0.1 Introduction and Motivation

This course is an introduction to continuous mathematical optimization in finite dimension. It is intended for third-year undergraduate mathematics students. The document is composed of the following chapters:

- Review of differential calculus
- Convex analysis
- Existence and uniqueness results
- Optimality conditions
- Gradient methods
- Newton's method

The origin of the word *optimization* comes from the Latin word *optimum*, which means "the best." In the *Larousse* dictionary, the verb *to optimize* is defined as: *To give something, a machine, a company, etc., the optimal performance by creating the most favorable conditions or by making the best possible use of it.*

In mathematics, an optimization problem consists of minimizing (or maximizing) a function of a single variable (or of several variables) over a given set. The formulation of such a problem involves three steps:

Step I: Identify the decision variables, which are the parameters of the problem that can be controlled. They are represented by a column vector

$$x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n.$$

Step II: Identify a mathematical function (a measure) for which we seek the smallest or largest value. This function is called the *objective function* or *cost function*, denoted by f.

Step III: Describe the constraints on the decision variables.

0.2 Mathematical Formulation

As mentioned earlier, the mathematical formulation of an optimization problem involves defining the set of decision variables that govern the situation to be modeled. These variables may be real, integer, or binary.

Next, one identifies the objective function, which is a linear or nonlinear mathematical function composed of the decision variables, representing the modeled physical system.

Finally, one describes the set of parameters that restrict the feasible model through equations or inequalities involving the decision variables. These are called the *constraints*.

Mathematically, an optimization problem can be represented as follows:

Objective function:	$\max f(x)$ or $\min f(x)$
Constraints:	$g(x) \leq 0$, $g(x) \geq 0$, or $g(x) = 0$
Bound constraints:	$l \leq x \leq u$
Sign constraints:	$x \leq 0$, $x \geq 0$

0.3 Application Examples

0.3.1 Example: Least Squares Problems

A common situation in biology is to have two sets of data of size n,

$$y_1, y_2, \ldots, y_n \quad \text{and} \quad x_1, x_2, \ldots, x_n,$$

obtained experimentally or measured from a population.

A regression problem consists of finding a function

$$y = g(\mu, x)$$

such that the error

$$E_i = y_i - g(\mu, x_i)$$

is as small as possible.

The principle of least squares consists in finding the parameters of g (an affine function, polynomial, exponential, etc.) that minimize the sum of the squares of the distances between y_i and $g(\mu, x_i)$, that is:

$$\min_{\mu \in \mathbb{R}^n} f(\mu) = \sum_{i=1}^{m} E_i(\mu)^2, \quad \text{with } E_i(\mu) = y_i - g(\mu, x_i).$$

Here, E_i is called the *residual* or *error*. The objective function (or cost function) to be minimized is the sum of squared residuals $E_i(\mu)^2$.

As an example, the following data represent the size of an antelope population at different times:

x_i	1	2	4	5	8
y_i	3	4	6	11	20

Here, time is measured in years, and the population is measured in hundreds.

It is common to model the evolution of a population using an exponential fit of the form:

$$g(\mu, x) = \mu_1 \cdot \exp(\mu_2 \cdot x).$$

Objective: Find the parameters μ_1 and μ_2 that minimize the least squares criterion.

0.4 Exercises

Exercise 0.1 We want to build a box by cutting out four squares from the corners of a rectangular sheet of cardboard and folding up the remaining flaps. The sheet measures 22 cm in length and 18 cm in width. The volume of the constructed box depends on the size of the squares cut out. For which value of x does the box attain the maximum possible volume?

Exercise 0.2 We want to construct a box from a rectangular sheet of cardboard by cutting six squares of side x at each corner and at the middle of the sides, and then folding the sides as shown in the figure. This sheet has dimensions 45×30 cm. The aim of this exercise is to determine the dimensions of the closed box that yield the maximum volume.

a) Determine the function to be optimized.

b) Justify the following relations:

$$p = 30 - 2x, \qquad l = \frac{45 - 3x}{2}.$$

c) Determine the set of admissible solutions for x.

d) Show that the volume expressed as a function of x can be written as

$$v(x) = 3x^3 - 90x^2 + 675x.$$

e) Find the value of x for which the volume is maximal.

Solutions to the Exercises

Solution 0.1 The decision variable is the side length of the cut square, denoted by x. The objective is to determine the maximal volume of the constructed box, which is written mathematically as

$$\max V(x) = x(22 - 2x)(18 - 2x),$$

with

$$0 \leq x \leq \tfrac{1}{2}\min(22, 18) = 9.$$

Solution 0.2

a) The function to be optimized is the volume, given by

$$V(x) = \tfrac{1}{2}\, x(45 - 3x)(30 - 2x).$$

b) The box width is obtained by cutting two squares from the 30 cm side, hence

$$p = 30 - 2x.$$

The box length is obtained by cutting three squares from the 45 cm side and folding the remaining length in two, hence

$$l = \frac{45 - 3x}{2}.$$

c) The admissible set D for x is

$$D = \left\{ 0 \leq x \leq \tfrac{1}{2} \min(45, 30) \right\} = \{ 0 \leq x \leq 15 \}.$$

d) The volume expressed as a function of x is

$$V(x) = 3x^3 - 90x^2 + 675x.$$

Chapter 1

Differential Calculus

In this chapter, we recall the main results of differential calculus, essentially the definition of the first-order and second-order differential.

What is Differential Calculus?

Differential calculus is a branch of mathematics that studies how functions change when their inputs change. It focuses on the concept of the **derivative**, which measures the rate of change of a function at a given point.

Definition of the Derivative

Let $f(x)$ be a real-valued function. The derivative of f at a point x is defined as

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h},$$

if this limit exists.

Geometric Meaning

The derivative represents the slope of the tangent line to the curve $y = f(x)$ at a point.

- If $f'(x) > 0$, the function is increasing at x.
- If $f'(x) < 0$, the function is decreasing at x.

Example

Consider the function $f(x) = x^2$. Then

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = \lim_{h\to 0} \frac{2xh + h^2}{h} = 2x.$$

Thus, the slope of the curve $y = x^2$ at a point x is $2x$.

Applications of Derivatives

Differential calculus has many applications:

- Finding maxima and minima (optimization problems).
- Studying concavity and inflection points.
- In physics: velocity and acceleration are derivatives of position.
- In economics: marginal cost and marginal revenue are derivatives of cost and revenue functions.

1.1 First-Order Differential

Let $U \subset \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^m$, such that for $x = (x_1, x_2, \ldots, x_n)^\top$ we have

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{pmatrix}.$$

Before giving the definition of differentiability, it is important to recall that of continuity:

Definition 1.1.1. A function f is said to be continuous at the point x_0 if

$$\forall \varepsilon > 0,\ \exists \delta > 0 : \|x - x_0\|_{\mathbb{R}^n} \leq \delta \ \Rightarrow\ \|f(x) - f(x_0)\|_{\mathbb{R}^m} \leq \varepsilon.$$

In dimension 1, a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x_0 \in \mathbb{R}$ if there exists a real number $f'(x_0)$ such that:

$$\lim_{h\to 0} \frac{|f(x_0 + h) - f(x_0) - f'(x_0)h|}{|h|} = 0.$$

In strictly higher dimensions, this definition is expressed as f

Example 1.1.2. Let $f(x) = Ax + b$, with $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $b \in \mathbb{R}^m$, and let $x_0 \in \mathbb{R}^n$. Then

$$f(x_0 + h) = A(x_0 + h) + b = f(x_0) + A(h),$$

hence

$$Df(x_0) = A.$$

Exercise 1.1.3. Show that if f is differentiable at a point x_0, then f is necessarily continuous at x_0.

Definition 1.1.4. Let U be an open subset of $\mathbb{R}^n$ and let f be a function defined from U into $\mathbb{R}^m$. We say that f is **differentiable on** U if f is differentiable at every point $x_0 \in U$. In this case, we call the **differential of** f the function

$$Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), \quad x \mapsto Df(x).$$

If, in addition, the differential Df is a continuous function on U, then we say that f is **continuously differentiable**, or that f is of class $\mathcal{C}^1$.

The **partial derivatives** of f are the functions denoted

$$\frac{\partial f_i}{\partial x_j}$$

defined by

$$\frac{\partial f_i(x)}{\partial x_j} = \lim_{t \to 0} \frac{f_i(x_1, \ldots, x_j + t, x_{j+1}, \ldots, x_n) - f_i(x_1, \ldots, x_n)}{t}.$$

We define the **Jacobian matrix** at the point x as the matrix of the linear map $Df(x)$ in the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. It is given by

$$[Df(x)]_{i,j} = \frac{\partial f_i}{\partial x_j}(x), \quad i = 1, \ldots, m, \ j = 1, \ldots, n.$$

If $m = 1$, we denote by $\nabla f(x)$ the transpose of $Df(x)$, called the **gradient** of f at the point x.

Remark 1.1.5. A function f may have partial derivatives without being differentiable. For example, let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2 + y^2}}, & \text{if } x^2 + y^2 \neq 0, \\ 0, & \text{if } (x,y) = (0,0). \end{cases}$$

The function f is zero at the point $(0,0)$, and its partial derivatives exist and are zero, although f is not differentiable at this point.

Indeed, by the previous theorem we should have

$$f(h_1, h_2) = f(0,0) + Df(0,0)(h) + \|h\|\varepsilon(h) = \|h\|\varepsilon(h),$$

where $\|h\| = \sqrt{h_1^2 + h_2^2}$. For $h = (3t, 4t)$ we get

$$\frac{f(3t, 4t)}{\|h\|} = \frac{12}{25},$$

which does not tend to zero as $t \to 0$. Hence f is not differentiable at $(0,0)$.

Lemma 1.1.6. *Let f be a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ differentiable at the point x. Then*

$$\lim_{t\to 0} \frac{f(x+th)-f(x)}{t} = Df(x)(h).$$

1.1.1 Proof of Lemma 1.1

Proof. Assume f is differentiable at the point x. For all t in some open interval $I \subset \mathbb{R}$ containing 0, we have

$$f(x+th) = f(x) + t\,Df(x)(h) + \|th\|\,\varepsilon(th),$$

with $\lim\limits_{t\to 0} \varepsilon(th) = 0$. Dividing by t and passing to the limit yields

$$\lim_{t\to 0} \frac{f(x+th)-f(x)}{t} = Df(x)(h).$$

□

Example 1.1.7 (Example 1.3)**.** Let f be the quadratic function

$$f(x) = \tfrac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle,$$

where $A \in M_n(\mathbb{R})$ is a symmetric matrix and $b \in \mathbb{R}^n$. Compute the differential of f at the point x.

We compute

$$Df(x)(h) = \lim_{t\to 0} \frac{f(x+th)-f(x)}{t}.$$

First expand:

$$\begin{aligned} f(x+th) &= \tfrac{1}{2}\langle x+th, A(x+th)\rangle - \langle b, x+th\rangle \\ &= f(x) + \tfrac{1}{2}\big(t\langle h, Ah\rangle + 2t\langle h, Ax\rangle\big) - t\langle b, h\rangle + \tfrac{1}{2}t^2\langle h, Ah\rangle \\ &= f(x) + t\langle Ax-b, h\rangle + \tfrac{1}{2}t^2\langle h, Ah\rangle. \end{aligned}$$

Thus

$$\frac{f(x+th)-f(x)}{t} = \langle Ax-b, h\rangle + \tfrac{1}{2}t\langle h, Ah\rangle,$$

and letting $t \to 0$ gives

$$Df(x)(h) = \langle Ax-b, h\rangle, \qquad \forall h \in \mathbb{R}^n.$$

Therefore, in matrix form,

$$Df(x) = (Ax-b)^\top.$$

Exercise 1.1.8 (Exercise 1.2)**.** Show that the following functions are differentiable and compute their differentials:

1. $f(x) = \langle x, Ax\rangle, \quad x \in \mathbb{R}^n$, with $A \in M_n(\mathbb{R})$.
2. $f(x,y) = x^2 - y^2, \quad (x,y) \in \mathbb{R}^2$.
3. $f(x) = \|x\|^2, \quad x \in \mathbb{R}^n$.

1.2 Second-Order Differential

Definition 1.2.1. Let f be a function defined on an open set $U \subset \mathbb{R}^n$ with values in $\mathbb{R}$. Assume f is differentiable at every point of U, so that $Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ is defined. If moreover Df is differentiable at a point $x \in U$, we say that f is twice differentiable at x. The second-order differential at x is denoted by $\nabla^2 f(x)$.

Theorem 1.2.2. *Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$. If f is twice differentiable at x, then there exists a symmetric matrix $\nabla^2 f(x)$ and a continuous function $\varepsilon(\cdot)$ with $\varepsilon(\|h\|) \to 0$ as $h \to 0$, such that*

$$f(x+h) = f(x) + Df(x) \cdot h + \tfrac{1}{2} h^\top \nabla^2 f(x)\, h + \|h\|^2 \varepsilon(h).$$

The matrix $\nabla^2 f(x)$ is called the Hessian *of f at x and is given by*

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

i.e.

$$[\nabla^2 f(x)]_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Example 1.2.3. Consider the quadratic form

$$q(x) = \tfrac{1}{2}\langle x, Hx\rangle - \langle b, x\rangle, \tag{1.1}$$

where H is a symmetric $n \times n$ matrix. We have already seen that

$$Dq(x) = (Hx - b)^\top,$$

and therefore

$$\nabla^2 q(x) = H.$$

1. First-Order Derivative

Example: Find the first derivative of

$$f(x) = 3x^3 - 5x^2 + 2x - 7.$$

Solution:

$$f'(x) = \frac{d}{dx}(3x^3) - \frac{d}{dx}(5x^2) + \frac{d}{dx}(2x) - \frac{d}{dx}(7).$$
$$f'(x) = 9x^2 - 10x + 2.$$

2. Second-Order Derivative

Example: Find the second derivative of

$$f(x) = \sin(x).$$

Solution:

$$f'(x) = \cos(x), \quad f''(x) = -\sin(x).$$

3. Gradient of a Scalar Function

Example: Let

$$f(x, y) = x^2 y + 3y^2.$$

Find $\nabla f(x, y)$.

Solution:

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}.$$

Compute each derivative:

$$\frac{\partial f}{\partial x} = 2xy, \quad \frac{\partial f}{\partial y} = x^2 + 6y.$$

Thus,

$$\nabla f(x, y) = \begin{bmatrix} 2xy \\ x^2 + 6y \end{bmatrix}.$$

4. Jacobian of a Vector Function

Example: Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$\mathbf{F}(x, y) = \begin{bmatrix} f_1(x, y) \\ f_2(x, y) \end{bmatrix} = \begin{bmatrix} x^2 y \\ e^x + y^2 \end{bmatrix}.$$

Find the Jacobian matrix $J_{\mathbf{F}}(x, y)$.

Solution: The Jacobian is

$$J_{\mathbf{F}}(x, y) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}.$$

Compute each partial derivative:

$$\frac{\partial f_1}{\partial x} = 2xy, \quad \frac{\partial f_1}{\partial y} = x^2,$$

$$\frac{\partial f_2}{\partial x} = e^x, \quad \frac{\partial f_2}{\partial y} = 2y.$$

Thus,

$$J_{\mathbf{F}}(x, y) = \begin{bmatrix} 2xy & x^2 \\ e^x & 2y \end{bmatrix}.$$

5. Second-Order Partial Derivatives (Hessian Matrix)

Example: Let

$$f(x, y) = x^3 + xy^2.$$

Find the Hessian matrix $H_f(x, y)$.

Solution: First-order derivatives:

$$\frac{\partial f}{\partial x} = 3x^2 + y^2, \quad \frac{\partial f}{\partial y} = 2xy.$$

Second-order derivatives:

$$\frac{\partial^2 f}{\partial x^2} = 6x, \quad \frac{\partial^2 f}{\partial x \partial y} = 2y,$$

$$\frac{\partial^2 f}{\partial y \partial x} = 2y, \quad \frac{\partial^2 f}{\partial y^2} = 2x.$$

Thus the Hessian is

$$H_f(x, y) = \begin{bmatrix} 6x & 2y \\ 2y & 2x \end{bmatrix}.$$

1.3 Exercises

Exercise 1.3 Let the following functions be defined on $\mathbb{R}^n$ (or $\mathbb{R}^2$ where indicated):

1. $f_1(x) = \frac{3}{2}x_1^2 - 2x_2^2 - \frac{3}{2}x_3^2 + x_1x_2 - 2x_2x_3 - 3x_1 - x_3$,
2. $f_2(x) = (x_1 - 1)^2 - 10(x_1 - x_2)^2$,
3. $f_3(x) = 5x_1^2 - 5x_2^2 - x_1x_2 - 11x_1 - 11x_2 - 11$.

For each $i = 1, 2, 3$:

1. Calculate $\nabla f_i(x)$ (the gradient or the transpose of $Df_i(x)$).
2. Calculate $\nabla^2 f_i(x)$ (the Hessian matrix).

2. Among these functions, which ones are quadratic? Justify your answer.

Exercise 1.4 Determine $Df(x)$ and $\nabla^2 f(x)$ for the quadratic function

$$f(x) = \tfrac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle + c,$$

where $A \in M_n(\mathbb{R})$ (an $n \times n$ matrix), $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Exercise 1.5 Give the second-order Taylor expansion of the function f in a neighborhood of the point x_0 for the following cases:

a) $f(x) = x_1 e^{-x_2} - x_2 - 1, \quad x_0 = (1, 0)^\top$.

b) $f(x) = x_1^4 - 2x_1^2x_2^2 - x_2^4, \quad x_0 = (1, 1)^\top$.

Exercise 1.6 Let $x(t) = (t,\ e^t,\ t^2)^\top$, $t \in \mathbb{R}$, and let $f(x) = x_1x_2^3 - x_1x_2 - x_3$ for $x = (x_1, x_2, x_3)^\top \in \mathbb{R}^3$. Find $\dfrac{d}{dt}f(x(t))$ in terms of t.

Exercise 1.7 The goal of this exercise is to recover the Taylor formula for a function $f : \mathbb{R}^n \to \mathbb{R}$. Let $f \in C^2$, let x and $x_0 \in \mathbb{R}^n$, and define

$$z(\alpha) = x_0 - \alpha\frac{(x - x_0)}{\|x - x_0\|}, \qquad \alpha \in \mathbb{R},$$

and the function $\Phi(\alpha) = f(z(\alpha))$.

1. Compute $\Phi'(\alpha)$ and $\Phi''(\alpha)$.
2. Noting that $f(x) = \Phi(\|x - x_0\|)$, deduce that

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \tfrac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0) + o\big(\|x - x_0\|^2\big).$$

Exercise 1.8 Let φ be a continuous function $\varphi : \mathbb{R} \to \mathbb{R}$, and define $f, g : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) = \int_0^{x-y} \varphi(t)\, dt, \qquad g(x,y) = \int_0^{x+y} \varphi(t)\, dt.$$

1. Show that f and g are of class C^1 on $\mathbb{R}^2$.
2. Denote by $Df(x,y)$ and $Dg(x,y)$ the differentials of f and g at the point (x,y). Compute
$$Df(x,y)(h,k) \quad \text{and} \quad Dg(x,y)(h,k)$$
for every $(h,k) \in \mathbb{R}^2$.

Exercise 1.9 Let $f : \mathbb{R}^n \to \mathbb{R}$ and $x : \mathbb{R} \to \mathbb{R}^n$ be two C^2 functions. Define the real-valued function $g(t) = f(x(t))$.

1. Compute $g''(t)$ in the case $x(t) = u + tv$, where u and v are two vectors in $\mathbb{R}^n$.
2. Compute $g''(t)$ for a general C^2 curve $x(t)$.

Solutions to the Exercises

Solution 1.1 Let the functions be

$$\begin{aligned}
f_1(x) &= \tfrac{3}{2}x_1^2 - 2x_2^2 - \tfrac{3}{2}x_3^2 + x_1x_2 - 2x_2x_3 - 3x_1 - x_3,\\
f_2(x) &= (x_1-1)^2 - 10(x_1-x_2)^2,\\
f_3(x) &= 5x_1^2 - 5x_2^2 - x_1x_2 - 11x_1 - 11x_2 - 11.
\end{aligned}$$

1. Calculate $\nabla f_i(x)$ and $\nabla^2 f_i(x)$ for $i = 1,2,3$.

$$\nabla f_1(x) = \begin{pmatrix} 3x_1 - x_2 - 3 \\ 4x_2 - x_1 - 2x_3 \\ 3x_3 - 2x_2 - 1 \end{pmatrix}, \qquad \nabla^2 f_1(x) = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 4 & 2 \\ 0 & 2 & 3 \end{pmatrix}.$$

$$\nabla f_2(x) = \begin{pmatrix} 22x_1 - 20x_2 - 2 \\ -20x_1 - 20x_2 \end{pmatrix}, \qquad \nabla^2 f_2(x) = \begin{pmatrix} 22 & -20 \\ -20 & 20 \end{pmatrix}.$$

$$\nabla f_3(x) = \begin{pmatrix} 10x_1 - x_2 - 11 \\ -x_1 - 10x_2 - 11 \end{pmatrix}, \qquad \nabla^2 f_3(x) = \begin{pmatrix} 10 & -1 \\ -1 & 10 \end{pmatrix}.$$

2. Since each $\nabla^2 f_i(x)$ is constant and symmetric, the functions f_i are quadratic for $i = 1,2,3$.