INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE

Optimization-II

Tutorials

4th year

Copyright © 2024 Frédéric de Gournay

Publié par INSA de Toulouse

frederic@degournay.fr

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the "License"). You may not use this file except in compliance with the License. You may obtain a copy of the License at `http://creativecommons.org/licenses/by-nc/3.0`. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "as is" basis, without warranties or conditions of any kind, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Première édition, Janvier 2024.

1 Refreshers

Exercise 1.1 will test your computational skills, you must know Leibniz formula and the relationship between the gradient and the Jacobian.

(TD1 : 20 mins) Exercise 1.1

Let $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$.

1. Let $\Psi : \mathbb{R}^m \mapsto \mathbb{R}$ be defined as
$$\Psi(y) = \sum_{i=1}^{m} y_i^4 + 1.$$
Compute $\nabla\Psi(y)$ and $H[\Psi]$. Is Ψ convex ?
2. Let $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $\varphi(x) = Ax - b$. Compute the jacobian of φ
3. Let $f(x) = \Psi(Ax - b)$ show that $\nabla f(x) = A^\star \nabla\Psi(Ax - b)$ by two different methods. The first method is a direct Taylor expansion (compute $f(x+h)-f(x)$ for small h), for the second method use the chain rule.

In Exercise 1.2, we prove a very important formula for least-square approximation. The result of this exercise and the method of proof must be learned by heart, as if your life depends on it[1].

(15 mins) Exercise 1.2

Let $A_1 \in \mathcal{M}_{n,n}(\mathbb{R})$ be a symetric matrix, $b_1 \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Let
$$f_1 : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle A_1 x, x\rangle - \langle b_1, x\rangle + c.$$

1. Show that $\nabla f_1(x) = A_1 x - b_1$ and $H[f_1](x) = A_1$.
2. Under which condition is f_1 convex ? strictly convex ?
3. Let $A_2 \in \mathcal{M}_{n,m}(\mathbb{R})$, $b_2 \in \mathbb{R}^m$ and
$$f_2 : x \in \mathbb{R}^n \mapsto \frac{1}{2}\|A_2 x - b_2\|^2$$
4. Show that $\nabla f_2(x) = A_2^\star(A_2 x - b_2)$ and $H[f_2](x) = A_2^\star A_2$.
5. Under which condition is f_2 convex ? strictly convex ?

In Exercise 1.3, we revise the natural scalar product on the space of matrices (Frobenius), we reprove some formula for the gradient of the determinant and the trace.

(15 mins) Exercise 1.3

We recall that $\mathcal{M}_{n,n}(\mathbb{R})$, the vector space of real square matrices of size n is Euclidean when endowed with the scalar product $\langle A, B\rangle = \mathrm{tr}(A^\top B)$. Show quickly that each of the following map is differentiable on its domain, compute their differential and their gradient.

1. $\mathrm{tr} : A \mapsto \mathrm{tr}(A)$
2. $\det : A \mapsto \det(A)$ when A is invertible (hint, first prove that $\nabla \det(I_n) = I_n$).

[1] Your life may not **really** depend on knowing Exercise 1.2 by heart, but your success in class certainly will.

3. $f : A \mapsto \ln|\det(A)|$ when A is invertible

Exercise 1.4 has exactly the same pedagogical goal as Exercise 1.3, it can be used to see if one masters the techniques of Exercise 1.3.

(15 mins) Exercise 1.4

We work in the space $\mathcal{M}_{m,n}(\mathbb{R})$ of real matrices of size $m \times n$ with the scalar product $\langle U, V\rangle = \mathrm{tr}(U^\top V)$. Let $B \in \mathcal{S}_n$ a symetric $n \times n$ real matrix. Let

$$f : A \in \mathcal{M}_{m,n}(\mathbb{R}) \mapsto \mathrm{tr}(ABA^\top)$$

Give the differential and the gradient of f, you must find $\nabla f(A) = 2AB$.

In Exercise 1.5, the goal is to get a quick intuition on convexity, the idea is to rely on sketches of the solutions and to design an argument from an intuition.

(TD1 : 15 mins) Exercise 1.5

Draw the following functions and prove that they are convex or prove that they are not

1. $f : x \mapsto x^2$
2. $f : x \mapsto e^{-x^2}$
3. $f : x \mapsto e^x$
4. $f : x \mapsto \sqrt{x}$
5. $f : x \mapsto e^{-x}$
6. $f : x \mapsto \frac{1}{x}$

The right approach to Exercise 1.6 is to compute Hessians and to discuss according to the sign of the eigenvalues.

(TD1 : 15 mins) Exercise 1.6

Discuss about the convexity or concavity of the following functions

1. $f(x, y) = (y - x^2)^2 - x^2$
2. $f(x, y) = x^4 + y^4 - (x - y)^2$
3. $f(x, y, z) = x^4 + 2y^2 + 3z^2 - yz - 23y + 4x - 5$

Exercise 1.7 is very simple, just go back to the definitions.

(5 mins) Exercise 1.7

Let E be a vector space with a norm $\|\bullet\|$, show that the norm is a convex function.

For Exercise 1.8, you need to come back to the definitions.

(5 mins) Exercise 1.8

Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a convex non-decreasing function and $h : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function. Show that $f = g \circ h$ is convex. Give an example where g and h are convex but $g \circ h$ is non-convex.

Exercise 1.9 is a little bit involved in the computational side. Hopefully we give you the result for the gradient and the Hessian, so that you can skip the first question if you can't solve it.

(15 mins) Exercise 1.9

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be defined as $f(x) = \log(e^{x_1} + \dots e^{x_n})$. For any $x \in \mathbb{R}^n$, denote $s \in \mathbb{R}^n$ such that $s_i = \frac{e^{x_i}}{e^{x_1}+\dots+e^{x_n}}$.

1. Show that

$$\nabla f(x) = s \quad \text{and} \quad \begin{cases} (H[f](x))_{ii} = s_i^2 - s_i s_i \\ (H[f](x))_{ij} = -s_i s_j \qquad \text{if } i \neq j \end{cases}$$

2. Show that $(H[f](x)h, h) = \sum_i s_i h_i^2 - (\sum_i s_i h_i)^2$.
3. From $\sum s_i = 1$ and $s_i > 0$, conclude that f is (strictly) convex.

For Exercise 1.10, remember the following result : If

- for every $i \in \mathcal{I}$, the function g_i is convex,
- for every $j \in \mathcal{E}$, the function g_i is affine.

then the set

$$\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in \mathcal{I} \text{ and } g_i(x) = 0, i \in \mathcal{E}\}$$

is convex. The goal of this exercise is to remember this result and to focus on the fact that there is no equivalence. Most of the ideas must originate from drawings.

(TD1 : 10 mins) Exercise 1.10

Draw the following sets, are they convex ?

1. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y \geq x^2\}$
2. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1,\ y \geq x^2\}$
3. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y \leq x^2,\ y \leq 0\}$
4. $X = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y = 2x\}$

Exercise 1.11 is solved by going back to the basis.

(5 mins) Exercise 1.11

Let g be a concave positive function. Show that $f = \frac{1}{g}$ is convex

2 Existence

The difficulty in showing existence of solution is the coercivity. Exercise 2.1 is in finite dimension and we have to prove coercivity by different ways.

(TD2 : 30 mins) Exercise 2.1

Show that the following problem $\inf_{M \in X} f(M)$ admits minima

1. $f : (x, y) \mapsto x^2 + y^2$, $X = \mathbb{R}^2$
2. $f : (x, y) \mapsto 3x$, $X = \{(x, y)$ such that $x^2 + y^2 \leq 1\}$
3. $f : (x, y) \mapsto x^4 + y^6 + x^2y^2$, $X = \mathbb{R}^2$
4. $f : (x, y) \mapsto 4x^2 + 6y^2 - 5xy$, $X = \mathbb{R}^2$

For Exercise 2.2, the coercivity is not trivial, you can use the following formula valid for each $x, y \in \mathbb{R}$ and each $\lambda > 0$

$$|xy| \leq \frac{1}{2}\left(\frac{x^2}{\lambda} + \lambda y^2\right)$$

(TD2 : 15 mins) Exercise 2.2

Let $f : (x, y) \mapsto \frac{1}{100}x^2 + 10^{10}y^2 - 100xy$, show that the problem $\inf_{M\in\mathbb{R}^2} f(M)$ admits minima.

In Exercise 2.3, the function is not coercive on $\mathbb{R}^2$ but is coercive on X. The notion of coercivity depends on the domain, even when the domain is unbounded.

(TD2 : 15 mins) Exercise 2.3

Let $X = \{M = (x, y) \text{ such that } x^2 \geq y^2\}$ and $f(x, y) = x^2 - y$

1. Show that f is not coercive on $\mathbb{R}^2$ and that $\inf_{\mathbb{R}^2} f = -\infty$.
2. Show that for every $(x, y) \in X$, then $f(x, y) \geq \frac{x^2+y^2}{2} - y$. Conclude that f is coercive on X and that there exists a global minimizer to $\inf_X f$.

Exercise 2.4 is a little bit on the theoretical side, but it shows that some standard problems of statistics have a solution.

(20 mins) Exercise 2.4

Let $A \in \mathcal{M}_{pn}(\mathbb{R})$ and $C \in \mathcal{M}_{qn}(\mathbb{R})$ and $b \in \mathbb{R}^n$. We consider the problem

$$\inf_X f \text{ if } f(x) = \frac{1}{2}\|Ax\|^2 - \langle b, x\rangle \text{ and } X = \{x \text{ such that } Cx = 0\}.$$

We suppose that both $\ker A$ and $\ker C$ are not trivial (not equal to $\{0\}$).

1. Suppose that there exists $u \in \ker A \cap \ker C$ such that $\langle u, b\rangle \neq 0$. Show that the problem does not admit minimizers.
2. Suppose that $\ker A \cap \ker C = \{0\}$
   (a) Show that the problem is equivalent to minimizing $g(x) = f(x) + \frac{1}{2}\|Cx\|^2$ on X.
   (b) Show that g can be written $g(x) = \frac{1}{2}\langle Bx, x\rangle - \langle b, x\rangle$ with $B = A^TA + C^TC$. Prove that B is a symetric definite positive matrix.
   (c) Show that there exists a minimizer to g over X, hence a minimizer of f over X.
3. If $I \in \mathbb{N}$ (number of classes) and for each $1 \leq i \leq I$, we are given $n_i \in \mathbb{N}$ (number of elements in each class). If we are given $Y_{ij} \in \mathbb{R}$ (random value of the (i, j) individual) for each $1 \leq i \leq I$ and $1 \leq j \leq n_i$. For any $\alpha \in \mathbb{R}^I$ and $\mu \in \mathbb{R}$, let
$$f(\alpha, \mu) = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \alpha_i - \mu)^2.$$
Apply the above result to show that the problem $\inf_X f$ admits a solution, where :
   (a) $X = \left\{(\alpha, \mu) \in \mathbb{R}^I \times \mathbb{R} \text{ such that } \sum_{i=1}^{I} \alpha_i = 0\right\}$
   (b) $X = \left\{(\alpha, \mu) \in \mathbb{R}^I \times \mathbb{R} \text{ such that } \sum_{i=1}^{I} n_i\alpha_i = 0\right\}$

In Exercise 2.5, we showcase the problems that one may encounter in infinite dimension.

(TD3 : 15 mins) Exercise 2.5

For the following problems, show that the function f is coercive and continuous and that X is closed but that the problems do not admit a minimizer.

- $X = \ell^2(\mathbb{R})$ and $f(x) = (\|x\|^2 - 1)^2 + \sum_{i=0}^{+\infty} \frac{x_i^2}{i+1}$
- $X = \{x \in \ell^2(\mathbb{R}) \text{ such that } \|x\| = 1\}$ and $f(x) = \sum_{i=0}^{+\infty} \frac{x_i^2}{i+1}$

3 First order conditions

The following exercise is about minimizing a quadratic function with linear constraints, it is very similar to the exercise done in class.

(TD3 : 25 mins) Exercise 3.1

Show that the following problem admits a solution and find it using KKT theorem.

$$\min_{2x+y \geq 5} x^2 + y^2$$

Exercise 3.2 is a very important exercise to bear in mind. It is an example of a convex problem (convex function and convex inequality constraints) for which the constraints are not qualified. In this example, there exists a global minimum but no KKT points. This exercise is a very good example of poor design in modelisation. Indeed, could we think of a worst way to describe the set X ? Do we need optimization theory to solve the problem ? Everything goes wrong because of poor design, and it shows how important it is to describe correctly the problem you want to solve. This exercise will be seen again in Exercise 5.2.

(TD3 : 20 mins) Exercise 3.2

Let $X = \{(x, y) \text{ such that } (x-1)^2 + y^2 \leq 1 \text{ and } (x+1)^2 + y^2 \leq 1\}$.

Show that $X = \{0, 0\}$ the constraints are not qualified and show that the conclusions of KKT theorem do not apply in this case (check that $\inf_X y$ is attained at 0 and that this point fails to be a KKT point).

In the following exercise, we do not apply KKT because the sets are open. It is a nice little reminder of minimization on open sets.

(20 mins) Exercise 3.3

Let: $f : (x, y) \in \mathbb{R}^2 \mapsto x^3 + 2xy + y^2$.

1. Is the function f convex ? concave ?
2. Give every local extrema of f and determine if they are local or global.
3. Define the following sets:

$$X_1 = \{(x, y) \in \mathbb{R}^2 \mid y - x < 0\} \quad \text{et} \quad \widetilde{X}_1 = \{(x, y) \in \mathbb{R}^2 \mid y - x > 0\}.$$

Give every local extrema of f on X_1, and then on $\widetilde{X}_1$.

Exercise 3.4 is another application of KKT, it is in $3d$, so it is a little bit more involved

on a computational point of view. Moreover, we study minimizers and maximizers, hence this exercise allows recapping the KKT formulas for maximisation. It also emphasizes the fact that, when looking for solutions to minimisation problems by the KKT method, solutions to maximisation problems is almost given for free.

(TD4 : 30 mins) Exercise 3.4

Let $X = \{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 1,\ x_1^2 + x_2^2 + x_3^2 \leq 1\}$.

Give the extrema of the function $f : x \in \mathbb{R}^3 \mapsto x_1$ on X.

In the following exercise, the computations are a little bit on the abstract side, the final formula looks awful but is not *that* difficult to obtain.

(TD4 : 30 mins) Exercise 3.5

Let A be a real $m \times n$ matrix and $b \in \mathbb{R}^n$. Let:

$$\inf_{x \in \mathbb{R}^n} \frac{1}{2}\|x\|^2 - \langle b, x\rangle \qquad \text{such that: } Ax = 0.$$

1. Show that this problem is equivalent to a projection problem on a certain convex set. Make explicit this convex set. Show existence and uniqueness of a global minimizer.
2. Show that the constraints are qualified everywhere and show that if A is of full rank (i.e., $A^\star$ is injective), then the optimal solution is given by:

$$x^\star = (I_n - A^\star(AA^\star)^{-1}A)b.$$

This exercise solves the problem of finding the symmetric matrix closest to a given matrix A. It is an orthogonal decomposition, when one proves that symmetric matrices are orthogonal to anti-symmetric matrices, one can see that this exercise fits nicely with the decomposition formula

$$X = \underbrace{\frac{X + X^\top}{2}}_{symmetric} + \underbrace{\frac{X - X^\top}{2}}_{anti-symmetric}.$$

(20 mins) Exercise 3.6

Let A be a square matrix of size n. Solve the following problem:

$$\inf_{X \in \mathcal{M}_{nn}(\mathbb{R})} f(X) = \frac{1}{2}\|X - A\|_F^2 \qquad \text{with: } X^\top = X,$$

in the case where the matrix A is (a) symmetric and then (b) non-symmetric.

This problem is all about solving KKT equations, a striking feature of to show that the computations for a certain problem can be helpful with another. If the student masters the following exercise, he can be assured he understands first order conditions.

(20 mins) Exercise 3.7

Let $f(x,y) = y - 2x$ and:

$$X = \{(x,y) \in \mathbb{R}^2 \mid y \geq x^2 \text{ et } x^2 + y^2 \leq 1\}.$$

1. Give the extremal points of f on X. Determine if they are local or global.
2. What happens if we replace $y \geq x^2$ by $y \leq x^2$?

4 Second order conditions

The exercise below is tedious but rather easy. It is a non-convex problem and the idea is to study thoroughly second order necessary and sufficient conditions. Note that in the following problem, we could avoid second order analysis by comparing the value of the function at the different KKT points, but it wouldn't be fun.

(TD5 : 60 mins) Exercise 4.1

Denote $X = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 2\}$ and consider the following problems

$$(P_1) \qquad \min_{x \in X} f_1(x) \quad \text{with } f_1(x) = x_1 + x_2.$$
$$(P_2) \qquad \min_{x \in X} f_2(x) \quad \text{with } f_2(x) = x_1 x_2.$$

1. Show that the problems (P_1) et (P_2) admit solutions.
2. Show that the constraint $x_1^2 + x_2^2 = 2$ is qualified in any point of X.
3. Determine every KKT point of f_1, by second order conditions determine if they are local minima or maxima of f_1 sur X. Are they global ?
4. Same question for f_2.

Exercise 4.2 is kind of similar to Exercise 4.1, with a little twist here, (spoiler: the minimizers are not global).

(TD6 : 60 mins) Exercise 4.2

Compute the KKT points of

$$(P) \qquad \inf_{(x,y) \in \mathbb{R}^2} f(x,y) = -\frac{1}{2}(x-4)^2 + 2y^2 \quad \text{with} \quad 1 - x^2 - y^2 \leq 0.$$

Compute the minimizers of (P) and determine if they are local or global.

The following exercise is a fun one, it is easy to be lost in one's own calculations... good luck !

(60 mins) Exercise 4.3

For any $k \in \mathbb{R}^\star$, we consider the following problem:

$$(P_k) \qquad \min_{(x,y) \in \mathbb{R}^2} f(x,y) = (x-1)^2 + y^2 \qquad \text{with} \qquad 2kx - y^2 \leq 0.$$

We denote $X_k = \{(x,y) \in \mathbb{R}^2 \mid 2kx - y^2 \leq 0\}$.

1. Show that the problem is not convex.

2. Show that the problem admits at least one global minimizer.
3. Solve the unconstrained problem. Show that the unique global minimizer of P_k is $(1,0)$ for $k \leq 0$
4. Show that the constraints are qualified at each point.
5. Find the KKT points of (P_k). You must obtain

	(x,y)	λ
$k<0$	$(1,0)$	0
$0<k<1$	$(0,0)$	$\frac{1}{k}$
	$(1-k, \sqrt{2k(1-k)})$	1
	$(1-k, -\sqrt{2k(1-k)})$	1
$k \geq 1$	$(0,0)$	$\frac{1}{k}$

6. Without computing the Hessian of the Lagrangian, prove that except for the point $(0,0)$ in the case $0<k<1$, every KKT point is a global minimum.
7. By computing second order conditions, prove that the point $(0,0)$ in the case $0<k<1$ is not a local minimum.

5 Duality

The following exercise can be solved directly or by looking at its dual. This exercise is a good introduction to dual problems.

(TD7 : 20 mins) Exercise 5.1

Consider the following problem

$$(P_1) \quad \left| \begin{array}{ll} \inf\limits_{(x)\in\mathbb{R}^2} & f(x) = 2x_1 + x_2 \\ \text{s.c.} & x_1^2 + x_2^2 \leq 1 \end{array} \right.$$

1. Show that the problem admits a solution and that the dual approach is possible.
2. Compute $(P_1^\star)$, the dual problem of (P_1).
3. Solve $(P_1^\star)$.
4. From the solutions of $(P_1^\star)$, deduce the solutions of (P_1)

The following exercise is a follow-up on Exercise 3.2, where the constraints are not qualified and the global minimum is not a KKT point. We study in this exercise the behavior of the dual function.

(TD7 : 20 mins) Exercise 5.2

We consider the problem $\min_{x\in X} f(x)$, where $X \subset \mathbb{R}^2$ and

$f(x_1,x_2) = x_2 \quad X = \{x = (x_1,x_2) \text{ such that } (x_1-1)^2+x_2^2 \leq 1 \text{ and } (x_1+1)^2+x_2^2 \leq 1\}$

1. Draw X, show that the problem is convex.

2. Show that $X = \{0\}$, give the global minimum of f over X.
3. Show that there is no KKT point. Show that the constraints are not qualified anywhere.
4. Compute the Lagrangian and the dual function, show that the latter is equal to

$$f^\star(\lambda) = -\frac{1}{2(\lambda_1+\lambda_2)} - \frac{(\lambda_2-\lambda_1)^2}{\lambda_1+\lambda_2}$$

5. Solve the dual problem, show that there is no duality gap but that there is no saddle point.

The following exercise is an abstract exercise that is solved twice, first using a primal approach and then using a dual approach. The objective is to understand the common computations between the two approaches.

(TD7 : 20 mins) Exercise 5.3

Consider the following problem with $a \neq 0$

$$\inf_{x\in\mathbb{R}^n} \langle a, x\rangle \quad \text{with } \|x\|_2^2 \leq 1.$$

1. Show that the problem is convex and admits a global minimizer. Show that the constraints are qualified.
2. Find every KKT point.
3. Compute the dual function and solve the dual problem.
4. Discuss the difference between the KKT search and the solution of the dual problem.

This exercise shows that, when there is no convexity, KKT points are not necessary saddle-points.

(TD8 : 30 mins) Exercise 5.4

Let

$$(P) \quad \left| \begin{array}{ll} \min\limits_{(x,y)\in\mathbb{R}^2} & -x^2+y^2 \\ \text{s.c.} & x+y \geq 0 \\ & y \geq x \end{array} \right.$$

1. Draw X, the set of constraints of (P). Show **very simply** that problem (P) admits an infinity of solutions. You will show that this set of solutions is $\{(x,|x|), x\in\mathbb{R}\}$.
2. Put the problem in standard form, compute the dual function and show that the associated Lagrangian does not admit saddle points.

The next exercise is a $3d$ convex optimization problem.

(20 mins) Exercise 5.5

Let

$$(P) \quad \left| \begin{array}{ll} \min\limits_{(x,y,z)\in\mathbb{R}^3} & x-y \\ \text{with:} & x+y+z\leq 1, \quad x^2+y^2+z^2\leq 1 \end{array} \right.$$

1. Show that problem (P) has a solution.
2. Solve problem (P) by a dual approach.

The next exercise is all about linear programming in standard duality.

(TD8 : 30 mins) Exercise 5.6: Linear programming

Consider the linear optimization problem

$$(\mathcal{LP}) \quad \left| \begin{array}{ll} \min\limits_{M=(x,y)\in\mathbb{R}^2} & 2x+y \\ \text{s.t.} & x+y=1, x\geq 0, y\geq 0 \end{array} \right.$$

1. Show that the problem admits a solution and that the constraints are qualified.
2. Compute $\mathcal{L}$, the Lagrangian of the problem and show that $\nabla_M \mathcal{L}$ does not depend on M.
3. Solve the KKT equations.
4. Compute the dual function and solve the dual problem.
5. How do we recover the solution of the primal problem from the solution of the dual problem ?

The following exercise is left as a training exercise

(TD9 : 30 mins) Exercise 5.7

Consider the optimization problem

$$(P) \quad \left| \begin{array}{ll} \min\limits_{(x,y)\in\mathbb{R}^2} & x^2+y^2-x \\ \text{s.t.} & x+y\geq 1 \end{array} \right.$$

1. Put the problem in standard form and give $(P^\star)$, its dual.
2. Solve $(P^\star)$.
3. Give the solutions of (P).

This exercise is a little bit on the theoritical side, but it covers any quadratic function with affine equality constraints.

(TD9 : 30 mins) Exercise 5.8: Quadratic optimization

Let A be a square real symetric and invertible matrix of size n. Let $b\in\mathbb{R}^n$ and C be a real matrix of size $p\times n$ and of full rank ($p<n$). Let $d\in\mathbb{R}^p$. We consider the following problem:

$$(P) \quad \left| \begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & f(x)=\frac{1}{2}\langle Ax,x\rangle+\langle b,x\rangle \\ \text{s.c.} & Cx-d=0 \end{array} \right.$$

1. Under which conditions is $f^\star$, the dual function associated to the problem (P) is well defined (not always equal to $-\infty$)? Compute it.
2. Solve the dual problem $(P^\star)$. Do we have existence and uniqueness of a solution of (P) ? Give the analytical expression of this solution and of the associated multiplicator
3. Replace the equality constraint by an inequality $Cx - d \leq 0$. Write the dual problem associated to the new constraint.

6 First order descent methods

The next exercise proves that the Wolfe step bisection algorithm converges in a finite number of iterations. It also proves the existence of a Wolfe step.

(TD10 : 35 mins) Exercise 6.1

The objective is to prove the convergence of the Wolfe step bisection algorithm. We recall that we are given a function ϕ which is C^1, bounded from below, with $\phi(0) = 0$ and $\phi'(0) < 0$. Morevoer, we are given $0 < \varepsilon_1 < \varepsilon_2 < 1$. If we denote

$W_1 = \{s \geq 0 \text{ such that } \phi(s) \leq \varepsilon_1 s\phi'(0)\}$ and $W_2 = \{s \geq 0 \text{ such that } \phi'(s) \geq \varepsilon_2\phi'(0)\}$

A Wolfe step is any element in $W_1 \cap W_2$. The bisection algorithm is the following

```
def Wolfe(s) # s is the initial step, it is >0.
  sm,sp=0,np.inf
  while True :
    if not s in W_1 :
      sp=s
      s=0.5*(sp+sm)
    elif not s in W_2 :
      sm=s
      s=2s if sp==np.inf else 0.5*(sp+sm)
    else :
      return s
```

We denote s^- the value of `sm` and s^+ the value of `sp` The objective is to prove that this algorithm converges in a finite number of iterations in the `while` loop, suppose it spends an infinite number of iterations. We denote s_k^+, s_k^-, s_k the values of s^+, s^- and s at iteration k

1. Show that for k, $s_k^+ \in W_1^c$ and $s_k^- \in W_2^c \cap W_1$
2. Show that for any k, $s_k^- \leq s_k \leq s_k^+$ and $(s_k^+)_k$ is decreasing and $(s_k^-)_k$ is increasing.
3. Show that it is not possible to have $s_k^+ = +\infty$ for each k.
4. Show that there exists $s^\star$ such that $s_k^- \leq s^\star \leq s_k^+$ and both $(s_k^+)_k$ and $(s_k^-)_k$ converge to $s^\star$.
5. By using $s^+ \in W_1^c$ and $s^- \in W^1$, show that $\phi(s^\star) = \varepsilon_1 s^\star \phi'(0)$.
6. By computing $\frac{\phi(s^\star)-\phi(s)}{s^\star - s}$ for $s = s_k^+$ or for $s = s_k^-$ (or both), show that
$$\phi'(s^\star) \geq \varepsilon_1 \phi'(0)$$
7. Show that the algorithm cannot spend an infinite number of iterations.

The exercise below is an introduction to the theory of convergence of the stochastic gradient.

(60 mins) Exercise 6.2: Stochastic gradient

The objective is to minimize the function $x \mapsto F(x) = \frac{1}{N}\sum_{i=1}^{N} F_i(x)$ with N very large. We suppose that at each iteration n, we draw batch of indices $\mathcal{B}_n \subset [1, N]$ uniformly and we are given the function $f_n = \frac{1}{|\mathcal{B}_n|}\sum_{i\in\mathcal{B}_n} F_i$. We suppose that the draws are independent. We then have, forall x deterministic

$$\mathbb{E}(f_n(x)) = F(x) \quad \mathbb{E}(\nabla f_n(x)) = \nabla F(x).$$

- We say that a random variable is $\mathcal{F}_n$-measurable if it only depends on $(\omega_1, \dots, \omega_n)$, where ω_i is the randomness that defines $\omega_i \mapsto \mathcal{B}_i(\omega_i)$. In simple words, X is $\mathcal{F}_n$-measurable iff it is determined by the first n draws. In probabilistic words, $(\mathcal{F}_p)_p$ is the filtration generated by the batches $(\mathcal{B}_n)_n$.
- For any n, m, if X is $\mathcal{F}_m$-mesurable, we define $\mathbb{E}_n(X)$,a $\mathcal{F}_n$-measurable random variable as

$$\mathbb{E}_n(X) = \begin{cases} X & \text{if } m \leq n \\ \int X(\omega_1, \dots \omega_m) d\mathbb{P}(w_{n+1}) \dots d\mathbb{P}(w_m) & \text{if } m > n \end{cases}.$$

  In other words $\mathbb{E}_n(X) = \mathbb{E}(X|\mathcal{F}_n)$ is the conditional expectation with respect to the σ-algebra $\mathcal{F}_n$.
- We have $\mathbb{E}_0 = \mathbb{E}$ and for every X and Y random variables, if Y is $\mathcal{F}_n$-measurable, then :

$$\mathbb{E}_n(XY) = \mathbb{E}_n(X)Y \text{ and } \mathbb{E}_n(\mathbb{E}_m(X)) = \mathbb{E}_{\min(n,m)}(X)$$

We suppose F has an L-Lipschitz gradient and is bounded from below, we suppose that $(x_n)_n$ is a sequence of random variables given by the following gradient method:

$$x_{n+1} = x_n - \alpha_{n+1}\nabla f_{n+1}(x_n), \text{ with } \quad \alpha_n < c.$$

Here α_n is the step and is not a random variable. Suppose that there exists constants A and B such that, for every θ

$$\mathbb{V}(\nabla f_n(x)) \leq A + B\|\nabla F(x)\|^2,$$

Where

$$\mathbb{V}(\nabla f_n(x)) = \mathbb{E}(\|\nabla f_n(x)\|^2) - \|\nabla F(x)\|^2$$

1. If x_0 is deterministic, show that x_n is $\mathcal{F}_n$-measurable for each n.
2. Obtain the following inequality

$$F(x_{n+1}) - F(x_n) \leq -\alpha_n\langle\nabla f_{n+1}(x_n), \nabla F(x_n)\rangle + \frac{L\alpha_n^2}{2}\|\nabla f_{n+1}(x_n)\|^2$$

3. Apply $\mathbb{E}_n$ on both sides of the inequality to obtain

$$\mathbb{E}_n(F(x_{n+1})) - F(x_n) \leq \left(-\alpha_n + \frac{L\alpha_n^2(B+1)}{2}\right)\|\nabla F(x_n)\|^2 + \frac{L\alpha_n^2}{2}A$$

4. Show that there exists a $c > 0$ such that if $0 < \alpha_n < c$ for all n, then

$$\mathbb{E}_n(F(x_{n+1})) - F(x_n) \leq -\frac{\alpha_n}{2}\|\nabla F(x_n)\|^2 + \frac{C_2\alpha_n^2}{2}$$

5. Apply the expectation $\mathbb{E}$ and prove that there exists $C, C_2 > 0$ such that

$$\sum_{n=1}^{K} \alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \leq C + C_2 \sum_{n=1}^{K} \alpha_n^2$$

6. Suppose that $\sum_{n=1}^{+\infty} \alpha_n^2 < +\infty$ and $\sum_{n=1}^{+\infty} \alpha_n = +\infty$. Show that

$$\lim_{K \to +\infty} \left(\frac{1}{\sum_{n=1}^{K} \alpha_n} \sum_{n=1}^{K} \alpha_n \mathbb{E}\left(\|\nabla F(x_n)\|^2\right) \right) = 0$$

7. At each iteration n, draw a number between 1 and n, call τ_n this drawing process, it must be independent of the drawing of the batches $\mathcal{B}$ and the law is given by

$$\mathbb{P}(\tau_n = k) = \frac{\alpha_k}{\sum_{k=1}^{n} \alpha_n} \quad \forall k = 1, \dots n$$

Show that

$$\lim_{n \to +\infty} \mathbb{E}\left(\|\nabla F(x_{\tau_n})\|^2\right) = 0$$

7 Second order descent methods

This exercise proves the formula for the BFGS algorithm. You will implement this formula.

(TD10 : 25 mins) Exercise 7.1

We want to solve $-B_k \nabla f(x_k)$ if B_k is defined by the BFGS recurence relationship

$$B_k = \left(I - \rho_k \sigma_k y_k^T\right) B_{k-1} \left(I - \rho_k y_k \sigma_k^T\right) + \rho_k \sigma_k \sigma_k^T. \text{ with } \rho_k = \frac{1}{\langle y_k, \sigma_k \rangle}$$

1. Let $q_k = -\nabla f(x_k)$ and define the sequences $(\alpha_i)_{1 \leq i \leq k}$ and $(q_i)_{0 \leq i \leq k}$ by the inverse recurrence for all $i = k, \dots 1$:

$$\begin{cases} \alpha_i & = \rho_i \langle \sigma_i, q_i \rangle \\ q_{i-1} & = q_i - \alpha_i y_i \end{cases}$$

Set $z_0 = B_0 q_0$ and compute the sequence $(z_i)_{i=1,k}$ by the following forward recurrence for all $i = 1, \dots, k$:

$$\begin{cases} \beta_i & = \rho_i \langle y_i, z_{i-1} \rangle \\ z_i & = z_{i-1} + (\alpha_i - \beta_i)\, \sigma_i \end{cases}$$

Show that for all i, we have $z_i = B_i q_i$

2. If $L = (\sigma_k, y_k, \rho_k)_k$ is saved in a sequence. Justify the following algorithm for computing $-B_k \nabla f(x_k)$:
   (a) $q = -\nabla f(x_k)$ and create an empty list called L_α
   (b) For (σ, y, ρ) in reversed order of L :
      i. Compute $\alpha = \rho \langle \sigma, q \rangle$ and append α to L_α
      ii. Set $q = q - \alpha y$

(c) Reverse the list of L_α.
(d) Set $q = B_0 q$.
(e) For $(\sigma, y, \rho), \alpha$ in (L, L_α) :
  i. Compute $\beta = \rho\langle y, q\rangle$.
  ii. Set $q = q + (\alpha - \beta)\sigma$

8 Convex non-smooth optimization

(TD 11 : 20 mins) Exercise 8.1

Let $\mathcal{C}$ be a non-empty convex subset of $\mathbb{R}^n$.
1. Recall the definition of $i_\mathcal{C}$, the characteristic function of $\mathcal{C}$
2. Under which condition is $i_\mathcal{C}$ lower-semi-continuous ?

(TD 11 : 40 mins) Exercise 8.2

Let $a \in \{+\infty, 2\}$. Are the following functions l.s.c on $\mathbb{R}$?

$$f_1(x) = \begin{cases} x^2 & \text{if } |x| \leq 2 \\ a & \text{if } |x| > 2 \end{cases} \quad \text{and } f_2(x) = \begin{cases} x^2 & \text{if } |x| < 2 \\ a & \text{if } |x| \geq 2 \end{cases}$$

(15 mins) Exercise 8.3

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the function defined by:

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{si} \quad x \neq 0 \\ -1 & \text{si} \quad x = 0. \end{cases}$$

1. Show that the function f is l.s.c on $\mathbb{R}$.
2. If $f(0) = 1$, is f still l.s.c ?

(TD 12 : 60 mins) Exercise 8.4

Consider the following functions:

- $f_1(x) = \frac{1}{2}(x-1)^2 + |3x - 2|$, $x \in \mathbb{R}$,
- $f_2(x) = \max(|x|, x^2 - 2)$, $x \in \mathbb{R}$,
- $f_3(x, y, z) = x^2 + xy + y^2 + \frac{1}{2}z + |y - z|$.

1. Show that f_i, $i = 1, \ldots, 3$, is convex and l.s.c on $\mathbb{R}$.
2. For any $x \in \mathbb{R}$, compute the subdifferential $\partial f_1(x)$. Deduce the global minimizer(s) of f_1.
3. For any $x \in \mathbb{R}$, compute the subdifferential $\partial f_2(x)$. Deduce the global minimizer(s) of f_2.
4. For any $M = (x, y, z) \in \mathbb{R}^3$, compute the subdifferential $\partial f_3(M)$. Deduce the global minimizer(s) of f_3.

(15 mins) Exercise 8.5

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a C^1 function and $X = \{x \text{ s.t.} g(x) \leq 0\}$. We suppose that the constraints are qualified everywhere, that is :

$$\forall x, \quad g(x) = 0 \Rightarrow \nabla g(x) \neq 0.$$

We let i_X be the characteristic function of X, that is

$$i_X(x) = \begin{cases} 0 & \text{if } g(x) \leq 0 \\ +\infty & \text{if } g(x) > 0 \end{cases}$$

1. We suppose that i_X is convex, does g have to be convex ?
2. Show that i_X is l.s.c.
3. Suppose that $g(x) < 0$, show that $\partial i_X(x) = \{0\}$.
4. Suppose that $g(x) = 0$, show that for all $t \geq 0$, then $t\nabla g(x) \in \partial i_X(x)$.
5. Suppose that $g(x) = 0$, show that for all $t \geq 0$, then $\partial i_X(x) = \{t\nabla g(x), t \in \mathbb{R}^+\}$.

(Constrained minimization on the unit disk : 60 mins) Exercise 8.6

In this exercise, we suppose that the result of Exercise 8.5 is known. Define the following minimization problem

$$(P) \qquad \inf_{(x,y)\in\mathbb{R}^2} g(x,y) = \frac{1}{2}(x-2)^2 + |y| \quad \text{such that:} \quad x^2 + y^2 \leq 1.$$

And denote $X = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

1. Transform the problem (P) in a unconstrained minimization problem :

$$(P') \qquad \inf_{(x,y)\in\mathbb{R}^2} f(x,y) = g(x,y) + h(x,y)$$

   where $h : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ is a convex function on $\mathbb{R}^2$ that has to be given.
2. Compute the subdifferential $\partial f(M)$ in each point $M = (x,y) \in dom(f)$.
3. Show that g does not admit minimizers on the interior of X (denoted $\mathring{X}$).
4. Give the global minimizers of g on X.
5. Can we apply the same method to solve the following maximization problem ?

$$\sup_{(x,y)\in\mathbb{R}^2} g(x,y) = \frac{1}{2}(x-2)^2 + |y| \quad \text{such that:} \quad x^2 + y^2 \leq 1.$$

(KKT Theorem-I : 20 mins) Exercise 8.7

The objective of this exercise is to prove KKT theorem for convex non differentiable functions. We will prove it in two different ways. We aim at minimizing a function f over a set of constraints X which is given by a finite number of inequalities only

$$X = \{x \text{ s.t. } g_i(x) \leq 0, \forall i \in \mathcal{I}\}.$$

We suppose that the functions f and $(g_i)_i$ are convex and that they are finite everywhere (hence they are continuous everywhere). We also suppose that there exists a point $x_0 \in X$ such that $g_i(x_0) < 0$ for every i (this is a variant of Slater's

condition). The KKT theorem is :

If $x^\star$ is a global minimizer of f over X, there exists $\lambda \geq 0$ such that $\lambda_i g_i(x^\star) = 0$ for all i and

$$0 \in \partial f(x^\star) + \sum_i \lambda_i \partial g_i(x).$$

1. Denote $\phi(x) = \max_i(f(x) - f(x^\star), g_i(x))$. Prove that $\phi \geq 0$ everywhere and that $\phi(z) = 0$ iff z is a global minimizer of f on X.
2. Show that for any x with $\phi(x) = 0$, we have

$$\partial\phi(x) = \left\{ \alpha\partial f(x) + \sum_{i\in\mathcal{A}_x} \gamma_i \partial g_i(x), \text{ where } \alpha + \sum_{i\in\mathcal{A}_x} \gamma_i = 1 \right\}$$

3. Conclude that if $x^\star$ is a global minimizer of f over X, there exists $\alpha \geq 0$ and $\gamma \geq 0$ such that $\gamma_i g_i(x^\star) = 0$ for all i and $\alpha + \sum_i \gamma_i = 1$

$$0 \in \alpha\partial f(x^\star) + \sum_i \lambda_i \partial g_i(x).$$

4. Let $x^\star$ be a global minimizer and α, γ be as above, we prove that $\alpha \neq 0$, to this end we suppose that $\alpha = 0$. Consider now the function $\psi : x \mapsto \sum_{i\in\mathcal{A}_x} \gamma_i g_i(x)$. Compute $\partial\psi(x)$ and prove that $x^\star$ is a global minimizer of ψ. Prove this is absurd and hence $\alpha \neq 0$.
5. Set $\lambda = \frac{\gamma}{\alpha}$ and prove KKT theorem.

(Subdifferential of a norm : 0 mins) Exercise 8.8

1. Show that the dual of the norm $\|\cdot\|_1$ is $\|\cdot\|_\infty$.
2. Give the subdifferential of the norm $\|\cdot\|_1$.
3. **Additional question:** Answer the same questions by switching the roles of $\|\cdot\|_\infty$ and $\|\cdot\|_1$.

(TD 13 : The Legendre-Fenchel transform : 60 mins) Exercise 8.9

Let f be a function from E to $\mathbb{R}$, where E is an Hilbert space, its Legendre-Fenchel transform is a function from E to $\mathbb{R}$, given by

$$f^\star(s) = \sup_{x\in E} \left(\langle s, x\rangle - f(x) \right),$$

1. Let A be a symmetric positive definite matrix and let f be the function defined by

$$f : \mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto \frac{1}{2}\langle Ax, x\rangle,$$

   where $\langle\cdot,\cdot\rangle$ denotes the usual Euclidean inner product or $\mathbb{R}^n$.
   (a) Show that f is convex and lower semicontinuous on $\mathbb{R}^n$.
   (b) Compute the Legendre–Fenchel transform of f.
2. Let f be the function defined by

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto \frac{1}{2}x^2 + |x|.$$

(a) Show that f is convex and lower semicontinuous on $\mathbb{R}$.
(b) Show that the Legendre–Fenchel transform of f is given by

$$f^\star(s) = \begin{cases} 0, & \text{if } s \in [-1,1], \\ \frac{1}{2}(|s|-1)^2, & \text{if } |s| > 1. \end{cases}$$

(The Heavy Ball method, TD 14 : 60 mins) Exercise 8.10

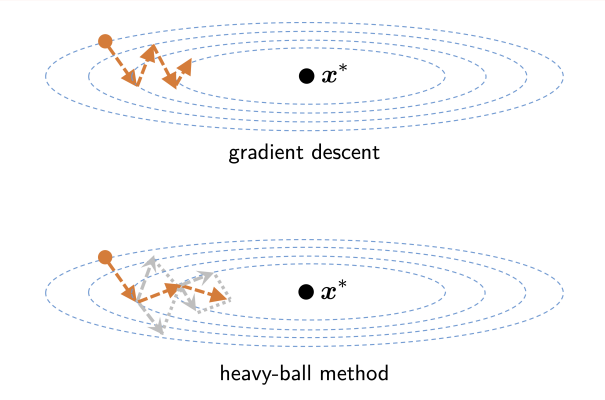
Part 1: From the ODE to the discrete scheme.
We consider the Heavy Ball (HB) differential equation:

$$\ddot{x}(t) + \alpha \dot{x}(t) + \nabla F(x(t)) = 0, \qquad \alpha > 0, \tag{1}$$

describing the motion of a body (a "heavy ball") in a potential F, subject to a friction proportional to its velocity.

1. By discretizing this ODE in a suitable way (finite-difference scheme, specifying explicitly the choice made for each term), show that one recovers the classical Heavy Ball scheme:

$$y_k = x_k + a(x_k - x_{k-1}), \qquad x_{k+1} = y_k - s\nabla F(x_k), \qquad a \in [0,1],\ s > 0. \tag{2}$$



2. Specify the relationship between the discretization step h of the ODE and the algorithmic step s, as well as between the friction α and the inertia a. Comment: what does s become as $h \to 0$? How does the inertia a vary when the friction α increases? decreases? Give a physical interpretation.

Part 2: 1D study: $f(x) = \frac{\mu}{2}x^2$.

1. Solve the Heavy Ball differential equation for this function, as a function of α and μ.
2. Show that the trajectory decays at a rate of the form $e^{-\tau(\alpha)t/2}$, where $\tau(\alpha)$ should be made explicit. Plot the graph of τ as a function of α.
3. Deduce the optimal friction for the Heavy Ball ODE, as well as the corresponding optimal value of the inertia for the algorithm found in question 1 of Part 1. Comment:
   - What does the optimal friction (and the associated optimal inertia) become for a **very flat** function (small μ)?
   - What does it become for a **very steep** function (large μ)?

Part 3: 2D study: $f(x) = \frac{\mu_1}{2}x_1^2 + \frac{\mu_2}{2}x_2^2$, with $\mu_1 < \mu_2$.

1. Solve the Heavy Ball ODE for this function. Deduce the decay rates of each component of the trajectory, as functions of α, μ_1, μ_2.
2. We now turn to the choice of the friction parameter α. We propose a worst-case analysis:
   (a) Suppose that $\alpha = 2\sqrt{\mu_2}$. Compute the decay rates of the two components x_1 and x_2.
   (b) Suppose that $\alpha = 2\sqrt{\mu_1}$. Compute the decay rates of the two components x_1 and x_2.
   (c) Conclude on the choice of α. Comment in particular on the role of the conditioning μ_2/μ_1, and compare with the usual tuning of the stepsize in gradient descent.

(Preparing labwork, TD 15 : 60 mins) Exercise 8.11

Let $\gamma > 0$, the objective is to compute the proximal operator to $x \mapsto \gamma\|x\|_1$.

1. Compute the proximal operator of the function $h : x \mapsto \gamma|x|$, $x \in \mathbb{R}$.
2. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be defined as:

$$f(x_1, \dots, x_n) = f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n),$$

   where the functions $f_i : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ are all convex and l.s.c.
   (a) Show that:

$$\partial f(x) = \partial f_1(x_1) \times \partial f_2(x_2) \times \cdots \times \partial f_n(x_n).$$

   (b) Deduce that:

$$prox_f(x) = (prox_{f_1}(x_1), prox_{f_2}(x_2), \dots, prox_{f_n}(x_n)).$$

3. Using the previous questions, compute the proximal operator of the function $f : x \mapsto \gamma\|x\|_1$.

(TD 16 : 60 mins) Exercise 8.12

Compute the proximal operators of the following functions

$$\begin{aligned}
(a)\ f : x &\mapsto x_1^2 - 3x_1x_2 + 3x_2^2,\ x = (x_1, x_2) \in \mathbb{R}^2,\\
(b)\ g : x &\mapsto \frac{1}{2}\|Ax - b\|^2,\ x \in \mathbb{R}^n,\\
(c)\ h : x &\mapsto \max(x^2 - 1, 0)\ x \in \mathbb{R}.
\end{aligned}$$

(Minimization of the difference of convex functions : 30 mins) Exercise 8.13

Let g and h be two convex l.s.c functions from $\mathbb{R}^n$ into $\mathbb{R}$. We consider the following problem:

$$(P) \qquad \text{Minimize } f(x) = g(x) - h(x),\ x \in \mathbb{R}^n.$$

1. Without additional hypothesis, is the problem (P) a convex problem ?

2. We say that x is a T-critical point of f iff:

$$\partial g(x) \cap \partial h(x) \neq \emptyset.$$

Show that if $x^\star$ is a global minimizer of f, then:
(i) $\partial h(x^\star) \subset \partial g(x^\star)$.
(ii) $x^\star$ is a T-critical point of f.

3. From now, on we suppose that both g and h are strongly convex on $\mathbb{R}^n$. We are interested in the following algorithm: start with any $x_0 \in \mathbb{R}^n$. If x_k is given,
   - Pick up $s_k \in \partial h(x_k)$.
   - x_{k+1} is the solution to the following problem:

$$(P_k): \min_{x\in\mathbb{R}^n} g(x) - \langle s_k, x\rangle.$$

   (a) Show that the problem (P_k) is strongly convex.
   (b) Write down the optimality condition for (P_k) which is satisfied for x_{k+1}.
   (c) Show that for any $k \in \mathbb{N}$,

$$g(x_k) + h(x_{k+1}) \geq g(x_{k+1}) + h(x_k) + \frac{c_g + c_h}{2}\|x_{k+1} - x_k\|^2,$$

   where c_g and c_h are the constants of strong convexity of g and h respectively.
   (d) Deduce that the proposed algorithm is indeed a descent algorithm and that, provided f is bounded from below, we have

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\|^2 < +\infty.$$

To conclude, if the sequences $(x_k)_{k\in\mathbb{N}}$ and $(s_k)_{k\in\mathbb{N}}$ are bounded, one can show that any accumulation point $\bar{x}$ of the sequence $(x_k)_{k\in\mathbb{N}}$ is a T-critical point of (P) (although the proof is somewhat technical).

9 Previous Exams

(CC1-2024-2025 : 45 mins) Exercise 9.1

Soit $a > 0$, on veut résoudre

$$(P) \qquad \inf_{x\in X} ax_1 - x_1x_2$$

avec: $X = \left\{x = (x_1, x_2) \in \mathbb{R}^2 \text{ et } x_1 \geq 0 \text{ et } x_2 \geq 0 \text{ et } x_1 + x_2^2 = 3\right\}$. *Pour information, il s'agit d'un problème en économie connu sous le nom "d'optimisation de la courbe d'offre". Il y a une demande de h unités d'un produit, x_1 est la quantité de produit que vous allez produire et x_2 est le prix auquel il est vendu. Le coût de production de x_1 unités est ax_1. De plus on sait que pour un prix x_2, les concurrents sont capable d'offrir αx_2^2 unités de produit. Dans cet exercice, on a choisi: $\alpha = 1$ et $h = 3$.*

1. Ecrire le problème sous forme standard et montrer qu'il n'est pas convexe.
2. **Existence d'un point de minimum global.**

   (a) Montrer que X est fermé borné.
   (b) En déduire l'existence d'un point de minimum global du problème (P).
3. Montrer que les trois contraintes ne peuvent pas être nulles/actives en même temps, et montrer que les contraintes sont qualifiées en tout point.
4. Ecrire le Lagrangien et les conditions vérifiées par les points de KKT.
5. On cherche tout d'abord les points de KKT de la forme $(0, x_2)$ (i.e. $x_1 = 0$).
   (a) Montrer que $(0, \bar{x}_2)$ soit un point de KKT de (P) si et seulement si $a^2 \geq 3$, et calculer $\bar{x}_2$.
   (b) Dans le cas où $a^2 > 3$, étudier les conditions d'ordre 2 en ce point de KKT montrer qu'il est un minimum local.
   Pour information, l'interprétation économique est la suivante : votre prix de production est a et si $\alpha a^2 \geq h$, alors les concurrents sont capable de saturer le marché pour ce prix là, il n'y a pas de place pour vous et vous produisez une quantité $x_1 = 0$ de marchandise.
6. On recherche les points de KKT où $x_1 \neq 0$. Montrer qu'il en existe un et qu'il vérifie $x_1 = 3 - x_2^2$ et $x_2 = \frac{a}{3} + \sqrt{\frac{a^2}{9} + 1}$. *Ici on n'étudie pas les conditions de second ordre pour conclure sur l'optimalité du point trouvé.*
7. On revient maintenant au cas $a^2 < 3$. Sans calcul supplémentaire, donner le point de minimum global de (P) dans ce cas.

(CC1-2024-2025 : 45 mins) Exercise 9.2

Soit $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice symétrique définie positive, $a \in \mathbb{R}^n$ avec $a \neq 0$ et $c \in \mathbb{R}$, on cherche à résoudre le problème suivant

$$(P_2) \qquad \inf_{x \in X} \langle a, x \rangle \text{ avec } X = \left\{ x \in \mathbb{R}^n \text{ tel que } \frac{1}{2}\langle Qx, x \rangle \leq c \text{ et } \langle a, x \rangle \leq 0 \right\}$$

1. *(0.5 pt)* Que se passe-t-il si on prend $c < 0$? et $c = 0$?
   Dans la suite, on suppose: $c > 0$.
2. **Quelques propriétés de (P_2).**
   (a) *(0.5 pt)* Démontrer l'existence d'un point de minimum global.
   (b) *(0.5 pt)* Montrer que le problème (P_2) est convexe.
   (c) *(0.5 pt)* Montrez que les contraintes sont qualifiées en tout point.
3. **Résolution par KKT.**
   (a) *(1 pt)* Ecrire le Lagrangien du problème et les équations de KKT.
   (b) *(3 pt)* Résoudre les équations de KKT, montrer qu'il n'y a qu'un seul point de minimum global au problème donné. *Indication, vous devrez montrer que le minimum global est (x, λ) avec*

$$x = -\frac{1}{\lambda_1} Q^{-1} a, \quad \lambda = (\lambda_1, 0) \quad \lambda_1 = \sqrt{\frac{\langle Q^{-1}a, a \rangle}{2c}}$$

   .
4. **Résolution par l'approche duale.**
   (a) *(2 pt)* Ecrire la fonction duale.
   (b) *(2 pt)* Résoudre le problème dual.
   (c) *(1 pt)* Pourquoi est-ce que la résolution du problème dual permet de résoudre le problème primal sans résoudre les points de KKT ?

(CC2-2024-2025 : 90 mins) Exercise 9.3

Exercice I:

1. Soit $Q \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice symétrique définie positive, $a \in \mathbb{R}^n$ et $c \in \mathbb{R}$.
   (a) On note f la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ donnée par
   $$f(x) = \frac{1}{2}\langle Qx, x\rangle + \langle a, x\rangle.$$
   Montrer que f est coercive sur $\mathbb{R}^n$
   (b) Montrer que $X = \{x \text{ tel que } \frac{1}{2}\langle Qx, x\rangle \leq c\}$ est borné.
2. Montrer que les contraintes sont qualifiées en tout point pour les ensembles X suivants
   (a) $X = \{(x, y) \text{ tel que } x^2 - y^4 - 1 \leq 0 \text{ et } x \geq 0\}$
   (b) $X = \{x \in \mathbb{R}^n \text{ tel que, pour tout } i, \sum_{j=1}^n x_j^2 \leq x_i + 1\}$

Exercice II: Soit le problème :

$$\max_{(x,y,z)\in\mathbb{R}^3} f(x, y, z) := xyz \quad \text{sous} \quad x + y + z = 3.$$

1. Trouver les points de KKT.
2. En étudiant les conditions d'ordre 2, montrer que le point $(1, 1, 1)^T$ est maximiseur local. *Indication : En notant* $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$,

$$\forall d \in \mathbb{R}^3, \ \textit{si} \ \left\langle d, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle = 0 \ \textit{alors} \ Ad = -d.$$

Exercice III: Soit

$$(P) \quad \left| \begin{array}{ll} \min\limits_{(x,y)\in\mathbb{R}^2} & \frac{1}{2}x^2 - y \\ \text{sous:} & x + e^y \leq 1, \end{array} \right.$$

1. Déterminer la fonction duale du problème (P).
2. Résoudre le problème dual.
3. En déduire une solution de (P).
4. Que se passe-t-il si on remplace la contrainte $x + e^y \leq 1$ par $x + e^y = 1$?

(CC3-2023-2024 : 90 mins) Exercise 9.4

1. Soit
   $$f(x) = x_1^2 + x_1x_2 + \frac{1}{2}x_2^2 + |3x_1 - x_2 + 1|, \ x = (x_1, x_2) \in \mathbb{R}^2.$$
   (a) Montrer que la fonction f est convexe semicontinue inférieurement.
   (b) Donner le sous-différentiel de f en tout point de $\mathbb{R}$.
   (c) En déduire le(s) point(s) de minimum global de f.
2. Soit $g(x) = \max(-2x, x^2), \ x \in \mathbb{R}$.
   (a) Justifier que l'opérateur proximal de g est bien défini de façon unique.

(b) Donner le sous-différentiel de g en tout point de $\mathbb{R}$.
(c) Calculer $prox_g(x)$, $x \in \mathbb{R}$.

3. Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe semi continue inférieurement. Montrer que:
$$p = prox_f(x) \Leftrightarrow x - p \in \partial f(p).$$

4. On se place dans $\mathbb{R}^2$ et on note: $y = (y_1, y_2) \in \mathbb{R}^2$. Soit X un ensemble convexe fermé.
$$f(y) = \iota_X(y) = \begin{cases} 0 & \text{si } y \in X \\ +\infty & \text{sinon.} \end{cases}$$
(a) Donner l'expression de l'opérateur proximal $prox_{\iota_X}$ de ι_X ? (vous écrirez explicitement le problème d'optimisation associé).
(b) Soit $X = \{y \in \mathbb{R}^2 \mid y_1 = y_2\}$. Montrer que:
$$prox_{\iota_X}(x_1, x_2) = \left(\frac{x_1 + x_2}{2}, \frac{x_1 + x_2}{2}\right).$$

5. Soit f une fonction convexe différentiable à gradient Lipschitz de constante $L > 0$. On suppose de plus f coercive. On s'intéresse à l'algorithme du gradient implicite:
$$x_{k+1} = x_k - h\nabla f(x_{k+1}),\ k \in \mathbb{N} \tag{5}$$
où $h > 0$ désigne le pas de l'algorithme. Soit $k \in \mathbb{N}$ et x_k l'itéré courant. On définit la fonction suivante:
$$g_k(y) = f(y) + \frac{1}{2h}\|y - x_k\|^2.$$
(a) Démontrer que la fonction g_k est strictement convexe.
(b) Montrer que x_{k+1} défini par l'algorithme (5) est l'unique minimiseur de g_k.
(c) En comparant $g_k(x_k)$ et $g_k(x_{k+1})$, montrer que (5) est un algorithme de descente pour f sans condition sur le pas h.
(d) **Comparaison avec la descente de gradient explicite.**
i. Donner une condition suffisante pour que l'algorithme de descente de gradient ($x_{k+1} = x_k - h\nabla f(x_k)$) soit un algorithme de descente pour f.
ii. En vous appuyant sur les questions 3 et 4(a), comparez les descentes de gradient explicite et implicite. Que vaut-il mieux utiliser en théorie ? En pratique ?

INSA TOULOUSE
135 avenue de Rangueil
31400 Toulouse

Tél : + 33 (0)5 61 55 95 13
www.insa-toulouse.fr

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE