OPTIMISATION

IDRISS MAZARI-FOUQUER

Contents

Date: Last updated: September 3, 2026.

General introduction, pre-requisites

The goal of this course is to provide sound theoretical foundations for continuous optimisation. Optimisation problems arise in a variety of contexts, whether it be in medicine, in mechanics, in finance or in artificial intelligence.

We will consider optimisation problems that write

$$\min_{x \in X} f(x)$$

where X is a subset of a (possibly infinite dimensional) vector space and $f : X \to \mathbb{R}$ is the so-called "objective function". The class will proceed along the following lines:

(1) We will first review basic concepts in optimisation, the main goal being to provide standard tools to establish existence (and possibly uniqueness) of minimisers x^* of f. The key words are coercivity and convexity; a certain familiarity with differential calculus is required, as is a certain dexterity with linear algebra.
(2) The second part of the class deals with unconstrained finite-dimensional optimisation, that is, when $X = \mathbb{R}^d$ for some $d \geq 1$. The main emphasis of this part will be on approximation procedures, the key concept being that of gradient descent. A reasonable familiarity with differential equations will be expected in the last section of this part.
(3) The third part of the class is devoted to constrained optimisation problems, that is, when X is a strict subset of a vector space (still finite-dimensional at this stage). The key-concepts that the student should master at the end of this section are Lagrange multipliers and basic duality. We will also place a great emphasis on approximation methods, the two main algorithms being the projected gradient algorithm and the Uzawa algorithm.
(4) The final part of this class should be seen as an introduction on more advanced topics in optimisation, with a particular emphasis on optimal control and back-propagation.

We use the following notational conventions:

(1) $M_d(\mathbb{R})$ denotes the set of $d \times d$ matrices, $M_{p,q}(\mathbb{R})$ denotes the set of $p \times q$ matrices.
(2) $S_d(\mathbb{R})$ denotes the set of symmetric matrices in $M_d(\mathbb{R})$. The transpose of a matrix M is written M^T.
(3) $S_d^+(\mathbb{R})$, resp. $S_d^{++}(\mathbb{R})$, resp. $S_d^-(\mathbb{R})$, resp. $S_d^{--}(\mathbb{R})$ denotes the set of symmetric positive, resp. definite positive, resp. negative, resp. symmetric negative, matrices.

Parts of these lectures are inspired by classes I gave at other points, parts (including many of the exercises) are lifted from my predecessors Guillaume Carlier, Pierre Cardaliaguet, Olga Mula and Yannick Viossat. Let us also mention that several of the computers sessions were thought out and constructed by Maxime Chupin and Guillaume Legendre for the "Numerical Optimization" class they gave in 2025-2026. Finally, a number of online references were used, in particular the substantial work of Herbin [**?**] regarding basics of numerical methods, and which is a trove of interesting exercises. The class should be self-contained and, as usual, there is no need to use other references: the lectures and exercise sessions should provide enough material.

References

[1] Nicolas Boumal. *An introduction to optimization on smooth manifolds.* Cambridge: Cambridge University Press, 2023.

[2] Nicolas Boumal, Dmitriy Drusvyatskiy, and Quentin Rebjock. Gradient descent can converge to any isolated saddle point.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

[4] Guillaume Carlier. *Classical and modern optimization.* Adv. Textb. Math. Hackensack, NJ: World Scientific, 2022.

[5] Guillaume Garrigos and Robert M. Gower. Handbook of Convergence Theorems for (Stochastic) Gradient Methods. Preprint, arXiv:2301.11235 [math.OC] (2023), 2023.

[6] Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *J. Mach. Learn. Res.*, 17:43, 2016. Id/No 153.

Part 1. A review of basic concepts

1. Existence of optimisers, optimality conditions

Throughout this entire chapter, unless stated otherwise, we consider a fixed, $\mathscr{C}^2$ function $f : \mathbb{R}^d \to \mathbb{R}$. The goal of this section is to fix the terminology, as well as some notations.

Definition 1.1. *A point $x^* \in \mathbb{R}^d$ is called:*

(1) *A* global minimiser of f *if*

$$\forall x \in \mathbb{R}^d \,, f(x^*) \leq f(x).$$

(2) *A* local minimiser of f *if*

$$\exists \varepsilon > 0 \,, \forall x \in \mathbb{R}^d \,, \|x - x^*\| \leq \varepsilon \Rightarrow f(x^*) \leq f(x).$$

1.1. The optimisation problem under consideration.

The goal of this class is to study the optimisation problem

$$\inf_{x \in \mathbb{R}^d} f(x). \tag{1.1}$$

A first question is whether or not a solution x^* actually exists, in which case we are allowed to write

$$\min_{x \in \mathbb{R}^d} f(x).$$

To this end, let us recall the definition of *coercivity*:

Definition 1.2. *A continuous function f is* coercive *if, for any $M \in \mathbb{R}$, the sub-level set $\{x \in \mathbb{R}^d : f(x) \leq M\}$ is bounded.*

The main point of coercivity is the following proposition:

Proposition 1.1. *Assume f is coercive. Then f has a global minimiser x^*. With a slight abuse of terminology, we will say that x^* solves (1.1), and dub it a* minimiser *of f in $\mathbb{R}^d$.*

Proof of Proposition 1.1. Let $x_0 \in \mathbb{R}^d$ be arbitrary and consider the set

$$E_0 := \{x \in \mathbb{R}^d : f(x) \leq f(x_0)\}.$$

By assumption, E_0 is a compact set and it is clear that if x^* solves

$$\inf_{x \in E_0} f(x) \tag{1.2}$$

then it solves (1.1). However, the existence of a solution to (1.2) follows from the Weierstrass theorem: every continuous function on a compact set reaches its extrema on this set. $\square$

1.2. Optimality conditions.

Now, consider (1.1), and assume that x^* is a solution. Naturally, one would like to have either an explicit or a good enough numerical approximation of the minimiser x^*. Unless we are quite lucky and an easy comparison argument provides an explicit value, this is a hopeless endeavour. The best we can do is to rely on *optimality conditions*. Optimality conditions allow to reduce the search of a minimiser to the resolution of a non-linear system of equations.

There are two optimality conditions. The first-order optimality condition reads

$$\nabla f(x^*) = 0 \tag{1.3}$$

while the second-order optimality condition writes

$$\nabla^2 f(x^*) \in S_d^+(\mathbb{R}).$$

For the sake of future references, let us single out the following definition:

Definition 1.3. *Let $f \in \mathscr{C}^1(\mathbb{R}^d;\mathbb{R})$. A point $x^* \in \mathbb{R}^d$ is called a* critical point *of f if*

$$\nabla f(x^*) = 0.$$

Proposition 1.2. *Assume f is $\mathscr{C}^2$ and let x^* be a critical point of f.*

(1) *If $\nabla^2 f(x^*) \in S_d^{++}(\mathbb{R})$, then x^* is a strict local minimiser of f.*
(2) *If $\nabla^2 f(x^*) \in S_d^{--}(\mathbb{R})$, then x^* is a strict local maximiser of f.*
(3) *If $\nabla^2 f(x^*)$ has at least one negative and one positive eigenvalue, x^* is a* saddle *point: there exist two orthogonal directions $\vec{e}_1\,,\vec{e}_2$ such that $t^* = 0$ is a local minimiser of $t \mapsto f(x+te_1)$, and a local maximiser of $t \mapsto f(x+te_2)$.*
(4) *If $\nabla^2 f(x^*) \in S_d^+(\mathbb{R})$, but not in $S_d^{++}(\mathbb{R})$, then we cannot conclude and further analysis is required.*

In the first two cases, we say that x^* is a non-degenerate critical point.

Proof of Proposition 1.2. We only prove the first and third points. We first assume that

$$\nabla^2 f(x^*) \in S_d^{++}(\mathbb{R}). \tag{1.4}$$

There are two ways to prove that x^* is a local minimiser, both of which naturally rely on Taylor expansions, and on the following consequence of (1.4) (see Exercise 1.1): there exists a constant $\underline{\lambda} > 0$ such that, for any $z \in \mathbb{R}^d$,

$$\langle \nabla^2 f(x^*) z, z\rangle \geq \underline{\lambda}\|z\|^2. \tag{1.5}$$

(1) <u>First approach: a proof by contradiction</u> Argue by contradiction and assume that there exists a sequence $\{x_k\}_{k\in\mathbb{N}}$ such that

$$\forall k \in \mathbb{N}\,, f(x_k) \leq f(x^*).$$

From the mean-value theorem, we know that for any $k \in \mathbb{N}$ there exists $\xi_k \in [x_k; x^*] = \{(1-t)x_k + tx^*\,, t \in [0;1]\}$ such that

$$f(x_k) = f(x^*) + \frac{1}{2}\langle \nabla^2 f(\xi_k)(x_k - x^*), x_k - x^*\rangle.$$

In particular, setting for any $k \in \mathbb{N}$ $z_k := \frac{x_k - x^*}{\|x_k - x^*\|}$,

$$\langle \nabla^2 f(\xi_k) z_k, z_k\rangle \leq 0.$$

As for any $k \in \mathbb{N}$ we have $\|z_k\| = 1$ we can (up to taking a subsequence) assume that $\{z_k\}_{k\in\mathbb{N}}$ converges to some $z_\infty\,, \|z_\infty\| = 1$. Since $x_k \underset{k\to\infty}{\to} x^*$, it follows that $\xi_k \underset{k\to\infty}{\to} x^*$. Passing to the limit in the previous inequality we obtain

$$\langle \nabla^2 f(x^*) z_\infty, z_\infty\rangle \leq 0,$$

in contradiction with (1.5).

(2) Second approach: continuity of the Hessian From (1.5) and the fact that $\nabla^2 f$ is continuous it is possible to show the following fact (see Exercise 1.1): there exists $\underline{\lambda}' > 0$ and $\varepsilon > 0$ such that

$$\forall x \in \mathbb{B}(x^*; \varepsilon), \forall z \in \mathbb{R}^d, \langle \nabla^2 f(x) z, z \rangle \geq \underline{\lambda}' \|z\|^2.$$

We can conclude as before: fix such an $\varepsilon > 0$. Then, for any $x \in \mathbb{B}(x^*; \varepsilon)$, there exists $\xi \in \mathbb{B}(x^*; \varepsilon)$ such that

$$f(x) - f(x^*) = \frac{1}{2} \langle \nabla^2 f(\xi)(x - x^*), (x - x^*) \rangle \geq \frac{\underline{\lambda}'}{2} \|x - x^*\|^2$$

and the conclusion follows.

□

The exercises of this chapter (in particular Exercise 1.3) contain several examples of optimisation problems that can be solved by hand. Such examples are usually limited to dimension 2 or 3, unless the problem has a very specific structure.

2. Convexity, strict convexity

The second crucial notion is that of convexity, which has two main interests:

(1) The first one is when studying global properties of a functions: in that case, assuming convexity helps in either proving uniqueness of the optimiser (in the strongly convex case) or at least to have some geometric structure on the set of minimisers.
(2) The second one is more local in nature: consider a function $f : \mathbb{R}^d \to \mathbb{R}$ that admits a non-degenerate local minimum $x^* \in \mathbb{R}^d$. Then, locally around x^*, by Taylor expansions, f can be suitably approximated by a (strongly) convex functions. Thus, we can hope that any global study of convex functions can translate to local studies around non-degenerate local minima (this is of particular importance when dealing with gradient descents).

2.1. **Various definitions of convexity and basic properties.**

Definition 2.1. *A set $K \subset \mathbb{R}^n$ is said to be convex if for all x and y in K, $tx + (1-t)y \in K$ for all t in $[0, 1]$.*

Definition 2.2. *Let $K \subset \mathbb{R}^n$ be a convex set and $f : K \to \mathbb{R}$ be a function.*

(1) *f is **convex** on K if*
$$\forall x\,, y \in K,\ t \in (0, 1)\,, f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$
(2) *f is **strictly convex** on K if*
$$\forall x \neq y \in K,\ t \in (0, 1)\,, f(tx + (1-t)y) < tf(x) + (1-t)f(y).$$
(3) *f is **strongly convex** on K if there exists $\alpha > 0$ such that*
$$\forall x, y \in K,\ t \in [0, 1]\,, f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\alpha}{2}t(1-t)\|x-y\|^2.$$
(4) *f is said to be **concave** if $-f$ is convex (and similar definitions for strictly or strongly concave).*

Of course, these definitions can be a tad annoying to work with. When f is more regular (say, $\mathscr{C}^1$ or $\mathscr{C}^2$), equivalent characterisations are available.

Proposition 2.1 (Equivalent characterisation of convexity in the $\mathscr{C}^1$ regime)**.** *Let $f \in \mathscr{C}^1(\mathbb{R}^d; \mathbb{R})$. Then the following propositions are equivalent:*

(1) *f is convex.*
(2) *(A convex function is above its tangent hyperplane) For any $x\,, y \in \mathbb{R}^d$,*
$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle. \tag{2.1}$$
(3) *(The gradient of a convex function is monotone) For any $x\,, y \in \mathbb{R}^d$,*
$$\langle \nabla f(y) - \nabla f(x), y - x\rangle \geq 0. \tag{2.2}$$

Proof of Proposition 2.1. Assume that f is convex and let $x\,, y \in \mathbb{R}^d$. Thus, for any $t \in [0; 1]$,
$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$
or, alternatively,
$$\frac{f(x + t(y-x)) - f(x)}{t} \leq f(y) - f(x).$$
As f is $\mathscr{C}^1$, passing to the limit $t \to 0$ gives
$$\langle \nabla f(x), y - x\rangle \leq f(y) - f(x),$$

which is exactly (2.1). Now, assuming (2.1), we obtain

$$\langle \nabla f(x), y-x\rangle \leq f(y)-f(x) \leq -\langle \nabla f(y), x-y\rangle$$

or, equivalently,

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

Now, assuming (2.2), let us show (2.1). Let $x\,,y\in\mathbb{R}^d$. Then, by the Taylor formula,

$$\begin{aligned} f(y)-f(x) &= \int_0^1 \langle \nabla f\left((1-t)x+ty\right), y-x\rangle dt \\ &= \int_0^1 \langle \nabla f\left((1-t)x+ty\right)-\nabla f(x), y-x\rangle dt + \langle \nabla f(x), y-x\rangle. \end{aligned}$$

However, setting $x_t:=(1-t)x+ty$, we have

$$x_t - x = t(y-x).$$

In particular,

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

whence

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

so that (2.1) is satisfied. Finally, assume (2.1). Let us show that f is convex. Let $x\,,y\in\mathbb{R}^d$ and $t\in[0;1]$. Consider the map

$$g: t\mapsto f\left((1-t)x+ty\right)-(1-t)f(x)-tf(y).$$

Then $g(0)=g(1)=0$ and, retaining the notation $x_t:=(1-t)x+ty$,

$$g'(t)=\langle \nabla f(x_t), y-x\rangle - (f(y)-f(x)).$$

Observe that for any $t_0\,,t_1\in[0;1]$ there holds

$$\begin{aligned} (t_0-t_1)(g'(t_0)-g'(t_1)) &= \langle \nabla f(x_{t_0})-\nabla f(x_{t_1}), (t_0-t_1)(y-x)\rangle \\ &= \langle \nabla f(x_{t_0})-\nabla f(x_{t_1}), x_{t_0}-x_{t_1}\rangle \\ &\geq 0 \end{aligned}$$

whence g' is non-decreasing. By the Rolle theorem, there exists $s\in(0;1)$ such that $g'(s)=0$. Consequently, g is non-increasing on $(0;s)$ and non-decreasing on $(s;1)$, and is thus maximal at either $t=0$ or $t=1$, thereby concluding the proof of convexity of f. □

We leave the following proposition as an exercise:

Proposition 2.2 (Equivalent characterisation of strong convexity in the $\mathscr{C}^1$ regime)**.** *Let $f\in\mathscr{C}^1(\mathbb{R}^d;\mathbb{R})$. Then the following propositions are equivalent:*

(1) *f is α-strongly convex.*
(2) *For any $x\,,y\in\mathbb{R}^d$,*

$$f(y)\geq f(x)+\langle \nabla f(x), y-x\rangle + \frac{\alpha}{2}\|y-x\|^2. \tag{2.3}$$

(3) *For any $x\,,y\in\mathbb{R}^d$, $x\neq y$*

$$\langle \nabla f(y)-\nabla f(x), y-x\rangle \geq \alpha\|y-x\|^2. \tag{2.4}$$

To conclude these reminders, we recall some characterisation of convex $\mathscr{C}^2$ functions:

Proposition 2.3. *Let $f\in\mathscr{C}^2(\mathbb{R}^d;\mathbb{R})$. Then the following properties are equivalent:*

(1) *f is convex.*
(2) *For any $x \in \mathbb{R}^d \,, \nabla^2 f(x) \in S_d^+(\mathbb{R})$.*

There is also a nice characterisation of α-strongly convex functions:

Proposition 2.4. *Let $f \in \mathscr{C}^2(\mathbb{R}^d;\mathbb{R})$ and $\alpha > 0$. Then the following properties are equivalent:*

(1) *f is α-strongly convex.*
(2) *For any $x \in \mathbb{R}^d$, the lowest eigenvalue $\lambda_1(\nabla^2 f(x))$ of the Hessian of f satisfies*
$$\lambda_1\left(\nabla^2 f(x)\right) \geq \alpha.$$

Unfortunately, there is no such nice characterisation of strict convexity, but merely an implication (we also refer to Exercise 1.5):

Proposition 2.5. *Let $f \in \mathscr{C}^2(\mathbb{R}^d;\mathbb{R})$. If*
$$\forall x \in \mathbb{R}^d \,, \nabla^2 f(x) \in S_d^{++}(\mathbb{R})$$
then f is strictly convex.

Similarly, we leave the proofs of these propositions as exercises.

2.2. **Convex function and minimisation.** The main result of this section is the following:

Theorem 2.1. *Let f be a convex function, and assume that the problem*
$$\min_{x \in \mathbb{R}^d} f(x) \tag{2.5}$$
has a solution x^*. Then:

(1) *If f is strictly convex, (2.5) has a unique solution.*
(2) *In general, the set of minimisers*
$$X := \{x \in \mathbb{R}^d : \, f(x) = f(x^*)\}$$
is a convex set.

Proof of Theorem 2.1. We begin with the fact that the set of minimisers X is convex: let $x \,, y$ be such that $f(x) = f(y) = f(x^*)$. Then, by convexity of f,
$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) = f(x^*),$$
whence $(1-t)x + ty \in X$. Now, suppose that f is strictly convex and assume by contradiction that f has two distinct minimisers $x \,, y$. By strict convexity, for any $t \in (0;1)$,
$$f((1-t)x + ty) < (1-t)f(x) + tf(y) = f(x^*),$$
in contradiction with the minimality of x^*. $\square$

2.3. **Several (useful) inequalities related to convex functions.** To conclude the first part, we will give and prove several inequalities related to convex functions– these might seem quite abstruse at first, but they will be coming in handy in later parts of the class. Furthermore, the proofs allow to get familiar with usual tricks when dealing with convex functions.

Proposition 2.6. *Let f be a convex function and assume that ∇f is μ-Lipschitz continuous. Then*

$$\forall x\,, y \in \mathbb{R}^d\,, f(x) - f(y) - \langle \nabla f(y), y - x\rangle \geq \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|^2. \tag{2.6}$$

In particular, if f admits a minimiser x^*, it follows that

$$\forall x \in \mathbb{R}^d\,, f(x) - f(x^*) \geq \frac{1}{2\mu}\|\nabla f(x)\|^2. \tag{2.7}$$

This is sometimes referred to as "co-coercivity of the gradient".

Proof of Proposition 2.6. The key is to introduce an auxiliary point z. Fix $x\,, y \in \mathbb{R}^d$. Then for any $z \in \mathbb{R}^d$

$$\begin{aligned} f(x) - f(y) &= f(x) - f(z) + f(z) - f(y) \\ &\leq \langle \nabla f(x), x - z\rangle \\ &+ f(z) - f(y) \\ &\text{by convexity} \\ &\leq \langle \nabla f(x), x - z\rangle + \langle \nabla f(y), z - y\rangle + \frac{\mu}{2}\|z - y\|^2 \\ &\text{by } \mu\text{-Lipschitzianity of the gradient.} \end{aligned}$$

We now minimise the right-hand side with respect to z. As $\varphi : z \mapsto \langle z, \nabla f(y) - \nabla f(x)\rangle + \frac{\mu}{2}\|y - z\|^2$ is a strictly convex function of z, if z^* is a critical point of φ, then it is a global minimiser of φ. As

$$\nabla\varphi(z^*) = 0 \Leftrightarrow z^* = y + \frac{1}{\mu}\left(\nabla f(x) - \nabla f(y)\right)$$

we obtain

$$f(x) - f(y) \leq \langle \nabla f(x), x\rangle - \langle \nabla f(y), y\rangle + \varphi(z^*).$$

Expanding, we deduce

$$\begin{aligned} f(x) - f(y) &\leq \langle \nabla f(x), x - y\rangle - \frac{1}{\mu}\langle \nabla f(x), \nabla f(x) - \nabla f(y)\rangle \\ &\quad + \frac{1}{\mu}\langle \nabla f(y), \nabla f(x) - \nabla f(y)\rangle + \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|^2 \\ &\leq \langle \nabla f(x), x - y\rangle - \frac{1}{2\mu}\|\nabla f(x) - \nabla f(y)\|^2, \end{aligned}$$

which is the desired inequality. □

The final inequality, the proof of which is to be found in Exercise 1.9, is the Polyak-Lojasiewicz inequality:

Proposition 2.7. *Let $f \in \mathscr{C}^1(\mathbb{R}^d; \mathbb{R})$ be an α-strongly convex function. Then*

$$\forall x \in \mathbb{R}^d\,, f(x) - \inf_{\mathbb{R}^d} f \leq \frac{1}{2\alpha}\|\nabla f(x)\|^2.$$

EXERCISE SHEET #1

Exercise 1.1. *Let $A \in S_d(\mathbb{R})$.*

(1) *Letting $\lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_d(A)$ be the eigenvalues of A, show that*
$$\lambda_1(A) = \inf_{\|z\|^2=1} \langle Az, z\rangle.$$
(2) *Show that for any two $A\,, B \in S_d(\mathbb{R})$ there holds*
$$|\lambda_1(A) - \lambda_1(B)| \leq \|A - B\|_{\mathrm{op}}$$
where $\|\cdot\|_{\mathrm{op}}$ stands for the standard operator norm on the set of matrices.

Exercise 1.2. *Let $A \in S_d(\mathbb{R})$ and $b \in \mathbb{R}^d$. We consider*
$$f : x \mapsto \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle.$$
(1) *Show that f is coercive if, and only if $A \in S_d^{++}(\mathbb{R})$.*
(2) *Show that f is convex if, and only if $A \in S_d^{+}(\mathbb{R})$.*
(3) *Show that f is strictly convex if, and only if $A \in S_d^{++}(\mathbb{R})$.*

Exercise 1.3. *Classify the critical points (local minimisers, local maximisers, saddle points, indeterminate critical points) of the following functions:*

(1) $f_1 : (x, y) \mapsto (x - y)^2 + (x + y)^3$,
(2) $f_2 : (x, y) \mapsto x^2 - 2y^2 + 3xy$,
(3) $f_3(x, y) \mapsto x^4 + y^3 - 3y - 2$.

Exercise 1.4 (Distance between two sets)**.** *Let A and B be two closed, nonempty subsets of $\mathbb{R}^d$.*

(1) *Show that if A is compact, then the problem*
$$\min_{a\in A,\, b\in B} \|a - b\|$$
has a solution (at least one).
(2) *Show with a counter-example that this problem need not have a solution if neither A nor B is assumed compact, even if A and B are convex.*

Exercise 1.5. *Give an example of a strictly convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that the equation*
$$\nabla^2\varphi(x) = 0$$
has infinitely many solutions.

Exercise 1.6 (Carathéodory theorem)**.** *Let $\Omega \subset \mathbb{R}^d$. We call the convex hull of Ω the smallest convex set containing Ω. We denote it by $C(\Omega)$.*

(1) *Show that*
$$C(\Omega) = \left\{\sum_{i=0}^{N} t_i x_i\,, N \in \mathbb{N}\,, \{t_i\}_{i=0,\dots,N} \in [0;1]^{N+1}\,, \sum_{i=0}^{N} t_i = 1\,, \{x_i\}_{i=0,\dots,N} \in \Omega^{N+1}\right\}.$$
(2) *We now want to show the Carathéodory theorem: for any $x \in C(\Omega)$, there exist $t_0, \dots, t_d \in [0;1]\,, x_0, \dots, x_d \in \Omega$ such that*
$$\sum_{i=0}^{d} t_i = 1\,, x = \sum_{i=0}^{d} t_i x_i.$$

(a) *Using an example, show why one needs at least* $(d+1)$ *points.*
(b) *Prove the Carathéodoory theorem; you can argue by descending induction, starting (for instance) from a point* $x \in C(\Omega)$ *that writes* $x = \sum_{i=0}^{d+1} t_i x_i$*, and showing that one of the vectors* $x_i's$ *can be expressed using the others.*
(c) *Deduce from the Carathéodory theorem that if* Ω *is compact, then so is* $C(\Omega)$*.*

Exercise 1.7 (Extreme points I: projection on closed convex sets)**.** *Let* $K \subset \mathbb{R}^d$ *be a closed convex set. Show that there exists a unique* $z \in K$*, denoted by* Π_K *and dubbed the orthogonal projection of* x *on* K*, such that*

$$\|x - \Pi_K(x)\| = \min_{z \in K} \|x - z\|$$

and that

$$\forall y \in K\,, \langle x - \Pi_K(x), y - \Pi_k(x)\rangle \leq 0.$$

Show that Π_K *is* 1*-Lipschitz.*

Exercise 1.8 (Extreme points II: The Krein-Milman theorem)**.** (1) *Give an example of a convex set* $K \subset \mathbb{R}^d$ *that has no extreme points.*
(2) *We assume that* K *is compact. Prove that* K *has extreme points.*
(3) *We now want to prove the (finite-dimensional) Krein-Milman theorem: any* $x \in K$ *is a convex combination of extreme points of* K*.*
   (a) *Let* $x \in \partial K$*. Show that there exists a hyperplane (called the supporting hyperplane)* $H = \{\varphi = 0\}$ *where* $\varphi \in (\mathbb{R}^d)'\,, \varphi \neq 0$ *(the dual of* $\mathbb{R}^d$*) such that* $x \in H$ *and* $\varphi(K) \subset (-\infty; 0]$ *(hint: think of the projection* Π_K*).*
   (b) *Let* $x \in K$*. Show that if* $x \in H$ *for some supporting hyperplane of* K *(i.e. associated with some* $y \in K$*) then* x *is an extreme point of* K *if, and only if,* x *is an extreme point of* $H \cap K$*.*
   (c) *Show the Krein-Milman theorem proceeding by induction on the dimension.*

Exercise 1.9 (Polyak-Lojasciewicz Inequality)**.** *Let* $f : \mathbb{R} \to \mathbb{R}$ *be an* α*- strongly convex function and let* x^* *be a minimiser of* f*. First, prove that*

$$\forall x \in \mathbb{R}^d\,, \|x - x^*\|^2 \leq \frac{2}{\alpha}(f(x) - f(x^*)).$$

Second, show the following inequality:

$$\forall x \in \mathbb{R}^d\,, f(x) - f(x^*) \leq \frac{1}{2\alpha}\left\|\nabla f(x)\right\|^2.$$

Finally, deduce that

$$\forall x \in \mathbb{R}^d\,, \|x - x^*\| \leq \frac{1}{\alpha}\|\nabla f(x)\|.$$

Exercise 1.10 (Legendre transform)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$*. The Legendre transform (or convex conjugate)* $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *is the function defined as:*

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \left(\langle y, x\rangle - f(x)\right)$$

(1) *Prove that* f^* *is convex.*
(2) *Assume* f *is* μ*-strongly convex. Prove that, for any given* y*, the supremum in the definition of* $f^*(y)$ *is attained at a* unique *point* $x^*(y)$*.*

(3) *Let* $f(x) = \frac{1}{2}\langle x, Qx\rangle$, *where* $Q \in \mathbb{R}^{n\times n}$ *is a symmetric positive definite matrix. Compute its Legendre transform* f^*. *Is* $f^*(y)$ *strictly convex? Strongly convex?*

(4) *Baby Fenchel-Moreau theorem: In the one-dimensional case* ($n = 1$), *assuming* f *is continuously differentiable and strictly convex, prove that* $f^{**} = f$. This result extends to the higher-dimensional case, but requires finer tools. However, just remember that, if f is convex, $f^{**} = f$.

Exercise 1.11 (Sub-differential of convex functions)**.** *For a convex function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *we define the subdifferential at a point* x *(where* $f(x) < +\infty$*), denoted* $\partial f(x)$, *as the set of all vectors* $g \in \mathbb{R}^n$ *satisfying:*

$$f(z) \geq f(x) + \langle g, z - x\rangle \quad \forall z \in \mathbb{R}^n$$

We assume that f *is a coercive convex function on* $\mathbb{R}^n$.

(1) *Prove that a point* x^* *is a global minimizer of* f *if and only if* $0 \in \partial f(x^*)$.

(2) *Prove that the subdifferential operator is monotone in the sense that, for any* $g_x \in \partial f(x)$ *and* $g_y \in \partial f(y)$, $\langle g_x - g_y, x - y\rangle \geq 0$.

(3) ***Fenchel-Young Inequality:*** *Prove that for any* $x, y \in \mathbb{R}^n$, *the following inequality holds:*

$$f(x) + f^*(y) \geq \langle y, x\rangle$$

(4) *Prove that the Fenchel-Young inequality is an equality if, and only if,* $y \in \partial f(x)$.

(5) *Prove that, for any* x, y, $y \in \partial f(x)$ *if, and only if,* $x \in \partial f^*(y)$.

Exercise 1.12. *Consider a company whose state is defined by a vector of capital stocks* $x \in \mathbb{R}^n$. *The firm controls its rate of investment* $v \in \mathbb{R}^n$, *meaning the capital evolves according to* $\dot{x} = v$. *To model dynamic decision-making, the firm seeks to minimise its total cost, given by the integral of an instantaneous* ***running cost*** $L(x, v)$ *over time:*

$$\min \int_0^\infty e^{-\rho t} L(x(t), v(t))\, dt$$

The running cost L *captures two facets:*

(1) *The expenses or penalties associated with maintaining the current capital stock* x *(e.g., depreciation, storage, or inefficiency).*

(2) *The financial burden of changing capital at rate* v.

We assume that $L(x, v)$ *is strictly convex with respect to the investment rate* v.

Solving this infinite-horizon dynamic problem requires the Maximum Principle, which will be covered later in the semester. However, the core of that theory relies on reducing the global integral problem to a sequence of instantaneous, static *optimisations.*

To perform this static step, we introduce a dual variable $p \in \mathbb{R}^n$, *representing the "shadow price" or marginal economic value of capital. For a fixed state* x *and a given shadow price* p, *we define the Hamiltonian* H *as the Legendre transform of* L *with respect to* v:

$$H(x, p) = \sup_{v\in\mathbb{R}^n} \left(\langle p, v\rangle - L(x, v)\right)$$

(1) *Suppose the firm's running cost is* $L(x, v) = \frac{1}{2}\langle v, Qv\rangle + C(x)$, *where* $Q \in \mathbb{R}^{n\times n}$ *is a symmetric positive definite matrix representing adjustment costs,*

and $C(x)$ is the penalty for the current capital state. Compute the Hamiltonian $H(x, p)$ explicitly and provide an economic interpretation of each of its terms.

(2) *Assume L is μ-strongly convex in v. Derive the explicit relationship between the optimal instantaneous investment rate v^* and the shadow price p.*

(3) *Prove that, if the company follows this optimal investment strategy, then $\dot{x} = \nabla_p H(x, p)$.*

(4) *Bonus (The Adjoint Equation): A fundamental result of the calculus of variations states that the optimal dynamic trajectory must satisfy the Euler-Lagrange equation. For our discounted cost, this yields:*

$$\frac{d}{dt}\left(\nabla_v L(x, \dot{x})\right) - \rho \nabla_v L(x, \dot{x}) = \nabla_x L(x, \dot{x})$$

Using this Euler-Lagrange equation, prove that the shadow price must evolve according to the differential equation $\dot{p} = \rho p - \nabla_x H(x, p)$.