Chapter 1 : Optimization of multivariable functions

MATHEMATICAL ANALYSIS 4 (2 CPI-Section B)-2024/2025

.

Pr H. Haddadou, ESI, Algiers

Plan

1. **Introduction and Concepts**
   - Introduction
   - Concepts of minimum, maximum and extremum
   - Types of continuous optimization
2. **Unconstrained Optimization (Free extrema)**
   - First-order optimality condition (necessary)
   - Second-order optimality condition (sufficient)
   - Summary on unconstrained optimization and the Hessian in 2 and 3 Dimensions
3. **Constrained Optimization**
   - Equality-constrained optimization problem statement
   - Necessary optimality condition for equality constraints (first order)
   - Sufficient optimality condition for equality constraints (second order)
   - Remark on extrema with inequality constraints
4. **Appendix : Convexity Analysis and Optimization**

Introduction and Concepts

Introduction

This chapter focuses on the study of extrema of functions of several variables.

Purpose : We are interested in minimizing or maximizing certain quantities.

Why ? "There is nothing in the world that happens without the will to minimize or maximize something" (A famous quote by Euler).

How ? By using **mathematical and numerical tools**.

Areas of application for optimization :

- Engineering / Artificial Intelligence, Data Analysis, Machine Learning, ...
- Economics / Finance, Organization, ... • Biology, ;

A mathematical optimization problem is presented as follows :

Minimizing (or maximizing) a function $f : \mathbb{R}^m \to \mathbb{R}$ on C consists of finding $v \in \mathbb{R}$ and $x^* \in C$ such that

$$\begin{cases} v = \underset{x\in C}{Min}\{f(x)\}, \\ f(x^*) = v, \end{cases} \qquad \left(resp. \begin{cases} v = \underset{x\in C}{Max}\{f(x)\}, \\ f(x^*) = v. \end{cases} \right)$$

Terminology :

- The function f is the **objective function**. • v is the optimal value of f.
- x^* is the optimal solution. • $C \subset D_f$ is called the set of constraints.

Concepts of Minimum, Maximum, and Extremum

Definition (Global Extremum)

Let D be a subset of $\mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$ and $p \in D$.. Then, a function f defined on D has a

- **global minimum** *at p if $\forall u \in D,\ f(p) \leq f(u)$,*
- **global maximum** *at p if $\forall u \in D,\ f(p) \geq f(u)$,*
- **global extremum** *at p if it has either a minimum or a maximum at p.*

Definition (Local Extremum)

A function f defined on a set D has a

- **local minimum** *at p if $\exists V(p) \subset \mathbb{R}^m$ such that*

$$\forall u \in V(p) \cap D, f(p) \leq f(u),$$

- **local maximum** *at p if $\exists V(p) \subset \mathbb{R}^m$ such that*

$$\forall u \in V(p) \cap D, f(p) \geq f(u),$$

- **local extremum** *at p if it has either a local minimum or a local maximum.*

Remark

A global (resp. local) maximum or minimum is called strict if it is unique in D (resp. in its neighborhood).

Remark

A local extremum may not necessarily be a global extremum.

Remark

A function f has an extremum

- *globally at p on D* **if**

$$f(u) - f(p) \text{ does not change sign } \forall u \in D,$$

- *locally at p on an open set D* **if** $\exists V(p) \subset \mathbb{R}^m$ *such that*

$$f(u) - f(p) \text{ does not change sign } \forall u \in V(p) \cap D.$$

Continuous optimization's type

There are two types of continuous

- Unconstrained Optimization : Unconstrained optimization refers to the process of finding the maximum or minimum of an objective function without any restrictions on the decision variables. Mathematically, it involves solving :

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

  where $f(x)$ is a continuous function, and x can take any value in $\mathbb{R}^n$ ($\mathbb{R}^n$ can be replaced by an open subset of $\mathbb{R}^n$).

- Constrained Optimization : Constrained optimization involves optimizing an objective function subject to constraints on the decision variables. These constraints can be equality constraints ($g_i(x) = 0$) and/or inequality constraints ($h_j(x) \leq 0$). The general formulation is :

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad g_i(x) = 0, \quad h_j(x) \leq 0$$

  where $g_i(x)$ and $h_j(x)$ define the feasible region. .

Unconstrained Optimization (Free Extrema)

Study of Extremum

To study analytically the extremum of a function f of class C^1 on $\mathbb{R}^m$ we can follow the folowing method :

1. Determine of what is called critical points by first-order optimality condition
2. Determine the nature of the critical points
   - By definition : We study sign of

     $$f(p+H)-f(p)$$

     with $H \in \mathbb{R}^m$.

   - If f is a fairly smooth function, we use second-order optimality condition Otherwise, we use Taylor's expansion in the neighborhood of the critical point up to a sufficiently high order.

First-Order Optimality Condition (necessary)

Definition (Critical Point)

Let D be an **open set** *in $\mathbb{R}^m$. A point $p \in D$ is called a* **critical point** *of a differentiable function $f : D \to \mathbb{R}$ if*

$$\nabla f(p) = 0.$$

The following result provides a necessary condition for a point to be an extremum :

Theorem (First-Order Optimality Condition (necessary))

Let D be an **open set** *in $\mathbb{R}^m$, $p \in D$ and $f : D \to \mathbb{R}$ a function of class C^1. Then,*

$$(f \text{ has a local extremum at } p \in D) \Rightarrow (p \text{ is a critical point of } f).$$

Remark

The converse is false.

Different Types of Critical Points

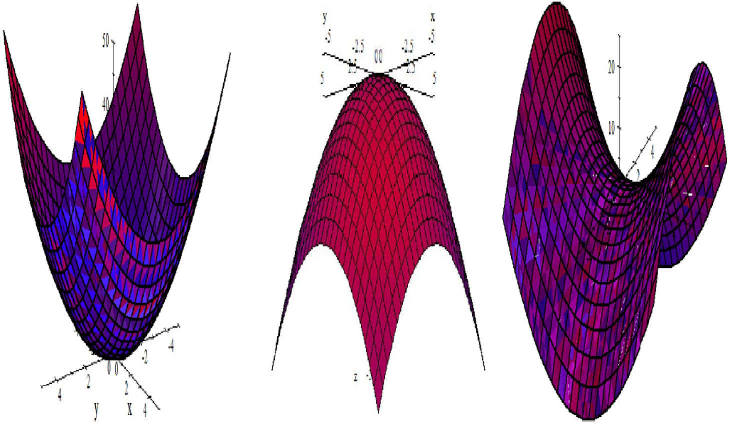


Figure – Existence of : Minimum Point - Maximum Point - Saddle Point

Study the Extremum by Using Definition

Example

Let $f(x,y) = x^2 + y^2 + xy$. We have $f \in C^1(\mathbb{R}^m)$.

Critical points of f : We have

$$\nabla f(x,y) = \begin{pmatrix} 2x+y \\ 2y+x \end{pmatrix}.$$

Thus,

$$((x,y)\text{is a critical point of } f) \Leftrightarrow \begin{cases} 2x+y=0, \\ 2y+x=0, \end{cases} \Leftrightarrow \begin{cases} x=0, \\ y=0. \end{cases}$$

Nature of the critical points : We have

$$f(x,y) - f(0,0) = (x + \frac{1}{2}y)^2 + \frac{3}{4}y^2 > 0; \ \ \forall (x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}.$$

Therefore, f has a single extremum at (0,0), which is a global minimum.

Example

Let $f(x,y) = x^2 + y^2 - 2x - 4y$. We have $f \in C^1(\mathbb{R}^2)$.

Critical points of f : We have

$$\nabla f(x,y) = \begin{pmatrix} 2x-2 \\ 2y-4 \end{pmatrix}$$

Thus,

$$((x,y)\text{is a critical point of } f) \Leftrightarrow \begin{cases} 2x-2=0, \\ 2y-4=0, \end{cases} \Leftrightarrow \begin{cases} x=1, \\ y=2. \end{cases}$$

Nature of the critical points : Making a change of variable $(x,y) = (h+1, k+2)$ to shift the origin, we get

$$\begin{aligned} f(x,y) - f(1,2) &= f((h+1,k+2)) - f(1,2) \\ &= (h+1)^2 + (k+2)^2 - 2(h+1) - 4(k+2) + 5 \\ &= h^2 + k^2 > 0. \end{aligned}$$

Therefore, f has a single extremum at (1,0), which is a global minimum.

Example

Let $f(x,y) = x^3 + y^2$. We have $f \in C^1(\mathbb{R}^2)$.

Critical points of f : We have

$$\nabla f(x,y) = \begin{pmatrix} 3x^2 \\ 2y \end{pmatrix}$$

Thus, $((x,y)$ is a critical point of $f) \Leftrightarrow \begin{cases} 3x^2 = 0, \\ 2y = 0, \end{cases} \Leftrightarrow \begin{cases} x = 0; \\ y = 0. \end{cases}$.

Nature of the critical points : We observe that

- $f(x,0) - f(0,0) = x^3 < 0$ if $x < 0$,
- $f(x,0) - f(0,0) = x^3 > 0$ if $x > 0$.

Therefore, $f(x,y) - f(0,0)$ changes sign in every neighborhood of $(0,0)$. We conclude that f does not even have a local extremum at $(0,0)$.

Study the Extremum by Using Second Partial Derivatives

We will generalize to multivariable functions what is well known for the extrema of real functions of one variable, namely, if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is of class C^2 in the neighborhood of a critical point p (i.e. $f'(p) = 0$), then :

- if $f''(p) > 0$, f has a local minimum at p,
- if $f''(p) < 0$, f has a local maximum at p,
- if $f''(p) = 0$, we cannot conclude, further calculations are needed (for example, the Taylor expansion of order > 2).

For multivariable functions, if f is C^2 in the neighborhood of a critical point p, we will be interested in a **"certain notion of positivity"** of the Hessian matrix.

Notions of "Positive Definite" and "Negative Definite" Matrices

Definition

A symmetric matrix $A \in M_m(\mathbb{R})$ is :

- *positive definite if $\forall X \in \mathbb{R}^m \setminus \{0\}$, ${}^tXAX > 0$,*
- *negative definite if $-A$ is positive definite (i.e., $\forall X \in \mathbb{R}^m \setminus \{0\}$, ${}^tXAX < 0$),*
- *indefinite if A is neither positive definite nor negative definite.*

The following rule is very useful to check if a matrix is "positive definite" :

Proposition (Sylvester's Criterion)

Let Δ_k be the k^{th} principal minor of order k. Then,

$$A \text{ matrix } A \text{ is positive definite} \quad \textbf{iff} \quad \forall k \in \{1, ..., n\},\ \det(\Delta_k) > 0.$$

Example

The matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ is positive definite.

Proposition

Let A be a symmetric matrix. Then,

- *A is positive definite* **iff** *all its eigenvaluesare strictly positive.*
- *A is negative definite* **iff** *all its eigenvaluesare strictly negative.*
- *A is invertible and undefined* **iff** *all eigenvalues of A are nonzero, and at least two of them have opposite signs.*

Example

Let $f(x,y) = x^2 + y^2 + 2y^3$. We have :

$$\nabla^2(f)(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2+12y \end{pmatrix}$$

In particular :

- $\nabla^2(f)(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ is a positive definite matrix.
- $\nabla^2(f)(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & -10 \end{pmatrix}$ is an invertible and indefinite matrix.
- $\nabla^2(f)(0,\frac{-1}{6}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ is a non-invertible matrix.

Fundamental Theorem

Theorem

Let D be an open subset of $\mathbb{R}^m$ and $f : D \subseteq \mathbb{R}^m \to \mathbb{R}$ be a C^2 function, and let p be a critical point of f (i.e., $\nabla f(p) = 0$). Then,

- *if $\nabla^2 f(p)$ is positive definite, then f has a strict local minimum at p,*
- *if $\nabla^2 f(p)$ is negative definite, then f has a strict local maximum at p,*
- *if $\nabla^2 f(p)$ is invertible and indefinite, then f has neither a local maximum nor a local minimum at p (i.e., p is a saddle point),*

Idea of the proof :

We use the second-order Taylor expansion of f in the neighborhood of p :
$f(u) = f(p) + \nabla f(p) \cdot (u-p) + \frac{1}{2}(u-p)^t \, \nabla^2 f(p) \, (u-p) + \|u-p\|^2 \varepsilon(u)$. $\square$

Remark

If $\nabla^2 f(p)$ is non-invertible, we cannot conclude. In this case, the critical point p is said to be degenerate (further analysis is required to determine its nature, i.e., to find the sign of $f(p+H) - f(p)$ for every H close enough to $0_{\mathbb{R}^m}$).

Examples

Example (Details given in class)

Let the function f be defined on $\mathbb{R}^3$ by

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

The only critical point of f is $p = (0, 0, 0)$. Using the Hessian of f, we deduce that f has a strict local minimum at $(0, 0, 0)$.

Example (Details given in class)

Let the function f be defined on $\mathbb{R}^3$ by

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

The function f has two critical points $(1, 1, 0)$ and $(-1, 1, 0)$. We deduce that :

- Since $Hess(f)(1, 1, 0)$ is positive definite, f has a strict local minimum at $(1, 1, 0)$,
- Since $Hess(f)(-1, 1, 0)$ is invertible and indefinite, f has neither a local maximum nor a local minimum at $(-1, 1, 0)$, which is a saddle point.

Particular case of functions with two variables

For a function of class C^2, let us define

- $c = \frac{\partial f}{\partial x}(\alpha, \beta)$, $d = \frac{\partial f}{\partial y}(p)$,
- $r = \frac{\partial^2 f}{\partial^2 x}(\alpha, \beta)$, $s = \frac{\partial^2 f}{\partial x \partial y}(\alpha, \beta)$, $t = \frac{\partial^2 f}{\partial^2 y}(\alpha, \beta)$

From the fundamental theorem on the Hessian and extrema, the following result is deduced

Corollary (Monge's Theorem (Discriminants))

Let D be an open set of $\mathbb{R}^2$, $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be an application of class C^2, and let $p = (\alpha, \beta) \in D$. Then, (α, β) is a critical point if and only if $c = d = 0$. Moreover,

- *if $rt - s^2 > 0$ and*
  - *$r > 0$ then f has a strict local minimum at p,*
  - *$r < 0$ then f has a strict local maximum at p,*
- *if $rt - s^2 < 0$, then f does not have an extremum at p (i.e., p is a saddle point),*
- *if $rt - s^2 = 0$, the critical point p is said to be degenerate, and no conclusion can be drawn.*

Examples

Example

Find and determine the nature of the critical points of the function $f(x, y) = 2x^3 - 6xy + 3y^2$.

Solution : We have $f \in \mathcal{C}^{\infty}\left(\mathbb{R}^2\right)$.

Critical points of f : A point $p = (x, y)$ is a critical point of f if and only if

$$\nabla f(p) = 0_{\mathbb{R}^2} \Leftrightarrow \left\{ \begin{array}{c} 6x^2 - 6y = 0, \\ 6y - 6x = 0 \end{array} \right. \Leftrightarrow p \in \{(0, 0), (1, 1)\}.$$

Nature of the critical points : We have

$$r = \frac{\partial^2 f}{\partial x^2}(x, y) = 12x, \quad t = \frac{\partial^2 f}{\partial y^2}(x, y) = 6, \quad \frac{\partial f}{\partial x \partial y}(x, y) = -6.$$

Thus,

p	r	t	s	$\Delta = rt - s^2$	**Conclusion**
$(0,0)$	0	6	-6	$-36 < 0$	No extremum at (0,0) ((0,0) is a saddle point)
$(1,1)$	$12 > 0$	6	-6	$6 \times 12 - 6^2 > 0$	(1,1) is a strict local minim

Additionally, we have

$$f(x,0) = 2x^3 \to -\infty \text{ as } x \to -\infty.$$

Thus, we conclude that f does not have a global minimum.

Example

Let's study the extrema of $f(x,y,z) = x^4 + x^2 + 4y^2 + 3z^2$ using Monge's theorem.

Solution : We have $f \in \mathcal{C}^\infty(\mathbb{R}^3)$ and

$$\nabla f(x,y,z) = \begin{pmatrix} 4x^3 + 4x \\ 4y \\ 6z \end{pmatrix} \Rightarrow \nabla^2 f(x,y,z) = \begin{pmatrix} 12x^2 + 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

Critical points of f : p is a critical point of f if and only if

$$\nabla f(p) = 0_{\mathbb{R}^2} \Leftrightarrow \begin{cases} 4x^3 + 4x = 0, \\ 4y = 0, \\ 6z = 0, \end{cases} \Leftrightarrow \begin{cases} 4x(x^2+1) = 0, \\ 4y = 0, \\ 6z = 0, \end{cases} \Rightarrow p = (0,0).$$

Nature of the critical points : The matrix $\nabla^2 f(x,y,z)$ is symmetric and positive definite (since $\det(\Delta_1) = 2 > 0$, $\det(\Delta_2) = 8 > 0$ *and* ($\det(\Delta_{)} = 48 > 0$). Thus, we deduce that f has a single local extremum, which is a local minimum at $p = (0,0,0)$.

Example

Consider the dataset

(x_i, y_i)	$(0,1)$	$(1,0)$	$(2,1)$	$(3,1)$

We are looking for a line $(\triangle)$ whose equation $y = \alpha x + \beta$ minimizes the sum of the squared distances between the points in the dataset and their projections onto the y-axis parallel to $(\triangle)$.
1) Find the function $f(\alpha, \beta)$ that expresses the sum of the squared distances between the points in the dataset and their projections onto $(\triangle)$.
2) Find the optimal line $(\triangle)$.
3) Represent (graphically) the points in the dataset.

Solution : 1) Let (x_i, y_i) be a point in the dataset. Its projection onto the y-axis parallel to $(\triangle)$ is the point $(x_i, \alpha x_i + \beta)$, so the distance between the point and its projection is $|\alpha x_i + \beta - y_i|$. Therefore,

$$\begin{aligned} f(\alpha, \beta) &= \sum_{i=1}^{4} |\alpha x_i + \beta - y_i|^2 \\ &= (\beta - 1)^2 + (\alpha + \beta)^2 + (2\alpha + \beta - 1)^2 + (3\alpha + \beta - 1)^2 \\ &= 14\alpha^2 + 12\alpha\beta - 10\alpha + 4\beta^2 - 6\beta + 3. \end{aligned}$$

2) a) Calculation of ∇f and $\nabla^2 f$: We have $f \in \mathcal{C}^{\infty}(\mathbb{R}^2)$ (the variables here are α and β) and

$$\nabla f(\alpha,\beta) = \begin{pmatrix} 28\alpha + 12\beta - 10 \\ 12\alpha + 8\beta - 6 \end{pmatrix}, \quad \nabla^2 f(\alpha,\beta) = \begin{pmatrix} 28 & 12 \\ 12 & 8 \end{pmatrix}.$$

b) Critical points of f :

$$p = (\alpha,\beta) \text{ is a critical point of } f \Leftrightarrow \nabla f(\alpha,\beta) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{cases} 28\alpha + 12\beta - 10 = 0, \\ 12\alpha + 8\beta - 6 = 0, \end{cases}$$

c) Nature of the critical point : We have

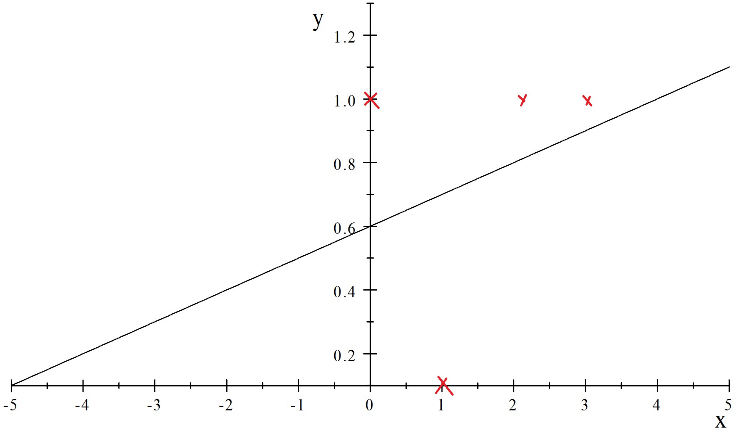
$$\nabla^2 f\left(\frac{1}{10}, \frac{3}{5}\right) = \begin{pmatrix} 28 & 12 \\ 12 & 8 \end{pmatrix},$$

which is symmetric and positive definite (since $\det(\Delta_1) = 28 > 0$ and $\det(\Delta_1) = 80 > 0$). Thus, we deduce that f has a single local extremum, which is a local minimum at $p = \left(\frac{1}{10}, \frac{3}{5}\right)$. Moreover, it is the global minimum because f is strictly convex ($\forall X \in \mathbb{R}^2; \nabla^2 f(X)$ is positive definite).

d) Equation of the line (Δ) : We deduce that

$$(\Delta) : y = \frac{1}{10}x + \frac{3}{5}.$$

3) Graphical representation :



Points of the data set X . The line (Δ)— black.

Remark on Degenerate Critical Points

To try to determine the nature of a degenerate critical point p, one can use other techniques such as going back to the definition, OR using the Taylor formula of order ≥ 2, that is, to determine the sign of $f(p+h) - f(p)$ for $h = (h_1, h_2)$ close to $0_{\mathbb{R}^2}$.

Example

Study the local extrema of $f(x, y) = x^4 + y^4 - 2(x-y)^2$ on $\mathbb{R}^2$.

Solution : The critical points are $(0,0), (\sqrt{2}, -\sqrt{2}), (-\sqrt{2}, \sqrt{2})$. Using the Monge's theorem, we get :

- At $(\sqrt{2}, -\sqrt{2})$ and $(\sqrt{2}, \sqrt{2})$, the function f has a local minimum,
- At $(0,0)$, we cannot conclude.

Moreover, if we set $p = (0,0)$ and $h = (h_1, h_2)$, we get

$$f(p+h) - f(p) = h_1^4 + h_2^4 - 2(h_1 - h_2)^2$$

Thus,
- For $h = (h_1, h_1)$, we get $f(p+h) - f(p) = h_1^4 + h_1^4 > 0$, for all $h_1 \neq 0$.
- For $h = (h_1, 0)$, $f(p+h) - f(p) = h_1^4 - 2h_1 = h_1^2(h_1^2 - 2) < 0$, for h_1 sufficiently close to 0.

Therefore, the function f does not have a local extremum at $(0,0)$.

Summary on Unconstrained Optimization and the Hessian in 2 and 3 Dimensions

Let p be a critical point of a function $f : \mathbb{R}^m \to \mathbb{R}$ of class $\mathcal{C}^2$.
Let $d_k = \det \Delta_k$, where Δ_k is the principal minor of $\nabla^2 f(p)$.
We have (excluding the case of degenerate points or saddle points) :

m	sign of d_j	Nature of the point p
1	$d_j > 0$	local minimum
1	$d_j < 0$	local maximum
2	$d_2 > 0,\ d_1 > 0$	local minimum
2	$d_2 > 0,\ d_1 < 0$	local maximum
3	$d_3 > 0,\ d_2 > 0,\ d_1 > 0$	local minimum
3	$d_3 < 0,\ d_2 > 0,\ d_1 < 0$	local maximum

Constrained Optimization

Equality-Constrained Optimization Problem Statement

In this section, we focus on constrained optimization where the objective function is optimized subject to one or more equality constraints.

Problem statement

Determine the local extrema of a function f in a set

$$A \doteq \{x \in D \text{ such that } g_1(x) = 0, ..., g_k(x) = 0\},$$

where D is an open set of $\mathbb{R}^m$, and f and $g_1, \ldots g_k : D \to \mathbb{R}$ are defined on D.

Definition

We say that a function f has a maximum (resp. minimum) at $p \in A$ subject to the constraints $g_1(x) = 0, ..., g_k(x) = 0$, if the restriction $f_{|A}$ has a maximum (resp. minimum) at the point p,
that is to say :

- $g_1(p) = \cdots = g_k(p) = 0,$
- $f(p) = \max\limits_{x \in D,\ g_1(x) = \cdots = g_k(x) = 0} f(x) \left(\textit{resp.} \quad f(p) = \min\limits_{x \in D,\ g_1(x) = \cdots = g_k(x) = 0} f(x) \right).$

Necessary optimality condition for equality constraints (first order)

Theorem (Lagrange multipliers)

Let D be an open set of $\mathbb{R}^m$, and let $p \in D$. Let f and $g_1, \dots g_k : \mathbb{R}^m \to \mathbb{R}$ be functions defined on D. Suppose that :

- *f and $g_1, \dots g_k$ are of class C^1 on D,*
- *$\{\nabla g_1(p), .., \nabla g_k(p)\}$ is linearly independent.*

If f p is a constrained extremum at p subject to the constraints $g_1(x) = 0, \cdots, g_k(x) = 0$, then

$$\exists \lambda_1,\ \lambda_k \in \mathbb{R} \ \text{ such that } \ \nabla f(p) + \lambda_1 \nabla g_1(p) + .. + \lambda_k \nabla g_k(p) = 0.$$

The real numbers $\lambda_1, \dots, \lambda_k$ are called the Lagrange multipliers of f at p.

Optimization with constraints viewed as unconstrained optimization !

Remark

*The Lagrange theorem can be seen as the necessary condition for the existence of unconstrained extrema (critical points) written for the function L (***called the Lagrangian***) defined from* $D \times \mathbb{R}^k \to \mathbb{R}$ *by*

$$L(x, \lambda_1, \cdots, \lambda_k) = f(x) + \lambda_1 g_1(X) + \cdots + \lambda_k g_k(X).$$

Thus, we have

$$\nabla L(x, \lambda_1, \cdots, \lambda_k) = \begin{pmatrix} \nabla f(X) + \lambda_1 \nabla g_1(X) + \cdots + \lambda_k \nabla g_k(X) \\ g_1(X) \\ \vdots \\ g_k(X) \end{pmatrix}$$

The Lagrange multiplier theorem translates as : **Suppose that** $\{\nabla g_1(p), .., \nabla g_k(p)\}$ *is linearly independent.* **If** f *has a constrained extremum at* p *under the constraints* $g_1(x) = \cdots = g_k(x) = 0$, **then,**

$$\exists \lambda_1,\ \lambda_k \in \mathbb{R} \text{ such that } \nabla L(p, \lambda_1, \cdots, \lambda_k)) = 0_{\mathbb{R}^{m+k}}.$$

Sufficient optimality condition for equality constraints (second order)

Consider the optimization problem of a function f of class $\mathcal{C}^2$ on $\mathbb{R}^m$ under the constraints $g_1(x) = \cdots = g_s(x) = 0$ where the g_k are of class $\mathcal{C}^1$ on $\mathbb{R}^m$.
Define $\mathcal{L}(x, \lambda) = f(x) - \lambda_1 g_1(x) - \cdots - \lambda_s g_s(x)$.

Definition

The bordered Hessian matrix of the Lagrangian (denoted by $\overline{H}_{\mathcal{L}}$) is the matrix of second partial derivatives of $\mathcal{L}$ with respect to x_i bordered by the first partial derivatives of the constraint functions $g_1, \cdots, g_s$:

$$\overline{H}_{\mathcal{L}} = \left(\begin{array}{ccc|ccc} 0 & \cdots & 0 & \frac{\partial g_1}{\partial x_1} & \cdots\cdots & \frac{\partial g_1}{\partial x_m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \frac{\partial g_s}{\partial x_1} & \cdots\cdots & \frac{\partial g_s}{\partial x_m} \\ \hline \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_s}{\partial x_1} & \frac{\partial^2 \mathcal{L}}{\partial x_1^2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_m} \\ & \vdots & & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_m} & \cdots & \frac{\partial g_s}{\partial x_m} & \frac{\partial^2 \mathcal{L}}{\partial x_1 \partial x_m} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial x_m^2} \end{array}\right)$$

We have the following restlt :

Proposition

Let (p, λ) be obtained by the Lagrange multipliers theorem and Δ_j be the principal minor of order j of $\overline{H}_{\mathcal{L}}(p, \lambda)$. We have

- *if the last $m - s$ determinants $\det(\Delta_j)$ have alternating signs and the sign of the last one is that of $(-1)^m$, then p is a strict local maximum point for the considered optimization problem,*
- *if the last $m - s$ determinants $\det(\Delta_j)$ all have the sign of $(-1)^s$, then p is a strict local minimum point for the considered optimization problem.*

Summary on Equality-Constrained Optimization and the Bordered Hessian for m= 2 and m=3

Let $0 \leq j \leq s$. and $f,\ g_j : \mathbb{R}^m \to \mathbb{R}$ be a functions of class $\mathcal{C}^2$ on an open set D of $\mathbb{R}^m$. Suppose that the conditions of the Lagrange theorem are satisfied and that $(p, \lambda_1, \cdots, \lambda_s)$ is a critical point of the Lagrangian $\mathcal{L}$.

Recall that :

- $d_k = \det \Delta_k$, where Δ_k is the principal minor of $\overline{H}_{\mathcal{L}}(p, \lambda_1, \cdots, \lambda_s)$,
- m is the dimension of the space,
- s is the number of equality constraints.

We have (excluding the case of degenerate points (i.e., when one of the last $m - s$ determinants $\det \Delta_k$ is zero) or saddle points (i.e., when all of the last $m - s$ determinants $\det \Delta_k$ are nonzero but do not satisfy the proposition on the bordered Hessian and extrema with equality constraints)) :

m	s	Sign of d_j	Nature of the point p
2	1	$d_3 > 0$	local maximum
2	1	$d_3 < 0$	local minimum
3	1	$d_4 < 0,\ d_3 < 0$	local minimum
3	1	$d_4 < 0,\ d_3 > 0$	local maximum
3	2	$d_4 > 0$	local minimum
3	2	$d_4 < 0$	local maximum

Steps for studying the extrema of a differentiable function under equality constraints

To study the extremasubject to an equality constraints, we can follow the following steps :

Method 1 : By reduction : If, based on the constraints, we can express some variables as functions of the others, we reduce the dimension. For example, if we optimize a function of two variables under a single constraint :

- If we can write for $(x, y) \in A$, $y = h(x)$ (or $x = h(y)$) where h is a sufficiently regular function, then the constrained problem reduces to studying the extrema of a one-variable function $f(x, h(x))$ (or $f(h(y), y)$),
- If we can write for $(x, y) \in A$, $x = h(t)$ and $y = k(t)$ where h and k are sufficiently regular functions, then the constrained problem reduces to studying the extrema of a one-variable function $f(h(t), k(t))$).

Method 2 : Using Lagrange multipliers (if the linear independence condition related to the constraints is satisfied)

- Determination of the critical points of $\mathcal{L}$,

- Study of the nature of these critical points using either the bordered Hessian $\overline{H}_{\mathcal{L}}$ or by definition (studying the sign of $f(h+p) - f(p)$ with h very close to $0_{\mathbb{R}^m}$ and $p+h$ satisfying all the constraints).

Examples

Example

Study the extrema of the function $f(x, y) = xy$ subject to the constraint $x + y = 6$ using the reduction method.

Solution : We observe that if a point $(x, y) \in \mathbb{R}^2$ satisfies the constraint $x + y = 6$, then $y = 6 - x$. By substitution, the study of the constrained extrema of f under the constraint $x + y = 6$ reduces to the study of the extrema over $\mathbb{R}$ of the function :

$$\Psi(x) \doteq f(x, 6 - x) = 6x - x^2.$$

Study of the extrema of Ψ : We have $\Psi \in C^\infty(\mathbb{R})$ and $\Psi'(x) = 6 - 2x$, as well as $\Psi''(x) = -2$. Thus, if Ψ admits an extremum at x, then $\Psi'(x) = 6 - 2x = 0$, i.e., $x = 3$. Since $\Psi''(x) < 0$, this implies that Ψ has a local maximum at $x = 3$.

Conclusion : We deduce that f has a maximum at $(3, h(3)) = (3, 3)$ under the constraint $x + y = 6$, and $f(3, 3) = 9$.

Example

Study the extrema of the function $f(x, y) = xy$ subject to the constraint $x + y = 6$ using the Lagrangian.

In this case, we are in dimension $m = 2$ and there is $s = 1$ equality constraint. Let $g(x, y) = x + y - 6$ and

$$\mathcal{L}(x, y, \lambda) \doteq f(x, y) - \lambda g(x, y) = xy + \lambda(x + y - 6)$$

Verification of the independence condition : We have $\forall (x, y) \in \mathbb{R}^2$ with $g(x, y) = 0$: $\nabla g\,(x, y) = \ ^t(1, 1)) \neq (0, 0)$ so the family $\{\nabla g((x, y)\}$ is independent.

Search for the critical points of the Lagrangian : If f has an extremum at $p = (x, y)$ subject to the constraint $g(x, y) = 0$, then there exists $\lambda \in \mathbb{R}$ such that : $\nabla \mathcal{L}(x, y, \lambda) = (0, 0, 0)$. We solve the system

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x}(x, y) = 0, \\ \dfrac{\partial \mathcal{L}}{\partial y}(x, y) = 0, \\ g(x, y) = 0, \end{cases} \Leftrightarrow \begin{cases} y + \lambda = 0, \\ x + \lambda = 0, \\ x + y - 6 = 0, \end{cases} \Leftrightarrow \begin{cases} x = 3; \\ y = 3, \\ \lambda = -3. \end{cases}$$

To determine the nature of the only critical point, we have two methods

▷ Determination of the nature of the critical points by the test :

Let $(h_1, h_2) \in \mathbb{R}^2$ with $g(3+h_1, 3+h_2) = 0$. We have

$$f(3+h_1, 3+h_2) - f(3,3) = (3+h_1)(3+h_2) - 9 = 9h_1h_2 + 3h_1 + 3h_2.$$

Since, $g(3+h_1, 3+h_2) = 0 \Leftrightarrow h_1 + h_2 = 0 \Leftrightarrow h_2 = -h_1$, thus

$$f(3+h_1, 3+h_2) - f(3,3) = -h_1^2 \leq 0.$$

Therefore, $(3,3)$ is a maximum point of f under the constraint $x+y=6$ and $f(3,3)=0$.

▷ Determination of the nature of the critical points by the bordered Hessian

Let's calculate $\overline{H}_{\mathcal{L}}$. We have

$$\overline{H}_{\mathcal{L}}(x,y,\lambda) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \Rightarrow \overline{H}_{\mathcal{L}}(3,3,-3) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Here $m=2$ and $s=1$ and $d_3 = 2 >$, so we deduce that $(3,3)$ is a maximum point of f under the constraint $x+y=6$ and $f(3,3)=0$.

Remark on Extrema with Inequality Constraints

The following result is very important (to determine whether a function has extrema, although it does not provide a practical method to compute them !).

Proposition

If f is a continuous function on a closed and bounded (compact) subset A of $\mathbb{R}^m$, then it has both a maximum and a minimum on A.

To find the extrema on a closed set A, knowing that $A = Int(A) \cup Fr(A)$, we can proceed as follows :

- Find the extrema of f in $Int(A)$,
- Find the extrema of f on $Fr(A)$,
- Compare the values of f at these points to determine the extrema of f on A.

Remark

There is a result providing a necessary condition for optimality in the case of inequality constraints (due to Karush-Kuhn-Tucker).

Example

Study the extrema of $f(x,y) = x^2 + xy + y^2$ on

$$A = \{(x,y) \in \mathbb{R}^2 \text{ such that } x^2 + y^2 \leq 1\}.$$

Solution : Since f is continuous on A, which is a closed and bounded subset of $\mathbb{R}^m$, it follows that f has both a maximum and a minimum on A. Moreover, we have $A = Int(A) \cup Fr(A)$ with

- $Int(A) = \{(x,y) \in \mathbb{R}^2 \text{ such that } x^2 + y^2 < 1\}$,
- $Fr(A) = \{(x,y) \in \mathbb{R}^2 \text{ such that } x^2 + y^2 = 1\}$.

Finding the Maximum and Minimum of f on A

- In $Int(A)$, f has $(0,0)$ as its minimum.
- On $Fr(A)$, f has extrema at $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, $(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$, and $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

Finally, we compare the values of $f(0,0)$ and

$$f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$$

to determine the extrema.

Appendix : Convexity Analysis and Optimization

Existence of an Optimum

Definition

A function $f : \mathbb{R}^m \to \mathbb{R}$ is said to be coercive on **an unbounded subset** *$C \subset D_f$ of $\mathbb{R}^m$ if*

$$\lim_{\substack{x \in C, \\ \|x\| \to +\infty}} f(x) = +\infty.$$

Proposition

Let C be a nonempty closed subset of $\mathbb{R}^m$ and let $f : \mathbb{R}^m \to \mathbb{R}$ be continuous on C. Then, we have :

- *If C is bounded, then f attains both a minimum and a maximum on C.*
- *If C is unbounded and f is coercive on C, then f attains at least a minimum on C.*

Convex Sets and Functions

Convex Sets : A set $C \subset \mathbb{R}^m$ is said to be convex if for all $\lambda \in [0, 1]$ and for all $x, y \in C$, we have $\lambda x + (1 - \lambda)y \in C$.

Geometrically :

- For $x, y \in \mathbb{R}^m$, the set $\{\lambda x + (1 - \lambda)y$ such that $\lambda \in [0, 1]\} \subset \mathbb{R}^m$ is called the line segment joining the two points x and y and is denoted by $[x, y]$.
- The convexity of C meaCns that the line segment connecting any two points x and y of C remains entirely within C.

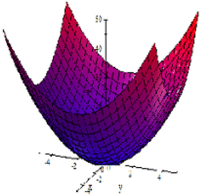
A conxex set A non-convex set A non-convex set

Convex Functions : Let $f : \mathbb{R}^m \to \mathbb{R}$ and let C be a convex set (with $C \subset D_f$). We say that

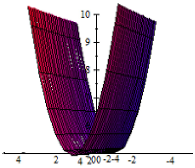
- f is convex on C if for every pair of points $x, y \in C$ and every $\lambda \in [0, 1]$, we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$.
- f is strictly convex on C if for every pair of distinct points $x, y \in C$ and every $\lambda \in]0, 1[$, we have $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$.

Remark : Strict convexity implies convexity.

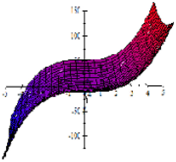
Geometrically : The convexity of f means that the line segment connecting $(x, f(x))$ and $(y, f(y))$ lies above (with respect to the axis representing the values of f) the portion of the graph of f corresponding to $[x, y]$.



A strictly convex function



A convex function



A non-convex function

Convexity and the Hessian

Proposition (Second-Order Characterization of the Convexity of a Function)

Let $f : \mathbb{R}^m \to \mathbb{R}$ *be a* **twice differentiable** *function on a* **convex open** *set* $C \subset D_f$. *Then,*

- $\big($*the Hessian matrix* $\nabla^2 f(x)$ *is positive semidefinite for all* $x \in C\big) \Longleftrightarrow$ $(f$ *is convex on* $C)$;
- $\big($*the Hessian matrix* $\nabla^2 f(x)$ *is positive definite for all* $x \in C\big) \Rightarrow$ $(f$ *is strictly convex on* $C)$.

Corollary

Let $f : \mathbb{R}^m \to \mathbb{R}$ *be a* **twice differentiable** *function on a* **convex open** *set* C. *Then,*

- $\left(sp\left(\nabla^2 f(x)\right) \subset \mathbb{R}_+,\ \forall x \in C\right) \Longleftrightarrow (f$ *is convex on* $C)$;
- $\left(sp\left(\nabla^2 f(x)\right) \subset \mathbb{R}_+^*,\ \forall x \in C\right) \Rightarrow (f$ *is strictly convex on* $C)$.

Convexity and Minimization

Proposition

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function on a convex set C. Then,

- *(x^* is a local minimum on C) $\Longleftrightarrow$ (x^* is a global minimum on C);*
- *if f is of class $\mathcal{C}^1$ on C and $x^* \in int(C)$, then*

$$(x^* \text{ is a global minimum on } C) \Longleftrightarrow (\nabla f(x^*) = 0).$$

Proposition

Let $f : \mathbb{R}^m \to \mathbb{R}$ be a strictly convex function on a convex set C. Then, f admits at most one minimum point on C.