widehat{x}) \mid z\rangle < 0.\end{aligned}$$

Then, there exists $\widehat{\mu} \in \mathbb{R}^N$ and $\widehat{\lambda} \in [0,+\infty[^q$ such that $\widehat{x}$ is a critical point of $\mathcal{L}(\cdot,\widehat{\mu},\widehat{\lambda})$ and the complementary slackness condition holds.

Remark: A sufficient condition for Mangasarian-Fromovitz conditions to be satisfied is that $\{\nabla g_i(\widehat{x}) \mid i \in \{1,\ldots,m\}\} \cup \{\nabla h_j(\widehat{x}) \mid j \in J(\widehat{x})\}$ is a family of linearly independent vectors.

Exercise 2

Let f be defined as

$$\left(\forall x = (x^{(i)})_{1\leq i\leq N} \in \mathbb{R}^N\right) \qquad f(x) = \sum_{i=1}^{N} \exp(x^{(i)})$$

with $N > 1$. We want to find a minimizer of f on $\mathbb{R}^N$ subject to the constraints

$$\sum_{i=1}^{N} x^{(i)} = 1$$

$$(\forall i \in \{1, \ldots, N\}) \quad x^{(i)} \geq 0.$$

1. What can be said about the existence/uniqueness of a solution to this problem ?
2. Apply the Lagrange multiplier method.

Exercise 3

By using the Lagrange multipliers method, solve the following problem

$$\underset{x=(x^{(i)})_{1\leq i\leq N}\in B}{\text{maximize}} \quad (x^{(N)})^3 - \frac{1}{2}(x^{(N)})^2$$

where B is the unit sphere, centered at 0, of $\mathbb{R}^N$.

A few algorithms

Problem

Let $\mathcal{H}$ be a Hilbert space.
Let $f : \mathcal{H} \to \mathbb{R}$ be differentiable.
Let C be a nonempty closed convex subset of $\mathcal{H}$.
We want to:

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

Problem

Let $\mathcal{H}$ be a Hilbert space.
Let $f : \mathcal{H} \to \mathbb{R}$ be differentiable.
Let C be a nonempty closed convex subset of $\mathcal{H}$.
We want to:

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

Objective: Build a sequence $(x_n)_{n\in\mathbb{N}}$ converging to a minimizer.

Principle of first-order methods

- If f is differentiable, then, at iteration n, we have

$$(\forall x \in \mathcal{H}) \qquad f(x) = f(x_n) + \langle \nabla f(x_n) \mid x - x_n \rangle + o(\|x - x_n\|).$$

  So if $\|x_{n+1} - x_n\|$ is small enough and x_{n+1} is chosen such that

$$\langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle < 0$$

  then $f(x_{n+1}) < f(x_n)$.
- In particular, the steepest descent direction is given by

$$x_{n+1} - x_n = -\gamma_n \nabla f(x_n), \qquad \gamma_n \in \,]0, +\infty[\,.$$

- To secure that the solution belongs to C we can add a projection step.
- A relaxation parameter λ_n can also be added.

Principle of first-order methods

The **projected gradient algorithm** has the following form:

$$x_{n+1} = x_n + \lambda_n\big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n\big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$.

Principle of first-order methods

The **projected gradient algorithm** has the following form:

$$x_{n+1} = x_n + \lambda_n\big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n\big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$.

Remark: x is a fixed point of the projected gradient iteration if and only if $x \in C$ and

$$(\forall y \in C) \qquad \langle \nabla f(x) \mid y - x \rangle \geq 0.$$

Principle of first-order methods

The projected gradient algorithm has the following form:

$$x_{n+1} = x_n + \lambda_n\big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n)\big)$$

where $\gamma_n \in \;]0,+\infty[$ and $\lambda_n \in]0,1]$.

Remark: x is a fixed point of the projected gradient iteration if and only if $x \in C$ and

$$(\forall y \in C) \qquad \langle \nabla f(x) \mid y - x\rangle \geq 0.$$

Proof: If x is a fixed point, then

$$\begin{aligned} & x = x + \lambda_n\big(P_C(x - \gamma_n \nabla f(x)) - x\big) \\ \Leftrightarrow \quad & x = P_C(x - \gamma_n \nabla f(x)). \end{aligned}$$

According to the characterization of the projection, for every $y \in C$,

$$\begin{aligned} & \langle x - \gamma_n \nabla f(x) - x \mid y - x\rangle \leq 0 \\ \Leftrightarrow \quad & \langle \nabla f(x) \mid y - x\rangle \geq 0. \end{aligned}$$

Principle of first-order methods

The projected gradient algorithm has the following form:

$$x_{n+1} = x_n + \lambda_n \big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n \big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$.

Remark:

- x is a fixed point of the projected gradient iteration if and only if $x \in C$ and
$$(\forall y \in C) \qquad \langle \nabla f(x) \mid y - x \rangle \geq 0.$$
- When f is convex, x is a fixed point of the projected gradient iteration if and only if x is a global minimizer of f over C.

Principle of first-order methods

The **projected gradient algorithm** has the following form:

$$x_{n+1} = x_n + \lambda_n \big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n \big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$.

Remark: If $C = \mathcal{H}$ and $\lambda_n = 1$, we recover the standard gradient descent iteration:

$$x_{n+1} = x_n - \gamma_n \nabla f(x_n).$$

Convergence

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in \,]0, +\infty[$, i.e.

$$(\forall (x, y) \in \mathcal{H}^2) \quad \|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|.$$

Assume that $\mathrm{Argmin}_{x \in C} f(x) \neq \varnothing$.
Assume that $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n \in \mathbb{N}} \lambda_n > 0$, and $\sup_{n \in \mathbb{N}} \lambda_n \leq 1$.
Then the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the projected gradient algorithm converges to a minimizer of f over C.

Convergence

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in \,]0, +\infty[$.
Assume that $\operatorname{Argmin}_{x \in C} f(x) \neq \varnothing$.
Assume that $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n \in \mathbb{N}} \lambda_n > 0$, and $\sup_{n \in \mathbb{N}} \lambda_n \leq 1$.
Then the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the projected gradient algorithm converges to a minimizer of f over C.
In addition, if f is strongly convex, the convergence is linear, i.e. there exists $\chi \in [0, 1[$ such that

$$(\forall n \in \mathbb{N}) \quad \|x_n - \widehat{x}\| \leq \chi^n \|x_0 - \widehat{x}\|$$

where $\widehat{x}$ is the unique minimizer of f over C.

Convergence

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in \,]0, +\infty[$.
Assume that $\operatorname{Argmin}_{x \in C} f(x) \neq \varnothing$.
Assume that $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $\sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and $\sup_{n\in\mathbb{N}} \lambda_n \leq 1$.
Then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the projected gradient algorithm converges to a minimizer of f over C.

Remark: If f is non convex with a β-Lipschtzian gradient, it can only be proved that $(f(x_n))_{n\in\mathbb{N}}$ is a converging sequence provided that $\gamma_n \leq 1/\beta$.

Example: Uzawa algorithm

Problem

Let $\mathcal{L}\colon \mathcal{H} \times [0,+\infty[^q \to \mathbb{R}$ be differentiable with respect to its second argument. We want to find a saddle point of $\mathcal{L}$

Solution

Set $\lambda_0 \in [0,+\infty[^q$

For $n = 1, \ldots$

$$\left\lfloor \begin{array}{l} \text{Set } \gamma_n \in \,]0,+\infty[,\ \rho_n \in]0,1] \\ x_n \in \mathrm{Argmin}\mathcal{L}(\cdot,\lambda_n) \\ \lambda_{n+1} = \lambda_n + \rho_n(P_{[0,+\infty[^q}(\lambda_n + \gamma_n \nabla_\lambda \mathcal{L}(x_n,\lambda_n)) - \lambda_n). \end{array} \right.$$

Principle of second-order methods

- If f is twice differentiable, then, at iteration n, we have

$$(\forall x \in \mathcal{H}) \quad f(x) = f(x_n) + \langle \nabla f(x_n) \mid x - x_n \rangle + \frac{1}{2} \big\langle (x - x_n) \mid \nabla^2 f(x_n)(x - x_n) \big\rangle + o(\|x - x_n\|^2).$$

- If $\nabla^2 f(x_n)$ is positive definite, the minimizer x_{n+1} of the quadratic term is given by Newton's iteration

$$\nabla f(x_n) + \nabla^2 f(x_n)(x_{n+1} - x_n) = 0$$
$$\Leftrightarrow \quad x_{n+1} = x_n - \big(\nabla^2 f(x_n)\big)^{-1} \nabla f(x_n).$$

Convergence

Convergence theorem

Let $f \in \mathcal{H} \;\; \to \;]-\infty, +\infty]$ be three times continuously differentiable in a neigborhood of a local minimizer $\widehat{x}$ and assume that $\nabla^2 f(\widehat{x})$ is positive definite.
Then, there exists $\epsilon \in \;]0, +\infty[$ such that, if $\|x_0 - \widehat{x}\| \leq \epsilon$, then $(x_n)_{n\in\mathbb{N}}$ converges to $\widehat{x}$.
In addition, the convergence is quadratic, i.e. there exists $\kappa \in \;]0, +\infty[$ such that

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - \widehat{x}\| \leq \kappa \|x_n - \widehat{x}\|^2.$$

Numerical behaviour

- Although the convergence of Newton's algorithm is faster than the gradient descent in terms of iteration number, the computational cost of each iteration is higher.

- To improve the convergence guarantees of Newton's algorithm, we may practically modify it as follows:

$$(\forall n \in \mathbb{N}) \qquad x_{n+1} = x_n - \gamma_n \big(\nabla^2 f(x_n) + \lambda_n \mathrm{Id}\big)^{-1} \nabla f(x_n),$$

  with $(\gamma_n, \lambda_n) \in \,]0, +\infty[^2$.

- Quasi-Newton algorithms read

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

  where H_n is a definite positive matrix providing some approximation to the Hessian.

Exercise 4

Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $y \in \mathcal{G}$ and let $\alpha \in \,]0, +\infty[$.
We want to minimize the function defined as

$$(\forall x \in \mathcal{H}) \qquad f(x) = \frac{1}{2}\|Lx - y\|^2 + \frac{\alpha}{2}\|x\|^2.$$

1. Give the form of the gradient descent algorithm allowing us to solve this problem.
2. How does Newton's algorithm read for this function ?
3. Study the convergence of the gradient descent algorithm by performing the eigendecomposition of L^*L.